This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Big Data for Twenty-First-Century Economic Statistics

Volume Authors/Editors: Katharine G. Abraham, Ron S. Jarmin, Brian Moyer, and Matthew D. Shapiro, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-80125-4 (cloth), 978-0-226-80139-1 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/big-data-twenty-first-century-economic-statistics

Conference Date: March 15-16, 2019

Publication Date: Februrary 2022

Chapter Title: Estimating the Benefits of New Products

Chapter Author(s): W. Erwin Diewert, Robert C. Feenstra

Chapter URL: https://www.nber.org/books-and-chapters/big-data-twenty-first-century-economic-statistics/estimating-benefits-new-products

Chapter pages in book: p. 437 – 473

15

Estimating the Benefits of New Products

W. Erwin Diewert and Robert C. Feenstra

15.1. Introduction

One of the more pressing problems facing statistical agencies and economic analysts is the new goods (and services) problem—that is, 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

W. Erwin Diewert is professor emeritus of the Vancouver School of Economics at the University of British Columbia, and a research associate of the National Bureau of Economic Research.

Robert C. Feenstra holds the C. Bryan Cameron Distinguished Chair in International Economics and is a professor of economics and director of the Center for International Data at the University of California–Davis, and a research associate of the National Bureau of Economic Research.

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 also thank Ninghui Li for her excellent research assistance. Financial support was received from a Digging into Data multi-country grant, provided by the United States NSF and the Canadian SSHRC. 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. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/books-and-chapters/big-data-21st-century-economic-statistics/estimating-benefits-new-products.

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 cor-

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 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 the demand curve is bounded, provided that the elasticity of substitution is greater than unity, and it can be computed with information on the expenditure

---

responding 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." (Hicks 1940, 114). Hofsten (1952, 95–97) extended Hicks's 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'" (Rothbarth 1941, 100).

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.

on the new goods and the elasticity. So Feenstra's methodology sidesteps 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 is sufficient information to compute the consumer benefits from new products.

In section 15.2, 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—that is, 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 15.2 are suggestive, they are not rigorous 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) when there are potentially many new and disappearing goods. Accordingly, in section 15.3 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 15.4, 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 15.4.1. 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 15.4.2–15.4.3, 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; that is, $N(N + 1)/2$ 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 15.4.4–15.4.5.

In section 15.4.6, 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 15.3, 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 about *six times the size* of the KBF gains, and their 95 percent confidence intervals do not overlap. 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 that is about three times 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 statistical agencies 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 15.5.[6]

15.2 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

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

6. The dataset on frozen juice products is listed in appendix A of our working papers (Diewert and Feenstra 2019a, 2019b). Certain results presented here are proved in appendixes B and C of Diewert and Feenstra (2019b).

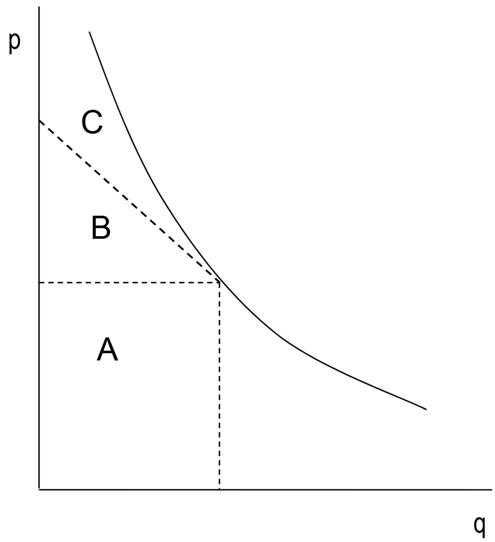

Fig. 15.1 Constant-elasticity demand

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 15.1 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 15.1, 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

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

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 15.1. 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

Table 15.1 **Consumer gains from a new product with share = 0.1 (% of expenditure)**

	$(B+C)/E_t$	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.

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 15.1, 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 15.1 for $s_{1t} = 0.1$ and various values of σ.

Column two in table 15.1 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 these results give us a first illustration of 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.

15.3 Utility-Based Approach

15.3.1 Utility Function Approach

We begin with a CES utility function for the consumer,[7] defined by

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

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 $t = 1, \ldots, T$ 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

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

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

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

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

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

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

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,

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

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

7. The CES function was introduced into the economics literature by Arrow et al. (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.

in (6) approaches zero, and a lower-bound of zero when the share of this product approaches one.[8]

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 1976). We maintain throughout the assumption that the observed quantities are optimally chosen for the prices; that is, that they correspond to the shares given in (6). When these shares are computed over the goods $i \in I$, we denote them as

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

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:

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

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

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

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,[9] 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:

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

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

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

8. 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 toward the price axis. Nevertheless, it is shown in appendix C of Diewert and Feenstra (2019b) that the Hicksian demand curve in (5) is strictly convex provided $s_{it} \le 0.5$.

9. 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 lies in between $s_{it}(I)$ and $s_{it-1}(I)$ for all goods $i \in I$.

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 $e(p_\tau, I)$, for $\tau = t-1, t$, 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.[10] The ratio $e(p_t,I)/e(p_{t-1},I)$ is still measured by the 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 rewritten as

$$s_{i\tau} \equiv \frac{p_{i\tau}q_{i\tau}}{\sum_{n\in I_\tau} p_{n\tau}q_{n\tau}} = s_{i\tau}(I)\lambda_\tau, \quad \tau = t-1, t;\ i \in I_t, \tag{12}$$

$$\text{with } \lambda_\tau \equiv \frac{\sum_{n\in I} p_{n\tau}q_{n\tau}}{\sum_{n\in I_\tau} p_{n\tau}q_{n\tau}}.$$

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

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

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 $\lambda_\tau(I) < 1$. 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, $\lambda_\tau(I)$ 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

10. Feenstra (1994) shows that we can instead define I as a nonempty 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.

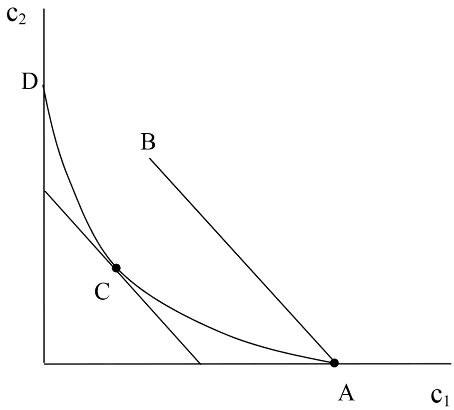

Fig. 15.2 CES indifference curve

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 15.2, 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:

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

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

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

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

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 15.1, 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 15.2. In the next section we compare these CES gains to an approximation of the measure of *total consumer utility* gain due to Hausman (2003).

15.3.2 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 15.2, but Hausman argues that it holds more generally for any Hicksian demand curves with constant utility. Letting $\eta_{1t}|_U$ 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:

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

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

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

In column six of table 15.1 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 the 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 nondecreasing, concave and homogeneous of degree one for $q \equiv (q_1, \ldots, q_N) \geq 0_N$, and twice continuously differentiable for $q \gg 0_N$. We suppose that the consumer faces positive prices $p_t \equiv (p_{1t}, \ldots, p_{Nt}) \gg 0_N$ 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*[11] with the period t quantity vector $q_t \gg 0_N$ solving (19) are

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

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

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

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

11. 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.

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) $$\begin{aligned} u'(q_{1t}) &= f_1(q_{1t}, q_{2t}) + f_2(q_{1t}, q_{2t})\partial q_2(q_{1t})/\partial q_1, \quad \text{differentiating (26)} \\ &= f_1(q_{1t}, q_{2t}) + f_2(q_{1t}, q_{2t})(-p_{1t}/p_{2t}), \quad \text{differentiating (25)} \\ &= 0, \quad \text{using (24)}, \end{aligned}$$

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 of Diewert and Feenstra (2019b), we calculate the second derivative $u''(q_{1t})$ and we show that it is nonpositive, so that the first term on the right of (30) is a nonnegative gain. Furthermore, we define an inverse demand function, $p_1 = D_1(q_1)$ that is consistent with our model; that is, 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 $u'''(\tilde{q}_1) \ge 0$. 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

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

$$= -\frac{1}{2} s_{1t}[D'(q_{1t})(q_{1t} / p_{1t})],$$

where the final term appearing in brackets in (31) is the *elasticity of the constant-utility inverse demand curve.* In appendix B of Diewert and Feenstra (2019b), 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 $\eta_{1t}|_U$, as shown in (7). More generally, we likewise expect that $[D'(q_{1t}(q_{1t}/p_{1t})])$ equals the inverse of $\eta_{1t}|_U$ 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.

15.3.3 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:[12]

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

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.[13] 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)

12. 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.

13. 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 $S \equiv \{q: q \gg 0_N \text{ and } Aq \gg 0_N\}$.

was a *flexible functional form*; that is, it can approximate an arbitrary twice continuously differentiable linearly homogeneous function to the accuracy of a second-order Taylor series approximation around an arbitrary positive quantity vector q^*. Since the Fisher quantity index gives exactly the correct utility ratio for the quadratic functional form defined by (32), he labeled 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 $p_t \gg 0_N$. The first order conditions (22) to maximize the utility function in (32) become

$$p_t = E_t A q_t / (q_t^T A q_t). \tag{33}$$

While these are the conditions for an interior maximum with $q_t \gg 0_N$, 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 $q_t \gg 0_N$ is

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

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},

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

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

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

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 approaches zero as the quantity q_{it} approaches zero in a neighborhood of the reservation price as $p_{it} \to p_{it}^*$ and $q_{it} \to 0$. 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 15.1. That conjecture can be established more formally, as we show in appendix C of Diewert and Feenstra (2019b). 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 of Diewert and Feenstra (2019b) 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.[14]

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 section 15.4, 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 15.3.2. Then our analysis of the Hausman approximation in that section continues to hold. Indeed, we show in appendix B of Diewert and Feenstra (2019b) that in this case the elasticity of the inverse demand curve is:

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

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

14. While we formally establish this result in appendix C of Diewert and Feenstra (2019b) 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.

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

15.4 Empirical Illustration Using CES and KBF Utility Functions

15.4.1 Scanner Data for Sales of Frozen Juice

We use the data from store number 5[15] in the Dominick's Finer Foods Chain of 100 stores in the Greater Chicago area on 19 varieties of frozen orange juice for three 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 sidestep the problems associated with this conversion by aggregating the weekly data into *pseudo-months*—which we simply refer to as "months"—that consist of four consecutive weeks.

Expenditure or sales shares, $s_{it} \equiv p_{it}q_{it} / \sum_{n=1}^{19} p_{nt}q_{nt}$, were computed for products $i = 1, \ldots, 19$ and months $t = 1, \ldots, 39$. We computed the sample average expenditure shares for each product. The bestselling products were products 1, 5, 11, 13, 14, 15, 16, 18, and 19. These products had a sample average share that exceeded 4 percent or a sample maximum share that exceeded 10 percent. There is tremendous volatility in product prices, quantities, and sales shares for both the bestselling and least popular products. 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 dataset.*

In the following sections, we will use this dataset 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.

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

In this section and the next, we will use the *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

15. This store is located in a northeast suburb of Chicago.

juice products because 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 ($i = 1, \ldots, N$; $t = 1, \ldots, T$):

- 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. We therefore allow for measurement error in the log shares due to the measurement error in prices, treating the quantities as accurate.[16]

Our goal is to estimate the elasticity of substitution for a CES direct utility function (3) that was discussed in section 15.3.1 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:

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

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,

16. See our working paper, Diewert and Feenstra (2019b), for other methods. We discuss there the more general technique from Feenstra (1994) that corrects for errors in prices, quantities, and expenditure shares.

(40) $$\ln s_{it} = \ln a_i + \frac{(\sigma - 1)}{\sigma}\ln q_{it} - \ln\left(\sum_{n=1}^{N} a_n q_{nt}^{(\sigma-1)/\sigma}\right) + u_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T,$$

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, for any variable x_{it} we difference the *logarithms* of x_{it} with respect to time; that is, define $\Delta \ln x_{it}$ as follows:

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

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

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

We apply this technique to obtain the *double-differenced log share* $\Delta^2 \ln s_{it}$, the *double-differenced log quantity* $\Delta^2 \ln q_{it}$, and the *double-differenced error variables* $\Delta^2 u_{it}$. Then using 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:

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

where the new residuals, $\Delta^2 u_{it}$, have means 0 and a constant covariance matrix with 0 covariances for observations that 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[17] the simplest possible system of estimating equations that one could imagine.

We have 15 product estimating equations of the form (43) that are estimated with STATA.[18] The resulting estimate for $(\sigma - 1)/\sigma$ was 0.849 (with a standard error of 0.006) and thus the corresponding estimated σ is equal to 6.62. The standard error on $(\sigma - 1)/\sigma$ was tiny using the present regression results so σ was very accurately determined using this method. The

17. 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 estimation of this system of equations; that is, we just used a standard systems nonlinear regression package that assumed intertemporal independence of the error terms.

18. The STATA code to obtain the results in this paper is available on request.

equation-by-equation R^2 for the 15 products $i = 1, \ldots, N - 1$ were as follows: 0.998, 0.996, 0.997, 0.990, 0.995, 0.994, 0993, 0.993, 0.990, 0.997, 0.991, 0.995, 0.997, 0.991, and 0.995. The average R^2 is 0.994, 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.[19] This double differencing method for estimating the elasticity of substitution worked much better than any other method that we tried.

15.4.3 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-Vartia $P_{SV}(I)$ 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 unity 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 $\sigma = 6.62$ gives the results shown in the third column of table 15.2. 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, which is reported along with its bootstrapped 95 percent confidence interval:[20]

$$G_{CES} = (\lambda_t / \lambda_{t-1})^{-1/(\sigma-1)}. \tag{44}$$

Recall that in month 9, products 2 and 4 make their appearance, and table 15.2 tells us that the effect of this increase in variety is to lower the price level and increase utility for month 9 by 0.83 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 0.40 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 0.9918 × 1.0040 × 0.9951 × 1.0044 × 0.9965 = 0.9918, for a decrease in the price level and increase in utility over the sample period of 0.83 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 canceling of the impact of availability of product 12 is a highly desirable feature of these CES

19. 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.

20. In our bootstrap, we resample with replacement the monthly observations across all products 500 times.

Table 15.2 **Changes in the price level and CES gains due to the availability of products, σ = 6.62**

	Availability	(λt / λt – 1)1 / (σ – 1)	GCES
9	2 and 4 new	0.9918	1.0083 [1.0075, 1.0091]
10	12 disappears	1.0040	0.9960 [0.9955, 0 .9963]
11	12 reappears	0.9951	1.0049 [1.0045, 1.0054]
20	12 disappears	1.0044	0.9956 [0.9952, 0.9960]
23	12 reappears	0.9965	1.0035 [1.0032, 1.0039]
Cumulative Gain		0.9918	1.0083 [1.0075, 1.0091]

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 reenters, 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 15.2 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 15.1, there are two potential problems with the Feenstra methodology: (i) the CES functional form is not fully flexible; and (ii) the reservation prices that 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.

15.4.4 Estimation of the KBF Utility Function

The quadratic or KBF utility function was introduced in section 15.3.3 above. Multiplying both sides of equation *i* in (33) by q_{it} and dividing by $p_t \cdot q_t = E_t$, we obtain the following *system of inverse demand share equations*:

$$(45) \qquad s_{it} \equiv \frac{p_{it}q_{it}}{p_t \cdot q_t} = \frac{q_{it}\sum_{n=1}^{N} a_{in}q_{nt}}{q_t^T A q_t}, \quad i = 1, \ldots, N,$$

where a_{in} is the element of A that is in row *i* and column *n* for $i, n = 1, \ldots, N$. 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:

(46) $A = bb^T + B$; $b \gg 0_N$; $B = B^T$; B 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 1 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 $B = 0_{N \times N}$, a matrix of 0's. The resulting utility function becomes $f(q) = (q^T bb^T q)^{1/2} = (b^T q b^T q)^{1/2} = b^T q$, 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 15.3.3), 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.[21] Write C as $[c^1, c^2, \ldots, c^N]$ where c^k is a column vector for $k = 1, \ldots, N$. If C is lower triangular, then the first $k - 1$ elements of c^k are equal to 0, $k = 2, 3, \ldots, N$. Thus, we have the following representation for B:

$$B = -CC^T = -\sum_{k=1}^{19} C^k C^{kT}, \tag{47}$$

where we impose the following restrictions on the vectors c^k in order to impose the restrictions $Bq^* = 0_N$ on B:[22]

$$c^{kT} q^* = 0; \; k = 1, \ldots, N. \tag{48}$$

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 the

21. $C = [c_{nk}]$ is a lower triangular matrix if $c_{nk} = 0$ for $k > n$; that is, there are 0's in the upper triangle. Wiley, Schmidt, and Bramble (1973) 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.

22. 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).

parameters in the C matrix. Furthermore, typically nonlinear estimation is not successful if one attempts to estimate all 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 $c^{1T} \equiv [c_1^1, c_2^1, \ldots, c_N^1]$ 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 $c^1 \equiv 0_N$.[23] In the third stage, we estimate $f(q) = (q^T[bb^T - c^1c^{1T} - c^2c^{2T}]q)^{1/2}$ where $c^{1T} \equiv [c_1^1, c_2^1, \ldots, c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0, c_2^2, \ldots, c_N^2]$ 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 $c^{1T} \equiv [c_1^1, c_2^1, \ldots, c_N^1]$, $c^{1T}q^* = 0$, $c^{2T} \equiv [0, c_2^2, \ldots, c_N^2]$, $c^{2T}q^* = 0$, $c^{3T} \equiv [0, 0, c_3^3, \ldots, c_N^3]$ and $c^{3T}q^* = 0$. At each stage, the log likelihood will generally increase.[24] We stop adding columns to the C matrix when the increase in the log likelihood becomes small (or the number of degrees of freedom becomes small). At stage k of this procedure, it turns out that we are estimating the substitution matrices of rank $k - 1$ that is the most negative semidefinite that the data will support. This is the same type of procedure that Diewert and Wales (1988) used 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.

15.4.5 The Estimation of KBF Preferences Using Price Equations

We considered two methods for estimating the KBF utility function. The first used a stochastic version of the share equations (45).[25] When we applied that method to predict prices for products that were actually available, it performed rather poorly, giving us little confidence that the reservation prices for products *not* available would be reliable. Accordingly, we switched from estimating share equations to the estimation of price equations. We considered the system of estimating equations using prices as the dependent variables, as was shown in (33):

$$(49) \quad p_{it} \equiv E_t \textstyle\sum_{j=1}^{19} a_{ij}q_{jt} / [\sum_{n=1}^{19}\sum_{m=1}^{19} a_{nm}q_{nt}q_{mt}] + \varepsilon_{it}, t = 1, \ldots, 39; i = 1, \ldots, 18,$$

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 (49) into a single nonlinear regression and we drop the observations that correspond to products i that were not available in period t.

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

24. 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 because the successive models are nested.

25. See our working paper, Diewert and Feenstra (2019b).

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 as compared to using shares as the dependent variable. 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 (49) evaluated at our new parameter estimates. The equation-by-equation R^2 comparing the predicted prices for the 19 products with the actual prices were as follows: 0.830, 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 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 matrix; that is, we set $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T} - c^5c^{5T}$, where $c^{5T} = [0, 0, 0, 0, c_5^5, \ldots, c_{19}^5]$ 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 $n = 5, 6, \ldots, 18$. 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 $c^{6T} = [0, 0, 0, 0, c_6^6, \ldots, c_{19}^6]$ 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 $n = 6, 7, \ldots, 18$. 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 to converge to the maximum likelihood estimates. Thus, we stopped our sequential estimation process at this point.

The parameter estimates for the rank 6 substitution matrix are listed below in table 15.3.

The estimated b_n in table 15.3 for $n = 1, \ldots, 18$ 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 15.3, it can be seen that the highest-quality products were products 6, 17, and 4

Table 15.3 **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

($b_6 = 2.09$, $b_{17} = 1.58$, $b_4 = 1.57$) 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 $k = 1, \ldots, 6$), 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 $i, j = 1, \ldots, 19$. The *predicted price* for product i in month t is calculated 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: 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 12 of the 19 equations had an R^2 greater than 0.70, while five of the equations had an R^2 less than 0.40 (product 12 had $R^2 = 0.284$).[26]

15.4.6 The Gains and Losses Due to Changes in Product Availability

In this section, we consider a 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 $p_t = (p_{1t}, \ldots, p_{Nt}) \geq 0_N$ and $q_t = (q_{1t}, \ldots, q_{Nt}) \geq 0_N$, respectively, for $t = 1, \ldots, T$. Sales or expenditures on the N products during period t are $E_t \equiv p_t \cdot q_t$ for $t = 1, \ldots, T$.[27] We assume that a linearly homogeneous utility function, $f(q_1, \ldots, q_N) = f(q)$, has been estimated where $q \geq 0_N$.[28] 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 refer to these as *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 $f(q)$:

$$p_{it}^* \equiv E_t[\partial f(q_t)/\partial q_i]/f(q_t), \quad i = 1, \ldots, N; t = 1, \ldots, T. \tag{50}$$

The prices defined by (50) are also Rothbarth's (1941) *virtual prices*; they are the prices that 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),[29] 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 $i = 1, \ldots, N$. 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:

26. The sample average expenditure shares of these low R^2 products were 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.

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

28. 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.

29. This assumes that observed prices are the dependent variables in the estimating equations.

(51) $$E_t^* \equiv \sum_{i=1}^N p_{it}^* q_{it},\ t = 1, \ldots, T$$
$$= \sum_{i=1}^N q_{it} E_t[\partial f(q_t) / \partial q_i] / f(q_t), \quad \text{using definition (50)}$$
$$= E_t,$$

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 consumer utility during period t*, U_t, as follows:

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

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:

(53) $$U_{\sim 1t} \equiv f(0, q_{\sim 1t})$$
$$\leq f(q_{1t}, q_{\sim 1t}) \text{ since } f(q) \text{ is nondecreasing in the components of } q$$
$$= U_t \text{ using definition (52).}$$

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

(54) $$E_{\sim 1t}^* \equiv \sum_{i=2}^N p_{it}^* q_{it}$$
$$= E_t - p_{1t}^* q_{1t} \quad \text{using (51)}$$
$$\leq E_t \text{ since } p_{1t}^* \geq 0 \text{ and } q_{1t} > 0.$$

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

(55) $$\lambda_1 \equiv E_{\sim 1t}^* / E_t \leq 1 \quad \text{using (54).}$$

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_{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:

$$(56) \qquad \Sigma_{i=2}^{N} p_{it}^{*} q_{it} / \lambda_1 = (1/\lambda_1)\Sigma_{i=2}^{N} p_{it}^{*} q_{it}$$

$$= (1/\lambda_1) E_{1t}^{*} \quad \text{using definition (54)}$$

$$= (E_t / E_{\sim 1t}^{*}) E_{\sim 1t}^{*} \quad \text{using definition (55)}$$

$$= E_t .$$

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:

$$(57) \qquad U_{A1t} \equiv f(0, q_{\sim 1t} / \lambda_1)$$

$$= (1/\lambda_1) f(0, q_{\sim 1t}) \quad \text{using the linear homogeneity of } f$$

$$= (1/\lambda_1)\, U_{\sim 1t} \quad \text{using definition (53).}$$

Note that λ_1 can be calculated using definition (55) and $U_{\sim 1t}$ can be calculated using definition (53). 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:

$$(58) \qquad \max_{q's}\{f(0,q_2,q_3,\ldots,q_N) \colon \Sigma_{i=2}^{N} p_{it}^{*} q_{it} \le E_t\} \equiv U_{1t}$$

$$\ge U_{A1t},$$

where U_{A1t} is defined by (57). The inequality in (58) follows because (56) shows that $q_{\sim 1t}/\lambda_1$ is a feasible solution for the utility maximization problem defined by (58). 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

$$(59) \qquad U_t \ge U_{1t} \ge U_{A1t}.$$

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 1 is not available. The *actual* constrained utility level, U_{1t}, will in general involve solving numerically the nonlinear programming problem defined by (58). For the KBF functional form, instead of maximizing $(q^T A q)^{1/2}$, we could maximize its square, $q^T A q$, and thus solving (58) 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 (58) because 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}—that is, define *the product* 1 *utility gain in period t* as

$$(60) \qquad G_{1t} \equiv U_t / U_{1t} \geq 1,$$

where the inequality follows from (59). The corresponding *product* 1 *approximate utility gain* is defined as

$$(61) \qquad G_{A1t} \equiv U_t / U_{A1t} \geq G_{1t} \geq 1,$$

where the inequalities follow again from (59). 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) assumed no measurement error in prices, so that $p_{it} = p_{it}^*$, and we have

$$(62) \quad G_{A1t} = \frac{U_t}{U_{A1t}} = \lambda_{1t}\frac{U_t}{U_{\sim 1t}} \qquad \text{from definitions (57) and (61)}$$

$$= \frac{\sum_{i=2}^{N} p_{it}^* q_{it}}{E_t}\frac{U_t}{U_{\sim 1t}} \qquad \text{from definition (55)}$$

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

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

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

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 (44) when applied to the case of new product 1. In other words, the earlier CES gain is identical to the 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 that 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—that is, define *the product* 1 *utility loss that can be attributed to the disappearance of product* 1 *in period* $t + 1$ as

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

where the inequality follows from (59). The corresponding *product* 1 *approximate utility loss* is defined as

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

where the inequalities again follow from (59). Thus, in general the approximate loss is a 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 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 15.4. 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 15.4. In the fourth column of table 15.4 we repeat the CES gain in utility from table 15.2 for period 9 due to the introduction of products

Table 15.4 **The gains and losses of utility due to changes in product availability**

Month	Availability	$G_{A,KBF}$ $L_{A,KBF}$	G_{CES} ($\sigma = 6.62$)
9	2 and 4 new	1.0013	1.0083
		[1.0009, 1.0040]	[1.0075, 1.0091]
10	12 disappears	0.9975	0.9959
		[0.9935, 0.9996]	[0.9955, 0.9963]
11	12 reappears	1.0030	1.0049
		[1.0005, 1.0088]	[1.0045, 1.0054]
20	12 disappears	0 .9988	0.9956
		[0.9968, 0.9998]	[0.9952, 0.9960]
23	12 reappears	1.0008	1.0035
		[1.0001, 1.0020]	[1.0032, 1.0039]
Cumulative Gain		1.0014	1.0083
		[1.0011, 1.0047]	[1.0075, 1.0091]

2 and 4, and the various impacts of the exit and entry of product 12. Thus, table 15.4 compares the gains and losses in utility for the KBF and CES models for the five months in which 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 increases consumers' utility by 0.83 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 over all several months.

Recall that in our earlier calculation of the CES gain (see table 15.2), the net effect on utility of the entry and exit of product 12 canceled out, so that the overall utility gains came only from the initial entry or products 2 and 4. That result roughly holds in the KBF case, too, 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 15.4) 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 six times higher than the KBF gains* in table 15.4 (and their bootstrapped 95 percent confidence intervals do not overlap). That is a surprising result because 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 15.3.3 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 that 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 about *six* to get the estimated KBF gain.

The resolution to these surprising empirical results is that the *KBF and CES demand curves must have different slopes 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 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 $s_{2t} \equiv 1 - s_{1t}$.[30] With these modifications, we can calculate $G_{H,KBF}$ and $G_{H,CES}$ for each product and 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 15.5, together with the bootstrapped 95 percent confidence intervals.[31]

From table 15.5, 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 0.17 percentage points using our estimated KBF utility function and about 0.46 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 three. 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 the Hausman approximation to the CES gains that are about three times as high on average as the Hausman approximation to the KBF gains, it must be that the elasticity of demand for the KBF function is about three times as high as for the CES.[32] With the results shown in table 15.5, it is not surprising that the CES gains (from products 2 and 4) *are six times higher than the KBF gains* in table 15.4: about three times

30. The KBF shares that we use for this exercise are fitted shares; that is, we use the actual quantities that are observed in period t, q_{it}, and the estimated prices $p_{it}^{*} \equiv f_1(q_t)E_t / f(q_t)$ where $f(q)$ is the estimated utility function. In the CES case, we use the observed shares for simplicity.

31. The bootstrap uses 500 draws with replication. In some cases, the estimated coefficient was below the 95 percent confidence interval obtained by dropping the top and bottom 2.5 percent of observations. In those cases, we dropped fewer observations at the bottom and more at the top (still dropping 5 percent in total), so that the coefficient was within the confidence interval.

32. In appendix B of Diewert and Feenstra (2019b), table B1, we report some average elasticities for each product that are quite similar to the elasticities of inverse demand.

Table 15.5 **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	0.0041	0.0042	11	0.0034	0.0034
	[0.0029, 0.0139]	[0.0039, 0.0046]		[0.0011, 0.0129]	[0.0031, 0.0037]
2	0.0008	0.0017	12	0.0021	0.0019
	[0.0006, 0.0052]	[0.0015, 0.0018]		[0.0004, 0.0057]	[0.0018, 0.0021]
3	0.0006	0.0026	13	0.0056	0.0221
	[0.0004, 0.0038]	[0.0024, 0.0029]		[0.0039, 0.0108]	[0.02037, 0.0239]
4	0.0008	0.0021	14	0.0009	0.0057
	[0.0004, 0.0020]	[0.0020, 0.0023]		[0.0004, 0.0108]	[0.0053, 0.0062]
5	0.0033	0.0095	15	0.0009	0.0017
	[0.0026, 0.0091]	[0.0088, 0.0103]		[0.0003, 0.0075]	[0.0016, 0018]
6	0.0001	0.0027	16	0.0031	0.01012
	[0.0001, 0.0013]	[0.0025, 0.0029]		[0.0016, 0.0121]	[0.0093, 0.0110]
7	0.0005	0.0030	17	0.0019	0.0021
	[0.0005, 0.0040]	[0.0028, 0.0033]		[0.0003, 0.0034]	[0.0020, 0.0024]
8	0.0010	0.0020	18	0.0011	0.0039
	[0.0002, 0.0069]	[0.0018, 0.0022]		[0.0007, 0.0047]	[0.0036, 0.0042]
9	0.0008	0.0020	19	0.0004	0.0041
	[0.0006, 0.0038]	[0.0019, 0.0022]		[0.0004, 0.0165]	[0.0037, 0.0044]
10	0.0005	0.0014	Mean	0.0017	0.0046
	[0.0003, 0.0031]	[0.0013, 0.0016]		[0.0017, 0.0041]	[0.0042, 0.0049]

within this difference comes from having more elastic demand for the KBF than for the CES utility functions (so that the Hausman linear approximation to the gains in the CES case are nearly three times as high as in the KBF case), while the other two times comes from CES demand curves being more convex (with gains about twice higher) than KBF demand.

15.5 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.[33] 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:

- The Feenstra CES methodology for dealing with 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 15.3.1.
- It is not a trivial matter to obtain an accurate estimate for σ. Section 15.4.2 developed one methodological approach to the estimation of the elasticity of substitution if purchasers of products have CES preferences. These methods 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 15.3.3 we discussed the utility function $f(q) \equiv (q^T A q)^{1/2}$, 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 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 it is a semiflexible functional form that is similar to the normalized quadratic semiflexible form introduced by Diewert and Wales (1987, 1988).
- In section 15.4.5 we estimated the unknown parameters in the A matrix using prices as the dependent variables. This approach generated satisfactory point estimates for the KBF functional form, but because of

33. Generally, it is challenging to estimate direct demand functions when there are new goods because the reservation prices for goods not available—which will influence the demand for available goods—are unknown. In some cases, the reservation prices can be solved as a function of observed prices and quantities for available goods, and therefore included in the estimation (see Feenstra and Weinstein (2017) for an application to a symmetric translog expenditure function). This problem does not arise when the inverse demand functions are estimated instead, because then the quantity for goods that are not available is simply zero, which can be used in the inverse demand equations for all goods that are available.

the large number of parameters, many of the individual estimates are insignificantly different from zero.

- The results presented in section 15.4.6 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 about six times greater than the KBF gains*: in rough terms, about three times within this difference comes from having more elastic demand for the KBF than for the CES utility functions (so that the Hausman linear approximation to the gains in the CES case are three times as high as in the KBF case), while the other two times comes from CES demand curves being more convex (with gains about twice higher) than KBF demand. Furthermore, the confidence intervals for these estimates of gains in the KBF and CES cases do not overlap.

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; that is, it is not a fully flexible functional form. With that simplification, they confirm that the translog case is 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.[34] 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 (2020), who also use a CES utility function. They assume that this functional form represents the "true" preferences, so that any observed deviation

34. Note, however, that Feenstra and Weinstein (2017) find another source of gains from new goods in the translog case, and that is a procompetitive effect on lowering the markups on existing goods. This procompetitive effect does not occur under a CES utility function because then markups are fixed. When this procompetitive effect is added to the gains from new products in the translog case, the *total* gains are comparable in size to what Broda and Weinstein (2006) estimate as the gains from new products in the CES case,

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 toward that good. They argue that the consumer gain from that price reduction is *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.

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* 63:225–50.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121 (2): 541–85.

Diewert, W. E. 1974. "Applications of Duality Theory." In *Frontiers of Quantitative Economics*, vol. 2, edited by M. D. Intriligator and D. A. Kendrick, 106–71. Amsterdam: North-Holland.

———. 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4:114–45.

———. 1980. "Capital and the Theory of Productivity Measurement." *American Economic Review* 70 (2): 260–67.

———. 1987. "Index Numbers." In *The New Palgrave: A Dictionary of Economics*, vol. 2, edited by J. Eatwell, M. Milgate, and P. Newman, 767–80. London: Macmillan.

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

———. 1999. "Index Number Approaches to Seasonal Adjustment." *Macroeconomic Dynamics* 3:48–67.

———. 2018. "Duality in Production." Discussion Paper 18-02, Vancouver School of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W. E., and R. Feenstra. 2019a. "Estimating the Benefits of New Products: Some Approximations." Discussion Paper 19-02, Vancouver School of Economics, University of British Columbia, Vancouver, Canada.

———. 2019b. "Estimating the Benefits of New Products." NBER Working Paper No. 25991, National Bureau of Economic Research, Cambridge, MA.

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, University of British Columbia, Vancouver, Canada.

Diewert, W. E., and R. J. Hill. 2010. "Alternative Approaches to Index Number Theory." In *Price and Productivity Measurement*, edited by W. E. Diewert, Bert M. Balk, Dennis Fixler, Kevin J. Fox, and Alice O. Nakamura, 263–78. Victoria, Canada: Trafford Press.

Diewert, W. E., and T. J. Wales. 1987. "Flexible Functional Forms and Global Curvature Conditions." *Econometrica* 55:43–68.
———. 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): 157–77.
———. 2010. "New Products with a Symmetric AIDS Expenditure Function." *Economic Letters* 2:108–11.
Feenstra, R. C., and M. D. Shapiro. 2003. "High Frequency Substitution and the Measurement of Price Indexes." In *Scanner Data and Price Indexes*, Studies in Income and Wealth, vol. 64, edited by R. C. Feenstra and M. D. Shapiro, 123–49. Chicago: University of Chicago Press.
Feenstra, Robert C., and Clinton Shiells. 1997. "Bias in U.S. Import Prices and Demand." In *The Economics of New Goods*, edited by Timothy Bresnahan and Robert Gordon, 249–76. Chicago: University of Chicago Press.
Feenstra, Robert C., and David E. Weinstein. 2017. "Globalization, Markups, and U.S. Welfare." *Journal of Political Economy* 125 (4): 1041–74.
Fisher, Irving. 1922. *The Making of Index Numbers*. Boston: Houghton-Mifflin.
Hardy, G. H., J. E. Littlewood, and G. Polyá. 1934. *Inequalities*. Cambridge: Cambridge University Press.
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. 1996. "Valuation of New Goods under Perfect and Imperfect Competition." In *The Economics of New Goods*, edited by T. F. Bresnahan and R. J. Gordon, 209–37. Chicago: University of Chicago Press.
———. 1999. "Cellular Telephone, New Products and the CPI." *Journal of Business and Economic Statistics* 17(2): 188–94.
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): 237–63.
Hicks, J. R. 1940. "The Valuation of the Social Income." *Economica* 7:105–24.
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 1939 in *Econometrica* 7:10–29.
Konüs, A. A., and S. S. Byushgens. 1926. "K probleme pokupatelnoi cili deneg." *Voprosi Konyunkturi* 2:151–72.
Marshall, A. 1887. "Remedies for Fluctuations of General Prices." *Contemporary Review* 51:355–75.
Redding, Stephen, and David E. Weinstein. 2020. "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences." *Quarterly Journal of Economics* 135 (1): 503–60.
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* 58:223–28.
Shephard, R. W. 1953. *Cost and Production Functions*. Princeton, NJ: Princeton University Press.
University of Chicago. 2013. *Dominick's Data Manual*. Chicago: James M. Kilts Center, University of Chicago Booth School of Business.
Vartia, Y. O. 1976. "Ideal Log-Change Index Numbers." *Scandinavian Journal of Statistics* 3:121–26.
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* 68:317–23.