Estimating the Benefits of New and Disappearing Products

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

and Robert Feenstra

University of California at Davis and NBER

Presented at the CRIW Pre-Conference:

Big Data for 21st Century Economic Statistics,

Cambridge, MA,

July 18, 2018

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 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 look at in section 12. We will also explain the problem with his first 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 implemented 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.

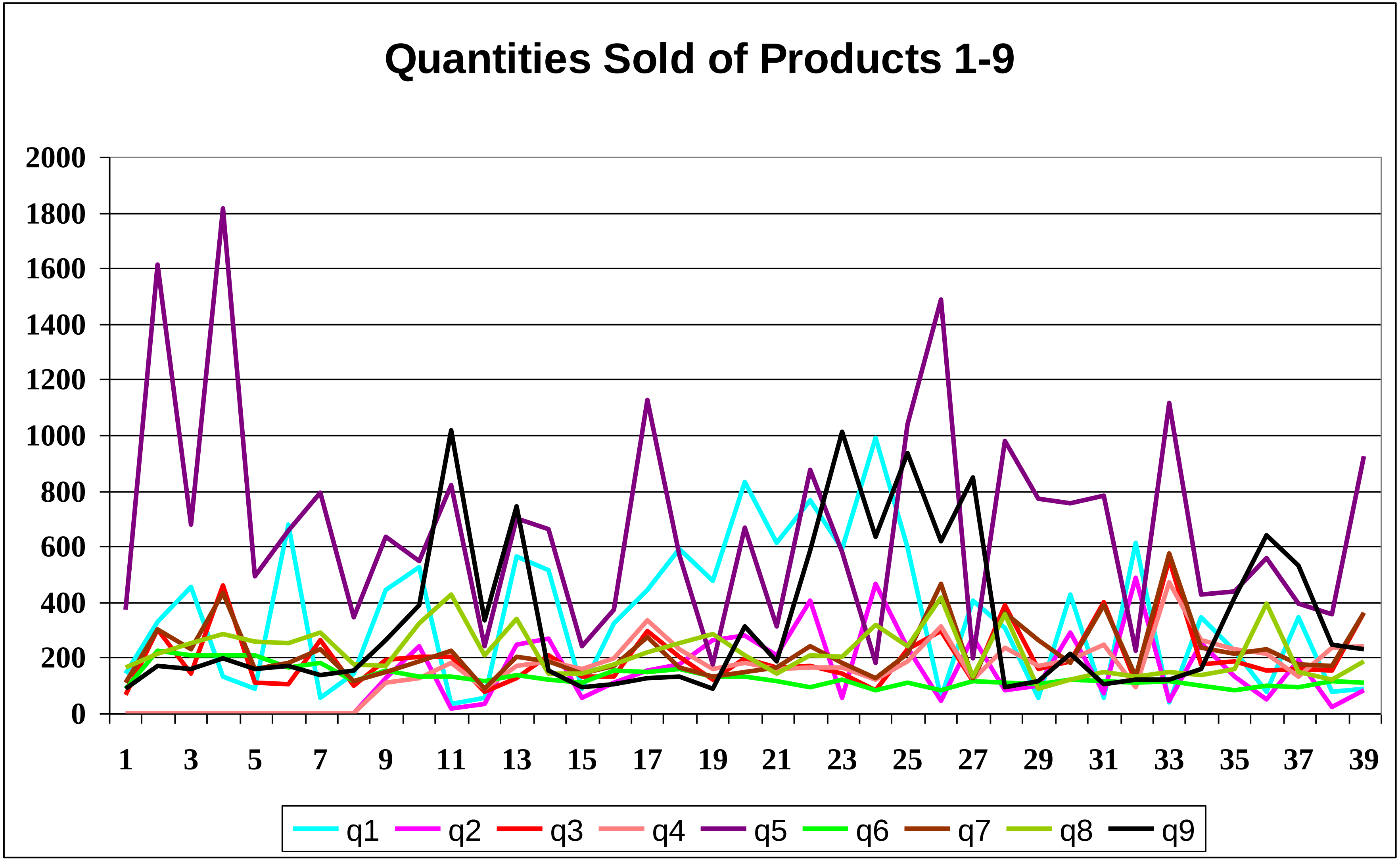
The Data

- **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)

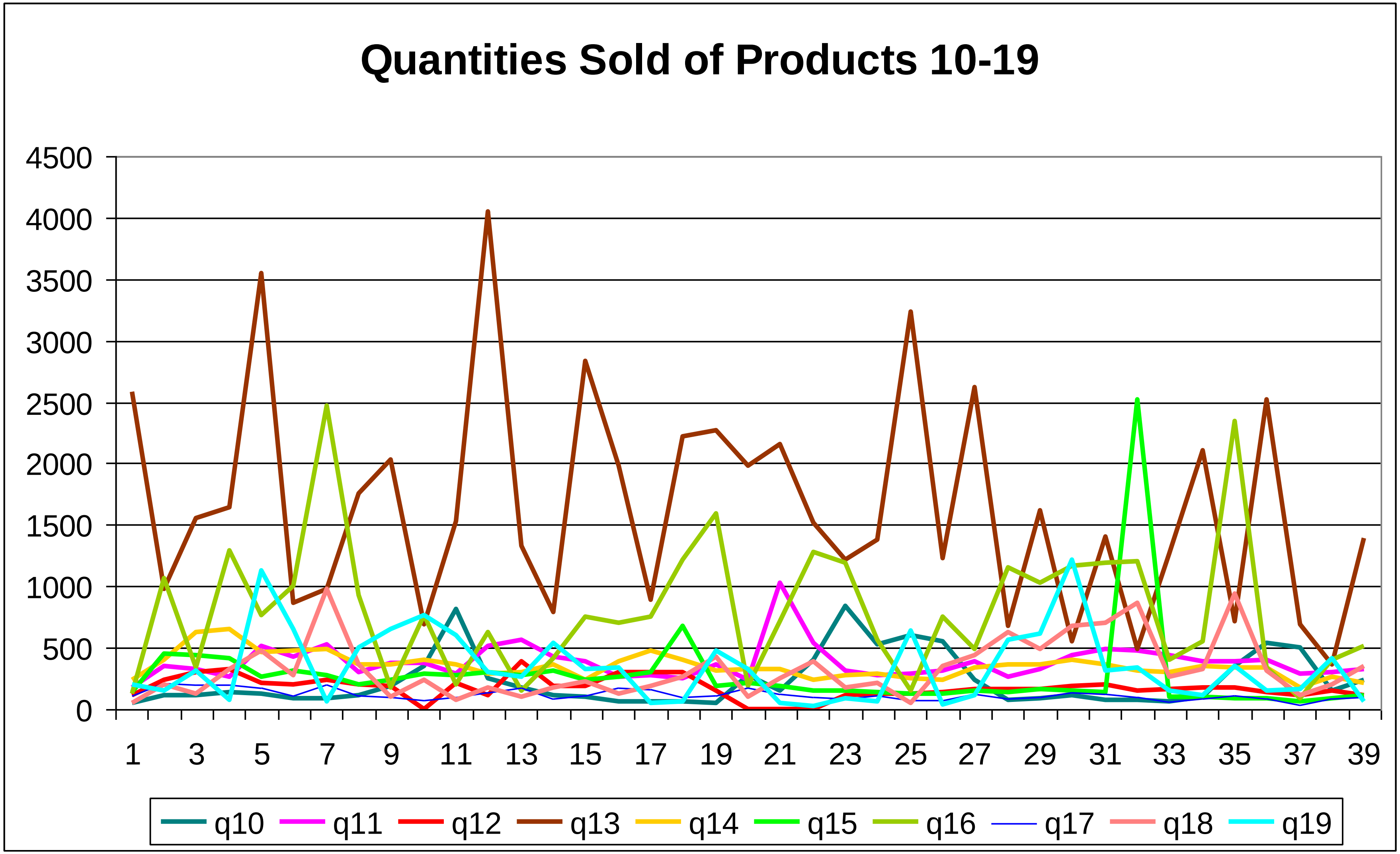
- **There were no sales of Products 2 and 4 for “months” 1-8 and there were no sales of Product 12 in “month” 10 and in “months” 20-22.**
- **Thus there is a new and disappearing product problem for 20 observations (out of 741 total observations on all 19 products and all 39 “months” in this data set.**
- **Later in the paper, we 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



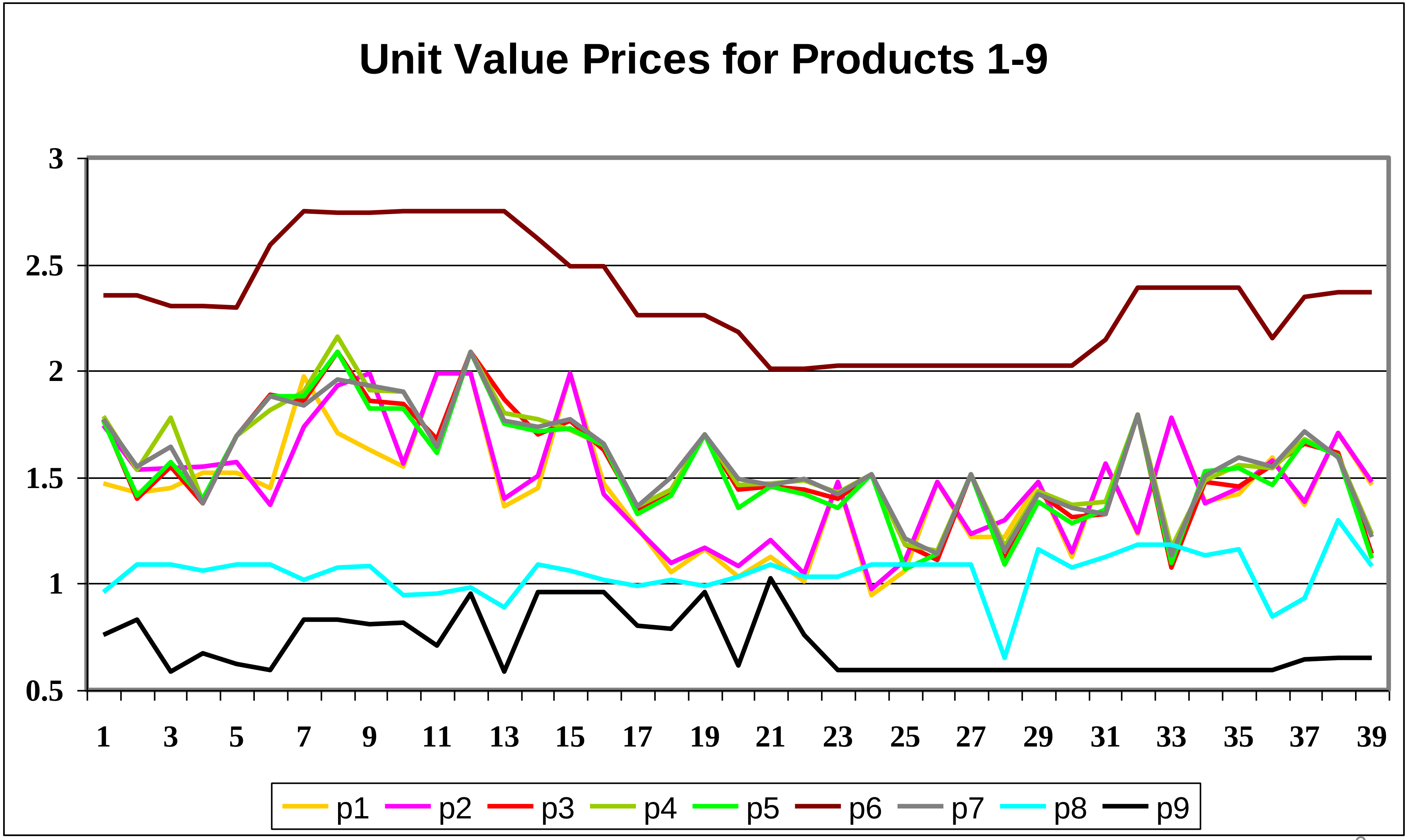
More Quantity Data

There is tremendous variation in the monthly quantities sold.



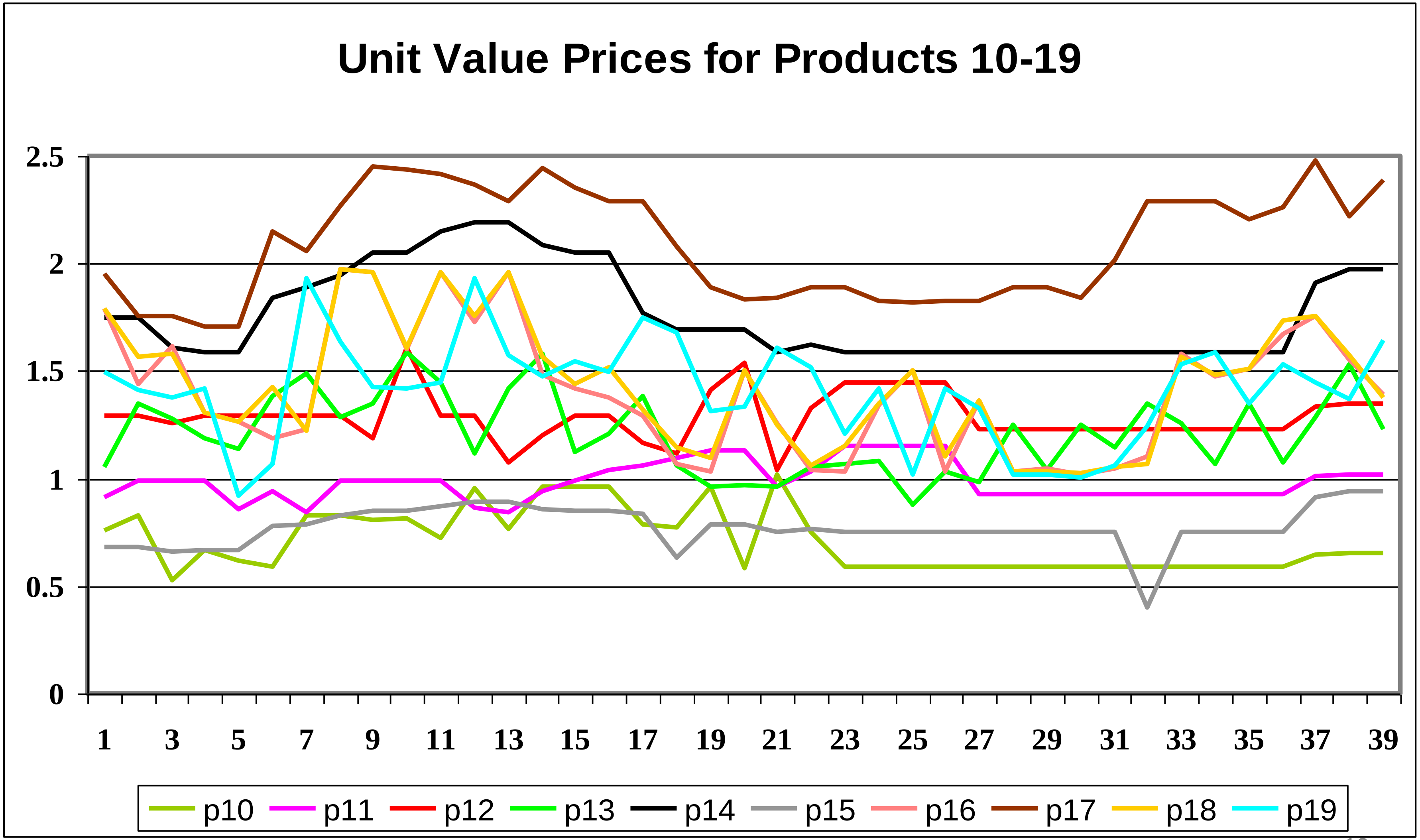
Price Data

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.



Quick Summary of the Paper (1)

- In order to implement the Feenstra methodology, we need an estimate of the elasticity of substitution σ
- We can estimate σ by estimating the CES cost function using expenditure shares as dependent variables and prices as independent variables; we can use a systems approach or stack the equations into one big equation (advantage of the latter is that we do not have to estimate 18x18/2 variance-covariance parameters). Either way, we get $\sigma = 3.8$ and a big gain from the introduction of new products using Feenstra's methodology.
- We can estimate σ by estimating the CES utility function using expenditure shares as dependent variables and quantities as independent variables; i.e., see equations (33) on the previous slide. The resulting estimate for $\sigma = 6.8$ (average R^2 was equal to 0.9439) and the gain from the introduction of new products using Feenstra's methodology is now equal to .67% (1.64%).
- Which specification should we use? Using quantities as independent variables leads to much higher fits. Why is this?

Quick Summary of the Paper (2)

- **Which CES specification should we use?**
- **A special case of the CES unit cost function occurs when r = 1. The resulting CES preferences are then Leontief no substitution preferences. If we fit the CES unit cost function model with r = 1 and then fit the CES unit cost function with a general r, we find that the fit of the estimating equations does not improve all that much (and the fits are not good; average R^2 was only 0.3767).**
- **A special case of the CES utility function occurs when s = 1. The resulting CES utility function is then a linear function where all elasticities of substitution are equal to + ∞. If we fit the CES utility function with s = 1 and then fit the CES utility function with a general s, again we find that the fits do not improve all that much but the average R^2 was 0.9439.**
- **If the products are highly substitutable, it is better to start with a linear utility function and tweak it versus starting with a no substitution utility function and tweak it.**

Quick Summary of the Paper (3)

- **How can we determine whether the Feenstra methodology overestimates the gains from product variety?**
- **By estimating a flexible functional form for the utility function! Once we have the estimated utility function in hand (and assuming that the function is well behaved when the level of any product consumed is equal to 0), then we can simply differentiate the utility function at the observed period t quantity vector with respect to any zero components of the observed quantity vector and get the Hicksian reservation price, which can then be used in an index number formula.**
- **With the estimated utility function in hand, we can also compare solutions to utility maximization problems where all products are available versus solutions when some products are not available and look at the difference in welfare.**
- **Bottom line: for our data set, the Feenstra methodology greatly overestimated the gains from increased product availability.**

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.**

Additional Slides if Time Permits

- **What is the flexible functional form that we actually estimated?**
- **Hausman's cost function methodology explained (and dismissed as impractical).**
- **The world's most parsimonious system of estimating equations for estimating the elasticity of substitution (section 5 of the paper)**

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

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

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

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

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

- **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 $f(q) = (q^TAq)^{1/2} = b^Tq$, 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 $k = 1,\ldots,K$. If C is lower triangular, then the first $k-1$ elements of c^k are equal to 0 for $k = 2,3,\ldots,N$.
- 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) $c^n \cdot q^* = c^{nT} q^* = 0$; $\qquad n = 1,\ldots,N.$

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

KBF and CES Gains from Changes in Product Availability:

Table 6: 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	1.00728
$L_{A12}{}^{10}$	0.99748	0.99643
$G_{A12}{}^{11}$	1.00304	1.00433
$L_{A12}{}^{20}$	0.99881	0.99615
$G_{A12}{}^{23}$	1.00078	1.00311
Product	1.00138	1.00728

(Anomalous results have been eliminated!)

- **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!**

The Hausman Approximate Consumer Surplus Methodology (cont)

- **In section 12 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 section 11 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).**
- **At the end of section 11 of the paper, we explain what the problem is with the rigorous cost function 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.**

A New Method for Estimating the Elasticity of Substitution

- **Our goal is to estimate the elasticity of substitution for a CES direct utility function f(q) that was discussed in Section5 in the main text. We now drop products that are not present in all T periods. So N is now 16. The CES utility function is:**

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

- **The purchasers' system of inverse demand equations is**

(43) $s_n^t \equiv p_n^t q_n^t/\Sigma_{i=1}^{N} p_i^t q_i^t = \beta_n(q_n^t)^s/\Sigma_{i=1}^{N} \beta_i(q_i^t)^s$;

$$t = 1,\ldots,T;\ n = 1,\ldots,N.$$

- **We take natural logarithms of both sides of the equations in (B2) and add error terms e_n^t in order to obtain the following *fundamental set of estimating equations*:**

(44) $\ln s_i^t = \ln\beta_i + s\ln q_i^t + \ln[\Sigma_{n=1}^{N} \beta_n \ln(q_n^t)^s] + e_{si}^t$.

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

The Feenstra Double Differencing Method for Estimating σ (see section 5)

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

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

(46) $ds_n^t \equiv \Delta s_n^t - \Delta s_N^t$; $n = 1,\ldots,N-1;\ t = 2,3,\ldots,T$

$$= \ln(s_n^t) - \ln(s_n^{t-1}) - \ln(s_N^t) - \ln(s_N^{t-1}).$$

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

(47) $dq_n^t \equiv \Delta q_n^t - \Delta q_N^t$; $n = 1,\ldots,N-1;\ t = 2,3,\ldots,T$

$$= \ln(q_n^t) - \ln(q_n^{t-1}) - \ln(q_N^t) - \ln(q_N^{t-1}).$$

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

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

The Feenstra Double Differencing Method for Estimating σ

- **The double differenced log shares ds_n^t satisfy the following system of (N – 1)(T – 1) estimating equations under our assumptions:**

(49) $ds_n^t = s\, dq_n^t + \varepsilon_n^t$; $\quad n = 1,\ldots,N-1;\ t = 2,3,\ldots,T$

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

The Feenstra Double Differencing Method for Estimating σ

- **Using the data listed in Appendix A, we have 15 product estimating equations of the form (B8) which we estimated using the NL system command in Shazam. Thus our N = 16 and our T = 39.**
- **The resulting estimate for s was 0.86491 (with a standard error of 0.0067) and thus the corresponding estimated σ is equal to 1/(1–s) = 7.4025, which is in line with our earlier estimates for σ when we estimated the CES utility function using Models 4 and 15. The standard error on s was tiny!**
- **The equation by equation R^2 were as follows: 0.9936, 0.9895, 0.9905, 0.9913, 0.9869, 0.9818, 0.9624, 0.9561, 0.9858, 0.9911, 0.9934, 0.994, 0.9906, 0.9921 and 0.9893.**
- **The average R^2 was 0.9859 which is very high for share equations or for transformations of share equations.**