High-Frequency Substitution and

the Measurement of Price Indexes

by

Robert C. Feenstra,
University of California, Davis
and NBER

Matthew D. Shapiro,
University of Michigan
and NBER

Revised, February 2001

Prepared for the CRIW conference "Scanner Data and Price Indexes," Sept. 15-16, 2000. The authors thank Roger Betancourt, Walter Oi, Marshall Reinsdorf, Julio Rotemberg and Jack Triplett for very helpful comments.

High-Frequency Substitution and the Measurement of Price Indexes

ABSTRACT

This paper investigates the use of high-frequency scanner data to construct price indexes. In the presence of inventory behavior, *purchases* and *consumption* by individuals differ over time. Cost-of-living indexes can still be constructed using data on purchases. For weekly data on canned tuna, the paper contrast two different types of price indexes: *fixed-base* and *chained* indexes. Only the former are theoretically correct, and in fact, the chained indexes have a pronounced upward bias for most regions of the U.S. This upward bias can be caused by consumers purchasing goods for inventory. The paper presents some direct statistical support for inventory behavior being the cause of the upward bias, though advertising and special displays also have a very significant impact on shopping patterns.

Robert C. Feenstra
Department of Economics
University of California, Davis
Davis, CA 95616
and NBER
rcfeenstra@ucdavis.edu

Matthew D. Shapiro
Department of Economics
University of Michigan
Ann Arbor MI 48109
and NBER
shapiro@umich.edu

1. Introduction

The availability of scanner data for a wide range of household products raises the possibility of improving the measurement of the Consumer Price Index (CPI). Scanner data have a number of potential advantages over price measurements based on survey sampling. Scanner data include the universe of products sold, whereas sampling techniques capture only a small fraction of the population. Scanner data are available at very high frequency, whereas the cost of survey sampling typically limits data to monthly or lower frequency. Finally, scanner data provide simultaneous information on quantities sold in addition to prices, while survey techniques typically collect separate data on price and quantity – typically at different frequencies and for different samples.

Ongoing research on using scanner data for measuring the CPI has attempted to mimic the CPI's monthly sampling frame and therefore abstracted from the high-frequency variation in prices and sales. Reinsdorf (1999), for example, uses either monthly unit-values, or the prices in the third week of each month, to construct *monthly* price indexes for coffee. The collection of prices on a single day, which are then used to construct monthly indexes, corresponds to current practice at the Bureau of Labor Statistics. This practice has, of course, been constrained by the fact that prices are not sampled at frequencies greater than one month, but this constraint is no longer relevant with scanner data.

This paper takes a step at toward using the higher-frequency data available from scanner data. It examines how consumer behavior at high-frequency, i.e. weekly purchases of canned tuna, affect the application of index number formula that have been typically been implemented for lower-frequency or time-average data.

Outside of price index research, it has been quite common to use the high-frequency variation in prices and sales available from scanner data. In the marketing literature, it is well recognized that a great deal of substitution occurs across weeks in response to changes in prices and advertising. For example, Van Heerde, Leeflang and Witting (1999) have found that store level data for tuna and toilet tissue contains a dip in sales in the weeks following a promotion, which is consistent with previous studies at the household level. There is also high substitution between different varieties of tuna, depending whether they are on sale or not. Given this evidence, it would be highly desirable to construct weekly price indexes in a way that takes this behavior into account.

In order to construct "true" or "exact" price indexes, we need to have a well-specified model of consumer demand, which includes the response to sales and promotions. Betancourt and Gautschi (1992) present a model that distinguishes between *purchases* and *consumption* by individuals; in the presence of inventory behavior, these differ over time. Only *purchases* are observed when using data from retail outlets, as we do. Despite this, we show in section 2 that by using the Betancourt and Gautschi framework, an exact price index can still be constructed that measures the true cost-of-living for an individual. This index must compare one planning horizon (e.g. a month or year) to another, and cannot be constructed by comparing one week to the next.

In Section 3, we introduce the data on canned tuna, which is drawn from the ACNielsen academic database. They consist of weekly data over 1993-94, for 316 varieties of tuna over 690 stores. In Sections 4 and 5, we examine how several price indexes perform using these high-frequency price and quantity data. We construct two different types of weekly price indexes. The first – a *fixed-base* index – compares each week in 1993 to the modal price in 1992, using as weights the average 1992 sales at the modal price. We consider different formulas for the price index, including the Laspeyres, Geometric, and Törnqvist. The fixed-based Laspeyres index corresponds to the arithmetic average of price relatives traditionally used in the CPI. The fixed-based Geometric index corresponds to the "geometric mean" formula now used to produce the elementary price indexes for the majority of the CPI.[1] The Törnqvist index uses changing expenditure weights to control for substitution among the goods. We calculate the fixed-base Törnqvist index using the average of the 1992 sales (at the modal price) and the current 1993 weekly sales as weights. Hence, it uses long-term (i.e., base period to present) price relatives. If we take one year as the planning horizon, then this formulation corresponds quite well to our theoretical model of section 2.

The second type of index we consider are *chained* formula, which update the weights continuously and cumulates period-by-period changes in the price indexes to get long-term changes. The chained Törnqvist constructs the week-to-week Törnqvist using average sales in adjacent weeks, and then cumulates these results.[2]

The fixed-base Törnqvist does not equal the chained Törnqvist in general, and for our sample of weekly tuna data, we find that the difference between these two indexes is rather large: the chained Törnqvist has a pronounced upward bias for most regions of the U.S.[3] The reason for this is that periods of low price (i.e. sales) attract high purchases only when they are accompanied by advertising, and this tends to occur in the *final weeks* of a sale. Thus, the *initial* price decline, when the sale starts, does not receive as much weight in the cumulative index as

---

[1] The BLS uses unweighted averages for both the arithmetic and geometric means of price relatives. The goods to be averaged are probability-sampled using expenditure weights. Given that we will use the universe of observations, we use base-period expenditure weights rather than probability sampling.

[2] Since Cobb-Douglas utility, which underlies the Geometric formula, implies constant expenditure shares, we do not compute a chained Geometric index.

[3] An upward bias of the chained index with high-frequency data has also been noted by Triplett (1999).

the final price *increase*, when the sale ends. The demand behavior that leads to this upward bias of the chained Törnqvist – with higher purchases at the end of a sale – means that consumers are very likely purchasing goods for inventory accumulation. The only theoretically correct index to use in this type of situation is a *fixed-base* index, as demonstrated in section 3. Thus, our empirical results reinforce our theoretical results in showing the validity of fixed-base indexes when using high-frequency data.

In Section 6, we directly investigate the extent to which the weekly purchases of tuna are consistent with inventory behavior. We find some statistical support for this hypothesis, more so in the Northern regions of the U.S. than in the South. We also find that advertising and special displays have a very pronounced impact on shopping patterns. Concluding remarks are given in section 7.

2. A Representative Consumer Model

The purchases of consumers from a retail outlet, as distinct from their consumption, has been modeled in a "household production" framework by Betancourt and Gautschi (1992). They have in mind any number of reasons why purchases differ from consumption, e.g. because the individual must spend time to transform the former to the latter. Here we focus on the inter-temporal decisions of a consumer purchasing a storable good, so that purchases and consumption differ due to inventory behavior. With this simplification, we initially summarize the two-stage decision problem presented by Betancourt and Gautschi, and then show how even sharper results can be obtained by considering a single-stage decision problem.

Suppose the consumer is making the purchases q_t of a single brand of tuna, over the planning horizon $t=1,\ldots,T$. Consumption of tuna over the same horizon is denoted by x_t, and the vector of purchases and consumption are $q = (q_1,\ldots,q_T)$ and $x= (x_1,\ldots,x_T)$. Purchases and consumption are related; for example, we might specify that the sums of each over the horizon are equal. This would not allow for the decay of items (e.g. losing them), or any other reason why the consumer might limit purchases even when the item is on sale. We capture the general relationship between purchases and consumption by the constraints $f(q, x) \leq 0$, where f is a vector of quasi-convex functions. Given consumption x, the individual then solves the first-stage problem:

$$\min_{q \geq 0} \sum_{t=1}^{T} p_t q_t \quad \text{subject to} \quad f(q, x) \leq 0 \, , \tag{1}$$

where the price of the item in period t is p_t, and the vector of prices is $p = (p_1, \ldots, p_T)$. It is assumed that consumers know the future prices with perfect foresight. The constraint set represented by $f(q, x) \leq 0$ includes the feasibility constraints (e.g., that time t consumption cannot exceed time t purchases plus storage), the effect of depreciation during storage, and so on.

Denote the solution to (1) as the costs C(p, x). As usual, the derivative of this cost function with respect to prices give the optimal level of purchases, $q^* = C_p(p, x)$. In the second-stage, the consumer maximizes utility subject to the constraint that these costs do not exceed the available income I:

$$\max_{x \geq 0} U(x, z) \quad \text{subject to} \quad C(p, x) \leq I \, , \tag{2}$$

where $z = (z_1, \ldots, z_N)$ is a vector of consumption of all other goods, which we take as exogenous.[4] By choosing N suitably large, this vector can include all goods that complement or substitute for canned tuna in all periods. Let us denote the optimal level of consumption obtained from (2) as $x^* = g(p, z, I)$. Then it follows that optimal purchases can be obtained as $q^* = C_p[p, g(p, z, I)]$.

[4] Note that income I is *net* of the cost of purchasing the goods z.

This slightly complex formula for purchases does not shows their relation to underlying utility, however, so we now consider a simpler derivation.

Consider combining (1) and (2) into a single-stage problem:

$$\min_{q,x\geq 0} \sum_{t=1}^{T} p_t q_t \quad \text{subject to } f(q, x) \leq 0 \text{ and } U(x, z) \geq \overline{U}, \tag{3}$$

where $\overline{U}$ is an exogenous level of utility. We can write the solution to (3) as an expenditure function, $E(p, z, \overline{U})$. Differentiating this function with respect to prices, and using the envelope theorem, we obtain optimal purchases $q^* = E_p(p, z, \overline{U})$. These must equal purchases computed from our two-stage results above, so that $E_p(p, z, \overline{U}) = C_p[p, g(p, z, E(p, z, \overline{U}))]$. Clearly, the single-stage problem gives a much simpler expression. In particular, the derivatives of the expenditure function are *fully observable* since they equal purchases. We might expect, therefore, that this information will be enough to "work back" and reveal enough properties of the expenditure function itself so as to construct a cost-of-living index, i.e. to determine the expenditure needed to achieve utility $\overline{U}$ at various prices. We now show that this is indeed the case, using some well-known results from the price index literature.

Let $\tau = 0,1$ denote two planning horizons, each of length T periods. For concreteness, we can say that the periods t denote weeks, and the planning horizons $\tau = 0,1$ are years. We then consider the problem of a consumer making weekly purchases in one year as compared to another. This formulation ignores the issue that at the end of the first year, the optimal purchases should depend on the prices in the beginning of the next year; by treating the two planning horizons as distinct, we are supposing that there is no overlap in the information used by the consumer to make decisions in one year versus the next. This is a simplification.

The price vectors p^τ differ across the years, as do the exogenous variables z^τ and the level of annual utility U^τ. We will specify that the expenditure function in year τ, $E(p^\tau, z^\tau, U^\tau)$, takes on a translog functional form over its price arguments:

$$\ln E = \alpha_0^\tau + \sum_{t=1}^{T} \alpha_t^\tau \ln p_t^\tau + \frac{1}{2} \sum_{s=1}^{T} \sum_{t=1}^{T} \gamma_{st} \ln p_s^\tau \ln p_t^\tau \qquad (4)$$

where, $\alpha_t^\tau = h_t(z^\tau, U^\tau)$, $t=0,1,\ldots,T$. (5)

Without loss of generality we can suppose that $\gamma_{st} = \gamma_{ts}$ in (4). The functions $h_t(z^\tau, U^\tau)$ in (5) are left unspecified, except for the requirement that the translog function is linearly homogeneous in prices, which is satisfied if,

$$\sum_{t=1}^{T} \alpha_t^\tau = 1 \text{ and } \sum_{t=1}^{T} \gamma_{st} = \sum_{t=1}^{T} \gamma_{ts} = 0. \qquad (6)$$

The first condition implies that the functions $h_t(z^\tau, U^\tau)$ must sum to unity over $t = 1,\ldots,T$, for $\tau = 0,1$. Additional properties on these functions can be imposed to ensure that the expenditure function is increasing in utility, and to obtain any desired properties with respect to the exogenous variables z^τ.

The formulation in (4)-(6) is quite general, and it is well-known that the translog function provides a second-order approximation to an arbitrary function around a point (Diewert, 1976). The form in which we have written the expenditure function emphasizes that changes in the exogenous variables z^τ and U^τ in (5) act as *shift parameters* to the function in (4). For example, changes in the value of the function $\alpha_0^\tau = h_0(z^\tau, U^\tau)$ have a neutral impact on the expenditure function in (4). More importantly, changes in value of α_t^τ, for t=1,…,T, have a *non-neutral* impact on the expenditure function in (4). The importance of this can be seen by differentiating the log of expenditure with respect to the log of prices p^τ, obtaining the share of annual expenditures spent on tuna in each week:

$$s_t^\tau \equiv \frac{p_t^\tau q_t^\tau}{\sum_{t=1}^{T} p_t^\tau q_t^\tau} = \alpha_t^\tau + \sum_{s=1}^{T} \gamma_{st} \ln p_s^\tau \quad . \tag{7}$$

Thus, changes in annual utility or in the exogenous variables z^τ, which affect α_t^τ, clearly have an impact on the share of expenditure spent on tuna each period. For example, the consumption of more beef might shift demand away from tuna in some periods. Seasonal effects on demand are incorporated too, because α_t^τ can change exogenously over time.[5] In summary, the expenditure function in (4)-(6) encompasses a very wide range of demand behaviour, across both products and time within the planning horizon.

We next need to specify how to measure the cost-of-living (COL). Normally, the cost-of-living index is measured as the ratio of expenditure needed to obtain a *fixed* level of utility at two different prices. In our application, we have the utility levels U^0 and U^1 in the two years, so which should be choose? We follow Caves, Christensen and Diewert (1982a,b) in considering a *geometric mean* of the ratio of expenditure levels needed to obtain each level of utility:

$$COL \equiv \left[\frac{E(p^1, z^1, U^1)}{E(p^0, z^1, U^1)} \; \frac{E(p^1, z^0, U^0)}{E(p^0, z^0, U^0)} \right]^{1/2} \quad . \tag{8}$$

[5] Seasonal effects in tuna purchases are found by Chevalier, Kashyap and Rossi (2000).

The first term on the right of (8) gives the ratio of annual expenditures need to obtain utility U^1, holding fixed the exogenous variable z^1 but with prices changing. Of course, the consumer does not actually face the prices p^0 with exogenous variables z^1, so the expenditure level $E(p^0, z^1, U^1)$ is not observed. Similarly, the second term on the right of (8) gives the ratio of annual expenditures needed to obtain utility U^0, holding fixed the exogenous variable z^0 and with prices changing. Again, the expenditure $E(p^1, z^0, U^0)$ is not observed.

Despite the fact that (8) consists partially of unobserved information, this geometric mean can indeed be measured with data on purchases and prices:

Theorem (Caves, Christensen and Diewert)

If the annual expenditure function takes the form in (4)-(6), and purchases are optimally chosen so that (7) holds, then the cost-of-living in (8) can be computed as a Törnqvist index:

$$COL = \exp\left[\sum_{t=1}^{T} \frac{1}{2}(s_t^0 + s_t^1)\ln(p_t^1 / p_t^0)\right]. \qquad (9)$$

We provide a brief proof in the Appendix. This result of Caves, Christensen and Diewert (1982a,b) demonstrates the generality of the Törnqvist index, in that it accurately measures the cost-of-living even when the "first order" parameters α_t^τ of the translog function are changing. In a producer context, such changes capture non-neutral technical change, while in our consumer context these changes can capture change in prices of exogenous commodities, seasonal effects, and even the effects of advertising if it shifts α_t^τ.

While our results so far were obtained for a single variety of tuna, purchased over time, they readily extend to multiple varieties. Thus, suppose that the price vector p^0 and p^1 in (8) include the prices of i = 1,…,N varieties over t = 1,…,T periods. Then the cost-of-living is still measured with a Törnqvist index, defined over varieties and time:

$$COL = \exp\left[\sum_{t=1}^{T}\sum_{i=1}^{N} \tfrac{1}{2}(s_{it}^{0} + s_{it}^{1})\ln(p_{it}^{1} / p_{it}^{0})\right], \quad (10)$$

where the expenditure shares are $s_{it}^{\tau} \equiv p_{it}^{\tau} q_{it}^{\tau} \Big/ \sum_{t=1}^{T}\sum_{i=1}^{N} p_{it}^{\tau} q_{it}^{\tau}$.

We shall refer to (10) as the "true" cost-of-living index, and contrast it with various other formulas traditionally used by the BLS. In order to implement any of these formulas, we need to decide what to use as the base period, when $\tau=0$. We will be interested in focusing on the effects of *sales* on consumer purchases, so we will choose the base period prices as the *mode* prices for each item in an initial year (i.e. the typical non-sale prices). Correspondingly, the expenditure share in the base period will be constructed using the average quantity at the modal price. It follows that our base period prices and expenditure shares will not differ over weeks, so we rewrite these as p_{i0} and s_{i0}. Then we can also drop the superscript "1" for the current year, and rewrite (10) simply as:

$$COL = \exp\left[\sum_{t=1}^{T}\sum_{i=1}^{N} \tfrac{1}{2}(s_{i0} + s_{it})\ln(p_{it} / p_{i0})\right]. \quad (10')$$

In the next section, we use summary statistics to begin to investigate the frequency of sales and advertising in the data for canned tuna, and the extent to which these affect demand. Price indexes are constructed in section 4, where we contrast (10') with alternative formulas. Finally, in section 5 we directly test for the influence of inventory behavior on demand.

3. Tuna Data

The data we shall use is taken from the ACNielsen academic database. It includes two years (1992-93) of weekly data, for 316 Universal Product Codes (UPC's) of canned tuna. There are 10 market areas, with a total of 690 stores; the smallest market area is Southwest (54 stores) and the largest is Northeast (86 stores). The data are drawn from a random sample of the large-scale ACNielsen SCANTRACK database. For each store, UPC product, and week, the database includes: the value of sales; quantity sold; and a host of marketing indicators. These indicators can be broken into two groups: advertising indicators, about whether there was a sale and what type of ads were used; and display indicators, about whether the product appeared in a special location within the store.

An example of the data for two actual products sold in one store, over the first six months of 1993, is shown in Table 1. We define the "typical" price for a product as the mode price in each year, which was 66¢ for product A and $1.29 for product B in 1993, as indicated at the bottom of Table 1. Both of these mode prices had fallen from the year before. We further define a "sale" as a week where the (average) price that week is at least 5% less than the annual mode price. The occurrences of sales are indicated in bold in Table 1. In some cases, a sale coincides with an advertisement for the product,[6] and these cases are shown in italics. Notice that for both products, there are several instances where the product first goes on sale *without* an advertisement, in which case the quantity does not increase by much, if at all. Following this, an ad occurs at the *end* of the sale, and this leads to a very marked increase in the quantity purchased. This particular pattern of purchases – which has a large increase at the end of the sale – is

[6] There are five different kinds of advertisements indicated in the database, such as featured ads, ads with coupons, etc., but we do not distinguish them. Similarly, there are a number of different kinds of displays, but we do not distinguish these in our analysis.

consistent with inventory behavior. When it occurs simultaneously with an ad, however, we will need to try and distinguish whether the behavior arises due to advance purchase and storage, or due to the information that consumers receive from the ad.

In Table 2 we report summary statistics of the data for three market areas: the Northeast (with 86 stores), Midwest (with 57 stores) and Southwest (with 54 stores), for 1992 and 1993. All values reported are averaged across the product and stores in each region. First, we report the modal prices for each year, ranging from $1.30 (in the Midwest, 1992) to $1.73 (in the Northeast, 1993). For each product, we then measure its price each week *relative to the mode for the year*. Sales are defined as a week with this "relative price" less than 0.95. The average value of this relative price, and unit-value (constructed as the sales-weighted average of the relative prices) are reported in the second the third rows. Naturally, the unit-values are below the average relative prices, indicating that consumers purchase more when prices are low. Following this, we report average prices (relative to the mode) and quantities (relative to the mean quantity at the mode price), during weeks with and without sales. Three cases are distinguished: (i) no display nor ad; (ii) a display but no ad; (iii) an advertisement (with or without a special display).

In weeks without sales, having either a display or an advertisement is seen to increase the quantity purchased by 1.5 – 2 times. Surprisingly, about the same impact is obtained from having a sale in the *absence* of both displays and ads. Larger impacts are obtained when either of these features accompanies a sales, and the combination of a sale and advertisement increases the quantity purchased by 6 – 13 times. Generally, sales occur in 15 – 23% of the weeks, and of these, somewhere between one-quarter and one-half of the sales last only one week, or last more than four weeks. Less than 1% of weeks *without* sales have ads, but 8 – 18% of the weeks *with* sales also have ads. At the bottom of Table 2 we report the frequency of such ads during sales: for sales lasting only one week, 57 – 67% have ads; for sales lasting longer, 6-27% have an ad in the *first* week, and 14-24% have an ad in the *last* week. In the Midwest, it is much more likely to see an ad at the *end* of a sale than the *beginning*, but the reverse holds in the Northeast, and there is no consistent pattern in the Southwest.

4. Formulas for Price Indexes

The individual tuna varieties (i.e. UPC codes) are denoted by the subscript i within each store. We will be using the modal price in 1992 as a "base period" price p_{i0}, and let q_{i0} denote the mean quantity purchased at that price in 1992. Then the Laspeyres index from the base period to the week t in 1993 is,

$$P_t^L \equiv \frac{\sum_i q_{i0} p_{it}}{\sum_i q_{i0} p_{i0}} = \sum_i w_{i0} (p_{it} / p_{i0}), \tag{11}$$

where the equality in (11) follows by defining the base period expenditure shares,

$w_{i0} \equiv p_{i0} q_{i0} / \sum_i p_{i0} q_{i0}$.

We will refer to (11) as a *fixed-base Laspeyre*s index. It can be distinguished from the *chained Laspeyres* which is constructed by first taking the week-to-week index,

$$P_{t-1,t}^L \equiv \sum_i w_{i0} (p_{it} / p_{it-1}), \tag{12}$$

using the same base period weights. The chained Laspeyres is then constructed by simply cumulating these week-to-week indexes:

$$CP_t^L \equiv CP_{t-1}^L \cdot P_{t-1,t}^L \quad , \quad \text{with}\, CP_0^L \equiv 1. \tag{13}$$

It is well known that the chained Laspeyres has an upward bias, because it does not satisfy the "time reversal" test.[7] For this reason BLS generally uses fixed-base formulas, constructed over the "long term relatives" p_{it}/p_{i0}. We will be constructing the chained Laspeyres for comparison purposes.[8]

An alternative to using the arithmetic mean in (11) is to use a weighted geometric mean of the prices for individual products. This results in the *fixed-base* Geometric index,

$$P_t^G = \exp\left[\sum_i w_{i0} \ln(p_{it} / p_{i0}) \right]. \tag{14}$$

Note that a the fixed-base Geometric formula in (14) would be identical to a chained version (constructed by defining and week-to-week geometric index $P_{t-1,t}^G$ and then cumulating). For this reason, we do not construct the chained Geometric.

The Laspeyres and Geometric indexes presume, respectively, zero and unit elasticity of substitution among varieties. To provide a better approximation to changes in the cost-of-living under more general assumption, we consider the superlative Törnqvist functional form. The fixed-based Törnqvist index is defined as,

---

7 Denoting any price index by $P(p_{t-1}, p_t)$, the "time reversal" test is satisfied if $P(p_{t-1}, p_t)P(p_t, p_{t-1}) = 1$. That is, when prices change from p_{t-1} to p_t and then back to p_{t-1}, we want the two-period chained index to be unity. However, this test is not satisfied for the Laspeyres formula in (12): it can be shown that $P^L(p_{t-1}, p_t)P^L(p_t, p_{t-1}) \geq 1$, so the index is upward biased.

8 An alternative formula for the chained Laspeyres would be to use the period t-1 weights in (12), so it becomes $\Sigma_i\, w_{it-1}\, (p_{it}/p_{it-1})$, which would then be cumulated as in (13). Results for this index are reported in note 12.

$$P_t^T = \exp\left[\sum_i \tfrac{1}{2}(w_{i0} + w_{it})\ln(p_{it} / p_{i0})\right], \qquad (15)$$

where $w_{it} \equiv p_{it}q_{it} / \sum_{i \in I_k} p_{it}q_{it}$ is the expenditure share of product i in week t.[9] It is important to compare this formula to the true cost-of-living index in (10'), which is also a Törnqvist formula: the only difference is that (10') is aggregated over varieties *and* time, whereas (15) is aggregated only over varieties, for a single week. If we average (15) over all the weeks in a year, then we would expect the result to be quite close to that calculated from (10').[10] Thus, an average of the Törnqvist indexes in (15) appears to be quite close to the true cost-of-living index in (10').

An alternative formulation of the Törnqvist is to first construct it on a week-to-week basis:

$$P_{t-1,t}^T = \exp\left[\sum_i \tfrac{1}{2}(w_{it} + w_{it-1})\ln(p_{it} / p_{it-1})\right], \qquad (16)$$

The *chained* Törnqvist is then obtained by cumulating (16):

$$CP_t^T \equiv CP_{t-1}^T \cdot P_{t-1,t}^T \quad , \quad \text{with}\, CP_0^T \equiv 1. \qquad (17)$$

The chained Törnqvist in (17) will generally *not equal* the fixed-base Törnqvist in (15), and therefore, we expect that the average values of the chained Törnqvist over a year might differ

---

[9] Note that the *fixed-base* formula does include current data in the expenditure weight.

[10] There is not an exact equality between taking a weighted average of (15) over all weeks in the year, versus computing (10') directly, because calculating a Törnqvist index in two stages is not the same as calculating it directly in one stage. This is shown by Diewert (1978), who nevertheless argues that "approximate" consistency between one-stage and two-stage Törnqvist indexes will obtain.

substantially from the exact index in (10'). Thus, we do not have the same justification for (17) as for the fixed-base index in (15), which we expect to be close to (10').[11]

5. Calculation of Price Indexes

We calculated the Laspeyres-ratio, Geometric, and Törnqvist indexes for 1993, using as the base period the mode price in 1992 and the average sales at that price. As an initial example, we show this calculation for the sample data over January – June 1993 in Table 1, with results in Table 3. Any of the fixed-base indexes have nearly the same values in the first week of January, and last week of June, because the prices for the two products were identical in those weeks (66¢ for product A and $1.29 for product B , respectively). The chained indexes, however, do not satisfy this property. The chained Laspeyres ends up with a value of 1.28, rising some 37% from its value in the first week of January. This is entirely due to the fact that the Laspeyres index does *not satisfy* "time reversal", so that when one product goes on sale and its price falls temporarily, the index *does not* return to its former value when the sale ends.

More surprisingly, the chained Törnqvist index shows an even greater upward bias, ending with a value of 1.45, which is nearly *twice* the value of the chained Laspeyres! The chained Törnqvist index does satisfy "time reversal" provided that the weekly expenditures are consistent with the maximization of a static (i.e. weekly) utility function. But this assumption is violated in the data for these two sample products: in periods when the prices are low, but there are no advertisements, the quantities *are not* high (see Table 1). Because the ads occur in the final period of the sales, the price *increases* following the sales receive much greater weight than the price *decreases* at the beginning of each sale. This leads to the dramatic upward bias of the

---

11 Alterman, Diewert and Feenstra (2000, chap. 4, Propositions 1,2) have identified some conditions under which a

chained Törnqvist. When averaged over all the weeks, the chained Törnqvist gives a value of 1.015 relative to the 1992 modal prices, and 1.144 relative to the 1993 model prices; both of these art substantially higher than the fixed-base Törnqvist and the other indexes.

The question arises as to whether this is a general feature of the data on canned tuna. To determine this, we report in Table 4 the values of prices indexes in 1993 computed for each store, and then averaged over the weeks in 1993 and over the ten regions of the U.S. The prices indexes are computed using either the modal prices in 1992 as the base, or the modal prices in 1993. In the first column of Table 4, we report the true COL index from (10'), constructed relative to each base, and averaged over all stores in each region. The values for this index show the drop in the cost-of-living (or conversely, the welfare gains) from having items periodically on sale during 1993. We are interested in comparing this true index to the others, so as to determine their bias.

From Table 4, we see that the fixed-base Laspeyres is always higher than the true index, and that the chained Laspeyres is considerably higher still.[12] Both of these are above the chained and fixed-base Törnqvist, respectively. In addition, the chained Törnqvist exceeds its fixed-base counterpart in many regions of the country: the upward bias of the chained Törnqvist is most apparent in the Northwest and Southwest, and occurs in seven out of the ten regions (all except the Midwest, South Southwest and Southeast). On the other hand, the average of the fixed-base

---

fixed-base Törnqvist index between two dates, and the chained index between the same dates, will be similar in magnitude. Since these indexes do not coincide in our data, the conditions they identify are not satisfied.

[12] When instead we use the alternative formula for the chained Laspeyres, described in note 8, then the upward bias of the index is much worse. This is because the weight w_{t-1} is much higher as the end of the sale than at the beginning, so the price *increase* at the end of the sale is given a much greater weight than the price *decrease* at the beginning. (This problem is ameliorated in the chained Törnqvist, because the weights are averaged over two periods.) For example, this alternative formula for the chained Laspeyres, then averaged over all weeks in 1993 and stores in a region, equals 20.6 for the Northeast , 2.1 for the Midwest, , and 3.1 for the Southwest. In one extreme case (a store in the East Northeast), the week-to-week Laspeyres calculated as in note 8 typically exceeds 1.2, so that the chained Laspeyres rises from unity to $1.2^{52} > 10{,}000$ during the 1993 year!

Törnqvist over the stores and weeks is quite close to the average of the true COL index over stores in each region. This result was expected, because the COL index itself was a Törnqvist index computed over product varieties and time, as in (10'), whereas our fixed-base Törnqvist has been computed over product varieties, and then averaged across weeks. Thus, they differ only in their respective weights.

The difficulty with using the true COL index (10') in practice is that it compares one planning horizon (e.g. a year) to another whereas the BLS may very well need to report price indexes at higher frequency (i.e. monthly). The fixed-base Törnqvist more than meets this requirement, since it constructed at weekly intervals. Furthermore, as we have shown, the average of the fixed-base Törnqvist is empirically quite close to the COL index. These results therefore lend support to fixed-base Törnqvist, even when applied to high frequency scanner data. Conversely, the upward bias of the chained Törnqvist makes it highly inappropriate to use at high frequency, and it appears that this bias is due to inventory behavior. To confirm this, it would be desirable to have some independent evidence on such behavior, as we explore econometrically in the next section.

6. Estimation of Inventory Behavior

To determine how demand for tuna responds to prices, we need to adopt a specific functional form. The static CES specification of the utility function leads to a demand curve for a variety of tuna as follows:

$$x_{it} = (p_{it} / P_t)^{-\rho} \tag{18}$$

where x_{it} is consumption (relative to some base), p_{it} is the price of a variety i and P_t is the price index for tuna at time t.

We have experimented with developing a full-blown model of inventory behavior for consumers, but it quickly gets very complicated. Simple models have the following implications:

- Consumers buy for future consumption when goods are on sale.
- Consumers will buy more when the next sale is more distant.
- If there is a cost of storage, consumers will defer purchases for storage until the last period of the sale.
- Sales are asymmetric: Consumers might want to sell back some of their inventory when prices are unusually high (a negative sale), but they cannot.

To make this concrete, consider the following formulation. Suppose that there is a per-period storage cost of *s* units of tuna. This would include depreciation or loss in storage, the shadow price of shelf space in the pantry, and interest.[13] Suppose that there is a sale – defined as a substantially lower-than-normal price, perhaps accompanied by an advertisement. The consumer expects the next sale to be H periods in the future. Then a consumer will purchase now to fulfill future demand. The shadow price of consumption h periods ahead for a variety i put into storage at the time of the sale t is $p_{it}(1+s)^h$. Assuming that the cost of storage (s) and that the time to the next sale (H) are not too high, the consumer will purchase sufficient quantity for all future needs until the next sale. Hence, quantity sold at the time of the sale will be

$$q_{it} = \sum_{h=0}^{H} \left(\frac{p_{it}(1+s)^h}{P_t} \right)^{-\rho} = \left(\frac{p_i}{P_t} \right)^{-\rho} \left(\frac{1-(1+s)^{-\rho(H+1)}}{1-(1+s)^{-\rho}} \right) . \qquad (19)$$

[13]There is a problem with units of measurement: depreciation and loss is in units of tuna, whereas interest and storage costs are in units of the numeraire.

If the shadow price $p_{it}(1+s)^h$ exceeds some future price $p_{i,t+H'}$ prior to the next sale, then the process is truncated at H'. If s is small, then the term in the power of H simplifies to be simply (H+1) itself, since by L'Hospital's rule:

$$\lim_{s \to 0} \left(\frac{1-(1+s)^{-\rho(H+1)}}{1-(1+s)^{-\rho}} \right) = H+1 \,. \tag{20}$$

In order to estimate (19), we take natural logs and make use of (20). We include both current prices, as well as leads and lags up to length L, obtaining the estimating equation:

$$\ln q_{it} = \beta_0 + \sum_{\ell=-L}^{L} \beta_{1\ell} \ln p_{it+\ell} + \sum_{\ell=-L}^{L} \beta_{2\ell} \ln P_{t+\ell} + \beta_3 \ln(1+H_{own,it}) + \beta_4 \ln(1+H_{any,t}) + \varepsilon_{it} \tag{21}$$

where q_{it} are weekly sales measured relative to the quantity at the mode price; p_{it} is the price in that week relative to its mode for the year; P_t in the fixed-base Törnqvist index for that store; $H_{own,it}$ is the number of weeks to the next sale of this product i; and $H_{any,t}$ is the number of weeks to the next sale of any variety of canned tuna, in that store. The inclusion of leads and lags for p_{it} and P_t (up to length L) allows for intertemporal substitution in consumption, as potentially distinct from inventory behavior. Note that the variables $\ln(1+H_{own,it})$ and $\ln(1+H_{any,t})$ are nonzero *only when* it is the last week of a sale; otherwise, they are not relevant to the inventory problem.

Estimates of (21) for each region of the U.S., over all weeks in 1993, are reported in Table 5. In the first set of estimates for each region we report the coefficients of (21), along with their standard errors. In the second set of estimates, we extend (21) to allow for indicator variables indicated whether that variety of tuna had a special display, whether it was advertised, and also an interaction term between advertising and the price of that variety relative to its mode. There are nearly 50,000 observations or more for each region, which pools over varieties of tuna, weeks and stores.[14]

Estimation is by ordinary least squares, including fixed-effects for each store, as recommended by Betancourt and Malanoski (1999).[15] We do not report the coefficients on the store fixed-effects, and we also do not report the coefficients on the lead and lag values of p_{it} and P_t. The inclusion of these leads and lag often increased the (absolute) values of the concurrent price elasticities, and the leads and lags themselves were sometimes significant though not always of positive sign.[16] Most importantly, the inclusion of the leads and lag of prices has little impact on the coefficients on the inventory terms $\ln(1+H_{own,it})$ and $\ln(1+H_{any,t})$: the estimates reported in Table 5 are for a single lead and lag, L=1, but similar results are obtained for L=0 or L=2. Coefficients on $\ln(1+H_{own,it})$ and $\ln(1+H_{any,t})$ that are significantly different from zero at the 5% level are indicated in bold.

Strong evidence of inventory behavior is found for the Northeast regions, at the top of Table 5: the East Northeast has a coefficient of 0.35 on $\ln(1+H_{own,it})$ while the Northeast has a

---

[14] In some cases the value of $H_{own,it}$ could not be measured, because the next sale of that variety was after the end of the sample, so these observations were omitted. Less frequently, the value of $H_{any,t}$ could not be measured because the next sale of any variety was after the end of the sample; these values of $H_{any,t}$ were set equal to zero.

[15] We considered using an instrumental variables estimator to take into account less than perfect foresight for H, but we obtained inadequate first-stage fits.

[16] Betancourt and Gautschi (1992) show generally that for retail purchases (as contrasted with consumption) there is a tendency to obtain complementarity rather than substitution in demand. This might explain the cross-price elasticities that were sometimes significantly negative.

coefficient of 0.33 on $\ln(1+H_{any,t})$. To interpret these, we can measure the total "inventory effect" on demand during the last week of a sale as:

$$\text{Inventory Effect} = \hat{\beta}_3 \; \overline{\ln(1+H_{own,it})} + \hat{\beta}_4 \; \overline{\ln(1+H_{any,it})} \quad . \tag{22}$$

Note that the sample average value of $\ln(1+H_{own,it})$ when this variable is positive is 2.4 (so the next sale of each product is 10 weeks away), and the average value of $\ln(1+H_{any,t})$ when this variable is positive is 1.2 (so the next sales of any product is 2.3 weeks away).

Using the coefficients in Table 5, for the East Northeast the "inventory effect" is exp(0.35*2.4-0.08*1.2)=2.2, indicating that the quantity demanded during the last week of a sale is more than twice as high as average. For the Northeast region, the "inventory effect" equals exp(0.03*2.4+0.22*1.2)=1.4, so demand is 40% higher at the end of sale. Other regions that show particularly strong inventory behavior are the Upper Midwest, for both inventory variables, and most regions of the South, for the variable reflecting sales in other varieties.

However, when we add the indicator variables for displays and advertising, along with the interaction between advertising and price, then the magnitude of inventory behavior is substantially reduced in all regions. In the cases where there is still some evidence of inventory behavior – such at the East Northeast and Southeast – a positive coefficient on one of the inventory variables is offset by a negative coefficient on the other. Indeed, when advertising is included then the only region that retains significant evidence of inventory behavior (without an offsetting negative effect) is the Northeast. As an example, in one store in that region a certain tuna product fell in price from $1.59 to 88¢ in one week and sales went from about 100 cans average to 20,000 in that week! This is the largest demand response in our dataset, and almost surely indicates that the purchases were for inventory. At the same time, we cannot rule out that some portion of the increased demand was in response to the *advertised* price of 88¢. Generally, when we take into account displays and advertising in Table 5, the extent of inventory behavior is reduced markedly.[17]

While these inventory regressions provide some direct evidence of inventory behavior, we also wish to know whether this can explain the upward bias in the chained Törnqvist index. To this end, in Figure 1 we graph the "inventory effect" against the index bias, measured as the difference between the chained Törnqvist and the true COL index (where both of these are averaged over all weeks in 1993, and using the modal price in 1993 as the base). The "inventory effect" in (22) is measured using the coefficients on the first row for each region in Table 5; that is, ignoring the advertising and display variables. The means for $\ln(1+H_{own,it})$ and $\ln(1+H_{any,t})$ in (22) are now computed over the entire sample, i.e. for both positive and zero observations. This will capture not only the average value of these variables when positive, but also the number of times that sales occur. We graph the average "inventory effects" against the index bias for the ten regions in Figure 1, for the 580 individual stores over which (22) could be estimated in Figure 2.

In Figure 1, there is only a weak positive relation between the "inventory effect" and the index bias; the correlation between these variables is 0.05. The Southwest and Northwest have the highest bias of the chained Törnqvist index, and these both have coefficients of about 0.15 on $\ln(1+H_{any,t})$ in Table 5: while these effects are significant, they are not the largest coefficients

[17] We have also re-estimated (15) while *excluding* all one-week sales. This allows us to determine what inventory behavior is associated with multi-week sales. Generally, the coefficients we obtain on $\ln(1+H_{own,it})$ and $\ln(1+H_{any,t})$ are lower than those reported in Table 5, which combines the one-week and multi-week sales.

that we find on inventory behavior. Conversely, the East Northeast region is shown as having the highest "inventory effect" in Figure 1, and while it has a non-negligible bias of the chained Törnqvist index in Table 4, this bias is not the largest across regions.

However, when we look across individual stores in Figure 2, the evidence for a positive relationship between inventory behavior and index bias is more apparent. The correlation between these two variables is 0.12, which is significantly different from zero at the 1% level (with N=580). Thus, there are some stores that display both strong inventory behavior and a pronounced upward bias of the chained Törnqvist index. The three stores shown in Figure 2 whose index bias exceeds unity are all in the Northeast and East Northeast regions; and in addition, 20 out of the top 25 stores with highest "inventory effects" are also in these two regions. Generally, the shopping patterns of the Northeast regions show marked inventory behavior and an upward bias of the chained Törnqvist, supporting the idea that such behavior causes the upward bias.

7. Conclusions

The data on tuna shows substantial high-frequency variation in price and substantial response of consumer demand to this variation in price, suggesting inventory behavior. A true cost-of-living index in this context, as derived in section 2, must compare all prices over one planning horizon to all prices in another, e.g. compare one year to the next. This differs from the conventional approach taken at the BLS, which is to compute price indexes in each month. Averaging over a month, as the BLS does, is a step toward aligning price measurement with the consumption rather than the shopping period. Yet, the month might not be the correct planning horizon. Moreover, even if it were, the results of section 2 show that the arithmetic average of prices is not the correct summary statistic to input into a cost-of-living index.

We find that the fixed-base Törnqvist, computed weekly and then averaged over a year, can adequately measure the true COL index (which is itself a Törnqvist formula). That is, the fixed-base Törnqvist captures the reduction in the cost-of-living that arises when consumers economize by substituting toward goods whose price is low. Conversely, the chained Törnqvist gives too much weight to price increases that follow the end of sales, and is upward biased.

The upward bias of the chained Törnqvist can be explained by purchases for storage rather than consumption. During sales, some of the increase in demand corresponds to purchases for storage, as supported by our regression results. In particular, we find that purchases are increasing in time to the next sale. This finding is consistent with a forward-looking consumer engaging in storage. This evidence of forward-looking behavior is somewhat undermined by accounting for advertisements. Nevertheless, we find a link between inventory behavior – especially in the Northeast – and the upward bias of the chained Törnqvist. It follow that the chained approach is to be avoided when using high frequency scanner data, and a fixed-base Törnqvist (or the true COL index) should be used instead.

Appendix

<u>Proof of Theorem:</u>

Taking the log of (8), we obtain,

$$\ln(COL) = \frac{1}{2}[\ln E(p^1, z^1, U^1) - \ln E(p^0, z^1, U^1) + \ln E(p^1, z^0, U^0) - \ln E(p^0, z^0, U^0)]$$

$$= \frac{1}{2}\left[\sum_{t=1}^{T} (\alpha_t^1 + \alpha_t^0)\ln(p_t^1 / p_t^0) + \sum_{s=1}^{T}\sum_{t=1}^{T} \gamma_{st} \ln p_s^1 \ln p_t^1 - \sum_{s=1}^{T}\sum_{t=1}^{T} \gamma_{st} \ln p_s^0 \ln p_t^0\right]$$

$$= \frac{1}{2}\left[\sum_{t=1}^{T} (\alpha_t^1 + \alpha_t^0)\ln(p_t^1 / p_t^0) + \sum_{s=1}^{T}\sum_{t=1}^{T} \gamma_{st} (\ln p_s^1 + \ln p_t^0)(\ln p_t^1 - \ln p_t^0)\right]$$

$$= \frac{1}{2}\left[\sum_{t=1}^{T} (s_t^1 + s_t^0)\ln(p_t^1 / p_t^0)\right]$$

where the second line follows by using the translog formula in (4) and (5), the third line using simple algebra, and the final line follows from the share formula in (7).

References

Alterman, William F., W. Erwin Diewert, and Robert C. Feenstra (2000) *International Price Indexes and Seasonal Commodities*. U.S. Dept. of Labor, Bureau of Labor Statistics, Washington, D.C.

Betancourt, Roger R., and David Gautschi (1992) "The Demand for Retail Products and the Household Production Model: New Views on Complementarity and Substitution," *Journal of Economic Behavior and Organization,* 17, 257-275.

Betancourt, Roger R., and Margaret Malanoski (1999) "An Estimable Model of Supermarket Behavior: Price, Distribution Services and Some Effects of Competition," *Empirica,* 26, 55-73.

Caves, Douglas W., Laurits R. Christensen and W. Erwin Diewert (1982a) "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity," *Econometrica*, 50(11), November, 1393-1414

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

Chevalier, Judith A., Anil K. Kashyap and Peter E. Rossi (2000) "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," National Bureau of Economic Research working paper no. 7981.

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

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

Reinsdorf, Marshall B. (1999) "Using Scanner Data to Construct CPI Basic Component Indexes, *Journal of Business and Economic Statistics,* 17(2), April, 152-160.

Triplett, Jack (1999) "Elementary Indexes for a Consumer Price Index," in Walter Lane, ed., *Proceedings of the Fourth Meeting of the International Working Group on Price Indices,"* U.S. Dept. of Labor, Bureau of Labor Statistics, Washington, D.C.

van Heerde, Harald J., Peter S.H. Leeflang and Dick R. Witting (1999) "The Estimation of Pre- and Post-promotion Dips with Store-level Scanner Data," *Journal of Marketing Research,* forthcoming.

Figure 1: Index Bias and Inventory Effect - Regions

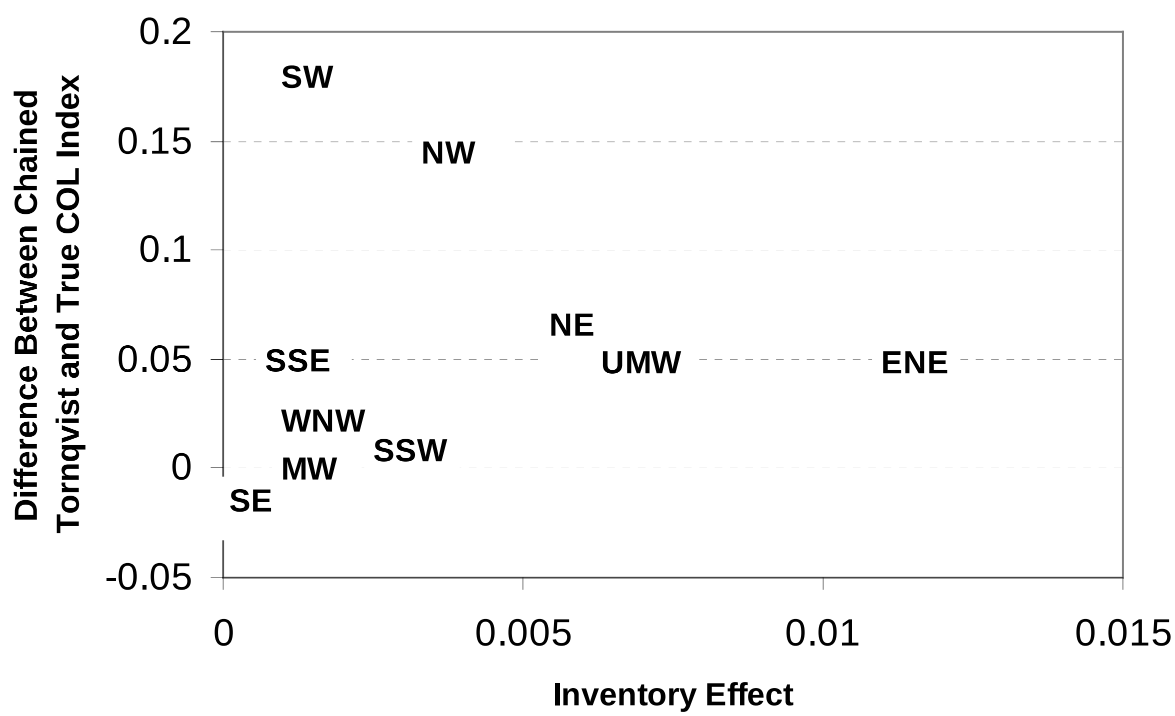

Figure 2: Index Bias and Inventory Effect - Stores

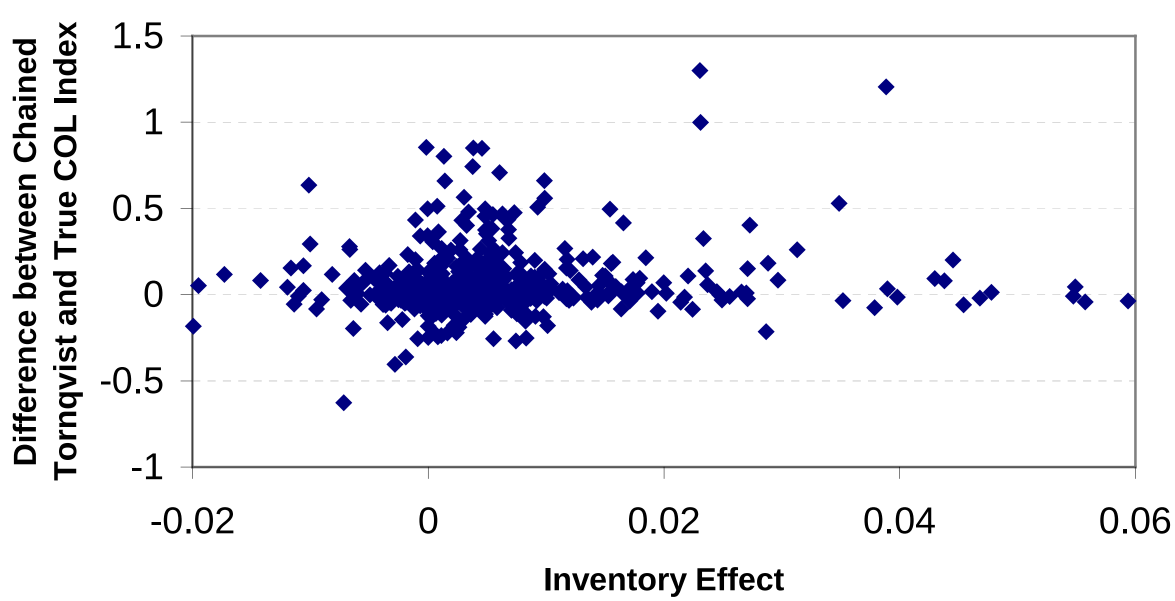

Table 1: Data for two sample products, January – June, 1993

	Product A			Product B		
Week Ending:	**Quantity**	**Price**	**Ad**	**Quantity**	**Price**	**Ad**
1/09/93	25	0.66	N	15	1.39	N
1/16/93	17	0.66	N	20	1.29	N
1/23/93	***150***	***0.59***	***Y***	14	1.29	N
1/30/93	***109***	***0.59***	***Y***	24	1.29	N
2/06/93	58	0.66	N	31	1.29	N
2/13/93	38	0.66	N	16	1.29	N
2/20/93	**7**	**0.33**	**N**	8	1.29	N
2/27/93	**5**	**0.33**	**N**	**15**	**1.19**	**N**
3/06/93	***213***	***0.49***	***Y***	**21**	**1.19**	**N**
3/13/93	43	0.66	N	***92***	***1.19***	***Y***
3/20/93	12	0.66	N	19	1.29	N
3/27/93	**5**	**0.33**	**N**	27	1.29	N
4/03/93	50	0.66	N	23	1.29	N
4/10/93	***231***	***0.49***	***Y***	22	1.29	N
4/17/93	15	0.66	N	15	1.29	N
4/24/93	18	0.66	N	28	1.39	N
5/01/93	**3**	**0.33**	**N**	12	1.39	N
5/08/93	**18**	**0.33**	**N**	8	1.39	N
5/15/93	***210***	***0.50***	***Y***	11	1.39	N
5/22/93	6	0.66	N	19	1.39	N
5/29/93	21	0.66	N	**18**	**1.19**	**N**
6/05/93	15	0.66	N	**43**	**1.19**	**N**
6/12/93	29	0.66	N	***81***	***1.19***	***Y***
6/19/93	6	0.66	N	13	1.39	N
6/26/93	4	0.66	N	15	1.39	N
Mode 92	**20.6**	**0.79**		**17.7**	**1.39**	
Mode 93	**23.8**	**0.66**		**19.9**	**1.29**	

Notes:
Data in bold are on sale, with price more than 5% below the yearly mode. Data in italics *also* have an advertisement, as tends to occur in the final week of each sale. Demand is exceptionally high in these final weeks when the product is advertised.

Table 2: Summary of data for Three Regions

	Northeast		Midwest		Southwest	
	1992	**1993**	**1992**	**1993**	**1992**	**1993**
Mode Price ($)	1.69	1.73	1.30	1.32	1.56	1.59
Relative to mode:						
Price	0.988	0.986	1.005	1.000	1.000	0.989
During weeks without sales:						
No displays or ads: Price	1.02	1.01	1.03	1.02	1.03	1.02
Quantity	0.98	0.99	0.98	0.98	0.98	0.99
With display, no ad: Price	1.02	1.02	1.02	1.03	1.04	1.04
Quantity	1.84	1.72	1.50	1.66	1.69	1.67
With advertisement: Price	1.01	1.01	1.02	1.04	1.03	1.04
Quantity	2.22	1.46	2.20	2.02	1.68	1.88
During weeks with sales:						
No displays or ads: Price	0.85	0.87	0.88	0.89	0.87	0.87
Quantity	1.94	1.79	1.97	1.40	1.87	2.09
With display, no ad: Price	0.74	0.76	0.84	0.81	0.82	0.78
Quantity	7.20	7.94	5.93	6.41	6.77	5.64
With advertisement: Price	0.66	0.68	0.76	0.75	0.73	0.70
Quantity	12.64	11.70	7.01	6.25	8.81	9.14
Frequency of no sale (%)	**82.6**	**80.8**	**84.7**	**81.1**	**80.0**	**77.5**
Frequency of sales (%)	**17.4**	**19.2**	**15.3**	**18.9**	**20.1**	**22.5**
Lasting one week	44.9	45.2	40.1	26.0	34.4	24.9
Lasting two weeks	19.1	15.1	16.5	15.2	22.1	15.6
Lasting three weeks	6.1	9.3	6.1	6.7	7.9	11.0
Lasting four weeks	5.6	9.6	7.0	9.2	7.5	15.6
More than four weeks	24.4	21.0	30.4	43.0	28.2	33.0
During weeks without sales:						
Freq. of no displays or ads (%)	98.4	98.7	96.9	97.2	97.6	98.7
Freq. of displays, not ads (%)	1.0	0.8	2.2	2.1	1.3	0.9
Freq. of advertisements (%)	0.6	0.5	0.9	0.7	1.2	0.4
During weeks with sales:						
Freq. of no displays or ads (%)	78.2	80.7	81.6	86.9	81.0	85.2
Freq. of displays, not ads (%)	3.4	2.9	7.0	5.3	2.6	1.9
Freq. of advertisements (%)	18.4	16.4	11.3	7.8	16.4	12.9
Freq. of ads during sales (%):						
For sales of one week only	71.7	65.1	56.9	62.1	59.7	67.4
At start (for sales > one week)	26.4	16.1	10.6	5.9	27.2	11.2
At end (for sales > one week)	19.2	14.4	20.4	18.1	24.1	18.1

Table 3: Price Indexes constructed over two sample products, January – June, 1993

Week Ending	Fixed- Base Laspeyres	Chained Laspeyres	Fixed- Base Geometric	Fixed-Base Tornqvist	Chained Tornqvist
1/09/93	0.934	0.934	0.931	0.927	0.927
1/16/93	0.891	0.894	0.890	0.894	0.885
1/23/93	***0.856***	***0.856***	***0.851***	***0.812***	***0.830***
1/30/93	***0.856***	***0.856***	***0.851***	***0.826***	***0.830***
2/06/93	0.891	0.897	0.890	0.886	0.886
2/13/93	0.891	0.897	0.890	0.883	0.886
2/20/93	**0.724**	**0.718**	**0.675**	**0.736**	**0.688**
2/27/93	**0.681**	**0.684**	**0.643**	**0.720**	**0.641**
3/06/93	***0.764***	***0.821***	***0.756***	***0.709***	***0.769***
3/13/93	***0.848***	***0.930***	***0.848***	***0.850***	***0.889***
3/20/93	0.891	0.977	0.890	0.897	0.947
3/27/93	**0.724**	**0.782**	**0.675**	**0.777**	**0.856**
4/03/93	0.891	1.094	0.890	0.884	1.044
4/10/93	***0.805***	***0.982***	***0.790***	***0.729***	***0.857***
4/17/93	0.891	1.118	0.890	0.893	1.015
4/24/93	0.934	1.170	0.931	0.945	1.071
5/01/93	**0.768**	**0.936**	**0.706**	**0.820**	**0.968**
5/08/93	**0.768**	**0.936**	**0.706**	**0.722**	**0.968**
5/15/93	***0.851***	***1.123***	***0.830***	***0.743***	***1.240***
5/22/93	0.934	1.272	0.931	0.954	1.433
5/29/93	**0.848**	**1.163**	**0.848**	**0.848**	**1.277**
6/05/93	**0.848**	**1.163**	**0.848**	**0.850**	**1.277**
6/12/93	***0.848***	***1.163***	***0.848***	***0.850***	***1.277***
6/19/93	0.934	1.280	0.931	0.949	1.452
6/26/93	0.934	1.280	0.931	0.955	1.452
Averages:					
Base 92	**0.848**	**0.997**	**0.835**	**0.842**	**1.015**
Base 93	**0.951**	**1.116**	**0.941**	**0.946**	**1.144**

Notes:
Data are for the two products in Table 1. Data in bold had one product on sale, with price more than 5% below the yearly mode. Data in italics *also* have an advertisement for that produce, as tends to occur in the final week of each sale. Demand is high in these final weeks when the product is advertised, so the largest increases in the indexes – especially the chained indexes – follow these weeks.

Table 4: Price Indexes constructed over Complete Sample (Average values over 1993)

	True COL Index	Fixed-base Laspeyres	Chained Laspeyres	Fixed-base Geometric	Fixed-base Tornqvist	Chained Tornqvist
			East Northeast			
1992 base	0.950	1.009	1.144	1.002	0.967	0.980
1993 base	0.940	0.991	1.127	0.987	0.959	0.986
			Northeast			
1992 base	0.937	0.987	1.204	0.976	0.941	1.006
1993 base	0.930	0.972	1.193	0.966	0.932	0.991
			Northwest			
1992 base	0.921	0.995	1.184	0.978	0.940	1.075
1993 base	0.954	1.007	1.224	0.996	0.971	1.097
			West Northwest			
1992 base	0.977	1.013	1.118	1.002	0.978	0.998
1993 base	0.954	0.985	1.095	0.975	0.956	0.974
			Midwest			
1992 base	0.970	1.004	1.145	0.998	0.972	0.956
1993 base	0.958	0.983	1.108	0.978	0.960	0.956
			Upper Midwest			
1992 base	0.937	0.958	1.033	0.949	0.951	0.995
1993 base	0.985	0.999	1.081	0.997	1.000	1.037
			South Southeast			
1992 base	0.949	0.972	1.036	0.965	0.955	0.993
1993 base	0.988	1.006	1.072	1.001	0.994	1.034
			South Southwest			
1992 base	0.970	1.006	1.104	0.997	0.978	0.976
1993 base	0.971	0.994	1.093	0.989	0.980	0.979
			Southeast			
1992 base	0.985	0.989	1.017	0.985	0.978	0.972
1993 base	1.000	1.007	1.043	1.005	0.998	0.994
			Southwest			
1992 base	0.964	1.029	1.204	1.017	0.983	1.138
1993 base	0.945	0.998	1.192	0.991	0.961	1.123
			Total U.S.			
1992 base	0.956	0.996	1.119	0.987	0.964	1.007
1993 base	0.962	0.994	1.122	0.989	0.971	1.014

Table 5: Regression Results over 1993 Sample
Dependent variable: Log of Quantity (relative to mode)

Log Relative Price	Log Price Index	Log(Weeks to Own Next Sale)	Log(Weeks to Any Next Sale)	Display	Ad	Log Rel. Price*Ad	R^2, N
East Northeast							
-3.82	0.85	**0.34**	**-0.08**				0.23
(0.04)	(0.03)	**(0.01)**	**(0.04)**				115,434
-2.25	0.84	**0.11**	**-0.11**	1.08	0.66	-1.15	0.29
(0.04)	(0.03)	**(0.01)**	**(0.04)**	(0.02)	(0.02)	(0.06)	115,434
Northeast							
-3.37	0.76	**0.03**	**0.33**				0.22
(0.03)	(0.03)	**(0.01)**	**(0.03)**				120,554
-2.31	0.81	-0.01	**0.22**	0.71	0.46	-0.71	0.26
(0.04)	(0.03)	(0.01)	**(0.03)**	(0.01)	(0.02)	(0.06)	120,554
Northwest							
-2.85	0.66	0.01	**0.14**				0.26
(0.04)	(0.04)	(0.02)	**(0.04)**				57,168
-2.42	0.67	0.01	0.07	0.58	0.33	-0.20	0.28
(0.05)	(0.04)	(0.02)	(0.04)	(0.03)	(0.02)	(0.06)	57,168
West Northwest							
-2.57	0.79	0.01	0.05				0.15
(0.04)	(0.06)	(0.02)	(0.04)				79,488
-2.32	0.85	0.01	0.04	0.57	0.41	0.34	0.18
(0.05)	(0.06)	(0.02)	(0.04)	(0.01)	(0.02)	(0.07)	79,488
Midwest							
-2.26	0.56	-0.01	0.10				0.09
(0.06)	(0.06)	(0.03)	(0.05)				48,537
-1.74	0.57	-0.01	0.02	0.70	0.40	-0.10	0.13
(0.07)	(0.06)	(0.03)	(0.05)	(0.02)	(0.03)	(0.10)	48,537

Notes: Coefficients on weeks to next sale that are significantly different from zero at the 5% level are indicated in bold. Standard errors are shown in parentheses. Regressions also included fixed effects by store, and lag and lead prices.

Table 5 (cont'd): Regression Results over 1993 Sample
Dependent variable: Log of Quantity (relative to mode)

Log Relative Price	Log Price Index	Log(Weeks to Own Next Sale)	Log(Weeks to Any Next Sale)	Display	Ad	Log Rel. Price*Ad	R^2, N
			Upper Midwest				
-2.21 (0.06)	0.38 (0.06)	**0.15** **(0.03)**	**0.19** **(0.05)**				0.08 46,906
-1.34 (0.07)	0.48 (0.06)	-0.01 (0.03)	**0.12** **(0.05)**	0.50 (0.02)	0.59 (0.03)	-1.03 (0.11)	0.12 46,906
			South South-East				
-2.04 (0.06)	0.40 (0.06)	-0.02 (0.03)	**0.18** **(0.05)**				0.10 59,340
-1.68 (0.07)	0.44 (0.06)	0.01 (0.03)	-0.003 (0.05)	0.51 (0.02)	0.20 (0.02)	-0.71 (0.09)	0.11 59,340
			South South-West				
-2.35 (0.06)	0.56 (0.04)	**-0.07** **(0.02)**	**0.30** **(0.04)**				0.10 76,876
-1.72 (0.06)	0.62 (0.04)	-0.03 (0.02)	**0.14** **(0.04)**	0.55 (0.02)	0.55 (0.03)	-0.86 (0.10)	0.13 76,876
			South East				
-2.06 (0.07)	0.27 (0.09)	**0.09** **(0.02)**	**-0.09** **(0.04)**				0.06 66,689
-1.43 (0.08)	0.29 (0.09)	**0.07** **(0.02)**	**-0.12** **(0.04)**	0.46 (0.02)	0.46 (0.02)	-0.27 (0.11)	0.08 66,689
			South West				
-3.26 (0.05)	0.62 (0.05)	-0.03 (0.02)	**0.15** **(0.05)**				0.24 57,121
-2.48 (0.05)	0.70 (0.05)	**-0.05** **(0.02)**	0.06 (0.05)	0.52 (0.03)	0.51 (0.03)	-0.62 (0.09)	0.26 57,121
			Total U.S.				
-3.08 (0.01)	0.73 (0.01)	**0.13** **(0.006)**	0.00 (0.01)				0.17 728,122
-2.19 (0.02)	0.78 (0.01)	**0.08** **(0.006)**	**-0.06** **(0.01)**	0.67 (0.006)	0.45 (0.007)	-0.78 (0.02)	0.20 728,122