Title
Estimating the Benefits of New Products

Authors
Diewert, Erwin
Feenstra, Robert C

Publication Date
2019-06-01

Peer reviewed

Estimating the Benefits of New Products*

by

W. Erwin Diewert
University of British Columbia and NBER

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

June 5, 2019

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. Hausman used a linear approximation to the demand curve to compute the reservation price, while Feenstra used a reservation price of infinity for a CES demand curve, which will lead to higher gains. The present paper evaluates these approaches, comparing the CES gains to those obtained using a quadratic utility function using scanner data on frozen juice products. We find that the CES gains from new frozen juice products are about five times greater than those obtained using the quadratic utility function.

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

JEL Classification Numbers: C33, C43, C81, D11, D60, E31.

* We thank the organizers and participants at the *Big Data for 21st Century Economic Statistics* conference, and especially Marshall Reinsdorf and Matthew Shapiro, for their helpful comments. We acknowledge the James A. Kilts Center, University of Chicago Booth School of Business, https://www.chicagobooth.edu/research/kilts/datasets/dominicks, for the use of the Dominick's Dataset. Financial support was received from a Digging into Data multi-country grant, provided by the United States NSF and the Canadian SSHRC.

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.[1] With these reservation (or virtual[2]) prices 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 reservation prices to be estimated*?

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 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. Additional bias formulae were developed by Diewert (1980; 498-501) (1987; 779) (1998; 51-54) and Hausman (2003; 26-28). Hausman proposes taking a *linear approximation to the demand curve* at the point of consumption, and computing the consumer surplus gain to a new product under this linear demand curve. Provided that the demand curve is convex, then this linear approximation will be a *lower bound* to the consumer surplus gain. We will compare that proposal to other methods of dealing with new goods.

Researchers have also relied 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

---

[1] "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.

[2] 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).

availability. The two main contributors in this area are Feenstra (1994) and Hausman (1996).[3] Feenstra assumes a *constant elasticity of substitution* (CES) utility or cost function, while Hausman assumes an *almost ideal demand system* (AIDS). The CES functional form is not fully flexible (in contrast to the AIDS), so that is one drawback of Feenstra's approach.[4] He adopts that case because it has a particularly simple form of the reservation prices: in the CES case, the demand curve never touches the price axis and so the reservation price is *infinity*. As we will show in the following sections, however, the area under demand curve is bounded provided that the elasticity of substitution is greater than unity, and it can be computed with information on the expenditure on the new goods and the elasticity. So Feenstra's methodology side-steps the issue of estimating the reservation prices, but instead, requires that we estimate the elasticity of substitution. Feenstra (1994) provides a robust double-differencing method to estimate that elasticity that can be applied to a dataset with many new and disappearing goods, as typically occur with scanner data.

To summarize, there are two problems with Feenstra's CES methodology for measuring the net benefits of changes in the availability of products: (i) the CES functional form is not fully flexible; and (ii) the reservation price that induces a potential consumer 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. Against these drawbacks, a benefit is that the elasticity of substitution can be estimated quite easily using the double-differencing method, and the elasticity along with the expenditure share on the items are sufficient information to compute the consumer benefits from new products.

In section 2, we begin with the simple example of a partial equilibrium, constant-elasticity demand curve, which has a reservation price of infinity. We show that the consumer surplus under a constant-elasticity demand curve is at least twice the consumer surplus under a linear approximation to the demand curve. This result is our first illustration of the extent to which a constant-elasticity case will lead to greater gains than a linear demand curve, i.e. by about a factor of at least two when the elasticity of demand is the *same* for the two demand curves and reasonably high. While these results in section 2 are suggestive, they are not rigorous

[3] See also Hausman (1999) (2003) and Hausman and Leonard (2002)

[4] See Diewert (1974) (1976) for the definition of a flexible functional form. Feenstra (2010) shows that the CES methodology discussed here to measure the gains from new goods can be extended to the AIDS case.

because they rely on a partial equilibrium demand curve with a single new good. Our general goal is to measure total consumer utility (not just consumer surplus) and when there are potentially many new and disappearing goods. Accordingly, in section 3 we examine a constant elasticity of substitution (CES) utility function, and show that the exact gains from new goods are still at least twice as high as those obtained from a linear approximation to that demand curve. In addition to the CES utility function, we also examine the *quadratic flexible functional form* that was initially due to Konüs and Byushgens (1926; 171). That utility function can be used to justify the Fisher (1922) price index, and so we will also call it the *KBF functional form*. The demand curves for both the CES and KBF demand curves are convex under weak conditions, but the CES demand is *more* convex.

In section 4, we turn to the econometric estimation of the demand system for the CES and KBF utility functions, using scanner data for frozen juice in one grocery store, as described in section 4.1. The estimation of the CES demand curves can be simplified using a double-differencing method due to Feenstra (1994), which eliminates all unknown parameters except the elasticity of substitution. In sections 4.2–4.4, we show that this method performs very well on the scanner data. In comparison, estimation of the demand curves corresponding to the quadratic utility function is more difficult because it inherently has more free parameters, i.e. N (N+1)/2 free parameters in a symmetric matrix with N goods. We solve this degrees of freedom problem by introducing a *semiflexible version* of the flexible quadratic functional form.[5] This new methodology is explained and implemented in sections 4.5–4.7.

In section 4.8, we compare the results obtained from the CES and KBF utility functions for the consumer benefits from new goods. According to our theoretical results in section 3, we would expect that the CES gains should be not much more than twice as high as the KBF gains (because the KBF gains exceed those from a linear approximation), provided that those demand curves have the same elasticity at the point of consumption. In fact, that is not what we find: the CES gains are *more than five times the size* of the KBF gains. The reason for this result is that the implied elasticities of demand for the two preferences systems, evaluated at the same point of consumption for the new goods, are actually quite different: the KBF preferences give *demand*

[5] 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 4.4 below, we also show how the correct curvature conditions can be imposed on our semiflexible quadratic functional form.

that is at least twice as elastic as the CES demand for the new varieties of frozen juice. This finding highlights an important difference between the CES and KBF utility functions: because the former has a single estimation parameter, and the latter has a whole matrix of parameters, it will not in general be the case that they have the same elasticity of demand when estimated. Indeed, this result is implied by the limitation that the CES utility function is not fully flexible. That theoretical limitation becomes an important simplification for estimation, however. We believe that it is practical for statistical agencies to implement the double-differenced estimation of the CES system, but it would be much more challenging for them to implement the estimation of the KBF system, at least for most datasets. In the end, we are left with a trade-off between the practicality of using the CES system against the challenge of estimating a more flexible utility function to obtain a more general measure of gains. Further conclusions are provided in section 5. The dataset is listed in Appendix A, so that other researchers can use it to test out possible improvements to our methods, and certain results are proved in Appendices B and C.

2. Constant-Elasticity Demand Curve

Consider a constant-elasticity demand curve of the form $q_1 = kp_1^{-\sigma}$, where q_1 denotes quantity of good 1, p_1 denotes its price, and $k > 0$ is parameter. In period t this good is newly available at the price of p_{1t} and the chosen quantity q_{1t}. The demand curve is illustrated in Figure 1 and it approaches the vertical axis as the price approaches infinity, which means that the reservation price of the good is *infinite*. But provided that the elasticity of demand σ is greater than unity, the area under the demand curve, as shown by the regions A+B+C in Figure 1, is bounded above. Region A is the expenditure on the good, while B+C is the consumer surplus. The consumer surplus is calculated as the area to the left of the demand curve between its price of p_{1t} and infinity, and relative to total expenditure E_t on all goods it equals:

(1) $$\frac{B+C}{E_t} = \frac{1}{E_t}\int_{p_{1t}}^{\infty} kp^{-\sigma}dp = \frac{p_{1t}q_{1t}}{E_t(\sigma-1)} = \frac{s_{1t}}{(\sigma-1)}, \quad \sigma > 1,$$

where $s_{1t} \equiv p_{1t}q_{1t} / E_t$ denotes the share of spending on good 1. We see that this expression for the consumer gains from the new good shrinks as the elasticity of substitution is higher, indicating that the new good is a closer substitute for an existing good.

One might worry that calculating the consumer gains this way, with a reservation price of infinity, results in gains that are too large. A suggestion given by Hausman (2003) is to use a linear approximation to the demand curve, as shown by the dashed line in Figure 1. The linear approximation to the demand function goes through the price axis at the reservation price p_1^*, where $p_1^* \equiv p_{1t} + \alpha q_{1t}$ and $\alpha \equiv (p_1^* - p_{1t})/q_{1t} > 0$ is the absolute value of the slope of the inverse constant-elasticity demand curve evaluated at $q_1 = q_{1t}$. Hausman took the area of the triangle below the linear approximation to the true demand curve but above the line $p_1 = p_{1t}$ as his lower-bound measure of the gain in consumer surplus that would occur due to the new product. That consumer surplus area is region B in Figure 1, which is less than the area under the constant elasticity demand curve, B+C. Indeed, we now show that the consumer surplus B following Hausman's method is less than one-half of the true consumer surplus region B+C.

The consumer surplus B relative to total expenditure on the product E_t is obtained by computing the area of that triangle,

$$(2) \qquad \frac{B}{E_t} = \frac{(p_1^* - p_{1t})q_{1t}}{2E_t} = \frac{\alpha (q_{1t})^2}{2E_t} = \frac{\alpha (q_{1t} / p_{1t}) p_{1t} q_{1t}}{2E_t} = \frac{s_{1t}}{2\sigma},$$

where the second equality follows from the definition of the slope $\alpha \equiv (p_1^* - p_{1t}) / q_{1t}$ of the inverse demand curve; the third equality from algebra; and the fourth equality because we have assumed the slope of the constant-elasticity demand curve and its linear approximation are equal at the point of consumption, so it follows that the inverse elasticity of demand must also be equal, $\alpha(q_{1t} / p_{1t}) = 1/\sigma$. Comparing equations (1) and (2), the ratio of the consumer surplus from the linear approximation to that from the constant-elasticity demand curve is *less than one-half,* $B/(B+C) = (\sigma - 1)/2\sigma < 1/2$. Those two measures of gain are summarized in Table 1 for $s_{1t} = 0.1$ and various values of σ.

Column two in Table 1 consists of the constant-demand elasticity gain in (1) and column three shows the Hausman approximate gain in (2), while column four takes their ratio. While there results give us a first illustration that the gains in the constant-demand-elasticity case, they lack rigor by dealing with consumer surplus for a partial equilibrium demand curve with only one new good. Accordingly, in the next section we extend our results to many new (and disappearing) goods while using a constant-elasticity-of-substitution (CES) utility function. We will find that the constant-demand-elasticity and CES cases give quite similar results.

Table 1: Consumer Gains from a New Product with Share= 0.1 (Percent of Expenditure)

σ	$(B+C)/E_t$	B/E_t	Ratio	G_{CES}	$G_{H,CES}$	Ratio
2	10.0	2.50	0.25	11.1	2.78	0.25
3	5.00	1.67	0.33	5.40	1.85	0.34
4	3.33	1.25	0.37	3.58	1.39	0.39
5	2.50	1.00	0.40	2.66	1.11	0.42
6	2.00	0.83	0.42	2.12	0.93	0.44
10	1.12	0.50	0.45	1.18	0.56	0.47

Notes: Column two computes the constant-demand-elasticity gain in (1); column three computes the Hausman gain (2) as a lower bound to the constant-demand-elasticity case; column four computes the ratio of the previous two columns; column five computes the CES gain (15); column six computes the Hausman gain (18) as a lower bound to the CES case; and column seven computes the ratio of the previous two columns.

3. Utility-based Approach

3.1 Utility Function Approach

We begin with a CES utility function for the consumer,[6] defined by,

$$(3) \qquad U_t = U(q_t, I_t) = \left[\sum_{i \in I_t} a_i q_{it}^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \quad \sigma > 1, \qquad t=1,\ldots,T.$$

where $a_i > 0$ are parameters and $I_t \subseteq \{1,\ldots,N\}$ denotes the set of goods or varieties that are available in period t=1,…,T at the prices p_{it}. We will treat this set of goods as changing over time due to new or disappearing varieties. The unit-expenditure function is defined as the minimum expenditure to obtain utility of one. For the CES utility function, the unit-expenditure function is:

$$(4) \qquad e(p_t, I_t) = \left[\sum_{i \in I_t} b_i p_{it}^{1-\sigma} \right]^{1/(1-\sigma)}, \quad \sigma > 1, \; b_i \equiv a_i^{\sigma}, \qquad t=1,\ldots,T.$$

It follows that total expenditure needed to obtain utility of U_t is $E_t = U_t \, e(p_t, I_t)$.

From Shepard's Lemma, we can differentiate the expenditure function with respect to p_{it} to obtain the Hicksian demand q_{it} for that good,

[6] The CES function was introduced into the economics literature by Arrow, Chenery, Minhas and Solow (1961), and in the mathematics literature it is known as a mean of order $r \equiv 1 - \sigma$; see Hardy, Littlewood and Polyá (1934; 12-13). Rather than being a utility function for a consumer, equation (1) could instead be a production function for a firm. In that case, we would replace utility U_t by output Y_t.

(5) $$q_{it}(p_t, U_t) = U_t\left[\sum_{i\in I_t} b_i p_{it}^{1-\sigma}\right]^{\frac{\sigma}{1-\sigma}} b_i p_{it}^{-\sigma}, \qquad t = 1,...,T;\ i\in I_t.$$

Multiplying by p_{it} and dividing by expenditure E_t to obtain expenditure shares,

(6) $$s_{it} \equiv \frac{p_{it}q_{it}}{E_t} = \frac{b_i p_{it}^{1-\sigma}}{\sum_{n\in I_t} b_n p_{nt}^{1-\sigma}}, \qquad t = 1,...,T;\ i\in I_t.$$

Notice that the quantity q_{it} approaches zero as $p_{it}\to\infty$, in which case the share in (5) also approaches zero provided that $\sigma > 1$. Differentiating $-\ln q_{it}$ from (5) with respect to $\ln p_{it}$, we obtain the (positive) Hicksian own-price elasticity corresponding to the CES utility function,

(7) $$\eta_{it}\big|_U \equiv -\frac{\partial \ln q_{it}}{\partial \ln p_{it}}\bigg|_U = \sigma(1-s_{it}).$$

This elasticity is not constant as was assumed for the partial equilibrium, constant-elasticity demand curve in the previous section. Rather, the elasticity in (7) varies between an upper-bound of σ when $p_{it}\to\infty$ and the share in (6) approaches zero, and a lower-bound of zero when the share of this product approaches one.[7]

Initially, we consider the case where there is no change in the set of goods over time, so $I_{t-1} = I_t \equiv I$. Our goal is to measure the ratio of the unit-expenditure functions with a formula depending only on observed prices and quantities, which will then correspond to an "exact" price index (Diewert, 1974). We maintain throughout the assumption that the observed quantities are optimally chosen for the prices, i.e. that they correspond to the shares given in (6). When these shares are computed over the goods $i\in I$, we denote them as:

(8) $$s_{i\tau}(I) \equiv p_{i\tau}q_{i\tau}\Big/\sum_{n\in I} p_{n\tau}q_{n\tau}, \qquad \tau = t-1, t;\ i\in I.$$

Then dividing $s_{it}(I)$ by $s_{it-1}(I)$ from (6), raising this expression to the power $1/(\sigma-1)$, making use of (4) and rearranging terms slightly, we obtain:

[7] The fact that the elasticity is close to zero for shares approaching unity suggests that the Hicksian CES demand curve cannot be globally convex for all shares: very inelastic demand must be concave in a region as prices rise and the demand curve bends towards the price axis. Nevertheless, it is shown in Appendix C that the Hicksian demand curve in (5) is strictly convex provided $s_{it} \le 0.5$.

$$(9) \qquad \left(\frac{s_{it}(I)}{s_{it-1}(I)}\right)^{\frac{1}{1-\sigma}} \frac{e(p_t, I)}{e(p_{t-1}, I)} = \left(\frac{p_{it}}{p_{it-1}}\right), \qquad i \in I.$$

To simplify (9) further, we make use of the weights $w_i(I)$ defined by,

$$(10) \qquad w_i(I) \equiv \left(\frac{s_{it}(I) - s_{it-1}(I)}{\ln s_{it}(I) - \ln s_{it-1}(I)}\right) \Bigg/ \sum_{n \in I} \left(\frac{s_{nt}(I) - s_{nt-1}(I)}{\ln s_{nt}(I) - \ln s_{nt-1}(I)}\right), \qquad i \in I.$$

The numerator in (10) is the logarithmic mean of the shares $s_{it}(I)$ and $s_{it-1}(I)$, and lies in-between these two shares,[8] while the denominator ensures that the weights $w_i(I)$ sum to unity. Then we take the geometric mean of both sides of (9) using the weights $w_i(I)$, to obtain:

$$(11) \qquad \frac{e(p_t, I)}{e(p_{t-1}, I)} \prod_{i \in I} \left(\frac{s_{it}(I)}{s_{it-1}(I)}\right)^{w_i(I)} = \frac{e(p_t, I)}{e(p_{t-1}, I)}, \text{ since } \prod_{i \in I} \left(\frac{s_{it}(I)}{s_{it-1}(I)}\right)^{w_i(I)} = 1,$$

$$= P_{SV}(I) \equiv \prod_{i \in I} \left(\frac{p_{it}}{p_{it-1}}\right)^{w_i(I)}, \qquad \text{using (9).}$$

The result on the first line of (11) that the product shown equals unity follows from taking the log of this expression and using the weights defined in (10), along with the fact that $\sum_{i \in I} s_{it-1}(I) = \sum_{i \in I} s_{it}(I) = 1$ from (8). Then it follows from (11) that the ratio of the unit-expenditure functions equals the term $P_{SV}(I)$ defined as shown, which is the price index due to Sato (1967) and Vartia (1967) constructed over the (constant) set of goods I.

With this result in hand, let us now consider the case where the set of goods is changing over time but some of the goods are available in both periods, so that $I_{t-1} \cap I_t \neq \varnothing$. We again let $e(p_\tau, I)$ denote the expenditure function defined over the goods within the set I, which is the set of goods available in both periods, $I \equiv I_{t-1} \cap I_t$. We refer to the set I as the "common" set of goods because they are available in both periods.[9] The ratio $e(p_t, I)/e(p_{t-1}, I)$ is still measured by the

---

[8] Treating $s_{it-1}(I)$ as a fixed number, it is straightforward to show using L'Hôpital's rule that as $s_{it}(I) \to s_{it-1}(I)$ then the numerator of (10) also approaches $s_{it-1}(I)$. So the Sato-Vartia weights are well defined even as the shares approach each other. The concavity of the natural log function can be used to show that the numerator of the Sato-Vartia weights lie in-between $s_{it}(I)$ and $s_{it-1}(I)$ for all goods $i \in I$.

[9] Feenstra (1994) shows that we can instead define I as a non-empty subset of the goods available in both periods, and obtain the same results as shown below, but we do not pursue that generalization here. Later in the paper, we will refer to the price index constructed with these common goods as the *maximum overlap* index.

Sato-Vartia index as in expression (11). Our interest, however, is in the ratio $e(p_t,I_t)/\, e(p_{t-1},I_{t-1})$ that incorporates new and disappearing goods. To measure this ratio we return to the share equation (6), which applies for all goods $i \in I_t$. Notice that these shares can be re-written as:

$$s_{i\tau} \equiv \frac{p_{i\tau} q_{i\tau}}{\sum_{n \in I_\tau} p_{n\tau} q_{n\tau}}, \qquad \tau = t-1, t;\ i \in I_t\ .\rho \tag{12}$$

Now we can proceed in the same fashion as (9), using (4), (6) and (12) to form the ratio,

$$\left(\frac{s_{it}(I)\lambda_t}{s_{it-1}(I)\lambda_{t-1}} \right)^{\frac{1}{1-\sigma}} \frac{e(p_t, I)}{e(p_{t-1}, I)} = \left(\frac{p_{it}}{p_{it-1}} \right), \qquad i \in I. \tag{13}$$

Once again, we take the geometric mean of both sides of (13) using the weights $w_i(I)$, and shifting the terms λ_t and λ_{t-1} to the right, we obtain in the same manner as equation (11):

$$\frac{e(p_t, I_t)}{e(p_{t-1}, I_{t-1})} = P_{SV}(I) \left(\frac{\lambda_t}{\lambda_{t-1}} \right)^{1/(\sigma-1)}. \tag{14}$$

This result shows that the exact price index for the CES utility and expenditure function is obtained by modifying the Sato-Vartia index, constructed over the common set of goods, by the ratio of the terms $\lambda_\tau(I) \leq 1$. Each of these terms can be interpreted as the *period τ expenditure on the goods in the common set I, relative to the period τ total expenditure.* Alternatively, $\lambda_t(I)$ is interpreted as *one minus the period t expenditure on new goods (not in the set I), relative to the period t total expenditure,* while $\lambda_{t-1}(I)$ is interpreted as *one minus the period t-1 expenditure on disappearing goods (not in the set I), relative to the period t-1 total expenditure.* When there is a greater expenditure share on new goods in period t than on disappearing goods in period t-1, then the ratio $\lambda_t(I)/\, \lambda_{t-1}(I)$ will be less than unity, which leads to a *fall* in the exact price index in (14) by an amount that depends on the elasticity of substitution.

The importance of the elasticity of substitution can be seen from Figure 2, where we suppose that the consumer minimizes the expenditure needed to obtain utility along the indifference curve AD. If initially only good 1 is available, then the consumer chooses point A with the budget line AB. When good 2 becomes available, the same level of utility can be obtained with consumption at point C. Then the drop in the cost of living is measured by the inward movement of the budget line from AB to the line through C, and this shift depends on the convexity of the indifference curve, or the elasticity of substitution.

To relate the CES result in (14) back to equation (1), suppose that: only good 1 is newly available in period t so that $\lambda_t(I)=1-s_{1t}$; there are no disappearing goods so that $\lambda_{t-1}(I)=1$; and the prices of all other goods do not change so that $P_{SV}=1$. We follow Hausman (2003) in constructing the expenditure that would be needed to give the consumer the same utility level U_t even if good 1 is not available. That expenditure level is $E_t^* \equiv U_t e(p_t, I_{t-1})$. Then taking the difference between E_t^* and E_t, we have the compensating variation for the loss of good 1:

$$\text{(15)} \qquad G_{CES} \equiv \frac{E_t^* - E_t}{E_t} = \frac{e(p_t, I_{t-1}) - e(p_t, I_t)}{e(p_t, I_t)} = (1-s_{1t})^{-1/(\sigma-1)} - 1,$$

using the formula for $e(p_t, I_{t-1})/e(p_t, I_t)$ from (14). Taking a second-order Taylor series expansion around $s_{1t} = 0$, this gain can be expressed as:

$$\text{(16)} \qquad G_{CES} = \left(1-s_{1t}\right)^{-1/(\sigma-1)} - 1 = \frac{s_{1t}}{(\sigma-1)} + \frac{\sigma \tilde{s}_{1t}^2}{2(\sigma-1)^2}, \quad \text{for } 0 \le \tilde{s}_{1t} \le s_{1t},$$

$$\ge \frac{s_{1t}}{(\sigma-1)}, \qquad \text{since } \tilde{s}_{1t}^2 \ge 0.$$

We see that the second line of (16) is identical to (1), which is therefore a lower-bound to the CES gains. In the fifth column of Table 1, we show the CES gains from (15), which are slightly above the constant-demand-elasticity gains from (1). Our results in this section show that the CES gains with many new (and disappearing) goods give a generalization of the simple, *consumer surplus* calculation of section 2. In the next section we compare these CES gains to an approximation of the measure of *total consumer utility* gain due to Hausman (2003).

3.2 Hausman Lower Bound to the Welfare Gain

Hausman (1999; 191) (2003; 27) proposed a very simple methodology for calculating a lower bound to the gain from the appearance of a new good. We illustrated that approach for a demand curve with elasticity of σ in section 2, but Hausman argues that it holds more generally for any Hicksian demand curves with constant utility. Letting $\eta_{1t}\big|_U$ denote the (positive) compensated demand derivative for good 1 when it first appears, we obtain the generalization of (2) by replacing σ with the Hicksian elasticity:

(17) $$G_H = \frac{s_{1t}}{2\,\eta_{1t}\big|_U}.$$

For the CES demand curve, we can calculate the lower bound to the welfare gain using the elasticity of demand for the CES system, as calculated in (7), and we obtain,

(18) $$G_{H,CES} = \frac{s_{1t}}{2\sigma(1-s_{1t})}\quad .$$

In column six of Table 1 we calculate the Hausman lower-bound gains in (18) using the Hicksian elasticities for CES demand, and in column seven we show the ratio of the CES gain in (15) and the Hausman lower-bound in (18). Similar to what we found for the constant-demand-elasticity case in the previous section, the Hausman lower bound calculation in (18) is less than one-half of the CES gains in (15), and approaches one-half of those gains for elasticities of substitution that are reasonably high.

We next derive the formula for the Hausman lower-bound formula in (17) for a *general form* of utility even when the Hicksian demand curves are not well-behaved and differentiable. That will turn out to be the case for quadratic utility that we consider in the next section, which will give rise to well-behaved inverse demand curves (prices as a function of quantities), but not necessarily well-behaved direct demand curves (quantities as a function of prices). So this derivation focusing on inverse demand curves will be important for the rest of the paper.

Denote the utility function by $U = f(q) \geq 0$, where $f(q)$ is non-decreasing, concave and homogeneous of degree one for $q \equiv (q_1,...,q_N) \geq 0_N$, and twice continuously differentiable for $q >> 0_N$. We suppose that the consumer faces positive prices $p_t \equiv (p_{1t},...,p_{Nt}) >> 0_N$ in period t and maximizes utility:

(19) $$\max_{q\geq 0} \{f(q) : p_t \cdot q \leq E_t\},$$

where $p_t \cdot q$ is the inner product. The first order necessary conditions for an *interior maximum*[10] with the period t quantity vector $q_t >> 0_N$ solving (19) are:

(20) $$\nabla f(q_t) = \lambda_t p_t\ ,$$

(21) $$p_t \cdot q_t = E_t,$$

[10] Since $f(q)$ is a concave function of q over the feasible region, these conditions are also sufficient for an interior maximum. In the following sections we will characterize the conditions for a maximum on the boundary of the feasible region, with some quantities equal to zero.

where $\nabla f(q_t)$ is the vector of partial derivatives $f_i(q_t) \equiv \partial f(q_t)/\partial q_i$ evaluated at q_t, and λ_t is the Lagrange multiplier on the budget constraint. Take the inner product of both sides of (21) with q_t and solve the resulting equation for $\lambda_t = q_t \cdot \nabla f(q_t)/p_t \cdot q_t = q_t \cdot \nabla f(q_t)/E_t$ where we have used (21). Euler's Theorem on homogeneous functions implies that $q_t \cdot \nabla f(q_t) = f(q_t)$ and so $\lambda_t = f(q_t)/E_t$. Using this result in equation (21), we obtain the first-order condition:

(22) $$\nabla f(q_t)/f(q_t) = p_t/E_t.$$

To simplify the notation in the rest of this section, we consider only N=2 commodities: good 1 is potentially new in period t, and good 2 represents all other expenditure. In addition, for this section we also scale the utility level so that it equals expenditure for period t:

(23) $$f(q_{1t}, q_{2t}) = E_t.$$

It follows that the first-order condition (22) becomes $\nabla f(q_t) = p_t$, and specializing to the case of two goods these conditions become:

(24) $$p_{it} = f_i(q_{1t},q_{2t}) \equiv \partial f(q_{1t},q_{2t})/\partial q_i\,, \qquad i=1,2.$$

We will derive a second-order Taylor series approximation to the utility loss if good 1 were removed, and compare that approximation to the Hausman measure defined by (17).

To make this calculation we reduce purchases of q_1 down to 0 in a linear fashion, holding prices fixed at their initial levels, p_{1t}, p_{2t}. 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 $q_1 = q_{1t}$ and ends up at $q_1 = 0$:

(25) $$q_2(q_1) \equiv [E_t - p_{1t}q_1]/p_{2t}.$$

The derivative of $q_2(q_1)$ evaluated at q_{1t} is $q_2'(q_{1t}) \equiv \partial q_2(q_{1t})/\partial q_1 = -(p_{1t}/p_{2t})$, a fact which we will use later. Define utility as a function of q_1 for $0 \le q_1 \le q_{1t}$, holding expenditures on the two commodities constant at E_t, as follows:

(26) $$U = u(q_1) \equiv f(q_1,q_2(q_1)) = f(q_1, [E_t - p_{1t}q_1]/p_{2t}).$$

We use the function $u(q_1)$ to measure the consumer loss of utility as we move q_1 from its original equilibrium level of q_{1t} to 0. Alternatively, the difference between the utility levels $u(q_{1t})$

and u(0) is the *gain of utility due to the appearance of product 1*, defined as a share of expenditure:

(27) $G_U \equiv [u(q_{1t}) - u(0)]/E_t$.

We express u(0) by a second-order Taylor series expansion around the point q_{1t}:

(28) $u(0) = u(q_1) + u'(q_1)(0 - q_1) + \frac{1}{2}u''(q_{1t})(0 - q_{1t})^2$.

The term $u'(q_{1t})$ is computed as:

(29) $u'(q_{1t}) = f_1(q_{1t},q_{2t}) + f_2(q_{1t},q_{2t})\partial q_2(q_{1t})/\partial q_1$, differentiating (26)

$= f_1(q_{1t},q_{2t}) + f_2(q_{1t},q_{2t})(-p_{1t}/p_{2t})$, differentiating (25)

$= 0$, using (24)

so this term vanishes as an envelope theorem result. It follows from (28) and (29) that a second-order approximation to the consumer gain from good 1 in (27) is,

(30) $G_H = -\frac{1}{2}u''(q_{1t})q_{1t}^2 / E_t$.

In Appendix B, we calculate the second derivative $u''(q_{1t})$ and we show that it is non-positive, so that the first term on the right of (30) is a non-negative gain. Furthermore, we define an inverse demand function, $p_1 = D_1(q_1)$ that is consistent with our model, i.e. holding other variables constant. The variables that Hausman holds constant are the utility level U_t and the price of product 2, p_{2t}. Endogenous variables are q_1, q_2 and E while the driving variable is p_1 which goes from p_{1t} to the reservation price $p_1^* = D_1(0)$ when q_1 goes from q_{1t} to 0. Because utility is held constant we regard this derived inverse demand curve as a Hicksian demand curve. We show that the slope of this inverse demand curve at q_{1t} equals $D'(q_{1t}) = u''(q_{1t})$ and so the inverse demand curve is convex if and only if $u'''(\tilde{q}_1) \geq 0$. Convexity of the demand curve implies that the Hausman approximation in (30) is a *lower bound* to the consumer gain from the introduction of good 1.

Substituting the result that $D'(q_{1t}) = u''(q_{1t})$ in (30), we have therefore established that the Hausman gain G_H due to the availability of good 1 is:

(31) $G_H = -\frac{1}{2}q_{1t}^2\ D'(q_{1t})/E_t$.

$= -\ \frac{1}{2}s_{1t}\ [D'(q_{1t})(q_{1t} / p_{1t})]$, where the final term appearing in brackets in (31) is the *elasticity of the constant-utility inverse demand curve.* In Appendix B we solve for this elasticity for particular utility functions, and in the CES case we find that it is precisely the inverse of the price elasticity of the Hicksian demand curve $\eta_{1t}\big|_U$, as shown in (7). More generally, we likewise expect that $[D'(q_{1t})(q_{1t} / p_{1t})]$ equals the inverse of $\eta_{1t}\big|_U$ whenever the Hicksian demand is well-behaved and differentiable. Our results in this section are therefore an alternative proof of the Hausman approximation in (17), but we have obtained these results even in cases where the Hicksian demand elasticity does not exist and instead the *inverse* demand functions are well-behaved and differentiable. This result will be very useful as we explore a quadratic utility function in the next section.

3.3 Konüs-Byushgens-Fisher (KBF) Utility Function

The functional form for the consumer's utility function f(q) that we will consider next is the following quadratic form:[11]

$$U = f(q) = (q^T A q)^{1/2}, \tag{32}$$

where the N by N matrix $A \equiv [a_{ik}]$ is symmetric (so that $A^T = A$) and thus has N(N+1)/2 unknown a_{ik} 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.[12] These conditions ensure that the utility function has indifference curves with the correct curvature.

Konüs and Byushgens (1926) showed that the Fisher (1922) "ideal" quantity index $Q_F(p_{t-1},p_t,q_{t-1},q_t) \equiv [(p_{t-1}\cdot q_t /p_{t-1}\cdot q_{t-1})(p_t\cdot q_t/ p_t\cdot q_{t-1})]^{1/2}$ is exactly equal to the aggregate utility ratio $f(q_1)/f(q_0)$, provided that the consumer maximizes the utility function defined by (32) in periods t-1 and t, where p_{t-1} and p_t are the price vectors with chosen quantities q_{t-1} and q_t. Diewert (1976) elaborated on this result by proving that the utility function defined by (32) 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

[11] 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.

[12] Diewert and Hill (2010) show that these conditions are sufficient to imply that the utility function defined by (32) is positive, increasing, linearly homogeneous and concave over the regularity region $S \equiv \{q: q >> 0_N \text{ and } Aq >> 0_N\}$.

utility ratio for the quadratic functional form defined by (32), he labelled the Fisher quantity index as a *superlative index* and we shall call (32) the *KBF functional form*.

Assume that all products are available in period t and consumers face the positive prices $p_t >> 0_N$. The first order conditions (22) to maximize the utility function in (32) become:

$$(33) \qquad p_t = E_t A q_t / (q_t^T A q_t) .$$

While these are the conditions for an interior maximum with $q_t >> 0_N$, we can obtain the condition for a zero optimal quantity $q_{it}=0$ if we impose that value on the right of (33) and then define the left-hand side for good i as the reservation price p_{it}^*. Then for all prices $p_{it} \geq p_{it}^*$, the consumer will optimally choose $q_{it} = 0$. We see that an advantage of the quadratic functional form is that the corresponding reservation price can be calculated very easily from (33), for any good where the quantity happens to equal 0 in the period under consideration.

In order to characterize demand, it is useful to work with the expenditure function. Assume for the moment that the matrix is of full rank, and denote $A^* = A^{-1}$. Then the minimum expenditure to obtain one unit of utility when the optimal $q_t >> 0_N$ is,

$$(34) \qquad e(p_t) = (p_t^T A^* p_t)^{1/2},$$

The total expenditure function is then $E_t = U_t e(p_t)$, and Hicksian demand is obtained by differentiating with respect to p_{it},

$$(35) \qquad q_{it}(p_t, U_t) = U_t \left[\frac{\sum_{n=1}^{N} a_{in}^* p_{nt}}{(p_t^T A^* p_t)^{1/2}} \right], \qquad i = 1,\ldots,N,$$

where a_{in}^* are the elements of A^*. Differentiating $-\ln q_{it}$ with respect to $\ln p_{it}$, we obtain the (positive) Hicksian elasticity,

$$(36) \qquad \eta_{it}\big|_U \equiv -\frac{\partial \ln q_{it}}{\partial \ln p_{it}}\bigg|_U = \frac{-a_{ii}^* p_{it}}{\sum_{n=1}^{N} a_{in}^* p_{nt}} + \frac{p_{it} \sum_{n=1}^{N} a_{in}^* p_{nt}}{p_t^T A^* p_t} = \frac{-a_{ii}^* p_{it}}{\sum_{n=1}^{N} a_{in}^* p_{nt}} + s_{it},$$

where s_{it} is the share of expenditure on good i. Notice that the denominator of the first ratio on the right of (36) must be positive to obtain positive demand in (35), but it aproaches zero as the quantity q_{it} approaches zero in a neighborhood of the reservation price as $p_{it} \rightarrow p_{it}^*$ and $q_{it} \rightarrow 0$.

Because the share then approaches zero, it follows that the Hicksian elasticity of demand in (36) remains positive if and only if $a^*_{ii} < 0,\ i = 1,\ldots,N$, which we assume is the case.

The fact that the KBF utility function has finite reservation prices suggests that it lies in-between the demand curves for the CES utility function (which have infinite reservation prices) and the linear approximation illustrated in Figure 1. That conjecture can be established more formally, as we show in Appendix C. We compute the second derivatives of the Hicksian demand curves for the quadratic utility function and show that so long as the demand curve is downward sloping, then it will be convex. In Appendix C we also compare the second derivative of the demand curve in the KBF case with that obtained in the CES case. Provided that the first derivatives of the demand curves are equal at the point of consumption (p_{it}, q_{it}), and that the expenditure share satisfies $s_{it} < 0.5$, then the second derivative of the CES Hicksian demand curves will *exceed* the second derivatives of those quadratic demand curves. This means that the demand curves for the quadratic utility function lie *in-between* the constant-elasticity demand curves considered in the previous section and the straight-line Hausman approximation.[13]

Using the expenditure function (34) with coefficients $A^* = A^{-1}$, where A is the matrix of coefficients for the direct utility function in (32), requires that the matrix A has full rank so that it is invertible. It is quite possible that A can have less than full rank, however, which means that there are certain goods in the utility function (or linear combinations of goods) that are perfect substitutes with other goods (or their combinations). In that case, at certain prices the demand for goods will not be uniquely determined, so we cannot work with demand as a function of prices or with the expenditure function. Instead, it makes sense to go back to the utility function in (32) and work with the *inverse demand functions* which are defined by (33), where prices (on the left) are a function of quantities and expenditure (on the right). The matrix of coefficients A will be of less than full rank in our empirical application of the KBF utility function, as we shall explain in sections 4.6 and 4.7, so we shall use the inverse demand functions in (33) for estimation. Fortunately, even in this case we can define a constant-utility Hicksian inverse demand curve, as we denoted by $p_{1t} = D(q_{1t})$ in section 3.2. Then our analysis of the Hausman approximation in that section continues to hold. Indeed, we show in Appendix B that in this case the elasticity of

---

[13] While we formally establish this result in Appendix C in a neighborhood of the consumption point, we expect that it will hold for all prices up to the reservation price, which is finite for the quadratic demand curves but infinite for the CES demand curve.

the inverse demand curve is:

(37) $$\frac{\partial \ln D_1(q_{1t})}{\partial \ln q_{1t}} = \frac{s_{1t}}{(1-s_{1t})^2}\left(\frac{a_{11}}{p_1^2}-1\right),$$

which can be used in (31) to obtain the Hausman approximation to the gain from good 1 in the KBF case:

(38) $$G_{H,KBF} = -\frac{1}{2}\left(\frac{s_{1t}}{1-s_{1t}}\right)^2\left(\frac{a_{11}}{p_1^2}-1\right).$$

4. Empirical Illustration using CES and KBF Utility Functions

4.1 Scanner Data for Sales of Frozen Juice

We use the data from Store Number 5[14] 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 and quadratic utility functions explained in the previous two sections. The micro data from the University of Chicago (2013) 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 Appendix A, the "monthly" data for quantities sold and the corresponding unit value prices for the 19 products are listed in Tables A1 and A2.[15] There were no sales of Products 2 and 4 for month 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, $s_{it} \equiv p_{it}q_{it}/\sum_{n=1}^{19} p_{nt}q_{nt}$, were computed for products i = 1,...,19

[14] This store is located in a North-East suburb of Chicago.

[15] In what follows, we will describe our 4 week "months" as months.

and months t = 1,...,39. 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%. There is tremendous volatility in product prices, quantities and sales shares for both the best selling and least popular products.

In the following sections, we will use this data set in order to estimate the elasticity of substitution σ for the CES utility and unit-expenditure functions, making differing assumptions on the errors underlying the price and expenditure share data.

4.2. Estimation of the CES Utility Function with Error in Prices

In this section and the next, we will use *double differencing approach* that was introduced by Feenstra (1994) to estimate the elasticity of substitution. His method requires that product shares be positive in all periods. In order to implement his method, we drop the products that are not present in all periods. Thus, we drop products 2, 4 and 12 from our list of 19 frozen juice products since products 2 and 4 were not present in months 1-8 and product 12 was not present in months 20-22. Thus in our particular application, the number of always present products in our sample will equal 16. We also renumber our products so that the original Product 13 becomes the Nth product in this section. This product had the largest average sales share. If we assume that purchasers are choosing all 19 products by maximizing CES preferences over the 19 products, then this assumption implies that they are also maximizing CES preferences restricted to the always present 16 products.

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

- q_{it} is the observed amount of product i sold in period t;
- p_{it} is the observed unit value price of product i sold in period t and
- s_{it} is the observed share of sales of product i in period t that is constructed using the quantities q_{it} and the corresponding observed unit value prices p_{it}.

In our particular application, N = 16 and T = 39. We aggregated over weekly unit values to construct pseudo-monthly unit value prices. Since there was price change within the monthly time period, the observed monthly 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 monthly quantities sold during the period do not suffer from this time aggregation bias. In this section, we will allow for measurement error in the log shares due to the measurement error in prices and, in the next section, we shall also add measurement error in the share due to changing tastes.

Our goal is to estimate the elasticity of substitution for a CES direct utility function (3) that was discussed in section 3.1 above. The system of share equations that corresponds to this consumer utility function was shown as (6) when expressed as a function of prices. An alternative expression for the shares as a function of quantities can be obtained by denoting the CES utility function by $f(q_t)$ and using the first-order condition (22) for good i multiplied by q_{it} to obtain the share equations:

(39) $$s_{it} \equiv \frac{p_{it}q_{it}}{E_t} = \frac{a_i q_{it}^{(\sigma-1)/\sigma}}{\sum_{n\in I_t} a_n q_{nt}^{(\sigma-1)/\sigma}}, \qquad i = 1,...,N;\ t = 1,...,T,$$

where T = 39 and N = 16. This system of share equations corresponds to the consumers' system of inverse demand equations for always present products, which give monthly unit value prices as functions of quantities purchased. We take natural logarithms of both sides of the equations in (39) and add error terms u_{it} to reflect the measurement error in prices and therefore in shares,

(40) $$\ln s_{it} = \ln a_i + \tfrac{(\sigma-1)}{\sigma}\ln q_{it} - \sum_{n=1}^{N} a_n q_{nt}^{(\sigma-1)/\sigma} + u_{it}, \qquad i = 1,...,N;\ t = 1,...,T$$

where by assumption the q_{it} are measured without error and the error terms u_{it} 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 (40) are the positive parameters a_i and the elasticity of substitution $\sigma > 1$.

The Feenstra double-differenced variables are defined in two stages. First, we difference the *logarithms* of the s_{it} with respect to time; i.e., define $\Delta\ln s_{it}$ as follows:

(41) $$\Delta\ln s_{it} \equiv \ln(s_{it}) - \ln(s_{it-1}), \qquad i = 1,...,N;\ t = 2,3,...,T.$$

Now pick product N as the numeraire product and difference the $\Delta\ln s_{it}$ with respect to product N, giving rise to the following *double differenced log variable*, $\Delta^2\ln s_{it}$:

(42) $$\Delta^2\ln s_{it} \equiv \Delta\ln s_{it} - \Delta\ln s_{Nt}, \qquad i = 1,...,N-1;\ t = 2,3,...,T$$
$$= \ln(s_{nt}) - \ln(s_{nt-1}) - \ln(s_{Nt}) - \ln(s_{Nt-1}).$$

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

(43) $\Delta^2 \ln q_{it} \equiv \Delta \ln q_{it} - \Delta \ln q_{Nt}$; $i = 1,...,N-1;\ t = 2,3,...,T$

$= \ln(q_{nit}) - \ln(q_{it-1}) - \ln(q_{Nt}) - \ln(q_{Nt-1})$.

Finally, define the *double-differenced error variables* $\Delta^2 u_{it}$ as follows:

(44) $\Delta^2 u_{it} \equiv u_{it} - u_{it-1} - u_{Nt} + u_{Nt-1}$, $i = 1,...,N-1;\ t = 2,3,...,T$.

Using definitions (41)-(44) and equation (40), it can be verified that the double-differenced log shares $\Delta^2 \ln s_{it}$ satisfy the following system of $(N-1)(T-1)$ estimating equations:

(45) $\Delta^2 \ln s_{it} = \frac{(\sigma-1)}{\sigma} \Delta^2 \ln q_{it} + \Delta^2 u_{it}$, $i = 1,...,N-1;\ t = 2,3,...,T$

where the new residuals, $\Delta^2 u_{it}$, have means 0 and a constant covariance matrix with 0 covariances for observations which are separated by two or more time periods. Thus we have a system of linear estimating equations with only one unknown parameter across all equations, namely, σ. This is almost[16] 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 (45) which we estimated using the NL system command in Shazam.[17] The resulting estimate for $(\sigma-1)/\sigma$ was 0.865 (with a standard error of 0.007) and thus the corresponding estimated σ is equal to 7.40. The standard error on $(\sigma-1)/\sigma$ was tiny using the present regression results so σ was very accurately determined using this method. The equation-by-equation R^2 for the 15 products $i = 1,\ldots,N-1$ were as follows: 0.994, 0.990, 0.991, 0.991, 0.987, 0.982, 0.962, 0.956, 0.986, 0.991, 0.993, 0.994, 0.991, 0.992 and 0.989. The average R^2 is 0.986, 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 $(N-1)(T-1) = 570$ observations.[18] This double differencing method for estimating the elasticity of substitution worked much better than any other method that we tried.[19]

---

[16] The variance covariance structure is not quite classical due to the correlation of residuals between adjacent time periods. We did not take this correlation into account in our empirical estimation of this system of estimating equations; i.e., we just used a standard systems nonlinear regression package that assumed intertemporal independence of the error terms.

[17] See White (2004).

[18] 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.

[19] See our working paper, Diewert and Feenstra (2019), for other methods.

4.3. Estimation of the CES Utility Function with Errors in Prices and Tastes

In the previous section, the error terms in equations (40) and (45) reflected time aggregation errors in forming the monthly unit value prices, which we assumed were reflected in the expenditure share but not in the quantities. But in reality, errors in the unit values can arise due to inaccurate measurement of quantities themselves, creating inaccurate unit values when dividing expenditure on a barcode item by the quantity. Such measurement error in quantities is therefore reflected in the unit values but not in the expenditure shares. We could expect, however, that other errors in expenditure shares could arise because our assumed CES functional form for the consumer's utility function may not be correct. One way to model that situation is to allow the consumer taste parameters to change over time, while retaining the rest of the CES structure. In that case we obtain an error in the share equations due to taste change. However, we will assume that the error in shares due to taste change is uncorrelated with the measurement error in prices.

We now make that measurement error in prices explicit by assuming that the natural log of the *unit values* p_{it} are related to the *true prices* ρ_{it} by:

(46) $$\ln p_{it} = \ln \rho_{it} + u_{it}\,, \qquad i = 1,...,N;\ t = 1,...,T$$

where u_{it} is the measurement error in the log unit values, which is assumed to be uncorrelated with the logarithms of the true prices, $\ln\rho_{it}$.

Consider the share equations (6) but replace the unit value prices p_{it} by the true prices ρ_{it}. In addition, we will allow the taste parameters b_i appearing in (6) to vary over time, and so we replace them by b_{it}, i = 1,…,N. We assume that the taste parameters have an error term:

(47) $$\ln b_{it} = \ln b_i + \varepsilon_{it}\,. \qquad i = 1,...,N;\ t = 1,...,T$$

With these changes to the share equation (6), we take natural logarithms to obtain:

(48) $$\ln s_{it} = \ln b_i + (1-\sigma)\ln \rho_{it} - \ln\left[\sum\nolimits_{n=1}^{N} b_{nt}\rho_{nt}^{(1-\sigma)}\right] + \varepsilon_{it}\,,\ \ i = 1,...,N;\ t = 1,...,T$$

As explained, the error term ε_{it} can arise due to movements in the share variable that does not reflect CES behavior with fixed taste parameters 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.[20]

We will make the usual assumption that the errors in the share equations (48) are uncorrelated with the "true" prices ρ_{it} in (46), i.e. these "true" prices are exogenous to the consumer.[21] Furthermore, we shall assume that the measurement errors u_{it} in the unit values is uncorrelated with the errors ε_{it} in the share variables. 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. Once again, we rely on the double-differencing method due to Feenstra (1994). As in the previous section, for any variable x we define the double difference over time and with respect to the product N as $\Delta^2 \ln x_{it} \equiv \Delta \ln x_{it} - \Delta \ln x_{Nt}$.

The share equation in (48) is simplified by taking first-differences over time to eliminate the nuisance parameter b_i, and then by taking an additional difference with respect to a reference product N to eliminate the summation term:[22]

(49) $\Delta^2 \ln s_{it} \equiv \Delta \ln s_{it} - \Delta \ln s_{Nt}$ $\quad$ $i = 1,...,N\text{-}1;\ t = 2,...,T$

$= (1-\sigma)\ \Delta^2 \ln \rho_{it} + \Delta^2 \varepsilon_{it},$ $\quad$ from (48)

$= (1-\sigma)\ \Delta^2 \ln p_{it} - (1-\sigma)\ \Delta^2 u_{it} + \Delta^2 \varepsilon_{it},$ $\quad$ from (46).

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 N) as:

(50) $M_i(x,y) \equiv (1/T)\ (\Sigma_t\, \Delta^2 x_{it}\ \Delta^2 y_{it})\ .$ $\quad$ $i=1,...,N\text{-}1,$

If x = y then the cross moment defined in (50) 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 (50) *for each* of the products i=1,…,N-1, so then using the panel nature of the dataset we have a cross-section of such moments for i = 1,...,N-1.

---

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

[21] 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.

[22] We assume that the reference product N 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.

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

(51) $\mathbf{E}[M_i(\varepsilon,\ln\rho)]=0$, $\mathbf{E}[M_i(u,\ln\rho)]=0$ and $\mathbf{E}[M_i(\varepsilon,u)]=0$, $i=1,...,N-1$,

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.

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

$$
\begin{aligned}
(52)\quad M_i(\varepsilon,u) &\equiv (1/T)\,[\Sigma_t\,(\Delta^2\varepsilon_{it}\,\Delta^2 u_{it})] \\
&= (1/T)[\Sigma_t\,\Delta^2\varepsilon_{it}\,(\Delta^2\ln p_{it} - \Delta^2\ln\rho_{it})] \\
&= (1/T)\,[\Sigma_t\,(\Delta^2\ln s_{it} - (1-\sigma)\Delta^2\ln p_{it} + (1-\sigma)\Delta^2 u_{it})\,\Delta^2\ln p_{it}] - M_i(\varepsilon,\ln\rho) \\
&= M_i(\ln s,\ln p) - (1-\sigma)M_i(\ln p,\ln p) + (1-\sigma)M_i(u,\ln p) - M_i(\varepsilon,\ln\rho) \\
&= M_i(\ln s,\ln p) - (1-\sigma)M_i(\ln p,\ln p) + (1-\sigma)M_i(u,\ln\rho) + (1-\sigma)M_i(u,u) - M_i(\varepsilon,\ln\rho),
\end{aligned}
$$

where the second line uses (48) to express the measurement error $\Delta^2 u_{it}$; the third follows by re-expressing that error $\Delta^2\varepsilon_{it}$ in full using (49), and combining the share error $\Delta^2\varepsilon_{it}$ with the term $\Delta^2\ln\rho_{it}$ to obtain $M_i(\varepsilon,\ln\rho)$; 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 (46). It is convenient to rewrite (52) as,

$$(53)\quad M_i(\ln p,\ln p) = \frac{1}{(1-\sigma)} M_i(\ln s,\ln p) + M_i(u,u) + \text{Error}_i, \qquad \text{for } i=1,\ldots,N,\ i\neq N,$$

where Error_i is defined as follows:

$$(54)\quad \text{Error}_i \equiv M_i(u,\ln\rho) - \frac{1}{(1-\sigma)}\,[M_i(\varepsilon,\ln\rho) + M_i(\varepsilon,u)].$$

What we have obtained in (53) is a simple linear regression involving moments of the data, which can be run over the products $i=1,\ldots,N-1$. The error in this regression, defined in (54), consists of a sum of the moment conditions that we have discussed in (51) and which we assumed are zero in expected value. It follows that minimizing the squared error by running OLS on (54) is a generalized method of moments estimator.

Examining regression (53) more closely, the dependent variable is the second moment of the log unit values (differenced with respect to time and with respect to product N). The first term on the right is the cross moment of the market shares and unit values, and the coefficient of this term is $1/(1-\sigma)$.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.

Running the OLS regression for the frozen juice data results in $\sigma = 7.99$ for weekly data, and $\sigma = 5.99$ 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 we also see that the estimate of $\sigma = 7.40$ from the monthly data in the previous section – using quantity on the right of the share equation as in (45) – neatly lies in-between the weekly and monthly consistent estimates obtained in this section. Accordingly, we are comfortable continuing to use the estimate of $\sigma = 7.40$ when we compute the gains and losses from new and disappearing varieties of frozen juice, as we do in the next section.

4.4 Estimation of the Changes in the CES CPI Due to Changing Product Availability

Recall that the Feenstra methodology to measure the exact CES price index used the Sato-Vatio $P_{SV}(I)$ in (11), expressed over the common products, and multiplied that index by the terms $(\lambda_t / \lambda_{t-1})^{1/(\sigma-1)}$ in (14) that captures new and disappearing products. This term will differ from 1 if the available products change from the previous period. For our dataset, the term λ_t is less than unity for months 9 (products 2 and 4 become available), 11 (product 12 becomes available), and 23 (product 12 again becomes available). The term λ_{t-1} is greater than unity for months 10 (product 12 becomes unavailable) and 20 (product 12 again becomes unavailable). Computing $(\lambda_t / \lambda_{t-1})^{1/(\sigma-1)}$ using our estimate of $\sigma = 7.403$ gives the results shown in the third column of Table 2. In the final column, we can *invert* this term to obtain the gain in CES utility (or loss if less than one) due to the availability of goods:[23]

$$G_{CES} = (\lambda_t / \lambda_{t-1})^{-1/(\sigma-1)}. \tag{55}$$

[23] The CES gain in (55) is slightly more general than the compensating variation gain in (11) for a single new good.

Table 2: Changes in the Price Level and CES Gains due to the Availability of Products, σ = 7.40

	Availability	$(\lambda_t / \lambda_{t-1})^{1/(\sigma-1)}$	G_{CES}
9	2 and 4 new	0.9928	1.0073
10	12 disappears	1.0036	0.9964
11	12 reappears	0.9957	1.0043
20	12 disappears	1.0039	0.9962
23	12 reappears	0.9969	1.0031
Cumulative Gain		0.9928	1.0073

Recall that in month 9, products 2 and 4 make their appearance, and Table 2 tells us that the effect of this increase in variety is to lower the price level and increase utility for month 9 by 0.73 percentage points. In month 10 when product 12 disappears from the store, this has the effect of increasing the price level and lowering utility by 0.36 percentage points. That product comes in and out of the dataset, and the overall effect on the price level of the changes in the availability of products is equal to 0.9928×1.0036×0.9957×1.0039× 0.9969 = 0.9928, for a decrease in the price level and increase in utility over the sample period of 0.73 percentage points. Notice that this overall effect just reflects the introduction of products 2 and 4 in month 9, since the net impact of the disappearance and reappearance of product 12 *cancels out* when cumulated. That cancelling of the impact of availability of product 12 is a highly desirable feature of these CES results, but it is not a necessary outcome because it depends on the shares of product 12: it just so happens that these shares are nearly equal when it exits and re-enters, leading to zero net impact. We will explore in later sections whether this desirable result continues to hold with other functional forms for utility.

These results in Table 2 are our first estimates of the gains from increased product availability in our frozen juice data. While they are promising results, as we mentioned in section 1, there are two potential problems with the Feenstra methodology: (i) the CES functional form is not fully flexible; and (ii) the reservation prices which induce consumers to demand 0 units of products that are not available in a period are infinite, which *a priori* seems implausible. Thus in the following section, we will introduce a flexible functional form that will generate finite reservation prices for unavailable products, and hence will provide an alternative methodology for measuring the net benefits of new and disappearing products.

4.5 Estimation of the KBF Utility Function

The quadratic or KBF utility function was introduced in section 3.3, above. Multiply both sides of equation i in (33) by q_{it} and $p_t \cdot q_t = E_t$, we obtain the following *system of inverse demand share equations*:

$$(56)\qquad s_{it} \equiv \frac{p_{it}q_{it}}{p_t \cdot q_t} = \frac{q_{it}\sum_{n=1}^{N} a_{in}q_{nt}}{q_t^T A q_t}, \qquad i = 1,\ldots,N,$$

where a_{in} is the element of A that is in row i and column n for i, n = 1,...,N. These equations will form the basis for our system of estimating equations in this and the following section. Note that they are nonlinear equations in the unknown parameters a_{ik}. It turns out to be useful to reparameterize the A matrix as follows:

$$(57)\qquad A = bb^T + B;\ b >> 0_N\ ;\ B = B^T\ ;\ B \text{ is negative semidefinite; } Bq^* = 0_N,$$

where q^* is a positive vector. The vector $b^T \equiv [b_1,\ldots,b_N]$ 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 (32) boils down to the *linear utility function* model, which means that the goods are perfect substitutes for each other. We will add the matrix B into our estimation as described below, but restrict it to be of less than full rank, so the matrix A will also be of less than full rank. As anticipated earlier (see the end of section 3.3), this means that A cannot be inverted and it will be necessary to work with the inverse demand curves of the KBF system, rather than the expenditure function or the associated Hicksian or Marshallian demand curves.

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.[24] Write C as $[c^1,c^2,...,c^N]$ where c^k is a column vector for k = 1,...,N. If C is lower triangular, then the first k−1 elements of c^k are equal to 0, k = 2,3,...,N. Thus we have the following representation for B:

$$(58) \qquad B = -CC^T = -\sum_{k=1}^{19} c^k c^{kT}$$

where we impose the following restrictions on the vectors c^k in order to impose the restrictions $Bq^* = 0_N$ on B:[25]

$$(59) \qquad c^{kT}q^* = 0\,; \quad k = 1,....,N.$$

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 estimated the parameters in the utility function $f(q) = (q^TAq)^{1/2}$ in stages. In the first stage, we estimated 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 starting $c^1 \equiv 0_N$.[26] In the third stage, we estimate $f(q) = (q^T[bb^T - c^1c^{1T} - c^2c^{2T}]q)^{1/2}$ where $c^{1T} \equiv [c_1^1,c_2^1,...,c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0,c_2^2,...,c_N^2]$ 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 $c^{1T} \equiv [c_1^1,c_2^1,...,c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0,c_2^2,...,c_N^2]$, $c^{2T}q^* = 0$, $c^{3T} \equiv [0,0,c_3^3,...,c_N^3]$ and $c^{3T}q^* = 0$. At each stage, the log likelihood will generally increase.[27] 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

---

[24] $C = [c_{nk}]$ 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.

[25] The restriction that C be lower 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. For additional materials on the properties of the KBF functional form, see Diewert (2018).

[26] We also use the constraint $c^{1T}q^*$ to eliminate one of the c_n^1 from the nonlinear regression.

[27] 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.

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.

4.6 The Estimation of KBF Preferences Using Share Equations

The estimating equations for the KBF utility function are the following stochastic version of the share equations (56) above:

$$(60) \qquad s_{it} = q_{it} \textstyle\sum_{j=1}^{19} a_{ij}q_{jt}/[\sum_{n=1}^{19} \sum_{m=1}^{19} a_{nm}q_{nt}q_{mt}] + \varepsilon_{it} \qquad t = 1,...,39; i = 1,...,19$$

where the error term vectors $\varepsilon_t^T = [\varepsilon_{1t},...,\varepsilon_{19t}]$ 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.[28] Because the shares in (60) sum to unity over the i=1,…,19 products for each t, and likewise the term on the right-hand side without the error sums to unity, it follows that the error terms ε_{it} sum to zero over the over the i=1,…,19 products for each t. So the variance-covariance matrix Σ of the errors is singular and we drop the last equation for product 19. In order to identify the parameters, the normalization $b_{19} = 1$ can be imposed. We also choose the reference vector $q^* = 1_{19}$ as a vector of ones.

It is possible to estimate (60) as a system of 18 equations, which we attempted in our working paper (see Diewert and Feenstra, 2017). But we found that for estimation, it is more convenient to stack the 18 estimating share equations listed in equations (60) into a single equation. In the first model, we estimated the 18 unknown parameters in the linear utility function with $A = bb^T$, where $b^T \equiv [b_1,b_2,...,b_{19}]$ and $b_{19} = 1$, using the single equation Nonlinear command in Shazam. The final log likelihood was 2379.4 and the R^2 was 0.982.

An advantage of the single equation approach is that we can now easily drop the 20 observations where the product was missing.[29] Thus for our next model, we dropped the 20 observations for products 2, 4 and 12 for the months when these products were missing, so the number of observations for this new model is equal to (36×18) − 20 = 682. We found that the parameter estimates for this new model were exactly the same as the corresponding parameter

[28] This is a slightly incorrect econometric specification since ε_{it} will automatically equal 0 if product i is not present during month t.

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

estimates that we obtained when using all the observations. However, the new log likelihood decreased to 2301.7 and the new R^2 decreased slightly to 0.981. In the models that follow, we continued to drop the 20 observations that correspond to the months when the products were missing.

In our next model, we set $A = bb^T - c^1c^{1T}$ with the normalizations $b_{19} = 1$ and $c^1_{19} = -\sum_{n=1}^{18} c^1_n$. We used the final estimates for the components of the b vector from the previous model as starting coefficient values for this model and we used $c^1_n = 0.001$ for n = 1,...,18 as starting values for the components of the c vector. The final log likelihood for this model was 2445.9, an increase of 144.2 for adding 18 new parameters to the previous model, and the R^2 increased to 0.988.

We continued on adding new columns c^k one at a time to the substitution matrix, using the finishing coefficient values from the previous nonlinear regression as starting values for the next nonlinear regression. Our final model added the column vector c^4 to the previous A matrix. Thus we had $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T}$ with $c^{4T} = [0,0,0,c^4_4,...,c^4_{19}]$ and the additional normalization $c^4_{19} = -\sum_{n=4}^{18} c^4_n$. As usual, we used the final estimates for the components of the b, c^1, c^2 and c^3 vectors from the previous model as starting coefficient values for this model and we used $c^4_n = 0.001$ for n = 4,...,18 as starting values for the nonzero components of the c^4 vector. The final log likelihood for this model was 2629.2, an increase of 14.7 for adding 15 new parameters to the previous model's parameters. Thus the increase in log likelihood is now less than one per additional parameter. The single equation R^2 increased to 0.992. The comparable R^2 for each separate product share equation were as follows:[30] 0.986, 0.993, 0.977, 0.985, 0.981, 0.954, 0.976, 0.858, 0.976, 0.969, 0.892, 0.928, 0.991, 0.920, 0.987, 0.957, 0.911 and 0.965. The average R^2 was 0.956, which is a relatively high average when estimating share equations.

Since the present model estimated 84 unknown parameters and we had only 682 degrees of freedom, we had only about 8 degrees of freedom per parameter at this stage. Moreover, the increase in log likelihood over the previous model was relatively small. Thus, we decided to stop

[30] 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.

adding columns to the C matrix at this point. With the estimated b and c vectors (denote them as $\hat{b}$ and $\hat{c}^k$ for k = 1,...,4), form the estimated A matrix as $\hat{A} \equiv \hat{b}\hat{b}^T - \hat{c}^1\hat{c}^{1T} - \hat{c}^2\hat{c}^{2T} - \hat{c}^3\hat{c}^{3T} - \hat{c}^4\hat{c}^{4T}$, and denote the ij element of $\hat{A}$ as $\hat{a}_{ij}$ for i,j = 1,...,19. The *expenditure share* for product i in month t is s_{it} defined as follows:

$$(61) \qquad s_{it}^* \equiv q_{it} \sum_{j=1}^{19} \hat{a}_{ij} q_{jt} / [\sum_{n=1}^{19} \sum_{m=1}^{19} \hat{a}_{nm} q_{nt} q_{mt}] , \qquad t = 1,...,39; i = 1,...,19.$$

The *predicted price* for product i in month t is defined using (33) as:

$$(62) \qquad p_{it}^* \equiv E_t \sum_{j=1}^{19} \hat{a}_{ij} q_{jt} / [\sum_{n=1}^{19} \sum_{m=1}^{19} \hat{a}_{nm} q_{nt} q_{mt}] , \qquad t = 1,...,39; i = 1,...,19$$

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 (62) 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 product prices were as follows:[31] 0.757, 0.823, 0.866, 0.897, 0.903, 0.758, 0.866, 0.002, 0.252, 0.122, 0.000, 0.001, 0.913, 0.672, 0.461, 0.724, 0.543, 0.815 and 0.423. The average R^2 is only 0.568 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 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.

[31] 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 are overstated.

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 that 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.

4.7 The Estimation of KBF Preferences Using Price Equations

Our next system of estimating equations used prices as the dependent variables, as was shown in (33):

$$p_{it} \equiv E_t \textstyle\sum_{j=1}^{19} a_{ij}q_{jt}/[\sum_{n=1}^{19} \sum_{m=1}^{19} a_{nm}q_{nt}q_{mt}] + \varepsilon_{it}\,, \quad t = 1,\ldots,39;\ i = 1,\ldots,18 \qquad (63)$$

where the A matrix was 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 the last model in the previous section. We stack up the estimating equations defined by (63) 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 the previous model as starting coefficient values for the present model. The initial log likelihood of our new model using these starting values for the coefficients was 415.6. The final log likelihood for this model was 518.9, an increase of 103.5. Thus switching from having shares to having prices as the dependent variables did significantly change our estimates. The single equation R^2 was 0.945. We used our estimated coefficients to form predicted prices p_{it}^* using equations (63) 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:[32] 0.830, 0.862, 0.900, 0.916, 0.899, 0.832, 0.913, 0.035, 0.244, 0.275, 0.024, 0.007, 0.870, 0.695, 0.421, 0.808, 0.618, 0.852 and 0.287. The average R^2 was 0.594. Of particular concern is product 12, which comes in and out of the sample, and which has a very low R^2 of only 0.007

Since the predicted prices are still not very close to the actual prices, we decided to press on and estimate a new model which added another rank 1 substitution matrix to the substitution

[32] See notes 30 and 31.

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 $c_{19}^5 = -\sum_{n=5}^{18} c_n^5$ We used the final estimates for the components of the b, c^1, c^2, c^3 and c^4 vectors from the previous model as starting coefficient values for the present model 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 518.9. The final log likelihood for this model was 550.3, an increase of 31.4. The single equation R^2 was 0.950.

Since the increase in log likelihood for the rank 5 substitution matrix over the previous rank 4 substitution matrix was fairly large, we decided to add another rank 1 matrix to the A matrix. Thus for our next model, we set $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T} - c^5c^{5T} - c^6c^{6T}$ where $c^{6T} = [0,0,0,0,c_6^6,...,c_{19}^6]$ with the additional normalization $c_{19}^6 = -\sum_{n=6}^{18} c_n^6$ We used the final estimates for the components of the b, c^1, c^2, c^3, , c^4 and c^5 vectors from the previous model as starting coefficient values for the new model along with $c_n^6 = 0.001$ for n = 6,7,...,18. The final log likelihood for this model was 568.9, an increase of 18.5. The single equation R^2 was 0.953. The present model had 111 unknown parameters that were estimated (plus a variance parameter). We had only 680 observations and 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 the rank 5 substitution matrix are listed below in Table 3.[33] The estimated b_n in Table 3 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 3, 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 $\hat{b}$ and $\hat{c}^k$ for k = 1,...,6), form the estimated A matrix as $\hat{A} \equiv \hat{b}\hat{b}^T - \hat{c}^1\hat{c}^{1T} - \hat{c}^2\hat{c}^{2T} - \hat{c}^3\hat{c}^{3T} - \hat{c}^4\hat{c}^{4T} - \hat{c}^5\hat{c}^{5T} - \hat{c}^6\hat{c}^{6T}$, and again denote the ij element of $\hat{A}$ as $\hat{a}_{ij}$ for i,j = 1,...,19. The *predicted price* for product i in month t is

[33] The standard errors for the estimated coefficients are equal to the coefficient estimate listed in Table 3 divided by the corresponding t statistic.

Table 3: Estimated Parameters for KBF Preferences

Coef	Estimate	t Stat	Coef	Estimate	t Stat	Coef	Estimate	t Stat
b_1	1.35	11.39	c_3^2	-0.08	-0.11	c_9^4	0.16	0.26
b_2	1.31	10.77	c_4^2	-0.71	-0.72	c_{10}^4	-0.03	-0.05
b_3	1.43	11.31	c_5^2	-0.10	-0.24	c_{11}^4	-0.61	-0.81
b_4	1.57	11.54	c_6^2	-0.64	-1.28	c_{12}^4	-1.59	-1.13
b_5	1.37	11.23	c_7^2	-0.61	-1.38	c_{13}^4	-0.23	-0.31
b_6	2.09	11.89	c_8^2	1.15	1.81	c_{14}^4	-0.16	-0.24
b_7	1.42	11.40	c_9^2	-0.39	-1.35	c_{15}^4	-0.67	-1.69
b_8	0.82	9.02	c_{10}^2	-0.54	-1.73	c_{16}^4	-0.22	-0.30
b_9	0.57	9.67	c_{11}^2	1.00	2.14	c_{17}^4	3.27	3.55
b_{10}	0.59	9.48	c_{12}^2	1.90	1.67	c_{18}^4	-0.35	-0.44
b_{11}	0.80	10.01	c_{13}^2	-0.46	-1.48	c_5^5	-0.06	-0.11
b_{12}	1.10	9.16	c_{14}^2	-0.73	-1.46	c_6^5	-0.04	-0.12
b_{13}	1.24	11.14	c_{15}^2	-0.32	-0.80	c_7^5	-0.10	-0.06
b_{14}	1.61	11.12	c_{16}^2	0.26	0.84	c_8^5	-0.25	-0.04
b_{15}	0.71	10.12	c_{17}^2	0.02	0.01	c_9^5	-0.62	-0.89
b_{16}	1.34	11.47	c_{18}^2	-0.50	-1.13	c_{10}^5	-0.56	-0.80
b_{17}	1.58	7.97	c_3^3	1.36	5.41	c_{11}^5	-0.11	-0.03
b_{18}	1.37	11.40	c_4^3	1.72	4.41	c_{12}^5	-0.31	-0.04
c_1^1	1.98	10.03	c_5^3	1.03	5.10	c_{13}^5	0.63	0.12
c_2^1	1.66	6.65	c_6^3	-0.43	-1.09	c_{14}^5	0.05	0.01
c_3^1	-0.25	-1.19	c_7^3	0.90	2.43	c_{15}^5	-0.08	-0.02
c_4^1	0.13	0.55	c_8^3	-0.46	-0.81	c_{16}^5	0.76	0.13
c_5^1	0.013	0.09	c_9^3	-0.01	-0.04	c_{17}^5	0.61	0.23
c_6^1	-0.01	-0.05	c_{10}^3	-0.08	-0.28	c_{18}^5	0.48	0.05
c_7^1	-0.38	-1.92	c_{11}^3	-0.59	-1.06	c_6^6	-0.01	-0.03
c_8^1	-0.43	-1.86	c_{12}^3	-0.14	-0.14	c_7^6	0.18	0.38
c_9^1	-0.02	-0.11	c_{13}^3	-0.02	-0.09	c_8^6	-0.76	-0.30
c_{10}^1	-0.28	-1.58	c_{14}^3	-0.45	-1.18	c_9^6	-0.08	-0.02
c_{11}^1	-0.96	-4.48	c_{15}^3	-0.46	-2.03	c_{10}^6	0.08	0.02
c_{12}^1	-0.88	-2.69	c_{16}^3	-0.01	-0.06	c_{11}^6	-0.44	-0.27
c_{13}^1	0.11	1.52	c_{17}^3	-2.16	-2.38	c_{12}^6	-0.95	-0.23
c_{14}^1	-0.22	-1.02	c_{18}^3	0.01	0.03	c_{13}^6	-0.60	-0.11
c_{15}^1	-0.13	-0.85	c_4^4	-0.50	-0.71	c_{14}^6	0.47	0.98
c_{16}^1	0.14	1.25	c_5^4	0.49	1.34	c_{15}^6	0.39	0.34
c_{17}^1	-0.68	-1.54	c_6^4	0.27	0.47	c_{16}^6	0.66	0.10
c_{18}^1	0.08	0.45	c_7^4	0.38	0.63	c_{17}^6	0.12	0.00
c_2^2	0.72	1.58	c_8^4	-0.11	-0.12	c_{18}^6	1.02	0.26

calculated as earlier in (62) but using the new $\hat{a}_{ij}$ estimates. The equation-by-equation R^2 that compares the predicted prices for the 19 products with the actual prices were as follows:[34] 0.827, 0.868, 0.900, 0.917, 0.896, 0.854, 0.905, 0.034, 0.328, 0.424, 0.052, 0.284, 0.865, 0.7280, 0.487, 0.814, 0.854, 0.848 and 0.321. The average R^2 was 0.642, which is a noticeable increase from the rank 4 model (average R^2=0.594), and now 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 (product 12 has R^2 = 0.284).[35]

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.

We can use these reservation prices in the calculation of exact price indexes for the KBF utility function. As noted earlier in section 3.3, the Fisher quantity index is exactly equal to the aggregate utility ratio for the KBF utility function in (32) provided that the quantities q_{t-1} and q_t are optimal for the prices p_{t-1} and p_t. Likewise, the Fisher price index defined by $P_F(p_{t-1},p_t,q_{t-1},q_t) \equiv [(p_t{\cdot}q_{t-1} /p_{t-1}{\cdot}q_{t-1})(p_t{\cdot}q_t/ p_{t-1}{\cdot}q_t)]^{1/2}$ is exactly equal to the ratio of expenditure functions in (34), $e(p_t)/e(p_{t-1})$, provided that quantities q_{t-1} and q_t minimize the expenditure needed to obtain utility of one at the prices p_{t-1} and p_t. Initially, we can compute these Fisher price indexes for our data by *ignoring* the products that are not available in two consecutive period t-1 and t, for t=2,…,39. We will refer to these indexes as the *Fisher maximum overlap price indexes,* denoted for simplicity by $P_{FM}(t-1,t)$ for t=2,…,39.

As a second calculation, we can make use of the reservation prices above for the *unavailable* products along with 0 quantities in that period, and we recompute the Fisher prices indexes while using these reservation prices. This procedure follows the suggestion of Hicks (1940), mentioned as the outset of our paper, for imputing the prices of unavailable products. We denote the *Fisher index with Hicksian reservation prices* by $P_{FH}(t-1,t)$ for t=2,…,39.

[34] See notes 30 and 31.
[35] 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.

A third Fisher index that we compute uses the *predicted prices for all products and all time periods* defined by equations (62). The predicted prices for unavailable products equal the reservation prices, of course, while for available products the predicted prices differ from actual prices due to the estimated error in the regression equation (63). Using these estimated prices for *all goods* ensures that the quantities used in the price index (including the 0 quantities for unavailable products) are *optimal* for those predicted prices. Denote the *Fisher index with predicted prices* by $P_F^*(t-1,t) \equiv [(p_t^* \cdot q_{t-1} / p_{t-1}^* \cdot q_{t-1})(p_t^* \cdot q_t / p_{t-1}^* \cdot q_t)]^{1/2}$ for t=2,…,39.

Feenstra's methodology for measuring the benefits and costs of changing product availability in the CES case makes use of a "maximum overlap" Sato-Vartia price index, which was denoted by $P_{SV}(I)$ and defined in (11) over the set of goods I that were available in periods t-1 and t. The result in (14) showed that by multiplying that maximum overlap index by the ratio $(\lambda_t / \lambda_{t-1})^{1/(\sigma-1)}$ we obtained the exact price index, which is lowered by the availability of new goods, and the CES gain in (55) was defined as the inverse of that ratio.

For the KBF utility function we can make a similar type of calculation. Since new goods contribute to lowering the exact price index, we expect that the Fisher price index using the Hicksian reservation prices will be *less than* the maximum overlap Fisher price index in periods when new goods appear. Taking the inverse ratio of these indexes, we obtain our first measure of gains for the KBF utility function,

$$G_{KBF}(t-1,t) = P_{FM}(t-1,t)/P_{FH}(t-1,t), \qquad t=2,\ldots,39. \tag{64}$$

A second measure of gains is obtained by taking the ratio of the maximum overlap price index with the Fisher index computed with *predicted prices* for all goods:

$$G_{KBF}^*(t-1,t) = P_{FM}(t\text{-}1,t)/P_F^*(t-1,t), \qquad t=2,\ldots,39, \tag{65}$$

These measures of gain are calculated for our frozen juice data set. If the availability of products is constant over periods $t-1$ and t, then $G_{KBF}(t-1,t)$ and $G_{KBF}^*(t-1,t)$ will be equal to 1. Thus the periods where these measures differ from unity in our data set are periods 9, 10, 11, 20 and 23, with these results shown in Table 4, below.

Table 4: Alternative Measures of Gain for the KBF Utility Function, Using Hicksian Reservation Prices for Unavailable Products and Using Predicted Prices for All Products

Month	Availability	G_{KBF}	G^*_{KBF}
9	2 and 4 new	1.0004	1.0016
10	12 disappears	0.9965	0.9988
11	12 reappears	1.0025	1.0015
20	12 disappears	0.9998	0.9971
23	12 reappears	0.9991	1.0001
Cumulative Gain		0.9983	0.9991

We expected $G_{KBF}(t-1,t)$ 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 G_{KBF} = 0.9991 which is less than unity, so the increased availability of product 12 in month 23 led to an *decrease* in utility rather than an *increase* as expected. Furthermore, the product of the 5 non-unitary values for G_{KBF} was 0.9983 (see the last row of Table 3) and so the overall increase in the availability of products led to a small *decrease* in utility over the sample period equal to 0.17 percentage points, rather than a *increase* 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, equal to their reservation prices). That is what we do in the measure of gains $G^*_{KBF}(t-1,t)$ defined in (65), and reported in the final column of Table 4.

Again, we expected G^*_{KBF} to be greater than 1 for periods 9, 11 and 23 when product availability increased and to be less than 1 for periods 10 and 20 when product availability decreased. Our expectations were realized: there were no anomalous results for the 5 periods, and in particular the month 23 value for G^*_{KBF} rose to 1.001, indicated a slight utility gain as product 12 reappears in the data, as compared to the month 23 value for G_{KBF} which was 0.9991. However, the product of the 5 non-unitary values for G^*_{KBF} turned out to be 0.9991 also, and so the overall increase in the availability of products led to a tiny *decrease* in utility over the sample period equal to 0.09 percentage points, rather than an *increase* as was expected. Unlike the CES results reported in Table 2, where the *overall* utility gain equaled the *initial* gain from the entry of products 2 and 4 in month 9, for the KBF preferences the repeated exit and entry of product 12 pulls down the initial gain (of 1.0016 in month 9) to become instead an overall loss.

The explanation for this anomalous result appears to be that the *maximum overlap* Fisher price index is not well-founded theoretically: because KBF preferences are not strongly separable over all goods (as are CES preferences), then if a good is not available in period t-1 it is theoretically incorrect to ignore it in period t when calculating the price index. In other words, we have not developed any result like in (14), for the CES case, that justifies using the "common" (i.e. maximum overlap) set of goods over two periods. We will address this problem in the following section, where we work directly with the utility function, to establish an analogue to the CES method for measuring the utility gain that is valid for the KBF or other functional forms.

4.8 The Gains and Losses Due to Changes in Product Availability Revisited

In this section, we consider framework for measuring the gains or losses in utility due to changes in the availability of products that can be applied to the KBF (or any other) utility function. 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 $p_t = (p_{1t},...,p_{Nt}) > 0_N$ and $q_t = (q_{1t},...,q_{Nt}) > 0_N$ respectively for t = 1,...,T. Sales or expenditures on the N products during period t are $E_t \equiv p_t \cdot q_t$ for t = 1,...,T.[36] We assume that a linearly homogeneous utility function, $f(q_1,...,q_N) = f(q)$, has been estimated where $q \geq 0_N$.[37] If product i is not available (or not sold) during period t, the corresponding price and quantity, p_{it} and q_{it}, 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 i, p_{it}^*, is defined as the prices satisfying the first-order conditions (22) using partial derivatives of the estimated utility function f(q):

---

[36] We also assume that $\sum_{i=2}^{19} p_{it}q_{it} > 0$ for t = 1,...,T.

[37] 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.

(66) $$p_{it}^* \equiv E_t\,[\partial f(q_t)/\partial q_i]/f(q_t)\,, \qquad i = 1,\ldots,N;\ t = 1,\ldots,T.$$

The prices defined by (66) 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 $p_{it}^* \geq 0$ for all i and t. If the estimated utility function fits the observed data exactly (so that all errors in the estimating equations are equal to 0),[38] then the predicted prices, p_{it}^*, for the available products will be equal to the corresponding actual prices, p_{it}.

Imputed expenditures on product i during period t are defined as $p_{it}^* q_{it}$ for i = 1,...,N. Note that if product n is not sold during period t, $q_{it} = 0$ and hence $p_{it}^* q_{it} = 0$ 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:

(67) $$\begin{aligned} E_t^* &\equiv \textstyle\sum_{i=1}^N p_{it}^* q_{it}\,, && t = 1,\ldots,T \\ &= \textstyle\sum_{i=1}^N q_{it}\, E_t\,[\partial f(q_t)/\partial q_i]/f(q_t), && \text{using definition (66)} \\ &= E_t \end{aligned}$$

where the last equality follows using the linear homogeneity of f(q) since by Euler's Theorem on homogeneous functions, we have $f(q) = \sum_{i=1}^N q_i\, \partial f(q)/\partial q_i$. 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 $q_{1t-1} = 0$) , but in period t, it becomes available and a positive amount is purchased (so that $q_{1t} > 0$). Our task is to define a measure of the increase in consumer 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 $q_{\sim 1t} \equiv [q_{2t},q_{3t},\ldots,q_{Nt}]$. Thus $q_t = [q_{1t},q_{\sim 1t}]$. 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*

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

consumer utility during period t, U_t, as follows:

$$(68)\qquad U_t \equiv f(q_t) = f(q_{1t},q_{\sim 1t}).$$

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

$$\begin{aligned}(69)\qquad U_{\sim 1t} &\equiv f(0,q_{\sim 1t}) \\ &\leq f(q_{1t},q_{\sim 1t}) \text{ since } f(q) \text{ is nondecreasing in the components of } q \\ &= U_t \text{ using definition (68).}\end{aligned}$$

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

$$\begin{aligned}(70)\qquad E^*_{\sim 1t} &\equiv \sum_{i=2}^{N} p^*_{it}\, q_{it} \\ &= E_t - p^*_{1t}\, q_{1t} \quad \text{using (67)} \\ &\leq E_t \quad \text{since } p^*_{1t} \geq 0 \text{ and } q_{1t} > 0.\end{aligned}$$

It will be useful to work with the ratio of $E^*_{\sim 1t}$ to E_t, defined as:

$$(75)\qquad \lambda_1 \equiv E^*_{\sim 1t}/E_t \leq 1 \text{ using (70).}$$

Notice that the scalar λ_1 is exactly the same as the term λ_t defined in (12), provided that we use the "common" set of goods $I \equiv \{2,\ldots,N\}$ in (12). In other words, this is the period t expenditure on the set of goods $\{2,\ldots,N\}$ that were also available in period t-1, relative to total expenditure. Then divide the vector of period t purchases excluding product 1, $q_{\sim 1t}$, by the scalar λ_1, and calculate the resulting imputed expenditures on the vector $q_{\sim 1t}/\lambda_1$ as equal to E_t:

$$\begin{aligned}(72)\qquad \sum_{i=2}^{N} p^*_{it}\, q_{it}/\lambda_1 &= (1/\lambda_1) \sum_{i=2}^{N} p^*_{it}\, q_{it} \\ &= (1/\lambda_1)\, E^*_{\sim 1t} \quad \text{using definition (70)} \\ &= (E_t / E^*_{\sim 1t})\, E^*_{\sim 1t} \quad \text{using definition (71)} \\ &= E_t.\end{aligned}$$

Using the linear homogeneity of f(q) in the components of q, we are able to calculate the utility level, U_{A1t}, that is generated by the vector $q_{\sim 1t}/\lambda_1$ as follows:

$$\begin{aligned}(73)\qquad U_{A1t} &\equiv f(0, q_{\sim 1t}/\lambda_1) \\ &= (1/\lambda_1)\, f(0,q_{\sim 1t}) \quad \text{using the linear homogeneity of f} \\ &= (1/\lambda_1)\, U_{\sim 1t} \quad \text{using definition (69).}\end{aligned}$$

Note that λ_1 can be calculated using definition (71) and $U_{\sim 1t}$ can be calculated using definition (69). Thus, U_{A1t} can also be readily calculated.

Consider the following (hypothetical) consumer's period t aggregate *utility maximization problem where product 1 is not available* and consumers face the imputed prices p_{it}^* 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:

$$(74) \quad \max_{q\text{'s}} \{f(0,q_2,q_3,\ldots,q_N) : \textstyle\sum_{i=2}^{N} p_{it}^* q_{it} \le E_t\} \equiv U_{1t}$$
$$\ge U_{A1t} ,$$

where U_{A1t} is defined by (73). The inequality in (74) follows because (72) shows that $q_{\sim 1t}/\lambda_1$ is a feasible solution for the utility maximization problem defined by (74). We also know that the actual utility level in period t, U_t exceeds the maximized utility level U_{1t} when good 1 is not available, so that we have:

$$(75) \quad U_t \ge U_{1t} \ge U_{A1t}.$$

We regard U_{A1t} as an approximation (and lower bound) to U_{1t}. Given that an estimated utility function f(q) is in hand, it is easy to compute the *approximate* utility level U_{A1t} when product one is not available. The *actual* constrained utility level, U_{1t}, will in general involve solving numerically the nonlinear programming problem defined by (74). For the KBF functional form, instead of maximizing $(q^TAq)^{1/2}$, we could maximize its square, q^TAq, and thus solving (74) 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 (74) since the strong separability of the CES functional form will imply that $U_{1t} = U_{A1t}$. In other words, for the CES utility function, when good 1 is not available then the consumer will *optimally choose* to inflate the purchases $q_{\sim 1t}$ by $(1/\lambda_1)$ in order to exhaust the budget E_t.

A reasonable measure of the gain in utility due to the new availability of product 1 in period t, G_{1t}, is the ratio of the completely unconstrained level of utility U_t to the product 1 constrained level U_{1t} i.e., define *the product 1 utility gain in period t* as:

$$(76) \quad G_{1t} \equiv U_t/U_{1t} \ge 1,$$

where the inequality follows from (75). The corresponding *product 1 approximate utility gain* is defined as:

(77) $$G_{A1t} \equiv U_t / U_{A1t} \geq G_{1t} \geq 1,$$

where the inequalities follow again from (75). Thus in general, the approximate gain is an upper bound to the true gain in utility due to the new availability of product 1 in period t.

Note that for the CES utility function we have $G_{A1t} = G_{1t}$ since $U_{1t} = U_{A1t}$. Furthermore, using the shares in (39) which assumed no measurement error in prices, so that $p_{it} = p_{it}^*$, we have:

(78) $$G_{A1t} = \frac{U_t}{U_{A1t}} = \lambda_{1t}\frac{U_t}{U_{\sim 1t}} \qquad \text{from definitions (73) and (77)}$$

$$= \frac{\sum_{i=2}^{N} p_{it}^* q_{it}}{E_t}\frac{U_t}{U_{\sim 1t}} \qquad \text{from definition (71)}$$

$$= \frac{\sum_{i=2}^{N} a_i q_{it}^{(\sigma-1)/\sigma}}{\sum_{i=1}^{N} a_i q_{it}^{(\sigma-1)/\sigma}}\frac{U_t}{U_{\sim 1t}} \qquad \text{from (39) with } p_{it} = p_{it}^*$$

$$= \left[\frac{\sum_{i=1}^{N} a_i q_{it}^{(\sigma-1)/\sigma}}{\sum_{i=2}^{N} a_i q_{it}^{(\sigma-1)/\sigma}}\right]^{1/(\sigma-1)} \qquad \text{from (3) with } \tfrac{\sigma}{\sigma-1} - 1 = \tfrac{1}{\sigma-1}$$

$$= \left(1 - \sum_{i=2}^{N} s_{it}\right)^{-1/(\sigma-1)} \qquad \text{from (39) once again.}$$

So for the CES case, the *approximate* measure of gain G_{A1t} equals the *true* gain G_{1t}, and these are exactly equal to the CES gain we defined earlier in (55) when applied to the case of new product 1. In other words, the earlier CES gain is identical to approximate measure of gain that we have proposed in this section when applied to that functional form. But our definitions in this section also apply to *any other* functional form for utility, including the KBF form, while recognizing that we are using the approximation (and upper bound) G_{A1t} rather than G_{1t}.

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. It turns out, however, 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 instead 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, is the ratio of the product 1 constrained level of utility U_{1t} 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

(79) $$L_{1,t+1} \equiv U_{1t}/U_t \leq 1,$$

where the inequality follows from (75). The corresponding *product 1 approximate utility loss* is defined as:

(80) $$L_{A1,t+1} \equiv U_{A1t}/U_t \leq L_{1,t+1} \leq 1,$$

where the inequalities again follow from (75). 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 $L_{A1,t+1} = L_{1,t+1}$.

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 for the KBF utility function are listed in the third column of Table 5. 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 in the KBF case is also listed in the third column of Table 5. In the fourth column of Table 5 we repeat the CES gain in utility from Table 2 for period 9 due to the introduction of products 2 and 4, and the various impacts of the exit and entry of product 12. Thus, Table 5 compares the gains and losses in utility for the KBF and CES models for the 5 months where there was a change in product availability.

In month 9, when products 2 and 4 become available, the CES model implies that the enhanced product availability increase consumers' utility by 0.73 percentage points while the KBF model implies a much smaller increase of 0.13 percentage points. Following that product introduction, we have the disappearance and reappearance of product 12 overall several months.

Table 5: The Gains and Losses of Utility Due to Changes in Product Availability

Month	Availability	$G_{A,KBF}$ $L_{A,KBF}$	G_{CES}
9	2 and 4 new	1.0013	1.0073
10	12 disappears	0.9975	0.9964
11	12 reappears	1.0030	1.0043
20	12 disappears	0.9988	0.9962
23	12 reappears	1.0008	1.0031
Cumulative Gain		1.0014	1.0073

Recall that in our earlier calculation of the CES gain (see Table 2), the net effect on utility of the entry and exit of product 12 cancelled out, so that the overall utility gains came only from the initial entry or products 2 and 4. That was not the case for our earlier calculation of the KBF utility gains (see Table 4), where the exit and entry of product 12 at its reservation prices had a noticeable and lasting impact on utility. That anomalous result not longer appears using our methodology of this section, where product 12 now has only a very small impact on overall utility, increasing the utility gain from 1.0013 (first row of the third column in Table 5) to 1.0014 (final row of the third column).

So product 12 has only a very minor effect on utility, and the principal impact comes from the month 9 introduction of products 2 and 4, *where the CES gains are more than five times higher than the KBF gains* in Table 5. That is a surprising result, since our argument throughout this paper has been that the CES gains are at least twice as high as the Hausman gains obtained from a linear approximation to the demand curve. We have noted in section 3.3 that the demand curves of the KBF utility function are convex, and since these convex demand curves lie above their linear approximation, the utility gain from a new product with KBF utility should *exceed* the utility gain along linear approximation. It follows CES gains should be *not much more than twice as high as the KBF gains, provided that those demand curves have the same elasticity at the point of consumption*. Instead, we are finding in our estimation that we must divide the CES gain by *more than five* to get the estimated KBF gain.

The resolution to these surprising empirical results is that the *KBF and CES demand curves must have different slslope at the point of consumption*. But there is nothing in our estimation that will guarantee that result, and in fact, our KBF utility function has *more elastic* demand on average for any products – including products 2 and 4 when they are introduced – than the estimated CES utility function. To illustrate the more elastic demand for the KBF function, we compute the Hausman approximation to the KBF gain as shown in (38) and Hausman approximation to the CES gain as shown in (18). To be more specific, we single out each product and regard it as a product 1 in the approximate formulae (18) and (38). The remaining products are aggregated into product 2. The share of this aggregate product 2 is simply $s_{2t} \equiv 1 - s_{1t}$.[39] With these modifications, we can calculate $G_{H,\,KBF}$ and $G_{H,\,CES}$ for each product and each time period. That is, we pretend that each product is newly introduced in each time period, and calculate the corresponding gains. Then we take the mean of these measures for each product over the 39 time periods for our estimated KBF and CES functional forms, as reported in Table 6.

Table 6: Gains from the Appearance of Each Product for the Estimated KBF and CES Utility Functions

Product	$G_{H,KBF}$	$G_{H,CES}$	Product	$G_{H,KBF}$	$G_{H,CES}$
1	0.00407	0.00230	11	0.00335	0.00053
2	0.00077	0.00294	12	0.00211	0.00070
3	0.00055	0.00403	13	0.00555	0.00457
4	0.00081	0.00125	14	0.00092	0.00461
5	0.00331	0.00091	15	0.00087	0.00120
6	0.00012	0.00505	16	0.00311	0.00323
7	0.00054	0.00064	17	0.00194	0.00382
8	0.00101	0.00185	18	0.00113	0.00420
9	0.00077	0.00396	19	0.00042	0.00372
10	0.00053	0.00444	Mean	0.00168	0.00265

From Table 6, it can be seen that averaging over all products and all time periods, the approximate gain in utility from the introduction of a product is about 0.168 percentage points using our estimated KBF utility function and about 0.265 percentage points using our estimated CES utility function. So the CES functional form gives a high estimate of the welfare gain by

[39] The shares that we use for this exercise are fitted shares; i.e., we use the actual quantities that are observed in period t, q_{it}, and the estimated prices $p_{it}^* \equiv f_i(q_t)E_t/f(q_t)$ where f(q) is the estimated utility function.

nearly a factor of two. The difference between them is explained *entirely* by the differing estimates of the inverse demand elasticities, as can be seen from equation (31). In order to have CES gains that are about twice as high on average as the KBF gains, it must be that the elasticity of demand for the KBF function is about twice as high as for the CES.[40]

With the results shown in Table 6 explaining that the Hausman approximation to the gains from a new product are about twice as high for the CES and the KBF functional forms, and further, that the actual CES gains are at least twice as high the Hausman approximation to the CES gains (as shown in Table 1), it is not surprising that the CES gains (from products 2 and 4) *are more than five times higher than the KBF gains* in Table 5: in very rough terms, about one-half of this difference comes from having more elastic demand for the KBF than for the CES demand functions (so that the Hausman linear approximation to the gains for the CES function are twice as high as for the KBF function), while the other half comes from CES demand curves being more convex (with infinite reservation price) than KBF demand.

5. Conclusions

Determining how to incorporate new goods into the calculation of price indexes is an important, unresolved issue for statistical agencies. That issue becomes particularly important with the increased availability of scanner data to measure prices and quantities, because new and disappearing products at the barcode level occur frequently in such data. Our goal in this paper has been to compare several empirical methods to deal with new and disappearing products: the proposal by Hausman (1999; 191) (2003; 27) to use a linear approximation to the demand curve to compute a lower bound to the consumer surplus, assuming that the true demand curve is convex; and with the estimation of two utility functions, the CES case and a quadratic utility function that we refer to as the KBF case. We have extended the approach of Hausman to apply to the analysis of inverse demand curve (prices as functions of quantities) rather than direct demand curves (quantities as functions of prices), as needed in the KBF case. Then we have illustrated our results using the barcode data for frozen juice from one grocery store. While obviously limited in its scope, there are several tentative conclusions that can be drawn from the computations undertaken in this paper:

[40] In Appendix B, Table B1, we report some average elasticities for each product that are quite similar to the elasticities of inverse demand.

- The Feenstra CES methodology for adjusting maximum overlap chained price indexes for changes in product availability is 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, as discussed in section 3.1
- It is not a trivial matter to obtain an accurate estimate for σ. Section 4.3 and 4.4 of the paper developed two methodological approaches to the estimation of the elasticity of substitution if purchasers of products have CES preferences. These method adapt Feenstra's (1994) double log differencing technique to the estimation of σ in a systems approach where only one parameter needs to be estimated for an entire system of transformed CES demand functions.
- 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, in section 3.3 we discussed 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).
- In section 4.5 we 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.[41] This semiflexible approach worked in a satisfactory manner. This approach also allowed us to drop the observations that

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

correspond to the 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), in section 4.6 above, 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 4.7, we switched to using actual prices as the dependent variables. This approach generated more satisfactory estimates for the KBF functional form.
- The results presented in sections 4.8 indicate that the Feenstra CES methodology for measuring the benefits of increases in product variety may overstate these benefits as compared to our semiflexible methodology. We find that the CES gains *are more than five times greater than the KBF gains*: in very rough terms, about one-half of this difference comes from having more elastic demand for the KBF than for the CES demand functions (so that the Hausman linear approximation to the gains for the CES function are twice as high as for the KBF function), while the other half comes from CES demand curves being more convex (with infinite reservation price) than KBF demand.

There is one other functional form that we have not explored in this paper but which deserves more attention when examining new goods, and that is the translog expenditure function. In its most general form this function is flexible, and under additional conditions the demand curves are convex with finite reservation prices for new goods. Feenstra and Shiells (1997) have examined the case of a single new good, and assuming that the translog and CES demand curves are tangent at the point of consumption, they argue that the gains from the new good in the translog case is *one-half* as large as the CES gains. Feenstra and Weinstein (2017) have examined a simplified *symmetric* translog expenditure function that has the same number of free parameters as the CES, i.e. it is not a fully flexible functional form. With that simplification, they confirm that the translog case are about one-half as large as the CES gains on a large dataset involving new imported products into the United States: they find that the gains from new imports are about one-half as large in the translog case as what Broda and Weinstein (2006) find in the CES case. Applying the translog functional form to scanner datasets would be a valuable exercise to see whether that method might be an alternative to the CES functional form, and we expect that the adjustment for new and disappearing goods will be about one-half as large in the translog case as for the CES.

Our approach can be compared to the recent work of Redding and Weinstein (2019), who also use a CES utility function. They assume that this functional form represents the “true” preferences, so that any observed deviation from the CES demand curves must represent a shift in tastes. For example, a good with a falling price and a very large increase in demand –a greater increase than what would be implied by the elasticity of substitution – must have a shift in tastes towards that good. They argue that the consumer gain from that price reduction are *greater* than what we would compute using constant tastes (which is the usual assumption of exact price indexes). So *in addition* to the CES correction for new goods, they would propose a further correction to allow for taste change. Our results in this paper show, in contrast, that once we move away from the CES case and consider alternative utility functions such as the KBF (or the translog case just mentioned), then the gains from new products will be less than that found for the CES utility function.

<u>Appendix A</u>: The Frozen Juice Data

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

Table A1: Monthly Quantities Sold for 19 Frozen Juice Products

Month t	q_{1t}	q_{2t}	q_{3t}	q_{4t}	q_{5t}	q_{6t}	q_{7t}	q_{8t}	q_{9t}
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_{10t}	q_{11t}	q_{12t}	q_{13t}	q_{14t}	q_{15t}	q_{16t}	q_{17t}	q_{18t}	q_{19t}
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 Juice Products

Month t	p_{1t}	p_{2t}	p_{3t}	p_{4t}	p_{5t}	p_{6t}	p_{7t}	p_{8t}	p_{9t}
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_{10t}	p_{11t}	p_{12t}	p_{13t}	p_{14t}	p_{15t}	p_{16t}	p_{17t}	p_{18t}	p_{19t}
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_{2t} and p_{4t} are not available for t =1,2,...,8 since products 2 and 4 were not sold during these months. However, in Table A.2, we filled in these missing prices with the estimated reservation prices that were estimated in section 4.4. Similarly, p_{12t} was missing for months t = 12, 20, 21 and 22 and again, we replaced these missing prices with the estimated reservation prices in Table A2. The estimated reservation prices appear in italics.

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).

<u>Appendix B</u>: Proof of results in section 3.2

In the main text, we compute the term $u'(q_{1t})$ as:

$$u'(q_{1t}) = f_1(q_{1t},q_{2t}) + f_2(q_{1t},q_{2t})\partial q_2(q_{1t})/\partial q_1 \qquad \text{differentiating (26)}$$

$$= f_1(q_{1t},q_{2t}) + f_2(q_{1t},q_{2t})(-p_{1t}/p_{2t})] \qquad \text{differentiating (25)}$$

It follows that,

(B1) $$u''(q_{1t}) = f_{11}(q_{1t},q_{2t}) + 2f_{12}(q_{1t},q_{2t})(-p_{1t}/p_{2t}) + f_{22}(q_{1t},q_{2t})(-p_{1t}/p_2)^2 \le 0,$$

where the inequality follows since the matrix of second order partial derivatives of $f(q_{1t},q_{2t})$ is negative semidefinite using the concavity of $f(q_1,q_2)$.

We can express the second derivative $u''(q_{1t})$ 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 of the marginal utility,* $\mu_{ij}(q_1,q_2)$, by multiplying $f_{ij}(q_1,q_2)$ by $q_j/f_i(q_1,q_2)$:

(B2) $$\mu_{ij}(q_1,q_2) \equiv [q_j/f_i(q_1,q_2)]f_{ij}(q_1,q_2), \qquad i,j = 1,2.$$

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:

(B3) $$f_{11}(q_{1t},q_{2t})q_{1t} + f_{12}(q_{1t},q_{2t})q_{2t} = 0;$$

(B4) $$f_{21}(q_{1t},q_{2t})q_{1t} + f_{22}(q_{1t},q_{2t})q_{2t} = 0.$$

Young's Theorem from calculus also implies that $f_{12}(q_{1t},q_{2t}) = f_{21}(q_{1t},q_{2t})$. Using this relationship along with (B3) and (B4) implies the following relationships between the second order partial derivatives of f:

(B5) $$f_{12}(q_{1t},q_{2t}) = f_{21}(q_{1t},q_{2t}) = f_{11}(q_{1t},q_{2t})(-q_{1t}/q_{2t});$$

(B6) $$f_{22}(q_{1t},q_{2t}) = f_{11}(q_{1t},q_{2t})(-q_{1t}/q_{2t})^2 .$$

Now substitute (B5) and (B6) into (B1) in order to obtain the following expression for $u''(q_{1t})$:

(B7) $$u''(q_{1t}) = f_{11}(q_{1t},q_{2t}) + 2f_{12}(q_{1t},q_{2t})(-p_{1t}/p_{2t}) + f_{22}(q_{1t},q_{2t})(-p_{1t}/p_{2t})^2$$

$$= f_{11}(q_{1t},q_{2t})[1 + 2(p_{1t}q_{1t}/p_{2t}q_{2t}) + (p_{1t}q_{1t}/p_{2t}q_{2t})^2]$$
$$= f_{11}(q_{1t},q_{2t})[1 + (s_{1t}/s_{2t})]^2$$

where $s_{it} \equiv p_{it}q_{it}/E_t$ for i = 1,2. Since $f_{11}(q_{1t},q_{2t}) \leq 0$, $u''(q_{1t}) \leq 0$ as well. Using (B2), we can write $f_{11}(q_{1t},q_{2t})$ in elasticity form as follows:

(B8) $$f_{11}(q_{1t},q_{2t}) = \mu_{11}(q_{1t},q_{2t})f_1(q_{1t},q_{2t})/q_{1t}$$
$$= \mu_{11}(q_{1t},q_{2t})p_{1t}/q_{1t}, \quad \text{using (24).}$$

Finally, substitute (B7) and (B8) into (30) and our second order approximation to the gain of utility due to the appearance of product 1 becomes:

(B9) $$G_U = -\tfrac{1}{2}s_{1t}\mu_{11}(q_{1t},q_{2t})[1 + (s_{1t}/s_{12})]^2.$$

To simplify this expression, we 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, $\partial D_1(q_{1t})/\partial q_1$, and then evaluate how the approximate Hausman loss defined by (12) compares to our approximate loss defined by (B9).

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

(B10) $$p_1 = d_1(q_1,q_2,E) \equiv Ef_1(q_1,q_2)/f(q_1,q_2);$$

(B11) $$p_2 = d_2(q_1,q_2,E) \equiv Ef_2(q_1,q_2)/f(q_1,q_2).$$

We want the partial equilibrium function, $p_1 = D_1(q_1)$ holding other variables constant. The variables that Hausman holds constant are the utility level U and the price of product 2, p_2. Endogenous variables in his framework are q_1, q_2 and E while the driving variable is p_1 which goes from p_{1t} to p_1^* while q_1 goes from q_{1t} to 0. We can adapt his framework in our direct utility function model as follows: regard $U_t \equiv f(q_{1t},q_{2t})$ and p_{2t} 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_{1t} to 0 is the following one:

(B12) $$f(q_1,q_2(q_1)) = f(q_{1t},q_{2t}) = E_t.$$

Thus we again scale utility so that initial utility $f(q_{1t},q_{2t})$ is equal to initial expenditure, E_t. Define $q_2(q_1)$ as the implicit function which satisfies (B12). The derivative of this implicit function is defined by differentiating $f(q_1,q_2(q_1)) = E_t$ with respect to q_1. Thus we find that:

(B13) $$q_2'(q_{1t}) = -\, f_1(q_{1t},q_{2t})/f_2(q_{1t},q_{2t}) = -p_{1t}/p_{2t}\ ,$$

where the second equation in (B13) follows from (B12) and (B10) and (B11) (our two inverse demand functions) evaluated at the initial equilibrium. We take the second inverse demand function defined by (B11) and set it equal to the constant, p_{2t}. We solve the resulting equation for expenditure as a function of q_1, $E(q_1)$:

(B14) $$\begin{aligned} E(q_1) &\equiv p_{2t}f(q_1,q_2(q_1))/f_2(q_1,q_2(q_1)) \\ &= p_{2t}\, E_t\,/f_2(q_1,q_2(q_1)), \quad \text{using (B12).} \end{aligned}$$

Differentiate (B14) with respect to q_1 in order to determine the derivative $E'(q_{1t})$. We find that

(B15) $$\begin{aligned} E'(q_{1t}) &= -\,(p_{2t}\, E_t\,/p_{2t}^{\,2})[f_{21}(q_{1t},q_{2t}) + f_{22}(q_{1t},q_{2t})q_2'(q_{1t})], \quad \text{using (B10)} \\ &= -\,(E_t\,/p_{2t})[f_{21}(q_{1t},q_{2t}) + f_{22}(q_{1t},q_{2t})(-p_{1t}/p_{2t})], \quad \text{using (B13).} \end{aligned}$$

We can now define our Hausman partial equilibrium inverse demand function $p_1 = D_1(q_1)$ by replacing q_2 and E in definition (B10) by $q_2(q_1)$ and $E(q_1)$:

(B16) $$\begin{aligned} D_1(q_1) &\equiv E(q_1)f_1(q_1,q_2(q_1))/f(q_1,q_2(q_1)) \\ &= E(q_1)f_1(q_1,q_2(q_1))/\, E_t, \quad \text{using (B12).} \end{aligned}$$

The derivative of the partial equilibrium inverse demand function defined by (B16) at q_{1t} is:

(B17) $$\begin{aligned} \partial D_1(q_{1t})/\partial q_1 &= -\,(p_{1t}/\, E_t)(\, E_t\,/p_{2t})[f_{21}(q_{1t},q_{2t}) + f_{22}(q_{1t},q_{2t})(-p_{1t}/p_{2t})] \\ &\quad + [E(q_{1t})/\, E_t][f_{11}(q_{1t},q_{2t}) + f_{12}(q_{1t},q_{2t})q_2'(q_{1t})], \text{ using (B15)} \\ &= [f_{21}(q_{1t},q_{2t})(-p_{1t}/p_{2t}) + f_{22}(q_{1t},q_{2t})(-p_{1t}/p_{2t})^2] + [f_{11}(q_{1t},q_{2t}) + f_{12}(q_{1t},q_{2t})q_2'(q_{1t})] \\ &= f_{11}(q_{1t},q_{2t}) + 2f_{12}(q_{1t},q_{2t})(-p_{1t}/p_{2t}) + f_{22}(q_{1t},q_{2t})(-p_{1t}/p_{2t})^2 \\ &= u''(q_{1t}) \qquad \text{where } u''(q_{1t}) \text{ was defined by (B1)} \\ &= f_{11}(q_{1t},q_{2t})[1 + (s_{1t}/s_{2t})]^{2,} \text{ using (B7).} \end{aligned}$$

Thus from (30), the Hausman lower-bound gains for this partial equilibrium demand derivative defined by (B17) turns out to be:

(B18) $$G_H \equiv -\tfrac{1}{2}[\partial D_1(q_{1t})/\partial q_1]q_{1t}^2/E_t$$
$$= -\tfrac{1}{2}q_{1t}^2 f_{11}(q_{1t},q_{2t})[1 + (s_{1t}/s_{2t})]^2/E_t, \text{ using (B17)}$$
$$= -\tfrac{1}{2}s_{1t}\mu_{11}(q_{1t},q_{2t})[1 + (s_{1t}/s_{2t})]^2, \text{ using (B8)}$$

where the elasticity marginal utility elasticity $\mu_{11}(q_{1t},q_{2t})$ is defined as $(q_{1t}/p_{1t})f_{11}(q_{1t},q_{2t})$. This is a rather surprising result: Hausman's first-order consumer surplus approximate approach to measuring the gain from a new product in (B18) turns out to be exactly equal to our second-order approximation gain in utility approach in (B9) when there are only 2 products.

We apply a modification of the above formulae to our data set using our estimated KBF and CES utility functions, as described in the main text. That is, we pretend that each product is newly introduced in each time period, and calculate the corresponding gains. Denote the mean of these measures for each product n over the 39 time periods for our estimated KBF and CES functional forms by $G_{H,KBF}$ and $G_{H,CES.}$ We also compute the CES and KBF marginal utility elasticities. These are obtained from (B2), which gives the following results for the CES and KBF utility functions using (B12):

(B19) $$\mu_{CES,nn} = -(1 - s_{nt})/\sigma\ , \quad \mu_{KBF,nn} = s_{nt}\left(\frac{a_{nn}}{p_n^2} - 1\right).$$

These means are listed in Table B1 below, and the Hausman approximate gains are also reported in Table 6 in the main text.

Table B1: Gains from the Appearance of Each Product for the Estimated KBF and CES Utility Functions, and the Marginal Utility Elasticities

Product	$G_{H,BF}$	$G_{H,CES}$	$\mu_{KBF,nn}$	$\mu_{CES,nn}$	Product	$G_{H,BF}$	$G_{H,CES}$	$\mu_{KBF,nn}$	$\mu_{CES,nn}$
1	0.00407	0.00230	-0.130	-0.139	11	0.00335	0.00053	-0.159	-0.140
2	0.00077	0.00294	-0.043	-0.143	12	0.00211	0.00070	-0.150	-0.143
3	0.00055	0.00403	-0.031	-0.141	13	0.00555	0.00457	-0.118	-0.115
4	0.00081	0.00125	-0.046	-0.142	14	0.00092	0.00461	-0.030	-0.136
5	0.00331	0.00091	-0.076	-0.130	15	0.00087	0.00120	-0.045	-0.143
6	0.00012	0.00505	-0.007	-0.141	16	0.00311	0.00323	-0.068	-0.130
7	0.00054	0.00064	-0.028	-0.141	17	0.00194	0.00382	-0.135	-0.142
8	0.00101	0.00185	-0.074	-0.143	18	0.00113	0.00420	-0.042	-0.139
9	0.00077	0.00396	-0.042	-0.143	19	0.00042	0.00372	-0.015	-0.139
10	0.00053	0.00444	-0.035	-0.144	Mean	0.00168	0.00265	-0.067	-0.139

From Table B1, it can be seen that averaging over all products and all time periods, the approximate gain in utility from the introduction of a product is about 0.168 percentage points using our estimated KBF utility function and about 0.265 percentage points using our estimated CES utility function. So the CES functional form gives a high estimate of the welfare gain by nearly a factor of two. The difference between them is explained entirely by the differing estimates of the marginal utility elasticities, which average -0.067 percentage points using our estimated KBF utility function and about 0.139 percentage points using our estimated CES utility function, or twice as high for the CES as compared to the KBF functional forms.

The average CES own marginal utility elasticity over all time periods and all products is − 0.139 and the corresponding KBF average elasticity is − 0.067. This explains why the CES loss is approximately twice as big as the KBF loss. However, note that for products 11, 12 and 13, the average KBF elasticity is larger in magnitude than the corresponding average CES elasticity. Furthermore, the KBF elasticities are quite variable as compared to the corresponding CES elasticities. This result follows from the properties of the above formula (B19), where $\mu_{CES,nn} \to -1/\sigma$ as $s_{nt} \to 0$, where the elasticity of substitution σ is common across goods. But for the KBF function, $\mu_{KBF,nn}$ depends on the parameters a_{nn} which can vary substantially across goods, and has the limit $\mu_{KBF,nn} \to 0$ as $s_{nt} \to 0$. So it is not surprising that the KBF marginal utility elasticities are usually smaller and generally more variable that the CES marginal utility elasticities.

Finally, we note the relationship between the marginal utility elasticities and the elasticity of inverse demand. From (B18) we have that,

(B20)
$$\begin{aligned}
[\partial D_1(q_{1t})/\partial q_1](q_{1t}/p_{1t}) &= \mu_{11}(q_{1t},q_{2t})[1 + (s_{1t}/s_{2t})]^2, \\
&= \mu_{11}(q_{1t},q_{2t})/(1 - s_{1t})^2 \,, \text{ using } s_{2t} = 1 - s_{1t} \\
&= -\,1/[\sigma(1 - s_{1t})], \text{ in the CES case from (B19)} \\
&= \frac{s_{1t}}{(1-s_{1t})^2}\left(\frac{a_{11}}{p_1^2} - 1\right), \text{ in the KBF case from (B19).}
\end{aligned}$$

The KBF inverse demand elasticity in the final line is used in (37) and (38) of the main text.

<u>Appendix C</u>: Proof of results in section 3.3

For the CES utility function, the derivative of the share is obtained from (6) as:

(C1) $$\frac{\partial \ln s_{it}}{\partial \ln p_{it}} = (1-\sigma) - (1-\sigma)\frac{b_{it}p_{it}^{1-\sigma}}{\sum_{i\in I_t} b_{it}p_{it}^{1-\sigma}} = (1-\sigma)(1-s_{it}).$$

Hicksian demand is shown by (5) and the derivative of demand is readily obtained from (7) as:

(C2) $$\left.\frac{\partial q_{it}}{\partial p_{it}}\right|_U = -\sigma\frac{q_{it}}{p_{it}}(1-s_{it}).$$

The second derivative of Hicksian demand is then:

$$\left.\frac{\partial^2 q_{it}}{\partial p_{it}^2}\right|_U = -\sigma\left.\frac{\partial q_{it}}{\partial p_{it}}\right|_U \frac{1}{p_{it}}(1-s_{it}) + \frac{\sigma q_{it}}{p_{it}^2}(1-s_{it}) + \frac{\sigma q_{it}s_{it}}{p_{it}^2}\frac{\partial \ln s_{it}}{\partial \ln p_{it}}$$

(C3) $$= -\sigma\left.\frac{\partial q_{it}}{\partial p_{it}}\right|_U \frac{1}{p_{it}}(1-s_{it}) + \frac{\sigma q_{it}}{p_{it}^2}(1-s_{it}) + \frac{\sigma q_{it}s_{it}}{p_{it}^2}(1-\sigma)(1-s_{it}).$$

Dividing out q_{it} / p_{it}^2 and using the Hicksian elasticity $\sigma(1-s_{it})$ from (7), we obtain,

(C4) $$\left.\frac{\partial^2 \ln q_{it}}{\partial \ln p_{it}^2}\right|_U = \sigma^2(1-s_{it})^2 + \sigma(1-s_{it}) + \sigma s_{it}(1-\sigma)(1-s_{it})$$
$$= \sigma(1-s_{it})[1+s_{it}+\sigma(1-2s_{it})],$$

which is positive for $s_{it} \le 0.5$. This condition ensures that the final term in (C4) is non-negative, so it follows that:

(C5) $$\left.\frac{\partial^2 \ln q_{it}}{\partial \ln p_{it}^2}\right|_U \ge \sigma(1-s_{it})(1+s_{it}) \ge 3s_{it}\sigma(1-s_{it}) = -3s_{it}\left.\frac{\partial \ln q_{it}}{\partial \ln p_{it}}\right|_U,$$

where the second inequality again uses $s_{it} \le 0.5$ so that $3s_{it} \le 1+s_{it}$, and the final equality uses the Hicksian elasticity in (7).

Turning now to the KBF utility function, Hicksian demand is shown in (35), and the derivative of this demand is obtained from (36) as:

(C6) $$\left.\frac{\partial q_{it}}{\partial p_{it}}\right|_U = \frac{q_{it}}{p_{it}}\left[\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}} - \frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right].$$

It follows that the second derivative is:

$$\begin{aligned}
\left.\frac{\partial^2 q_{it}}{\partial p_{it}^2}\right|_U &= \left.\frac{\partial q_{it}}{\partial p_{it}}\right|_U \frac{1}{p_{it}}\left[\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}} - \frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right] - \frac{q_{it}}{p_{it}^2}\left[\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}} - \frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right] \\
&\quad + \frac{q_{it}}{p_{it}^2}\left[\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}} - \left(\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}}\right)^2 - \frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t} - \frac{a^*_{ii}p_{it}^2}{p_t^T A^* p_t} + 2\left(\frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right)^2\right] \\
&= \frac{q_{it}}{p_{it}^2}\left[\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}} - \frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right]^2 - \frac{q_{it}}{p_{it}^2}\left[\left(\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}}\right)^2 + \frac{a^*_{ii}p_{it}^2}{p_t^T A^* p_t} - 2\left(\frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right)^2\right] \\
&= \frac{q_{it}}{p_{it}^2}\left[-2\left(\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}}\right)\frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t} - \frac{a^*_{ii}p_{it}^2}{p_t^T A^* p_t} + 3\left(\frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right)^2\right] \\
&= 3s_{it}\frac{q_{it}}{p_{it}^2}\left[-\left(\frac{a^*_{ii}p_{it}}{\sum_{n=1}^N a^*_{in}p_{nt}}\right) + \left(\frac{p_{it}\sum_{n=1}^N a^*_{in}p_{nt}}{p_t^T A^* p_t}\right)\right] = -3s_{it}\frac{1}{p_{it}}\left.\frac{\partial q_{it}}{\partial p_{it}}\right|_U .
\end{aligned}$$

It follows that for the KBF utility function,

(C7) $$\left.\frac{\partial^2 \ln q_{it}}{\partial \ln p_{it}^2}\right|_U = -3s_{it}\left.\frac{\partial \ln q_{it}}{\partial \ln p_{it}}\right|_U .$$

Comparing this result with (C5), it can be seen that the second derivative of the KBF function is higher than for the CES function at the consumption point when their compensated demand elasticities (and therefore their slopes of demand) are equal.

References

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.

Broda, Christian and David E. Weinstein (2006), "Globalization and the Gains From Variety," *Quarterly Journal of Economics*: 121(2), 541–585.

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

Diewert, W.E. (1999), "Index Number Approaches to Seasonal Adjustment", *Macroeconomic Dynamics* 3, 48-67.

Diewert, W. E. (2018), "Duality in Production", Discussion Paper 18-02, Vancouver School of Economics, University of British Columbia, Vancouver, B.C., Canada, V6T 1L4.

Diewert, W.E. and R. Feenstra (2019), "Estimating the Benefits of New Products: Some Approximation", Discussion Paper 19-02, Vancouver School of Economics, University of British Columbia, Vancouver, B.C., Canada, V6T 1L4.

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. (2010) "New Products with a Symmetric AIDS Expenditure Function," *Economic Letters*, 2, 108-111.

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.

Feenstra, Robert C. and Clinton Shiells (1997), "Bias in U.S. Import Prices and Demand," in Timothy Bresnahan and Robert Gordon, eds. *The Economics of New Goods*, NBER and Univ. of Chicago Press, 1997, 249-276.

Feenstra, Robert C. and David E. Weinstein (2017),"Globalization, Markups, and U.S. Welfare," *Journal of Political Economy*: 125(4), August, 1041-1074.

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

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.

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

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.

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.

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

Redding, Stephen and David E. Weinstein (2019), "A Unified Approach to Estimating Demand and Welfare," NBER Working Paper no. 22479, revised.

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.

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

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.

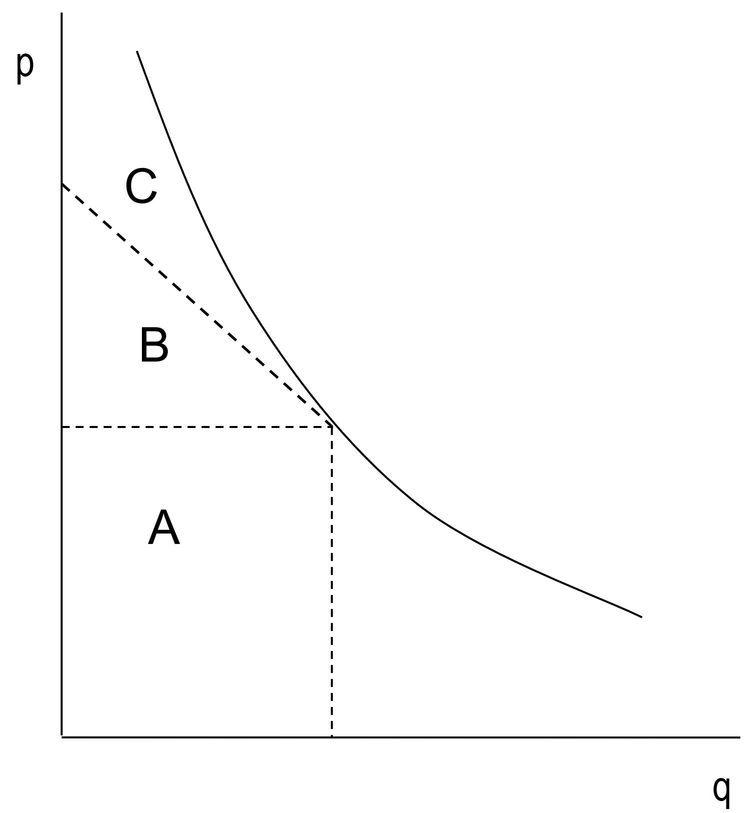

Figure 1: Constant-Elasticity Demand

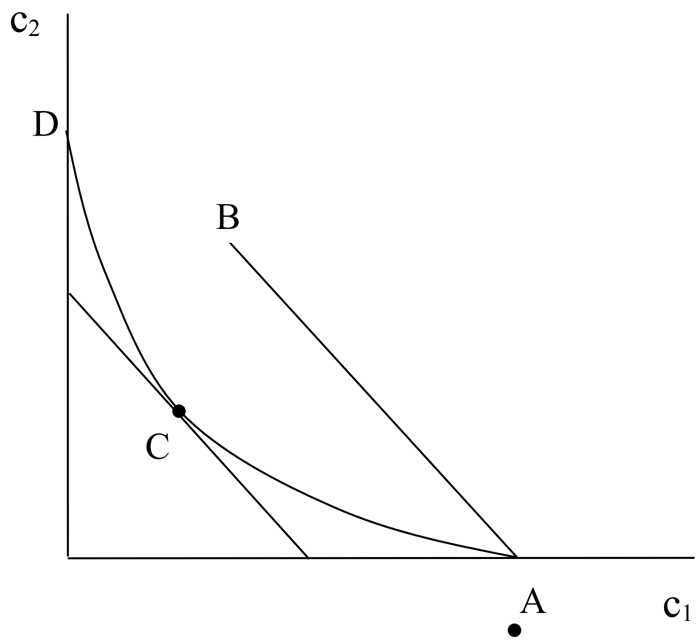

Figure 2: CES Indifference Curve