Estimating the Benefits and Costs of New and Disappearing Products

December 18, 2017

Erwin Diewert and Robert Feenstra,[1]
Discussion Paper 17-10,
Vancouver School of Economics,
University of British Columbia,
Vancouver, B.C., Canada, V6T 1L4.

Abstract

A major challenge facing statistical agencies is the problem of adjusting price and quantity indexes for changes in the availability of commodities. This problem arises in the scanner data context as products in a commodity stratum appear and disappear in retail outlets. Hicks suggested a reservation price methodology for dealing with this problem in the context of the economic approach to index number theory. Feenstra and Hausman suggested specific methods for implementing the Hicksian approach. The present paper evaluates these approaches and suggests some alternative approaches to the estimation of reservation prices. The various approaches are implemented using some scanner data on frozen juice products that are available online.

Keywords

Hicksian reservation prices, virtual prices, Laspeyres, Paasche, Fisher, Törnqvist and Sato-Vartia price indexes, new goods, welfare measurement, Constant Elasticity of Substitution (CES) preferences, Konüs, Byushgens and Fisher (KBF) preferences, duality theory, consumer demand systems, flexible functional forms.

JEL Classification Numbers

C33, C43, C81, D11, D60, E31.

---

[1] W. Erwin Diewert: Vancouver School of Economics, University of British Columbia, Vancouver B.C., Canada, V6T 1Z1 and the School of Economics, UNSW Sydney, NSW 2052, Australia (erwin.diewert@ubc.ca); Robert Feenstra, University of California at Davis, Davis, Davis, CA, USA, 95617 (rcfeenstra@uc.davis.edu). The authors thank Jan de Haan, Kevin Fox, Jerry Hausman, Alan Woodland and the participants at the EMG Workshop, December 1, 2017 for helpful comments. The authors gratefully acknowledge the financial support of a Trans Atlantic Platform Digging into Data grant. The project title is: Online Prices for Computing Standards of Living Across Countries (OPSLAC).

1. Introduction

One of the more pressing problems facing statistical agencies and economic analysts is the new goods (and services) problem; i.e., how should the introduction of new products and the disappearance of (possibly) obsolete products be treated in the context of forming a consumer price index? Hicks (1940) suggested a general approach to this measurement problem in the context of the economic approach to index number theory. His approach was to apply normal index number theory but estimate hypothetical prices that would induce utility maximizing purchasers of a related group of products to demand 0 units of unavailable products.[2] With these virtual (or reservation or imputed) prices[3] in hand, one can just apply normal index number theory using the augmented price data and the observed quantity data. The practical problem facing statistical agencies is: *how exactly are these virtual prices to be estimated*?

Economists have been worrying about the new goods problem at least since the early contributions of Lehr (1885; 45-46) and Marshall (1887; 373-374), who independently introduced the concept of *chained index numbers* in order to deal with this problem.[4] These authors suggested that the best way to deal with the problem was to use the price and quantity data for adjacent periods and use a suitable index number formula on the set of products that were present in both periods. Keynes (1930; 105-106) endorsed the idea of restricting index number comparisons to the set of products that were present in both periods being compared but he preferred to use this *maximum overlap method* [5] in the context of fixed base indexes. He rejected the idea of using chained indexes because he felt that chained indexes would suffer from a *chain drift problem*.[6] Indeed, we will find that the problem of chain drift is a serious one when calculating price indexes using scanner data on the sales of a retail outlet.

Following up on the contribution of Hicks, many authors developed bounds or rough approximations to the bias that might result from omitting the contribution of new goods in the consumer price index context. Thus Rothbarth (1941) attempted to find some

[2] "The same kind of device can be used in another difficult case, that in which new sorts of goods are introduced in the interval between the two situations we are comparing. If certain goods are available in the II situation which were not available in the I situation, the p_1's corresponding to these goods become indeterminate. The p_2's and q_2's are given by the data and the q_1"s are zero. Nevertheless, although the p_1's cannot be determined from the data, since the goods are not sold in the I situation, it is apparent from the preceding argument what p_1's ought to be introduced in order to make the index-number tests hold. They are those prices which, in the I situation, would *just* make the demands for these commodities (from the whole community) equal to zero." J.R. Hicks (1940; 114). Hofsten (1952; 95-97) extended Hicks' methodology to cover the case of disappearing goods as well.

[3] Rothbarth introduced the term "virtual prices" to describe these hypothetical prices in the rationing context: " I shall call the price system which makes the quantities actually consumed under rationing an optimum the 'virtual price system.'". E. Rothbarth (1941; 100).

[4] See Diewert (1993a; 52-63) for additional material on the early history of the new goods problem.

[5] Keynes (1930; 94) called this the *highest common factor method*.

[6] Keynes noted that chained index numbers failed Walsh's (1901; 389) *multiperiod identity test* which is the following test: $P(p^1.p^2,q^1,q^2)P(p^2.p^3,q^2,q^3)P(p^3.p^1,q^3,q^1) = 1$ where $P(p^1.p^2,q^1,q^2)$ is the bilateral index number formula which is being used. The divergence of the product of the 3 indexes from 1 serves as a measure of the amount of chain drift.

bounds for the bias while Hofsten (1952; 47-50) discussed a variety of approximate methods to adjust for quality change in products, which is essentially the same problem as adjusting an index for the contribution of a new product. Diewert (1980; 498-501) developed some bounds for the bias in a maximum overlap Fisher (1922) index relative to the bias that would result from using the Fisher formula where 0 prices and quantities were used in the Fisher formula for the base period when a new product was not available.[7] Additional bias formulae were developed by Diewert (1987; 779) (1998; 51-54) and Hausman (2003; 26-28). These approximations relied on information (or guesses) about expenditure shares, elasticities or ratios of virtual prices to actual prices. We will examine the Hausman approximate formula in more detail in section 11 below.

We turn now to methods that rely on some form of econometric estimation in order to form estimates of the welfare cost (or changes in the true cost of living index) of changes in product availability. The two main contributors in this area are Feenstra (1994) and Hausman (1996).[8] Econometric methods for adjusting price and quantity indexes will be the main focus of this study. We will apply various econometric methods in order to adjust a consumer price index for changes in the availability of products. We will also obtain econometric estimates for the virtual prices for unavailable products for each period in our sample period. We will test out our suggested methods on a scanner data set that is available on line.[9] The data set is listed in Appendix A so that researchers can use this data set to test out possible improvements to our suggested methodology.

Feenstra's (1994) methodology rests on the properties of the CES unit cost function. His methodology is explained in section 2. In section 3, we adapt his methodology to the case of a CES utility function. Section 4 introduces our scanner data set which we use to test out Feenstra's methodology. Section 5 estimates a CES unit cost function using our data set while section 6 estimates a CES direct utility function. Both systems of estimating equations use the sales shares of the 19 products in our sample as the dependent variables in a systems regression approach. If either the CES unit cost function model or the CES homogeneous utility model were to fit the sample data perfectly, we would obtain exactly the same results. However, neither model fits the data exactly. We find that the CES utility function model fits the data much better than the CES unit cost function model.

There are two problems with the CES unit cost function methodology:

- The CES functional form is not fully flexible[10] and
- The reservation price that induces a potential purchaser to *not* purchase a product is equal to plus infinity, which seems high. Thus the CES methodology may overstate the benefits of increases in product availability.

[7] Diewert (1980; 501) concluded that both Fisher price indexes would probably have an upward bias but the index which used zeros would definitely have a larger bias than the maximum overlap Fisher index. The similar type of argument appears in Diewert (1987; 779).
[8] See also Hausman (1999) (2003) and Hausman and Leonard (2002)
[9] The data are described in section 4 below.
[10] See Diewert (1974) (1976) for the definition of a flexible functional form.

Thus in section 7, we replace the CES utility function with a flexible functional form which was initially due to Konüs and Byushgens (1926; 171). This utility function is $u = f(q) \equiv (q^TAq)^{1/2}$ where A is a symmetric matrix of parameters and q^T is the row vector transpose of the column vector of quantities purchased, q. Konüs and Byushgens showed that if purchasers maximized this utility function in two periods where they faced the price vectors p^1 and p^2 and the utility maximizing vectors were q^1 and q^2, then the utility ratio, $f(q^2)/f(q^1)$, is equal to the Fisher (1922) quantity index, $Q_F(p^1,p^2,q^1,q^2) \equiv [p^{1T}q^2p^{2T}q^2/ p^{1T}q^1p^{2T}q^1]$.[11] Thus we will call this functional form for f the *KBF functional form*. The advantage in working with this flexible functional form is that when some component of the q vector is equal to 0, the resulting utility function is still well defined and the corresponding reservation price can be calculated by partially differentiating the estimated utility function with respect to the quantity variable that happens to equal 0 in the period under consideration. In fact, Diewert (1980; 501-503) suggested exactly this methodological approach to the estimation of reservation prices but in the end, he suggested that it would be difficult to estimate all of the N(N+1)/2 unknown parameters in the A matrix. In the present paper, we solve this degrees of freedom problem by introducing a *semiflexible version* of the flexible KBF functional form.[12] This new methodology is explained in section 7. In section 8, we attempt to estimate the KBF functional form using the usual systems approach to the estimation of consumer demand functions. However, the nonlinearity in our estimating share equations causes our nonlinear estimating procedure to come to a premature halt as we increase the rank of the A matrix. Hence in section 9, we drop the systems approach to the estimation of the unknown parameters in favour of the one big equation approach. The latter approach has the advantage of being able to drop the observations where a product was missing.

Although the implied fits in the product share equations were quite good using our one big equation approach, when we moved from predicted shares generated by our estimates to predicted prices, we found that predicted prices did not match up well with actual prices for the observations where products were present. Thus in section 10, we moved from shares as the dependent variables to using prices as the dependent variables. We continued to estimate higher rank A matrices using the one big equation approach with prices as the dependent variables until we estimated a rank 7 A matrix with 111 unknown parameters. We then used our estimated A matrix in order to define virtual or reservation prices for the unavailable products. We were also able to quantify the effects of the changing availability of products and compare the results of the KBF estimation with the earlier CES benefit measures. We found that the CES methodology did indeed give much

[11] Konüs and Byushgens (1926; 169-172) also introduced the KBF unit cost function, $c(p) \equiv (p^TBp)^{1/2}$ where B is a symmetric matrix of parameters. They showed that this unit cost function functional form is exact for the Fisher price index. If A or B is of full rank, then $B = A^{-1}$. For a description of the contributions of Konüs and Byushgens to index number theory and duality theory, see Diewert (1993a; 47-51). For a description of the regularity conditions that the matrices A and B must satisfy for the KBF f(q) or c(p) to be well behaved, see Diewert and Hill (2010). Diewert (1976) generalized the KB results to more general functional forms for f and c.

[12] Our new semiflexible functional form has properties that are similar to the semiflexible generalization of the Normalized Quadratic functional form introduced by Diewert and Wales (1987) (1988). In section 7 below, we also show how the correct curvature conditions can be imposed on our semiflexible KBF functional form.

higher estimates for the gains from increases in product availability as compared to our KBF methodology.

However, due to the fact that our estimated KBF preferences did not fit the data exactly, we found that occasionally our estimated gain from having an additional product had the wrong sign. Thus in section 11, we developed an alternative methodological approach based on our estimated KBF utility function (which is well behaved by construction) that was free from anomalous results. This utility function based approach is an alternative to Hausman's (1996) expenditure or cost function approach to measuring the gains from increases in product availability.

In section 12, we compare Hausman's approximate approach to a variant of our approach where we use a second order approximation to the estimated utility function. To keep things simple, we consider only the case of two products in this section. We obtain a rather surprising equivalence result.

Section 13 concludes.

Appendix B tries out Feenstra's *double differencing method* for estimating the elasticity of substitution but we apply it to the direct utility estimation of the CES functional form rather than estimating the dual CES unit cost function. We find that this method for estimating the elasticity of substitution worked very well on our scanner data set.

2. Feenstra's CES Unit Cost Function Methodology

In this section, we will explain Feenstra's (1994) CES cost function methodology that he proposed to measure the benefits and costs to consumers due to the appearance of new products and the disappearance of existing products.

The methodology assumes that purchasers of a group of N products all have the same linearly homogeneous, concave and nondecreasing utility function $f(q)$, where the nonnegative vector of purchased products is $q \equiv (q_1,...,q_N) \geq 0_N$ and $u = f(q) \geq 0$ is the utility that the vector of purchases q generates. Given that purchasers face the positive vector of prices $p \equiv (p_1,...,p_N)$ at an outlet, the unit cost function $c(p)$ that is dual to the utility function f is defined as the minimum cost of attaining the utility level that is equal to one:

$$(1)\ c(p) \equiv \min_q\{f(q) \geq 1;\ q \geq 0_N\}.$$

If the unit cost function $c(p)$ is known, then using duality theory, it is possible to recover the underlying utility function $f(q)$.[13] Feenstra assumed that the unit cost function has the following CES functional form:

---

[13] It can be shown that for $q >> 0_N$, $f(q) = 1/\max_p\{c(p): \Sigma_{n=1}^N p_nq_n \leq 1\ ;\ p \geq 0_N\}$; see Diewert (1974; 110-112) (1993b; 129) on the duality between linearly homogeneous aggregator functions f(q) and unit cost functions c(p).

$$(2)\ c(p) \equiv \alpha_0 \left[\sum_{n=1}^{N} \alpha_n p_n^{1-\sigma}\right]^{1/(1-\sigma)} \qquad \text{if } \sigma \neq 1;$$
$$\equiv \alpha_0 \prod_{n=1}^{N} p_n^{\alpha_n} \qquad \text{if } \sigma = 1$$

where the α_i and σ are nonnegative parameters with $\sum_{i=1}^{N} \alpha_i = 1$. The unit cost function defined by (2) is a *Constant Elasticity of Substitution (CES) utility function* which was introduced into the economics literature by Arrow, Chenery, Minhas and Solow (1961)[14]. The parameter σ is the *elasticity of substitution*;[15] when $\sigma = 0$, the unit cost function defined by (2) becomes linear in prices and hence corresponds to a fixed coefficients aggregator function which exhibits 0 substitutability between all commodities. When $\sigma = 1$, the corresponding aggregator or utility function is a Cobb-Douglas function. When σ approaches $+\infty$, the corresponding aggregator function f approaches a linear aggregator function which exhibits infinite substitutability between each pair of inputs. The CES unit cost function defined by (2) is of course *not* a fully flexible functional form (unless the number of commodities N being aggregated is 2) but it is considerably more flexible than the zero substitutability aggregator function (this is the special case of (2) where σ is set equal to zero) that is exact for the Laspeyres and Paasche price indexes.

In order to simplify the notation, we set $r \equiv 1 - \sigma$. Under the assumption of cost minimizing behavior on the part of purchasers of the N products for periods $t = 1,...,T$, Shephard's (1953; 11) Lemma tells us that the observed period t consumption of commodity i, q_i^t, will be equal to $u^t \partial c(p^t)/\partial p_i$ where $\partial c(p^t)/\partial p_i$ is the first order partial derivative of the unit cost function with respect to the ith commodity price evaluated at the period t prices and $u^t = f(q^t)$ is the aggregate (unobservable) level of period t utility. Denote the share of product i in total sales of the N products during period t as $s_i^t \equiv p_i^t q_i^t / p^t \cdot q^t$ for $i = 1,...,N$ and $t = 1,...,T$ where $p^t \cdot q^t \equiv \Sigma_{n=1}^{N} p_n^t q_n^t$. Note that the assumption of cost minimizing behavior during each period implies that the following equations will hold:

$$(3)\ p^t \cdot q^t = u^t c(p^t)\ ; \qquad t = 1,...,T$$

where c is the CES unit cost function defined by (2).

Using the CES functional form defined by (2) and assuming that $\sigma \neq 1$ (or $r \neq 0$),[16] the following equations are obtained using Shephard's Lemma:

$$(4)\ q_i^t = u^t \alpha_0 \left[\sum_{n=1}^{N} \alpha_n (p_n^t)^r\right]^{(1/r)-1} \alpha_i (p_i^t)^{r-1}; \qquad i = 1,\ldots,N;\ t = 1,...,T$$

---

[14] In the mathematics literature, this aggregator function or utility function is known as a mean of order $r \equiv 1 - \sigma$; see Hardy, Littlewood and Polyá (1934; 12-13).

[15] Let c(p) be an arbitrary unit cost function that is twice continuously differentiable. The Allen (1938; 504) Uzawa (1962) *elasticity of substitution* $\sigma_{nk}(p)$ between products n and k is defined as $c(p)c_{nk}(p)/c_n(p)c_k(p)$ for $n \neq k$ where the first and second order partial derivatives of c(p) are defined as $c_n(p) \equiv \partial c(p)/\partial p_n$ and $c_{nk}(p) \equiv \partial^2 c(p)/\partial p_n \partial p_k$. For the CES unit cost function defined by (2), $\sigma_{nk}(p) = \sigma$ for all pairs of products; i.e., the elasticity of substitution between all pairs of products is a constant for the CES unit cost function.

[16] When $\sigma = 1$, we have the case of Cobb-Douglas preferences. In the remainder of this paper, we will assume that $\sigma > 1$ (or equivalently, that $r < 0$).

$= u^t c(p^t)\ \alpha_i\ (p_i^t)^{r-1}/\sum_{n=1}^N \alpha_n\ (p_n^t)^r$.

Premultiply equation i for period t in (4) by $p_i^t/p^t{\cdot}q^t$. Using (2) and (3), the resulting equations can be rewritten as follows:

$$(5)\quad s_i^t = \alpha_i\ (p_i^t)^r/\sum_{n=1}^N \alpha_n\ (p_n^t)^r\ ; \qquad i = 1,\ldots,N;\ t = 1,\ldots,T.$$

The NT share equations defined by (5) can be used as estimating equations using a nonlinear regression approach. We will implement this approach later in the paper. Note that the positive scale parameter α_0 cannot be identified using equations (5), which of course is normal: utility can only be estimated up to an arbitrary scaling factor. Henceforth, we will assume $\alpha_0 = 1$. The share equations (5) are homogeneous of degree one in the parameters $\alpha_1,\ldots,\alpha_N$ and thus the identifying restriction on these parameters, $\sum_{i=1}^N \alpha_i = 1$, can be replaced with an equivalent restriction such as $\alpha_N = 1$.

Suppose that all N products are available in all T periods in our sample and we have estimated the unknown parameters which appear in equations (5). Then the *period t CES price index* (relative to the level of prices for period 1), P_{CES}^t, can be defined as the following ratio of unit costs in period t relative to period 1:

$$(6)\ P_{CES}^t \equiv [\sum_{n=1}^N \alpha_n\ (p_n^t)^r]^{(1/r)} / [\sum_{n=1}^N \alpha_n\ (p_n^1)^r]^{(1/r)}\ ; \qquad t = 1,\ldots,T.$$

Suppose further that the observed price and quantity data vectors , p^t and q^t for t = 1,...,T, satisfy equations (3) where c(p) is defined by (2) and the quantity data vectors q^t satisfy the Shephard's Lemma equations (4). Thus the observed price and quantity data are assumed to be consistent with cost minimizing behavior on the part of purchasers where all purchasers have CES preferences that are dual to the CES unit cost function defined by (2). Then Sato (1976) and Vartia (1976) showed that the sequence of CES price indexes defined by (6) *could be numerically calculated just using the observed price and quantity data*; i.e., it would not be necessary to estimate the unknown α_n and σ (or r) parameters in equations (6). The logarithm of the period t fixed base *Sato-Vartia Index* P_{SV}^t is defined by the following equation:

$$(7)\ \ln P_{SV}^t \equiv \Sigma_{n=1}^N w_n^t \ln(p_n^t/p_n^1)\ ; \qquad t = 1,\ldots,T.$$

The weights w_n^t that appear in equations (7) are calculated in two stages. The first stage set of weights is defined as $w_n^{t*} \equiv (s_n^t - s_n^1)/(\ln s_n^t - \ln s_n^1)$ for n = 1,...,N and t = 1,...,T provided that $s_n^t \neq s_n^1$. If $s_n^t = s_n^1$, then define $w_n^{t*} \equiv s_n^t = s_n^1$. The second stage weights are defined as $w_n^t \equiv w_n^{t*}/\Sigma_{i=1}^N w_i^{t*}$ for n = 1,...,N and t = 1,...,T. Note that in order for $\ln P_{CES}^t$ to be well defined, we require that $s_n^t > 0$, $s_n^1 > 0$, $p_n^t > 0$ and $p_n^1 > 0$ for all n = 1,...,N and t = 1,...,T; i.e., all prices and quantities must be positive for all products and for all periods.

Now we can explain Feenstra's (1994) model where "new" commodities can appear and "old" commodities can disappear from period to period.

Feenstra (1994) assumed CES preferences with $\sigma > 1$ (or equivalently, $r < 0$). He applied the reservation price methodology first introduced by Hicks (1940); i.e., Hicks assumed that the consumer had preferences over all goods, but for the goods which had not yet appeared, there was a reservation price that would be just high enough that consumers would not want to purchase the good in the period under consideration.[17] This assumption works rather well with CES preferences, *because we do not have to estimate these reservation prices*; they will all be equal to $+\infty$ when $\sigma > 1$.

Feenstra allowed for new products to appear and for existing products to disappear from period to period.[18] Feenstra assumed that the set of commodities that are available in period t is I(t) for t = 1,...,T. The (imputed) prices for the unavailable commodities in each period are set equal to $+\infty$ and thus if $r < 0$, an infinite price p_n^t raised to a negative power generates a 0; i.e., if product n is unavailable in period t, then $(p_n^t)^r = (\infty)^r = 0$ if r is negative.

The CES period t true price level under these conditions when $r < 0$ turns out to be the following CES unit cost function that is defined over only products that are available during period t:

$$(8)\ c(p^t) \equiv [\textstyle\sum_{n=1}^N \alpha_n (p_n^t)^r]^{(1/r)} = [\sum_{i\in I(t)} \alpha_i (p_i^1)^r]^{1/r} .$$

Using equations (4) for this new model and multiplying the period t demand q_i^t by the corresponding price p_i^t for the items that are actually available leads to the following equations which describe the purchasers' nonzero expenditures on product i in period t:

$$(9)\ p_i^t q_i^t = u^t [\textstyle\sum_{n\in I(t)} \alpha_n (p_n^t)^r]^{(1/r)-1} \alpha_i (p_i^t)^r ; \qquad t = 1,...,T;\ i\in I(t)$$
$$= u^t c(p^t)\, \alpha_i (p_i^t)^r / \textstyle\sum_{n\in I(t)} \alpha_n (p_n^t)^r .$$

In each period t, the sum of observed expenditures, $\Sigma_{n\in I(t)}\, p_n^t q_n^t$, equals the period t utility level, u^t, times the CES unit cost $c(p^t)$ defined by (8):

$$(10)\ \Sigma_{n\in I(t)}\, p_n^t q_n^t = u^t c(p^t) = u^t[\textstyle\sum_{i\in I(t)} \alpha_i (p_i^1)^r]^{1/r} ; \qquad t = 1,...,T.$$

Recall that the ith sales share of product i in period t was defined as $s_i^t \equiv p_i^t q_i^t / \Sigma_{n\in I(t)}\, p_n^t q_n^t$ for t = 1,...,T and $i\in I(t)$. Using these share definitions and equations (10), we can rewrite equations (9) in the following form:

$$(11)\ s_i^t = \alpha_i (p_i^t)^r / \textstyle\sum_{n\in I(t)} \alpha_n (p_n^t)^r ; \qquad t = 1,...,T;\ i\in I(t)$$
$$= \alpha_i (p_i^t)^r / c(p^t)^r$$

[17] The same logic is applied to disappearing products.

[18] In many cases, a "new" product is not a genuinely new product; it is just a product that was not in stock in the previous period. Similarly, in many cases, a disappearing product is not necessarily a truly disappearing product; it is simple a product that was not in stock for the period under consideration. Many retail chains rotate products, temporarily discontinuing some products in favour of competing products in order to take advantage of manufacturer discounted prices for selected products.

where the second set of equations follows using definitions (8).

Now we can work out Feenstra's (1994) model for measuring the benefits and costs of new and disappearing commodities. Start out with the period t CES exact price level defined by (8) and define the CES fixed base price index for period t, P_{CES}^t, as the ratio of the period t CES price level to the corresponding period 1 price level:[19]

$$
\begin{aligned}
(12)\ P_{CES}^t &\equiv c(p^t)/c(p^1)\ ; & t = 1,\ldots,T\\
&= [\textstyle\sum_{i\in I(t)} \alpha_i\,(p_i^t)^r]^{1/r} / [\textstyle\sum_{i\in I(1)} \alpha_i\,(p_i^1)^r]^{1/r}\\
&= [\text{ Index } 1]\times[\text{Index } 2]\times[\text{Index } 3]
\end{aligned}
$$

where the three indexes in equations (12) are defined as follows:

$$(13)\ \text{Index } 1 \equiv [\textstyle\sum_{i\in I(t)\cap I(1)} \alpha_i\,(p_i^t)^r]^{1/r} / [\textstyle\sum_{i\in I(1)\cap I(t)} \alpha_i\,(p_i^1)^r]^{1/r}\ ;$$
$$(14)\ \text{Index } 2 \equiv [\textstyle\sum_{i\in I(t)} \alpha_i\,(p_i^t)^r]^{1/r} / [\textstyle\sum_{i\in I(1)\cap I(t)} \alpha_i\,(p_i^t)^r]^{1/r}\ ;$$
$$(15)\ \text{Index } 3 \equiv [\textstyle\sum_{i\in I(1)\cap I(t)} \alpha_i\,(p_i^1)^r]^{1/r} / [\textstyle\sum_{i\in I(1)} \alpha_i\,(p_i^1)^r]^{1/r}\ .$$

Note that Index 1 defines a CES price index over the set of commodities that are available in both periods t and 1. Denote the CES cost function c^{t*} that has the same α_n parameters as before but is now defined over only products that are available in periods 1 and t:

$$(16)\ c^{t*}(p) \equiv [\textstyle\sum_{i\in I(t)\cap I(1)} \alpha_i\,(p_i)^r]^{1/r}\ ; \qquad t = 1,2,\ldots,T.$$

The period t expenditure share equations that correspond to equations (11) using the unit cost function defined by (16) are the following ones:

$$
\begin{aligned}
(17)\ s_i^{t*} &\equiv p_i^t q_i^t / \textstyle\sum_{n\in I(t))\cap I(1)} p_n^t q_n^t & t = 1,\ldots,T;\ i\in I(1)\cap I(t)\\
&= \alpha_i\,(p_i^t)^r / \textstyle\sum_{n\in I(t))\cap I(1)} \alpha_n\,(p_n^t)^r\\
&= \alpha_i\,(p_i^t)^r / c^{t*}(p^t)^r
\end{aligned}
$$

where the third equality follows using definitions (16).

Note that Index 1 is equal to $c^{t*}(p^t)/c^{t*}(p^1)$ and the Sato-Vartia formula (7) (restricted to commodities n that are present in periods 1 and t) can be used to calculate this index using the observed price and quantity data for the products that are available in both periods 1 and t.

We turn now to the evaluation of Indexes 2 and 3. It turns out that we will need an estimate for the elasticity of substitution σ (or equivalently of r) in order to find empirical expressions for these indexes. It is convenient to define the following *observable expenditure or sales ratios*:

[19] In the algebra which follows, the prices and quantities of period 1 can be replaced with the prices and quantities of any period. Feenstra (1994) developed his algebra for $c(p^t)/c(p^{t-1})$.

(18) $\lambda^t \equiv \sum_{n\in I(t)} p_n{}^t q_n{}^t / \sum_{n\in I(1)\cap I(t)} p_n{}^t q_n{}^t$; t = 1,,...,T.
(19) $\mu^t \equiv \sum_{n\in I(1)\cap I(t)} p_n{}^1 q_n{}^1 / \sum_{n\in I(1)} p_n{}^1 q_n{}^1$; t =1,,...,T.

We assume that there is at least one product that is present in periods 1 and t for each t. Let product i be any one of these common products for a given t. Then the share equations (11) and (17) hold for this product. These share equations can be rearranged to give us the following two equations:

(20) $\alpha_i(p_i{}^t)^r = [\sum_{n\in I(t)} \alpha_n (p_n{}^t)^r]\, p_i{}^t q_i{}^t / [\sum_{n\in I(t)} p_n{}^t q_n{}^t]$;
(21) $\alpha_i(p_i{}^t)^r = [\sum_{n\in I(1)\cap I(t)} \alpha_n (p_n{}^t)^r]\, p_i{}^t q_i{}^t / [\sum_{n\in I(1)\cap I(t)} p_n{}^t q_n{}^t]$.

Equating (20) to (21) leads to the following equations:

(22) $\sum_{n\in I(t)} \alpha_n (p_n{}^t)^r / \sum_{n\in I(1)\cap I(t)} \alpha_n (p_n{}^t)^r = \sum_{n\in I(t)} p_n{}^t q_n{}^t / \sum_{n\in I(1)\cap I(t)} p_n{}^t q_n{}^t$
$= \lambda^t$

where the last equality follows using definition (18). Now take the 1/r root of both sides of (22) and use definition (14) in order to obtain the following equality:

(23) Index 2 = $[\lambda^t]^{1/r} = [\sum_{i\in I(t)} p_i{}^t q_i{}^t / \sum_{i\in I(1)\cap I(t)} p_i{}^t q_i{}^t]^{1/r}$.[20]

Again assume that product i is available in periods 1 and t. Rearrange the share equations (11) and (17) for t = 1 and product i and we obtain the following two equations:

(23) $\alpha_i(p_i{}^1)^r = [\sum_{n\in I(1)} \alpha_n (p_n{}^1)^r]\, p_i{}^1 q_i{}^1 / [\sum_{n\in I(1)} p_n{}^1 q_n{}^1]$;
(24) $\alpha_i(p_i{}^1)^r = [\sum_{n\in I(1)\cap I(t)} \alpha_n (p_n{}^1)^r]\, p_i{}^1 q_i{}^1 / [\sum_{n\in I(1)\cap I(t)} p_n{}^1 q_n{}^1]$.

Equating (23) to (24) leads to the following equations:

(25) $\sum_{n\in I(1)\cap I(t)} \alpha_n (p_n{}^1)^r / \sum_{n\in I(1)} \alpha_n (p_n{}^1)^r = \sum_{n\in I(1)\cap I(t)} p_n{}^1 q_n{}^1 / \sum_{n\in I(1)} p_n{}^1 q_n{}^1$
$= \mu^t$

where the last equality follows using definition (19). Now take the 1/r root of both sides of (25) and use definition (15) in order to obtain the following equality:[21]

---

[20] If new products become available in period t that were not available in period 1, then $\lambda^t > 1$. Recall that $r = 1 - \sigma$ and $r < 0$. Index 2 evaluated at period t prices equals $(\lambda^t)^{1/r} = (\lambda^t)^{1/(1-\sigma)}$ and thus is an increasing function of σ for $1 < \sigma < +\infty$. With $\lambda^t > 1$, the limit of $(\lambda^t)^{1/(1-\sigma)}$ as σ approaches 1 is 0 and the limit of $(\lambda^t)^{1/(1-\sigma)}$ as σ approaches $+\infty$ is 1. Thus the gains in utility from increased product variety are huge if σ is slightly greater than 1 and diminish to no gains at all as σ becomes very large. Suppose that λ^t =1.05 and σ = 1.01, 1.1, 1.5, 2, 3, 5, 10 and 100. Then Index 2 will equal 0.0076, 0.614, 0.907, 0.952, 0.976, 0.988, 0.995 and 0.9995 respectively. Thus the gains from increased product variety are very sensitive to the estimate for the elasticity of substitution. The gains are gigantic if σ is close to 1.

[21] If some products that were available in period 1 become unavailable in period t, then $\mu^t < 1$. Index 3 evaluated at period 1 prices equals $(\mu^t)^{1/r} = (\mu^t)^{1/(1-\sigma)}$ and is an decreasing function of σ for $1 < \sigma < +\infty$. With $\mu^t < 1$, the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches 1 is $+\infty$ and the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches $+\infty$ is 1. Thus the losses in utility from decreased product variety are huge if σ is slightly greater than 1 and

(26) Index 3 = $[\mu^t]^{1/r} = [\sum_{n\in I(1)\cap I(t)} p_n^1 q_n^1 / \sum_{n\in I(1)} p_n^1 q_n^1]^{1/r}$.

Thus if r is known or has been estimated, then Index 2 and Index 3 can readily be calculated as simple ratios of sums of observable expenditures raised to the power1/r. Note that $[\sum_{i\in I(t)} p_i^t q_i^t / \sum_{i\in I(1)\cap I(t)} p_i^t q_i^t] \geq 1$. If period t has products that were not available in period 1, then the strict inequality will hold and since $1/r < 0$, it can be seen that Index 2 will be less than unity. Thus Index 2 is a measure of how much the true cost of living index is *reduced* in period t due to the introduction of products that were not available in period 1. Similarly, $[\sum_{i\in I(1)\cap I(t)} p_i^1 q_i^1 / \sum_{i\in I(1)} p_i^1 q_i^1] \leq 1$. If period 1 has products that are not available in period t, then the strict inequality will hold and since$1/r < 0$, it can be seen that Index 3 will be greater than unity, Thus Index 3 is a measure of how much the true cost of living index is *increased* in period t due to the disappearance of products that were available in period 1 but are not available in period t.

Turning briefly to the problems associated with estimating r (and the α_n) when not all products are available in all periods, it can be seen that the initial estimating share equations (5) are now replaced by the following equations:

(27) $s_n^t = \alpha_n (p_n^t)^r / \Sigma_{k=1}^N \alpha_k (p_k^t)^r$; $\qquad$ t = 1,...,T; $n\in I(t)$.

In the next section, we obtain an alternative set of share equations that could be used in order to estimate the elasticity of substitution.

3. The Primal Approach to the Estimation of CES Preferences

It turns out that estimating the purchaser’s utility function directly (rather than estimating the dual unit cost function) is advantageous when estimates of reservation prices for products that are not available are required. In the case of CES preferences, this advantage is not apparent since the CES reservation prices are automatically set equal to infinity. But it turns out that there may be advantages in estimating the CES utility function directly because of econometric considerations as we shall see later. Thus in this section, we will derive the purchaser demand functions that are consistent with the maximization of a CES utility function.

Using the same notation that was used in the beginning of the previous section, we assume that the purchaser utility function f(q) is defined as the following *CES utility function*:

(28) $f(q_1,...,q_N) \equiv [\Sigma_{n=1}^N \beta_n q_n^s]^{1/s}$

---

diminish to no gains at all as σ becomes very large. Suppose that μ^t =0.95 and σ takes on the same values as in the previous footnote. Then Index 3 will equal 168.9, 1.670, 1.108, 1.053, 1.026, 1.013, 1.0057 and 1.00052 respectively. Thus the losses are gigantic if σ is close to 1 and negligible if σ is very large.

where the parameters β_n are positive and sum to 1 and s is a parameter which satisfies the inequalities $0 < s \le 1$. Thus f(q) is a mean of order s.

Assume that all products are available in a period and purchasers face the positive prices $p \equiv (p_1,...,p_N) >> 0_N$. The first order necessary (and sufficient) conditions (provided that $s \le 1$) that can be used to solve the unit cost minimization problem defined by (1) are the following conditions:

$$(29)\ p_n = \lambda\beta_n q_n^{s-1}\ ; \qquad n = 1,...,N;$$
$$(30)\ 1 = [\Sigma_{n=1}^N \beta_n q_n^s]^{1/s}.$$

Multiply both sides of equation n in (29) by q_n and sum the resulting N equations. This leads to the equation $\Sigma_{n=1}^N p_n q_n = \lambda\Sigma_{n=1}^N \beta_n q_n^s$. Solve this equation for λ and use this solution to eliminate the λ in equations (29). The resulting equations (where equation n is multiplied by q_n) are the following ones:

$$(31)\ p_n q_n/\Sigma_{i=1}^N p_i q_i = \beta_n q_n^s/\Sigma_{i=1}^N \beta_i q_i^s; \qquad n = 1,...,N.$$

The equations (29) and (30) can be used to obtain an explicit solution for $q_1,...,q_N$ and λ as functions of the price vector p.[22] Use these solution functions to form the unit cost function, c(p) equal to $\Sigma_{n=1}^N p_n q_n(p)$. This function turns out to be the following one:[23]

$$(32)\ c(p) = [\Sigma_{n=1}^N \beta_n^{1/(1-s)}\, p_n^{s/(s-1)}]^{(s-1)/s}\ .$$

It can be seen that c(p) is proportional to a mean of order r where $r = s/(s-1)$. Thus if f(q) is the CES utility function defined by (28), then the corresponding elasticity of substitution is $\sigma = 1 - r = 1 - [s/(s-1)] = -\, 1/(s-1) = 1/(1-s)$. Note that our assumption that s satisfies $0 < s \le 1$ implies that σ satisfies $1 < \sigma \le \infty$.

If purchasers maximize the CES utility function defined by (28) when they face the positive price vector p, the utility maximizing q will satisfy the share equations (31). If we evaluate equations (31) using the period t price and quantity data, we obtain the following system of estimating equations, assuming that all products are available in all periods:

$$(33)\ s_n^t \equiv p_n^t q_n^t/\Sigma_{i=1}^N p_i^t q_i^t = \beta_n (q_n^t)^s/\Sigma_{i=1}^N \beta_i (q_i^t)^s; \qquad t = 1,...,T;\ n = 1,...,N.$$

It can be seen that the right hand sides of equations (33) are homogeneous of degree 0 in the parameters $\beta_1,...,\beta_N$ so a normalization of these parameters is required for the identification of the parameters. The normalization $\Sigma_{n=1}^N \beta_n = 1$ can be replaced by an equivalent normalization such as $\beta_N = 1$.

[22] Under these conditions, the first order necessary conditions (29) and (30) for solving the unit cost minimization problem are also sufficient conditions.

[23] Explicit solutions for the $q_n(p)$ can be obtained by using Shephard's Lemma; i.e., $q_n(p) = \partial c(p)/\partial p_n$ for n = 1,...,N where c(p) is defined by (32).

We now consider the case where not all products are available in all periods. The parameter s is assumed to be greater than 0 (and less than or equal to 1 so that the resulting CES utility function is concave). If product n is not available in period t, we can set $q_n^t = 0$ and $(q_n^t)^s = (0)^s = 0$ and thus product n will drop out of the utility function. Thus if we simply set quantities equal to 0 when the corresponding products are not available in a period, the overall CES utility function evaluated at the period t quantity data (with the appropriate 0 values inserted), $f(q^t)$, will be equal to $[\Sigma_{n\in I(t)} \beta_n (q_n^t)^s]^{1/s}$, the utility function f^t which is defined over just the products that are actually available during period t; i.e., the following equations will be satisfied where we define u_{CES}^t as the *period t aggregate CES utility or quantity* (or volume) *level*:

$$(34)\ u_{CES}^t = f(q^t) \equiv [\Sigma_{n=1}^N \beta_n (q_n^t)^s]^{1/s} = [\Sigma_{n\in I(t)} \beta_n (q_n^t)^s]^{1/s} ; \qquad t= 1,...,T$$

where the last equality follows under the assumption that s > 0.

Once the aggregate utility or quantity levels u_{CES}^t have been defined by equations (34), the corresponding *CES fixed base quantity index* can be defined as follows:[24]

$$\begin{aligned}(35)\ Q_{CES}^t &\equiv u_{CES}^t/u_{CES}^1 ; \qquad t = 1,...,T\\ &\equiv f(q^t)/f(q^1) ;\\ &= [\sum_{i\in I(t)} \beta_i (q_i^t)^s]^{1/s} / [\sum_{i\in I(1)} \beta_i (q_i^1)^s]^{1/s}\\ &= [\text{ Index } 1^*]\times[\text{Index } 2^*]\times[\text{Index } 3^*]\end{aligned}$$

where the above indexes are defined as follows:

$$(36)\ \text{Index } 1^* \equiv [\sum_{i\in I(t)\cap I(1)} \beta_i (q_i^t)^s]^{1/s} / [\sum_{i\in I(1)\cap I(t)} \beta_i (q_i^1)^s]^{1/s} ;$$
$$(37)\ \text{Index } 2^* \equiv [\sum_{i\in I(t)} \beta_i (q_i^t)^s]^{1/s} / [\sum_{i\in I(1)\cap I(t)} \beta_i (q_i^t)^s]^{1/s} ;$$
$$(38)\ \text{Index } 3^* \equiv [\sum_{i\in I(1)\cap I(t)} \beta_i (q_i^1)^s]^{1/s} / [\sum_{i\in I(1)} \beta_i (q_i^1)^s]^{1/s} .$$

Note that Index 1^* defines a CES quantity index over the set of commodities that are available in both periods t and 1. Denote the CES utility function f^{t*} that has the same β_n parameters as in definition (28) but is now defined over only products that are available in periods 1 and t:

$$(39)\ f^{t*}(q) \equiv [\sum_{i\in I(t)\cap I(1)} \beta_i (q_i)^s]^{1/s} ; \qquad t = 1,2,...,T.$$

Utility maximizing behavior on the part of purchasers will imply that the following counterparts to equations (17) will hold:

$$\begin{aligned}(40)\ s_i^{t*} &\equiv p_i^t q_i^t/\Sigma_{n\in I(t)\,)\cap I(1)}\, p_n^t q_n^t \qquad t = 1,...,T;\ i\in I(1)\cap I(t)\\ &= \beta_i (q_i^t)^s/\sum_{n\in I(t)\,)\cap I(1)} \beta_n (q_n^t)^s\\ &= \beta_i (q_i^t)^s/f^{t*}(q^t)^s\end{aligned}$$

[24] In the algebra which follows, the period 1 quantity vector can be replaced by the quantity vector of any period.

where the third equality follows using definitions (39).

Note that Index 1^* is equal to $f^{t*}(q^t)/f^{t*}(q^1)$ and the modified Sato-Vartia formula (7) (restricted to commodities n that are present in periods 1 and t and where quantities and prices are interchanged in the formula) can be used to calculate this index using the observed price and quantity data for the products that are available in both periods 1 and t.

As usual, we assume that there is at least one product that is present in periods 1 and t for each t. Let product i be any one of these common products for a given t. Then the ith share equation in (33) and (40) for period t can be rearranged to give us the following two equations:

(41) $\beta_i(q_i^t)^s = [\sum_{n\in I(t)} \beta_n (q_n^t)^s]\, p_i^t q_i^t/[\sum_{n\in I(t)} p_n^t q_n^t]$;
(42) $\beta_i(q_i^t)^s = [\sum_{n\in I(1)\cap I(t)} \beta_n (q_n^t)^s]\, p_i^t q_i^t/[\sum_{n\in I(1)\cap I(t)} p_n^t q_n^t]$.

Equating (41) to (42) leads to the following equations:

(43) $\sum_{n\in I(t)} \beta_n (q_n^t)^s/\sum_{n\in I(1)\cap I(t)} \beta_n (q_n^t)^s = \sum_{n\in I(t)} p_n^t q_n^t/\sum_{n\in I(1)\cap I(t)} p_n^t q_n^t$
$= \lambda^t$

where the last equality follows using definition (18). Now take the 1/s root of both sides of (43) and use definition (37) in order to obtain the following equality:[25]

(44) Index $2^* = [\lambda^t]^{1/s} = [\sum_{i\in I(t)} p_i^t q_i^t/\sum_{i\in I(1)\cap I(t)} p_i^t q_i^t]^{1/s}$.

Again assume that product i is available in periods 1 and t. Rearrange the share equations (33) and (40) for t = 1 and product i and we obtain the following two equations:

(45) $\beta_i(q_i^1)^s = [\sum_{n\in I(1)} \beta_n (q_n^1)^s]\, p_i^1 q_i^1/[\sum_{n\in I(1)} p_n^1 q_n^1]$;
(46) $\beta_i(q_i^1)^s = [\sum_{n\in I(1)\cap I(t)} \beta_n (q_n^1)^s]\, p_i^1 q_i^1/[\sum_{n\in I(1)\cap I(t)} p_n^1 q_n^1]$.

Equating (45) to (46) leads to the following equations:

(47) $\sum_{n\in I(1)\cap I(t)} \beta_n (q_n^1)^s/\sum_{n\in I(1)} \beta_n (q_n^1)^s = \sum_{n\in I(1)\cap I(t)} p_n^1 q_n^1/\sum_{n\in I(t)} p_n^1 q_n^1$
$= \mu^t$

where the last equality follows using definition (19). Now take the 1/s root of both sides of (47) and use definition (38) in order to obtain the following equality:[26]

[25] If new products become available in period t that were not available in period 1, then $\lambda^t > 1$. Recall that the elasticity of substitution in terms of s is equal to $\sigma \equiv 1/(1-s)$ where s satisfies $0 < s \le 1$. Thus as s increases, σ also increases. With $\lambda^t > 1$, the limit of $(\lambda^t)^{1/s}$ as s approaches 0 is $+\infty$ and the limit of $(\lambda^t)^{1/s}$ as s approaches 1 is $\lambda^t > 1$. Thus Index 2^* is a decreasing function of s for $0 < s \le 1$. Suppose that $\lambda^t = 1.05$ and s = 0.005, 0.01, 0.1, 0.3, 0.5, 0.9, 0.99 and 0.999. Then $\sigma = 1/(1-s)$ is equal to 2, 10, 100 and 1000. The corresponding Index 2^* values are 17292.6, 131.5, 1.629, 1.177, 1.103, 1.056, 1.051 and 1.050. Thus the value of Index 2^* is very large when σ is close to 1 and its value declines to λ^t as σ approaches plus infinity.

(48) Index $3^* = [\mu^t]^{1/s} = [\sum_{n\in I(1)\cap I(t)} p_n^1 q_n^1 / \sum_{n\in I(t)} p_n^1 q_n^1]^{1/s}$.

If s is known or has been estimated, then Index 2^* and Index 3^* can readily be calculated as simple ratios of sums of observable expenditures raised to the power1/s. Note that $[\sum_{i\in I(t)} p_i^t q_i^t / \sum_{i\in I(1)\cap I(t)} p_i^t q_i^t] \geq 1$. If period t has products that were not available in period 1, then the strict inequality will hold and since $1/s > 0$, it can be seen that Index 2 will be greater than unity. Similarly, $[\sum_{i\in I(1)\cap I(t)} p_i^1 q_i^1 / \sum_{i\in I(1)} p_i^1 q_i^1] \leq 1$. If period 1 has products that are not available in period t, then the strict inequality will hold and since$1/s > 0$, it can be seen that Index 3 will be less than unity.

The interpretations of Index 2^* and Index 3^* are not as simple as were the interpretations for Index 2 and Index 3. These indexes reflect the effects of changes in the availability of products but they also reflect increases and decreases in utility that are due to changes in total expenditures that vary across periods. In section 6 below, we will explain how the methodology developed in this section can be modified to provide valid counterparts to Feenstra's unit cost function methodology that was explained in section 2 above. However, what is true is that the utility ratio decomposition that is defined by (35) above can be implemented using observable prices and quantities provided an estimate for s or σ is available. In this respect, the utility function decomposition (35) is similar to Feenstra's unit cost function decomposition defined earlier by (12).

Note that the purchasers' system of utility maximizing nonzero share equations for each period that is defined by equations (33) can be rewritten as the following system of equations:[27]

(49) $s_i^t = \beta_i (q_i^t)^s / \sum_{n\in I(t)} \beta_n (q_n^t)^s$; $\quad t = 1,...,T;\ i\in I(t)$.

Recall that the purchasers' system of cost minimizing share equations using the CES unit cost function defined by (2) was given by equations (11); $s_i^t = \alpha_i (p_i^t)^r / \sum_{n\in I(t)} \alpha_n (p_n^t)^r$ for $t = 1,...,T$ and $i\in I(t)$. Equations (11) and (49) have exactly the same dependent variables but they have totally different independent variables: period t prices for equations (11) and period t quantities for equations (49). In Sections 5 and 6 below, we will use some scanner data and estimate both systems of equations and see which system fits the data best. In the following section, we will explain our data set.

4. Scanner Data for Sales of Frozen Juice

[26] If some products that were available in period 1 become unavailable in period t, then $\mu^t < 1$. Index 3^* evaluated at period 1 prices equals $(\mu^t)^s = (\mu^t)^{1/(1-\sigma)}$ and is an increasing function of σ for $1 < \sigma < +\infty$. With $\mu^t < 1$, the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches 1 is 0 and the limit of $(\mu^t)^{1/(1-\sigma)}$ as σ approaches $+\infty$ is $\mu^t < 1$. Suppose that $\mu^t = 0.95$ and σ equals 1.005, 1.010, 1.111, 1.429, 2, 10, 100 and 1000.. Then Index 3^* will equal 0.000035, 0.0059, 0.5987, 0.8428, 0.9025, 0.9446, 0.9495 and 0.9500.

[27] Because of the separability properties of the CES utility function, the assumption of utility maximizing behavior on the part of CES purchasers will imply that the share equations (40) and (49) will hold simultaneously.

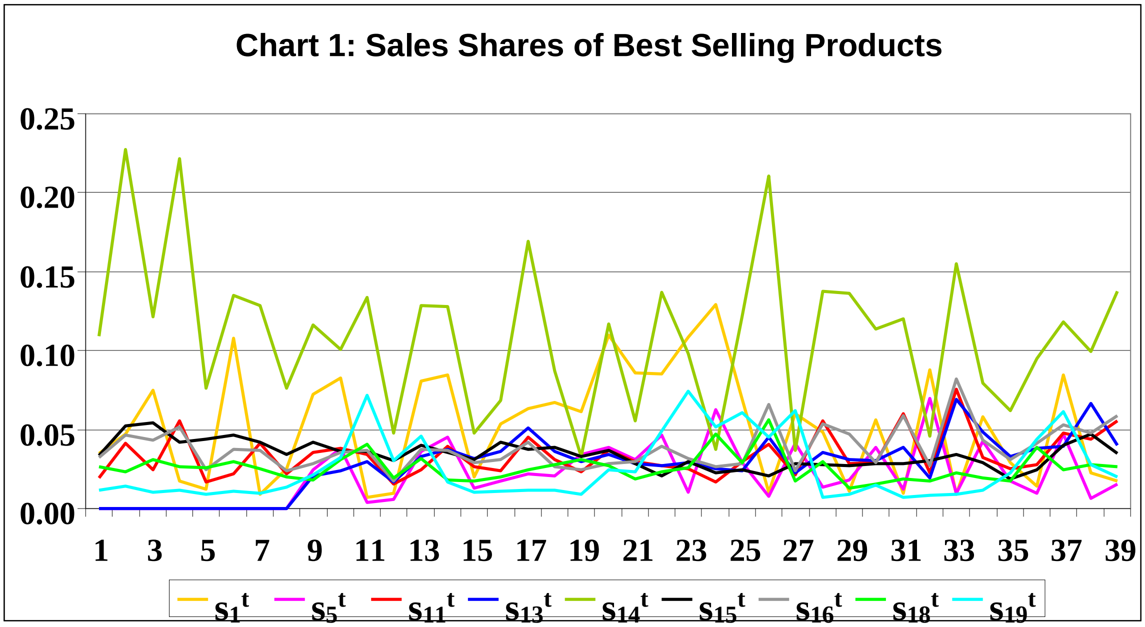
We will use the data from Store Number 5[28] in the Dominick's Finer Foods Chain of 100 stores in the Greater Chicago area on 19 varieties of frozen orange juice for 3 years in the period 1989-1994 in order to test out the CES models explained in the previous two sections; see the University of Chicago (2013) for the micro data.

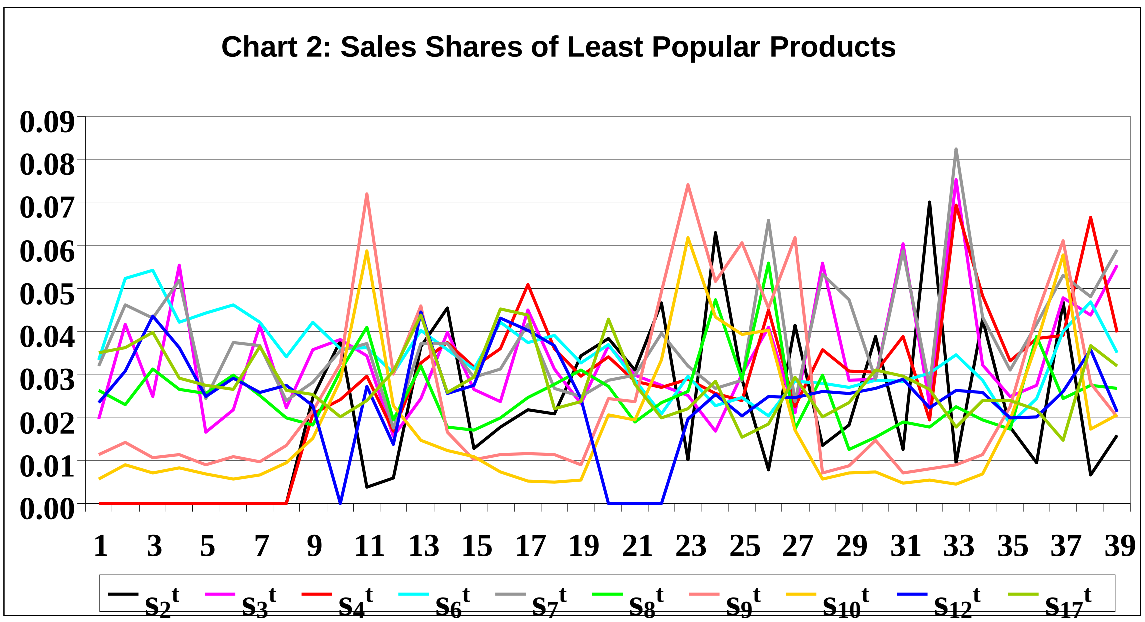
The micro data are weekly quantities sold of each product and the corresponding unit value price. However, our focus is on calculating a monthly index and so the weekly price and quantity data need to be aggregated into monthly data. Since months contain varying amounts of days, we are immediately confronted with the problem of converting the weekly data into monthly data. We decided to side step the problems associated with this conversion by aggregating the weekly data into *pseudo-months* that consist of 4 consecutive weeks.

In the Appendix, the "monthly" data for quantities sold and the corresponding unit value prices for the 19 products are listed in Tables A1 and A2. There were no sales of Products 2 and 4 for "months" 1-8 and there were no sales of Product 12 in "month" 10 and in "months" 20-22. Thus there is a new and disappearing product problem for 20 observations in this data set. Later in this paper, we will impute Hicksian reservation prices for these missing products and these estimated prices are listed in Table A2 in italics. The corresponding imputed quantity for a missing observation is set equal to 0.

Expenditure or sales shares, $s_i^t \equiv p_i^t q_i^t / \Sigma_{n=1}^{19} p_n^t q_n^t$, were computed for products i = 1,...,19 and "months" t = 1,...,39.[29] We computed the sample average expenditure shares for each product. The best selling products were products 1, 5, 11, 13, 14, 15, 16, 18 and 19. These products had a sample average share which exceeded 4% or a sample maximum share that exceeded 10%. These shares are listed in Table A3 in the Appendix. The remaining 10 products are the lesser selling products and these shares are listed in Table A4 in the Appendix. See Charts 1 and 2 below for plots of these shares.

[28] This store is located in a North-East suburb of Chicago.
[29] In what follows, we will describe our 4 week "months" as months.





It can be seen that there is tremendous volatility in the sales shares, both for the best selling and least popular products. In Charts 3 and 4 below, we plot the relative prices for the best selling and least popular products. The relative price for product i in period t is defined as $p_{Ri}^t \equiv p_i^t/p_i^1$ for i = 1,...,19 and t = 1,...,39.[30]

[30] For convenience, the imputed reservation prices for products 2, 4 and 12 were used in Chart 4.

Chart 3: Normalized Prices for Best Selling Products

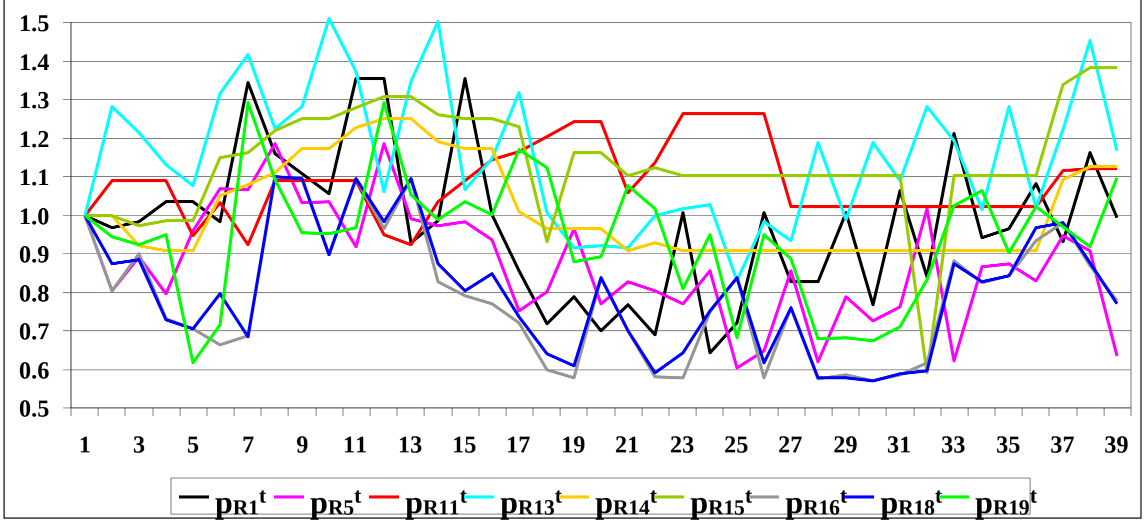
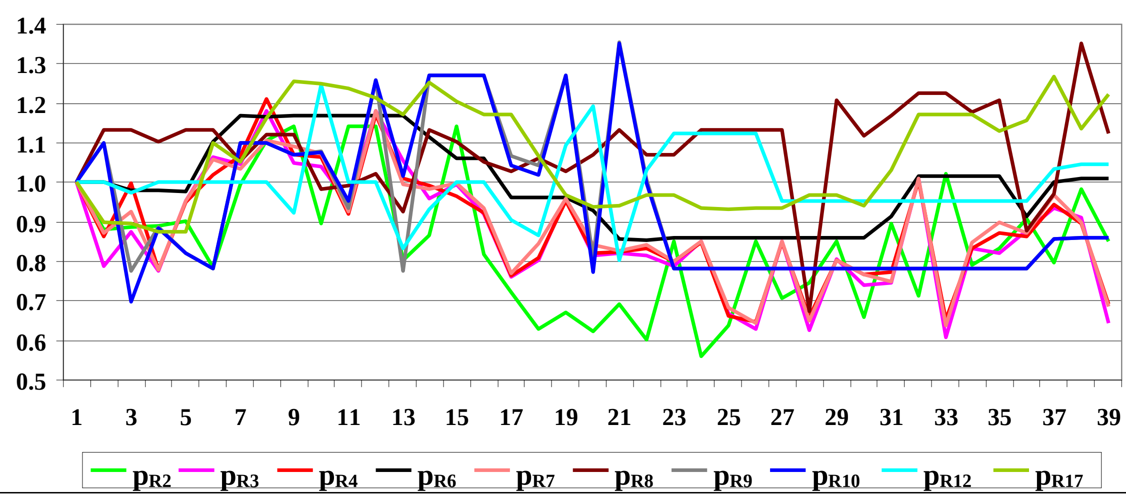


Chart 4: Normalized Prices for Least Popular Products

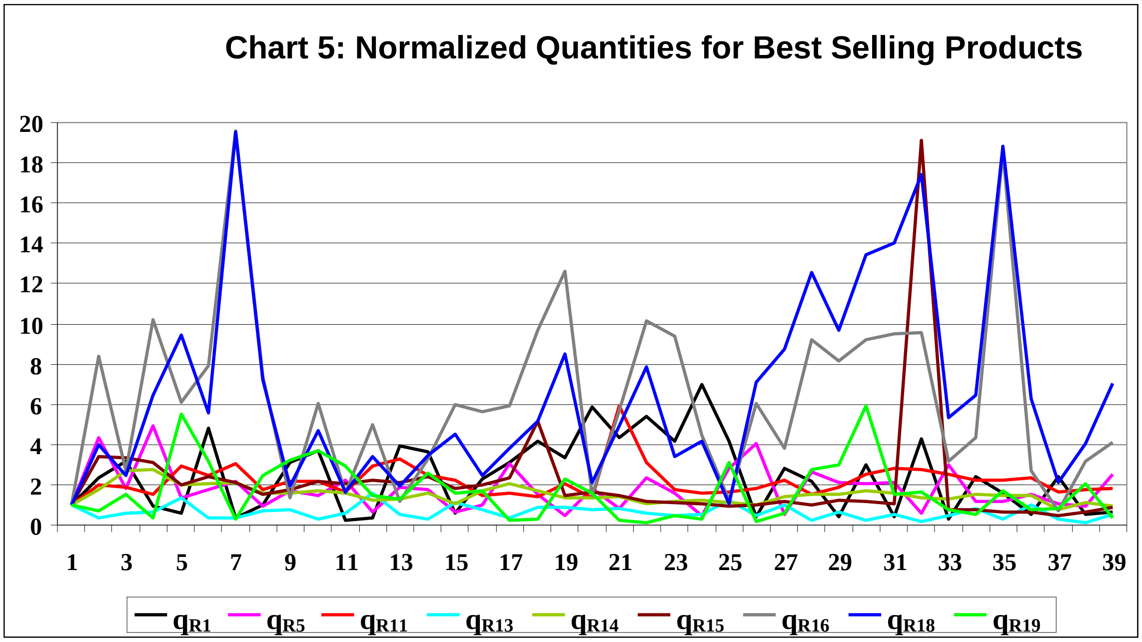


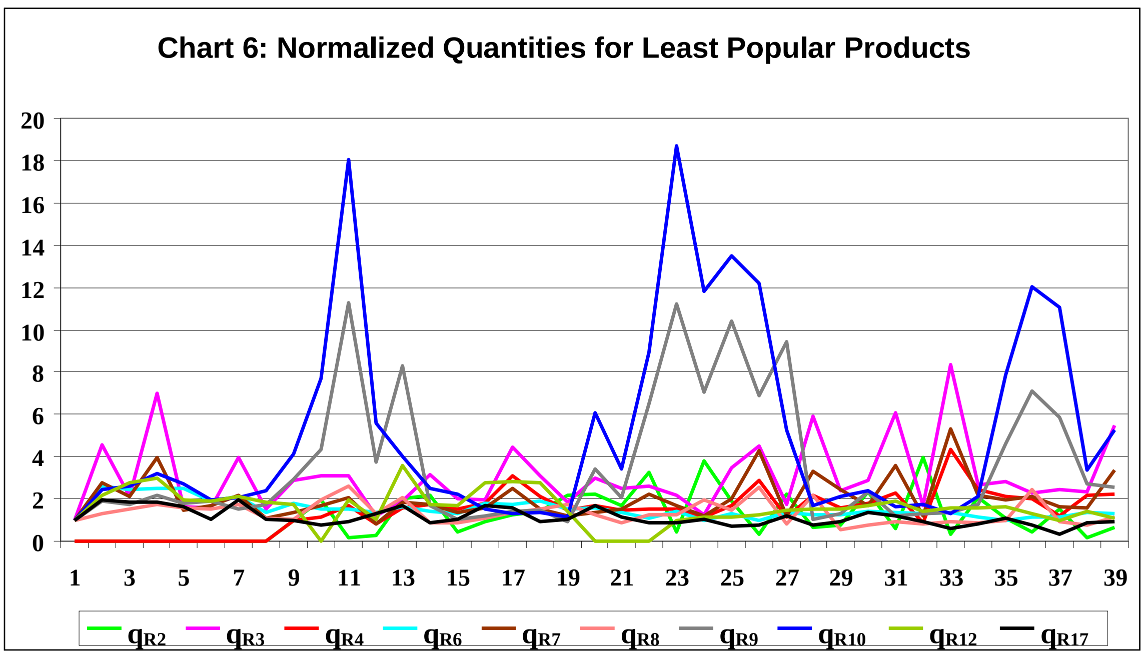
It can be seen that there is also tremendous volatility in product prices for both the best selling and least popular products.

Finally, in Charts 4 and 5 below, we plot the relative quantities for the best selling and least popular products. The relative quantity for product i in period t is defined as $q_{Ri}^t \equiv q_i^t/q_i^1$ for i = 1,...,19 and t = 1,...,39.[31]

---

[31] However, $q_2^t = 0$ and $q_4^t = 0$ for t =1,...,8. Thus we define q_{R2}^t as q_2^t/q_2^9 and q_{R4}^t as q_4^t/q_4^9 for t = 1,...,8.





It can be seen that the volatility of quantities (relative to month 1) greatly exceeds the volatility of prices (relative to month 1). When a product goes on sale at say ½ of its normal price, the volume sold of the product can easily increase 10 fold or more.

In the following section, we will use this data set in order to implement Feenstra's unit cost function methodology for the treatment of new and disappearing products.

5. The Estimation of CES Preferences Using the Unit Cost Function Approach

We assume that purchaser preferences are defined by the utility function that is dual to the CES unit cost function defined by (2) in Section 2 above. Recall that the system of estimating equations (5) was obtained for this model.[32] In this section, we compare two methods for estimating the elasticity of substitution for this model: the first method uses the nonlinear equations in (5), and the second method is a simplified version of the estimator proposed by Feenstra (1994).

Using the first method, we have a nonlinear system of 19 estimating equations where the ith equation for period t is $s_i^t = \alpha_i(p_i^t)^r/\sum_{n=1}^N \alpha_n(p_n^t)^r$ for i = 1,…,19 and t = 1,...,39. We add error terms, ε_i^t, to these equations where $(\varepsilon_1^t,..., \varepsilon_{19}^t)$ is assumed to be distributed as a multivariate normal random variable with mean vector 0_{19} and variance-covariance matrix Σ for t = 1,...,39. In order to identify the α_i, we impose the following normalization:

(50) $\alpha_{19} = 1.$

Since the shares s_i^t sum to one for each period t, all 19 error terms ε_i^t for i = 1,...,19 cannot be distributed independently so we dropped the equation for product 19 from our list of estimating equations.[33]

We chose to estimate the key parameter r in two stages. In the first stage, we set r = 1, so that we obtained the following system of nonlinear estimating equations:

(51) $s_i^t = [\alpha_i p_i^t/\sum_{n=1}^N \alpha_n p_n^t] + \varepsilon_i^t$; t = 1,...,39; i = 1,...,18.

We used the nonlinear regression software package in Shazam[34] to estimate the unknown α_i in equations (51). The final log likelihood turned out to be 2034.884. The equation by equation R^2 values were as follows:[35] 0.6138, 0.1277, 0.5476, 0.4875, 0.2376, 0.1191, 0.5014, 0.0172, 0.0761, 0.047, 0.0016, 0.4232, 0.6578, 0.0012, 0.5826, 0.2973, 0.1481 and 0.2323. Thus the fits for this preliminary regression were not very good but this is to be expected: *Model 1* defined by equations (51) corresponds to preferences that exhibit *no* substitution between products, which is implausible for closely related products.

---

[32] The actual estimating equations are defined by equations (27), which take into account the prices and quantities which are missing due to product unavailability. We will explain how we dealt with the problem of unavailable commodities in subsequent footnotes.

[33] There is a problem with this stochastic specification: for 20 observations, the price of a product that is not available is taken to be $+\infty$ in the nonlinear regression model that corresponds to equations (5). If $r < 0$, then $(+\infty)^r = 0$ and the corresponding quantity and expenditure share will also be 0 so in our regression model, there will be 20 observations (out of a total of 702) that will automatically have 0 error terms. In order to apply standard nonlinear regression software for systems of equations, we have temporarily ignored this problem.

[34] See White (2004). For the 20 observations where prices and quantities were not available, we set the corresponding prices, quantities and shares equal to 0 in the system of estimating defined by equations (51).

[35] The R^2 concept that we used is the square of the correlation coefficient between the dependent variables and the corresponding predicted variables.

The estimated α_n coefficients generated by the above special case of CES preferences were used as starting coefficient values (along with r = 1 as a starting value) in the unit cost CES model defined by the normalization (50) and equations (52):

$$(52)\quad s_i^t = [\alpha_i (p_i^t)^r / \sum_{n=1}^N \alpha_n (p_n^t)^r] + \varepsilon_i^t ; \qquad t = 1,...,39;\ i = 1,...,18.$$

Again, Shazam was used to estimate the 19 unknown parameters in equations (52).[36] The final log likelihood for this *Model 2* was 2195.039, an increase of 160.155 over the previous Model 1 regression for adding one parameter. The estimate for r was −2.8041 with a standard error equal to 0.12939. Hence the resulting point estimate for the elasticity of substitution is $\sigma \equiv 1 - r = 3.8041$. Thus there is a considerable amount of substitution between the 19 frozen juice products. The equation by equation R^2 values were as follows: 0.6357, 0.7179, 0.6407, 0.8221, 0.3619, 0.0051, 0.6652, 0.0268, 0.0572, 0.0206, 0.0109, 0.4286,0.7419, 0.0781, 0.8050, 0.3370, 0.1589 and 0.2673. These R^2 values are considerably higher than the corresponding ones from the first regression model. However, the average R^2 was only equal to 0.3767 which is not very satisfactory.

While the estimate of $\sigma = 3.8041$ for the elasticity of substitution seems like a reasonable amount of substitution, it is considerably lower than the estimate that we shall obtain in the next section from the utility function approach. Rather than proceed with this initial estimate, we now present an alternative method for estimating the elasticity that is a simplified version of the estimator in Feenstra (1994). A key feature of this estimator is that it takes into account measurement error in the prices, which can arise because we are aggregating the prices over time, i.e. over weeks in our initial data, and over months in the dataset that we use to construct all of our price indexes. So the prices in the data are actually *unit values* of each product defined over these time intervals. We will find that the alternative estimator of Feenstra (1994), which controls for this measurement error, provides an estimate of σ that is much higher than 3.8041.

We begin by noting that the prices p_i^t that are listed in Appendix A are not the true, minute-by-minute selling prices in the store. The prices p_i^t are *aggregates over time* – weekly or monthly. We refer to these aggregates over time as *unit values*, and we assume that they are related to the *true prices* ρ_i^t by:

$$(53)\quad \ln p_i^t = \ln\rho_i^t + u_i^t ,$$

[36] Recall that we have 20 observed prices p_i^t that are entered as zeros. But $(0)^r$ is not well defined if $r < 0$. Thus for t = 1,...,8, the 0 price terms $(p_2^t)^r$ and $(p_4^t)^r$ on the right hand side of equations (52) were replaced by $\delta^t(p_2^t)^r$ and $\delta^t(p_4^t)^r$ where we set $p_2^t = p_4^t = 1$ for t = 1,...,8 and where δ^t is a dummy variable which takes on the value 0 for t = 1,...,8 and is equal to 1 for t = 9,10,...,39. Similarly, the 0 price terms $(p_{12}^t)^r$ for t = 10 and t = 20, 21 and 22 were replaced by $\delta_{12}^t(1)^r$ where δ_{12}^t is a dummy variable which takes on the value 0 for t = 10, 20, 21 22 and is equal to 1 for other periods t. Thus we set the prices for the 20 missing observations equal to 1 but we nullified terms involving these prices using our dummy variables. With these modifications, the Shazam system nonlinear regression package worked well. The starting log likelihood for the nonlinear regression model defined by this modification of (52) was equal to the final log likelihood for the model defined by (51) which is a check that our modified model behaved in the desired manner.

where u_i^t is the measurement error in the log unit values. We assume that the measurement error in the log unit values,ln p_i^t, is uncorrelated with the logarithms of the true prices, $\ln\rho_i^t$.

Next, consider the share equations (27) but replace the unit value prices p_i^t by the true prices ρ_i^t. Take the natural logarithms of these equations and add error terms to obtain the following equations:

$$(54)\quad \ln s_i^t = \ln\alpha_i + r\ln\rho_i^t - \ln[\Sigma_{n\in I(t)}\, \alpha(\rho_n^t)^r] + \varepsilon_i^t\ ; \qquad t = 1,...,T;\ i\in I(t).$$

The error term ε_i^t can arise due to movements in the share variable that does not reflect CES behavior on the part of the representative consumer. A good example for our frozen juice data – or other scanner data – would be sales that lead to shopping for inventories, which is behavior that lies outside our model.[37]

We will make the usual assumption that the errors in the share equations (54) are uncorrelated with the prices in (53), i.e. these “true” prices are exogenous to the consumer.[38] Furthermore, we shall assume that the measurement errors u_i^t in the unit values is uncorrelated with the errors ε_i^t in the share variables. This latter assumption is motivated by the fact that barcode data typically includes some rather extreme values for the (daily, weekly or monthly) unit values. While the most extreme values can be filtered out, there is inevitably some remaining measurement error that does not appear to be reflected in the sales data, i.e. it is true measurement error.

The challenge now is to obtain a consistent estimate for the elasticity of substitution in the presence of (independent) errors in both the share and the unit value data. Feenstra (1994) argues that the *panel* nature of the dataset – over products and time – provides a method to obtain such a consistent estimate. To show this, the share equation in (54) is simplified by taking first-differences over time to eliminate the nuisance parameters α_i, and then by taking an additional difference with respect to a reference product k to eliminate the summation term:[39]

$$(55)\quad \Delta\ln s_i^t - \Delta\ln s_k^t = r(\Delta\ln\rho_i^t - \Delta\ln\rho_k^t) + \Delta\varepsilon_i^t - \Delta\varepsilon_k^t\ ,$$

$$= r(\Delta\ln p_i^t - \Delta\ln p_k^t) - r(\Delta u_i^t - \Delta u_k^t) + \Delta\varepsilon_i^t - \Delta\varepsilon_k^t; \qquad t = 2,...,T,\ i\in I(t)\cap I(t\text{-}1),\ i\neq k,$$

[37] Feenstra and Shapiro (2003) analyze inventory stockpiling behavior for canned tuna.

[38] The estimator in Feenstra (1994) allows for upward sloping supply curves, so that prices become endogenous, but we ignore that feature of the estimator here.

[39] We assume that the reference product k is available in every period, and in practice, we choose it as the product with highest cumulative sales that is available in every period. In our data set, this is product 13. Our estimation method is somewhat sensitive to the choice of the reference product. The ideal reference product has a large share in every period and a small period to period variance in the shares.

where the second line of (55) is obtained by substituting from (53) for the unit values. This *double differencing* focuses attention on the only parameter that is needed to compute the Feenstra price index, namely, the elasticity of substitution $\sigma \equiv 1 - r$.

To proceed further, it is convenient to define second and cross-moments of the errors and data. These will be used to express our assumptions about terms being uncorrelated, and they will be used in the estimation. For any two variables x and y, define their cross-moment in the data (differenced over time and differenced with respect to product k) as:

$$(56)\quad M_i(x,y) \equiv (1/T_i)\,[\Sigma_{t\in T(i)}\,(\Delta x_i^t - \Delta x_k^t)(\Delta y_i^t - \Delta y_k^t)]\,, \qquad i=1,\ldots,N,\ i\neq k,$$

where the variables Δx_i^t and Δy_i^t are both available in the periods $t\in T(i)$, and the number of such periods is denoted by T_i. For example, if the products x and y are available in *all* periods, then $T(i)=\{2,\ldots,T\}$ and $T_i=T-1$, to allow for the first differencing. These products might be available in fewer periods, however. If x = y then the cross moment defined in (56) becomes a second moment of the variable x. For whatever choice of the variables x and y that we make, the moments are constructed by averaging over time as in (56) *for each* of the products $i=1,\ldots,N$, $i\neq k$, so that because of the panel nature of the dataset, we have a cross-section of such moments.

With this definition, our assumptions that certain terms are uncorrelated can be expressed conveniently as,

$$(57)\quad E[M_i(\varepsilon,\ln\rho)]=0,\ \ E[M_i(u,\ln\rho)]=0 \text{ and } E[M_i(\varepsilon,u)]=0, \qquad i=1,\ldots,N,\ i\neq k,$$

where E denotes the expected value. The first of these assumptions is that prices are exogenous to the consumer; the second is that the measurement error in the unit values is uncorrelated with the true prices, and the third is that the errors in the shares and in the unit values are uncorrelated. We now show how these moment conditions can be combined to obtain a consistent estimate of the elasticity of substitution, in what Feenstra (1994) refers to as a generalized method of moment (GMM) estimator.

The cross-moment between the errors in shares and in unit values can be written as:

$$\begin{aligned}
(58)\quad M_i(\varepsilon,u) &\equiv (1/T_i)\,[\Sigma_{t\in T(i)}\,(\Delta\varepsilon_i^t - \Delta\varepsilon_k^t)\,(\Delta u_i^t - \Delta u_k^t)]\\
&= (1/T_i)\{\Sigma_{t\in T(i)}\,(\Delta\varepsilon_i^t - \Delta\varepsilon_k^t)[(\Delta\ln p_i^t - \Delta\ln p_k^t) - (\Delta\ln\rho_i^t - \Delta\ln\rho_k^t)]\}\\
&= \{(1/T_i)\,\Sigma_{t\in T(i)}\,[\Delta\ln s_i^t - \Delta\ln s_k^t - r(\Delta\ln p_i^t - \Delta\ln p_k^t) + r(\Delta u_i^t - \Delta u_k^t)]\\
&\qquad \times (\Delta\ln p_i^t - \Delta\ln p_k^t)\} - M_i(\varepsilon,\ln\rho)\\
&= M_i(\ln s,\ln p) - rM_i(\ln p,\ln p) + rM_i(u,\ln p) - M_i(\varepsilon,\ln\rho)\\
&= M_i(\ln s,\ln p) - rM_i(\ln p,\ln p) + rM_i(u,\ln\rho) + rM_i(u,u) - M_i(\varepsilon,\ln\rho),
\end{aligned}$$

where the second line uses (54) to express the measurement error $(\Delta u_i^t - \Delta u_k^t)$; the third follows by combining the term $(\Delta\ln\rho_i^t - \Delta\ln\rho_k^t)$ with the share error $(\Delta\varepsilon_i^t - \Delta\varepsilon_k^t)$ to obtain

$M_i(\varepsilon,\ln\rho)$, and then re-expressing that error in full using (55); the fourth line follows from definition of the various cross-moments; and the last line follows because $M_i(u,\ln p) = M_i(u,\ln\rho) + M_i(u,u)$, from (53). It is convenient to rewrite (58) as,

(59) $M_i(\ln p,\ln p) = (1/r)M_i(\ln s,\ln p) + M_i(u,u) + Error_i$, for $i=1,\ldots,N$, $i\neq k$,

where $Error_i$ is defined as follows:

(60) $Error_i \equiv M_i(u,\ln p) - (1/r)[M_i(\varepsilon,\ln p)+ M_i(\varepsilon,u)]$.

What we have obtained in (59) is a simple linear regression involving moments of the data, which can be run over the products $i=1,\ldots,N$, $i\neq k$. The error in this regression, defined in (60), consists of a sum of the moment conditions that we have discussed in (57). By running OLS on the regression (59) we will be making the sum over products of these squared errors as small as possible, which is the goal of the GMM estimator.

Examining this regression more closely, the dependent variable is the second moment of the log unit values (differenced with respect to time and with respect to product k). The first term on the right is the cross moment of the market shares and unit values, and the coefficient of this term is (1/r). The second term on the right is the sample variance of the measurement error in the unit values for the products. That variance is not observed in the data, but we assume that this (population) variance is *constant* across the products, so that this second term is replaced by a *constant term* in the regression. As already mentioned, the remaining terms are errors that are minimized by running OLS. Feenstra (1994) discusses how more efficient estimates can be obtained by running *weighted* least squares, and how the correct standard error of (1/r) can be obtained from a slightly different version of (59).[40]

Running the OLS regression for the frozen juice data result in $\sigma \equiv 1 - r = 7.9891$ for weekly data, and $\sigma = 5.9900$ from monthly data. Thus, we see that aggregating over time from weeks to months does result in a lower estimate of the elasticity of substitution. But the estimate of 5.9900 from the monthly data is still considerably higher than the estimate of $\sigma = 3.8041$ that was obtained from the direct estimation of (52), as described above. When computing the Feenstra Indexes 2 and 3, we use the estimate of $\sigma = 5.9900$ obtained from monthly data, for consistency with the monthly data used in later estimates.

[40] Specifically, instead of constructing the second moments of the data initially, the regression can be run over all $i\neq k$, and $t\in T(i)$ by using $(\Delta\ln p_i^t - \Delta\ln p^t)^2$ as the dependent variable, which is regressed on a constant term and on $(\Delta\ln p_i^t - \Delta\ln p_k^t)$ $(\Delta\ln s_i^t - \Delta\ln s_k^t)$. Instrumental variables (IV) which are indicator variables for each product are used to estimate this regression. In the first stage we regress $(\Delta\ln p_i^t - \Delta\ln p_k^t)$ $(\Delta\ln s_i^t - \Delta\ln s_k^t)$ on these IV, over all $i\neq k$, and $t\in T(i)$, and the predicted values from these regression equal the averages $M_i(l\ s,\ln p)$ repeated T_i times for each product $i\neq k$. Thus, this IV regression is equivalent to a WLS version of (59), where T_i is used as the weight for each product. This IV regression gives the estimate (standard error) for σ of 5.9845 (0.847), which is nearly equal to the unweighted estimate of 5.9900 obtained from running OLS on (59). Feenstra (1994) further discusses how more efficient estimates of σ can be obtained from another weighted regression.

We cannot use the results of the Feenstra method of moments estimator for σ in order to directly compute the CES monthly price levels because we need estimates for the nuisance parameters (the estimated α_n^*) in order to do this. However, we can define the CES price level for month t, P_{CES}^{t*}, as follows using our earlier Model 2 estimated coefficients:

$$(61)\ P_{CES}^{t*} \equiv [\sum_{n\in I(t)} \alpha_n^* (p_n^t)^r]^{1/r} ; \qquad t = 1,...,39.$$

Note that the prices of products that were not available in period t do not appear in the terms on the right hand side of equation t in equations (61). The normalized CES price index for month t is defined as $P_{CES}^t \equiv P_{CES}^{t*}/P_{CES}^{1*}$ for t = 1,...,39. This econometrically based price index for frozen juice is listed in Table 1 below along with a number of "traditional" price indexes.

Table 1: CES and Feenstra Price Indexes and Sato-Vartia, Fisher, Laspeyres, Paasche Fixed Base and Chained Maximum Overlap Price Indexes

t	P_{CES}^t	P_{FEEN}^t	P_{SV}^t	P_F^t	P_{FCh}^t	P_l^t	P_{LCh}^t	P_P^t	P_{PCh}^t
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00508	0.99711	0.99711	1.00218	1.00218	1.08991	1.08991	0.92151	0.92151
3	0.99521	1.00504	1.00504	1.02342	1.01124	1.06187	1.12136	0.98637	0.91193
4	0.94121	0.93679	0.93679	0.93388	0.94265	1.00174	1.06797	0.87061	0.83202
5	0.91406	0.93730	0.93730	0.93964	0.93715	0.98198	1.11998	0.89913	0.78417
6	1.00557	1.04223	1.04223	1.03989	1.04075	1.13639	1.27665	0.95159	0.84844
7	1.08119	1.08505	1.08505	1.05662	1.10208	1.22555	1.42086	0.91097	0.85481
8	1.15826	1.25882	1.25882	1.15739	1.26987	1.17446	1.75897	1.14057	0.91676
9	1.12605	1.22598	1.23745	1.15209	1.24778	1.17750	1.73986	1.12722	0.89487
10	1.13885	1.22532	1.23111	1.14617	1.24137	1.21100	1.78937	1.08481	0.86120
11	1.11631	1.19653	1.20887	1.14088	1.22950	1.19184	1.85291	1.09210	0.81584
12	1.09576	1.18382	1.19602	1.12760	1.22009	1.21172	2.00384	1.04932	0.74288
13	1.07865	1.17805	1.19020	1.10698	1.20731	1.15736	2.16323	1.05880	0.67380
14	1.09705	1.20718	1.21962	1.13419	1.23863	1.19572	2.29212	1.07582	0.66934
15	1.02962	1.13129	1.14295	1.05579	1.15978	1.12363	2.30484	0.99205	0.58359
16	1.01928	1.12715	1.13877	1.05099	1.15371	1.09373	2.32686	1.00993	0.57204
17	0.95211	1.05805	1.06896	0.98640	1.08568	1.07191	2.27306	0.90771	0.51855
18	0.86187	0.95922	0.96911	0.89490	0.98385	0.96788	2.12683	0.82742	0.45512
19	0.86582	0.96690	0.97687	0.89032	0.99122	0.97566	2.19851	0.81244	0.44690
20	0.86905	0.96291	0.96805	0.89016	0.99104	1.04652	2.35818	0.75716	0.41649
21	0.87341	0.97380	0.97900	0.89453	1.00061	1.01001	2.46345	0.79225	0.40643
22	0.84816	0.93449	0.93948	0.85466	0.95983	0.96827	2.42222	0.75438	0.38034
23	0.85830	0.94742	0.95627	0.88842	0.97730	0.94697	2.52523	0.83349	0.37823
24	0.87495	0.93760	0.94636	0.88930	0.96178	0.95666	2.59808	0.82668	0.35604
25	0.77932	0.85935	0.86738	0.80421	0.88017	0.83788	2.52526	0.77189	0.30678
26	0.81396	0.89729	0.90567	0.84644	0.91938	0.92401	2.82064	0.77539	0.29967
27	0.87388	0.95499	0.96391	0.88641	0.98171	0.92853	3.20399	0.84620	0.30080
28	0.79552	0.87825	0.88645	0.81528	0.90580	0.90110	3.25314	0.73763	0.25221
29	0.83771	0.92684	0.93550	0.85705	0.95671	0.91523	3.55936	0.80258	0.25715
30	0.83003	0.91453	0.92307	0.84508	0.94446	0.92571	3.60564	0.77147	0.24739
31	0.85502	0.93991	0.94869	0.87333	0.97386	0.94494	3.80130	0.80715	0.24949
32	0.90137	0.99276	1.00204	0.89973	1.00016	1.04403	4.32811	0.77538	0.23112
33	0.89378	1.00742	1.01683	0.92673	1.02452	1.01783	5.40982	0.84377	0.19402
34	0.93133	1.03371	1.04336	0.95385	1.05227	0.99801	5.91196	0.91165	0.18729

35	**0.98200**	**1.08487**	**1.09500**	**0.98690**	**1.10820**	**1.05351**	**6.39424**	**0.92451**	**0.19206**
36	**0.94110**	**1.06466**	**1.07461**	**0.96237**	**1.08529**	**1.00318**	**6.63992**	**0.92322**	**0.17739**
37	**1.03787**	**1.17261**	**1.18356**	**1.04948**	**1.18995**	**1.09380**	**7.44751**	**1.00696**	**0.19013**
38	**1.07081**	**1.19584**	**1.20701**	**1.09545**	**1.21560**	**1.16242**	**7.84172**	**1.03234**	**0.18844**
39	**0.91069**	**1.03639**	**1.04607**	**0.94999**	**1.05918**	**1.02873**	**7.11030**	**0.87729**	**0.15778**

The chained Sato-Vartia index is listed as P_{SV}^t in Table 1. Each chain link is defined over the set of products that are available in both periods t and t−1. The logarithm of the chain link going from period t−1 to period t is defined as follows:

(62) $\ln P_{LSV}^t \equiv \Sigma_{n\in I(t-1)\cap I(t)}\, w_n^t \ln(p_n^t/p_n^{t-1})$; $\qquad$ t = 2,3,...,T.

The weights w_n^t that appear in equations (62) are calculated in two stages. The first stage weight for product n in period t is defined as $w_n^{t*} \equiv (s_n^t - s_n^{t-1})/(\ln s_n^t - \ln s_n^t)$ for $n\in I(t-1)\cap I(t)$ and t = 2,...,T provided that $s_n^t \neq s_n^{t-1}$. If $s_n^t = s_n^{t-1}$, then define $w_n^{t*} \equiv s_n^t = s_n^{t-1}$. The second stage weights are defined as $w_n^t \equiv w_n^{t*}/\Sigma_{i\in I(t-1)\cap I(t)}\, w_i^{t*}$ for $n\in I(t-1)\cap I(t)$ and t = 2,...,T. These chain links P_{LSV}^t are cumulated into the chained Sato-Vartia price index $P_{SV}^t \equiv P_{SV}^{t-1}\times P_{LSV}^t$ for t = 2,3,...,39 that is listed in Table 1. This index ends up at the level 1.04607 in month 39 which is well above the corresponding econometrically determined P_{CES}^{39} level equal to 0.91069.

The Sato-Vartia indexes, P_{SV}^t, are equal to Feenstra's Index 1 in his decomposition of the CES price index using index numbers. We can also compute his Index 2 and Index 3 terms once we use his preferred estimator for the parameter r , $r^* = -4.9900$, which we now have. Using this estimated r^*, his Index 2 and Index 3 for month t in the present context when we are computing chained indexes are defined as follows:

(63) $\text{Index}_2^t \equiv [\sum_{i\in I(t)} p_i^t q_i^t/\sum_{i\in I(t-1)\cap I(t)} p_i^t q_i^t]^{1/r^*}$;
(64) $\text{Index}_3^t \equiv [\sum_{n\in I(t-1)\cap I(t)} p_n^{t-1} q_n^{t-1}/\sum_{n\in I(t-1)} p_n^{t-1} q_n^{t-1}]^{1/r^*}$.

The above indexes will be equal to 1 if the available products remain the same going from period t−1 to period t. There are 5 periods where the number of available products changes from the previous period: months 9, 10, 11, 20 and 23. Index_2^t will be less than unity for months 9 (products 2 and 4 become available), 11 (product 12 becomes available), and 23 (product 12 again becomes available). Index_3^t will be greater than unity for months 10 (product 12 becomes unavailable) and 20 (product 12 again becomes unavailable). Using $r^* = -4.9900$ and the data tabled in the Appendix, we can calculate Index_2^t and Index_3^t for these 5 months. The results are listed in Table 2.

Table 2: Indexes Measuring the Effects of Changes in the Price Level due to the Availability of Products when σ = 5.9900

Month t	Index_2^t	Index_3^t
9	**0.99073**	**1.00000**
10	**1.00000**	**1.00460**
11	**0.99448**	**1.00000**
20	**1.00000**	**1.00495**

23	**0.99603**	**1.00000**

In month 9, products 2 and 4 make their appearance and Table 2 tells us that the effect on the CES price level of this increase in variety is to lower the price level for month 9 by about 0.93 percentage points. In month 10 when product 12 disappears from the store, this disappearance has the effect of increasing the price level for frozen juice by 0.46 percentage points. The overall effect on the price level of the changes in the availability of products is equal to 0.99073×1.00460×0.99448×1.00495×0.99603 = 0.99075, a decrease in the price level over the sample period of about 0.93 percentage points. This is a noticeable reduction in the price level.

The indexes listed in Table 2 are chain links. For the 5 months when one of the two indexes is not equal to 1, these links can be multiplied with the corresponding Sato-Vartia chain link in order to obtain the overall Feenstra chain link index. The Feenstra chain links can be cumulated and the resulting indexes are the P_{FEEN}^t that are listed in Table 1 above. Note that P_{FEEN}^{39} ends up at 1.03639 which is lower than the corresponding Sato-Vartia index, P_{SV}^{39} = 1.04607. Recall that the cumulative effects of changes in the availability of products was 0.99075. This factor times P_{SV}^{39} is equal to P_{FEEN}^{39}.

With very small errors in the estimating equations, the Feenstra index for month t, P_{FEEN}^t, should equal the econometrically based CES index, P_{CES}^t. But it can be seen that in general, the Feenstra indexes are well above the corresponding CES indexes.[41] However, exact equality between the two indexes will hold only if all purchasers have common CES preferences and minimize the cost of their purchases in each month. If these assumptions were satisfied, all of the errors in the econometric model that we estimated would be 0 but of course, this is far from being the case. In fact, the CES unit cost function model does not fit the data very well and so P_{FEEN}^t is not equal to P_{CES}^t.

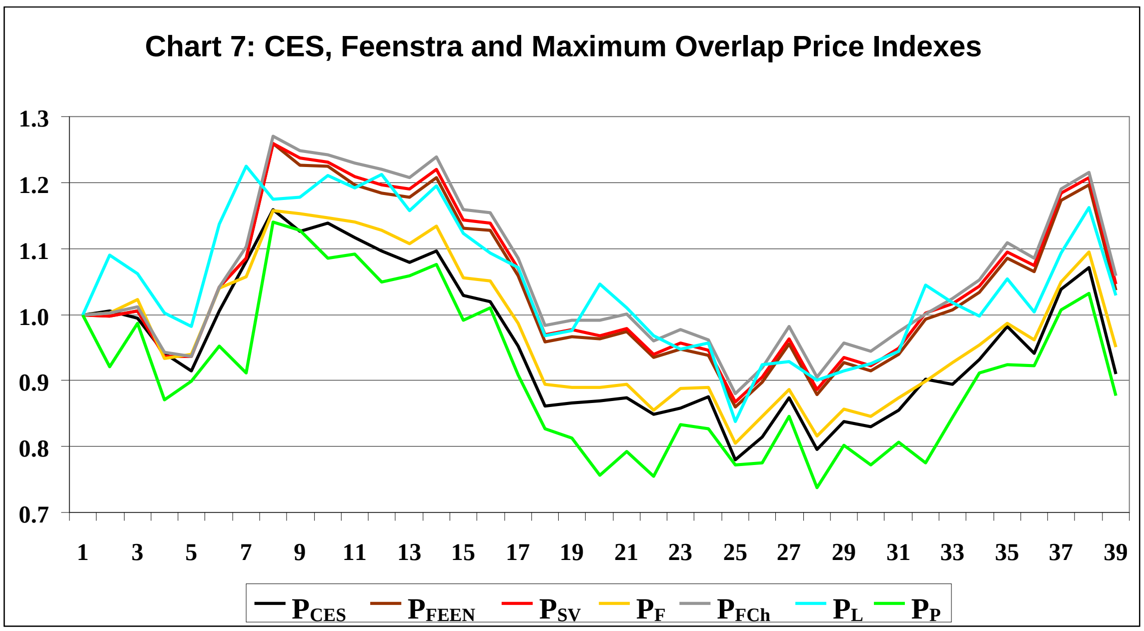
It is of some interest to compare P_{CES}^t and P_{FEEN}^t to traditional fixed base and chained Laspeyres, Paasche and Fisher price indexes. It should be noted that these indexes cannot take into account the effects of changes in the availability of products. The chain links for these indexes are calculated for each period t using the usual formulae but restricting the scope of the index to products that are available in periods t−1 and t. These *maximum overlap chain links* are then cumulated into the Chained Laspeyres, Paasche and Fisher indexes P_{LCh}^t, P_{PCh}^t and P_{FCh}^t that are listed in Table 1 above.

Calculating traditional fixed base indexes is a tricky business when the base period does not include all products, which is the case with our data. Thus for months 1 to 9, we calculated fixed base Laspeyres, Paasche and Fisher indexes, excluding products 2 and 4, which were not available in months 1 to 8. In month 9, all products were available. In the subsequent months, all products were available except for months 10 and 20-22. Excluding these 4 months (and months 1 to 9), we calculated fixed base Laspeyres, Paasche and Fisher indexes relative to month 9 and then linked the resulting indexes (at month 9) to their fixed base counterparts that were constructed for months 1 to 9. We are missing indexes for months 9 and 20-22. For month 10, we used the Laspeyres, Paasche

[41] This indicates that the maximum overlap Sato Vartia indexes suffer from an upward chain drift problem.

and Fisher indexes going from month 9 to 10, excluding product 12 (which is missing for month 10) and used these links to our earlier index levels established for month 9. For months 20-22, we calculated fixed base Laspeyres, Paasche and Fisher indexes over the 4 months 19-22 excluding product 12 and then linked these indexes for months 20-22 to their earlier counterpart index levels for month 19. The resulting sequence of indexes, P_L^t, P_P^t and P_F^t are listed in Table 1 above.

Looking at Table 1, it can be seen that the chained Laspeyres and chained Paasche indexes are complete disasters. P_{LCh}^t ended up at 7.11030 for month 39 (too high) and P_{PCh}^t ended up at 0.15778 (too low). Their fixed base counterparts, P_L^t and P_P^t, ended up at 1.02873 and 0.87729. This is a fairly substantial gap and indicates that these indexes are subject to substitution bias. The chained Fisher index P_{FCh}^t ended up at 1.05918 and its fixed base counterpart P_F^t ended up at 0.94999. This index is somewhat comparable to P_{CES}^t which ended up at 0.91069.[42] The series in Table 1 are plotted in Chart 7 except for the chained Laspeyres and Paasche indexes which are more or less off the Chart!



The fixed base Paasche indexes are on the lowest curve on Chart 7. P_{CES}^t is slightly below the fixed base maximum overlap (fixed base) Fisher index P_F^t and they are the second and third curve from the bottom. The highest curves are P_{FCh}^t, the *chained* maximum overlap Fisher index, P_{SV}^t, the Sato-Vartia maximum overlap *chained* index followed by the *chained* Feenstra indexes, P_{FEEN}^t. It seems very likely that these indexes suffer from a fairly substantial *upward chain drift*. The wide gap between the fixed base maximum overlap Laspeyres and Paasche indexes indicates that these two indexes suffer from

[42] Remember that the fixed base Fisher index does not allow for the fact that increased availability of products over the sample period reduced the final price level by almost 2 percentage points. Thus if we reduce P_F^{39} by 1 percentage point, the maximum overlap fixed base Fisher index ends up at approximately 0.94 which is getting closer to P_{CES}^{39} = 0.91.

substantial substitution bias. The most reasonable indexes are the econometrically determined CES price index, P_{CES}^t, and the fixed base maximum overlap Fisher index, P_F^t.

The fact that the chained Fisher index ended up higher than its fixed base counterpart is a priori surprising; this fact indicates *upward chain drift* when we would expect downward chain drift. However, Feenstra and Shapiro (2003; 125) also found upward chain drift using chained Törnqvist price indexes on weekly ACNielson scanner data.[43] It is somewhat surprising that this upward chain drift that was found using weekly unit value data persists when monthly unit value data are used.[44]

We turn now to the estimation of CES preferences using the system of share equations that correspond to the maximization of a primal CES utility function.

6. The Estimation of CES Preferences Using the Utility Function Approach

Recall that the system of estimating equations for purchasers maximizing a CES utility function turned out to be equations (49). These share equations hold for products that are present in each period. In order to use the system command for estimating a system of nonlinear regression equations in Shazam, we will assume that these equations hold for all products.[45] Recall that we assumed that s satisfied the inequalities $0 < s \leq 1$. To start off our estimation procedure, we will estimate a preliminary *Model 3* where we assume $s = 1$. Thus we obtain the following system of estimating equations:

(65) $s_i^t = [\beta_i q_i^t/\sum_{n=1}^{19} \beta_n q_n^t] + \varepsilon_i^t$ $\qquad t = 1,...,39;\ i = 1,...,19$

where the error term vectors, ε^t, are assumed to be distributed as a multivariate normal random variable with mean vector 0_{19} and variance-covariance matrix Σ for $t = 1,...,39$.[46] Note that the CES utility function collapses down to a linear utility function. Thus all products are perfect substitutes in this model. In order to identify the β_n, we impose the following normalization:

(66) $\beta_{19} = 1$.

Since the shares s_i^t sum to one for each period t, all 19 error terms ε_i^t for $i = 1,...,19$ cannot be distributed independently so we dropped the equation for product 19 from our list of estimating equations for Model 3.

---

[43] For our data set, the maximum overlap chained Törnqvist indexes were fairly close to our chained Fisher indexes. The maximum overlap chained Törnqvist index ended up 1.5% higher than P_{FCh}^{39}.
[44] Feenstra and Shapiro (2003; 125) suggested the following cure for the chain drift problem: "The only theoretically correct index to use in this type of situation is a fixed base index, as demonstrated in section 5.3." However, this proposed solution does not treat all periods in a symmetric manner and it does not deal with the problem of entering and exiting products.
[45] When a product is missing, we will assume that the resulting quantity and share is equal to 0. If $s > 0$, this means that equations (65) will hold for $t = 1,..., 39$ and $n = 1,...,19$.
[46] This is a slightly incorrect econometric specification since ε_n^t will automatically equal 0 if product n is not present during month t.

As usual, we used the nonlinear regression software package in Shazam to estimate the unknown β_i in equations (65). The final log likelihood turned out to be 3074.316, which is a huge increase from the final log likelihoods for Models 1 and 2; recall that the dependent variables are the same in all 3 Models and thus the log likelihoods are comparable. The equation by equation R^2 values for Model 3 were as follows: 0.9676, 0.9809, 0.9666, 0.9779, 0.9581, 0.9494, 0.9724, 0.7750, 0.9648, 0.9762, 0.8291, 0.9168, 0.9846, 0.9292, 0.9653, 0.9559, 0.9065 and 0.9554. These R^2 are very much higher than the corresponding R^2 for Models 1 and 2.

The estimated β_n^* coefficients were as follows: 1.098, 1.101, 1.241, 1.248, 1.247, 1.939, 1.277, 0.825, 0.513, 0.520, 0.784, 1.017, 1.028, 1.416, 0.439, 1.099, 1.699 and 1.094. Of course, $\beta_{19} \equiv 1$. These coefficients reflect the marginal utility and hence the quality of one unit of each product relative to product 19. Thus they could be regarded as *quality adjustment parameters* for the first 18 products.

We used the above β_n^* coefficients as starting values (along with s = 1) in *Model 4* which is the following nonlinear regression model:

$$(67)\ s_i^t = [\beta_i (q_i^t)^s / \textstyle\sum_{n\in I(t)} \beta_n (q_n^t)^s] + \varepsilon_i^t \qquad t = 1,...,39;\ i = 1,...,18.$$

We again make the normalization (66). Using the nonlinear systems estimation package in Shazam, we estimated the above model.[47] The final log likelihood turned out to be 3239.160, which is an increase of 164.844 from the final log likelihood for Model 3. The estimated s was $s^* = 0.85374$ (standard error = 0.0065) and so the estimated elasticity substitution is $\sigma = 1/(1-s^*) = 6.8371$. This is very much larger than the estimated σ for Model 2. The estimated β_n^* coefficients were as follows: 1.095, 0.998, 1.136, 1.139, 1.135, 1.530, 1.168, 0.759, 0.555, 0.547, 0.801, 0.892, 1.093, 1.287, 0.465, 1.185, 1.304 and 1.049. These β_n^* estimates (and $\beta_{19} \equiv 1$) are proportional to the marginal rates of substitution between the 19 products when purchasers with CES preferences purchase one unit of each product.[48] The equation by equation R^2 values for Model 4 were as follows: 0.9748, 0.9873, 0.9716, 0.9904, 0.9637, 0.9600, 0.9766, 0.7746, 0.9678, 0.9792, 0.8057, 0.9387, 0.9863, 0.9207, 0.9821, 0.9527, 0.8996 and 0.9583. The average R^2 is equal to 0.9439. These R^2 are quite high considering that the dependent variables are

---

[47] The 0 values for some q_i^t do not cause problems if the parameter s is positive during the iterations for the nonlinear systems regression. However, temporary negative estimates for s can occur. In order to eliminate problems with the nonlinear estimation software, we set $q_i^t = 1$ for the 20 observations where there were missing products. Thus for t = 1,...,8, the 0 quantity terms $(q_2^t)^s$ and $(q_4^t)^r$ on the right hand side of equations (58) were replaced by $\delta^t(1)^s$ and $\delta^t(1^t)^s$ where δ^t is a dummy variable which takes on the value 0 for t = 1,...,8 and is equal to 1 for t = 9,10,...,39. Similarly, the 0 quantity terms $(q_{12}^t)^s$ for t = 10 and t = 20, 21 and 22 were replaced by $\delta_{12}^t(1)^s$ where δ_{12}^t is a dummy variable which takes on the value 0 for t = 10, 20, 21, 22 and is equal to 1 for other periods t.

[48] Let β be the vector of the 19 β_n. Evaluate the CES utility function when one unit of each product is purchased, which give rise to the utility level $f(1_{19})$ where 1_{19} is a vector of ones of dimension 19. Then the vector β is proportional to the vector of first order partial derivatives of the utility function, $\nabla_q f(1_{19})$. Thus the β_n reflect the relative marginal utilities of the 19 products when one unit of each product is purchased. The highest quality products appear to be Products 6, 17 and 14 while the lowest quality products are 15, 10 and 9.

shares. Thus the CES directly estimated utility function model fits the data much better than the CES unit cost function model.

Note that the dependent variables for the CES cost function model, Model 2, and for the CES direct utility function model, Model 4, are exactly the same. If the price and quantity data were exactly consistent with CES preferences for purchasers and utility maximizing behavior, then the results for Models 2 and 4 should be consistent; i.e., the same estimate for the elasticity of substitution would be obtained and the fits would be perfect for both models.[49] Of course, the data are not exactly consistent with purchasers maximization of a CES utility function so we are faced with the problem of choosing between Models 2 and 4. From a purely economic perspective, prices are exogenous (neglecting possible time aggregation error) and purchases (quantities) are endogenous and so perhaps we should favour estimating Model 2. However, from the perspective of choosing a model which best describes the data, it is clear that estimating Model 4 is preferable. The issue of model choice is important since as we have seen, the Model 4 estimate of the elasticity of substitution is much larger than the Model 2 estimate (although not much larger than the $\sigma = 5.9900$ estimate that was obtained using the Feenstra econometric specification) and hence as will be shown below, the gains and losses from changes in product availability will be smaller using the estimated σ implied by Model 4.[50]

Define the CES utility or quantity level for month t, Q_{CES}^{t*}, as follows using our Model 4 estimated coefficients:

$$(68)\ Q_{CES}^{t*} \equiv [\textstyle\sum_{n\in I(t)} \beta_n^{*} (q_n^{t})^{s*}]^{1/s*} ; \qquad t = 1,...,39.$$

The normalized CES quantity index for month t based on the estimation of a CES utility function is defined as $Q_{CES}^{t} \equiv Q_{CES}^{t*}/Q_{CES}^{1*}$ for t = 1,...,39. This econometrically based quantity index for frozen juice is listed in Table 4 below. It ends up at the level 1.39043 in month 39.

The chained Sato-Vartia quantity index defined over products that are present in successive periods is listed as Q_{SV}^{t} in Table 4. The period t link index, Q_{LSV}^{t}, is defined over only the products that are present in periods t and t−1. The logarithm of the chain link going from period t−1 to period t is defined as follows:

$$(69)\ \ln Q_{LSV}^{t} \equiv \textstyle\sum_{n\in I(t-1)\cap I(t)} w_n^{t} \ln(q_n^{t}/q_n^{t-1}) ; \qquad t = 2,3,...,T.$$

[49] We checked that our software was correct by taking our estimated coefficients for Model 4 and using them to define a CES direct utility function. We then calculated shadow prices period t imputed equilibrium prices for product n as $p_n^{t*} \equiv \beta_n(q_i^{t})^{s-1}$ for $n\in I(t)$ and t = 1,...,39. If $n\notin I(t)$, then $p_n^{t*} \equiv +\infty$. Using these imputed prices and actual quantities with $q_n^{t} \equiv 0$ for $n\notin I(t)$, we used the Model 2 software with this new data set and estimated the elasticity of substitution for this artificial data set. We obtained exactly the same estimate for σ as we obtained for Model 4 and the R^2 was 1 for all 18 estimating equations.

[50] From a numerical perspective, it is possible to explain why Model 4 fits the data better than Model 2. Product shares of sales are normalizations of product sales. Sales are equal to price times quantity. Quantities vary roughly 10 times as much as the variations in prices. Thus sales (and hence shares) of a product will be more highly correlated to quantities than to prices and so the Model 4 fits will be much higher than the Model 2 fits.

The weights w_n^t that appear in equations (69) are exactly the same as the weights that appeared in equations (62). These chain links Q_{LSV}^t are cumulated into the chained Sato-Vartia quantity index $Q_{SV}^t \equiv Q_{SV}^{t-1} \times Q_{LSV}^t$ for t = 2,3,...,39 that is listed in Table 4. This index ends up at the level 1.14374 in month 39, which is well below the corresponding CES level, which was 1.39043.[51]

The Sato-Vartia quantity indexes, Q_{SV}^t, are equal to Feenstra's Index 1^* in our adaptation of his decomposition of the CES price index to the utility function context; see (35) above. We can also compute the utility function counterparts to our Index 2^* and Index 3^* terms once we have an estimator for the parameter s which we now have. These utility function counterparts to Index 2^* and Index 3^* for month t in the present context when we are computing chained quantity indexes are defined as follows:

(70) $\text{Index}_2^{t*} \equiv [\sum_{i \in I(t)} p_i^t q_i^t / \sum_{i \in I(1) \cap I(t)} p_i^t q_i^t]^{1/s*}$;
(71) $\text{Index}_3^{t*} \equiv [\sum_{n \in I(t-1) \cap I(t)} p_n^{t-1} q_n^{t-1} / \sum_{n \in I(t-1)} p_n^{t-1} q_n^{t-1}]^{1/s*}$.

The above indexes will be equal to 1 if the available products remain the same going from period t−1 to period t. There are 5 periods where the number of available products changes from the previous period: months 9, 10, 11, 20 and 23. Index_2^{t*} will be greater than unity for months 9 (products 2 and 4 become available), 11 (product 12 becomes available), and 23 (product 12 again becomes available). Index_3^{t*} will be less than unity for months 10 (product 12 becomes unavailable) and 20 (product 12 again becomes unavailable). Using our estimated s = 0.85374 and the data tabled in the Appendix, we can calculate Index_2^{t*} and Index_3^{t*} for these 5 months. The results are listed in Table 3.

Table 3: Indexes Measuring the Effects of Changes on Utility due to the Changing Availability of Products

Month t	Index_2^{t*}	Index_3^{t*}
9	**1.0559**	**1.0000**
10	**1.0000**	**0.9735**
11	**1.0329**	**1.0000**
20	**1.0000**	**0.9716**
23	**1.0235**	**1.0000**

The overall effect on the utility level of the changes in the availability of products over the sample period is equal to 1.0559×0.9735×1.0329×0.9716×1.0235 = 1.0559, an increase in the utility level over the sample period of about 5.6%.[52]

[51] This difference is due to three factors: (i) the CES index is essentially a fixed base index whereas the chained Sato-Vartia index almost certainly suffers from some downward chain drift; (ii) there are error terms in the econometric model which indicates that the assumptions required for the exactness of the Sato-Vartia index are not precisely satisfied and (iii) the theoretical exact equality is Q_{CES}^t equal to Q_{SV}^t times the cumulated effects of Index_2^{t*} times Index_3^{t*} (which is Index^{t*}).

[52] The changes in utility that are implied by Index_2^{t*} and Index_3^{t*} are no longer just changes in the true cost of living due to the change in the availability of products but they also incorporate changes in expenditure due to the changing availability of products.

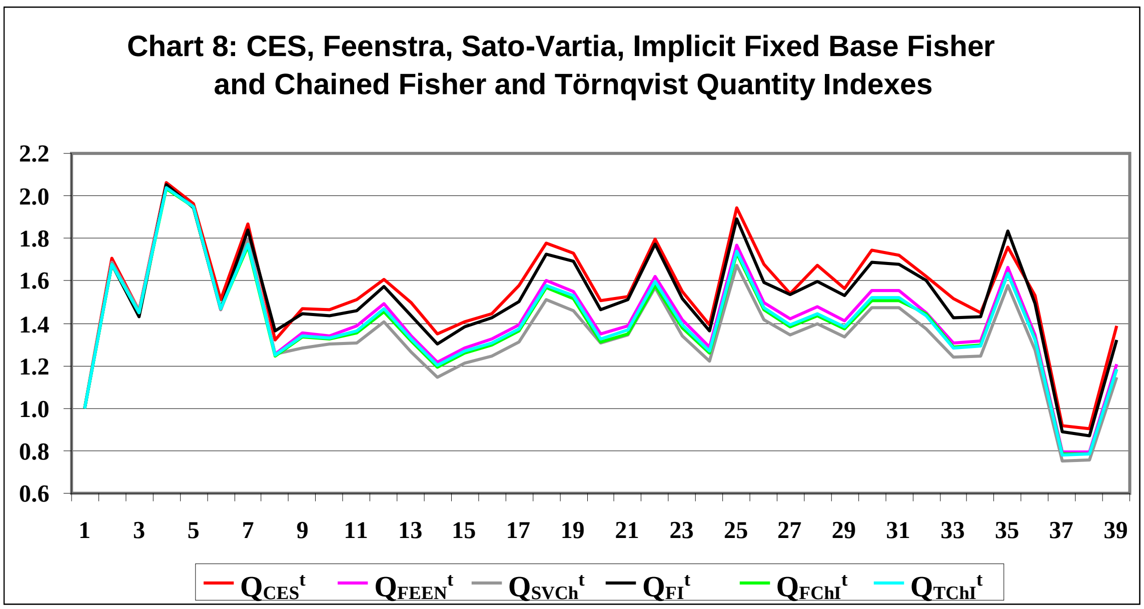
The indexes listed in Table 3 are chain links. These indexes are multiplied together and cumulated into an overall Index^{t*} which is listed in Table 4 below. For the 5 months when one of the two indexes is not equal to 1, the Index_2^{t*} and Index_3^{t*} links can be multiplied with the corresponding Sato-Vartia quantity chain link in order to obtain the overall Feenstra chain link quantity index. The Sato-Vartia quantity chain link indexes that are defined over products that are present in periods t−1 and t are cumulated together to form the chained Sato-Vartia quantity index, Q_{SVCh}^t, and this index is also listed in Table 4. The Feenstra chain links can be cumulated and the resulting indexes are the Q_{FEEN}^t that are listed in Table 4 below. Note that Q_{FEEN}^{39} ends up at 1.20762 which is substantially below the econometric index Q_{CES}^{39} = 1.39043. However, Q_{FEEN}^{39} does end up well above the corresponding Sato-Vartia chained quantity index, Q_{SVCh}^{39} = 1.14374. Recall that the cumulative effects of changes in the availability of products was 1.0559 = Index^{39*}. This factor times Q_{SVCh}^{39} is equal to Q_{FEEN}^{39}.

Table 4: The Econometric CES Quantity Index Q_{CES}^t, the Feenstra Methodology Chained Quantity Index Q_{FEEN}^t, the Maximum Overlap Chained Sato-Vartia Quantity Index Q_{SVCh}^t, the Implicit Fixed Base Fisher Quantity Index Q_{FI}^t, the Maximum Overlap Chained Implicit Fisher and Törnqvist Quantity Indexes, Q_{FChI}^t and Q_{TChI}^t and Index^{t*}

t	Q_{CES}^t	Q_{FEEN}^t	Q_{SVCh}^t	Q_{FI}^t	Q_{FChI}^t	Q_{TChI}^t	Index^{t*}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.70822	1.68506	1.68506	1.67654	1.67654	1.67959	1.00000
3	1.45934	1.45737	1.45737	1.43119	1.44843	1.45157	1.00000
4	2.06216	2.04785	2.04785	2.05425	2.03514	2.03919	1.00000
5	1.96224	1.94849	1.94849	1.94364	1.94881	1.94802	1.00000
6	1.50967	1.46626	1.46626	1.46955	1.46834	1.46708	1.00000
7	1.86844	1.79165	1.79165	1.83988	1.76398	1.77489	1.00000
8	1.32019	1.25591	1.25591	1.36598	1.24499	1.25137	1.00000
9	1.46936	1.35322	1.28155	1.44452	1.33374	1.34011	1.05593
10	1.46559	1.34105	1.30456	1.43629	1.32615	1.33234	1.02797
11	1.51089	1.38673	1.30604	1.46004	1.35480	1.36467	1.06178
12	1.60650	1.49331	1.40642	1.57386	1.45456	1.46740	1.06178
13	1.49921	1.34230	1.26419	1.43404	1.31487	1.32408	1.06178
14	1.35077	1.21896	1.14803	1.30245	1.19262	1.20208	1.06178
15	1.40904	1.28550	1.21070	1.38278	1.25881	1.26769	1.06178
16	1.44601	1.32412	1.24707	1.42560	1.29867	1.30661	1.06178
17	1.57935	1.39406	1.31294	1.50114	1.36386	1.37377	1.06178
18	1.77812	1.60334	1.51004	1.72526	1.56928	1.58030	1.06178
19	1.72916	1.55007	1.45987	1.68995	1.51792	1.52888	1.06178
20	1.50465	1.34887	1.30758	1.46230	1.31345	1.32761	1.03158
21	1.52512	1.38649	1.34405	1.51266	1.35229	1.36703	1.03158
22	1.79491	1.61959	1.57001	1.77473	1.58028	1.59666	1.03158
23	1.54974	1.41773	1.34274	1.51642	1.37851	1.39228	1.05585
24	1.39064	1.28832	1.22018	1.36238	1.25971	1.26775	1.05585
25	1.94418	1.76604	1.67262	1.89281	1.72944	1.73931	1.05585
26	1.67637	1.49690	1.41773	1.59158	1.46532	1.47133	1.05585
27	1.53951	1.41916	1.34410	1.53355	1.38467	1.39196	1.05585

28	**1.67520**	**1.47625**	**1.39816**	**1.59504**	**1.43565**	**1.44583**	**1.05585**
29	**1.56533**	**1.41286**	**1.33812**	**1.53249**	**1.37286**	**1.38351**	**1.05585**
30	**1.74361**	**1.55386**	**1.47167**	**1.68660**	**1.50914**	**1.52125**	**1.05585**
31	**1.72272**	**1.55555**	**1.47327**	**1.67916**	**1.50583**	**1.51999**	**1.05585**
32	**1.62124**	**1.44942**	**1.37276**	**1.60409**	**1.44302**	**1.43754**	**1.05585**
33	**1.51424**	**1.30870**	**1.23947**	**1.42693**	**1.29072**	**1.28548**	**1.05585**
34	**1.45108**	**1.31558**	**1.24599**	**1.43000**	**1.29625**	**1.29322**	**1.05585**
35	**1.76007**	**1.66510**	**1.57703**	**1.83588**	**1.63494**	**1.63344**	**1.05585**
36	**1.52947**	**1.34723**	**1.27596**	**1.49490**	**1.32559**	**1.32217**	**1.05585**
37	**0.91877**	**0.79504**	**0.75299**	**0.89098**	**0.78581**	**0.78270**	**1.05585**
38	**0.90334**	**0.79668**	**0.75454**	**0.87230**	**0.78609**	**0.78361**	**1.05585**
39	**1.39043**	**1.20762**	**1.14374**	**1.32139**	**1.18517**	**1.18376**	**1.05585**

We also compare Q_{CES}^t, Q_{SV}^t and Q_{FEEN}^t to the fixed base implicit Fisher quantity index[53] Q_{FI}^t and the maximum overlap chained Implicit Fisher and Törnqvist quantity indexes, Q_{FChI}^t and Q_{TChI}^t, which are also listed in Table 4. These 6 quantity indexes are plotted on Chart 8.



It can be seen that the CES quantity index that is based on the direct estimation of the CES utility function, Q_{CES}^t, is the (mostly) highest curve on Chart 8. The second highest line, the fixed base implicit Fisher quantity index Q_{FI}^t, provides the closest approximation to Q_{CES}^t. The 4 remaining quantity indexes are chained indexes and they all suffer from

[53] This index uses the maximum overlap Fisher price index P_F^t as the deflator for total expenditures. The implicit Fisher and Törnqvist chained quantity indexes are constructed by deflating total period t expenditures e^t by the maximum overlap Fisher and Törnqvist chained maximum overlap prices indexes, P_{FCh}^t and P_{TCh}^t respectively. All three of these implicit quantity indexes are normalized to equal 1 in month 1. The maximum overlap chained Fisher and Törnqvist price indexes approximate each other closely and thus Q_{FChI}^t closely approximates Q_{TChI}^t.

some downward chain drift. The Feenstra quantity index, Q_{FEEN}^t, is the closest to Q_{CES}^t, followed by the chained Sato-Vartia quantity index, Q_{SV}^t and Q_{FCh}^t. Finally, the chained Fisher and Törnqvist implicit quantity indexes, Q_{FCh}^t and Q_{TCh}^t, are the lowest curves on Chart 8 and the furthest from Q_{CES}^t.[54]

We will now define the implicit price indexes that correspond to Q_{CES}^t, Q_{SV}^t and Q_{FEEN}^t. The month t (unnormalized) price index that corresponds to Q_{CES}^t, P_{UCES}^{t*}, is defined as period t expenditure e^t divided by Q_{CES}^t:

$$(72)\ P_{UCES}^{t*} \equiv e^t/Q_{CES}^t; \qquad t = 1,...,T.$$

The normalized CES implicit price index for month t is defined as $P_{UCES}^t \equiv P_{UCES}^{t*}/P_{UCES}^{1*}$ for t = 1,...,T. P_{UCES}^t is listed in Table 5 below.

Forming the implicit price index that corresponds to the chained Sato-Vartia index Q_{SV}^t is more complicated. Denote the month t Sato-Vartia chain link index as Q_{LSV}^t for t = 2,3,...,T. Denote R^t as the month t adjusted expenditure ratio for t = 2,3,...,T. If the products purchased in months t−1 and t are the same, then R^t equals the expenditure ratio, e^t/e^{t-1}. However, if the number of products purchased in months t−1 and t is different, then $R^t \equiv [\Sigma_{n\in I(t-1)\cap I(t)}\ p_n^tq_n^t]/[\Sigma_{n\in I(t-1)\cap I(t)}\ p_n^{t-1}q_n^{t-1}]$. The month t Sato-Vartia chain link implicit price index P_{LUSV}^t is defined as R^t/Q_{LSV}^t for t = 2,3,...,T. These chain links are cumulated together to form the Sato-Vartia implicit chained price index, $P_{USVI}^t \equiv P_{USVI}^{t-1}\times P_{LUSV}^t$ for t = 2,...,T where $P_{USVI}^1 \equiv 1$. This index is listed in Table 5.

The month t (unnormalized) price index that corresponds to Q_{FEEN}^t, P_{UFEEN}^{t*}, is defined as period t expenditure e^t divided by Q_{FEEN}^t:

$$(73)\ P_{UFEEN}^{t*} \equiv e^t/Q_{FEEN}^t; \qquad t = 1,...,T.$$

The normalized Feenstra implicit price index for month t is defined as $P_{UFEEN}^t \equiv P_{UFEEN}^{t*}/P_{UFEEN}^{1*}$ for t = 1,...,T. P_{UFEEN}^t is also listed in Table 5.

Finally, Index^{t**} is an index which represents the cumulative effects on the cost of living of changes in the availability of products. This index for month t is the ratio of the month t Feenstra implicit chained price index, P_{UFEEN}^t, divided by the chained Sato-Vartia implicit price index, P_{USVI}^t:

$$(74)\ \text{Index}^{t**} \equiv P_{UFEEN}^t/\ P_{USVI}^t; \qquad t = 1,...,T.$$

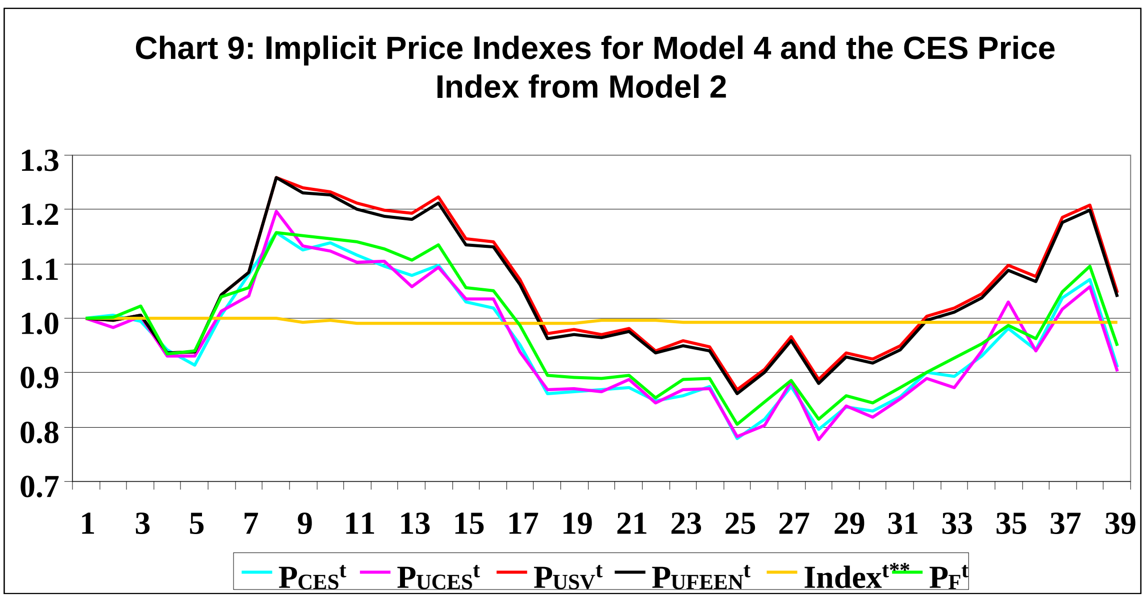
Index^{t**} is also listed in Table 5 along with P_{CES}^t (the CES unit cost function index that resulted from Model 2 above and is listed in Table 1) and the fixed base Fisher index P_F^t that is also listed in Table 5.

[54] Diewert (1978) showed that the Fisher and Törnqvist indexes approximate each other to the second order around an equal price and quantity point. However, the changes in prices and quantities in our data set were so large, it was uncertain whether the approximation $P_{FCh}^t \approx P_{TCh}^t$ would hold. Evidently it does hold.

Table 5: Unit Cost and Implicit CES Price Indexes Estimated from Models 2 and 4, P_{CES}^t and P_{UCES}^t, the Implicit Chained Sato-Vartia Price Index P_{USV}^t, the Implicit Chained Feenstra Price Index P_{UFEEN}^t, $Index^{t**}$ and the Fixed Base Fisher Index P_F^t

t	P_{CES}^t	P_{UCES}^t	P_{USVI}^t	P_{UFEEN}^t	$Index^{t**}$	P_F^t
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.00508	0.98359	0.99711	0.99711	1.00000	1.00218
3	0.99521	1.00369	1.00504	1.00504	1.00000	1.02342
4	0.94121	0.93029	0.93679	0.93679	1.00000	0.93388
5	0.91406	0.93074	0.93730	0.93730	1.00000	0.93964
6	1.00557	1.01226	1.04223	1.04223	1.00000	1.03989
7	1.08119	1.04046	1.08505	1.08505	1.00000	1.05662
8	1.15826	1.19753	1.25882	1.25882	1.00000	1.15739
9	1.12605	1.13261	1.23964	1.22981	0.99207	1.15209
10	1.13885	1.12326	1.23253	1.22757	0.99597	1.14617
11	1.11631	1.10248	1.21177	1.20119	0.99127	1.14088
12	1.09576	1.10469	1.19889	1.18842	0.99127	1.12760
13	1.07865	1.05886	1.19305	1.18264	0.99127	1.10698
14	1.09705	1.09361	1.22255	1.21187	0.99127	1.13419
15	1.02962	1.03612	1.14569	1.13569	0.99127	1.05579
16	1.01928	1.03616	1.14150	1.13154	0.99127	1.05099
17	0.95211	0.93755	1.07152	1.06217	0.99127	0.98640
18	0.86187	0.86830	0.97143	0.96295	0.99127	0.89490
19	0.86582	0.87013	0.97921	0.97066	0.99127	0.89032
20	0.86905	0.86511	0.96942	0.96502	0.99546	0.89016
21	0.87341	0.88722	0.98038	0.97593	0.99546	0.89453
22	0.84816	0.84505	0.94080	0.93653	0.99546	0.85466
23	0.85830	0.86932	0.95785	0.95027	0.99208	0.88842
24	0.87495	0.87123	0.94792	0.94042	0.99208	0.88930
25	0.77932	0.78295	0.86881	0.86193	0.99208	0.80421
26	0.81396	0.80363	0.90717	0.89998	0.99208	0.84644
27	0.87388	0.88298	0.96550	0.95785	0.99208	0.88641
28	0.79552	0.77627	0.88792	0.88089	0.99208	0.81528
29	0.83771	0.83908	0.93705	0.92963	0.99208	0.85705
30	0.83003	0.81745	0.92460	0.91728	0.99208	0.84508
31	0.85502	0.85125	0.95025	0.94273	0.99208	0.87333
32	0.90137	0.89022	1.00369	0.99574	0.99208	0.89973
33	0.89378	0.87329	1.01851	1.01045	0.99208	0.92673
34	0.93133	0.93999	1.04508	1.03681	0.99208	0.95385
35	0.98200	1.02941	1.09681	1.08813	0.99208	0.98690
36	0.94110	0.94062	1.07638	1.06786	0.99208	0.96237
37	1.03787	1.01774	1.18552	1.17613	0.99208	1.04948
38	1.07081	1.05781	1.20900	1.19943	0.99208	1.09545
39	0.91069	0.90282	1.04779	1.03950	0.99208	0.94999

The price indexes listed in Table 5 are plotted in Chart 9. It can be seen that the CES price index that was estimated using Model 2, P_{CES}^t, is fairly close to the implicit CES price index, P_{UCES}^t, that was estimated using the parameter estimates from Model 4. Recall that the elasticity of substitution σ that was estimated using Model 2 was 3.8 using the first econometric specification and was 6.0 using the Feenstra econometric specification while the σ that was estimated using Model 4 was somewhat larger at 6.8. These differences in σ did not make much difference to the overall true cost of living that was estimated by the various econometric models. However, *with a higher σ, the gains and losses from new and disappearing products are smaller*; i.e., Index^{t**} ends up at 0.99208, which indicates the overall reduction in the true cost of living at the end of month 39 due to changes product availability was 0.79 percentage points using the Model 4 parameter estimates whereas the decrease in the cost of living using the initial and final Model 2 estimates for σ was 1.7 and 0.93 percentage points respectively. The larger σ from Model 4 causes Index^{t**} to be closer to 1 and this causes the implicit chained Sato-Vartia price index $P_{USVI}{}^t$ to be very close to the implicit chained Feenstra price index $P_{UFEEN}{}^t$. If the Model 4 equations fit the data perfectly, then $P_{UFEEN}{}^t$ would be equal to $P_{UCES}{}^t$ for each month t. However, the fit is not exact and $P_{UFEEN}{}^t$ (and $P_{USVI}{}^t$) lie well above $P_{UCES}{}^t$ as t increases. The fixed base Fisher price index $P_F{}^t$ discussed earlier provides a fairly close approximation to the econometrically estimated CES price indexes $P_{CES}{}^t$ and $P_{UCES}{}^t$.



The methodology that was explained in this section is rather involved. The benefit from using this methodology, which relied on estimating the CES utility function, is that we obtained an estimate for the elasticity of substitution which is more reliable[55] than the estimate that was obtained by estimating the CES unit cost function using our initial econometric specification. However, once we have a new estimate for the elasticity of substitution, we can use the methodology that was developed in section 5 in order to obtain new estimates for $\mathrm{Index}_2{}^t$ and $\mathrm{Index}_3{}^t$ defined by (63) and (64). The point estimate for s from Model 4 is $s^* \equiv 0.87374$ and thus the corresponding $r = s^*/(s^*-1) = -6.92016$

[55] The log likelihood that was obtained by estimating Model 4 was much larger than the log likelihood that was obtained estimating Model 2.

and the corresponding $\sigma = 1/(1-s^*) = 6.8371$. Using $r^* = -6.92016$, we use the formula (63) and (64) to evaluate $\mathrm{Index}_2{}^t$ and $\mathrm{Index}_3{}^t$ and we obtain the following counterpart to Table 2 in section 5.

Table 6: Indexes Measuring the Effects of Changes in the Price Level due to the Availability of Products when σ = 6.8371

Month t	$\mathrm{Index}_2{}^t$	$\mathrm{Index}_3{}^t$
9	**0.9933**	**1.0000**
10	**1.0000**	**1.0033**
11	**0.9960**	**1.0000**
20	**1.0000**	**1.0036**
23	**0.9971**	**1.0000**

The overall effect on the price level of the changes in the availability of products is equal to $0.9933\times1.0033\times0.9960\times1.0036\times0.9971 = 0.9933$, a decrease in the price level over the sample period of about 0.67%. This estimate for the gains from changes in product availability is similar to our direct estimate for the Model 4 benefit which was a decrease in the price level over the sample period of 0.79%.

In Appendix B, we will adapt Feenstra's double differencing methodology to obtain an alternative method for estimating the elasticity of substitution in the context of the direct estimation of the CES utility function.

Potential problems with the Feenstra methodology for measuring the gains from increased product availability are the following:

- The reservation prices which induce purchasers to demand 0 units of products that are not available in a period are infinite, which a priori seems implausible and
- The CES functional form is not fully flexible.

Thus in the following section, we will introduce a flexible functional form that will generate finite reservation prices for new and unavailable products and hence will provide an alternative methodology for measuring the benefits of new products (and the losses for disappearing products).

7. The Konüs-Byushgens-Fisher Utility Function

The functional form for a purchaser's utility function f(q) that we will introduce in this section is the following one:[56]

(75) $f(q) = (q^TAq)^{1/2}$

[56] We assume that vectors are column vectors when matrix algebra is used. Thus q^T denotes the row vector which is the transpose of q.

where the N by N matrix $A \equiv [a_{nk}]$ is symmetric (so that $A^T = A$) and thus has N(N+1)/2 unknown a_{nk} elements. We also assume that A has one positive eigenvalue with a corresponding strictly positive eigenvector and the remaining N−1 eigenvalues are negative or zero.[57] These conditions will ensure that the utility function has indifference curves with the correct curvature.

Konüs and Byushgens (1926) showed that the Fisher (1922) quantity index $Q_F(p^0,p^1,q^0,q^1) \equiv [p^0\cdot q^1 p^1\cdot q^1/p^0\cdot q^0 p^1\cdot q^0]^{1/2}$ is exactly equal to the aggregate utility ratio $f(q^1)/f(q^0)$ provided that all purchasers maximized the utility function defined by (75) in periods 0 and 1 where p^0 and p^1 are the price vectors prevailing during periods 0 and 1 and aggregate purchases in periods 0 and 1 are equal to q^0 and q^1. Diewert (1976) elaborated on this result by proving that the utility function defined by (75) was a *flexible functional form*; i.e., it can approximate an arbitrary twice continuously differentiable linearly homogeneous function to the accuracy of a second order Taylor series approximation around an arbitrary positive quantity vector q^*. Since the Fisher quantity index gives exactly the correct utility ratio for the functional form defined by (75), he labelled the Fisher quantity index as a *superlative index*.

Assume that all products are available in a period and purchasers face the positive prices $p \equiv (p_1,...,p_N) >> 0_N$. The first order necessary (and sufficient) conditions (provided that s $\leq$ 1) that can be used to solve the unit cost minimization problem defined by (2) when the utility function f is defined by (66) are the following conditions:

(76) $p = \lambda Aq/(q^TAq)^{1/2}$;
(77) $1 = (q^TAq)^{1/2}$.

Multiply both sides of equation n in (76) by q_n and sum the resulting N equations. This leads to the equation $p\cdot q = \lambda(q^TAq)^{1/2}$. Solve this equation for λ and use this solution to eliminate the λ in equations (76). The resulting equations (where equation n is multiplied by q_n) are the following *system of inverse demand share equations*:

(78) $s_n \equiv p_nq_n/p\cdot q = q_n \Sigma_{k=1}^N a_{nk}q_j/q^TAq$;   n = 1,...,N

where a_{nk} is the element of A that is in row n and column j for n, k = 1,...,N. These equations will form the basis for our system of estimating equations in subsequent sections. Note that they are nonlinear equations in the unknown parameters a_{nk}.

It turns out to be useful to reparameterize the A matrix in definition (75). Thus we set A equal to the following expression:

(79) $A = bb^T + B$; $b >> 0_N$; $B = B^T$; B is negative semidefinite; $Bq^* = 0_N$.

[57] Diewert and Hill (2010) show that these conditions are sufficient to imply that the utility function defined by (75) is positive, increasing, linearly homogeneous and concave over the regularity region $S \equiv \{q: q >> 0_N$ and $Aq >> 0_N\}$.

The vector $b^T \equiv [b_1,...,b_N]$ is a row vector of positive constants and so bb^T is a rank one positive semidefinite N by N matrix. The symmetric matrix B has N(N+1)/2 independent elements b_{nk} but the N constraints Bq^* reduce this number of independent parameters by N. Thus there are N independent parameters in the b vector and N(N−1)/2 independent parameters in the B matrix so that $bb^T + B$ has the same number of independent parameters as the A matrix. Diewert and Hill (2010) showed that replacing A by $bb^T + B$ still leads to a flexible functional form.

The reparameterization of A by $bb^T + B$ is useful in our present context because we can use this reparameterization to estimate the unknown parameters in stages. Thus we will initially set $B = O_{N\times N}$, a matrix of 0's. The resulting utility function becomes $f(q) = (q^Tbb^Tq)^{1/2} = (b^Tqb^Tq)^{1/2} = b^Tq$, a linear utility function. Thus this special case of (79) boils down to the *linear utility function* model that we have already estimated as Model 3 above.

The matrix B is required to be negative semidefinite. We can follow the procedure used by Wiley, Schmidt and Bramble (1973) and Diewert and Wales (1987) and impose negative semidefiniteness on B by setting B equal to $-CC^T$ where C is a lower triangular matrix.[58] Write C as $[c^1,c^2,...,c^N]$ where c^k is a column vector for k = 1,...,K. If C is lower triangular, then the first k−1 elements of c^k are equal to 0 for k = 2,3,...,N. Thus we have the following representation for B:

$$(80)\ B = -CC^T$$
$$= -\sum_{n=1}^{N} c^n c^{nT}$$

where we impose the following restrictions on the vectors c^n in order to impose the restrictions $Bq^* = 0_N$ on B:[59]

$$(81)\ c^n\cdot q^* = c^{nT}q^* = 0\ ; \qquad n = 1,....,N.$$

If the number of products N in the commodity group under consideration is not small, then typically, it will not be possible to estimate all of the parameters in the C matrix. Furthermore, typically nonlinear estimation is not successful if one attempts to estimate all of the parameters at once. Thus we will estimate the parameters in the utility function $f(q) = (q^TAq)^{1/2}$ in stages. In the first stage, we estimate the linear utility function $f(q) = b^Tq$. In the second stage, we estimate $f(q) = (q^T[bb^T - c^1c^{1T}]q)^{1/2}$ where $c^{1T} \equiv [c_1^1,c_2^1,...,c_N^1]$ and $c^{1T}q^* = 0$. For starting coefficient values in the second nonlinear regression, we use the final estimates for b from the first nonlinear regression and set the

[58] $C = [c_{nk}]$ is a lower triangular matrix if $c_{nk} = 0$ for k > n; i.e., there are 0's in the upper triangle. Wiley, Schmidt and Bramble showed that setting $B = -CC^T$ where C was lower triangular was sufficient to impose negative semidefiniteness while Diewert and Wales showed that any negative semidefinite matrix could be represented in this fashion.
[59] The restriction that C be upper triangular means that c^N will have at most one nonzero element, namely c_N^N. However, the positivity of q^* and the restriction $c^{NT}q^* = 0$ will imply that $c^N = 0_N$. Thus the maximal rank of B is N−1.

starting $c^1 \equiv 0_N$.[60] In the third stage, we estimate $f(q) = (q^T[bb^T - c^1c^{1T} - c^2c^{2T}]q)^{1/2}$ where $c^{1T} \equiv [c_1^1,c_2^1,...,c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0,c_2^1,...,c_N^1]$ and $c^{2T}q^* = 0$. The starting coefficient values are the final values from the second stage with $c^2 \equiv 0_N$. In the fourth stage, we estimate $f(q) = (q^T[bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T}]q)^{1/2}$ where $c^{1T} \equiv [c_1^1,c_2^1,...,c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0,c_2^1,...,c_N^1]$, $c^{2T}q^* = 0$, $c^{3T} \equiv [0,0,c_3^1,...,c_N^1]$ and $c^{3T}q^* = 0$. At each stage, the log likelihood will generally increase.[61] We stop adding columns to the C matrix when the increase in the log likelihood becomes small (or the number of degrees of freedom becomes small). At stage k of this procedure, it turns out that we are estimating the substitution matrices of rank k−1 that is the most negative semidefinite that the data will support. This is the same type of procedure that Diewert and Wales (1988) used in order to estimate normalized quadratic preferences and they termed the final functional form a *semiflexible functional form*. The above treatment of the KBF functional form also generates a semiflexible functional form.

Instead of developing the above theory for the KBF utility function, we could develop the analogous theory for the dual KBF unit cost function, $c(p) \equiv (p^TA^*p)^{1/2}$ where $A^* = b^*b^{*T} - C^*C^{*T}$ where C^* is a lower triangular N by N matrix that satisfies $C^{*T}p^* = 0_N$ for the reference price vector p^*. The special case of this unit cost function where $C^* = O_{N\times N}$ leads to the Leontief (no substitution) unit cost function, $c(p) = b^{*T}p$ which we estimated as Model 1 in section 5 above. However, this model did not fit the data very well at all, which is not surprising since it is unlikely that there would be zero substitutability between closely related products. Model 3, which assumed that the products were perfectly substitutable fit the data much better than Model 1. Hence we will not estimate the KBF unit cost function model in this study since it is unlikely to fit the data very well.[62] Furthermore, a major goal of our econometric efforts is to estimate reservation prices that will induce purchasers of the group of products under consideration that result when a product is not available. This can be done rather easily if we estimate the purchasers' utility function rather than their dual unit cost function.

8. The Systems Approach to the Estimation of KBF Preferences

Our system of nonlinear estimating equations for *Model 5* is the following stochastic version of equations (75) above where $A = bb^T - c^1c^{1T}$:

$$(82)\quad s_i^t = q_i^t\, \Sigma_{k=1}^{19}\, a_{ik}q_k^t/[\Sigma_{n=1}^{19}\Sigma_{m=1}^{19}\, a_{nm}q_n^tq_m^t] + \varepsilon_i^t \qquad t = 1,...,39;\ i = 1,...,19$$

where $b^T = [b_1,...,b_{19}]$, $c^{1T} = [c_1^1,...,c_{19}^1]$ and the error term vectors, $\varepsilon^{tT} = [\varepsilon_1^t,...,\varepsilon_{19}^t]$ are assumed to be distributed as a multivariate normal random variable with mean vector 0_{19}

[60] We also use the constraint $c^{1T}q^*$ to eliminate one of the c_n^1 from the nonlinear regression.
[61] If it does not increase, then the data do not support the estimation of a higher rank substitution matrix and we stop adding columns to the C matrix. The log likelihood cannot decrease since the successive models are nested.
[62] If the A matrix in (75) has full rank N, then it can be shown that the dual unit cost function is equal to $c(p) = (p^TA^{-1}p)^{1/2}$.

and variance-covariance matrix Σ for t = 1,...,39.[63] In order to identify the parameters, we impose the following normalization:

(83) $b_{19} = 1$.

We also require another normalization on the elements of c^1; i.e., we need to satisfy the constraint $c^1 \cdot q^* = 0$ for some positive vector q^*. We chose q^* to equal the sample mean of the observed q^t vectors; i.e., we set $q^* \equiv (1/19)\Sigma_{t=1}^{19} q^t$. We used the constraint $c^1 \cdot q^* = 0$ to solve for $c_{19}^1 = -\Sigma_{n=1}^{18} c_n^1 q_n^*/q_{19}^*$ and we substituted this c_{19}^1 into equations (82). Since the shares s_i^t sum to one for each period t, all 19 error terms ε_i^t for i = 1,...,19 cannot be distributed independently so we dropped the equation for product 19 from our list of estimating equations for Model 5.

We used the nonlinear regression software package in Shazam to estimate the 36 unknown b_n and c_n^1 in equations (82). The starting values for the b_n were the final estimates for the β_n from Model 3 above and the starting values for the c_n^1 were set at 0.01 for each n = 1,...,18. The initial log likelihood was 3074.663 and the final log likelihood was 3216.919, a gain of 142.603 for adding 18 new parameters to the linear utility model. The equation by equation R^2 values were as follows: 0.9661, 0.9787, 0.9623, 0.9889, 0.9608, 0.9521, 0.9628, 0.8002, 0.9657, 0.9752, 0.8337, 0.9224, 0.9867, 0.8936, 0.9673, 0.9555, 0.9064 and 0.9599. These are fairly high R^2 considering that the dependent variables are shares.

In order to determine the effects of changing the reference quantity vector q^*, we reestimated the above model but chose q^* to equal 1_{19}, a vector of ones of dimension 19. Thus in this case, we set the last component of the vector c^1 equal to the following expression:

(84) $c_{19}^1 = -\Sigma_{n=1}^{18} c_n^1$.

The estimated b and c^1 vectors changed when we reestimated Model 5 with the new normalization (84) but the predicted values for each observation turned out to be identical to the predicted values generated by our initial Model 5 and thus the R^2 for each equation did not change and the final log likelihood also did not change. Thus it appears that the choice of q^* does not matter, as long as the chosen reference vector q^* is strictly positive. Thus in subsequent models where we added additional columns to the C matrix, we chose q^* to equal 1_{19}. This choice of q^* led to simpler programming codes for our subsequent nonlinear regressions.

Our system of nonlinear estimating equations for *Model 6* are equations (82) where $A = bb^T - c^1c^{1T} - c^2c^{2T}$ with $c^{2T} = [0,c_2^2,...,c_{19}^2]$ and the normalizations $b_{19} = 1$, $c_{19}^1 = -\Sigma_{n=1}^{18} c_n^1$ and $c_{19}^2 = -\Sigma_{n=2}^{18} c_n^2$. Thus there are 18 + 18 + 17 unknown parameters to estimate in the A matrix. However, the nonlinear maximum likelihood estimation package in Shazam

[63] Again this is a slightly incorrect econometric specification since ε_n^t will automatically equal 0 if product n is not present during month t.

did not converge for this model. The problem is that the error specification that is used in the system command for the Nonlinear estimation option in Shazam also estimates the elements of the variance covariance matrix Σ. Thus for our Model 6, it is necessary to estimate the 53 unknown parameters in the A matrix plus 19×18/2 = 171 unknown variances and covariances. This proved to be a too difficult task for Shazam.

Thus in the following section, we will develop an alternative estimation strategy: we will stack up our 18 product estimating equations into a single estimating equation. In this setup, we will only have to estimate a single variance parameter instead of estimating 171 such parameters. The cost of using this strategy will be a somewhat incorrect variance specification; i.e., it is not likely that all product equations will have exactly the same variance but it will turn out that the predicted values for the product shares are quite close to the actual product shares so a somewhat incorrect variance specification will not be too troublesome.

9. The Single Equation Approach to the Estimation of KBF Preferences Using Share Equations

For *Model 7*, we stacked the first 18 Model 5 estimating share equations listed in equations (82) into a single equation and estimated the 18 unknown parameters in $A = bb^T$ with $b^T \equiv [b_1,b_2,...,b_{19}]$ and $b_{19} = 1$ using the single equation Nonlinear command in Shazam. The final log likelihood was 2379.380 and the R^2 was 0.9818. The estimated b_n were similar to the corresponding estimates that we got using the systems approach to estimate Model 3 but there were some differences.

An advantage of our present single equation approach is that we can now easily drop the 20 observations where the product was missing.[64] Thus for *Model 8*, we dropped the 20 observations for products 2, 4 and 12 for the months when these products were missing. Thus the number of observations for this new model is equal to (39×18) − 20 = 682. We found that the parameter estimates for this new model were exactly the same as the corresponding parameter estimates that we obtained for Model 7. However, the new log likelihood decreased to 2301.735 and the new R^2 decreased to 0.9814 (from the previous 0.9818).

In the models which follow, we will continue to drop the 20 observations that correspond to the months when the products were missing. Thus when we refer to the estimating equations (82), we are now assuming that the 20 missing product observations have been dropped from equations (82). Moreover, we also drop the 39 observations that correspond to the 19th product.[65]

Model 8 is the same as Model 5 above except we now use the single equation estimating strategy. Thus we have where $A = bb^T - c^1c^{1T}$ with the normalizations $b_{19} = 1$ and $c_{19}^1 = -\Sigma_{n=1}^{18} c_n^1$. We used the final estimates for the components of the b vector from Model 7

[64] The error terms will automatically be 0 for these 20 observations.

[65] Since the shares within one period must sum to 1, the corresponding error terms cannot all be independently distributed and thus we drop one set of shares from the estimating equations.

as starting coefficient values for Model 8 and we used $c_n^1 = 0.001$ for n = 1,...,18 as starting values for the components of the c vector. The final log likelihood for this model was 2445.888, an increase of 144.153 for adding 18 new parameters to the Model 7 parameters. The R^2 increased to 0.9884.

Model 9 adds a new column to the A matrix. Thus we have $A = bb^T - c^1c^{1T} - c^2c^{2T}$ with $c^{2T} = [0,c_2^2,...,c_{19}^2]$ and the normalizations $b_{19} = 1$, $c_{19}^1 = -\Sigma_{n=1}^{18} c_n^1$ and $c_{19}^2 = -\Sigma_{n=2}^{18} c_n^2$. We used the final estimates for the components of the b and c^1 vectors from Model 8 as starting coefficient values for Model 9 and we used $c_n^2 = 0.001$ for n = 2,...,18 as starting values for the nonzero components of the c^2 vector. The final log likelihood for this model was 2565.896, an increase of 120.008 for adding 17 new parameters to the Model 8 parameters. The R^2 increased to 0.9907.

Model 10 adds another column to the A matrix. Thus we have $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T}$ with $c^{3T} = [0,0,c_3^3,...,c_{19}^3]$ and the normalizations $b_{19} = 1$, $c_{19}^1 = -\Sigma_{n=1}^{18} c_n^1$, $c_{19}^2 = -\Sigma_{n=2}^{18} c_n^2$ and $c_{19}^3 = -\Sigma_{n=3}^{18} c_n^3$. We used the final estimates for the components of the b, c^1 and c^2 vectors from Model 9 as starting coefficient values for Model 10 and we used $c_n^3 = 0.001$ for n = 3,...,18 as starting values for the nonzero components of the c^3 vector. The final log likelihood for this model was 2614.526, an increase of 48.630 for adding 16 new parameters to the Model 9 parameters. The R^2 increased to 0.9919.

Finally, *Model 11* adds another column to the A matrix. Thus we have $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T}$ with $c^{4T} = [0,0,0,c_4^4,...,c_{19}^4]$ and the additional normalization $c_{19}^4 = -\Sigma_{n=4}^{18} c_n^4$. As usual, we used the final estimates for the components of the b, c^1, c^2 and c^3 vectors from Model 10 as starting coefficient values for Model 11 and we used $c_n^4 = 0.001$ for n = 4,...,18 as starting values for the nonzero components of the c^4 vector. The final log likelihood for this model was 2629.182, an increase of 14.656 for adding 15 new parameters to the Model 10 parameters. Thus the increase in log likelihood is now less than one per additional parameter. The single equation R^2 increased to 0.9922. However, this single equation R^2 is not comparable to the equation by equation R^2 that we obtained using the systems approach in the previous section. The comparable R^2 for each separate product share equation are as follows:[66] 0.9859, 0.9930, 0.9773, 0.9853, 0.9814, 0.9543, 0.9755, 0.8581, 0.9760, 0.9694, 0.8923, 0.9278, 0.9908, 0.9202, 0.9874, 0.9566, 0.9111 and 0.9653. The average R^2 is 0.9560 which is a relatively high average when estimating share equations.[67]

[66] These equation by equation R^2 are the squares of the correlation coefficients between the actual share equations for product n and the corresponding predicted values from the nonlinear regression. We included the 20 zero share and quantity product observations since our model correctly predicts these 0 shares. These 0 share observations were also included in the Model 4 systems regression in the previous section.

[67] Note that the KBF Model 11 average R^2, 0.9560, is above the Model 4 direct CES utility function average R^2, which was 0.9439. The present model is much more flexible and hence is likely to generate more reliable estimates of elasticities of demand More importantly for our purposes is the fact that the present model will generate finite reservation prices for the missing products (rather than the rather high infinite reservation prices that the CES model generates).

Since Model 11 estimates 84 unknown parameters and we have only 682 degrees of freedom, we have only about 8 degrees of freedom per parameter at this stage. Moreover, the increase in log likelihood going from Model 10 to 11 was relatively small. Thus we decided to stop adding columns to the C matrix at this point.

With the estimated b and c vectors in hand (denote them as b^* and c^{k*} for k = 1,2,3,4), form the estimated A matrix as follows:

(85) $A^* \equiv b^*b^{*T} - c^{1*}c^{1*T} - c^{2*}c^{2*T} - c^{3*}c^{3*T} - c^{4*}c^{4*T}$

and denote the ij element of A^* as a_{ij}^* for i,j = 1,...,19. The *predicted expenditure share* for product i in month t is s_i^{t*} defined as follows:

(86) $s_i^{t*} \equiv q_i^t\, \Sigma_{k=1}^{19}\, a_{ik}^* q_k^t / [\Sigma_{n=1}^{19} \Sigma_{m=1}^{19}\, a_{nm}^* q_n^t q_m^t]$; t = 1,...,39; i = 1,...,19.

The *predicted price* for product i in month t is defined as follows:

(87) $p_i^{t*} \equiv e^t\, \Sigma_{k=1}^{19}\, a_{ik}^* q_k^t / [\Sigma_{n=1}^{19} \Sigma_{m=1}^{19}\, a_{nm}^* q_n^t q_m^t]$; t = 1,...,39; i = 1,...,19

where $e^t \equiv p^t \cdot q^t$ is period t sales or expenditures on the 19 products during month t.[68] We calculated the predicted prices defined by (87) for all products and all months.

Of particular interest are the predicted prices for products 2 and 4 for months 1-8 and for product 12 for months 10 and 20-22 when these products were not available. The predicted prices for products 2 and 4 for the first 8 months in our sample period were 1.62, 1.56, 1.60, 1.52, 1.61, 1.52, 1.70. 1.97 and 1.85, 1.46, 1.80, 1.37, 1.77, 1.83, 1.88, 2.27 respectively. The predicted prices for product 12 for months 10 and 20-22 were 1.37, 1.20, 1.22 and 1.28. These prices are rather far removed from the infinite reservation prices implied by the CES model.

However, there is a problem with our model: even though the predicted expenditure shares are quite close to the actual expenditure shares, *the predicted prices are not particularly close to the actual prices*. Thus the equation by equation R^2 for the 19 products were as follows:[69] 0.7571, 0.8209, 0.8657, 0.8969, 0.9025, 0.7578, 0.8660, 0.0019, 0.2517, 0.1222, 0.0000, 0.0013, 0.9125, 0.6724, 0.4609, 0.7235, 0.5427, 0.8148 and 0.4226. The average R^2 is only 0.5681 which is not very satisfactory. How can the R^2 for the share equations be so high while the corresponding R^2 for the fitted prices are so low? The answer appears to be the following one: when a price is unusually low, the corresponding quantity is unusually high and vice versa. Thus the errors in the fitted price equations and the corresponding fitted quantity equations tend to offset each other and so

---

[68] The predicted price p_i^{t*} is also equal to $[e^t \partial f(q^t)/\partial q_i]/f(q^t)$ where $f(q) \equiv (q^T A^* q)^{1/2}$. This follows from the first order necessary conditions for the month t utility maximization problem (with no errors) which are $p^{t*}/e^t = \nabla f(q^t)/f(q^t)$ where p^{t*} is the month t vector of predicted prices.

[69] For the 20 observations where the product was not available, we used the predicted prices as actual prices in computing these R^2. Thus for products 2, 4 and 12, the R^2 listed above are overstated.

the fitted share equations are fairly close to the actual shares whereas the errors in the fitted price and quantity equations can be rather large but in opposite directions.

The above poor fits for the predicted prices caused us to re-examine our estimating strategy. The primary purpose of our estimation of preferences is to obtain "reasonable" predicted prices for products which are not available. *Our primary purpose is not the prediction of expenditure shares; it is the prediction of reservation prices*! Thus in the following section, we will switch from estimating share equations to the estimation of price equations.

10. The Single Equation Approach to the Estimation of KBF Preferences Using Price Equations

Our *Model 12* system of estimating equations uses prices as the dependent variables:

(88) $p_i^t \equiv e^t \Sigma_{k=1}^{19} a_{ik}q_k^t/[\Sigma_{n=1}^{19}\Sigma_{m=1}^{19} a_{nm}q_n^tq_m^t] + \varepsilon_i^t$; $\qquad t = 1,...,39; i = 1,...,18$

where the A matrix is defined as $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T}$ and the vectors b and c^1 to c^4 satisfy the same restrictions as in Model 11 in the previous section. We stack up the estimating equations defined by (88) into a single nonlinear regression and we drop the observations that correspond to products i that were not available in period t.

We used the final estimates for the components of the b, c^1, c^2, c^3 and c^4 vectors from Model 11 as starting coefficient values for Model 12. The initial log likelihood of our new model using these starting values for the coefficients was 415.576. The final log likelihood for this model was 518.881, an increase of 103.305. Thus switching from having shares to having prices as the dependent variables did significantly change our estimates. The single equation R^2 was 0.9453. We used our estimated coefficients to form predicted prices p_i^{t*} using equations (87) evaluated at our new parameter estimates. The equation by equation R^2 comparing the predicted prices for the 19 products with the actual prices were as follows:[70] 0.8295, 0.8621, 0.9001, 0.9163, 0.8988, 0.8319, 0.9134, 0.0350, 0.2439, 0.2754, 0.0236, 0.0068, 0.8704, 0.6951, 0.4211, 0.8082, 0.6180, 0.8517 and 0.2868. The average R^2 was 0.5941.

Since the predicted prices are still not very close to the actual prices, we decided to press on and estimate *Model 13* which added another rank 1 substitution matrix to the substitution matrix; i.e., we set $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T} - c^5c^{5T}$ where $c^{5T} = [0,0,0,0,c_5^5,...,c_{19}^5]$ and the additional normalization $c_{19}^5 = -\Sigma_{n=5}^{18} c_n^5$.

We used the final estimates for the components of the b, c^1, c^2, c^3 and c^4 vectors from Model 12 as starting coefficient values for Model 13 along with $c_n^5 = 0.001$ for n = 5,6,...,18. The initial log likelihood of our new model using these starting values for the coefficients was 518.881. The final log likelihood for this model was 550.346, an

[70] Again, for the 20 observations where the product was not available, we used the predicted prices as actual prices in computing these R^2. As usual, these R^2 are just the squares of the correlation coefficients between the 39 predicted prices and the actual prices for product i for i = 1,...,19.

increase of 31.465. The single equation R^2 was 0.9501. We used our estimated coefficients to form predicted prices p_i^{t*} using equations (87) evaluated at our new parameter estimates. The equation by equation R^2 comparing the predicted prices for the 19 products with the actual prices were as follows: 0.8295, 0.8621, 0.9001, 0.9163, 0.8988, 0.8319, 0.9134, 0.0350, 0.2439, 0.2754, 0.0236, 0.0068, 0.8704, 0.6951, 0.4211, 0.8082, 0.6180, 0.8517 and 0.2868.

Since the increase in log likelihood for Model 13 over Model 12 was fairly large, we decided to add another rank 1 matrix to the A matrix. Thus for *Model 14,* we set $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T} - c^5c^{5T} - c^6c^{6T}$ where $c^{6T} = [0,0,0,\ 0,0,c_6^{\,6},...,c_{19}^{\,6}]$ and the additional normalization $c_{19}^{\,6} = -\,\Sigma_{n=6}^{18}\, c_n^{\,6}$.

We used the final estimates for the components of the b, c^1, c^2, c^3, , c^4 and c^5 vectors from Model 13 as starting coefficient values for Model 14 along with $c_n^{\,6} = 0.001$ for n = 6,7,...,18. The final log likelihood for this model was 568.877, an increase of 18.531. The single equation R^2 was 0.9527.

Model 14 had 111 unknown parameters that were estimated (plus a variance parameter). We had only 680 observations and so we decided to call a halt to our estimation procedure. Also convergence of the nonlinear estimation was slowing down and so it was becoming increasingly difficult for Shazam to converge to the maximum likelihood estimates. Thus we stopped our sequential estimation process at this point.

The parameter estimates for Model 14 are listed below in Table 7.[71]

Table 7: Estimated Parameters for Model 14

Coef	Estimate	t Stat	Coef	Estimate	t Stat	Coef	Estimate	t Stat
b_1^*	1.3450	11.388	c_3^{2*}	-0.0780	-0.113	c_9^{4*}	0.1525	0.256
b_2^*	1.3138	10.769	c_4^{2*}	-0.7121	-0.724	c_{10}^{4*}	-0.0321	-0.053
b_3^*	1.4318	11.311	c_5^{2*}	-0.0973	-0.242	c_{11}^{4*}	-0.6147	-0.812
b_4^*	1.5697	11.541	c_6^{2*}	-0.6352	-1.275	c_{12}^{4*}	-1.5855	-1.128
b_5^*	1.3709	11.226	c_7^{2*}	-0.6146	-1.378	c_{13}^{4*}	-0.2332	-0.311
b_6^*	2.0885	11.886	c_8^{2*}	1.1453	1.811	c_{14}^{4*}	-0.1605	-0.242
b_7^*	1.4180	11.403	c_9^{2*}	-0.3882	-1.351	c_{15}^{4*}	-0.6687	-1.690
b_8^*	0.8216	9.021	c_{10}^{2*}	-0.5408	-1.728	c_{16}^{4*}	-0.2246	-0.302
b_9^*	0.5692	9.670	c_{11}^{2*}	0.9956	2.140	c_{17}^{4*}	3.2700	3.547
b_{10}^*	0.5880	9.476	c_{12}^{2*}	1.9022	1.674	c_{18}^{4*}	-0.3506	-0.436
b_{11}^*	0.8010	10.010	c_{13}^{2*}	-0.4551	-1.480	c_5^{5*}	-0.0555	-0.105
b_{12}^*	1.0962	9.162	c_{14}^{2*}	-0.7303	-1.455	c_6^{5*}	-0.0444	-0.118
b_{13}^*	1.2411	11.136	c_{15}^{2*}	-0.3204	-0.795	c_7^{5*}	-0.0952	-0.056
b_{14}^*	1.6071	11.124	c_{16}^{2*}	0.2584	0.842	c_8^{5*}	-0.2548	-0.038
b_{15}^*	0.7145	10.115	c_{17}^{2*}	0.0199	0.007	c_9^{5*}	-0.6205	-0.887
b_{16}^*	1.3384	11.465	c_{18}^{2*}	-0.5013	-1.128	c_{10}^{5*}	-0.5634	-0.792
b_{17}^*	1.5759	7.968	c_3^{3*}	1.3620	5.405	c_{11}^{5*}	-0.1094	-0.028

[71] The standard errors for the estimated coefficients are equal to the coefficient estimate listed in Table 7 divided by the corresponding t statistic.

b_{18}^{*}	**1.3699**	**11.400**	**c_4^{3*}**	**1.7166**	**4.405**	**c_{12}^{5*}**	**-0.3085**	**-0.036**
c_1^{1*}	**1.9832**	**10.031**	**c_5^{3*}**	**1.0262**	**5.104**	**c_{13}^{5*}**	**0.6261**	**0.120**
c_2^{1*}	**1.6598**	**6.653**	**c_6^{3*}**	**-0.4277**	**-1.090**	**c_{14}^{5*}**	**0.0516**	**0.013**
c_3^{1*}	**-0.2507**	**-1.186**	**c_7^{3*}**	**0.8958**	**2.431**	**c_{15}^{5*}**	**-0.0774**	**-0.024**
c_4^{1*}	**0.1313**	**0.552**	**c_8^{3*}**	**-0.4633**	**-0.809**	**c_{16}^{5*}**	**0.7559**	**0.134**
c_5^{1*}	**0.0126**	**0.088**	**c_9^{3*}**	**-0.0097**	**-0.041**	**c_{17}^{5*}**	**0.6127**	**0.225**
c_6^{1*}	**-0.0106**	**-0.050**	**c_{10}^{3*}**	**-0.0785**	**-0.277**	**c_{18}^{5*}**	**0.4772**	**0.054**
c_7^{1*}	**-0.3807**	**-1.914**	**c_{11}^{3*}**	**-0.5885**	**-1.064**	**c_6^{6*}**	**-0.0093**	**-0.028**
c_8^{1*}	**-0.4251**	**-1.856**	**c_{12}^{3*}**	**-0.1383**	**-0.137**	**c_7^{6*}**	**0.1776**	**0.380**
c_9^{1*}	**-0.0179**	**-0.114**	**c_{13}^{3*}**	**-0.0220**	**-0.093**	**c_8^{6*}**	**-0.7621**	**-0.300**
c_{10}^{1*}	**-0.2753**	**-1.576**	**c_{14}^{3*}**	**-0.4538**	**-1.183**	**c_9^{6*}**	**-0.0805**	**-0.015**
c_{11}^{1*}	**-0.9620**	**-4.477**	**c_{15}^{3*}**	**-0.4603**	**-2.033**	**c_{10}^{6*}**	**0.0788**	**0.016**
c_{12}^{1*}	**-0.8816**	**-2.693**	**c_{16}^{3*}**	**-0.0116**	**-0.064**	**c_{11}^{6*}**	**-0.4361**	**-0.270**
c_{13}^{1*}	**0.1146**	**1.524**	**c_{17}^{3*}**	**-2.1645**	**-2.382**	**c_{12}^{6*}**	**-0.9471**	**-0.231**
c_{14}^{1*}	**-0.2175**	**-1.016**	**c_{18}^{3*}**	**0.0091**	**0.033**	**c_{13}^{6*}**	**-0.6016**	**-0.114**
c_{15}^{1*}	**-0.1262**	**-0.854**	**c_4^{4*}**	**-0.5049**	**-0.708**	**c_{14}^{6*}**	**0.4660**	**0.979**
c_{16}^{1*}	**0.1367**	**1.247**	**c_5^{4*}**	**0.4895**	**1.341**	**$c_{15}^{6(}$**	**0.3859**	**0.335**
c_{17}^{1*}	**-0.6792**	**-1.544**	**c_6^{4*}**	**0.2658**	**0.466**	**$c_{16}^{6(}$**	**0.6562**	**0.103**
c_{18}^{1*}	**0.0849**	**0.450**	**c_7^{4*}**	**0.3802**	**0.625**	**c_{17}^{6*}**	**0.1162**	**0.002**
c_2^{2*}	**0.7173**	**1.584**	**c_8^{4*}**	**-0.1078**	**-0.118**	**c_{18}^{6*}**	**1.0227**	**0.258**

The estimated b_n^* in Table 7 for n = 1,...,18 plus $b_{19} = 1$ are proportional to the vector of first order partial derivatives of the KBF utility function f(q) evaluated at the vector of ones, $\nabla_q f(1_{19})$. Thus the b_n^* can be interpreted as estimates of the relative quality of the 19 products. Viewing Table 7, it can be seen that the highest quality products were products 6, 17 and 4 ($b_6^* = 2.09$, $b_{17}^* = 1.58$, $b_4^* = 1.57$) and the lowest quality products were products 9, 10 and 15 ($b_9^* = 0.57$, $b_{10}^* = 0.59$, $b_{15}^* = 0.71$).

With the estimated b^* and c^* vectors in hand (denote them as b^* and c^{k*} for k = 1,...,6), form the estimated A matrix as follows:

(89) $A^* \equiv b^*b^{*T} - c^{1*}c^{1*T} - c^{2*}c^{2*T} - c^{3*}c^{3*T} - c^{4*}c^{4*T} - c^{5*}c^{5*T} - c^{6*}c^{6*T}$

and denote the ij element of A^* as a_{ij}^* for i,j = 1,...,19. The *predicted price* for product i in month t is defined as follows:

(90) $p_i^{t*} \equiv e^t\, \Sigma_{k=1}^{19}\, a_{ik}^* q_k^t / [\Sigma_{n=1}^{19}\Sigma_{m=1}^{19}\, a_{nm}^* q_n^t q_m^t]$; t = 1,...,39; i = 1,...,19

where $e^t \equiv p^t \cdot q^t$ is period t sales or expenditures on the 19 products during month t. We calculated the predicted prices defined by (90) for all products and all months.

Of particular interest are the predicted prices for products 2 and 4 for months 1-8 and for product 12 for months 10 and 20-22 when these products were not available. The predicted prices for products 2 and 4 for the first 8 months in our sample period were 1.62, 1.56, 1.60, 1.52, 1.61, 1.52, 1.70. 1.97 and 1.85, 1.46, 1.80, 1.37, 1.77, 1.83, 1.88, 2.27 respectively. The predicted prices for product 12 for months 10 and 20-22 were 1.37, 1.20, 1.22 and 1.28. These predicted prices will be used as our "best" reservation prices for the missing products in the remainder of the paper.

The equation by equation R^2 that compares the predicted prices for the 19 products with the actual prices were as follows:[72] 0.8274, 0.8678, 0.9001, 0.9174, 0.8955, 0.8536, 0.9047, 0.0344, 0.3281, 0.4242, 0.0516, 0.2842, 0.8650, 0.7280, 0.4872, 0.8135, 0.8542, 0.8479 and 0.3210. The average R^2 for Model 14 was 0.6424. Twelve of the 19 equations had an R^2 greater than 0.70 while 5 of the equations had an R^2 less than 0.40.[73]

The *month t utility level* or aggregate quantity level implied by the KBF model, Q_{KBF}^t, is defined as follows:

$$(91)\ Q_{KBF}^t \equiv (q^{tT}A^*q^t)^{1/2}\ ; \qquad t = 1,...,39.$$

The corresponding *KBF (unnormalized) implicit price level*, P_{KBF}^{t*}, is defined as period t sales of the 19 products, e^t, divided by the period t aggregate KBF quantity level, Q_{KBF}^t:

$$(92)\ P_{KBF}^{t*} \equiv e^t/Q_{KBF}^t\ ; \qquad t = 1,...,39.$$

The month t *KBF price index*, P_{KBF}^t, is defined as the month t KBF price level divided by the month 1 KBF price level; i.e., $P_{KBF}^t \equiv P_{KBF}^{t*}/P_{KBF}^{1*}$ for t = 1,...,39. The KBF price index is listed below in Table 8. These econometrically based KBF price indexes can be compared to our econometrically based CES price indexes P_{UCES}^t that are listed above in Table 5.

Now that we have imputed prices for the unavailable products, we can compute fixed base and chained Fisher indexes using these prices for the unavailable products along with the corresponding 0 quantities. Denote these Fisher indexes for month t that use our imputed prices as P_{FI}^t and P_{FICh}^t for t = 1,...,39. These indexes are also listed in Table 8.

The econometrically based CES price indexes that are listed in Table 5 above used a systems approach to the estimation of the CES utility function. But we used expenditure shares as the dependent variables in this systems approach. What happens if we estimate a CES utility function using prices as the dependent variables and using our one big estimating equation approach that we used in this section? We conclude this section by answering this question.

[72] As usual, the R^2 for the 39 product n equations was defined as the square of the correlation coefficient between the actual product n prices and their predicted counterparts using equations (90). For the prices of the 20 observations where a product was not available, we used the predicted prices in place of the actual prices. Thus the R^2 is overstated for products 2, 4 and 12.
[73] The sample average expenditure shares of these low R^2 products was 0.026, 0.026, 0.043, 0.025 and 0.050 respectively. Thus these low R^2 products are relatively unimportant compared to the high expenditure share products.

Equations (67) above used shares as the dependent variables when we estimated the CES utility function. Take the ith share equation defined by (67) and multiply both sides of the period t equation by e^t/q_i^t. This leads to the following system of estimating equations:

$$(93)\ p_i^t = [e^t/q_i^t][\beta_i (q_i^t)^s/\sum_{n=1}^{19} \beta_n (q_n^t)^s] + \varepsilon_i^t \qquad t = 1,...,39;\ i = 1,...,18.$$

Now stack the above 702 equations into a single estimating equation and drop the 20 observations where $q_i^t = 0$. We assume (for simplicity) that all errors ε_i^t are independently distributed normal variables with means 0 and variances σ^2. As usual, we set $\beta_{19} = 1$. Call this *Model 15*. We estimated the resulting nonlinear system of equations for Model 15 using the Nonlinear option in Shazam. For starting values for the regression, we used the estimated β_n from Model 3 and set s = 1 (which corresponds to the linear utility function). The initial log likelihood was equal to 165.039 and the final log likelihood was equal to 483.834. This final log likelihood can be compared to our final log likelihood from the KBF Model 14, which was substantially higher at 568.877. The single equation R^2 was 0.9393, which is below the single equation R^2 from Model 14, which was 0.9527. The 18 estimated β_n^* were as follows (with t statistics in brackets): 0.99 (15.2), 0.92 (15.2), 1.04 (15.6), 1.05 (15.6), 1.22 (15.7), 1.49 (15.7), 1.08 (15.7), 0.71 (15.2), 0.50 (14.5), 0.47 (14.6), 0.73 (15.2), 0.85 (15.4), 1.09 (15.0), 1.31 (15.8), 0.52 (15.0), 1.13 (15.6), 1.31 (15.7) and 1.01 (15.6). Thus the highest quality products are 6, 14 and 17 while the lowest quality products are 10, 9 and 15. The estimated parameter s was $s^* = 0.85365$ (157.24). This is virtually identical to our estimate for s from Model 4 (which used the systems approach to CES utility function estimation) which was 0.85374. *Thus the corresponding elasticity of substitution was virtually identical whether we use the systems approach or the single equation approach to estimation*. However, it turns out that the price index that is implied by our newly estimated CES utility function is not identical to the CES implicit price index that appeared in Table 5 above. The parameter s is virtually identical in the two CES utility function models but the estimated β_n coefficients differ enough to generate somewhat different implicit price indexes as we shall see.

The *month t utility level* or aggregate quantity level implied by the *New Single* equation CES Model 15, Q_{CESN}^t, is defined as follows:

$$(94)\ Q_{CESN}^t \equiv [\sum_{n=1}^{19} \beta_n^* (q_n^t)^{s^*}]^{1/s^*}\ ; \qquad t = 1,...,39.$$

The corresponding *New CES (unnormalized) implicit price level*, P_{CESN}^{t*}, is defined as period t sales of the 19 products, e^t, divided by the period t aggregate New CES quantity level, Q_{CESN}^t:

$$(95)\ P_{CESN}^{t*} \equiv e^t/Q_{CESN}^t\ ; \qquad t = 1,...,39.$$

The month t *New CES price index*, P_{CESN}^t, is defined as the month t CESN price level divided by the month 1 CESN price level; i.e., $P_{CESN}^t \equiv P_{CESN}^{t*}/P_{CESN}^{1*}$ for t = 1,...,39. The CESN price index is listed below in Table 8. This econometrically based CES implicit price index can be compared to our earlier utility function econometrically based

CES implicit price index P_{UCES}^t that was listed above in Table 5. For convenience, we list it again in Table 8.

It will turn out that we can define estimates of the change in the true cost of living index due to changes in the availability of products in our KBF framework in a manner that is similar to that used by Feenstra. In order to accomplish this task, we need to define various Fisher price indexes that make use of the predicted prices that result from the estimation of Model 14. The first of these additional Fisher indexes is P_{FI}^t which uses the predicted or imputed prices for the missing products (along with the associated 0 quantities) along with the actual prices and quantities for the remaining products to produce a fixed base Fisher price index. Using the same data, we can produce a chained Fisher price index, P_{FICh}^t. These indexes are listed in Table 8 below. The next two Fisher price indexes are the fixed base and chained maximum overlap Fisher indexes P_F^t and P_{FCh}^t that were defined earlier in Section 5 above. These indexes are listed in Table 1 and are listed again in Table 8 below. The final two Fisher indexes are the fixed base and chained Fisher price indexes, P_{FP}^t and P_{FPCh}^t, that use the *predicted prices* for all products and all time periods defined by equations (90), which in turn are generated by the Model 14 estimated KBF utility function. It turns out that these indexes are identical and are also equal to the corresponding KBF price indexes, P_{KBF}^t, that are directly defined by the estimated utility function; see equations (92), which define the $P_{KBF}^{t*} \equiv e^t/Q_{KBF}^t$ which in turn are normalized to define the P_{KBF}^t. Thus we have $P_{KBF}^t = P_{FP}^t = P_{FPCh}^t$ for all t. All of these indexes are listed in Table 8.

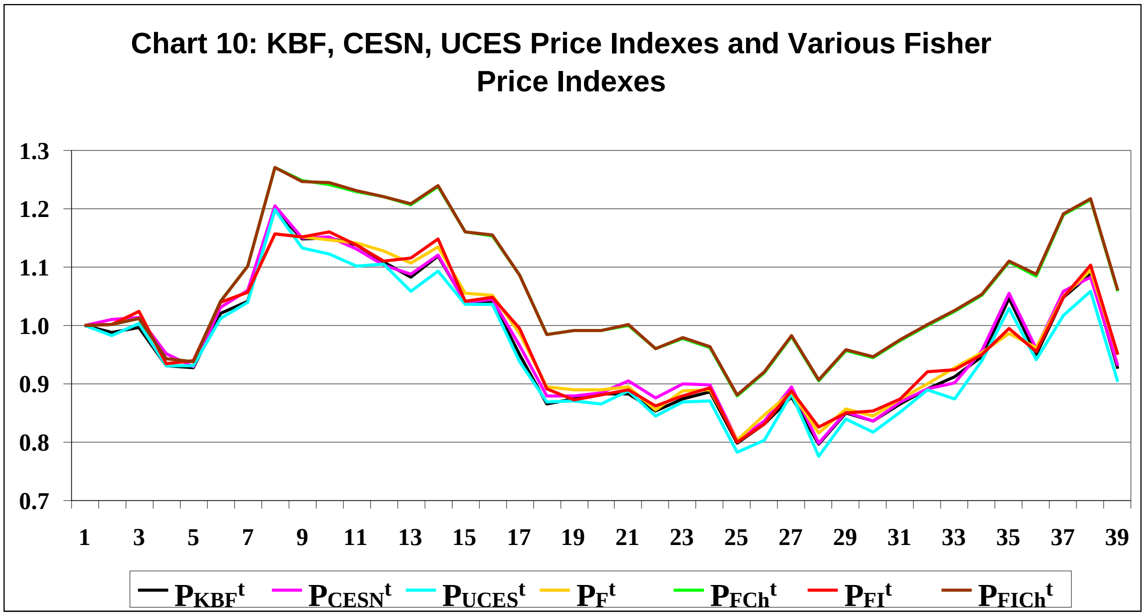
Table 8: The KBF Implicit Price Index, Alternative CES Implicit Price Indexes and Various Fisher Price Indexes using KBF Imputed Prices for Unavailable Products

Month	P_{KBF}^t	P_{CESN}^t	P_{UCES}^t	P_F^t	P_{FCh}^t	P_{FI}^t	P_{FICh}^t	P_{FP}^t
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	0.98816	1.01088	0.98359	1.00218	1.00218	1.00218	1.00218	0.98816
3	0.99734	1.01402	1.00369	1.02342	1.01124	1.02342	1.01124	0.99734
4	0.93078	0.95254	0.93029	0.93388	0.94265	0.93388	0.94265	0.93078
5	0.92749	0.92981	0.93074	0.93964	0.93715	0.93964	0.93715	0.92749
6	1.02000	1.03087	1.01226	1.03989	1.04075	1.03989	1.04075	1.02000
7	1.04222	1.05991	1.04046	1.05662	1.10208	1.05662	1.10208	1.04222
8	1.19800	1.20432	1.19753	1.15739	1.26987	1.15739	1.26987	1.19800
9	1.14801	1.14958	1.13261	1.15209	1.24778	1.15165	1.24727	1.14801
10	1.14946	1.15154	1.12326	1.14617	1.24137	1.16081	1.24528	1.14946
11	1.13863	1.13028	1.10248	1.14088	1.22950	1.13876	1.23033	1.13863
12	1.10858	1.10373	1.10469	1.12760	1.22009	1.10951	1.22091	1.10858
13	1.08290	1.08843	1.05886	1.10698	1.20731	1.11511	1.20813	1.08290
14	1.11953	1.12064	1.09361	1.13419	1.23863	1.14803	1.23948	1.11953
15	1.04018	1.03928	1.03612	1.05579	1.15978	1.04086	1.16056	1.04018
16	1.04081	1.04520	1.03616	1.05099	1.15371	1.04836	1.15449	1.04081
17	0.94930	0.96493	0.93755	0.98640	1.08568	0.99410	1.08642	0.94930
18	0.86479	0.87948	0.86830	0.89490	0.98385	0.89105	0.98452	0.86479
19	0.87354	0.87896	0.87013	0.89032	0.99122	0.87308	0.99189	0.87355
20	0.88231	0.88500	0.86511	0.89016	0.99104	0.88051	0.99193	0.88231
21	0.88333	0.90573	0.88722	0.89453	1.00061	0.88920	1.00150	0.88333

22	**0.85408**	**0.87541**	**0.84505**	**0.85466**	**0.95983**	**0.86217**	**0.96068**	**0.85408**
23	**0.87493**	**0.90000**	**0.86932**	**0.88842**	**0.97730**	**0.87981**	**0.97902**	**0.87493**
24	**0.88535**	**0.89766**	**0.87123**	**0.88930**	**0.96178**	**0.89357**	**0.96347**	**0.88535**
25	**0.79866**	**0.80271**	**0.78295**	**0.80421**	**0.88017**	**0.80050**	**0.88172**	**0.79866**
26	**0.83066**	**0.83631**	**0.80363**	**0.84644**	**0.91938**	**0.83026**	**0.92100**	**0.83066**
27	**0.87815**	**0.89563**	**0.88298**	**0.88641**	**0.98171**	**0.88749**	**0.98344**	**0.87815**
28	**0.79681**	**0.79846**	**0.77627**	**0.81528**	**0.90580**	**0.82665**	**0.90739**	**0.79681**
29	**0.85006**	**0.85211**	**0.83908**	**0.85705**	**0.95671**	**0.85086**	**0.95839**	**0.85006**
30	**0.83602**	**0.83604**	**0.81745**	**0.84508**	**0.94446**	**0.85383**	**0.94612**	**0.83602**
31	**0.86528**	**0.86982**	**0.85125**	**0.87333**	**0.97386**	**0.87411**	**0.97557**	**0.86528**
32	**0.89165**	**0.89127**	**0.89022**	**0.89973**	**1.00016**	**0.92038**	**1.00192**	**0.89165**
33	**0.91245**	**0.90245**	**0.87329**	**0.92673**	**1.02452**	**0.92404**	**1.02632**	**0.91245**
34	**0.94661**	**0.95715**	**0.93999**	**0.95385**	**1.05227**	**0.95012**	**1.05412**	**0.94660**
35	**1.04573**	**1.05575**	**1.02941**	**0.98690**	**1.10820**	**0.99422**	**1.11014**	**1.04573**
36	**0.95051**	**0.95953**	**0.94062**	**0.96237**	**1.08529**	**0.95568**	**1.08719**	**0.95051**
37	**1.04791**	**1.05833**	**1.01774**	**1.04948**	**1.18995**	**1.04808**	**1.19204**	**1.04791**
38	**1.08860**	**1.08352**	**1.05781**	**1.09545**	**1.21560**	**1.10279**	**1.21773**	**1.08860**
39	**0.92639**	**0.93053**	**0.90282**	**0.94999**	**1.05918**	**0.95071**	**1.06104**	**0.92639**

It can be seen that the first three econometrically based price indexes are all quite close to each other; in fact, it is difficult to distinguish $P_{KBF}{}^t$ from $P_{CESN}{}^t$ while our Section 5 CES price index $P_{UCES}{}^t$ tends to be slightly below the first two indexes (which both use prices as the dependent variables in their regressions). The two chained indexes based on actual price data, the maximum overlap chained Fisher index, $P_{FCh}{}^t$, and the chained Fisher index that uses the estimated reservation prices from Model 14, $P_{FICh}{}^t$, suffer from a considerable amount of upward chain drift (most of which occurs between months 8 and 9). The Fisher fixed base and chained indexes that use predicted prices from Model 14 everywhere, $P_{FP}{}^t$ and $P_{FPCh}{}^t$, are both exactly equal to $P_{KBF}{}^t$ as theory requires. The fixed base Fisher index that used the KBF reservation prices for unavailable products, $P_{FI}{}^t$, was generally a bit above $P_{CESN}{}^t$ throughout the sample period and ended up at 0.95071.[74] Chart 10 below plots the first 7 price indexes listed in Table 8.

[74] The sample means of P_{KBF}, P_{CESN}, P_{UCES}, P_{FI} and P_{FICh} were 0.9658, 0.9730, 0.9540, 0.9743 and 1.059 respectively. The correlation coefficients of P_{CESN}, P_{UCES}, P_{FI} and P_{FICh} with P_{KBF} were 0.9968, 0.9940, 0.9877 and 0.9366 respectively.



It can be seen that the two chained Fisher indexes are well above the other indexes. It can also be seen that the remaining indexes are not all that different. Thus in particular, the easy to calculate fixed base maximum overlap Fisher price index $P_F{}^t$ provides a satisfactory approximation to the theoretically more desirable fixed base Fisher index $P_{FI}{}^t$ that used imputed reservation prices for the missing products.

Some other conclusions that we can draw from the above Table and Chart is that $P_{KBF}{}^t$ and $P_{CESN}{}^t$ approximate each other quite closely. Both of these indexes are based on estimating a direct purchaser utility function, using observed prices as the dependent variables. $P_{UCES}{}^t$ is also fairly close to our preferred index $P_{KBF}{}^t$ where $P_{UCES}{}^t$ is also based on estimating a CES purchaser utility function but sales shares were used as the dependent variables in the estimating equations.

Feenstra's methodology for measuring the benefits and costs of changing product availability basically assumes that with the help of some econometric estimation (i.e., the estimation of the elasticity of substitution), it is possible to calculate the purchaser's true cost of living index. It is also possible to calculate an exact index for the cost of living index for the maximum overlap universe. Thus dividing the true cost of living by the maximum overlap cost of living, Feenstra obtains an index that can be interpreted as the net benefits of the changing availability of products between the two periods being compared. We can apply a variant of this methodology in the present situation. Having estimated reservation prices for the missing products, we can calculate a comprehensive Fisher chain link index going from period t−1 to period t, which is $P_{FICh}{}^t/P_{FICh}{}^{t-1}$. Holding product availability constant, we can calculate the corresponding chain link for the maximum overlap Fisher index for the products that are present in both periods, which is $P_{FCh}{}^t/\ P_{FCh}{}^{t-1}$. These indexes are listed in Table 8 above. The ratio of these two indexes is defined as follows:

(96) $I_{KBF}^{t} \equiv [P_{FICh}^{t}/P_{FICh}^{t-1}]/[P_{FCh}^{t}/P_{FCh}^{t-1}]$; $\qquad$ t = 2,3,...,T.

This index can be interpreted as a "correction" index which when multiplied by the readily calculated maximum overlap index $P_{FCh}^{t}/P_{FCh}^{t-1}$ gives us the "true" chain link index $P_{FICh}^{t}/P_{FICh}^{t-1}$, or it can be interpreted as the amount of bias in the maximum overlap chain link index due to changes in the availability of products. This index can be calculated for our data set using the information on P_{FICh}^{t} and P_{FCh}^{t} listed above in Table 8. When the availability of products increases (decreases) going from period t−1 to t, we expect I_{KBF}^{t} to be less (greater) than one and $1 - I_{KBF}^{t}$ is an estimate of the percentage decrease (increase) in the cost of living due to the increased (decreased) availability of products. If the availability of products is constant over periods t −1 and t, then I_{KBF}^{t} will be equal to 1. Thus the periods where I_{KBF}^{t} differs from 1 in our data set are periods 9, 10, 11, 20 and 23. The values for I_{KBF}^{t} for these periods are listed in Table 9 below.

Table 9: Alternative Bias Indexes for Fisher Maximum Overlap Chain Link Indexes Using KBF Imputed Prices for Unavailable Products and Using KBF Imputed Prices for All Products

t	I_{KBF}^{t}	I_{KBF}^{t*}
9	**0.99960**	**0.99836**
10	**1.00355**	**1.00124**
11	**0.99754**	**0.99847**
20	**1.00021**	**1.00294**
23	**1.00086**	**0.99988**
Product	**1.00176**	**1.00088**

We expected I_{KBF}^{t} to be less than 1 for periods 9, 11 and 23 when product availability increased and to be greater than 1 for periods 10 and 20 when product availability decreased. However, the month 23 value was $I_{KBF}^{23} = 1.00086$ which is greater than unity so the increased availability of product 12 in month 23 led to an *increase* in the cost of living rather than a *decrease* as expected. The product of the 5 nonunitary values for I_{KBF}^{t} t was 1.00176 (see the last row of Table 9) and so the overall increase in the availability of products led to a small *increase* in the cost of living over the sample period equal to 0.176 percentage points, rather than a *decrease* as was expected. Since our estimated KBF utility function is not exactly consistent with the observed data, these kinds of counterintuitive results can occur.

One method for eliminating anomalous results is to replace all observed prices by their predicted prices (and of course use predicted prices for the missing product prices). The comprehensive predicted Fisher chain link index going from period t−1 to period t using actual quantities q_i^{t} and predicted prices p_i^{t*} defined by definitions (90) above is $P_{FPCh}^{t}/P_{FPCh}^{t-1} = P_{FP}^{t}/P_{FP}^{t-1} = P_{KBF}^{t}/P_{KBF}^{t-1}$. Define P_{FPMCh}^{t} as the *maximum overlap chained Fisher price index* that uses actual quantities q_i^{t} and the predicted prices p_i^{t*} defined by (90) above. Holding product availability constant, we can calculate the corresponding chain link for this maximum overlap Fisher index using predicted prices for the products that are present in both periods, which is $P_{FPMCh}{}^{t}/P_{FPMCh}{}^{t-1}$. The ratio of these two link indexes is defined as $I_{KBF}{}^{t*}$:

$$(97)\ I_{KBF}{}^{t*} \equiv [P_{FPCh}{}^{t}/P_{FPCh}{}^{t-1}]/[P_{FPMCh}{}^{t}/P_{FPMCh}{}^{t-1}]\ ; \qquad t = 2,3,...,T.$$

This index can also be interpreted as a "correction" index which when multiplied by the maximum overlap index using predicted prices, $P_{FPMCh}{}^{t}/P_{FPMCh}{}^{t-1}$, gives us the "true" chain link index $P_{FPCh}{}^{t}/P_{FPCh}{}^{t-1}$ which is exactly consistent with our Model 14 estimated KBF utility function. Alternatively, it can be interpreted as an estimator for the amount of bias in the maximum overlap chain link Fisher index using predicted prices due to changes in the availability of products. When the availability of products increases (decreases) going from period t−1 to t, we expect $I_{KBF}{}^{t*}$ to be less (greater) than one and 1 − $I_{KBF}{}^{t}$ is an estimate of the percentage decrease (increase) in the cost of living due to the increased (decreased) availability of products. As was the case with $I_{KBF}{}^{t}$, if the availability of products is constant over periods t −1 and t, then $I_{KBF}{}^{t*}$ will be equal to 1. Thus the periods where $I_{KBF}{}^{t*}$ differs from 1 in our data set are again periods 9, 10, 11, 20 and 23. The values for $I_{KBF}{}^{t*}$ for these periods are listed in Table 9 above.

Again, we expected $I_{KBF}{}^{t*}$ to be less than 1 for periods 9, 11 and 23 when product availability increased and to be greater than 1 for periods 10 and 20 when product availability decreased. Our expectations were realized; there were no anomalous results for the 5 periods. However, the product of the 5 nonunitary values for $I_{KBF}{}^{t*}$ was 1.00088 (see the last row and column of Table 9) and so the overall increase in the availability of products led to a tiny *increase* in the cost of living over the sample period equal to 0.088 percentage points, rather than a *decrease* as was expected. Since our estimated product prices are not entirely reliable, this type of anomalous result can occur.

It is useful to compare our present estimates for the net benefits of increasing product availability (equal to a *net increase* in the true cost of living of either 0.176 or 0.088 percentage points) to our earlier net benefits that were generated by estimating the CES unit cost (see Table 2) and the CES utility function (see Table 6). Using the estimates from Tables 2 and 6, we found that the overall *net reduction* in the true cost of living using the CES functional form was about 0.93 for the unit cost function approach and 0.67 to 0.79 percentage points for the utility function approach.[75]

Two important tentative conclusions can be drawn from the results in this section:

- The CES methodology that is widely used to estimate the gains from increased product availability *likely* overstates the benefits from such increases.

[75] Since the Model 4 and 15 estimated σ coefficients are virtually identical, the net gains from increased product variety will be similar. Using the Model 2 estimates for σ, the estimated net reduction in the true cost of living index over the sample period due to increased product availability was approximately 1.64 percentage points. However Model 2 used expenditure share equations to estimate the CES unit cost function and we found that this model did not fit the data nearly as well as Models 4 and 15 so the resulting estimated gains from increasing product availability are not as plausible as the Model 4 and 15 results.

- The chain drift problem associated with the use of the chained Sato Vartia index in the Feenstra methodology can be substantial and in our example, was much more important than the gains and losses due to changes in product availability. All of the chained indexes that we evaluated in the course of our calculations that used actual price and quantity data were subject to substantial chain drift.

In the following section, we will develop an alternative methodology for estimating the gains and losses from changes in product availability that is based on the economic approach to index number theory. This approach utilizes the estimated well behaved utility function so it has the drawback of being very much dependent on the econometric estimation of the utility function. It has the advantage of being a much more transparent approach that is anomaly free.

11. The Gains and Losses Due to Changes in Product Availability Revisited

In this section, we consider a somewhat more general framework for measuring the gains or losses in utility due to changes in the availability of products. We suppose that we have data on prices and quantities on the sales of N products for T periods. The vectors of observed period t prices and quantities sold are $p^t \equiv [p_1^t,...,p_N^t] > 0_N$ and $q^t \equiv [q_1^t,...,q_N^t] > 0_N$ respectively for t = 1,...,T. Sales or expenditures on the N products during period t are $e^t \equiv p^t \cdot q^t = \Sigma_{n=1}^N p_n^t q_n^t > 0$ for t = 1,...,T.[76] We assume that a linearly homogeneous utility function, $f(q_1,...,q_N) = f(q)$, has been estimated where $q \geq 0_N$.[77] If product n is not available (or not sold) during period t, we assume that the corresponding observed price and quantity, p_n^t and q_n^t, are set equal to zeros.

We calculate *reservation prices* for the unavailable products. We also need to form *predicted prices* for the available commodities, where the predicted prices are consistent with our econometrically estimated utility function and the observed quantity data, q^t. The period t *reservation or predicted price* for product n, p_n^{t*}, is defined as follows, using the observed period t expenditure, e^t, the observed period t quantity vector q^t and the partial derivatives of the estimated utility function f(q) as follows:

$$(98)\quad p_n^{t*} \equiv e^t[\partial f(q^t)/\partial q_n]/f(q^t) ; \qquad n = 1,...,N;\ t = 1,...,T.$$

The prices defined by (98) are also Rothbarth’s (1941) *virtual prices*; they are the prices which rationalize the observed period t quantity vector as a solution to the period t utility maximization problem. Since f(q) is nondecreasing in its arguments and $e^t > 0$, we see that $p_n^{t*} \geq 0$ for all n and t.[78] If the estimated utility function fits the observed data exactly

[76] We also assume that $\Sigma_{n=2}^N p_n^t q_n^t > 0$ for t = 1,...,T.
[77] We assume that f(q) is a differentiable, positive, linearly homogeneous, nondecreasing and concave function of q over a cone contained in the positive orthant. The domain of definition of the function f is extended to the closure of this cone by continuity and we assume that observed quantity vectors q^t are contained in the closure of this cone.
[78] We also assume that $f(q^t) > 0$.

(so that all errors in the estimating equations are equal to 0),[79] then the predicted prices, p_n^{t*}, for the available products will be equal to the corresponding actual prices, p_n^t.

Imputed expenditures on product n during period t are defined as $p_n^{t*}q_n^t$ for n = 1,...,N. Note that if product n is not sold during period t, $q_n^t = 0$ and hence $p_n^{t*}q_n^t = 0$ as well. *Total imputed expenditures* for all products sold during period t, e^{t*}, are defined as the sum of the individual product imputed expenditures:

$$
\begin{aligned}
(99)\ e^{t*} &\equiv \Sigma_{n=1}^N p_n^{t*}q_n^t\ ; && t = 1,...,T \\
&= \Sigma_{n=1}^N q_n^t e^t[\partial f(q^t)/\partial q_n]/f(q^t) && \text{using definitions (98)} \\
&= e^t
\end{aligned}
$$

where the last equality follows using the linear homogeneity of f(q) since by Euler's Theorem on homogeneous functions, we have $f(q) = \Sigma_{n=1}^N q_n\ \partial f(q)/\partial q_n$. Thus period t imputed expenditures, e^{t*}, are equal to period t actual expenditures, e^t.

The above material sets the stage for the main acts: namely how to measure the welfare gain if product availability increases and how to measure the welfare loss if product availability decreases.

Suppose that in period t−1, product 1 was not available (so that $q_1^{t-1} = 0$) , but in period t, it becomes available and a positive amount is purchased (so that $q_1^t > 0$). Our task is to define a measure of the increase in purchaser welfare that can be attributed to the increase in commodity availability.

Define the vector of purchases of products during period t excluding purchases of product 1 as $q_{\sim 1}^t \equiv [q_2^t,q_3^t,...,q_N^t]$. Thus $q^t = [q_1^t,q_{\sim 1}^t]$. Since by assumption, an estimated utility function f(q) is available, we can use this utility function in order to define the *aggregate level of purchaser utility during period t*, u^t, as follows:

$$(100)\ u^t \equiv f(q^t) = f(q_1^t,q_{\sim 1}^t).$$

Now exclude the purchases of product 1 and define the (diminished) utility, $u_{\sim 1}^t$, the utility generated by the remaining vector of purchases, $q_{\sim 1}^t$, as follows:

$$
\begin{aligned}
(101)\ u_{\sim 1}^t &\equiv f(0,q_{\sim 1}^t) \\
&\leq f(q_1^t,q_{\sim 1}^t) && \text{since f (q) is nondecreasing in the components of q} \\
&= u^t && \text{using definition (90).}
\end{aligned}
$$

Define the *period t imputed expenditures on products excluding product 1*, $e_{\sim 1}^{t*}$, as follows:

$$
\begin{aligned}
(102)\ e_{\sim 1}^{t*} &\equiv \Sigma_{n=2}^N p_n^{t*}q_n^t \\
&= e^t - p_1^{t*}q_1^t && \text{using (99)}
\end{aligned}
$$

[79] This assumes that observed prices are the dependent variables in the estimating equations.

$\leq e^t$ since $p_1^{t*} \geq 0$ and $q_1^t > 0$.

Define the ratio of e^t to $e_{\sim 1}^{t*}$ as follows:

$$(103)\ \lambda_1 \equiv e^t/e_{\sim 1}^{t*}$$
$$\geq 1 \qquad \text{using (102) and } e_{\sim 1}^t > 0.$$

Multiply the vector of period t purchases excluding product 1, $q_{\sim 1}^t$, by the scalar λ_1 and calculate the resulting imputed expenditures on the vector $\lambda_1 q_{\sim 1}^t$:

$$(104)\ \Sigma_{n=2}^N\ p_n^{t*}(\lambda_1 q_n^t) = \lambda_1 \Sigma_{n=2}^N\ p_n^{t*} q_n^t$$

$$(105)\ \Sigma_{n=2}^N\ p_n^{t*}(\lambda_1 q_n^t) = \lambda_1 \Sigma_{n=2}^N\ p_n^{t*} q_n^t$$
$$= \lambda_1\ e_{\sim 1}^t \qquad \text{using definition (102)}$$
$$= [e^t/e_{\sim 1}^{t*}]e_{\sim 1}^t \qquad \text{using definition (103)}$$
$$= e^t.$$

Using the linear homogeneity of f(q) in the components of q, we are able to calculate the utility level, u_{A1}^t, that is generated by the vector $\lambda_1 q_{\sim 1}^t$ as follows:

$$(106)\ u_{A1}^t \equiv f(0,\lambda_1 q_{\sim 1}^t)$$
$$= \lambda_1 f(0,q_{\sim 1}^t) \qquad \text{using the linear homogeneity of f}$$
$$= \lambda_1 u_{\sim 1}^t \qquad \text{using definition (101).}$$

Note that λ_1 can be calculated using definition (103) and $u_{\sim 1}^t$ can be calculated using definition (101). Thus u_{A1}^t can also be readily calculated.

Consider the following (hypothetical) purchaser's period t aggregate *utility maximization problem where product 1 is not available* and purchasers face the imputed prices p_n^{t*} for products 2,...,N and the maximum expenditure on the N−1 products is restricted to be equal to or less than actual expenditures on all N products during period t, which is e^t:

$$(107)\ \max_{q\text{'s}}\ \{f(0,q_2,q_3,...,q_N) : \Sigma_{n=2}^N\ p_n^{t*} q_n \leq e^t\} \equiv u_1^t$$
$$\geq u_{A1}^t$$

where u_{A1}^t is defined by (103). The inequality in (107) follows because (104) shows that $\lambda_1 q_{\sim 1}^t$ is a feasible solution for the utility maximization problem defined by (107).

Now consider the following *period t unconstrained utility maximization problem* using imputed prices and actual expenditure e^t:

$$(108)\ \max_{q\text{'s}}\ \{f(q_1,q_2,q_3,...,q_N) : \Sigma_{n=1}^N\ p_n^{t*} q_n \leq e^t\}.$$

The first order necessary conditions[80] for the observed period t quantity vector q^t to solve (108) are as follows:

$$(109)\ \nabla f(q^t) = \lambda^* p^{t*} ;$$
$$(110)\ p^{t*}\cdot q^t = e^t$$

where $\nabla f(q^t)$ is the vector of first order partial derivatives of f evaluated at q^t and λ^* is the optimal Lagrange multiplier. Take the inner product of both sides of (109) with q^t and solve the resulting equation for $\lambda^* = q^t\cdot\nabla f(q^t)/p^{t*}\cdot q^t = q^t\cdot\nabla f(q^t)/e^t$ where we have used (99), which also shows that q^t satisfies the constraint (110). Euler's Theorem on homogeneous functions implies that $q^t\cdot\nabla f(q^t) = f(q^t)$ and so $\lambda^* = f(q^t)/e^t$. Replace λ^* in equations (109) by $f(q^t)/e^t$ and we find that the resulting equations are equivalent to equations (98). Thus q^t solves (108) and we have the following results:

$$\begin{aligned}(111)\ f(q^t) &= \max_{q\text{'s}} \{f(q_1,q_2,q_3,\ldots,q_N) : \Sigma_{n=1}^N p_n^{t*}q_n \le e^t\} \\ &= u^t \\ &\ge u_1^t\end{aligned}$$

where u_1^t is the optimal level of utility that is generated by a solution to the constrained period t utility maximization problem defined by (107). The inequality in (111) follows since any optimal solution for (107) is only a feasible solution for the unconstrained utility maximization problem defined by (108). The inequalities (107) and (111) imply the following inequalities:

$$(112)\ u^t \ge u_1^t \ge u_{A1}^t.$$

We regard u_{A1}^t as an approximation to u_1^t (and it is also a lower bound for u_1^t). Given that an estimated utility function f(q) is on hand, it is easy to compute the approximate utility level u_{A1}^t when product one is not available. The actual constrained utility level, u_1^t, will in general involve solving numerically the nonlinear programming problem defined by (107). For the KBF functional form, instead of maximizing $(q^TAq)^{1/2}$, we could maximize its square, q^TAq, and thus solving (107) would be equivalent to solving a quadratic programming problem with a single linear constraint. For the CES functional form, it turns out that there is no need to solve (107) since the strong separability of the CES functional form will imply that $u_1^t = u_{A1}^t$ and the latter utility level can be readily calculated.

A reasonable measure of the gain in utility due to the new availability of product 1 in period t, G_1^t, is the ratio of the completely unconstrained level of utility u^t to the product 1 constrained level u_1^t; i.e., define *the product 1 utility gain for period t* as

$$(113)\ G_1^t \equiv u^t/u_1^t \ge 1$$

[80] Since f(q) is a concave function of q over the feasible region, these conditions are also sufficient.

where the inequality follows from (111). The corresponding *product 1 approximate utility gain* is defined as:

(114) $G_{A1}{}^t \equiv u^t/u_{A1}{}^t \geq G_1{}^t \geq 1$

where the inequalities in (114) follow from the inequalities in (112). Thus in general, the approximate gain is an upper bound to the true gain $G_1{}^t$ in utility that is due to the new availability of product 1 in period t.

Now consider the case where product 1 is available in period t but it becomes unavailable in period t+1. In this case, we want to calculate an approximation to the loss of utility in period t+1 due to the unavailability of product 1 in period t+1. However, it turns out that our methodology will not provide an answer to this measurement problem using the price and quantity data for period t+1: we have to approximate the loss of utility that will occur in period t due to the unavailability of product 1 in period t+1 by looking at the loss of utility which would occur in period t if product 1 became unavailable. Once we redefine our measurement problem in this way, we can simply adapt the inequalities that we have already established for period t utility to the *loss* of utility from the unavailability of product 1 from the previous analysis for the *gain* in utility.

A reasonable measure of the hypothetical loss of utility due to the unavailability of product 1 in period t, $L_1{}^t$, is the ratio of the product 1 constrained level of utility $u_1{}^t$ to the completely unconstrained level of utility u^t to the product 1. We apply this hypothetical loss measure to period t+1 when product 1 becomes unavailable; i.e., define *the product 1 utility loss that can be attributed to the disappearance of product 1 in period t+1* as

(115) $L_1{}^{t+1} \equiv u_1{}^t/u^t \leq 1$

where the inequality follows from (111). The corresponding *product 1 approximate utility loss* is defined as:

(116) $L_{A1}{}^{t+1} \equiv u_{A1}{}^t/u^t \leq L_1{}^{t+1} \leq 1$

where the inequalities in (116) follow from the inequalities in (112). Thus in general, the approximate loss is an lower bound to the "true" loss $L_1{}^{t+1}$ in utility that can be attributed to the disappearance of product 1 in period t+1. As was the case with our approximate gain measure, if f(q) is a CES utility function, then $L_{A1}{}^t = L_1{}^t$.

If f(q) is the linear utility function that we estimated in Model 3 above, then it can be shown that all of the above gain and loss measures are equal to unity; i.e., there are no utility gains and losses from changes in product availability because each product is a perfect substitute for every other product. Thus the closer f(q) is to a linear function, the smaller will be the gains and losses due to changes in product availability.

It is straightforward to adapt the above analysis from product 1 to product 12 and to compute the approximate gains and losses in utility that occur due to the disappearance of product 12 in period 10, its reappearance in period 11, its disappearance in period 20 and its final reappearance in period 23. These approximate losses and gains are denoted by L_{A12}^{10}, G_{A12}^{11}, L_{A12}^{20} and G_{A12}^{23} and are listed in Table 10 for both our final KBF Model 14 and for our CES Model 15. It is also straightforward to adapt the above analysis to situations where two new products appear in a period, which is the case for our products 2 and 4 which were missing in periods 1-8 and make their appearance in period 9. The approximate utility gain due to the new availability of these products is denoted by $G_{A2,4}^{9}$ and this measure is also listed in Table 10 using the estimated utility functions for both our final KBF and CES models. We also list the product of these five approximate gain and loss estimates for both models in Table 10.

Table 10: Gains and Losses of Utility that can be Attributed to Changes In Product Availability Holding Expenditure Constant

	KBF	CES
$G_{A2,4}^{9}$	1.00127	1.00746
L_{A12}^{10}	0.99748	0.99512
G_{A12}^{11}	1.00304	1.00529
L_{A12}^{20}	0.99881	0.99644
G_{A12}^{23}	1.00078	1.00296
Product	1.00138	1.00724

The CES model implies that the net effect of changes in product availability is to increase purchasers' utility by approximately 0.724 percentage points while the KBF model implies a much smaller increase of 0.138 percentage points.[81] This is only one set of experimental calculations but the above results indicate that the net gains in utility predicted for increases in the availability of products by the CES model can substantially overstate the benefits of increased product variety. The results in the present section reinforce the results that we obtained in the previous section; i.e., the Feenstra methodology tends to overstate the benefits from increased product variety.

We conclude this section with a brief discussion of Hausman's (2003; 40) perfectly valid cost (or expenditure) function approach to the estimation of reservation prices[82] and we explain why we did not use it in the present study.

[81] Recall that the estimated reduction in the true cost of living that was generated by the CES Model 4 due to increased product availability was 0.79 percentage points which is approximately equal to a utility increase of 0.79 percentage points. This agrees nicely with our present estimate of a 0.72 percentage point utility increase.

[82] Hausman (1996; 217) (1999; 190) and Hausman and Leonard (2002; 248) for expositions and applications of his cost function methodology. Note that he did not assume homotheticity so his cost function framework was more general than the unit cost function approach that we are using. We believe that the assumption of homothetic preferences which can be represented by a linearly homogeneous utility function is an appropriate one for a statistical agency since the resulting price levels are independent of the levels of demand, which is a very useful property for macroeconomic applications of the resulting price indexes.

Instead of attempting to estimate a direct utility function, we could attempt to estimate a more general unit cost function than the CES unit cost function. Denote the more general unit cost function as $c(p)$ where $p \equiv [p_1,p_2...,p_N] \equiv [p_1,p_{\sim 1}]$ where $p_{\sim 1}$ is the set of prices excluding the price of product 1. Assuming that $c(p)$ is positive, nondecreasing, linearly homogeneous and concave over the positive orthant[83] and assuming all products are present in period t, the estimating equations for period t are the following ones:

$$(117)\ q_n^t = c_n(p^t)e^t/c(p^t) + \varepsilon_n^t\ ; \qquad n = 1,...,N$$

where q^t and p^t are the observed quantity and price vectors for period t, e^t is total expenditure on the N commodities during the period and $c_n(p^t) \equiv \partial c(p^t)/\partial p_n$ for $n = 1,...,N$. Now suppose product 1 is not available during period t. Then the N period t estimating equations are replaced by the following N equations:

$$(118)\ q_n^t = c_n(p_1^{t*},p_{\sim 1}^t)e^t/c(p_1^{t*},p_{\sim 1}^t) + \varepsilon_n^t\ ; \qquad n = 1,...,N$$

where $q_1^t = 0$ and p_1^{t*} is the reservation price that will drive demand for product 1 down to 0 in period t. It can be seen that p_1^{t*} is effectively an *extra unknown parameter* which must be estimated along with the other parameters in the unit cost function $c(p)$. Typically, the resulting estimating equations become very nonlinear and difficult to estimate and so it becomes necessary (as a practical matter) to drop *all* N estimating equations defined by (118). Thus the econometrician is reduced to using the estimating equations for periods where *all products in the group of products are available*. In many situations, this will greatly reduce the available degrees of freedom and in some cases, lead to no degrees of freedom at all if every period has a missing product. Contrast this situation with the methodology that we have used for Models 5-15: we only needed to drop the missing product estimating equations using our primal approach instead of having to drop all estimating equations for any period which had one or more missing products.[84]

In the following section, we turn to a discussion of another approach that Hausman took to generate estimates for the gains and losses due to changes in product availability.

11. Our Approximate Loss from Decreased Product Availability versus Hausman's Approximate Loss for the Case of Two Products

[83] We extend the domain of definition of $c(p)$ to the nonnegative orthant by continuity.

[84] There is another reason why we did not pursue Hausman's cost function methodology very far in this paper. The simplest unit cost function is a linear one but this corresponds to a zero elasticity of substitution model which as we have seen fits the data rather poorly in the present context where we expect closely related products to exhibit a considerable degree of substitutability. We could have generalized the linear unit cost function by assuming the KBF functional form for the unit cost function. But because the linear cost function fits the data so poorly, we suspect that a semiflexible KBF functional form would not fit the data as well as the KBF semiflexible functional form for the utility function. This utility functional form starts off with the perfect substitutes case which fits the data much better than the linear (no substitution at all) cost function.

Hausman (1999; 191) (2003; 27) presented a very simple and "easy" to implement methodology for calculating the approximate loss of consumer surplus due to the disappearance of a product. The framework is a partial equilibrium one where he drew an inverse demand curve for say product 1 as $p_1 = D_1(q_1)$ where q_1 is the quantity of product 1 purchased when its price is p_1. Hausman formed a first order Taylor series approximation to this demand curve around the point (p_1^*, q_1^*) which corresponds to a period when product 1 was available. He assumed that the demand curve is downward sloping and when $q_1 = 0$, the corresponding virtual demand price is p_1^{**}. The linear approximation to the actual inverse demand function goes through the p_1 axis at the point ρ_1^* where $\rho_1^* \equiv p_1^* + \alpha q_1^*$ and $\alpha \equiv - \partial D_1(q_1^*)/\partial q_1 > 0$ is the absolute value of the slope of the inverse demand curve evaluated at $q_1 = q_1^*$. Hausman took the area of the triangle below the linear approximation to the true inverse demand function but above the line $p_1 = p_1^*$ as his approximate measure of the loss in consumer surplus that would occur if product 1 were no longer available during the period under consideration. We scale the utility level $f(q_1^*,q_2^*)$ so that it equals expenditure e^* for the period. Thus we have:

$$(119)\quad f(q_1^*,q_2^*) = e^* \equiv p_1^*q_1^* + p_2^*q_2^*.$$

Define the Hausman approximate loss measure as a fraction of the period t expenditure e^* as follows:

$$\begin{aligned}(120)\quad L_H &\equiv - (1/2)(\rho_1^* - p_1^*)q_1^*/e^* \\ &= - (1/2)\alpha(q_1^*)^2/e^* \\ &= (1/2)s_1^*\eta\end{aligned}$$

where s_1^* is the share of product 1 in total expenditures, $p_1^*q_1^*/e^*$, and the *inverse elasticity of demand at the observed equilibrium point* is defined as

$$(121)\quad \eta \equiv [q_1^*/p_1^*]\partial D_1(q_1^*)/\partial q_1 = - [q_1^*/p_1^*]\alpha < 0.$$

When Hausman turned this approximate loss measure into an approximate gain measure due to increases in product availability, he found very large gains for his empirical examples. Note that large magnitude estimates for the inverse elasticity of demand η will translate into large losses of consumer surplus if product 1 is made unavailable.

We now adapt our loss model presented in the previous section to the case of only 2 commodities. We will derive first and second order Taylor series approximations to our loss measure and compare these approximations to the Hausman approximate loss measure defined by (120). We assume that the utility function $f(q_1,q_2)$ is twice continuously differentiable in this section.

We suppose that purchasers have maximized the utility function $f(q_1,q_2)$ in a period where they face prices $p_1^* > 0$ and $p_2^* > 0$ where f satisfies our usual regularity conditions plus differentiability. The optimal quantities are $q_1^* > 0$ and $q_2^* > 0$. These prices and quantities satisfy equations (119) and the optimality conditions (98) which we rewrite using our present notation as follows:

(122) $p_1^*f(q_1^*,q_2^*) = e^*f_1(q_1^*,q_2^*)$;
(123) $p_2^*f(q_1^*,q_2^*) = e^*f_2(q_1^*,q_2^*)$

where $e^* \equiv p_1^*q_1^* + p_2^*q_2^* > 0$ is observed expenditure in the period under consideration and $f_n(q_1^*,q_2^*) \equiv \partial f(q_1^*,q_2^*)/\partial q_n$ for n = 1,2. Using our utility scaling assumption (119), it can be seen that equations (122) and (123) simplify to $p_1^* = f_1(q_1^*,q_2^*)$ and $p_2^* = f_2(q_1^*,q_2^*)$. Now consider a model where we reduce purchases of q_1 down to 0. We do this in a linear fashion holding prices fixed at their initial levels, p_1^*, p_2^*. Thus we travel along the budget constraint until it intersects the q_2 axis. Hence q_2 is an endogenous variable; it is the following function of q_1 where q_1 starts at $q_1 = q_1^*$ and ends up at $q_1 = 0$:

(124) $q_2(q_1) \equiv [e^* - p_1^*q_1]/p_2^*$.

The derivative of $q_2(q_1)$ is $q_2'q_1) \equiv \partial q_2(q_1)/\partial q_1 = -(p_1^*/p_2^*)$, a fact which we will use later. Define utility as a function of q_1 for $0 \leq q_1 \leq q_1^*$, holding expenditures on the two commodities constant at e^*, as follows:

(125) $u(q_1) \equiv f(q_1,q_2(q_1))$.

We use the function $u(q_1)$ to measure the purchaser loss of utility as we move q_1 from its original equilibrium level of q_1^* to 0. Thus our *loss of utility due to the disappearance of product 1* as a fraction of optimal expenditure is defined as follows:

(126) $L \equiv [u(q_1^*) - u(0)]/e^*$.

Using our scaling of utility assumption (119), we can observe $u(q_1^*) = f(q_1^*,q_2^*) = e^*$. We approximate u(0) by a first order Taylor series approximation around the point q_1^*:

$$
\begin{aligned}
(127)\ u(0) &\approx u(q_1^*) + u'(q_1^*)(0 - q_1^*) \\
&= u(q_1^*) - q_1^*[f_1(q_1^*,q_2^*) + f_2(q_1^*,q_2^*)\partial q_2(q_1)/\partial q_1] && \text{differentiating (125)} \\
&= u(q_1^*) - q_1^*[f_1(q_1^*,q_2^*) + f_2(q_1^*,q_2^*)(-p_1^*/p_2^*)] && \text{differentiating (124)} \\
&= u(q_1^*) - q_1^*[p_1^* + p_2^*(-p_1^*/p_2^*)]f(q_1^*,q_2^*)/e^* && \text{using (122) and (123)} \\
&= u(q_1^*).
\end{aligned}
$$

Thus to the accuracy of a first order approximation to the true loss of utility that can be attributed to the disappearance of product 1, we have $L \approx 0$.

The second order derivative of $u(q_1)$ evaluated at q_1^* is given by the following expression:

$$
\begin{aligned}
(128)\ u''(q_1^*) &= f_{11}(q_1^*,q_2^*) + 2f_{12}(q_1^*,q_2^*)(-p_1^*/p_2^*) + f_{22}(q_1^*,q_2^*)(-p_1^*/p_2^*)^2 \\
&\leq 0
\end{aligned}
$$

where the inequality follows since the matrix of second order partial derivatives of $f(q_1^*,q_2^*)$ is negative semidefinite using the concavity of $f(q_1,q_2)$. Thus to the accuracy of a second order approximation to the true loss of utility that can be attributed to the disappearance of product 1, we have:

(129) $L \approx (1/2)\, u''(q_1^*)q_1^{*2}/e^* \le 0$.

If the underlying utility function is linear (so that all products are perfect substitutes), then it can be seen that there is no approximate loss of utility due to the disappearance of product 1 (since all of the second order partial derivatives of $f(q_1,q_2)$ are equal to 0 in this case). In the case of a linear utility function, there is no loss of utility if we take away the possibility of purchasing product 1 since the two products are perfect substitutes. In this case, the approximate loss of utility is equal to the actual loss of utility which in turn is equal to 0.

We can express the approximate loss defined by (129) in elasticity and share form if we make a few definitions. We know that $f_i(q_1,q_2) \equiv \partial f(q_1,q_2)/\partial q_i$ is the marginal utility of product i for i = 1,2. Thus $f_{ij}(q_1,q_2) \equiv \partial^2 f(q_1,q_2)/\partial q_i \partial q_j$ is the derivative of marginal utility i with respect to q_j. We can turn this second order partial derivative of the utility function into a *unit free elasticity* $\varepsilon_{ij}(q_1,q_2)$ by multiplying $f_{ij}(q_1,q_2)$ by $q_j/f_i(q_1,q_2)$:

(130) $\varepsilon_{ij}(q_1,q_2) \equiv [q_j/f_i(q_1,q_2)]f_{ij}(q_1,q_2)$; $\qquad$ i,j = 1,2.

We also need to make use of some identities that the second order partial derivatives of the linearly homogeneous utility function f satisfies. Using Euler's Theorem on homogeneous functions, the following two identities hold:

(131) $f_{11}(q_1^*,q_2^*)q_1^* + f_{12}(q_1^*,q_2^*)q_2^* = 0$;
(132) $f_{21}(q_1^*,q_2^*)q_1^* + f_{22}(q_1^*,q_2^*)q_2^* = 0$.

Young's Theorem from calculus also implies that $f_{12}(q_1^*,q_2^*) = f_{21}(q_1^*,q_2^*)$. Using this relationship along with (131) and (132) implies the following relationships between the second order partial derivatives of f:

(133) $f_{12}(q_1^*,q_2^*) = f_{21}(q_1^*,q_2^*) = f_{11}(q_1^*,q_2^*)(-q_1^*/q_2^*)$;
(134) $f_{22}(q_1^*,q_2^*) = f_{11}(q_1^*,q_2^*)(-q_1^*/q_2^*)^2$.

Now substitute (133) and (134) into (128) in order to obtain the following expression for $u''(q_1^*)$:

$$
\begin{aligned}
(135)\ u''(q_1^*) &= f_{11}(q_1^*,q_2^*) + 2f_{12}(q_1^*,q_2^*)(-p_1^*/p_2^*) + f_{22}(q_1^*,q_2^*)(-p_1^*/p_2^*)^2 \\
&= f_{11}(q_1^*,q_2^*)[1 + 2(p_1^*q_1^*/p_2^*q_2^*) + (p_1^*q_1^*/p_2^*q_2^*)^2] \\
&= f_{11}(q_1^*,q_2^*)[1 + (s_1^*/s_2^*)]^2
\end{aligned}
$$

where $s_n^* \equiv p_n^*q_n^*/e^*$ for n = 1,2. Since $f_{11}(q_1^*,q_2^*) \le 0$, $u''(q_1^*) \le 0$ as well. Using (130), we can write $f_{11}(q_1^*,q_2^*)$ in elasticity form as follows:

(136) $f_{11}(q_1^*,q_2^*) = \varepsilon_{11}(q_1^*,q_2^*)f_1(q_1^*,q_2^*)/q_1^*$

$= \varepsilon_{11}(q_1^*,q_2^*)p_1^*f(q_1^*,q_2^*)/q_1^*e^*$ using (122)
$= \varepsilon_{11}(q_1^*,q_2^*)p_1^*/q_1^*$ using (119).

Finally, substitute (135) and (136) into (129) and our second order approximation to the loss of utility due to the withdrawal of product 1 becomes the following expression:

(137) $L \approx (1/2)\varepsilon_{11}(q_1^*,q_2^*)s_1^*[1 + (s_1^*/s_2^*)]^2 \le 0$.

If q_1^* is small, then the above second order approximation to the loss of utility will be quite accurate. If $f_{11}(q_1^*,q_2^*) = 0$,[85] then the elasticity $\varepsilon_{11}(q_1^*,q_2^*)$ equals 0 as well and the approximate loss will be equal to 0. Formula (137) is our counterpart to Hausman's approximate loss function defined by (120).

We conclude this section by considering some alternative partial equilibrium models for the (inverse) demand function for product 1, $p_1 = D_1(q_1)$. We can then calculate the resulting partial derivative of this function at our observed equilibrium point, $\partial D_1(q_1^*)/\partial q_1$, and then evaluate how the approximate Hausman loss defined by (120) compares to our approximate loss defined by (137).

The two inverse demand functions that give us virtual (or equilibrium) prices as functions of quantities purchased and total expenditure on the two products e are the following functions:

(138) $p_1 = d_1(q_1,q_2,e) \equiv ef_1(q_1,q_2)/f(q_1,q_2)$;
(139) $p_2 = d_2(q_1,q_2,e) \equiv ef_2(q_1,q_2)/f(q_1,q_2)$.

We want the partial equilibrium function, $p_1 = D_1(q_1)$ holding other variables constant. But what exactly are these other variables that one should hold constant?

The simplest choice of variables to hold constant is to hold q_2 and e constant. In this case, $D_1(q_1) = d_1(q_1,q_2,e)$ where q_2 and e are held constant. In this case, $\partial D_1(q_1^*)/\partial q_1$ is equal to the following expression:

(140) $\partial D_1(q_1^*)/\partial q_1 = [e^*f_{11}(q_1^*,q_2^*)/f(q_1^*,q_2^*)] - [e^*f_1(q_1^*,q_2^*)^2/f(q_1^*,q_2^*)^2]$
$= f_{11}(q_1^*,q_2^*) - (p_1^*)^2/e^*$ using (119) and (122)
$= \varepsilon_{11}(q_1^*,q_2^*)(p_1^*/q_1^*) - (p_1^*)^2/e^*$ using (119) and (130).

Thus the elasticity η defined by (121) above becomes the following expression:

(141) $\eta \equiv [q_1^*/p_1^*]\partial D_1(q_1^*)/\partial q_1$
$= [q_1^*/p_1^*][\varepsilon_{11}(q_1^*,q_2^*)(p_1^*/q_1^*) - (p_1^*)^2/e^*]$ using (140)
$= \varepsilon_{11}(q_1^*,q_2^*) - (p_1^*q_1^*/e^*)$
$= \varepsilon_{11}(q_1^*,q_2^*) - s_1^*$.

[85] This condition means that the marginal utility of product 1 is constant as q_1 increases. It also means that locally, products 1 and 2 are perfect substitutes.

Since $\varepsilon_{11}(q_1^*,q_2^*) \leq 0$ and $s_1^* > 0$, we see that $\eta < 0$. Thus holding q_2 and e constant leads to the following Hausman type approximate loss due to the unavailability of product 1:

(142) $L_H \equiv (1/2)s_1^*\eta = (1/2)s_1^*[\varepsilon_{11}(q_1^*,q_2^*) - s_1^*] < 0.$

This measure of approximate loss will tend to be larger in magnitude than our measure of approximate loss defined by (137) if $\varepsilon_{11}(q_1^*,q_2^*)$ is close to 0.

However, holding q_2 and e constant is not what Hausman had in mind as constant variables. He worked in a cost function framework so specializing his more general framework to our homogeneous preferences model leads to a model where expenditure e is a function of prices and the utility level; i.e., $e = c(p_1,p_2)u$ where $c(p_1,p_2)$ is the unit cost function that is dual to the utility function $u = f(q_1,q_2)$. Thus we have the following equilibrium relationships in the period where both products are available:

(143) $e^* = c(p_1^*,p_2^*)u^*$; $q_1^* = c_1(p_1^*,p_2^*)u^*$; $q_2^* = c_2(p_1^*,p_2^*)u^*$

where e^* is total expenditure, $q_n^* > 0$ is optimal demand for product n for n = 1,2 and $c_n(p_1^*,p_2^*) \equiv \partial c(p_1^*,p_2^*)/\partial p_n$ for n = 1,2. Hausman holds utility constant and increases the price of product 1 to $p_1^{**} > p_1^*$ where p_1^{**} is the virtual price that drives the Hicksian demand for product 1 down to 0 so that $0 = c_1(p_1^{**},p_2^*)u^*$. The higher price of product 1 means that purchasers now have to spend $e^{**} \equiv c(p_1^{**},p_2^*)u^* > e^*$ to achieve the same utility level u^* that they attained before product 1 was withdrawn from the marketplace. Thus the Hausman exact loss in measured as the expenditure difference, $e^{**} - e^*$ whereas our exact loss concept was a utility difference.

The variables that Hausman holds constant are the utility level u and the price of product 2, p_2. Endogenous variables are q_1, q_2 and e while the driving variable is p_1 which goes from p_1^* to p_1^{**} while q_1 goes from q_1^* to 0. We can model his framework in our direct utility function model as follows: regard $u^* \equiv f(q_1^*,q_2^*)$ and p_2^* as fixed exogenous variables, p_1, q_2 and e as endogenous variables and q_1 as the driving exogenous variable. The constraint that utility remain constant as we decrease q_1 from q_1^* to 0 is the following one:

(144) $f(q_1,q_2(q_1)) = f(q_1^*,q_2^*) = e^*.$

Thus we again scale utility so that initial utility $f(q_1^*,q_2^*)$ is equal to initial expenditure, e^*. Define $q_2(q_1)$ as the implicit function which satisfies (144). The derivative of this implicit function is defined by differentiating $f(q_1,q_2(q_1)) = e^*$ with respect to q_1. Thus we find that:

(145) $q_2'(q_1^*) = - f_1(q_1^*,q_2^*)/f_2(q_1^*,q_2^*) = -p_1^*/p_2^*$

where the second equation in (145) follows from (144) and (138) and (139) (our two inverse demand functions) evaluated at the initial equilibrium. We take the second inverse demand function defined by (139) and set it equal to the constant, p_2^*. We solve the resulting equation for expenditure as a function of q_1, $e(q_1)$:

$$(146)\ e(q_1) \equiv p_2^* f(q_1,q_2(q_1))/f_2(q_1,q_2(q_1))$$
$$= p_2^* e^*/f_2(q_1,q_2(q_1)) \qquad \text{using (144).}$$

Differentiate (146) with respect to q_1 in order to determine the derivative $e'(q_1^*)$. We find that

$$(147)\ e'(q_1^*) = -[p_2^* e^*/p_2^{*2}][f_{21}(q_1^*,q_2^*) + f_{22}(q_1^*,q_2^*)q_2'(q_1^*)] \qquad \text{using (138)}$$
$$= -[e^*/p_2^*][f_{21}(q_1^*,q_2^*) + f_{22}(q_1^*,q_2^*)(-p_1^*/p_2^*)] \qquad \text{using (145).}$$

We can now define our Hausman partial equilibrium first (inverse) demand function $p_1 = D_1(q_1)$ by replacing q_2 and e in definition (138) by $q_2(q_1)$ and $e(q_1)$:

$$(148)\ D_1(q_1) \equiv e(q_1)f_1(q_1,q_2(q_1))/f(q_1,q_2(q_1))$$
$$= e(q_1)f_1(q_1,q_2(q_1))/e^* \qquad \text{using (144).}$$

Calculate the derivative of the partial equilibrium inverse demand function defined by (148) at q_1^*:

$$(149)\ \partial D_1(q_1^*)/\partial q_1 = -[p_1^*/e^*][e^*/p_2^*][f_{21}(q_1^*,q_2^*) + f_{22}(q_1^*,q_2^*)(-p_1^*/p_2^*)]$$
$$+ [e(q_1^*)/e^*][f_{11}(q_1^*,q_2^*) + f_{12}(q_1^*,q_2^*)q_2'(q_1^*)] \qquad \text{using (147)}$$
$$= [f_{21}(q_1^*,q_2^*)(-p_1^*/p_2^*) + f_{22}(q_1^*,q_2^*)(-p_1^*/p_2^*)^2] + [f_{11}(q_1^*,q_2^*) + f_{12}(q_1^*,q_2^*)q_2'(q_1^*)]$$
$$= f_{11}(q_1^*,q_2^*) + 2f_{12}(q_1^*,q_2^*)(-p_1^*/p_2^*) + f_{22}(q_1^*,q_2^*)(-p_1^*/p_2^*)^2$$
$$= u''(q_1^*) \qquad \text{where } u''(q_1^*) \text{ was defined by (128)}$$
$$= f_{11}(q_1^*,q_2^*)[1 + (s_1^*/s_2^*)]^2 \qquad \text{using (135).}$$

Thus the Hausman approximate loss for this partial equilibrium demand derivative defined by (149) turns out to be:

$$(150)\ L_H \equiv (1/2)q_1^*[\partial D_1(q_1^*)/\partial q_1]/e^*$$
$$= (1/2)q_1^* f_{11}(q_1^*,q_2^*)[1 + (s_1^*/s_2^*)]^2/e^* \qquad \text{using (149)}$$
$$= (1/2)\, s_1^* \varepsilon_{11}(q_1^*,q_2^*)[1 + (s_1^*/s_2^*)]^2$$

where the elasticity marginal utility elasticity $\varepsilon_{11}(q_1^*,q_2^*)$ is defined as $(q_1^*/p_1^*)f_{11}(q_1^*,q_2^*)$. This is a rather surprising result: Hausman's first order triangle consumer surplus approximate approach to measuring the loss due to the withdrawal of a product turns out to be exactly equal to our second order approximation loss of utility approach when there are only 2 products!

12. Conclusion

There are many tentative conclusions that can be drawn from the computations undertaken in this paper:

- The Feenstra CES methodology for adjusting maximum overlap chained price indexes for changes in product availability is very much dependent on having accurate estimates for the elasticity of substitution. The gains from increasing product availability are very large if the elasticity of substitution σ is close to one and fall rapidly as the elasticity increases.
- It is not a trivial matter to obtain an accurate estimate for σ. When applying traditional consumer demand theory to actual data, it is commonplace to have expenditure shares as the dependent variables and product prices as the independent variables. When this framework was applied to our grocery store data set using the CES functional form for the unit cost function, we found that the equation by equation fit was poor. An alternative econometric specification was used to estimate a CES utility function where sales shares are functions of quantities in this specification. We found that this specification fit the data much better and the resulting estimate for σ was much larger than the corresponding estimate for σ when we used the CES unit cost function specification.
- A major purpose of the present paper was the estimation of Hicksian reservation prices for products that were not available in a period. In the CES framework, these reservation prices turn out to be infinite. But typically, it does not require an infinite reservation price to deter a consumer from purchasing a product. Thus we estimated the utility function $f(q) \equiv (q^TAq)^{1/2}$, which was originally introduced by Konüs and Byushgens (1926). They showed that this functional form was exactly consistent with the use of Fisher (1922) price and quantity indexes so we called this functional form the KBF functional form. The use of this functional form leads to finite reservation prices, which can be readily calculated once the utility function has been estimated.
- We indicated how the correct curvature conditions on this functional form could be imposed and we showed that this functional form is a semiflexible functional form which is similar to the normalized quadratic semiflexible functional form introduced by Diewert and Wales (1987) (1988).
- We initially estimated the KBF functional form using expenditure shares as dependent variables and quantities as the conditioning variables. We used the usual systems approach to the estimation of a system of inverse demand equations. However, we found that existing algorithms for the nonlinear systems of equations bogged down using this approach because the approach requires the estimation of the elements of a symmetric variance-covariance matrix plus the elements of the symmetric matrix A.
- Thus we stacked the estimating equations into a single (big) equation and estimated the unknown parameters in the A matrix using sales shares as the dependent variables using a semiflexible approach. This approach required the estimation of only one variance parameter.[86]
- The one big equation semiflexible approach worked in a satisfactory manner. This approach also allowed us to drop the observations that correspond to the

[86] Of course, this approach has the disadvantage of not accounting adequately for heteroskedasticity and possible correlation between the various product equation error terms.

unavailable products. We ended up getting useful estimates for the parameters in the A matrix.

- However, when we used our estimated utility function to construct fitted prices for the available products (and estimated reservation prices for the unavailable products), we found that the fitted prices were not nearly as close to the actual prices as were the fitted sales shares to the actual sales shares. This was an unsatisfactory development since if the fitted prices are not close to the actual prices for products that are present, it is unlikely that the reservation prices for unavailable products would be close to the "true" reservation prices.
- Thus in section 10 above, we switched from the one big equation approach that had shares as dependent variables to a one big equation approach that had actual prices as the dependent variables. Models 14 and 15 estimated the KBF and CES utility functions using this alternative approach.
- It turned out that the one big equation Model 15 generated almost the same estimate for the elasticity of substitution for a CES utility function that was generated by the systems approach to CES utility function estimation, which was Model 4. This was encouraging. These two CES utility function models generated implicit price indexes, $P_{UCES}{}^{t}$ for period t using Model 4 and $P_{CESN}{}^{t}$ for period t using Model 15. These two CES price indexes are plotted on Chart 9. They do not differ by all that much.
- The KBF utility function estimated by Model 14 generated an implicit price index, $P_{KBF}{}^{t}$ for period t and this index is also plotted on Chart 9. $P_{KBF}{}^{t}$ is quite close to its CES counterpart implicit (econometrically based) price index, $P_{CESN}{}^{t}$. This shows that the CES utility function model using prices as the dependent variables generates price and quantity indexes which are fairly close to the much more complicated KBF price and quantity indexes, which is also an encouraging result.
- However, the results presented in sections 10 and 11 indicate that the Feenstra CES methodology for measuring the benefits of increases in product variety may substantially overstate these benefits as compared to our semiflexible methodology.
- Another major conclusion that follows from our analysis is that the chain drift problem that arises in the scanner data context is perhaps a much bigger problem than adjusting price indexes for changes in product variety.[87] Our estimated adjustments for changes in product variety were rather small as compared to the large amount of chain drift we found in all of our chained indexes that used actual price and quantity data.[88]
- In section 11, we developed a utility function based methodology for measuring the net gains from net increases in product availability that is a counterpart to Hausman's expenditure or cost function based methodology.
- In section 12, we restricted our model to the two product case and approximated our utility based measure of the gains from increased product availability by a second order Taylor series approximation. We then compared our approximate

[87] Thus Keynes (1930; 106) was right to worry about the use of chained indexes generating chain drift.
[88] See the Australian Bureau of Statistics (2016) and Diewert and Fox (2017) for a review of the use of multilateral methods that could be used to control the chain drift problem. These papers did not address the issues raised by changes in product availability which is the focus of the present paper.

measure to the approximate consumer surplus (or expenditure function) based Hausman model of the gains from increased product availability and found that our approximate measure coincided with his approximate measure in the 2 product case. Whether this equality persists in the N product case is an open question.

Appendix A: The Frozen Juice Data

Here is a listing of the "monthly" quantities sold of 19 varieties of frozen juice (mostly orange juice) from Dominick's Store 5 in the Greater Chicago area, where a "month" consists of sales for 4 consecutive weeks.

Table A1: "Monthly" Quantities Sold for 19 Frozen OJ Products

Month t	q_1^t	q_2^t	q_3^t	q_4^t	q_5^t	q_6^t	q_7^t	q_8^t	q_9^t
1	142	0	66	0	369	85	108	163	90
2	330	0	299	0	1612	223	300	211	171
3	453	0	140	0	675	206	230	250	158
4	132	0	461	0	1812	210	430	285	194
5	87	0	107	0	490	210	158	256	159
6	679	0	105	0	655	163	182	250	170
7	53	0	260	0	793	178	232	287	135
8	141	0	100	0	343	117	115	174	154
9	442	123	191	108	633	153	145	168	265
10	524	239	204	125	544	129	184	320	390
11	34	19	204	179	821	131	225	427	1014
12	52	32	79	85	243	117	89	209	336
13	561	247	124	172	698	139	200	340	744
14	515	266	206	187	660	120	188	144	153
15	87	56	131	161	240	109	144	141	93
16	325	111	130	195	372	151	169	176	105
17	444	154	294	331	1127	146	271	219	127
18	588	175	203	229	569	159	165	250	133
19	476	264	122	156	175	130	131	282	85
20	830	276	198	181	669	132	149	205	309
21	614	208	166	156	309	115	165	141	186
22	764	403	172	165	873	94	240	206	585
23	589	55	144	163	581	118	181	204	1010
24	988	467	81	122	178	81	128	315	632
25	593	236	230	184	1039	111	215	240	935
26	55	42	296	313	1484	81	465	413	619
27	402	273	113	121	199	114	127	129	849
28	307	81	390	236	976	107	359	357	95
29	57	96	157	168	771	105	262	85	116
30	426	289	188	191	755	121	181	121	211
31	56	70	399	246	783	116	387	147	105
32	612	487	110	94	222	109	130	129	118
33	40	42	552	470	1114	114	574	150	120
34	342	253	177	265	424	98	235	139	157

35	224	132	185	230	437	84	211	160	413
36	78	51	152	214	557	97	231	395	637
37	345	189	161	130	395	95	173	146	528
38	76	22	155	237	355	113	172	121	246
39	89	80	363	242	921	111	363	185	231

Month t	q_{10}^t	q_{11}^t	q_{12}^t	q_{13}^t	q_{14}^t	q_{15}^t	q_{16}^t	q_{17}^t	q_{18}^t	q_{19}^t
1	45	174	109	2581	233	132	126	107	50	205
2	109	351	239	983	405	452	1060	207	198	149
3	118	325	303	1559	629	442	343	199	123	313
4	143	263	322	1638	647	412	1285	195	324	75
5	121	514	210	3552	460	265	769	175	471	1130
6	89	424	206	865	482	314	1001	113	279	652
7	93	531	232	981	495	280	2466	206	976	59
8	108	307	201	1752	366	201	932	109	362	503
9	185	376	189	2035	366	233	170	103	98	658
10	346	381	0	694	399	290	764	81	236	760
11	811	286	210	1531	363	273	201	98	81	598
12	252	511	112	4054	292	295	626	138	171	297
13	180	569	392	1330	296	277	145	181	98	268
14	113	424	187	786	367	317	414	93	172	535
15	99	388	186	2828	242	242	755	109	226	323
16	68	259	299	1981	392	263	708	177	124	344
17	58	271	305	888	478	306	750	169	191	54
18	60	245	303	2217	403	681	1216	97	259	61
19	52	360	155	2266	309	190	1588	113	424	473
20	274	232	0	1983	320	214	183	181	105	323
21	154	1027	0	2152	328	190	720	122	245	49
22	402	539	0	1514	242	155	1280	95	394	23
23	841	309	109	1216	271	145	1186	94	170	94
24	531	272	126	1379	288	143	558	112	208	66
25	607	290	127	3240	254	125	153	77	53	634
26	549	314	138	1227	235	128	758	81	354	40
27	236	391	162	2626	334	155	483	130	437	118
28	75	265	164	681	361	135	1158	83	628	562
29	94	329	163	1620	362	159	1030	97	483	608
30	107	436	185	546	395	154	1161	144	672	1210
31	72	494	205	1408	368	142	1195	129	701	314
32	79	482	156	490	318	2522	1208	100	870	337
33	59	436	169	1265	300	103	401	61	267	151
34	96	391	171	2112	353	100	546	85	323	112
35	354	389	175	715	343	83	2342	117	941	346
36	541	406	141	2523	344	85	340	83	314	155
37	498	283	109	684	177	64	91	33	107	169
38	151	305	151	366	259	89	396	94	203	415
39	237	321	118	1392	218	118	515	100	353	67

It can be seen that there were no sales of Products 2 and 4 for months 1-8 and there were no sales of Product 12 in month 10 and in months 20-22. Thus there is a new and disappearing product problem for 20 observations in this data set.

The corresponding monthly unit value prices for the 19 products are listed in Table A2.

Table A2: "Monthly" Unit Value Prices for 19 Frozen OJ Products

Month t	p_1^t	p_2^t	p_3^t	p_4^t	p_5^t	p_6^t	p_7^t	p_8^t	p_9^t
1	1.4700	*1.7413*	1.7718	*1.7831*	1.7618	2.3500	1.7715	0.9624	0.7553
2	1.4242	*1.5338*	1.3967	*1.5378*	1.4148	2.3500	1.5460	1.0900	0.8300
3	1.4463	*1.5433*	1.5521	*1.7782*	1.5734	2.3000	1.6413	1.0900	0.5856
4	1.5200	*1.5476*	1.3753	*1.3872*	1.4004	2.3000	1.3793	1.0623	0.6701
5	1.5200	*1.5688*	1.6900	*1.6933*	1.6900	2.2929	1.6900	1.0900	0.6208
6	1.4457	*1.3659*	1.8854	*1.8155*	1.8821	2.5895	1.8761	1.0900	0.5900
7	1.9753	*1.7326*	1.8546	*1.9018*	1.8793	2.7500	1.8332	1.0140	0.8300
8	1.7040	*1.9262*	2.0900	*2.1594*	2.0900	2.7415	1.9600	1.0778	0.8300
9	1.6299	1.9900	1.8575	1.9085	1.8195	2.7437	1.9315	1.0796	0.8089
10	1.5505	1.5615	1.8410	1.8980	1.8253	2.7500	1.8987	0.9469	0.8148
11	1.9900	1.9900	1.6763	1.6420	1.6169	2.7500	1.6402	0.9549	0.7061
12	1.9900	1.9900	2.0900	2.0900	2.0900	2.7500	2.0900	0.9828	0.9509
13	1.3649	1.3977	1.8682	1.7993	1.7476	2.7500	1.7625	0.8900	0.5866
14	1.4506	1.5073	1.6992	1.7691	1.7120	2.6200	1.7389	1.0900	0.9600
15	1.9900	1.9900	1.7648	1.7186	1.7317	2.4900	1.7706	1.0609	0.9600
16	1.4712	1.4224	1.6305	1.6483	1.6498	2.4900	1.6578	1.0139	0.9600
17	1.2599	1.2559	1.3500	1.3618	1.3264	2.2600	1.3626	0.9900	0.8053
18	1.0567	1.0936	1.4213	1.4440	1.4096	2.2600	1.4962	1.0200	0.7880
19	1.1596	1.1683	1.7000	1.7000	1.7000	2.2600	1.7000	0.9900	0.9600
20	1.0301	1.0823	1.4442	1.4660	1.3573	2.1800	1.4930	1.0305	0.6120
21	1.1281	1.2025	1.4536	1.4700	1.4580	2.0104	1.4635	1.0900	1.0234
22	1.0125	1.0472	1.4437	1.4860	1.4168	2.0079	1.4900	1.0308	0.7609
23	1.4800	1.4800	1.3969	1.4263	1.3570	2.0200	1.4188	1.0307	0.5900
24	0.9450	0.9738	1.5100	1.5100	1.5100	2.0200	1.5100	1.0900	0.5900
25	1.0594	1.1084	1.1844	1.1794	1.0661	2.0200	1.2077	1.0900	0.5900
26	1.4800	1.4800	1.1127	1.1559	1.1414	2.0200	1.1404	1.0900	0.5900
27	1.2160	1.2293	1.5100	1.5100	1.5100	2.0200	1.5100	1.0900	0.5900
28	1.2174	1.3010	1.1100	1.1729	1.0923	2.0200	1.1537	0.6494	0.5900
29	1.4800	1.4800	1.4278	1.4341	1.3872	2.0200	1.4201	1.1631	0.5900
30	1.1285	1.1453	1.3092	1.3659	1.2811	2.0200	1.3580	1.0764	0.5900
31	1.5621	1.5600	1.3231	1.3803	1.3454	2.1457	1.3270	1.1244	0.5900
32	1.2363	1.2396	1.7900	1.7900	1.7900	2.3900	1.7900	1.1800	0.5900
33	1.7800	1.7800	1.0770	1.1653	1.0963	2.3900	1.1322	1.1800	0.5900
34	1.3830	1.3775	1.4778	1.4867	1.5261	2.3900	1.5043	1.1327	0.5900
35	1.4171	1.4518	1.4543	1.5537	1.5382	2.3900	1.5952	1.1631	0.5900
36	1.5910	1.5786	1.5532	1.5398	1.4620	2.1500	1.5465	0.8458	0.5900
37	1.3687	1.3859	1.6586	1.6811	1.6694	2.3492	1.7132	0.9334	0.6464
38	1.7100	1.7100	1.6161	1.6002	1.5986	2.3700	1.5945	1.3000	0.6500
39	1.4603	1.4793	1.1428	1.2318	1.1204	2.3700	1.2161	1.0822	0.6500

Month t	p_{10}^t	p_{11}^t	p_{12}^t	p_{13}^t	p_{14}^t	p_{15}^t	p_{16}^t	p_{17}^t	p_{18}^t	p_{19}^t
1	0.7553	0.9095	1.2900	1.0522	1.7500	0.6800	1.7900	1.9536	1.7900	1.4939
2	0.8300	0.9900	1.2900	1.3500	1.7500	0.6800	1.4400	1.7578	1.5637	1.4117
3	0.5280	0.9900	1.2567	1.2776	1.6112	0.6616	1.6126	1.7528	1.5827	1.3792
4	0.6685	0.9900	1.2900	1.1900	1.5900	0.6700	1.3081	1.7095	1.3033	1.4200
5	0.6203	0.8600	1.2900	1.1342	1.5900	0.6700	1.2620	1.7094	1.2607	0.9233
6	0.5900	0.9386	1.2900	1.3842	1.8386	0.7809	1.1895	2.1489	1.4238	1.0674
7	0.8300	0.8393	1.2900	1.4900	1.8900	0.7900	1.2303	2.0555	1.2249	1.9300
8	0.8300	0.9900	1.2900	1.2886	1.9442	0.8291	1.9709	2.2717	1.9699	1.6333
9	0.8088	0.9900	1.1900	1.3496	2.0500	0.8500	1.9600	2.4521	1.9600	1.4278
10	0.8123	0.9900	*1.6087*	1.5900	2.0500	0.8500	1.6045	2.4394	1.6057	1.4213
11	0.7201	0.9900	1.2900	1.4443	2.1464	0.8693	1.9600	2.4165	1.9600	1.4451
12	0.9519	0.8624	1.2900	1.1177	2.1900	0.8900	1.7284	2.3697	1.7579	1.9300
13	0.7683	0.8392	1.0765	1.4161	2.1900	0.8900	1.9600	2.2900	1.9600	1.5737
14	0.9600	0.9419	1.2034	1.5822	2.0855	0.8581	1.4810	2.4470	1.5627	1.4748
15	0.9600	0.9900	1.2900	1.1207	2.0500	0.8500	1.4155	2.3524	1.4374	1.5472
16	0.9600	1.0403	1.2900	1.2071	2.0500	0.8500	1.3793	2.2900	1.5192	1.4954
17	0.7881	1.0600	1.1671	1.3867	1.7668	0.8363	1.2925	2.2900	1.3198	1.7467
18	0.7693	1.0954	1.1179	1.0587	1.6900	0.6332	1.0697	2.0818	1.1456	1.6800
19	0.9600	1.1300	1.4100	0.9647	1.6900	0.7900	1.0330	1.8900	1.0922	1.3131
20	0.5834	1.1300	*1.5388*	0.9677	1.6900	0.7900	1.5000	1.8353	1.5000	1.3311
21	1.0214	0.9632	*1.0364*	0.9629	1.5900	0.7500	1.2542	1.8367	1.2507	1.6082
22	0.7542	1.0334	*1.3301*	1.0506	1.6239	0.7642	1.0378	1.8900	1.0599	1.5200
23	0.5900	1.1500	1.4500	1.0693	1.5900	0.7500	1.0352	1.8900	1.1490	1.2094
24	0.5900	1.1500	1.4500	1.0820	1.5900	0.7500	1.3423	1.8293	1.3476	1.4200
25	0.5900	1.1500	1.4500	0.8743	1.5900	0.7500	1.5000	1.8212	1.5000	1.0178
26	0.5900	1.1500	1.4500	1.0347	1.5900	0.7500	1.0331	1.8270	1.1024	1.4200
27	0.5900	0.9300	1.2300	0.9812	1.5900	0.7500	1.3609	1.8277	1.3589	1.3242
28	0.5900	0.9300	1.2300	1.2500	1.5900	0.7500	1.0296	1.8900	1.0339	1.0153
29	0.5900	0.9300	1.2300	1.0406	1.5900	0.7500	1.0489	1.8900	1.0344	1.0204
30	0.5900	0.9300	1.2300	1.2500	1.5900	0.7500	1.0194	1.8372	1.0219	1.0071
31	0.5900	0.9300	1.2300	1.1474	1.5900	0.7500	1.0485	2.0130	1.0533	1.0597
32	0.5900	0.9300	1.2300	1.3500	1.5900	0.4023	1.1019	2.2900	1.0672	1.2422
33	0.5900	0.9300	1.2300	1.2567	1.5900	0.7500	1.5768	2.2900	1.5630	1.5311
34	0.5900	0.9300	1.2300	1.0672	1.5900	0.7500	1.4765	2.2900	1.4829	1.5900
35	0.5900	0.9300	1.2300	1.3500	1.5900	0.7500	1.5100	2.2054	1.5082	1.3474
36	0.5900	0.9300	1.2300	1.0735	1.5900	0.7500	1.6709	2.2599	1.7327	1.5279
37	0.6464	1.0146	1.3335	1.2864	1.9099	0.9103	1.7535	2.4782	1.7560	1.4474
38	0.6500	1.0200	1.3500	1.5300	1.9700	0.9400	1.5549	2.2212	1.5702	1.3701
39	0.6500	1.0200	1.3500	1.2288	1.9700	0.9400	1.3916	2.3875	1.3794	1.6400

The actual prices p_2^t and p_4^t are not available for t =1,2,...,8 since products 2 and 4 were not sold during these months. However, in the above Table, we filled in these missing prices with the imputed reservation prices that were estimated in Section xx. Similarly, p_{12}^t was missing for months t = 12, 20, 21 and 22 and again, we replaced these missing prices with the corresponding estimated imputed reservation prices in Table A2. The imputed prices appear in italics in the above Table.

The following Table lists the sales shares of the best selling products along with total sales $e^t \equiv p^t \cdot q^t$ of the 19 products in each month. The best selling products were products 1, 5, 11, 13, 14, 15, 16, 18 and 19.

Table A3: Total Sales or Expenditures and Sales Shares of Best Selling Products

t	e^t	s_1^t	s_5^t	s_{11}^t	s_{13}^t	s_{14}^t	s_{15}^t	s_{16}^t	s_{18}^t	s_{19}^t
1	5968.15	0.0350	0.0000	0.0196	0.0000	0.1089	0.0335	0.0321	0.0263	0.0114
2	10027.65	0.0469	0.0000	0.0417	0.0000	0.2274	0.0523	0.0463	0.0229	0.0142
3	8741.65	0.0750	0.0000	0.0249	0.0000	0.1215	0.0542	0.0432	0.0312	0.0106
4	11449.37	0.0175	0.0000	0.0554	0.0000	0.2216	0.0422	0.0518	0.0264	0.0114
5	10899.81	0.0121	0.0000	0.0166	0.0000	0.0760	0.0442	0.0245	0.0256	0.0091
6	9120.39	0.1076	0.0000	0.0217	0.0000	0.1352	0.0463	0.0374	0.0299	0.0110
7	11602.34	0.0090	0.0000	0.0416	0.0000	0.1285	0.0422	0.0367	0.0251	0.0097
8	9435.49	0.0255	0.0000	0.0222	0.0000	0.0760	0.0340	0.0239	0.0199	0.0136
9	9932.26	0.0725	0.0246	0.0357	0.0208	0.1160	0.0423	0.0282	0.0183	0.0216
10	9824.99	0.0827	0.0380	0.0382	0.0242	0.1011	0.0361	0.0356	0.0308	0.0323
11	9941.34	0.0068	0.0038	0.0344	0.0296	0.1335	0.0362	0.0371	0.0410	0.0720
12	10591.60	0.0098	0.0060	0.0156	0.0168	0.0480	0.0304	0.0176	0.0194	0.0302
13	9474.14	0.0808	0.0364	0.0245	0.0327	0.1288	0.0404	0.0372	0.0319	0.0461
14	8816.27	0.0847	0.0455	0.0397	0.0375	0.1282	0.0357	0.0371	0.0178	0.0167
15	8713.10	0.0199	0.0128	0.0265	0.0318	0.0477	0.0312	0.0293	0.0172	0.0103
16	8942.04	0.0535	0.0177	0.0237	0.0359	0.0686	0.0421	0.0313	0.0200	0.0113
17	8837.16	0.0633	0.0219	0.0449	0.0510	0.1692	0.0373	0.0418	0.0245	0.0116
18	9214.45	0.0674	0.0208	0.0313	0.0359	0.0870	0.0390	0.0268	0.0277	0.0114
19	8979.63	0.0615	0.0344	0.0231	0.0295	0.0331	0.0327	0.0248	0.0311	0.0091
20	7768.64	0.1101	0.0385	0.0368	0.0342	0.1169	0.0370	0.0286	0.0272	0.0243
21	8075.62	0.0858	0.0310	0.0299	0.0284	0.0558	0.0286	0.0299	0.0190	0.0236
22	9052.47	0.0855	0.0466	0.0274	0.0271	0.1366	0.0209	0.0395	0.0235	0.0492
23	8040.42	0.1084	0.0101	0.0250	0.0289	0.0981	0.0297	0.0319	0.0262	0.0741
24	7230.79	0.1291	0.0629	0.0169	0.0255	0.0372	0.0226	0.0267	0.0475	0.0516
25	9084.74	0.0692	0.0288	0.0300	0.0239	0.1219	0.0247	0.0286	0.0288	0.0607
26	8040.22	0.0101	0.0077	0.0410	0.0450	0.2107	0.0204	0.0660	0.0560	0.0454
27	8112.81	0.0603	0.0414	0.0210	0.0225	0.0370	0.0284	0.0236	0.0173	0.0617
28	7761.02	0.0482	0.0136	0.0558	0.0357	0.1374	0.0279	0.0534	0.0299	0.0072
29	7838.74	0.0108	0.0181	0.0286	0.0307	0.1365	0.0271	0.0475	0.0126	0.0087
30	8506.49	0.0565	0.0389	0.0289	0.0307	0.1137	0.0287	0.0289	0.0153	0.0146
31	8752.06	0.0100	0.0125	0.0603	0.0388	0.1204	0.0284	0.0587	0.0189	0.0071
32	8613.56	0.0878	0.0701	0.0229	0.0195	0.0461	0.0302	0.0270	0.0177	0.0081
33	7892.10	0.0090	0.0095	0.0753	0.0694	0.1547	0.0345	0.0823	0.0224	0.0090
34	8140.59	0.0581	0.0428	0.0321	0.0484	0.0795	0.0288	0.0434	0.0193	0.0114
35	10813.31	0.0294	0.0177	0.0249	0.0331	0.0622	0.0186	0.0311	0.0172	0.0225
36	8586.06	0.0145	0.0094	0.0275	0.0384	0.0948	0.0243	0.0416	0.0389	0.0438
37	5580.64	0.0846	0.0469	0.0479	0.0392	0.1182	0.0400	0.0531	0.0244	0.0612
38	5702.95	0.0228	0.0066	0.0439	0.0665	0.0995	0.0470	0.0481	0.0276	0.0280
39	7491.91	0.0174	0.0158	0.0554	0.0398	0.1377	0.0351	0.0589	0.0267	0.0200

The following Table lists the sales shares of the 10 least selling products, which were products 2, 3, 4, 6, 7, 8, 9, 10, 12 and 17.

Table A4: Sales Shares of Least Popular Products

t	s_2^t	s_3^t	s_4^t	s_6^t	s_7^t	s_8^t	s_9^t	s_{10}^t	s_{12}^t	s_{17}^t
1	0.0000	0.0196	0.0000	0.0335	0.0321	0.0263	0.0114	0.0057	0.0236	0.0350
2	0.0000	0.0417	0.0000	0.0523	0.0463	0.0229	0.0142	0.0090	0.0308	0.0363
3	0.0000	0.0249	0.0000	0.0542	0.0432	0.0312	0.0106	0.0071	0.0436	0.0399
4	0.0000	0.0554	0.0000	0.0422	0.0518	0.0264	0.0114	0.0084	0.0363	0.0291
5	0.0000	0.0166	0.0000	0.0442	0.0245	0.0256	0.0091	0.0069	0.0249	0.0275
6	0.0000	0.0217	0.0000	0.0463	0.0374	0.0299	0.0110	0.0058	0.0291	0.0266

7	**0.0000**	**0.0416**	**0.0000**	**0.0422**	**0.0367**	**0.0251**	**0.0097**	**0.0067**	**0.0258**	**0.0365**
8	**0.0000**	**0.0222**	**0.0000**	**0.0340**	**0.0239**	**0.0199**	**0.0136**	**0.0095**	**0.0275**	**0.0262**
9	**0.0246**	**0.0357**	**0.0208**	**0.0423**	**0.0282**	**0.0183**	**0.0216**	**0.0151**	**0.0226**	**0.0254**
10	**0.0380**	**0.0382**	**0.0242**	**0.0361**	**0.0356**	**0.0308**	**0.0323**	**0.0286**	**0.0000**	**0.0201**
11	**0.0038**	**0.0344**	**0.0296**	**0.0362**	**0.0371**	**0.0410**	**0.0720**	**0.0587**	**0.0273**	**0.0238**
12	**0.0060**	**0.0156**	**0.0168**	**0.0304**	**0.0176**	**0.0194**	**0.0302**	**0.0227**	**0.0136**	**0.0309**
13	**0.0364**	**0.0245**	**0.0327**	**0.0404**	**0.0372**	**0.0319**	**0.0461**	**0.0146**	**0.0445**	**0.0438**
14	**0.0455**	**0.0397**	**0.0375**	**0.0357**	**0.0371**	**0.0178**	**0.0167**	**0.0123**	**0.0255**	**0.0258**
15	**0.0128**	**0.0265**	**0.0318**	**0.0312**	**0.0293**	**0.0172**	**0.0103**	**0.0109**	**0.0275**	**0.0294**
16	**0.0177**	**0.0237**	**0.0359**	**0.0421**	**0.0313**	**0.0200**	**0.0113**	**0.0073**	**0.0431**	**0.0453**
17	**0.0219**	**0.0449**	**0.0510**	**0.0373**	**0.0418**	**0.0245**	**0.0116**	**0.0052**	**0.0403**	**0.0438**
18	**0.0208**	**0.0313**	**0.0359**	**0.0390**	**0.0268**	**0.0277**	**0.0114**	**0.0050**	**0.0368**	**0.0219**
19	**0.0344**	**0.0231**	**0.0295**	**0.0327**	**0.0248**	**0.0311**	**0.0091**	**0.0056**	**0.0243**	**0.0238**
20	**0.0385**	**0.0368**	**0.0342**	**0.0370**	**0.0286**	**0.0272**	**0.0243**	**0.0206**	**0.0000**	**0.0428**
21	**0.0310**	**0.0299**	**0.0284**	**0.0286**	**0.0299**	**0.0190**	**0.0236**	**0.0195**	**0.0000**	**0.0278**
22	**0.0466**	**0.0274**	**0.0271**	**0.0209**	**0.0395**	**0.0235**	**0.0492**	**0.0335**	**0.0000**	**0.0198**
23	**0.0101**	**0.0250**	**0.0289**	**0.0297**	**0.0319**	**0.0262**	**0.0741**	**0.0617**	**0.0197**	**0.0221**
24	**0.0629**	**0.0169**	**0.0255**	**0.0226**	**0.0267**	**0.0475**	**0.0516**	**0.0433**	**0.0253**	**0.0283**
25	**0.0288**	**0.0300**	**0.0239**	**0.0247**	**0.0286**	**0.0288**	**0.0607**	**0.0394**	**0.0203**	**0.0154**
26	**0.0077**	**0.0410**	**0.0450**	**0.0204**	**0.0660**	**0.0560**	**0.0454**	**0.0403**	**0.0249**	**0.0184**
27	**0.0414**	**0.0210**	**0.0225**	**0.0284**	**0.0236**	**0.0173**	**0.0617**	**0.0172**	**0.0246**	**0.0293**
28	**0.0136**	**0.0558**	**0.0357**	**0.0279**	**0.0534**	**0.0299**	**0.0072**	**0.0057**	**0.0260**	**0.0202**
29	**0.0181**	**0.0286**	**0.0307**	**0.0271**	**0.0475**	**0.0126**	**0.0087**	**0.0071**	**0.0256**	**0.0234**
30	**0.0389**	**0.0289**	**0.0307**	**0.0287**	**0.0289**	**0.0153**	**0.0146**	**0.0074**	**0.0268**	**0.0311**
31	**0.0125**	**0.0603**	**0.0388**	**0.0284**	**0.0587**	**0.0189**	**0.0071**	**0.0049**	**0.0288**	**0.0297**
32	**0.0701**	**0.0229**	**0.0195**	**0.0302**	**0.0270**	**0.0177**	**0.0081**	**0.0054**	**0.0223**	**0.0266**
33	**0.0095**	**0.0753**	**0.0694**	**0.0345**	**0.0823**	**0.0224**	**0.0090**	**0.0044**	**0.0263**	**0.0177**
34	**0.0428**	**0.0321**	**0.0484**	**0.0288**	**0.0434**	**0.0193**	**0.0114**	**0.0070**	**0.0258**	**0.0239**
35	**0.0177**	**0.0249**	**0.0331**	**0.0186**	**0.0311**	**0.0172**	**0.0225**	**0.0193**	**0.0199**	**0.0239**
36	**0.0094**	**0.0275**	**0.0384**	**0.0243**	**0.0416**	**0.0389**	**0.0438**	**0.0372**	**0.0202**	**0.0219**
37	**0.0469**	**0.0479**	**0.0392**	**0.0400**	**0.0531**	**0.0244**	**0.0612**	**0.0577**	**0.0261**	**0.0147**
38	**0.0066**	**0.0439**	**0.0665**	**0.0470**	**0.0481**	**0.0276**	**0.0280**	**0.0172**	**0.0357**	**0.0366**
39	**0.0158**	**0.0554**	**0.0398**	**0.0351**	**0.0589**	**0.0267**	**0.0200**	**0.0206**	**0.0213**	**0.0319**

The specific products (and their package size in ounces) are as follows: 1 = Florida Gold Valencia (12); 2 = Florida Gold Pulp Free (12); 3 = MM Country Style OJ (12); 4 = MM Pulp Free Orange (12); 5 = MM OJ (12); 6 = MM OJ (16); 7 = MM OJ W/CA (12); 8 = MM Fruit Punch (12); 9 = HH Lemonade (12); 10 = HH Pink Lemonade (12); 11 = Dom Apple Juice (12); 12 = Dom Apple Juice (16); 13 = HH OJ (12); 14 = HH OJ (16); 15 = HH OJ (6); 16 = Tropicana SB OJ (12); 17 = Tropicana OJ (16); 18 = Tropicana SB Home Style OJ (12); 19 = Citrus Hill OJ (12)

Appendix B: The Feenstra Double Differencing Approach to the Estimation of a CES Utility Function

In this Appendix, we drop the products that are not present in all periods in order to simplify the econometric estimation procedure. Thus we drop products 2, 4 and 12 from our list of 19 frozen juice products. Thus in our particular application, the number of products N will equal 16. We also renumber our products so that the original Product 13 becomes the Nth product in this Appendix. This product has the largest average sales share.

There are 3 sets of variables in the model ($i = 1,...,N$; $t = 1,...,T$):

- q_i^t is the observed amount of product i sold in period t;
- p_i^t is the observed unit value price of product i sold in period t and
- s_i^t is the observed share of sales of product i in period t that is constructed using the quantities q_i^t and the corresponding observed unit value prices p_i^t.

As we aggregate over time to construct the time period t unit value prices and if there is price change within the time period, then the observed unit value prices will have some time aggregation errors in them. Any time aggregation error will carry over into the observed sales shares. Interestingly, as we aggregate over time, the aggregated quantities sold during the period do not suffer from this time aggregation bias.

Our goal is to estimate the elasticity of substitution for a CES direct utility function f(q) that was discussed in Sections 3 and 6 in the main text. This function is defined as follows:

(B1) $f(q_1,...,q_N) \equiv [\Sigma_{n=1}^N \beta_n q_n^s]^{1/s}$

where the parameters β_n are positive and sum to 1 and s is a parameter which satisfies the inequalities $0 < s \leq 1$. The corresponding elasticity of substitution is defined as $\sigma \equiv 1/(1-s)$. The system of share equations which corresponds to this purchaser utility function was derived as equations (33) in the main text which we repeat here:

(B2) $s_n^t \equiv p_n^t q_n^t/\Sigma_{i=1}^N p_i^t q_i^t = \beta_n (q_n^t)^s/\Sigma_{i=1}^N \beta_i (q_i^t)^s$; $\quad t = 1,...,T; n = 1,...,N.$

This system of share equations corresponds to the purchasers' system of inverse demand equations, which give purchase unit value prices as functions of quantities purchased. We take natural logarithms of both sides of the equations in (B2) and add error terms e_n^t in order to obtain the following *fundamental set of estimating equations* for *Model 16*:

(B3) $\ln s_i^t = \ln\beta_i + s\ln q_i^t + \ln[\Sigma_{n=1}^N \beta_n \ln(q_n^t)^s] + e_{si}^t$; $\quad i = 1,...,N; t = 1,...,T$

where the q_i^t are measured without error and the error terms have 0 means and a classical (singular) covariance matrix for the shares within each time period and the error terms are uncorrelated across time periods. The unknown parameters in (B3) are the positive parameters β_n and the positive parameter s where $0 < s \leq 1$.

The error terms in equations (B3) reflect not only time aggregation errors in forming the monthly unit value prices but they also reflect the fact that our assumed CES functional form for the purchasers' utility function may not be correct and the maximization of this utility function may take place with errors. Note that we are also assuming that the error terms are multiplicative error terms on the observed shares (before taking logs) rather than the additive error terms that we have assumed in Section 6.

The Feenstra double differenced variables are defined in two stages. First we difference the of the *logarithms* of the s_n^t with respect to time; i.e., define Δs_n^t as follows:

(B4) $\Delta s_n^t \equiv \ln(s_n^t) - \ln(s_n^{t-1})$; $\qquad$ n = 1,...,N; t = 2,3,...,T.

Now pick product N as the numeraire product and difference the Δs_n^t with respect to product N, giving rise to the following *double differenced log variable,* ds_n^t:

(B5) $ds_n^t \equiv \Delta s_n^t - \Delta s_N^t$; $\qquad$ n = 1,...,N−1; t = 2,3,...,T
$= \ln(s_n^t) - \ln(s_n^{t-1}) - \ln(s_N^t) - \ln(s_N^{t-1})$.

Define the *double differenced log quantity variables* in a similar manner:

(B6) $dq_n^t \equiv \Delta q_n^t - \Delta q_N^t$; $\qquad$ n = 1,...,N−1; t = 2,3,...,T
$= \ln(q_n^t) - \ln(q_n^{t-1}) - \ln(q_N^t) - \ln(q_N^{t-1})$.

Finally, define the *double differenced error variables* ε_n^t as follows:

(B7) $\varepsilon_n^t \equiv e_n^t - e_n^{t-1} - e_N^t + e_N^{t-1}$; $\qquad$ n = 1,...,N−1; t = 2,3,...,T.

Using definitions (B5)-(B7) and equations (B3), it can be verified that the double differenced log shares ds_n^t satisfy the following system of (N−1)(T−1) estimating equations under our assumptions:

(B8) $ds_n^t = s\, dq_n^t + \varepsilon_n^t$; $\qquad$ n = 1,...,N−1; t = 2,3,...,T

where the new residuals, ε_{si}^t, have means 0 and a constant (N−1) by (N−1) covariance matrix within a time period but are uncorrelated across time periods. Thus we have a classical system of linear estimating equations with only one unknown parameter across all equations, namely the parameter s. This is the simplest possible system of estimating equations that one could imagine!

Using the data listed in Appendix A, we have 15 product estimating equations of the form (B8) which we estimated using the NL system command in Shazam. thus our N = 16 and our T = 39. The resulting estimate for s was 0.86491 (with a standard error of 0.0067) and thus the corresponding estimated σ is equal to $1/(1-s) = 7.4025$, which is in line with our earlier estimates for σ when we estimated the CES utility function using Models 4 and 15. The standard error on s was tiny using the present regression results so σ was very accurately determined using this method. The equation by equation R^2 were as follows: 0.9936, 0.9895, 0.9905, 0.9913, 0.9869, 0.9818, 0.9624, 0.9561, 0.9858, 0.9911, 0.9934, 0.994, 0.9906, 0.9921 and 0.9893. The average R^2 is 0.9859 which is very high for share equations or for transformations of share equations. The results are all the more remarkable considering that we have only one unknown parameter in the entire system of $(N-1)(T-1) = 570$ equations.[89]

Why are the fits so good in Model 16 as compared to our earlier Models 4 and 15 which also estimated the CES utility function? It seems likely that the explanation is the assumption of multiplicative error terms in Model 16 as opposed to the assumption of additive error terms which were assumed in Models 4 and 15.

References

Allen, R.G.C. (1938), Mathematical Analysis for Economists, London: Macmillan.

Arrow, K.J., H.B. Chenery, B.S. Minhas and R.M. Solow (1961), "Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics* 63, 225-250.

Australian Bureau of Statistics (2016), "Making Greater Use of Transactions Data to Compile the Consumer Price Index", Information Paper 6401.0.60.003, November 29, Canberra: ABS.

Diewert, W.E., (1974), "Applications of Duality Theory," pp. 106-171 in M.D. Intriligator and D.A. Kendrick (ed.), *Frontiers of Quantitative Economics*, Vol. II, Amsterdam: North-Holland.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics* 4, 114-145.

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

Diewert, W.E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433-528 in *The Measurement of Capital*, Dan Usher (ed.), Studies in Income and Wealth, Vol. 45, National Bureau of Economics Research, Chicago: University of Chicago Press.

Diewert, W.E. (1987), "Index Numbers", pp. 767-780 in *The New Palgrave A Dictionary of Economics*, Vol. 2, J. Eatwell, M. Milgate and P. Newman (eds.), London: The Macmillan Press.

Diewert, W.E. (1993a), "The Early History of Price Index Research", pp. 33-65 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

[89] The results are dependent on the choice of the numeraire product. Ideally, we want to choose the product that has the largest sales share and the lowest share variance.

Diewert, W.E. (1993b), "Duality Approaches To Microeconomic Theory", in *Essays in Index Number Theory*, pp. 105-175 in Volume I, Contributions to Economic Analysis 217, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North Holland.

Diewert, W.E. (1998), "Index Number Issues in the Consumer Price Index", *The Journal of Economic Perspectives* 12:1, 47-58.

Diewert, W.E. and K.J. Fox (2017), "Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data", Discussion Paper 17-02, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1L4.

Diewert, W.E. and R.J. Hill (2010), "Alternative Approaches to Index Number Theory", pp. 263-278 in *Price and Productivity Measurement*, W.E. Diewert, Bert M. Balk, Dennis Fixler, Kevin J. Fox and Alice O. Nakamura (eds.), Victoria Canada: Trafford Press.

Diewert, W.E. and T.J. Wales (1987), "Flexible Functional Forms and Global Curvature Conditions", *Econometrica* 55, 43-68.

Diewert, W.E. and T.J. Wales (1988), "A Normalized Quadratic Semiflexible Functional Form", *Journal of Econometrics* 37, 327-42.

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

Feenstra, R.C. and M.D. Shapiro (2003), "High Frequency Substitution and the Measurement of Price Indexes", pp. 123-149 in *Scanner Data and Price Indexes*, R.C. Feenstra and M.D. Shapiro (eds.), Studies in Income and Wealth, Volume 64, Chicago: University of Chicago Press.

Fisher, I. (1922), *The Making of Index Numbers*, Houghton-Mifflin, Boston.

Hardy, G.H., J.E. Littlewood and G. Polyá (1934), *Inequalities*, Cambridge: Cambridge University Press.

Hausman, J.A. (1996), "Valuation of New Goods under Perfect and Imperfect Competition", pp. 20 -236 in *The Economics of New Goods*, T.F. Bresnahan and R.J. Gordon (eds.), Chicago: University of Chicago Press.

Hausman, J.A. (1999), "Cellular Telephone, New Products and the CPI", *Journal of Business and Economic Statistics* 17:2, 188-194.

Hausman, J. (2003), "Sources of Bias and Solutions to Bias in the Consumer Price Index", *Journal of Economic Perspectives* 17:1, 23-44.

Hausman, J.A. and G.K. Leonard (2002), "The Competitive Effects of a New Product Introduction: A Case Study", *Journal of Industrial Economics* 50:3, 237-263.

Hicks, J.R. (1940), "The Valuation of the Social Income", *Economica* 7, 105-124.

Hofsten, E. von (1952), *Price Indexes and Quality Change*, London: George Allen and Unwin.

Keynes, J.M. (1930), *Treatise on Money*, Vol. 1. London: Macmillan.

Konüs, A.A. (1924), "The Problem of the True Index of the Cost of Living", translated in *Econometrica* 7, (1939), 10-29.

Konüs, A.A. and S.S. Byushgens (1926), "K probleme pokupatelnoi cili deneg", *Voprosi Konyunkturi* 2, 151-172.

Lehr, J. (1885), *Beiträge zur Statistik der Preise*, Frankfurt: J.D. Sauerländer.

Marshall, A. (1887), "Remedies for Fluctuations of General Prices', *Contemporary Review* 51, 355-375.

Rothbarth, E. (1941), "The measurement of Changes in Real Income under Conditions of Rationing", *Review of Economic Studies* 8, 100-107.

Sato, K. (1976), "The Ideal Log-Change Index Number", *Review of Economics and Statistics* 58, 223-228.

Shephard, R.W. (1953), *Cost and Production Functions*, Princeton: Princeton University Press.

University of Chicago (2013), *Dominick's Data Manual*, James M. Kilts Center, University of Chicago Booth School of Business.

Uzawa, H. (1962), "Production Functions with Constant Elasticities of Substitution", *Review of Economic Studies* 29, 291-299.

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

Walsh, C.M. (1901), *The Measurement of General Exchange Value*, New York: Macmillan.

White, K.J. (2004), *Shazam: User's Reference Manual, Version 10*, Vancouver, Canada: Northwest Econometrics Ltd.

Wiley, D.E., W.H. Schmidt and W.J. Bramble (1973), "Studies of a Class of Covariance Structure Models", *Journal of the American Statistical Association* 68, 317-323.