Best Prices

Author(s): Judith A. Chevalier and Anil K Kashyap

Best Prices: Price Discrimination and Consumer Substitution[†]

By Judith A. Chevalier and Anil K Kashyap*

This paper proposes a method for aggregating prices when retailers use periodic sales to price-discriminate amongst heterogeneous customers. In the motivating model, loyal customers buy one brand and do not strategically time purchases, while Bargain Hunters always pay the lowest price available, the "best price." In the model, the best price is part of an exact price index. Accounting for the best price also substantially improves the empirical match between conventional price aggregation strategies and actual prices paid by consumers. The methodology improves inflation measurement while imposing little burden on the data-collection agency. (JEL C43, D12, E31, L13, L81, M31)

Official price indexes provide a crucial input to both public and private decision-making. Unfortunately, the modern retail environment creates a challenge for defining and computing cost of living metrics due to existence of a large number of substitute products offered at rapidly changing prices. Retail firms are continually innovating in their attempts to exploit consumer data in order to improve promotion strategies and facilitate price discrimination between different groups of consumers.[1] In this paper, we focus on a particular price discrimination strategy—temporary discounts. Temporary discounts create a challenge for price measurement when goods are close substitutes. When retailers stock a large number of varieties and strategically set prices for price discrimination purposes, the

relative prices of close substitutes can be quite volatile. Thus, the price aggregation methodology will have important implications for price measurement.

Most cost-of-living discussions take an exact index as a benchmark. An exact index tracks the relative cost of obtaining a given level of utility at different points in time. One challenge to constructing an exact index is consumer heterogeneity. Price discrimination occurs when there are different types of consumers to discriminate among. Thus, the price of maintaining a given level of utility over a specific period of time is not well-specified without aggregating utilities across different consumers.

Further, if goods are imperfect substitutes and the relative prices of the goods are volatile due to discounting behavior, aggregation methodologies will be very sensitive to the weight of each good in the price index. Finally, if consumers respond to price discounts by stockpiling goods, the appropriate period of time over which to measure prices and construct a price index is not obvious. An important conceptual question, then, is how to weight discounted prices versus regular prices in constructing price aggregates both across products and through time.

In order to examine these issues, we present a very simple model of sales motivated by price discrimination. Our model of sales is similar in spirit to Varian (1980); Salop and Stiglitz (1982); Sobel (1984); and Pesendorfer (2002). In particular, in our model, we posit that some consumers are active shoppers who chase discounts, substitute across products in a narrowly defined product category, and potentially use storage to maintain smooth consumption whilst concentrating their purchases to take advantage of discounts. Other customers are passive and brand loyal, and retailers will employ strategies to charge these two groups different prices. The model predicts frequent temporary discounts and long periods of constant regular prices, as observed empirically in many retail settings.

Our model implies that changes in unit values (the quantity-weighted average price of goods purchased), measured over time, are, or approach, the exact index. The intuition behind this result is simple. The retailer's second degree price discrimination motive induces the retailer to set prices so that only consumers with very weak brand preferences switch among products to save money, while consumers with strong brand preferences buy their most preferred goods. Thus, substitution between brands, when it is actually observed, is not associated with substantial utility consequences. The strategic behavior of retailers with some second degree price discrimination ability renders changes in unit values closer to an exact index than it would be absent strategic retailer behavior.

We introduce the concept of a "best price," defined as the lowest price charged for any good in the narrow product category during a short multi-week time window. The model predicts that best prices should be the relevant prices for discount-chasing consumers. As we describe below, while these specific predictions are particular to our model, the general ideas are quite durable. This leads to our price measurement proposal.

While changes in unit values would appropriately track the cost of achieving a given level of utility in our model, a real-time variable weight series is impractical for a statistical agency to construct. We demonstrate that our structural model implies a simple methodology to approximate unit values without the need for

high-frequency, quantity data. In our model, the level of the unit value can be approximated by a weighted average of a fixed weight price aggregate and the best price.

We examine the empirical validity of our model using detailed store-level data. We use a national dataset of supermarket prices for 2001 to 2011 provided by IRI. The IRI dataset covers stores in 50 markets around the country. Prices for individual products at many of the IRI stores display the now-familiar pattern of very infrequent regular price changes combined with frequent temporary discounts. We show that first, consistent with our model, a disproportionate fraction of goods are sold at temporary discounts. Second, unit values are well-approximated by a linear combination of a conventional fixed weight price aggregate and the best available price within the group of close substitutes. At this lowest level of aggregation, the US Bureau of Labor Statistics (BLS) samples one price per outlet and does not attempt to address within-outlet substitution at all. The degree of substitution embedded in existing BLS-style price aggregates will not adequately account for migration of customers to the "best price." Finally, we demonstrate that inflation for the prices paid by consumers between 2001 and 2011 differs systematically from the inflation rate that is implied by a BLS-type methodology for estimating inflation. Importantly, we find that inflation in unit values (the prices actually paid by consumers) is well-approximated by a combination of inflation as measured by a fixed-weight index and best-price inflation. We demonstrate that this is not captured by current US price aggregation methodologies.

As noted above, US statistical agencies cannot make, or at least have not made, use of high-frequency, quantity data. This makes direct construction of any index relying on contemporaneous quantity data (such as the Tornqvist or unit value index) impossible. In theory, the adoption of a model-driven weighting scheme that well-approximated substitution patterns would lead to a price index that more closely approximates an exact index. Our paper proposes such a weighting scheme to aggregate prices in a way that accounts for the specific (and important) issue of what to do about temporary discounts. Moreover, the associated incremental data collection requirements are modest.

Our paper proceeds as follows. Section I briefly reviews some of the relevant literature. Section II presents our simple model of a price-discriminating retailer facing two types of customers. We derive testable empirical implications of the model. Section III describes the data. Section IV establishes a number of new facts about pricing and purchase patterns that are consistent with the model. Section V concludes.

I. Literature Review and Institutional Setting

During the 1990s, statistical agencies in a number of countries, including the Bureau of Labor Statistics (BLS) in the United States, adopted the geometric mean formula in the calculation of the most basic components of the Consumer Price Index (CPI). The geometric mean formula is a constant elasticity of substitution aggregate that implies an elasticity of substitution of one between varieties. The change in the geometric mean of prices over time (the Jevons index) is a closer approximation to

a true cost of living index than the Laspeyres fixed-weight index. One undesirable feature of the Laspeyres index is that the Laspeyres index does not reflect any substitution in response to relative price changes. A very comprehensive history of BLS practices and some critiques that have been brought up in periodic reviews of BLS practices can be found in Reinsdorf and Triplett (2009). Mismeasurement issues (current to more recent CPI changes) are addressed comprehensively in Shapiro and Wilcox (1997).

There are important criticisms of the geometric mean formulation for our purposes. As mentioned above, the geometric mean formulation implies a cross-price elasticity of substitution of one across varieties. In practice, the BLS collects prices using an enumerator method that collects one price in the lowest item stratum per store (e.g., one price for "peanut butter" per store, not prices for several substitute peanut butters within a store). Both our own evidence and past evidence suggest that cross-price elasticities of substitution between different brands within an outlet are often much greater than one in absolute value. The sample of prices obtained across outlets may bear a relationship to the distribution of substitute variety prices within an outlet, but the empirical relationship between the assumptions implicit in the price collection methodology and actual substitution patterns is unknown. Handbury, Watanabe and Weinstein (2013) demonstrate that, in the Japanese context, at the lowest level of item aggregation, substitution patterns differ from those implied by the geometric mean formula, leading to systematic bias in price indices. Feenstra and Shapiro (2003) demonstrate that substitution is induced not only by price, but from promotion activity. They found, in a supermarket context, that consumers substitute heavily to products that are being promoted even when the promoted prices are not the lowest.

Such observations have led to many suggested alternatives to the BLS approach. Caves, Christensen, and Diewert (1982) show, for a representative consumer with a fairly general form of homothetic utility, that a Tornqvist index approximates the exact cost of living index. The Tornqvist index combines the current period expenditure weights with a base period expenditure weights. Feenstra and Shapiro (2003) furthermore propose calculating the Tornqvist over a series of periods to account for the tendency of shoppers to store goods.

Nakamura et al. (2015); Diewert (1995); and Reinsdorf (1999, 2003) endorse the approach suggested by Walsh (1921) and Davies (1924, 1932) who advocate the use of unit values (the current period quantity-weighted average of prices) to aggregate prices of very similar products (such as different brands of a consumer product). Changes in unit values over time correspond to an exact cost of living index only when all consumers have identical preferences and experience no disutility in substituting between alternative brands, sizes, and varieties. Nakamura (1999) demonstrates assumptions under which the unit value representation would, even with heterogeneous consumers, more closely reflect variations in consumer utility than would the BLS measurement approach.

Importantly, the approaches advocated by Diewert (1995) and Reinsdorf (1999, 2003) arose in part out of enthusiasm for the possibility that statistical agencies would immediately begin to exploit the same scanner datasets that retailers and manufacturers were using. The full adoption of scanner methodologies has been

slow; Australia uses scanner data for some parts of its CPI construction and according to Chessa, Verburg, and Willenborg (2017), as of 2017, there were seven European countries that have incorporated scanner data into their CPI procedures. Other statistical agencies such as the BLS have chosen not to rely on commercial provision of high-frequency, quantity data. The impediments to employing commercial scanner data, some of which are detailed in Bradley et al. (1997); Reinsdorf (1999); Triplett (2003); and Chessa, Verburg, and Willenborg (2017) have not yet been fully resolved. The BLS continues to primarily obtain retail data by having BLS agents visit stores to sample prices. Thus, while the suggestion of using time-varying product weights at the lowest aggregation level (such as in constructing unit values or the Tornqvist) is simple in theory, implementing it in practice requires information on quantities and expenditure amounts. This is a problem for brick and mortar retail categories, where scanner data are not used, as well as for online retail categories, where frequent scraped prices are available but where quantity data are not available. The fact that our price measurement proposal does not rely on high-frequency, quantity data is an important advantage of our approach.

A second important consideration that is brought to the fore by price-discriminating retailer behavior is the issue of consumer heterogeneity. The extant utility maximization models that provide foundation for various cost of living constructions rely on representative consumer models. Neither the Tornqvist nor the unit value construction necessarily approaches an exact index with heterogeneous consumers. As Fisher and Griliches (1995) point out, index construction requires data on individual willingnesses to pay that cannot be extracted from aggregate data without resorting to strong distributional assumptions. Nakamura (1999) uses a two-type consumer model to demonstrate the poor performance of the fixed-weight index relative to usage of unit values in measuring real output.

Redding and Weinstein (2016) demonstrate that essentially all standard price indices can be derived as a special case of their unified price index. Moreover, their model is able to accommodate consumer heterogeneity of some forms. For instance, if the different types of agents in the economy have CES demand functions that differ only in the elasticity of demand, their unified index remains valid. In our model, however, the individual consumers do not have CES preferences and instead care about the ordinal level of prices. So, our setup cannot be mapped into their framework. Nevertheless, under certain assumptions in our environment, we show that an exact index can be approximated in the presence of two types.

While our price proposal is a substantial departure from BLS practices in categories such as the supermarket categories we study, our proposal is quite similar to current BLS practices in some other parts of the CPI. To guide price sampling of air fares, the propensity of airline ticket buyers to purchase the lowest available coach fare on a particular day from a particular airline is estimated by the BLS using Department of Transportation data on tickets. These data are not obtained in real time but are used in picking the weights on various fare categories. Informed by the DOT data, price indices are constructed under the assumption that roughly half of the discount coach tickets will be purchased at the lowest available fare. A similar type of reasoning is used by the BLS in situations where an online retailer

features multiple sellers (such as Amazon Marketplace). There, purchases are assumed to occur at the lowest price for a given item by any seller. Thus, the BLS has implemented practices similar to that which we propose for identical products sold at different prices; our results suggest that this method is useful in measuring prices for close substitute products. A methodology such as the one we propose, then, could be helpful in a variety of e-commerce settings in which price data are readily obtainable but quantity data are more difficult to obtain.[2]

Both our proposal and these current BLS practices save BLS labor relative to plausible alternative collection methodologies that potentially capture within-strata substitution. Specifically, one could imagine an alternative collection strategy in which a within-outlet within-strata cross-price elasticities of demand are estimated, and prices are collected to form a CES price index. This would necessitate collecting the prices of all relevant close substitutes in the store. Instead, our methodology, like the BLS methodology for airline prices, offers a less costly shortcut by requiring only the lowest ordinal price in the category along with whichever one would normally be randomly sampled.

II. A Model of Price Discriminating Retailers and Heterogeneous Consumers

We present a simple model that is similar in spirit to Varian (1980), Salop and Stiglitz (1982), Sobel (1984), and Pesendorfer (2002). Consumers are heterogeneous in their brand loyalty and in their willingness to engage in product storage. The firm knows about this heterogeneity and accounts for it in price setting. We show that even in a stable demand environment, intermittent price discounts arise as an optimal strategy. Although the firm bears no costs of any kind in changing prices and consumer preferences are stable, the firm will iterate between a small number of prices, even in the face of changing marginal costs and some types of demand changes. The "regular" price will change infrequently but temporary discounts will be utilized to price discriminate. We show how to aggregate prices in this environment and discuss the robustness of our proposed measure to alternative assumptions about consumer behavior.

A. *Model Assumptions*

Consider a retailer selling two differentiated products: A and B. We will focus on a single retailer for simplicity.[3] The monopoly assumption is not necessary, and we discuss the oligopoly case as a possible extension after we have analyzed the basic setup.

Assume that all customers have unit demand in each period but differ in their preferences for the two goods. A fraction $\alpha/2$ of the customers value product A

---

[2] For a description of the airline methodology, see http://www.bls.gov/cpi/cpifacaf.htm. We thank a referee for pointing this out to us. The authors received helpful clarification of these methodologies through correspondence with BLS staff.

[3] It would be fairly straightforward to generalize to two retailers competing in geographic space. In models of this type such as Lal and Matutes (1994); Pesendorfer (2002); and Hosken and Reiffen (2007), consumer reservation prices would be determined by the price that would trigger consumer travel to another store.

at V^H and product B at V^L, where $V^H > V^L$. We call these consumers the Loyal A types. For convenience, we consider the symmetric case where a fraction $\alpha/2$ of the customers, the Loyal B types, value product B at V^H and product A at V^L. The remaining share of consumers, $(1 - \alpha)$, are Bargain Hunters who value both products at $V^M = (V^L + V^H)/2$. Thus, the Loyals and Bargain Hunters have the same mean willingness to pay for the two products, but Loyals have a preference between the two and the Bargain Hunters are willing to substitute between them freely. We normalize the total number of consumers to be 1 and suppose that the consumers can shop for N periods. Thus, in each period, the Loyal i types have utility $U^H = \max(0, V^H - P^i, V^L - P^{-i})$. The Bargain Hunter's utility from consuming is $U^B = \max(0, \max_i(V^M - P^i))$.

We allow for the possibility that the Bargain Hunters may choose to store goods. So their total utility consists of the utility from consuming minus any disutility associated with storage. In Chevalier and Kashyap (2014), we solve for a complete characterization of their optimal decision with rational expectations about future prices, when they are subject to a generic convex storage cost function. Here, we simplify the problem by assuming that they can store goods for free up to k^* periods and then face infinite storage costs.

The Loyals are assumed to be inattentive shoppers—they do not wait for and/or stock up during bargain periods, while the Bargain Hunters are strategic. These assumptions will deliver the empirical patterns described in Aguiar and Hurst (2007); they document that some consumers in a local area pay systematically lower prices for the same goods than other consumers.

The seller has a constant returns to scale technology of producing A and B, and the marginal cost of producing either is c. We assume that the marginal cost is below the reservation values for both types of consumers for both products.

B. *Model Results—Retailer Behavior*

To understand the optimal pricing strategy we start with the observation that total profits for the retailer over a horizon of N periods depend on the total amount of A and B sold and the prices charged.

Each period the Loyals arrive at the retailer to shop. If the price of either good is less than or equal to their reservation value, they purchase one unit of the good that delivers more surplus—but, they never engage in storage.

In contrast, the Bargain Hunters engage in storage whenever it pays to do so. Given the fixed capacity assumption that we have made about their storage costs, their shopping strategy becomes very simple. They seek to time their purchases so that when there is a discount on a good they buy enough of it to last until the next time that there will be a sale.

In this simple model, it is easy to see that the retailer has only two relevant choices to maximize profits:

- Always charge V^H, thus selling to all Loyal types every period and never serving the Bargain Hunters; or

- Hold a sale on one (and only one) of the products at a price of V^M every k^*th period in order to allow the Bargain Hunters to stock up and consume every period.

Notice that this follows because there is no benefit to charging a price between V^M and V^H, as charging such a price would lower the profits from the Loyals without achieving a sales gain from Bargain Hunters. Also, given the free storage technology available to the Bargain Hunters, there is no reason for the retailer to offer a good at a price of V^M, except in the special case where k^* equals 1.[4]

Because the Bargain Hunters can store for k^* periods, if the retailer holds a sale every k^*th period, then the undiscounted total profits for this retailer over all N periods is

$$(1) \quad N\frac{k^*-1}{k^*}\alpha(V^H - c) + \frac{N}{k^*}\frac{\alpha}{2}(V^M - c) + \frac{N}{k^*}\frac{\alpha}{2}(V^H - c) + N(1-\alpha)(V^M - c).$$

The four terms in (1) are very intuitive. The first term represents the profits from selling to the Loyals only, which will occur during all the periods with regular prices. The second term is the profits from the Loyals during the periods where they are able to buy their preferred good at a discount. During these sale periods, the other Loyals still pay V^H, so that explains the third term. The last term is the profits from the Bargain Hunters. Notice as long as the sale price is low enough to induce storage, the Bargain Hunters will buy enough to consume every period, even though purchases take place only every kth period.

The optimal strategy for the retailer is to charge V^H for one good every period and to charge V^M for the other good every k^* periods, as long as

$$(2) \quad (1-\alpha)k^*(V^M - c) > \frac{\alpha}{2}(V^H - V^M).$$

This expression shows that it is optimal to have a sale every k^* periods if the incremental profits from having one sale over k^* periods to attract the Bargain Hunters is greater than the one-period loss in revenues from the Loyals from holding the sale. Notice that here, since the retailer does not discount either good to a price below V^M, and since $V^H > V^M > V^L$, we need no additional conditions to insure that the wrong-type Loyals will not substitute to the less-preferred good to take advantage of the sale.

From the retailer's point of view the relevant "price plan" is the full sequence of high and low prices that prevail over the cycle of N periods. Note that even with unchanging cost and demand parameters, for many parameter combinations, the firm changes prices from period to period as it optimally cycles between capturing rents from Loyals and capturing the demand of the Bargain Hunters. Also, note that

---

[4] The more complex model in Chevalier and Kashyap (2014) allows the storage cost of the Bargain Hunters to have a more general form of convexity than the simple model here. This creates a situation in which the choice of the periodicity of sales is non-trivial. In this simple model, it is easy to show that, if the parameters are such that it is ever profit-maximizing for the Bargain Hunters to consume, it is profit-maximizing for the retailer to hold sales frequently enough to induce the Bargain Hunters to consume every period.

changing marginal cost will not change the prices charged, unless the increase in marginal costs is large enough to induce a switch from a price plan with occasional discounts to serve the Bargain Hunters to a price plan that does not serve the Bargain Hunters. In this sense, our model is consistent with an extensive literature that suggests that regular price changes are infrequent. Furthermore, note that in any individual period, the quantity sold for an individual good could be as low as $\alpha/2$ and as large as $\alpha/2 + k^*(1 - \alpha)$, if the good is on sale and the Bargain Hunters are storing until a sale expected in k^* periods. Thus, the volatility in quantity across products and periods occurs despite demand and supply primitives that are constant through time.

It is useful to compare the outcomes of this model to the models proposed in Kehoe and Midrigan (2010); Eichenbaum, Jaimovich, and Rebelo (2011); and Pesendorfer (2002). In all of these models, a firm will charge a fixed regular price and sometimes offer a discount. In all of these papers, the decision to discount is driven by some change in the cost or demand environment. In contrast, in our model, a sale would occur every k^* weeks with no change in the cost or demand environment.

Another important distinction between all of these models and ours is that none of the others explicitly examine a retailer managing a portfolio of close substitute products. Indeed, the fact that cost changes are important in these models implies that the prices for close substitute products would tend to be *positively* correlated. In our model, the time series of prices for close substitute products are *negatively* correlated (unless a common shock leads to the change in V^L and V^H).

C. *Model Results—Implications for Price Measurement*

We take as our objective aggregating prices such that the change in our price aggregate over time would form an exact price index. Recall that an exact price index is one that measures the change in the cost of achieving a given level of utility over time. Here, we make the natural assumption that utility is properly measured as the weighted average utility of the two types.

A perhaps surprising feature of our model is that, given the assumptions we have made, the conditions described in Diewert (1995) and Reinsdorf (2003) hold, so that the change in unit values over time corresponds to an exact price index. Unit values measure the weighted average prices consumers actually pay; in general, the change in unit values over time does not form an exact price index because, as prices change, consumers undertake substitution and the utility achieved from period to period is not constant.

In our model, changes in the unit value do constitute an exact index because of two important features of our environment. The first is that the Bargain Hunters are indifferent to switching between products and, up to the storage constraint, are indifferent to the timing of their purchases. Second, due to the strategic behavior of the retailer, the Loyal customers never substitute to the less-preferred product. Thus, Bargain Hunters engage in substitution, but it is without utility consequence. Substitution by the Loyal types would have utility consequences, but it does not occur.

While these results are of course special to our model, the insight about retailer price discrimination applies more generally. Retailers have an incentive to use strategies to extract the high willingness to pay from more loyal types while encouraging substitution only among types for whom substitution has a low utility cost. Thus, the change in the price paid (unit value) is a better approximation of the exact index than it would be in the absence of strategic retailer behavior.

In our model, with optimal pricing, Bargain Hunters earn no net consumer surplus. They pay their reservation value V^M. Nevertheless, the presence of Bargain Hunters creates a surplus for the Loyal consumers. Because of the existence of Bargain Hunters, Loyals sometimes obtain a product at a price lower than their maximum willingness to pay.[5]

From (1) it is easy to see that the average revenue per unit (unit value) equals

$$(3) \qquad \frac{\alpha + 2k^*(1-\alpha)}{2k^*} V^M + \frac{\alpha(2k^*-1)}{2k^*} V^H.$$

A simple rearrangement of the above expression forms the basis of our price measurement proposal. The expression can be rearranged to be equal to

$$(4) \qquad \alpha\left(\frac{1}{2k^*} V^M + \frac{2k^*-1}{2k^*} V^H\right) + (1-\alpha) V^M.$$

The two terms in (4) have natural interpretations.[6] The second term is simply the lowest price available to consumers. We will dub this the "best price" in our empirical analysis. The first term is a fixed weight aggregate measuring the average price over the entire period. The coefficients in (4) represent the population shares of the Loyals and Bargain Hunters, respectively.

Thus, because temporal changes in the unit value in our model are equivalent to the exact price index, the change in (4) represents the change in the cost of achieving a given level of utility over time. The formula shown in (4) forms the basis of our price measurement proposal. If the patterns of consumer and retailer behavior sufficiently conform to our model, a price index can be constructed by measuring the changes over time in a price aggregate formed using:

- an estimate of the share of Bargain Hunters in the marketplace,
- a conventional fixed weight price aggregate, and
- a measure of the best price.

---

[5] With more general costs of storage, the prices paid would have to be lowered to compensate consumers for storing goods and that would drive a wedge between prices paid and consumer utility.

[6] The idea here, of thinking of the price as the average over the entire planning horizon, is closely related to an insight first provided in Feenstra and Shapiro (2003). They advocate for the use of the Tornqvist index rather than unit values. Reinsdorf (2003), in his discussion of their paper, notes that, under the storage assumption that we make, unit values are a reasonable alternative. Of course, the Tornqvist does not emerge as an exact index in our model. More generally, the primary drawback of the Tornqvist, from our perspective, is the necessity of having current period weights that cannot be collected by BLS price collectors and is not available in many ecommerce settings.

By providing a simple functional form for product substitution, the formula does not require time-varying quantity data or real-time expenditure weights. To summarize our framework suggests

$$(5) \qquad unitvalue_t \ = \ \alpha\,fixedweightagg_t + (1 - \alpha)bestprice_t,$$

and that changes in the unit value over time windows longer than the k^* sale cycle horizon would measure changes in the cost of achieving a fixed level of utility over time.

Note also from equation (5) that, as α approaches 0, the weighted average price will approach the "best price"—the lowest price posted for any of the substitute products within the k^* period planning cycle. If α is fairly small and V^H remains constant for a long period of time, then a time series of the weighted average price will resemble a fixed increment over the time series of the "best price." If α is very large, then the weighted average price will more closely resemble the "regular" price. For intermediate values of α, the price paid resembles a weighted average of the "best" price and the fixed price index as illustrated in (4).

These observations also have implications for inflation measurement. If the environment is stable, then the model only predicts that the unit value price aggregate will be lower than a price aggregate calculated using fixed weights or using any methodology assuming a constant and small elasticity of substitution. The *change* in prices paid might be higher or lower than the change implied by any particular index, so that conventionally measured inflation would not necessarily be too low or too high. If there are shocks, for example to the willingness to pay of Bargain Hunters, these shocks can shift whether sales are held and the depth of sales; this would drive a wedge between the unit values and other price aggregates.

We can take the log difference of equation (5) to compute an approximate 12 month inflation formula. Specifically,

$$(6) \qquad \ln\big(unitvalue_t\big) - \ln\big(unitvalue_{t-12}\big)$$

$$= \ \ln(\alpha\,fixedweightagg_t + (1 - \alpha)bestprice_t)$$

$$- \ln(\alpha\,fixedweightagg_{t-12} + (1 - \alpha)bestprice_{t-12}).$$

Both (5) and (6) suggest estimating equations that we will explore in the empirical analysis.

It is important to be clear about the constructs that we use going forward in our empirical work. Our model suggests a way to aggregate the level of prices extant at any given point in time to infer the level of utility for agents in the model. Thus, we will refer to our methodology and alternative ways of aggregating prices at a given moment in time as price aggregates. Substantial literature has explored issues of chaining price changes over time and constructing price indexes. Here, we simply define alternative inflation measures to be log differences in the price aggregates.

D. *Model Extensions and Observations*

The basic model demonstrates the attractiveness of a temporary discounting strategy in the retail environment. In our model, we assume both heterogeneity in brand preferences and heterogeneity in the willingness to engage in storage. In an alternative simple model with no storage but brand taste heterogeneity, the retailer would still have an incentive to put only one type of good on sale each period, thus price-discriminating between Loyals and Bargain Hunters. In our model, it is the difference in storage propensity between Loyals and Bargain Hunters which leads to our result that sales can be held only episodically.

The model can be extended in a number of ways. Most importantly, from a measurement perspective, one could incorporate more complex demand. In particular, one could maintain the assumption that there is a mass of consumers that are indifferent between the substitute products, but build a different model of the cross-price elasticity of the Loyals. In that case, the appropriate price index would still be the weighted average of the two types, but the Loyal type's price aggregate would be altered to reflect their substitution patterns.

Furthermore, there are numerous ways to expand the model to allow for additional types of consumers. In particular, as mentioned above, there may be some consumers who inter-temporally substitute actively, but are highly brand loyal. In these more complex cases, the formulation and intuition in (4) is particularly helpful. In a model with more types, the pricing outcomes will be as in (4), but the weights on the lower prices versus the higher prices will reflect the behavior of the different consumer types.

Also, while there is a long tradition of two-type models in the literature on temporary discounts (Varian 1980; Sobel 1984), the model can be extended to allow for a continuum of consumers' willingness to pay. As in Varian (1980), as long as enough consumers do not move all of their purchases into sale periods, intertemporal variation in prices will enhance the retailer's ability to price discriminate.

Our model also abstracts from competition among multiple stores. Obviously, a model with perfect competition would limit the opportunity to price discriminate. An extension of our model along the lines of Lal and Matutes (1994) would also preserve the basic insights of our model. Lal and Matutes (1994) model multiproduct retailers with travel costs between stores. The ability of consumers to shop both stores, at a cost, disciplines the overall rents across products that stores can extract from each consumer. The ability to extract different levels of rents from different consumer types and from different products is preserved. Ellison (2005), in a different context, provides an extension of Lal and Matutes (1994) in which price discrimination across consumer types emerges in an equilibrium with spatial competition. The easiest way to understand the effects of introducing competition into our model is to view competition as impacting consumer willingness to pay. Spatial competition opens additional scope for price discrimination as retailers may benefit from price discriminating between consumers who will and won't travel to obtain a price discount.

Finally, we could derive a similar set of predictions if we assumed that all consumers viewed all products in a category as perfect substitutes, but consumers

differed in their ability to store goods. In this case, the retailer has essentially a single brand to manage, but will still want to periodically use discounts to sort between consumers who must buy frequently because of high storage costs and those that can be more patient.[7] A price aggregate similar to the one we proposed would again be informative, though depending on how the storage costs are modeled the prices that are posted might have to compensate the consumers for the storage costs. In that case, the prices do not directly reveal the consumer surplus that accompanies a sale.

More generally, the retailer in our setup is trying to design a pricing strategy that induces consumers to sort towards different purchases. In this sense, the insights from the literature on self-selecting contracts triggered by the seminal work of Mussa and Rosen (1978) also apply. While our model is very particular in its two-type setup, we believe that our basic insight is fairly general. A unit value index is not in general an exact index when products are heterogeneous because substitution induced by price changes is typically not utility-neutral. Nevertheless, whenever a retailer successfully undertakes a second degree pricing strategy between substitute brands, the unit value index becomes a closer approximation for the exact index, as the customers with the highest utility cost of brand-switching are strategically prevented from switching.

Summing up, the model comfortably explains the familiar price pattern observed for individual goods of a regular price with intermittent sale prices. It suggests a way to aggregate prices even without high-frequency data on quantity purchased and makes four predictions that we will test. First, at the store-level, a disproportionate fraction of goods are sold at temporary discounts. Second, unit values should be well-approximated by a linear combination of a conventional fixed weight price aggregate and the best available price within the group of close substitutes over the course of several weeks. Third, the degree of substitution embedded in existing BLS-style price aggregates will not adequately account for the migration of consumers to the "best price." Fourth, inflation as experienced by consumers will depend substantially on the inflation rate for the "best price."

III. The Data

We are interested in examining both the incidence of sales and consumer responses. To accomplish this, our project requires a dataset that contains both prices and quantities. We use IRI's "IRI Marketing Data Set," which is described in Bronnenberg, Kruger, and Mela (2008). Since the time of that publication, IRI has released additional data so that it covers retailers in 50 market areas for the period 2001 to 2011. We undertake two sets of analyses. First, to explore the details of our model on micro data, we undertake a store-level analysis. For these purposes, we focus on two categories of products: ground coffee and peanut butter. We look at

---

[7] Distinguishing between models where consumers differ in their brand preferences and their ability to store is not easy, though if all heterogeneity was due to storage, relative price movements between different goods would be indeterminate. Empirically, it seems that discounts on competing brands, say Coke and Pepsi, are staggered, which is consistent with the basic model we propose.

substitution patterns for a single large store in each of nine cities. Second, to more closely approximate BLS inflation measurement procedures, we construct national indices using a broader set of products that we sample nationally.

For the store-level analysis, we choose the peanut butter and coffee categories for several reasons. First, the categorization of these products in the IRI dataset matches closely a product that is tracked by the BLS. This allows us to mimic the BLS sampling procedures for each of them. Reinsdorf (1999) examines coffee for a similar reason—to match the BLS data to scanner data. The categories are reasonably representative of all the grocery products in the IRI sample. Bronnenberg, Kruger, and Mela (2008) describe the 30 categories tracked by IRI and calculate the percent of total volume in the category sold "on any deal." The median category has 37 percent of its volume sold "on any deal." They report that the share sold on deal is 40.8 percent for coffee and 32.9 percent for peanut butter. In our store-level analysis of these two products, we examine results for a sample of nine cities, one chosen randomly from each of the nine census divisions. In each city, we sample from the largest chain—though we discard any chain where private label brands dominate the national brands. The details of our rules for selecting brands and stores are discussed in the online Appendix. Here, we summarize the main elements of our procedures.

Whenever possible, we try to mimic BLS methodologies for examining prices. In particular, although the IRI data are reported at a weekly frequency, the BLS indices are constructed at a monthly frequency. Therefore, in our analyses, we aggregate data to a monthly frequency to match CPI construction methodologies. The exception is the summary information in Table 1.

Our store-level analysis proceeds starting with a set of specific universal product codes (UPC) for coffee and peanut butter. An example of the most disaggregated definition of a good that we would consider would be a "Peter Pan 18 ounce jar of creamy peanut butter." The IRI dataset, like most scanner datasets, contain a large number of products, many of which have a short lifespan. To focus our analyses, first, we restrict attention to benchmark sizes (11 to 13 ounces for coffee and 15 to 18 ounces for peanut butter) that are most typical in the data. Next, we identify the best selling products and focus on the top ten UPCs in each category. In calculating market shares, we inspect the data and splice together UPCs that are replacements for each other. In addition, we consolidate all UPCs within a given brand that have very high price correlation (as is typically in the case for different flavors of the same product). We then retain only UPCs for which we have at least six years of data. Finally, we convert the prices into price per ounce and exclude any private label UPCs or super premium UPCs. Collectively, we believe these steps create coherent categories of similar products where the model's assumptions about substitution are reasonable and where sampling variation will not produce spurious price volatility.

For some of our calculations, we require definitions for "sale prices" and "regular prices." We extend the methodology proposed by Kehoe and Midrigan (2010). They propose measuring a sale as a price cut that is reversed within five weeks. We adopt a similar definition. The data contain some miniscule apparent price changes; there are cases where the price in a week appears to be less than a cent or two lower than

the price in the previous week. As in most scanner datasets, the price series is actually constructed by dividing total revenues by total unit sales. There may be product scanning input errors or situations in which a consumer uses a cents off coupon or a store coupon, and any of these would create tiny shifts in measured prices that do not reflect changes in posted prices. We thus set a tolerance for a price change—requiring the price change to be "large" enough to be considered either a sale price or a change in the regular price. We set this tolerance at $0.002 per ounce. A regular price is set equal to the actual price if the product is not determined to be on sale using the above methodology. If the item is on sale, the regular price is defined to be the most recent past price for the item for which the item was not on sale.

It is worth emphasizing that the theoretical construct in our model is a fixed weight aggregation that uses quantity weights rather than expenditure weights. Thus, in our micro analysis of coffee and peanut butter, we use within-store quantity weights in constructing our fixed weight, geometric mean aggregates, and CES aggregates. While our model calls for fixed weights, over the decade of our sample, there are some changes in the UPCs and products offered. Furthermore, some brands gain in popularity while others decline. Thus, instead of a truly fixed quantity weight, we use a methodology similar to the BLS to calculate and update weights for the fixed-weight index. We start by identifying base volume weights in a "pre-period" of our dataset. Each quarter, we construct a new weight consisting of 15/16 of the previous weights and 1/16 of the prior quarter's weights.

For our national index analysis, we expand our analysis to consider 21 additional products from IRI. These represent all of the categories of products tracked in the IRI data that we viewed as sufficiently homogeneous to allow us to identify a "best price" and for which sufficient continuous data were available.[8] Because the BLS samples on a monthly basis, we assign each week in our database to a month. As the BLS does, each month, we will sample from particular stores in a particular week. As with coffee and peanut butter, we focus on the most popular sized products. We identify a set of candidate UPCs in each store based on historical expenditure weights, and then sample a particular UPC in each store using these weights. The online Appendix describes the way we approximate the BLS procedures for rotating new stores into the sample. Most of these products do not correspond exactly to a category that the BLS tracks in an identifiable way. The reasons for this are varied. For example, one of the categories that we examine is frankfurters. This does not correspond exactly to a category that the BLS tracks. This is in part because the BLS categorizations classify meats by animal of origin. Thus, frankfurters contribute to the pork, beef, and turkey items, and there is no BLS item for frankfurters. In order to try to mimic the BLS methodology, we need to know roughly how many prices to collect for each product. By choosing approximately the same number of prices as the BLS, our sampling error should be comparable to that of the BLS. Unfortunately, even for specific products that the BLS tracks, the BLS does not provide detailed data about the number of prices sampled. Based on the total number of food and beverage

---

[8] An example of a category that we excluded was diapers. Ounces seemed like a poor measure for diapers, and using the number of diapers in a package is also misleading since diapers are sized like clothing, and the products for larger size children tend to be more expensive.

prices collected monthly by the BLS (see http://www.bls.gov/cpi/cpirr2011.pdf), and given the weights of peanut butter and coffee in the total food and beverage category, we estimate that the BLS would collect roughly 240 monthly quotes for peanut butter and 224 for coffee. Lacking better information on exactly how many prices the BLS would sample in our other categories, we collect 240 prices per month for all categories in our national index analysis.

In order to make sure that our calculations are geographically representative, we divide the United States into the four US census regions and allocate our price estimates to each region in the same proportion as they are allocated in the CPI (using data from http://www.bls.gov/cpi/cpiri2011.pdf). Within each region, for each product category, in order to conduct our national analysis, we calculate category-level store expenditures for each store in the IRI sample. Within each store-category, we calculate expenditures for each UPC. As described in more detail in the online Appendix, we use these measures to construct store and item sampling weights. We use these sampling probabilities to draw the 240 items sampled for each category. Importantly, we collect only one UPC per store, and use equal weights in aggregation, consistent with BLS practices.

These choices mean that we focus on only 2 categories in our store-level analysis and a broader set of 23 categories in our national analysis. There is always a trade-off between the comprehensiveness of categories and the care that can be paid to the details of each category. For example, Chevalier, Kashyap, and Rossi (2003) document that in some categories UPCs are discontinued only to have the same product appear with a new UPC. Hence, splicing series by hand is the only sure way to capture all the same sales of these types of similar items. But, the task of splicing the data to capture UPC changes, as well as grouping UPCs within brands, requires a substantial investment in learning about the category and cleaning data. Of course, this could, in principle, be done on a large scale, but it is very costly. The Bureau of Labor Statistics devotes some portion of its field force to cleaning data to create consistency across categories in construction of the CPI. For our peanut butter and coffee categories for the store-level analysis, we undertake this type of splicing. In our national analysis of 23 categories, we use a series of rules to select the sizes and products that are included. Details can be found in the online Appendix.

IV. Results—Store-Level Data

Before turning to the main predictions of our model, we document that high frequency price variation is quite important for understanding the actual prices that consumers pay. Figures 1 and 2 show, for the 2003–2005 period in the Charlotte store, the monthly fixed weight price aggregate for peanut butter and coffee, respectively, along with a fixed weight aggregate of regular prices, the unit value, and the best price. The graphs show that both products have a substantial period of time when the regular price is nearly constant. Movements in the fixed weight price aggregate clearly reflect the occasional temporary discounts (captured by movements in the best-price series). As expected, unit values embody temporary discounts much more than the fixed weight aggregate. Although not shown, the geometric mean comoves very closely with the fixed weight price aggregate, but sits at a lower level; the
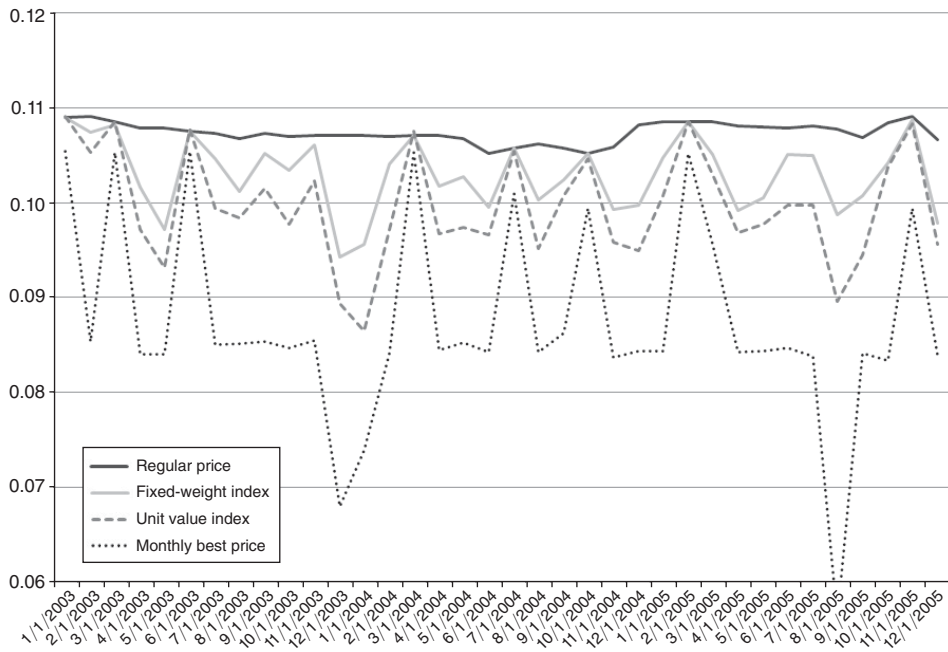
Peanut butter, Charlotte 2003−2005

FIGURE 1. MONTHLY PEANUT BUTTER REGULAR PRICES, FIXED-WEIGHT AGGREGATE, UNIT VALUES, AND BEST PRICES

Notes: The figure shows a fixed weight aggregate of regular prices (defined in the text and produced using our modification of the Kehoe and Midrigan 2010 algorithm for detecting sales), a fixed weight aggregate of posted prices, unit values, and the monthly best price.

difference with the fixed weight aggregate would be almost imperceptible at the scale of the graph.

Next, we explore whether our model adequately explains the pattern of temporary discounts and purchases. Recall that we identified three store-level predictions of our model. First, a disproportionate fraction of goods should be sold while the price is temporarily low. Second, unit values should be well-approximated by a linear combination of a conventional fixed weight price aggregate and the best available price within the group of close substitutes over the course of several weeks. Third, the geometric mean price aggregate used by the BLS in constructing the price index should not adequately account for the migration of consumers to the "best price."

A. *Purchase Responses to Sales*

We first examine the effect of sales on prices paid by consumers and quantities purchased. Table 1 shows total weeks and ounces with temporary discounts for each of the nine cities. It also shows, for weeks in which there is a discount, the mean size of the discount in each city. It is unsurprising that quantity sold increases substantially when a product experiences a price reduction. The combination of
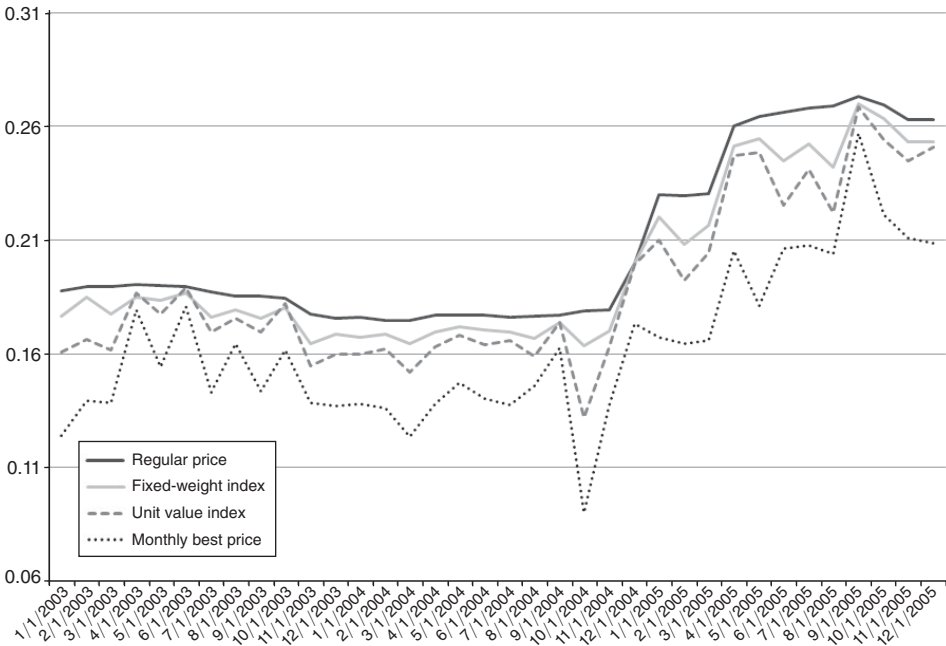
Coffee, Charlotte 2003–2005

FIGURE 2. MONTHLY COFFEE REGULAR PRICES, FIXED-WEIGHT AGGREGATE, UNIT VALUES, AND BEST PRICES

Notes: The figure shows a fixed-weight aggregate of regular prices (defined in the text and produced using our modification of the Kehoe and Midrigan 2010 algorithm for detecting sales), a fixed-weight aggregate of posted prices, unit values, and the monthly best price.

frequent, staggered discounts along with consumers who readily switch brands and time purchases means that a substantial fraction of all of the units sold are sold at prices below the "regular" price.[9] In all of the cities, the fraction of ounces sold at discount is substantially greater than the share of item-weeks with discounts, suggesting ubiquitous strategic shopping behavior by consumers for both products in all locations. For peanut butter, on average, the fraction of units bought on sale is roughly twice the fraction of item-weeks with sales. For coffee, responsiveness is lower, with the fraction of units sold on sale being roughly 1.3 times the fraction of item-weeks on sale. Under the strong assumption that sale weeks are chosen randomly, the ratio of the quantity sold in sale versus non-sale weeks can be compared to the mean size of discounts during sales to calculate a rough estimate of the elasticity of demand. The last column of the table shows the resulting estimates and they are large.

[9] The large share of transactions that take place at sale prices is not surprising. Kehoe and Midrigan (2010) mentions this finding as one of their observations about the Dominick's data that is distributed by the University of Chicago. Bronnenberg, Kruger, and Mela's (2008) IRI dataset covers 30 categories of goods at grocery chains in 47 different geographical markets (that has since expanded to 50 markets). Their Table shows the fraction of products that are sold on any deal and the mean percentage is 36.8 percent; more than 30 percent are sold on deal in 25 of the 30 categories they study. Griffith et al. (2009) also finds that about 29.5 percent of total food expenditure from a large sample of British households is on sale items. Hence, the findings for our categories are very typical of what happens in grocery stores.

TABLE 1—SHARE OF OUNCES SOLD AND SHARE OF WEEKS AT REGULAR AND SALE PRICES:
TOTALS FOR SAMPLE CITIES

	Ounces sold (%)		Weeks (%)		Average discount (%)	Demand elasticity
	Regular price	Sale price	Regular price	Sale price		
Panel A. Peanut butter						
Charlotte	60.03	39.97	75.91	24.09	17%	5.38
Chicago	33.92	66.08	59.17	40.83	21%	7.56
Hartford	50.08	49.92	92.45	7.55	27%	10.02
Houston	63.49	36.51	74.57	25.43	12%	3.90
Knoxville	65.24	34.76	73.13	26.87	11%	3.85
Los Angeles	49.49	50.51	65.83	34.17	13%	6.98
New York	37.49	62.51	78.63	21.37	21%	7.84
St. Louis	34.88	65.12	67.73	32.27	26%	5.65
West Texas	46.26	53.74	68.60	31.40	19%	4.52
Panel B. Coffee						
Charlotte	31.51	68.49	54.23	45.77	9	10.01
Chicago	43.27	56.73	52.01	47.99	13	3.02
Hartford	18.56	81.44	49.01	50.99	12	9.00
Houston	42.89	57.11	57.83	42.17	6	7.38
Knoxville	44.59	55.41	56.10	43.90	7	4.48
Los Angeles	41.48	58.52	50.42	49.58	14	2.85
New York	13.16	86.84	43.79	56.21	16	8.06
St. Louis	31.88	68.12	52.71	47.29	11	4.93
West Texas	40.32	59.68	50.98	49.02	9	5.16

Notes: Totals are given for the store identified in the online Appendix for each city. The sale and regular prices are calculated as described in the text based on our modification of the Kehoe and Midrigan (2010) algorithm for detecting sales. The average discount is the simple average overall sales for any items on sale in a given week.

There are a number of reasons why these patterns for coffee and peanut butter may differ. For example, the demographic characteristics of peanut butter buyers may differ from that of coffee buyers or coffee companies may be more successful in differentiating themselves in a way that creates brand loyalty, and people may be more willing to store peanut butter. The greatest potential impact of sales on price measurement occurs in Hartford for peanut butter, where, over our 11-year sample, only 8 percent of product-weeks have sales, but 50 percent of the total ounces sold are purchased at a sale price.

B. *Best Prices*

We next turn to the most important prediction of our model, that the variable weight (unit value) price aggregate can be well-approximated as an appropriately weighted average of fixed weight price aggregate and a measure of the " best price," the lowest price obtainable within a narrow item category (or stratum, to use the BLS language). In order to test this implication, we do need to operationalize some theoretical constructs of our model. First, we must make an assumption about the horizon over which Bargain Hunters can be expected to look for sales and stockpile.

Given that the BLS publishes monthly data, we will assume that shoppers operate on a monthly cycle. Second, we must define the universe of items within the store over which to seek the best price. Here, we use all products that meet the popularity and availability screens described above.

We conduct a simple regression of the unit value price aggregate on the fixed weight price aggregate and the "best price" series. Note that if equation (5) is a good approximation for our data, the weights on the two price series should add up to one and the constant term should be roughly zero. These conditions are *not* hardwired to hold. Consumer preferences and, therefore, market shares could drift away from the fixed weights of the fixed-weight indices; or consumers' demand might be very different than supposed by our model that emphasizes a prominent role for best prices.

Results are shown in Table 2 for both products in each of the cities. The first important finding is that the best-price coefficient is positive and significant in all cases. That is, the prices actually paid by consumers (unit values), are not well-approximated by the fixed weight aggregate. Instead, controlling for the best price adds explanatory power. This is not surprising given our findings in Table 1.

For most of the product/city combinations, the constant term appears small and the fixed weight and best-price coefficients roughly sum to one. In about half of the cases for each product, the estimated value of the constant term is not significantly different from zero at standard confidence levels. In about half of the cases for each product, we also cannot reject that the hypothesis that the coefficients for best price and the fixed-weight index sum to one. In all cases, the model fits well; the minimum R^2 among the 18 product-city specifications is 0.86.

The coefficient estimates on the best price vary between 0.2 and 0.7, suggesting that bargain hunting tendencies differ noticeably. Interestingly there is a strong positive (0.80) correlation between the best-price coefficients for the two products within stores. To the extent that demographic characteristics of the store customers are driving the importance of the best price, this correlation is reassuring. Some of these patterns are easier to observe when results are depicted graphically. Figure 3 shows the best-price coefficient as a dark bar with the fixed weight coefficient as a lighter bar. The distance of the sum from one can be easily seen. Above the coefficient estimates is a simple plot of the constant term. Following the layout of Table 2, the left panel shows the results for peanut butter, and the right panel coffee.

An important question to consider is whether the BLS adequately accounts for sale-chasing behavior. The BLS constructs a fixed expenditure share geometric mean index within item strata. This methodology, therefore, allows for a limited amount of cross-item substitution. This substitution differs substantially from that implied by our model. The BLS methodology effectively assumes a cross-price elasticity of demand of one between items within the strata. Whether or not this elasticity is sufficient to capture shopping patterns is an empirical issue, as is the question of whether ordinal prices as stressed by our model matter.

To assess the importance of these possibilities, Figure 4 depicts an alternative set of regression results in which a BLS-type geometric mean index and our Best Price indicator are both allowed to account for movements in the unit values. In all cases, the best price measure still has significant explanatory power for the unit values. Indeed, as can be seen by comparing figures, the coefficients for best price

TABLE 2—STRUCTURAL ESTIMATES OF PRICE COEFFICIENTS: EXPLAINING STORE-LEVEL UNIT VALUES WITH
FIXED-WEIGHT PRICE AGGREGATES AND BEST PRICES

		Peanut butter coefficients		Coffee coefficients
Charlotte	Fixed weight	0.804 (0.024)	Fixed weight	0.737 (0.038)
	Best price	0.234 (0.022)	Best price	0.292 (0.040)
	Constant	−0.0038 (0.002)	Constant	−0.0040 (0.004)
Chicago	Fixed weight	0.542 (0.032)	Fixed weight	0.648 (0.031)
	Best price	0.548 (0.039)	Best price	0.386 (0.042)
	Constant	−0.006 (0.003)	Constant	0.0058 (0.006)
Hartford	Fixed weight	0.484 (0.044)	Fixed weight	0.437 (0.035)
	Best price	0.587 (0.029)	Best price	0.667 (0.038)
	Constant	−0.005 (0.004)	Constant	−0.0162 (0.005)
Houston	Fixed weight	0.646 (0.045)	Fixed weight	0.831 (0.017)
	Best price	0.319 (0.030)	Best price	0.206 (0.017)
	Constant	0.0065 (0.004)	Constant	−0.007 (0.002)
Knoxville	Fixed weight	0.664 (0.037)	Fixed weight	0.678 (0.028)
	Best price	0.289 (0.032)	Best price	0.306 (0.031)
	Constant	0.007 (0.003)	Constant	0.0086 (0.002)
Los Angeles	Fixed weight	0.687 (0.047)	Fixed weight	0.716 (0.038)
	Best price	0.316 (0.032)	Best price	0.291 (0.033)
	Constant	0.0042 (0.004)	Constant	0.0053 (0.008)
New York	Fixed weight	0.414 (0.037)	Fixed weight	0.348 (0.043)
	Best price	0.590 (0.040)	Best price	0.697 (0.047)
	Constant	0.005 (0.004)	Constant	0.0009 (0.008)
St. Louis	Fixed weight	0.808 (0.073)	Fixed weight	0.646 (0.023)
	Best price	0.403 (0.045)	Best price	0.375 (0.020)
	Constant	−0.0258 (0.007)	Constant	−0.002 (0.003)
West Texas–New Mexico	Fixed weight	0.669 (0.067)	Fixed weight	0.915 (0.031)
	Best price	0.401 (0.044)	Best price	0.183 (0.026)
	Constant	−0.007 (0.007)	Constant	−0.0257 (0.006)

Notes: For each city and category we run a single regression motivated by equation (5) in the text. The dependent variable is the unit value for the dominant brands in the sampled store. The brands for each store are listed in online Appendix Table A1. The independent variables are the fixed-weight aggregate for those brands, the monthly best price amongst those brands, and a constant. Standard errors are in parentheses.
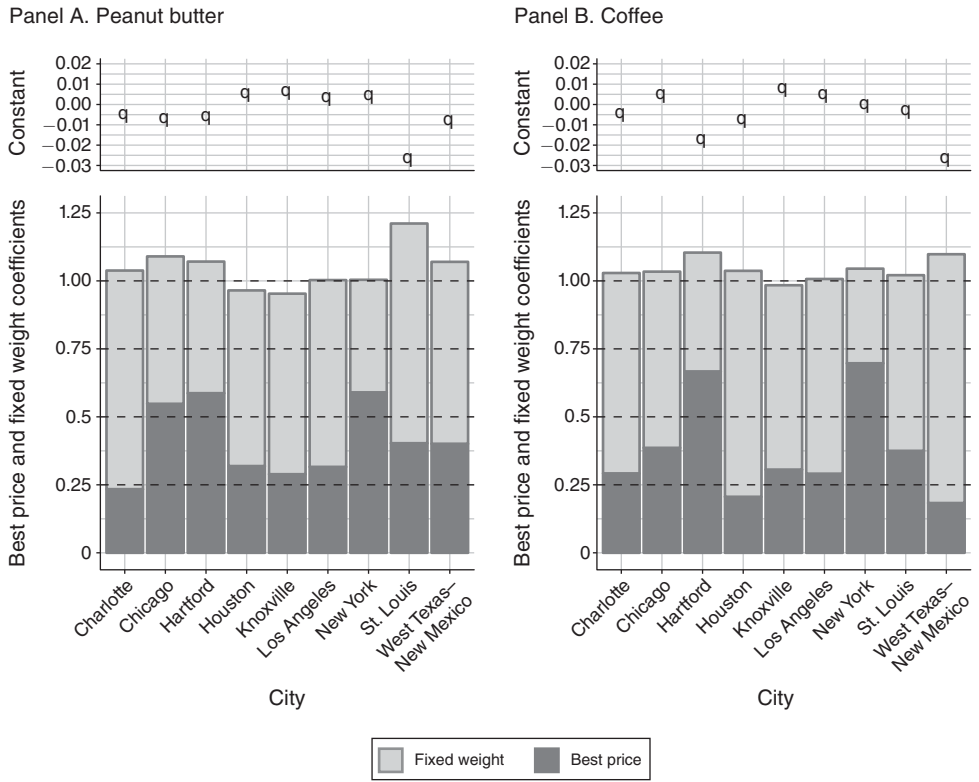
Panel A. Peanut butter   Panel B. Coffee

FIGURE 3. STORE-LEVEL ANALYSIS FOR PEANUT BUTTER AND COFFEE—RELATIONSHIP OF UNIT VALUES TO FIXED WEIGHT AND BEST PRICE AGGREGATES
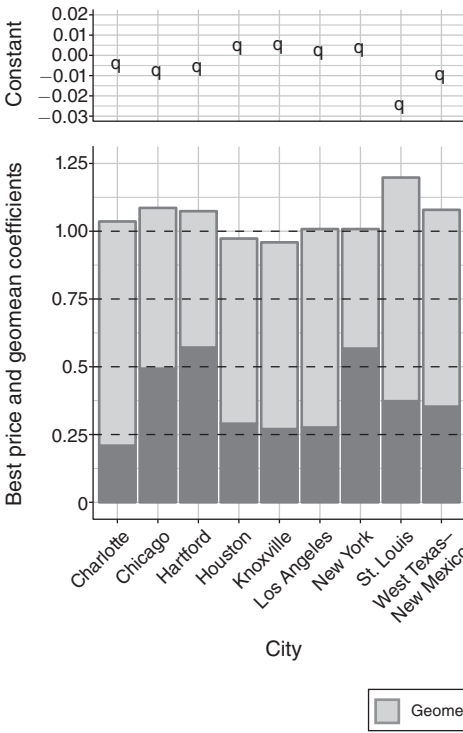
Notes: Relationship of unit values to fixed weight and best price aggregates based on estimates reported in Table 2. The upper part depicts the constant term coefficient in the regression of unit values on the fixed weight aggregate and best price. The lower part shows the best price coefficient and the fixed weight coefficient. A key prediction of the model is that the two coefficients sum to one.

when using the geometric weight index in the regression are nearly identical to the coefficients for best price when using the fixed-weight index in the regression. This suggests that the geometric mean does not sufficiently account for deal-chasing behavior on the part of strategic shoppers.

There are two distinct reasons why best-price coefficient could be positive and significant in these regressions. First, the substitution elasticity of one that is implicit in the geometric mean may not be large enough to capture the substitution patterns in the data. Alternatively, the CES may not be the correct functional form to capture the substitution patterns in our data. It is well-understood that the CES well-approximates aggregate data when the individual consumers have logit demand. In both the logit and the CES, cardinal price differences drive decisions; our model suggests that the ordinarily lowest price in the category should play an important role in driving purchasing.

To address these two alternatives, we investigate whether a CES price aggregate would perform better if it had a higher elasticity of substitution. Recall that a
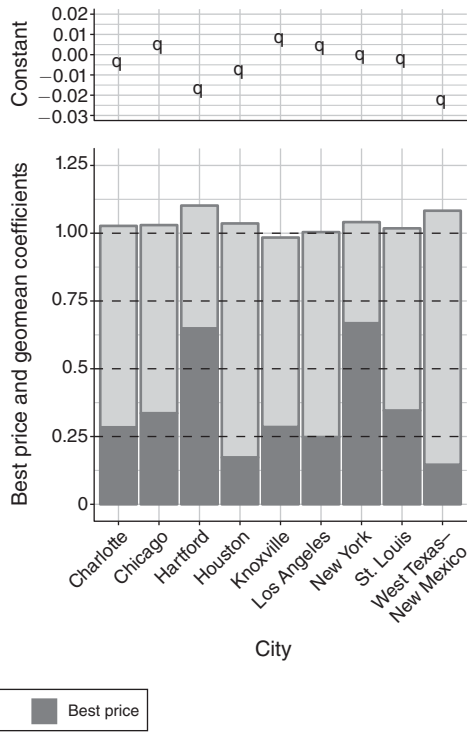
Panel A. Peanut butter

Panel B. Coffee

FIGURE 4. STORE-LEVEL ANALYSIS FOR PEANUT BUTTER AND COFFEE—RELATIONSHIP OF UNIT VALUES TO GEOMETRIC MEAN AND BEST PRICE AGGREGATES

Notes: The relationship of unit values to geometric mean and best price aggregates is based on estimates reported in online Appendix Table A5. The upper part depicts the constant term coefficient in the regression of unit values on the geometric mean aggregate and best price. The lower part shows the best price coefficient and the geometric mean coefficient.

geometric mean aggregation would match actual consumer behavior if consumers had a cross-price elasticity of one across varieties. To do this, we follow Shapiro and Wilcox (1997) who focus on matching the assumed elasticity of substitution to the substitution patterns present in the data. Specifically, we conduct a grid search to find the CES price aggregate that best matches the observed unit values. Specifically, we find the σ such that the CES price aggregate best matches unit values. The CES price aggregate is $\left(\sum_{i \in \Omega_{st}} \alpha_{is} p_{ist}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$, where Ω_{st} denotes the set of all products i sampled in store s at time t and the α coefficients are the quantity weights.[10] For each of the city-products, we search over elasticities of substitution from 0.5 to 10 in increments

---

[10] In their study, Shapiro and Wilcox (1997) start with the proposition that the statistical agency would want to use a superlative index, such as a Tornqvist, but recognize the problem that the Tornqvist requires real-time, quantity data. They investigate the ability of the CES to match the Tornqvist by conducting a grid search for the best match CES. In our (very different) context, we extend this idea to find the CES that best fits the unit value and then measure whether the best price is still important in explaining movement in the unit value.
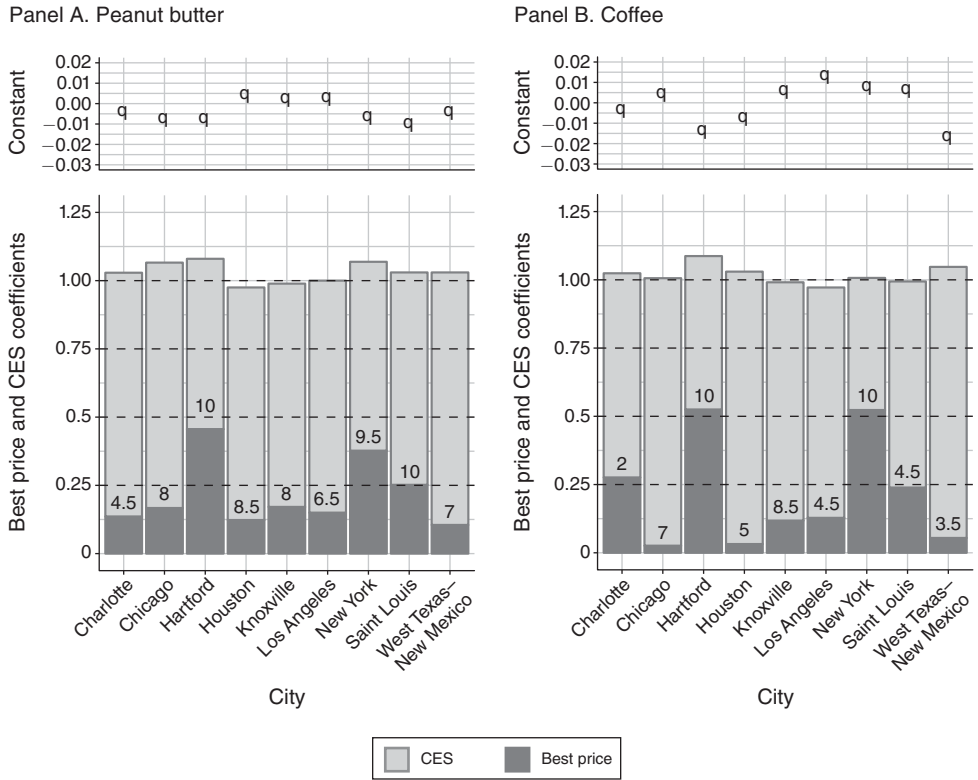
Panel A. Peanut butter

Panel B. Coffee

FIGURE 5. STORE-LEVEL ANALYSIS FOR PEANUT BUTTER AND COFFEE—RELATIONSHIP OF UNIT VALUES TO BEST FIT CES AGGREGATE AND BEST PRICE AGGREGATE

Notes: The relationship of unit values to best-fitting CES and best price aggregates is based on estimates reported in the online Appendix Table A6. Peanut butter is in left panel and coffee is in right. The CES parameter that best fits is noted in text above the bar, while the bar height measures the coefficient for the CES parameter in the best price regression.The upper part depicts the constant term coefficient in the regression of unit values on the best fit CES aggregate and best price. The lower part shows the best price coefficient and the best fit CES coefficient.

of 0.5 in order to find the CES price index for that city-product that most closely approximates the unit value aggregate. We then use these preferred CES parameters to reestimate regressions akin to the ones reported in Figure 3 and Figure 4. The results are depicted in Figure 5.

The numbers over the bars are marked to show the elasticity that most closely approximates the unit value. As in earlier figures, the dark bars show the coefficients on the best price and the lighter bars show best fit CES aggregate. Not surprisingly, for all city-product pairs, the best matching CES aggregate has an elasticity of substitution that is higher than one. For peanut butter, the preferred specification yielded an elasticity estimate greater than 4.5 in each city. For coffee, the best fit CES aggregate has an elasticity ranging from two to ten. These findings accord with Ivancic, Diewert, and Fox (2009), who suggest (in matching to the Tornqvist) that estimates of elasticities of substitution in the lowest item strata for food products are typically greater than one and are often greater than three. Consistent with the patterns in Tables 1 and 2, coffee is on average less elastic than peanut butter, and, as suggested by the raw data, there is a positive correlation across the two products

within cities (the correlation of the best fitting elasticity of substitution for coffee and the peanut butter within cities is 0.7).

For all of the city-product pairs except coffee in Chicago, the coefficient on the Best Price measure is positive and statistically different from zero at at least the 10 percent confidence level. This suggests that the substitution patterns in the data are not well-approximated by a CES specification; the ordinal price metric of the best price is typically required to match the substitution patterns in the variable weight index. In the one exception for the 16 city-product pairs, coffee in Chicago, the variable weight index is closely approximated by a CES price index with an elasticity of substitution equal to 7.[11]

As a final point of comparison, we also compute Tornqvist indices for peanut butter and coffee in each of our cities. In our model, a Tornqvist does not approximate an exact index due to consumer heterogeneity. In contrast, in many models with homogeneous consumers, a Tornqvist aggregate would approximate an exact index and hence is another potential benchmark. In the online Appendix, we show that a regression where the Tornqvist index is the dependent variable and the geometric mean and the best price are the controls, uncovers the best-price coefficient that is typically positive and statistically different from zero. This suggests that the BLS's preferred geometric mean index would better approximate the Tornqvist if the best price measure were also considered.

C. *Inflation Measurement*

Finally, we assess what the patterns we have documented imply for the measurement of inflation. To do this, we first use our scanner data to broaden our sample of products and to aggregate prices nationally. We use all categories available from our data vendor where the product size variation is small, products are fairly homogenous, and sufficient data exists in our database. This results in 21 additional grocery categories plus coffee and peanut butter. It is certainly clear, from our analysis using the city-level data discussed above, that the *level* of aggregated prices will differ depending on whether prices are aggregated using time-varying expenditure weights (unit values), or with fixed arithmetic weights (as are used for price index construction in many countries including Japan), or with geometric weights (as is used by the BLS in the United States).

Nonetheless, our price-level results do not necessarily have important implications for tracking inflation. If the mismeasurement in levels is constant over time, inflation may be well-measured with BLS procedures. Previous research, however, suggests several reasons to expect meaningful time variation in the frequency or depth of sales. Given the strong purchase response to sales, this would lead to time variation in unit-value inflation and best-price inflation that may not be fully captured in a BLS-style geometric mean inflation aggregate. This time variation in the frequency

---

[11] It is worth noting that, for the unit price index for coffee in Chicago, a regression on the CES index and a constant yields an R^2 of 0.98. A regression of the variable weight index on the fixed-weight index and the best price yields an R^2 of 0.96. Thus, although the best fit CES in that one case does not require the additional information contained in the best price, our best price methodology still yields a very good approximation of the variable-weight index.

or depth of sales could occur either due to cyclical factors as in Kryvtsov and Vincent (2014) or due to long run shifts in selling technologies as in Basker (2015) and Nakamura (1998), or due to consumer substitution patterns differing in high- and low-inflation environments as in Handbury, Watanabe, and Weinstein (2013). Indeed, Nakamura and Steinsson (2008) document that the fraction of BLS price quotes that are sales increased dramatically in the processed food, unprocessed food, household furnishings, and apparel categories over the 1988 to 2005 period. For the processed food and apparel categories, sales frequencies have doubled. Furthermore, the average size of sales has also increased substantially over the period in three of the four aforementioned categories. This continues a trend documented by Pashigian (1988) in the frequency and depth of sales that begins in the 1960s. If any of these forces are at work, inflation in unit values may be better approximated by a methodology such as ours that measures discounts explicitly, rather than a geometric mean methodology.

We also construct best-price indices. Our model suggests that because of the option to stockpile, the best price in a given week is a conservative measure of the price savings available to a determined bargain hunter. The BLS only collects prices in a given store in one week. So, the construction of a multi-week best price would require additional enumerator store trips and thus would be extremely expensive for the BLS relative to the current methodology. In order to create a practical proposal, we depart from the model's prediction that emphasizes a multi-week best price and collect instead the best price that is available during the week when the quote for the regular index would be gathered. The one week best price is selected from the candidate UPCs that were identified in that store as being potentially eligible for inclusion in the regular index.

Finally, we compute a national unit value index. We use this as our benchmark and assess the extent it could be approximated by the other indices, as it utilizes quantity data unavailable to the BLS. To compute this, we sample stores and weeks identically to the method used to calculate the fixed-weight, geometric mean, and best-price indices. For each sampled store, for the entire set of candidate UPCs from which we sampled for the price indices, we measure total expenditures for the week and divide by total volume in ounces. Thus, to obtain a sampled price for each store, we use a unit value approach; however, prices across stores and cities are aggregated using fixed weights.

Unit-value approaches that use transaction data can be influenced by product entry, exit, and availability. Our methodology attempts to create product categories that are homogeneous enough that entry and exit do not cause large changes to the unit values. This issue is a source of considerable study in countries where scanner data unit value methodologies have been incorporated into the CPI (see, for example, Chessa, Verburg, and Willenborg 2017). In practice, it is possible that products that become unavailable may have systematically higher or lower prices than the products that replace them. In our sample, there does not seem to be any clear pattern regarding the relative prices of the replacement prices.[12]

---

[12] In the 23 categories that we study, we examined the price changes of products for which the sampled product continued from period to period versus the price changes of products for which a replacement from the previous

We first summarize inflation rates in order to get an overall sense of biases. To do this, we divide our data into two time periods. The first period begins at the beginning of our data and ends at the start of the recession in October 2007. The second period covers the recession through end of the dataset November 2007 to December 2011.

Averaging across our products, using the change in the geometric mean price aggregates, prices rose 12.72 percent over the pre-recession period. We do identify a bias, in that unit value inflation exceeds geometric mean inflation by 1.9 percentage points over the period, about 0.3 percentage points per year. During the post-recession period, the bias reverses. Overall prices rise 19.9 percent using the change in geometric mean price aggregates, but geometric mean inflation exceeds unit value inflation by 0.9 percentage points, about 0.2 percentage points per year.

The reversal itself is not very surprising. Over the long-run, we would expect that all prices would be co-integrated so that inflation by all measures would be comparable. Over shorter intervals, shopping and pricing patterns could fluctuate. So the relevance for inflation measurement would turn on the gaps over shorter periods, such as phases of the business cycle.

These biases are comparable to previously identified ones that have attracted considerable attention. For instance, the Boskin Commission report to the US Senate (Boskin et al. 1996) estimated that upper level substitution bias, which arises because consumers can switch across types of products such as apples and bananas when prices change, would be about 0.15 percent per year. The commission also estimated outlet substitution bias, which arises because the same product is often sold at discount stores for cheaper prices than in regular stores, to be about 0.1 percent per year. Indeed, the Federal Reserve officials have recently referred to estimates of inflation (as measured by the personal consumption expenditure deflator) as being meaningfully below its 2 percent target when the measured rate was 1.7.[13]

To examine the inflation patterns for all 23 categories more systematically, we calculate rolling monthly 12-month inflation rates as the log change in unit values. Recall that, in our model, the unit value is identically equal to a weighted average of the fixed-weight index and the best price. To examine the fit of our model, we estimate a specification based on equation (6). Specifically, we estimate the following equation using nonlinear least squares:

$$(7) \qquad \ln(unitvalue_t) - \ln(unitvalue_{t-12})$$

$$= \gamma + \ln(\alpha \, fixedweightagg_t + (1 - \alpha) bestprice_t)$$

$$- \ln(\alpha \, fixedweightagg_{t-12} + (1 - \alpha) bestprice_{t-12}) + \epsilon_t,$$

where ϵ is a postulated residual.

---

sampled price had to be undertaken due to product unavailability. In 10 of our 23 categories, new introductions had on average lower prices than the products they replaced and in the other 13, new introductions had higher prices.

[13] For one example, see Evans (2017).
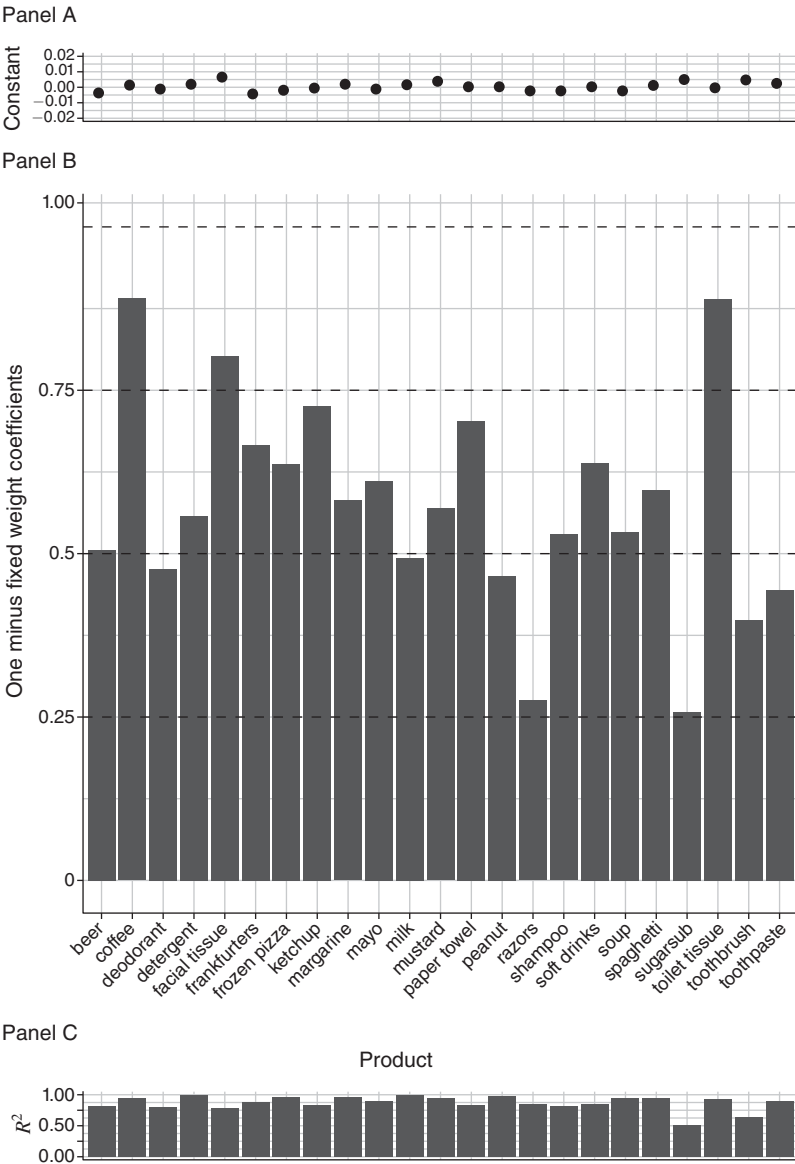
Panel A

Panel B

Panel C

FIGURE 6. NATIONAL ANALYSIS—STRUCTURAL ESTIMATES OF INFLATION USING FIXED WEIGHT
AND BEST PRICE AGGREGATES

Notes: Estimates are from the nonlinear least squares model described in text equation (7).
Panel B shows one minus the coefficient for the fixed-weight price index (corresponding to the
fit of the best price index). Panel A shows the constant term, which the model predicts should
be near zero. Panel C shows the R^2 of the regression.

The coefficient α is not constrained in our estimation procedure. If our model is
correctly specified, our estimate of α should be between zero and one and can be
interpreted as the share of Loyal consumers in the marketplace. The coefficient for
the constant term, γ, should be zero and the explanatory power of the specification
should be high. A summary of the results for all 23 products is given in Figure 6.

The figure shows the constant terms in panel A, 1 minus the alpha coefficients (that govern the importance of the best price) in panel B, and the R^2 values in panel C. The alpha coefficients range from 0.2 to 0.7 and average 0.4. Thus, the implied coefficient on the best-price index averages 0.6. The coefficient for the fixed-weight index is, in all cases, both statistically significantly different from zero and statistically significantly different from one. This means that $1 - \alpha$, the best-price coefficient, is significantly different from 0 in all cases. The constant terms are in general quite small and average 0.0002, although they are statistically significantly negative in some cases and positive in others. Finally, the overall explanatory power of the specifications is high. The minimum R^2 value is 0.53, but more than half are greater than 0.85.[14]

While the results in Figure 6 suggest that our model performs fairly well and that the best price is an important component to approximating unit value inflation, it is possible that the BLS-style geometric mean aggregate tracks unit value inflation adequately on its own. To examine this, we repeat the specification in (7), replacing the fixed-weight aggregate with the geometric mean aggregate. The results are presented in Figure 7 for all 23 products. As was the case in the store-level specification for coffee and peanut butter, the results for the geometric mean specifications are actually quite similar to the fixed-weight specifications. While the explanatory power of the 2 specifications are similar, it is interesting to note that, for 22 of the 23 products, the fixed-weight specification has a lower mean squared error and higher R^2 than does the corresponding geometric mean specification. That is, the model-implied specification fits better. Overall, in measuring unit value inflation, the inclusion of a best-price measure offers additional explanatory power over the geometric mean-type aggregate similar to that used by the BLS.

As discussed above, others have advocated the use of either unit values or the Tornqvist as an alternative to the BLS's geometric mean formulation at the lowest item stratum. Clearly, such proposals have been thus far unsuccessful due in part to difficulties in obtaining real-time, quantity data. Our proposed demand formulation fits the data fairly well and has the advantage of requiring only one extra price to be obtained at each visited store from the enumerator. Thus, it is relatively simple to construct.

V. Conclusion

We provide a simple model of consumer heterogeneity and show how that heterogeneity can motivate temporary price discounts by retailers. Margins that vary dramatically over time, even when consumer preferences are stable, are a natural outcome of our model. By the very nature of second-degree price discrimination, the seller optimally sets prices to encourage deal-driven brand switching only by those consumers who have the least disutility from switching brands. Indeed, in our stark model, only those with no disutility from switching brands actually switch them. Our empirical findings suggest that many consumers have a low disutility of

---

[14] We obtain similar results with alternative specifications. For example, we also examine specifications with a Newey-West correction for the standard errors and find results consistent with those reported here.
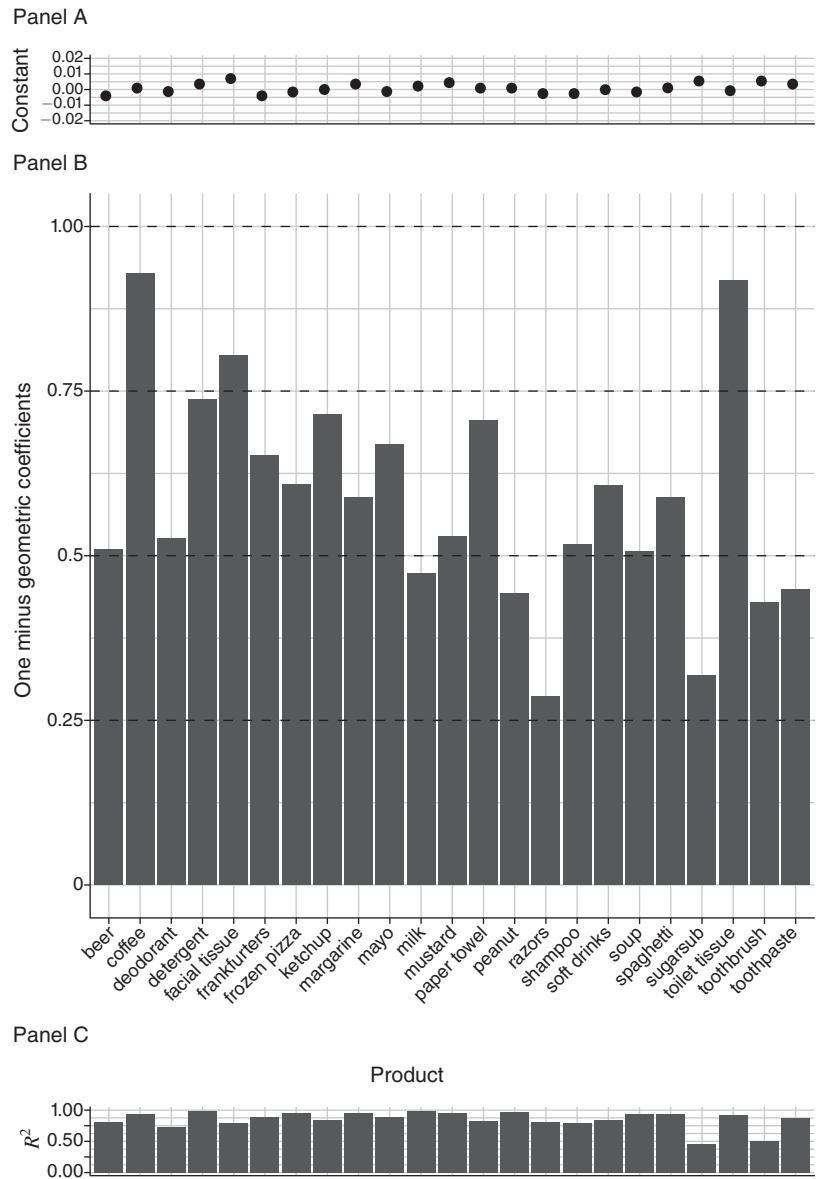
Panel A

Panel B

Panel C

FIGURE 7. NATIONAL ANALYSIS—ESTIMATES OF INFLATION USING GEOMETRIC MEAN AND BEST PRICE AGGREGATES

Notes: Estimates from nonlinear least squares model are described in text equation (7), replacing the fixed-weight aggregate with the geometric mean. Panel B shows 1 minus the coefficient for the geometric mean price index (corresponding to the fit of the best price index). Panel A shows the constant term. Panel C shows the R^2 of the regression.

switching brands. We show that the share of all goods in our sample that are sold at sale prices is much larger than the number of product-weeks in which sales occur.

We demonstrate that, from our model, the unit value emerges as the exact index; changes in the unit value index are associated with changes in consumer utility

as long as storage costs are negligible. We show that our model can be used as a structural model of prices paid. In particular, even without high-frequency, quantity data, unit values can be approximated as a weighted average of the fixed-weight index and the "best price." If approximating the unit value index is the goal, the best price is a crucial input.

This raises two questions that we begin to address here. First, what weights should a statistical agency use if attempting to construct a weighted average of the best-price and the fixed-weight index? Our parameter estimates vary by city and product. Thus, we suggest that weights could be approximated using short time series of quantity data not obtained in real time, essentially as we have done. Perhaps one criticism of our approach is that approximating the weights in the way we have proposed would be difficult or unacceptable and that it would be difficult to implement a weighting system that varies across products and cities. The limitation that the "best-fit" model would vary across cities and products is, in fact, not unique to our formulation. In our store-level section, we demonstrated that the "best-fit" CES demand system also varied across cities and products. This implies that the geometric mean formulation used by the BLS is a much better approximation for some products than for others. Time and cost limitations could determine the extent to which a "one-size-fits-all" best-price parameter could be used, just as the geometric mean methodology employs a "one-size-fits-all" CES demand parameter of one.

Second, is this approach truly practical for a statistical agency? In cases of physical collection, our approach only requires the enumerator to record a second price on each trip to the store. This strikes us as not very costly, while providing significantly more information about a price that is relevant to a very large number of consumers. As the BLS has shown by adopting a similar methodology for airline tickets, this methodology can also be useful in an ecommerce setting in which high-frequency data on prices are readily available but quantity estimates are only infrequently available. We believe that the BLS can construct price relatives using the exact methodologies that they currently do, but allowing a weight on the best price relative as well as the price relatives for randomly selected items that they selected using their current methodologies. While we exclude generics and premium products in our analysis, the BLS ought to be able to handle them precisely as they currently do in calculating price relatives for the randomly selected item.[15]

Finally, our evidence suggests that the disconnect between unit values and BLS-style indices remains when inflation is computed. Inflation in unit values will depend on shopping and discounting strategies, while a BLS-style inflation indicator will not fully capture these effects. Hence, the gap in measured inflation rates for these two constructs will vary whenever the depth or frequency of discounts changes. Prior research suggests that business cycle conditions and the level of inflation impact the depth and frequency of sales. Our results

---

[15] One subtle issue is the challenge that generics pose for the best-price calculation. If included, generics would frequently constitute the best price. One shortcut that could be examined, building off of the results in Feenstra and Shapiro (2003), is to create a strategy that utilizes promoted prices in lieu of finding the actual lowest price. While the best price in a category often corresponds to the promoted price, substantial transactions occur for products with heavily promoted prices even if they aren't the best price.

suggest that a best-price indicator helps explain unit value inflation in a parsimonious way.

Clearly, the importance of strategic consumer responses to temporary sales is of paramount importance in some sectors, and of more limited importance in others. While we agree that this kind of phenomenon would not be present for some types of goods and services, it definitely is relevant for many other products. The following categories of products account for about 44 percent of the CPI: food and beverages, household furnishings and operations, apparel, new and used vehicles, motor vehicles parts and equipment, motor vehicle maintenance and repair, airline fares, medical care commodities, recreation, college tuition and fees, technical and business school tuition, information and information processing, and tobacco and smoking products. In each of these categories, discounting is routine so that the kind of issues we have identified could apply.

Furthermore, as noted by Varian (1989), sellers almost always *want* to engage in price discrimination and price discrimination schemes involve substantial computational costs. Both the consolidation of the retailing sector over the last decades and the rapid decline in IT costs suggest that data-driven price discrimination schemes are likely to become more rather than less important in the future. Thus, if economists are to successfully model price-setting, and statistical agencies are to successfully measure prices and inflation, confronting price discrimination appears to be an inevitable challenge.

<h1 style="text-align:center">REFERENCES</h1>

Aguiar, Mark, and Erik Hurst. 2007. "Life-Cycle Prices and Production." *American Economic Review* 97 (5): 1533–59.

Basker, Emek. 2015. "Change at the Checkout: Tracing the Impact of a Process Innovation." *Journal of Industrial Economics* 63 (2): 339–70.

Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and D. Jorgenson. 1996. *Toward a More Accurate Measure of the Cost of Living*. Advisory Commission to Study the Consumer Price Index. Washington, DC, December.

Bradley, Ralph, Bill Cook, Sylvia G. Leaver, and Brent R. Moulton. 1997. "An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI." In *Proceedings of the Third Meeting of the International Working Group on Price Indices*, edited by Bert M. Balk, 169–82. Voorburg: Statistics Netherlands.

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela. 2008. "Database Paper—The IRI Marketing Data Set." *Marketing Science* 27 (4): 745–48.

Caves, Douglas W., Laurits R. Christensen, and Walter Erwin Diewert. 1982. "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers." *Economic Journal* 92 (365): 73–86.

Chessa, Antonio G., Johan Verburg, and Leon Willenborg. 2017. "A Comparison of Price Index Methods for Scanner Data." https://www.bundesbank.de/Redaktion/EN/Downloads/Bundesbank/Research_Centre/Conferences/2017/2017_05_10_ottawa_group_07_1_paper.html?__blob=publicationFile.

Chevalier, Judith A., and Anil K Kashyap. 2019. "Best Prices: Price Discrimination and Consumer Substitution: Dataset." *American Economic Journal: Economic Policy* https://doi.org/10.1257/pol.20150362.

Chevalier, Judith A., and Anil K Kashyap. 2014. "Best Prices: Price Discrimination and Consumer Substitution." National Bureau of Economic Research (NBER) Working Paper 20768.

Chevalier, Judith A., Anil K Kashyap, and Peter E. Rossi. 2003. "Why Don't Prices Rise during Periods of Peak Demand? Evidence from Scanner Data." *American Economic Review* 93 (1): 15–37.

Davies, George R. 1924. "The Problem of a Standard Index Number Formula." *Journal of the American Statistical Association* 19 (146): 180–88.

Davies, George R. 1932. "Index Numbers in Mathematical Economics." *Journal of the American Statistical Association* 27 (177): 58–64.

Diewert, W. Erwin. 1995. "Axiomatic and Economic Approaches to Elementary Price Indexes." National Bureau of Economic Research (NBER) Working Paper 5104.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo. 2011. "Reference Prices, Costs, and Nominal Rigidities." *American Economic Review* 101 (1): 234–62.

Ellison, Glenn. 2005. "A Model of Add-On Pricing." *Quarterly Journal of Economics* 120 (2): 585–637.

Evans, Charles L. 2017. "The Times They are a-Changin." Speech, DZ Bank-OMFIF International Capital Markets Conference, Frankfurt, March 29, 2007. https://www.chicagofed.org/publications/speeches/2017/03-29-2017-times-are-a-changin-charles-evans-frankfurt.

Feenstra, Robert C., and Matthew D. Shapiro. 2003. "High-Frequency Substitution and the Measurement of Price Indexes." In *Scanner Data and Price Indexes*, edited by Robert C. Feenstra and Matthew D. Shapiro, 123–50. Chicago: University of Chicago Press.

Fisher, Franklin M., and Zvi Griliches. 1995. "Aggregate Price Indices, New Goods, and Generics." *Quarterly Journal of Economics* 110 (1): 229–44.

Griffith, Rachel, Aviv Nevo, Ephraim Leibtag, and Andrew Leicester. 2009. "Consumer Shopping Behavior: How Much Do Consumers Save?" *Journal of Economic Perspectives* 23: 99–120.

Handbury, Jessie, Tsutomu Watanabe, and David E. Weinstein. 2013. "How Much Do Official Price Indexes Tell Us About Inflation?" National Bureau of Economic Research (NBER) Working Paper 19504.

Hosken, Daniel, and David Reiffen. 2007. "Pricing Behavior of Multiproduct Retailers." *B.E. Journal of Theoretical Economics* 7 (1): Article 23.

Ivancic, Lorraine, W. Erwin Diewert, and Kevin J. Fox. 2009. "Using a Constant Elasticity of Substitution Index to estimate a Cost of Living Index: from Theory to Practice." University of New South Wales Australian School of Business School of Economics Discussion Paper 2010/15.

Kehoe, Patrick J., and Virgiliu Midrigan. 2010. "Prices Are Sticky After All." National Bureau of Economic Research (NBER) Working Paper 16364.

Kryvtsov, Oleksiy, and Nicolas Vincent. 2014. "On the Importance of Sales for Aggregate Price Flexibility." Bank of Canada Working Paper 2014–45.

Lal, Rajiv, and Carmen Matutes. 1994. "Retail Pricing and Advertising Strategies." *Journal of Business* 67 (3): 345–70.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and product quality." *Journal of Economic Theory* 18 (2): 301–17.

Nakamura, Alice O., W. Erwin Diewert, John S. Greenlees, Leonard I. Nakamura, and Marshall B. Reinsdorf. 2015. "Sourcing Substitution and Related Price Index Biases." In *Measuring Globalization: Better Trade Statistics for Better Policy*, edited by Susan N. Houseman and Michael J. Mandel, 21–88. Kalamazoo: W. E. Upjohn Institute for Employment Research.

Nakamura, Emi, and Jón Steinsson. 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics* 123 (4): 1415–64.

Nakamura, Emi, Jón Steinsson, Patrick Sun, and Daniel Villar. 2016. "The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation." National Bureau of Economic Research (NBER) Working Paper 22505.

Nakamura, Leonard I. 1998. "The retail revolution and food-price mismeasurement." *Business Review* 23 (3): 3–14.

Nakamura, Leonard I. 1999. "The Measurement of Retail Output and the Retail Revolution." *Canadian Journal of Economics* 32 (2): 408–25.

Pashigian, B. Peter. 1988. "Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing." *American Economic Review* 78 (5): 936–53.

Pesendorfer, Martin. 2002. "Retail Sales: A Study of Pricing Behavior in Supermarkets." *Journal of Business* 75 (1): 33–66.

Redding, Stephen J., and David E. Weinstein. 2016. "A Unified Approach to Estimating Demand and Welfare." National Bureau of Economic Research (NBER) Working Paper 22479.

Reinsdorf, Marshall B. 1999. "Using Scanner Data to Construct CP1 Basic Component Indexes." *Journal of Business and Economic Statistics* 17 (2): 152–60.

Reinsdorf, Marshall B. 2003. "High-Frequency Substitution and the Measurement of Price Indexes: Comment." In *Scanner Data and Price Indexes*, edited by Robert C. Feenstra and Matthew D. Shapiro, 146–50. Chicago: University of Chicago Press.

Reinsdorf, Marshall, and Jack E. Triplett. 2009. "A Review of Reviews: Ninety Years of Professional Thinking About the Consumer Price Index." In *Price Index Concepts and Measurement*, edited by W. Erwin Diewert, John S. Greenlees, and Charles R. Hulten, 17–83. Chicago: University of Chicago Press.

Salop, Steven C., and Joseph E. Stiglitz. 1982. "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents." *American Economic Review* 72 (5): 1121–30.

Shapiro, Matthew D., and David W. Wilcox. 1997. "Alternative Strategies for Aggregating Prices in the CPI." *Federal Reserve Bank of St. Louis Review* 79: 113–25.

Sobel, Joel. 1984. "The Timing of Sales." *Review of Economic Studies* 51 (3): 353–68.

Triplett, Jack E. 2003. "Using Scanner Data in Consumer Price Indexes. Some Neglected Conceptual Considerations." In *Scanner Data and Price Indexes*, edited by Robert C. Feenstra and Matthew D. Shapiro, 151–62. Chicago: University of Chicago Press.

Varian, Hal R. 1980. "A Model of Sales." *American Economic Review* 70 (4): 651–59.

Varian, Hal R. 1989. "Price discrimination." In *Handbook of Industrial Organization*, Vol. 1, edited by Richard Schmalensee and Robert Willig, 597–654. Amsterdam: North-Holland.

Walsh, Correa Moylan. 1921. *The Problem of Estimation: A Seventeenth-Century Controversy and Its Bearing on Modern Statistical Questions, Especially Index-Numbers.* London: P. S. King and Son.