Estimating the Benefits and Costs of New and Disappearing Products
(Preliminary Version)

by W. Erwin Diewert, University of British Columbia and UNSW

Presented at the

Center for Applied Economic Research,

EMG Workshop 2017,

Pioneer International Theatre, AGSM Building, UNSW Sydney.

December 1, 2017.

Introduction

- **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 and that 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.**
- **With these virtual (or reservation or imputed) 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* or *virtual prices to be estimated*?**

Introduction (cont)

- **There are two approaches to solving this estimation problem that have been suggested in the literature on this topic:**
- **(1) Feenstra's (1994) Approach. This approach assumes that purchasers of a group of (related) products have CES preferences. His approach is quite clever and requires only observed data on the prices and quantities purchased for two consecutive periods plus an estimate of the elasticity of substitution σ between all pairs of products. The practical problem boils down to: how exactly should we estimate this elasticity of substitution? In the early part of this paper, we will look at alternative methods for estimating σ.**
- **(2) Hausman's (1996) Econometric Approach which involved estimating the AIDS expenditure function and calculating reservation prices. Hausman (2003) also suggested a simple consumer surplus approach which we will look at in section 11. We will also explain the problem with his approach.**

Introduction (cont)

- **There are two major problems with Feenstra's CES approach:**
  (1) the CES functional form is not flexible and
  (2) the Feenstra reservation prices for missing products are equal to $+\infty$. This seems to be a rather high reservation price; typically, it does not take an infinite price to deter potential purchasers from buying a product. Thus there is a good possibility that the Feenstra methodology exaggerates the benefits of increasing product variety.
- **We will implement Feenstra's methodology using a data set on frozen juice sales in a Chicago grocery store. This data set is available on line.**
- **We will also implement an alternative methodology and compare the results with the Feenstra results.**
- **The new alternative methodology makes use of a new semiflexible functional form that is exact for the Fisher index.**

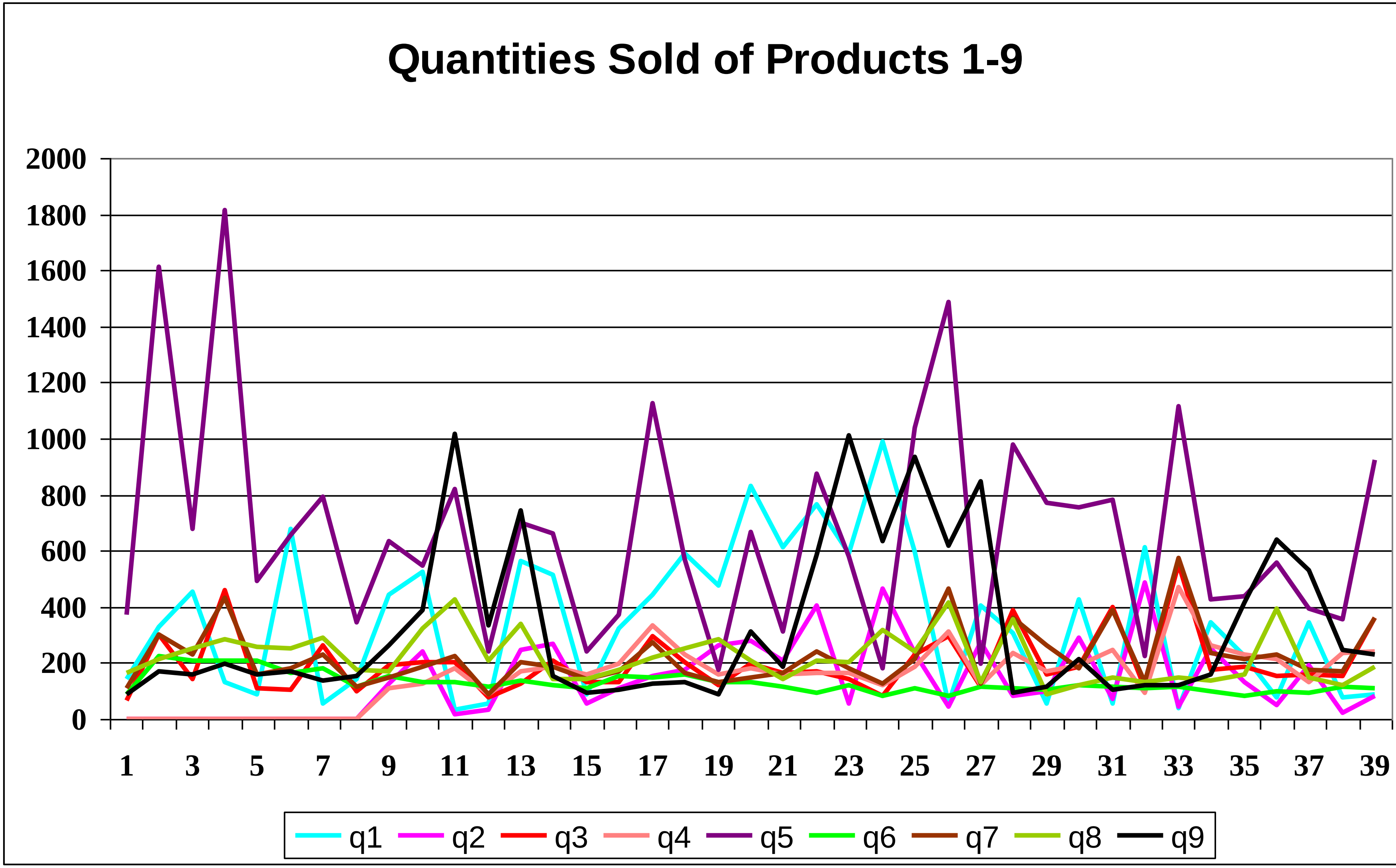
The Data

- **We will use the data from Store Number 5 in the Dominick's Finer Foods Chain of 100 stores in the Greater Chicago area on 19 varieties of frozen orange juice for 3 years in the period 1989-1994 in order to test out the CES models explained in the previous two sections; see the University of Chicago (2013) for the micro data.**
- **The micro data are weekly quantities sold of each product and the corresponding unit value price.**
- **The weekly price and quantity data need to be aggregated into monthly data. Since months contain varying amounts of days, we are immediately confronted with the problem of converting the weekly data into monthly data. We decided to side step the problems associated with this conversion by aggregating the weekly data into pseudo-months that consist of 4 consecutive weeks. We ended up with data for 39 "months".**

The Data (cont)

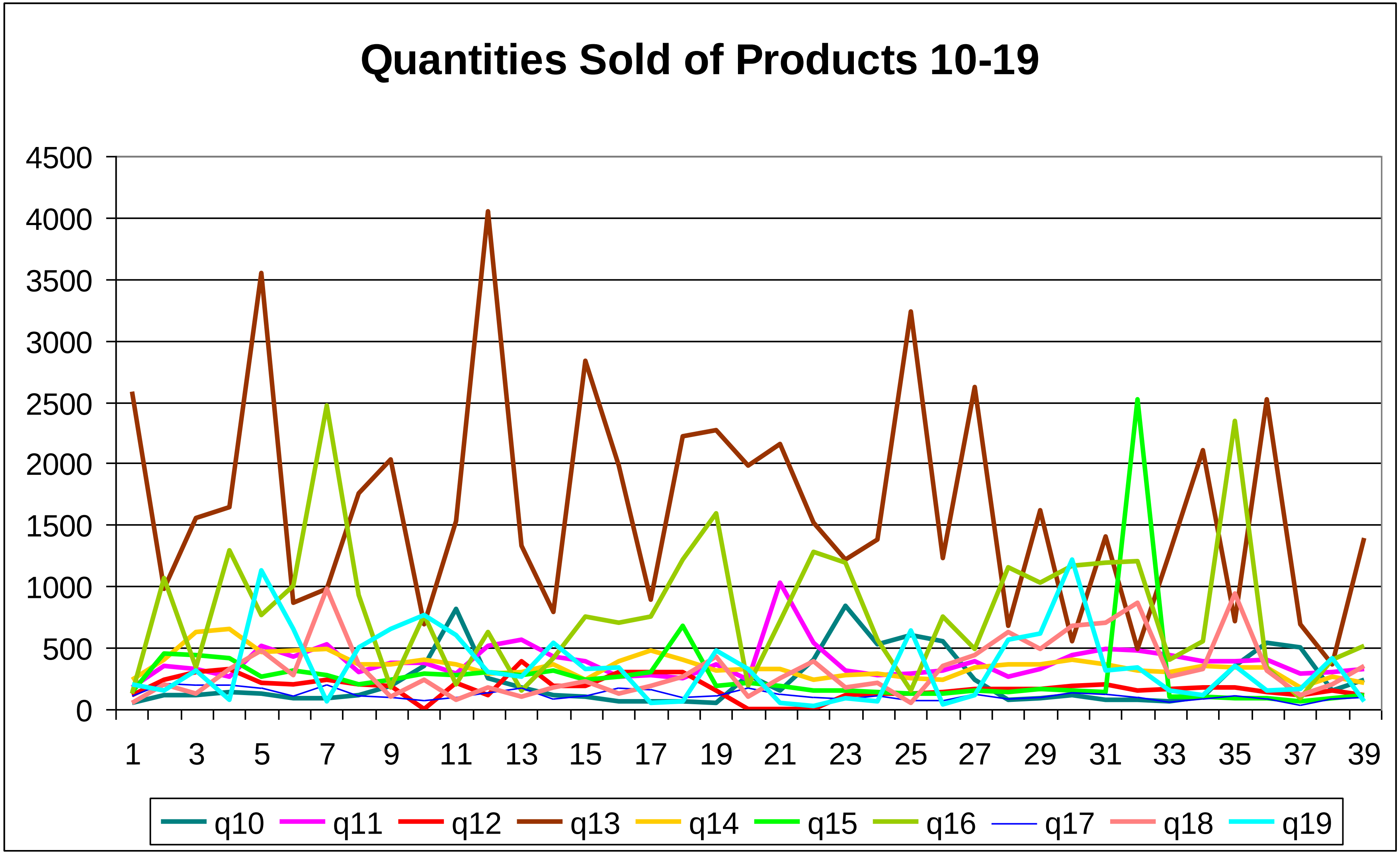
- **There were no sales of Products 2 and 4 for "months" 1-8 and there were no sales of Product 12 in "month" 10 and in "months" 20-22.**
- **Thus there is a new and disappearing product problem for 20 observations (out of 741 total observations on all 19 products and all 39 "months" in this data set.**
- **Later in the paper, we will impute Hicksian reservation prices for these missing products.**
- **The corresponding imputed quantity for a missing observation is set equal to 0.**
- **In the following slides, we plot the prices and (normalized) quantities in our data set so that one can see the tremendous variability in the data (even when it has been aggregated into "months").**

Quantity Data



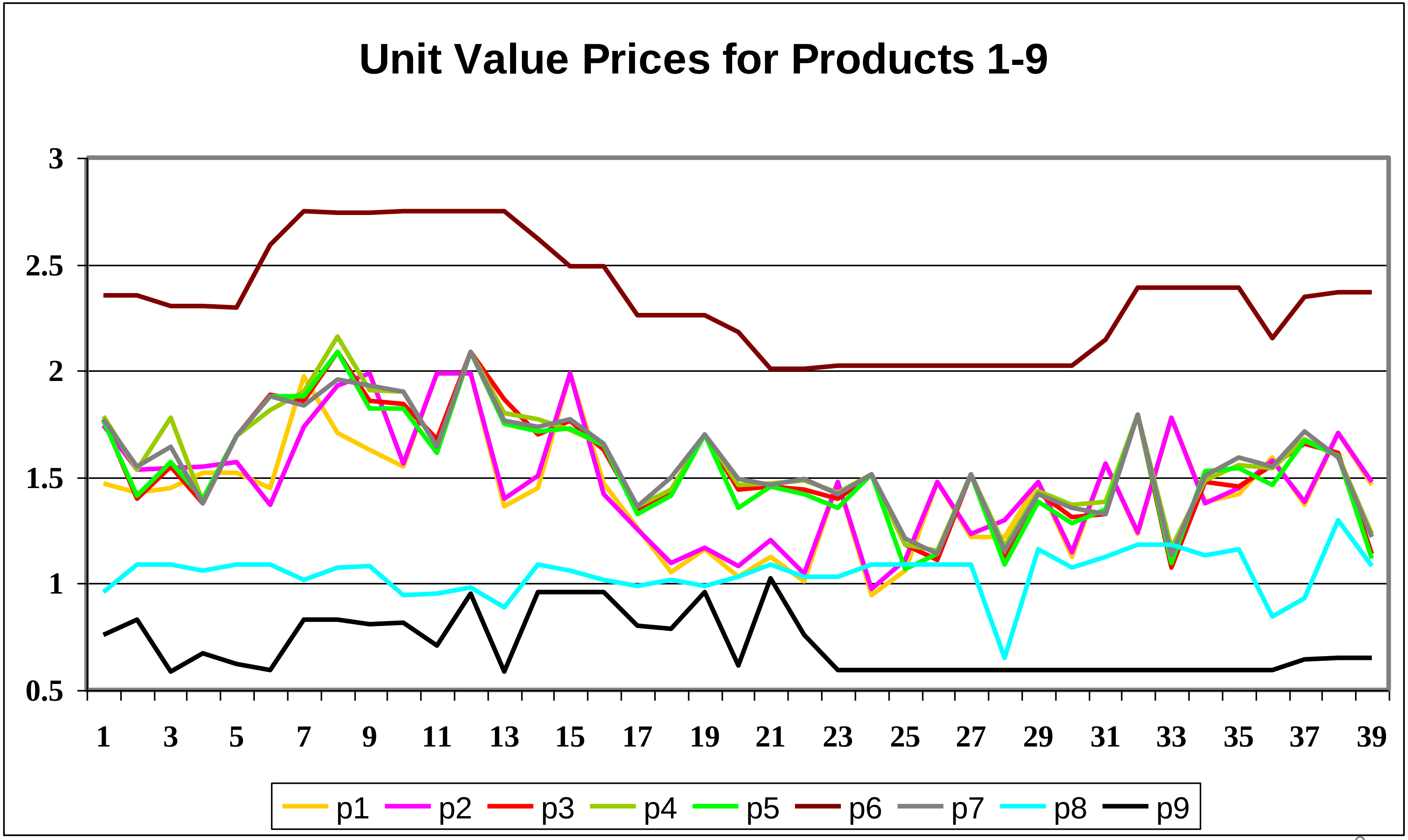
More Quantity Data

There is tremendous variation in the monthly quantities sold.



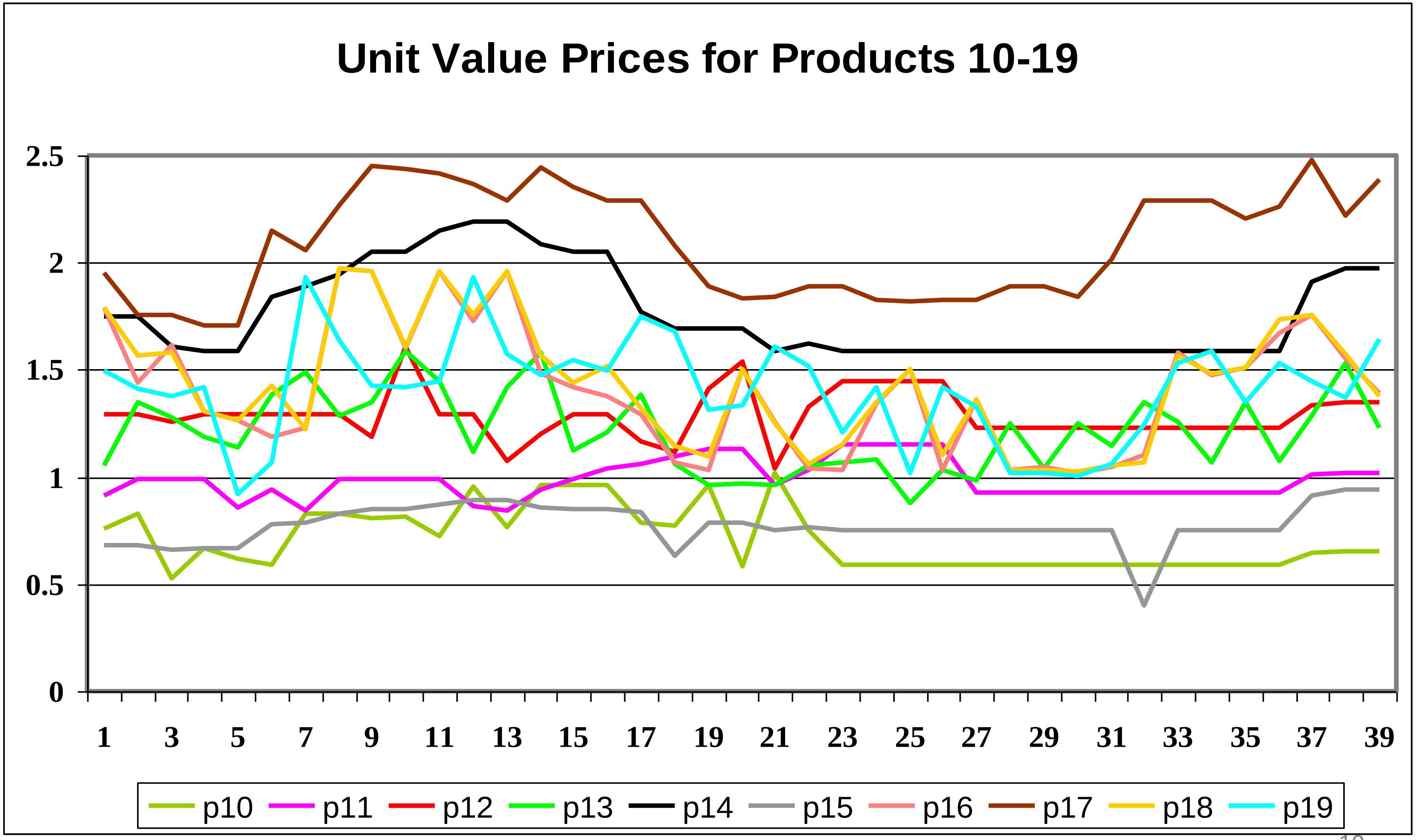
Price Data

Note: Imputed Reservation Prices are used for Products 2 and 4 for “months” 1-8 below.



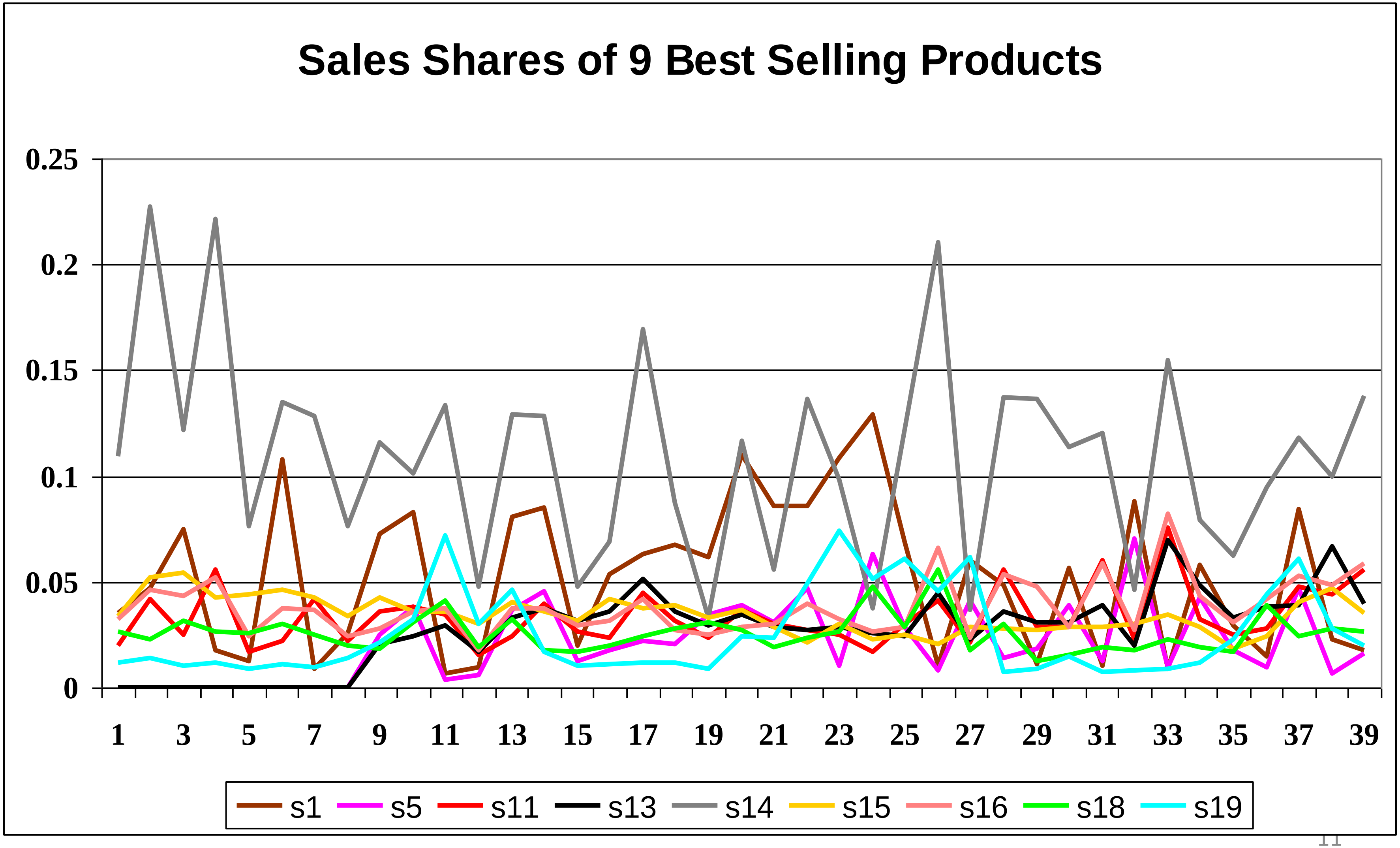
More Price Data

Prices vary much less than the quantity variation.



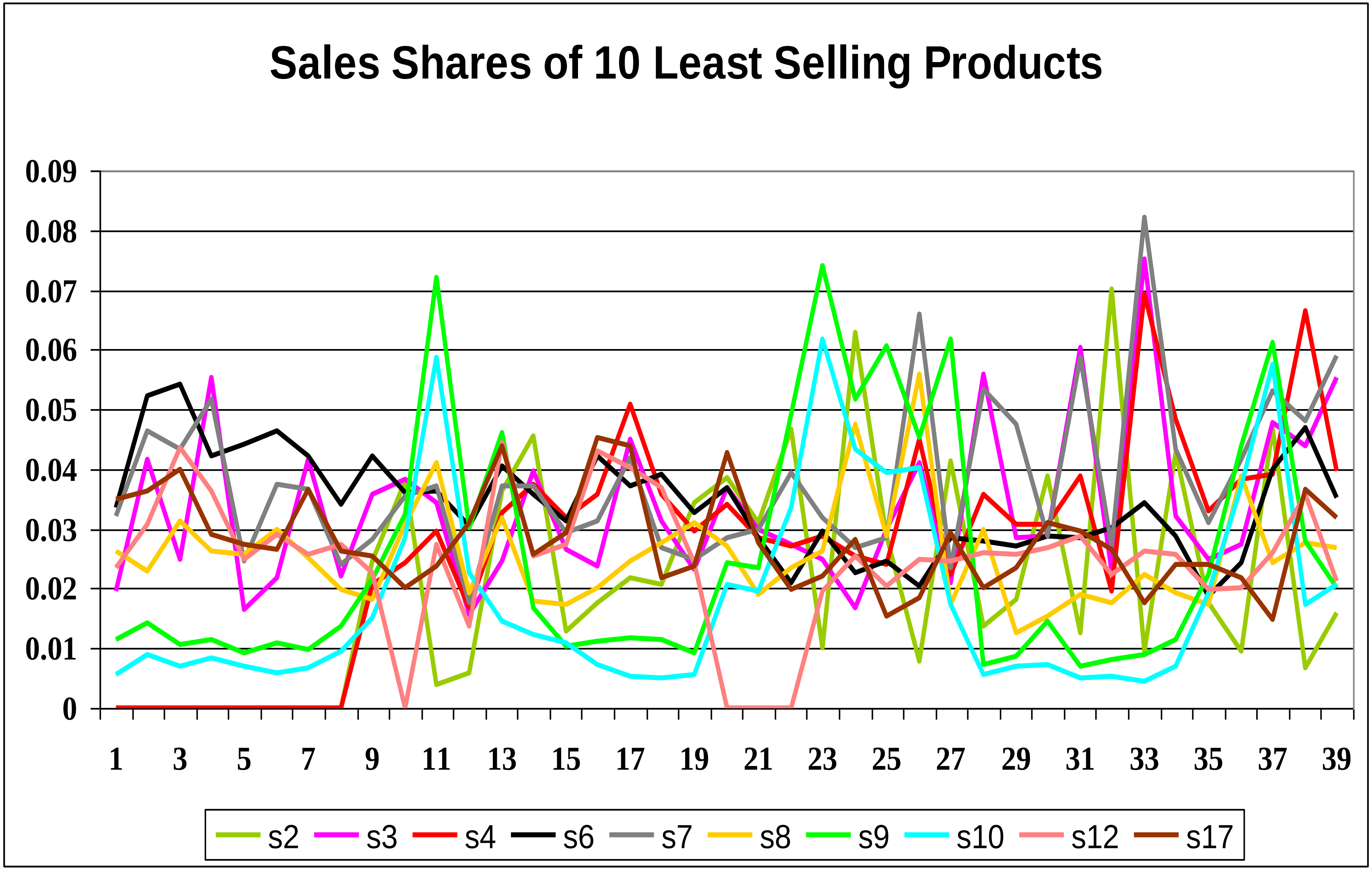
Sales Share Data

Shares fluctuate more than prices but less than quantities



Sales Shares of Less Selling Products

The cycles in the volatility differ across products.



The Feenstra CES Unit Cost Function Model

- The **unit cost function** that is dual to a linearly homogeneous utility function is defined as follows:

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

- Feenstra assumed that the unit cost function has the following **CES functional form**:

(2) $c(p) \equiv \alpha_0 [\sum_{n=1}^{N} \alpha_n p_n^{1-\sigma}]^{1/(1-\sigma)}$ where $\sigma > 1$

$= \alpha_0 [\sum_{n=1}^{N} \alpha_n p_n^{r}]^{1/r}$ where $r = 1 - \sigma < 0$

and where $\sum_{n=1}^{N} \alpha_n p_n = 1$ (or $\alpha_N = 1$).

- The resulting system of estimating **share equations** is:

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

- The unknown parameters to be estimated are the α_n and $r < 0$.
- Feenstra assumes that the reservation price for say product i in period t that is not present in that period is equal to $+\infty$. But since $r < 0$, $(p_i^t)^r = (+\infty)^r = 0$ and hence s_i^t defined by (5) = 0 as well, which is the correct answer!

The CES Unit Cost Function Model: Systems Approach

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

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

- **We used the nonlinear regression software package in Shazam to estimate the unknown α_i in equations (51).**
- **The final log likelihood turned out to be 2034.884. The equation by equation R^2 values were as follows: 0.6138, 0.1277, 0.5476, 0.4875, 0.2376, 0.1191, 0.5014, 0.0172, 0.0761, 0.047, 0.0016, 0.4232, 0.6578, 0.0012, 0.5826, 0.2973, 0.1481 and 0.2323.**
- **Thus the fits for this preliminary regression were not very good but this is to be expected: *Model 1* defined by equations (51) corresponds to preferences that exhibit *no* substitution between products, which is implausible for closely related products.**
- **Note that 0 prices for missing products are not a problem for this very simple no substitution model.**

The CES Unit Cost Function Model: Systems Approach

- **Shazam was used to estimate the 19 unknown parameters in equations (5), the actual CES unit cost function. The final log likelihood for this *Model 2* was 2195.039, an increase of 160.155 over the previous Model 1 regression for adding one parameter.**
- **The estimate for r was –2.8041 with a standard error equal to 0.12939. Hence the resulting point estimate for the elasticity of substitution is $\sigma \equiv 1 - r = 3.8041$. Thus there is a considerable amount of substitution between the 19 frozen juice products.**
- **The equation by equation R^2 values were as follows: 0.6357, 0.7179, 0.6407, 0.8221, 0.3619, 0.0051, 0.6652, 0.0268, 0.0572, 0.0206, 0.0109, 0.4286,0.7419, 0.0781, 0.8050, 0.3370, 0.1589 and 0.2673.**
- **These R^2 values are considerably higher than the corresponding ones from the first regression model. However, the average R^2 was only equal to 0.3767 which is not very satisfactory.**

The Feenstra CES Index Number Methodology

- **The CES true cost of living index between periods t and 1 is the following unit cost function ratio:**

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

- **As long as $r < 0$, the above price index is well defined even if the price for a missing product is assigned the value $+\infty$. Thus once we have estimates for r and the α_n, we can calculate the above CES price index for our sample data.**
- **Feenstra assumed that the set of commodities that are available in period t is I(t) for $t = 1,\ldots,T$.**
- **Feenstra showed how the CES price index could be calculated with just a knowledge of observable prices and quantities for the two periods being compared plus an estimate for r (or equivalently for $\sigma = 1 - r$).**
- **In the following slide, we show how the Feenstra methodology works.**

The Feenstra CES Index Number Methodology (cont)

- **Feenstra's CES unit cost index number methodology works as follows. Start off with the definition of the CES price index:**

(12) $P_{CES}{}^{t} \equiv c(p^t)/c(p^1)$; $\quad t = 1,\ldots,T$

$$= [\sum_{i\in I(t)} \alpha_i (p_i{}^t)^r]^{1/r} / [\sum_{i\in I(1)} \alpha_i (p_i{}^1)^r]^{1/r}$$

$$= [\text{ Index } 1]\times[\text{Index } 2]\times[\text{Index } 3]$$

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

(13) **Index 1** $\equiv [\sum_{i\in I(t)\cap I(1)} \alpha_i (p_i{}^t)^r]^{1/r} / [\sum_{i\in I(1)\cap I(t)} \alpha_i (p_i{}^1)^r]^{1/r}$;

(14) **Index 2** $\equiv [\sum_{i\in I(t)} \alpha_i (p_i{}^t)^r]^{1/r} / [\sum_{i\in I(1)\cap I(t)} \alpha_i (p_i{}^t)^r]^{1/r}$;

(15) **Index 3** $\equiv [\sum_{i\in I(1)\cap I(t)} \alpha_i (p_i{}^1)^r]^{1/r} / [\sum_{i\in I(1)} \alpha_i (p_i{}^1)^r]^{1/r}$.

- **Index 1 is the CES maximum overlap index that compares the prices of period t with the corresponding prices of period 1 over the set of products that are available in both periods. Feenstra applies a known result and notes that Index 1 can be calculated using observable data and the Sato-Vartia index which is exact for cost minimizing consumers that have CES preferences.**

The Feenstra CES Index Number Methodology (cont)

- **Feenstra interprets Index 2 as an index which will adjust the maximum overlap index for any products that happen to be available in period t but not in period 1. We expect this index to be equal to or less than 1. Feenstra shows that it is equal to:**

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

- **Similarly, Feenstra interprets Index 3 as an index which will adjust the maximum overlap index for any products that happen to be available in period 1 but not in period t. We expect this index to be equal to or greater than 1. Feenstra shows that it is equal to:**

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

- **Note that we require an estimate for r (or sigma) in order to calculate Index 2 and Index 3.**
- **Recall that our estimate for r was –2.8041 and which leads to the estimate for the elasticity of substitution** $\sigma \equiv 1 - r = 3.8041$**.**
- **In the following slide, we see what Index 2 and Index 3 look like for our data using r = –2.8041.**

Index 2 and Index 3 using the CES Unit Cost Function

Table 2: Indexes Measuring the Effects of Changes in the Price Level due to the Availability of Products when σ = 3.8041

Month t	$Index_2^t$	$Index_3^t$
9	0.9836	1.0000
10	1.0000	1.0082
11	0.9902	1.0000
20	1.0000	1.0088
23	0.9930	1.0000

- In month 9, products 2 and 4 make their appearance and Table 2 tells us that the effect on the CES price level of this increase in variety is to lower the price level for month 9 by about 2%. In month 10 when product 12 disappears from the store, this disappearance has the effect of increasing the price level for frozen juice by 0.82 percentage points.
- The overall effect on the price level of the changes in the availability of products is equal to 0.9836×1.0082×0.9902×1.0088×0.9930 = 0.9836, a decrease in the price level over the sample period of about 1.64%.

Estimating the CES Utility Function

- **We now assume that the purchaser utility function f(q) is defined as the following *CES utility function*:**

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

- **where the parameters β_n are positive and sum to 1 and s is a parameter which satisfies the inequalities $0 < s \leq 1$. Thus f(q) is a mean of order s.**
- **In order for f(q) to be concave in q, we need $s \leq 1$.**
- **In order for f(q) to be well defined if some component of q becomes 0, we need $s \geq 0$.**
- **The unit cost function that is dual to f(q) defined by (28) turns out to be the following one:**

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

- **It can be seen that c(p) is proportional to a mean of order r where $r = s/(s-1)$.**
- **The elasticity of substitution that corresponds to the f(q) defined (28) is $\sigma = 1/(1-s)$. Thus if s satisfies $0 < s \leq 1$, then σ satisfies $1 < \sigma \leq \infty$.**

Estimating the CES Utility Function (cont)

- **If the data were exactly consistent with CES utility maximizing purchasers, then the results of estimating the parameters of the direct utility function would be exactly consistent with the results of estimating the parameters of the CES unit cost function.**
- **However, there are "errors" in the data so this equality of results will not hold in practice.**
- **Thus it is of interest to estimate the parameters of the direct CES utility function and see which model fits the data best.**
- **The share equations that are correspond to the CES f(q) defined by (28) are the following ones:**

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

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

Estimating the CES Utility Function (cont)

- **To start off our estimation procedure, we will estimate a preliminary *Model 3* where we assume s = 1. Thus we obtain the following system of estimating equations:**

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

- **where the error term vectors, ε^t, are assumed to be distributed as a multivariate normal random variable with mean vector 0_{19} and variance-covariance matrix Σ for t = 1,...,39.**
- **Note that the CES utility function collapses down to a linear utility function when s = 1.**
- **Thus all products are perfect substitutes in this model.**
- **Since the shares s_i^t sum to one for each period t, all 19 error terms ε_i^t for i = 1,...,19 cannot be distributed independently so we dropped the equation for product 19 from our list of estimating equations for Model 3.**

Estimating the CES Utility Function (cont)

- **The final log likelihood turned out to be 3074.316, which is a huge increase from the final log likelihoods for Models 1 and 2; recall that the dependent variables are the same in all 3 Models and thus the log likelihoods are comparable.**
- **The equation by equation R^2 values for Model 3 were as follows: 0.9676, 0.9809, 0.9666, 0.9779, 0.9581, 0.9494, 0.9724, 0.7750, 0.9648, 0.9762, 0.8291, 0.9168, 0.9846, 0.9292, 0.9653, 0.9559, 0.9065 and 0.9554.**
- **These R^2 are very much higher than the corresponding R^2 for Models 1 and 2.**
- **The estimated β_n^* coefficients were as follows: 1.098, 1.101, 1.241, 1.248, 1.247, 1.939, 1.277, 0.825, 0.513, 0.520, 0.784, 1.017, 1.028, 1.416, 0.439, 1.099, 1.699 and 1.094. Of course, $\beta_{19} \equiv 1$.**
- **These coefficients reflect the marginal utility and hence the quality of one unit of each product relative to product 19. Thus they could be regarded as *quality adjustment parameters* for the first 18 products.**

Estimating the CES Utility Function (cont)

- **We used the Model 3 β_n^* coefficients as starting values (along with s = 1) in *Model 4* which is the following nonlinear regression model:**

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

- **The final log likelihood turned out to be 3239.160, which is an increase of 164.844 from the final log likelihood for Model 3.**
- **The estimated s was $s^* = 0.85374$ (standard error = 0.0065) and so the estimated elasticity substitution is $\sigma = 1/(1-s^*) = 6.8371$.**
- **This is very much larger than the estimated σ for Model 2. This was a surprise (initially)!**
- **The equation by equation R^2 values for Model 4 were as follows: 0.9748, 0.9873, 0.9716, 0.9904, 0.9637, 0.9600, 0.9766, 0.7746, 0.9678, 0.9792, 0.8057, 0.9387, 0.9863, 0.9207, 0.9821, 0.9527, 0.8996 and 0.9583. The average R^2 was equal to 0.9439.**
- **Thus the CES directly estimated utility function model fits the data much better than the CES unit cost function model.**

A Check on the Consistency of our Software

- **We checked that our software was correct by taking our estimated coefficients for Model 4 and using them to define a CES direct utility function.**
- **We then calculated shadow prices period t imputed equilibrium prices for product n as $p_n^{t*} \equiv \beta_n(q_i^t)^{s-1}$ for $n \in I(t)$ and $t = 1,\ldots,39$. If $n \notin I(t)$, then $p_n^{t*} \equiv +\infty$.**
- **Using these imputed prices and actual quantities with $q_n^t \equiv 0$ for $n \notin I(t)$, we used the Model 2 software with this new data set and estimated the elasticity of substitution for this artificial data set.**
- **We obtained exactly the same estimate for σ as we obtained for Model 4 and the R^2 was 1 for all 18 estimating equations.**
- **In retrospect, the big difference between the CES unit cost and direct utility function Models is not surprising in a situation where there are errors in optimization: the CES unit cost function is very close to the Leontief (no substitution) Model while the CES utility function model is very close to the Linear Utility Function Model (with perfect substitutability).**

Index 2 and Index 3 using the CES Utility Function σ

- The point estimate for s from Model 4 is $s^* \equiv 0.87374$ and thus the corresponding $r = s^*/(s^*-1) = -6.92016$ and the corresponding $\sigma = 1/(1-s^*) = 6.8371$.
- Using $r^* = -6.92016$, we use the formula (55) and (56) to evaluate Index_2^t and Index_3^t and we obtain the following counterpart to Table 2 in section 5.

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

Month t	Index_2^t	Index_3^t
9	0.9933	1.0000
10	1.0000	1.0033
11	0.9960	1.0000
20	1.0000	1.0036
23	0.9971	1.0000

- The overall effect on the price level of the changes in the availability of products is equal to 0.9933×1.0033×0.9960×1.0036×0.9971 = 0.9933, a decrease in the price level over the sample period of about 0.67%.
- As σ increases, the net gains from increased product availability decline rapidly.
- Using σ = 3.8041, the gains were 1.64%, much bigger than 0.67%.

The Konüs-Byushgens-Fisher Utility Function

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

(66) $f(q) = (q^T A q)^{1/2}$

- **where the N by N matrix $A \equiv [a_{nk}]$ is symmetric (so that $A^T = A$) and thus has N(N+1)/2 unknown a_{nk} elements.**
- **We also assume that A has one positive eigenvalue with a corresponding strictly positive eigenvector and the remaining N−1 eigenvalues are negative or zero.**
- **Konüs and Byushgens (1926) showed that the Fisher (1922) quantity index $Q_F(p^0,p^1,q^0,q^1) \equiv [p^0\cdot q^1 p^1\cdot q^1/p^0\cdot q^0 p^1\cdot q^0]^{1/2}$ is exactly equal to the aggregate utility ratio $f(q^1)/f(q^0)$ provided that all purchasers maximized the utility function defined by (66) in periods 0 and 1 where p^0 and p^1 are the price vectors prevailing during periods 0 and 1 and aggregate purchases in periods 0 and 1 are equal to q^0 and q^1.**

The Konüs-Byushgens-Fisher Utility Function (cont)

- **The following inverse demand share equations can be used as the basis for a system of estimating equations for this functional form:**

(69) $s_n \equiv p_n q_n/p\cdot q = q_n \Sigma_{k=1}^{N} a_{nk} q_j/q^T A q$; $n = 1,\ldots,N.$

- **It turns out to be useful to reparameterize the A matrix in definition (66). Thus we set A equal to the following expression:**

(70) $A = bb^T + B;$

$b >> 0_N$; $B = B^T$; **B is negative semidefinite;** $Bq^* = 0_N$.

- **The vector** $b^T \equiv [b_1,\ldots,b_N]$ **is a row vector of positive constants and so** bb^T **is a rank one positive semidefinite N by N matrix.**
- **If B is a matrix of 0's, then** $f(q) = (q^T A q)^{1/2} = b^T q$**, a linear utility function. Thus a special case of the KBF functional form is the linear utility function which implies all products are perfect substitutes.**
- **We need to impose negative semidefiniteness on B.**

The Konüs-Byushgens-Fisher Utility Function (cont)

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

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

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

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

- **We add the c^n columns one at a time and stop when the increase in the log likelihood slows down (or stops).**

The Konüs-Byushgens-Fisher Utility Function (cont)

- **This is the same type of procedure that Diewert and Wales (1988) used in order to estimate normalized quadratic preferences and they termed the final functional form a *semiflexible functional form*.**
- **The above treatment of the KBF functional form also generates a semiflexible functional form.**
- **Instead of developing the above theory for the KBF utility function, we could develop the analogous theory for the dual KBF unit cost function, $c(p) \equiv (p^T A^* p)^{1/2}$ where $A^* = b^* b^{*T} - C^* C^{*T}$ where C^* is a lower triangular N by N matrix that satisfies $C^{*T} p^* = 0_N$ for the reference price vector p^*.**
- **The special case of this unit cost function where $C^* = O_{N\times N}$ leads to the Leontief (no substitution) unit cost function, $c(p) = b^{*T} p$.**
- **Since we already estimated the Leontief unit cost function (as Model 1) and it performed very poorly relative to the linear utility function (Model 3), we decided not to pursue the estimation of the KBF unit cost function.**

The KBF Utility Function: the Systems Approach

- **We started off by estimating the b vector and a rank 1 C matrix.**
- **The starting values for the b_n were the final estimates for the β_n from Model 3 above and the starting values for the c_n^1 were set at 0.01 for each n = 1,...,18.**
- **The initial log likelihood was 3074.663 and the final log likelihood was 3216.919, a gain of 142.603 for adding 18 new parameters to the linear utility model.**
- **The equation by equation R^2 values were as follows: 0.9661, 0.9787, 0.9623, 0.9889, 0.9608, 0.9521, 0.9628, 0.8002, 0.9657, 0.9752, 0.8337, 0.9224, 0.9867, 0.8936, 0.9673, 0.9555, 0.9064 and 0.9599. So far, so good; these R^2 values are relatively high.**
- **However, when we attempted to estimate the KBF utility function for a rank 2 substitution matrix, Shazam failed to converge. The problem is that the system of equations maximum likelihood estimation procedure involves the estimation of a symmetric variance covariance matrix which in our case contains (19)(18)/2 = 171 independent parameters.**

The KBF Utility Function: the One Big Equation Approach

- **In order to deal with the nonconvergence problem, we decided to stack up our 18 estimating equations into one big nonlinear regression model which involves estimating a single variance parameter instead of the 171 parameters required for the systems approach.**
- **We stopped adding columns to the lower triangular C matrix at 4 columns, which was our *Model 11*.**
- **For this model, we have $A = bb^T - c^1c^{1T} - c^2c^{2T} - c^3c^{3T} - c^4c^{4T}$ with $c^{4T} = [0,0,0,c_4{}^4,\ldots,c_{19}{}^4]$ and the additional normalization $c_{19}{}^4 = -\sum_{n=4}^{18} c_n{}^4$.**
- **The final log likelihood for this model was 2629.182, an increase of 14.656 for adding 15 new parameters to the Model 10 parameters.**
- **Thus the increase in log likelihood is now less than one per additional parameter so we decided to stop adding columns at this point.**
- **The single equation R^2 increased to 0.9922.**

KBF Utility Function: the One Big Equation Approach (cont)

- **However, this single equation R^2 is not comparable to the equation by equation R^2 that we obtained using the systems approach in the previous section.**
- **The comparable R^2 for each separate product share equation are as follows: 0.9859, 0.9930, 0.9773, 0.9853, 0.9814, 0.9543, 0.9755, 0.8581, 0.9760, 0.9694, 0.8923, 0.9278, 0.9908, 0.9202, 0.9874, 0.9566, 0.9111 and 0.9653.**
- **The average R^2 is 0.9560 which is a relatively high average when estimating share equations.**
- **Once we have an estimated A matrix, it is straightforward to form the reservation prices for the 20 observations where we have missing products.**
- **We explain the relevant algebra on the next slide.**

KBF Utility Function: the One Big Equation Approach (cont)

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

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

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

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

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

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

- **The predicted prices for products 2 and 4 for the first 8 months in our sample period were 1.62, 1.56, 1.60, 1.52, 1.61, 1.52, 1.70. 1.97 and 1.85, 1.46, 1.80, 1.37, 1.77, 1.83, 1.88, 2.27 respectively.**
- **The predicted prices for product 12 for months 10 and 20-22 were 1.37, 1.20, 1.22 and 1.28. These prices are not infinite!**

KBF Utility Function: the One Big Equation Approach (conc)

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

KBF Utility Function: the One Big Equation Approach II

- **Our interest is not in predicting shares; our interest is in finding predicted prices for the observations when quantities are equal to 0.**
- **Thus we turned to another econometric specification of the KBF utility function model where prices replaced shares as the dependent variables in the One Big Regression Approach.**
- **Thus the new estimating equations become:**

$$(79)\ p_i^t \equiv e^t\, \Sigma_{k=1}^{19}\, a_{ik}^{*} q_k^t / [\Sigma_{n=1}^{19} \Sigma_{m=1}^{19}\, a_{nm}^{*} q_n^t q_m^t] + \varepsilon_i^t\ ;$$

$$t = 1,\ldots,39;\ i = 1,\ldots,18.$$

- **However, the observations that correspond to missing products are dropped from the stacked estimating equations defined by (79). This is an advantage of the one Big Regression Approach.**
- **As before, $A = bb^T - CC^T$ where C is a lower triangular matrix.**
- **With the new dependent variables, we were able to estimate a rank 6 substitution matrix.**

KBF Utility Function: the One Big Equation Approach II (cont)

- **The final log likelihood for the rank 6 substitution matrix *Model 14* was 568.877, an increase of 18.531 over the rank 5 substitution Model 13. The single equation R^2 was 0.9527.**
- **Model 14 had 111 unknown parameters that were estimated (plus a variance parameter). We had only 680 observations and so we decided to call a halt to our estimation procedure.**
- **The equation by equation R^2 that compares the predicted prices for the 19 products with the actual prices were as follows: 0.8274, 0.8678, 0.9001, 0.9174, 0.8955, 0.8536, 0.9047, 0.0344, 0.3281, 0.4242, 0.0516, 0.2842, 0.8650, 0.7280, 0.4872, 0.8135, 0.8542, 0.8479 and 0.3210.**
- **The average R^2 for Model 14 was 0.6424. For Model 11, it was 0.5681 so by switching from shares as the dependent variables to prices as the dependent variables, we have improved the accuracy of our estimated predicted prices.**

KBF Utility Function: the One Big Equation Approach II (cont)

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

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

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

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

- **The month t *KBF price index*, P_{KBF}^{t}, is defined as the month t KBF price level divided by the month 1 KBF price level; i.e.,**
- **$P_{KBF}^{t} \equiv P_{KBF}^{t*}/P_{KBF}^{1*}$ for t = 1,...,39.**
- **These econometrically based KBF price indexes can be compared to our econometrically based CES price indexes P_{UCES}^{t} that are defined in a similar manner using the results of Model 4, which estimated a direct CES utility function.**
- **However, before we make this comparison, we estimate one more model.**

CES Utility Function; One Big Equation; Prices as Dependent Variables (instead of shares)

- **This leads to the following system of estimating equations:**

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

- **Now stack the above 702 equations into a single estimating equation and drop the 20 observations where $q_i^t = 0$. Call this *Model 15*.**
- **The final log likelihood was equal to 483.834, which is below the final LL from the KBF Model 14 which was 568.877.**
- **The single equation R^2 was 0.9393, which is below the single equation R^2 from Model 14, which was 0.9527.**
- **The estimated parameter s was $s^* = 0.85365$. This is virtually identical to our estimate for s from Model 4 (which used the systems approach to CES utility function estimation with shares as dependent variables) which was 0.85374.**
- **Since the estimated s for Model 15 is the same as it was for Model 4, the Feenstra gains and losses from changes in product availability will not change.**

CES Utility Function; One Big Equation; Prices as Dependent Variables (cont)

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

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

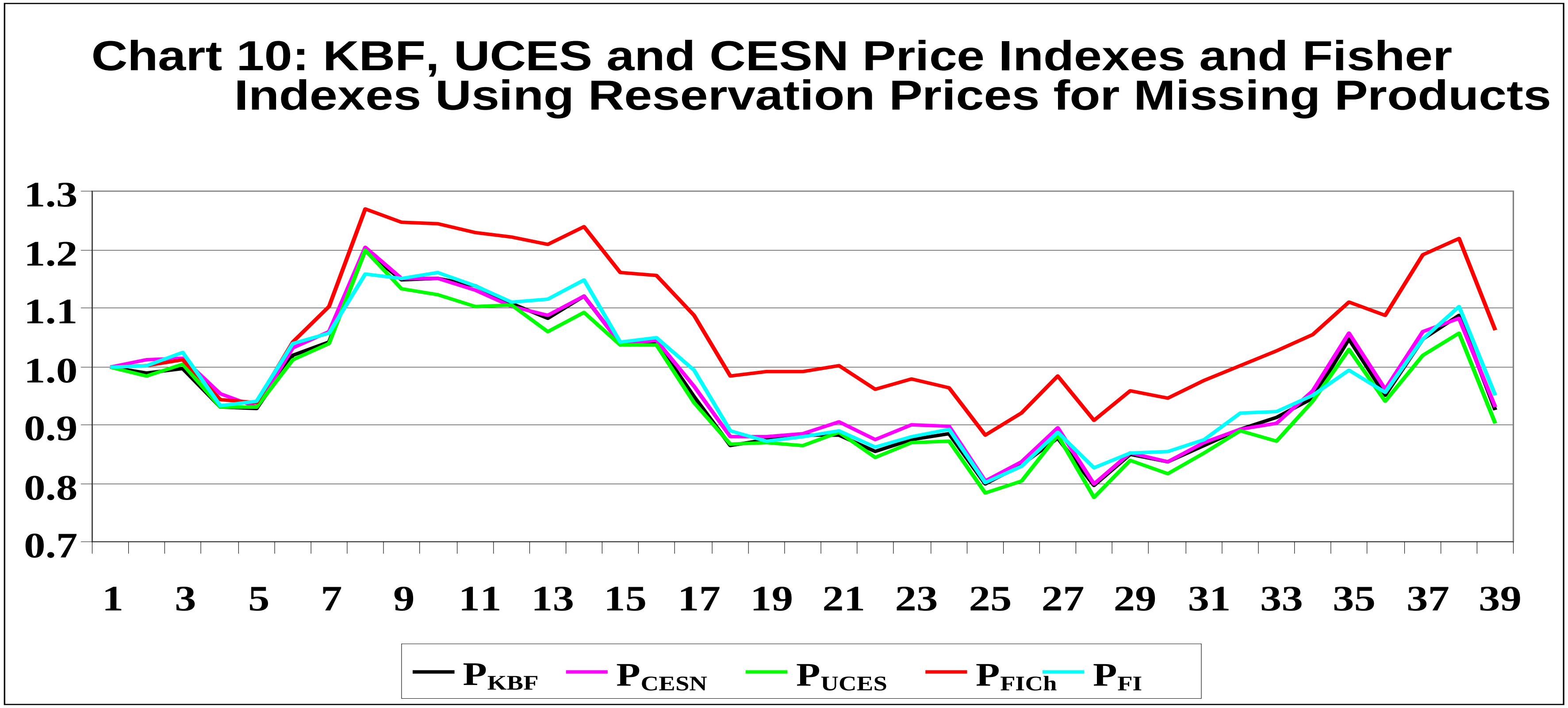
- **The corresponding *New CES (unnormalized) implicit price level*, $P_{CESN}{}^{t*}$, is defined as period t sales of the 19 products, e^t, divided by the period t aggregate quantity level, $Q_{CESN}{}^{t}$:**

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

- **The month t *New CES price index*, $P_{CESN}{}^{t}$, is defined as the month t CESN price level divided by the month 1 CESN price level; i.e.,**
- $P_{CESN}{}^{t} \equiv P_{CESN}{}^{t*}/P_{CESN}{}^{1*}$ **for** $t = 1,\ldots,39.$
- **The CESN price index will be compared to its econometric counterpart indexes $P_{KBF}{}^{t}$ (Model 14) and $P_{UESN}{}^{t}$ (Model 4).**

Comparison of 3 Econometric Based Price Indexes with Chained and Fixed Base Fisher Indexes that use Reservation Prices

P_{KBF}^t and P_{CESN}^t are very close to each other. The Model 4 CES price index, P_{UESN}^t, is pretty close to the Model 14 and 15 indexes. The fixed base Fisher index that uses reservation prices for the missing products is fairly close as well. The chained Fisher index that uses reservation prices has upward chain drift.



KBF Gains from Increased Product Availability: Approach 1

- **Having estimated reservation prices for the missing products, we can calculate a comprehensive Fisher chain link index going from period t–1 to period t, which is $P_{FICh}^{t}/P_{FICh}^{t-1}$.**
- **We can calculate the corresponding chain link for the maximum overlap Fisher index for the products that are present in both periods, which is $P_{FMCh}^{t}/ P_{FMCh}^{t-1}$.**
- **The ratio of these two indexes is defined as follows:**

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

- **This index can be interpreted as a "correction" index which when multiplied by the readily calculated maximum overlap index $P_{FMCh}^{t}/P_{FMCh}^{t-1}$ gives us the "true" chain link index $P_{FICh}^{t}/P_{FICh}^{t-1}$, or it can be interpreted as the amount of bias in the maximum overlap chain link index due to changes in the availability of products.**
- **We list this index below in Table 8 for the periods when there is a change in product availability (months 9, 10, 11, 20 and 23).**

KBF Gains from Increased Product Availability: Approach 2

- **The index I_{KBF}^{t} for t = 23 was greater than 1 rather than less than 1. This anomalous result is due to the fact that our estimated KBF utility function does not fit the data exactly.**
- **Thus we consider an alternative methodology that will (perhaps) eliminate counterintuitive results.**
- **We replace all observed prices by their predicted prices (and use predicted prices for the missing product prices).**
- **The comprehensive Fisher chain link index going from period t–1 to period t using predicted prices is $P_{FPCh}^{t}/P_{FPCh}^{t-1}$.**
- **Holding product availability constant, we can calculate the corresponding chain link for the maximum overlap Fisher index using predicted prices for the products that are present in both periods, which is $P_{FPMCh}^{t}/P_{FPMCh}^{t-1}$. Their ratio is:**

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

- **This index is also listed below in Table 8.**

KBF Gains from Changes in Product Availability: Approaches 1 and 2

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

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

- I_{KBF}^{t*} has all "signs" correct but the product of the availability factors is still greater than 1 when it should be less than 1 due to the net increase in product availability over the sample period.
- Thus we consider yet another methodological approach to measuring the gains and losses due to changes in product availability. This approach also makes use of the estimated f.

KBF Gains from Changes in Product Availability: Approach 3

- Consider a period t where all products are being purchased but in the prior period, one or more products were not available.
- Give purchasers in period t their total observed expenditure e^t but now ask them to maximize the estimated utility function but restricting them to not purchase the commodities that were not available in the previous period. Call this utility level $u^{t*} \leq u^t$ where $u^t = f(q^t)$, the estimated period t utility.
- Take the ratio $u^t/u^{t*} \equiv G^t \geq 1$, which is a measure of the period t gain in real income due to increased product availability.
- Now suppose that following period t, some products become unavailable. Again give purchasers in period t their total observed expenditure e^t but now ask them to maximize the estimated utility function but restricting them to not purchase the commodities that will not be available in the next period. Call this utility level $u^{t**} \leq u^t$.
- Take the ratio $u^{t**}/u^t \equiv L^{t+1} \leq 1$, which is a measure of potential period t+1 loss in real income due to decreased product availability.
- In the paper, we developed approximations to the above gain and loss measures (which are be exact for the CES utility function).

KBF and CES Gains from Changes in Product Availability: Approach 3

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

	KBF	CES	
$G_{A2,4}^{9}$	1.00127	1.00746	(Anomalous results have been eliminated!)
L_{A12}^{10}	0.99748	0.99512	
G_{A12}^{11}	1.00304	1.00529	
L_{A12}^{20}	0.99881	0.99644	
G_{A12}^{23}	1.00078	1.00296	
Product	1.00138	1.00724	

- **Since there is a net gain in product availability over the sample period, both estimated utility functions register a net gain.**
- **But the net gain from the KBF utility function is only about 1/5 of the gain that accrued to the CES utility function using Approach 3. The CES approach consistently overestimates!**
- **Our new CES gain in utility of 1.00724 ≈ 1/1.0067, the Model 4 estimated increase in utility.**

The Hausman Approximate Consumer Surplus Methodology

- **Hausman (1999; 191) (2003; 27) presented a very simple and easy to implement methodology for calculating the approximate loss of consumer surplus due to the disappearance of a product.**
- **The framework is a partial equilibrium one where he drew an inverse demand curve for say product 1 as $p_1 = D_1(q_1)$ where q_1 is the quantity of product 1 purchased when its price is p_1.**
- **Hausman formed a first order Taylor series approximation to this demand curve around the point (p_1^*, q_1^*) which corresponds to a period when product 1 was available.**
- **He assumed that the demand curve is downward sloping and when $q_1 = 0$, the corresponding virtual demand price is p_1^{**}.**
- **The linear approximation to the actual inverse demand function goes through the p_1 axis at the point ρ_1^* where $\rho_1^* \equiv p_1^* + \alpha q_1^*$ and $\alpha \equiv -\partial D_1(q_1^*)/\partial q_1 > 0$ is the absolute value of the slope of the inverse demand curve evaluated at $q_1 = q_1^*$.**

The Hausman Approximate Consumer Surplus Methodology (cont)

- **We scale the utility level $f(q_1^*, q_2^*)$ so that it equals expenditure e^* for the period. Thus we have:**

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

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

(110) $L_H \equiv -(1/2)(\rho_1^* - p_1^*)q_1^*/e^*$

$$= -(1/2)\alpha(q_1^*)^2/e^*$$

$$= (1/2)s_1^*\eta$$

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

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

- **The practical problem are what exactly are we holding constant when we calculate the elasticity defined by (111)?**

The Hausman Approximate Consumer Surplus Methodology (cont)

- **In section 11 of the paper, we work out a second order approximation to the gain in consumer surplus due to the new availability of a commodity using our Approach 3 methodology for the case of only two products.**
- **Remarkably, we show that our second order approximation is exactly equal to Hausman's first order approach provided that we place the "right" interpretation on Hausman's elasticity η.**
- **It is not known if this equality extends to the N commodity case.**
- **In addition to the above approximation approach, Hausman develops a rigorous approach to the estimation of the gains from increased product variety that is based on the estimation of an expenditure function (rather than a utility function as in our KBF approach).**
- **On page 57 of the paper, we explain what the problem is with the rigorous Hausman approach: whenever there is a missing product, his approach requires the econometrician to estimate the corresponding virtual price as an extra parameter and the resulting equations can become very nonlinear and messy.**

Conclusion: The Important Points to Take Away!

- When dealing with scanner data where there are periodic sales of products, chain drift is a huge problem.
- Multilateral index number theory can be used to deal with the chain drift problem; see the ABS (2016) and Diewert and Fox (2017)
- It is not a trivial matter to estimate the elasticity of substitution in the CES context. Estimation of the CES unit cost function may give very different results from estimation of the CES direct utility function.
- The CES methodology developed by Feenstra for measuring the gains from increased product availability appears to overestimate the gains by a substantial amount.
- The KBF utility function can be estimated and it can be used to calculate “reasonable” reservation prices but it is probably too labour intensive (and subject to many econometric uncertainties) to be adopted by statistical agencies as a practical approach to the estimation of reservation prices.