CONSUMER AND MARKET DEMAND

AGRICULTURAL POLICY RESEARCH NETWORK

The Scope of the Unit Value Problem

Timothy K.M. Beatty
Food and Resource Economics Group
University of British Columbia

Research Project Number CMD-07-06

PROJECT REPORT

December 2007

Department of Rural Economy
Faculty of Agriculture & Forestry,
and Home Economics
University of Alberta
Edmonton, Canada

The Scope of the Unit Value Problem

Timothy K.M. Beatty
Food and Resource Economics Group
University of British Columbia

Acknowledgements: Agriculture and Agri-Food Canada Consumer and Market Demand Agriculture Policy Research Network. This research was also supported by the Canada Research Chair program. Thanks to Ian Crawford, Tirtha Dhar and Krishna Pendakur for helpful discussion.

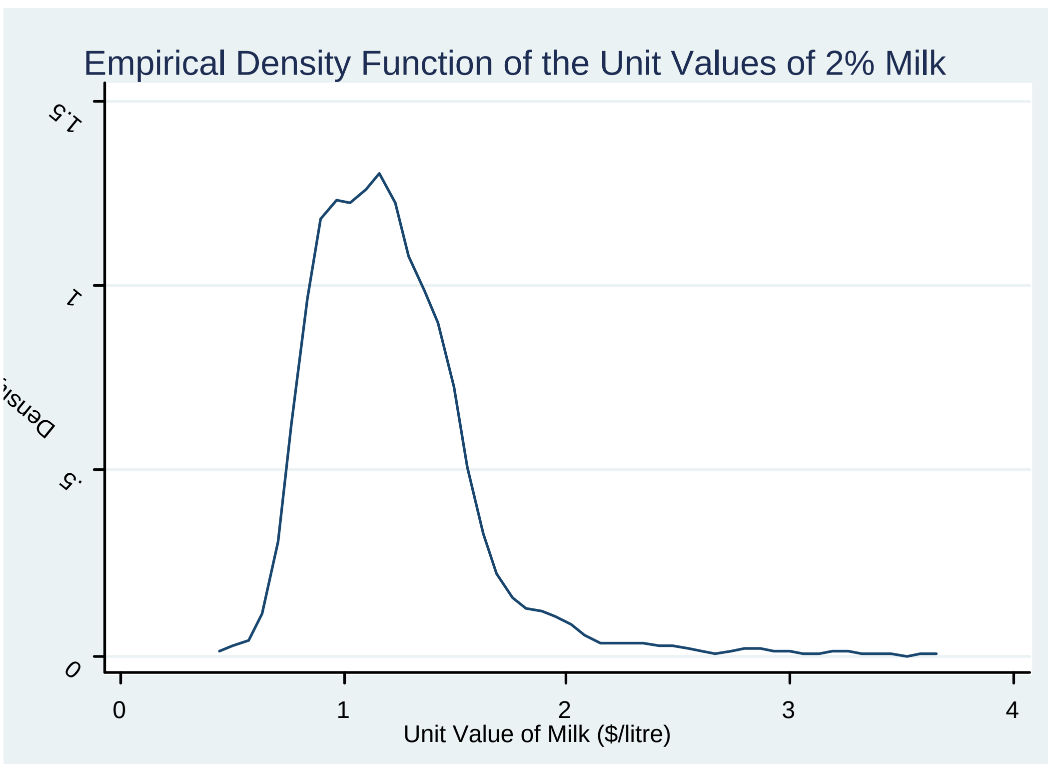
The Scope of the Unit Value Problem

This paper investigates the scope of the unit value problem in household level expenditure data. Specifically, to what extent do unit values and prices coincide. I test whether self-reported prices vary with covariates thought to be important in household demand behavior. In addition, I measure the expenditure elasticity of quality for 196 distinct food aggregates. Approximately half of the expenditure elasticities estimated are significantly different from zero, suggesting that unit values and prices do not coincide in most cases. Finally, a generalization of the standard quality demand regression is proposed, in which the expenditure elasticity of quality is allowed to vary over the range of total food expenditure. Some food aggregates display significant nonlinearities, providing further evidence of the difference between prices and unit values.

Introduction

In many expenditure surveys, households report both expenditure and quantity information on a wide range of commodities. Dividing expenditure by quantity yields a unit value. In practice, these unit values vary more than one would expect, given that they are, at least from the point of view of an outside observer, self-reported prices for the same good. This is the case even when the good of interest is by all measures homogeneous. For example, in the 2001 Statistics Canada Family Food Expenditure Survey, the 5th and 95th quantiles of the distribution of unit values of 2% milk were \$0.98 and \$2.95 respectively. Figure 1 illustrates the empirical density function of unit values of 2% milk. This variability calls into question whether unit values can be treated as exogenous prices in a typical demand analysis.

Figure 1. Price Dispersion of 2% Milk



One way to think about this issue is that consumers choose amongst goods of varying qualities, all of which are aggregated into a single expenditure category or aggregate. The resulting unit values, obtained by dividing expenditure on all goods in a category by physical quantity in that category, may be endogenous. Consider the food aggregate “beef”, high income consumers may choose better cuts of meat than low income consumers. These are different commodities, but from the researcher’s point of view, both types of consumers are observed purchasing beef for different prices. More importantly, each household is effectively choosing the price it faces. In this way, unit values are endogenous. Simply treating a unit value as if it were an exogenous price in a demand system may yield biased and inconsistent estimates.

This article makes four main contributions. Virtually all of the previous work on the topic has focused on a small number of relatively aggregated commodities. In contrast, I investigate the scope of the unit problem by considering a larger and more comprehensive set of food categories or aggregates. Second, I examine whether aggregating over commodities exacerbates or ameliorates the problem. Third, I extend the analysis of the unit value problem to a semiparametric setting to test whether quality elasticities are constant over the relevant range. Finally, in contrast to0 much of the work on the topic of unit values, the present study deals with unit values in a North American context.

The notion that unit values and prices might differ in important ways has a long history in economics, dating back over half a century to pioneering work by Houthakker (1952) , Theil (1952) and Prais and Houthakker (1955). They related the unit values of a wide variety of food items to household characteristics, notably expenditure. Seminal work by Deaton (1988, 1997) recasts and greatly extends the earlier work. Deaton’s framework provides the analytical apparatus through which the problem is currently viewed. This approach was recently generalized in an important paper by Crawford Laisney and Preston (2003). With the exception of the pioneering volume by Prais and Houthakker (1955), most of this literature has concerned itself with a small number of highly aggregated commodities in a developing county context. In contrast, the current work considers households in North America.

These issues were brought to the forefront in the agricultural economics literature by Cox and Wohlgenant (1986) and Nelson (1991) . More recently Dong, Shonkwiler and Capps (1998) and Dong and Kaiser (2005) have looked at the unit value issue with regards to Beef and Cheese respectively. The latter paper pays particular attention to the question of nonconsumption and the resulting effect on demand estimation. This paper differs from previous work in that "quality" and the expenditure elasticity of quality are the main focus rather than estimating complete demand relationships. In contrast to previous work, I pay particular attention to the potential endogeneity of expenditure (LaFrance (1991)) Finally, rather than focusing on a single food category, I consider an exhaustive set of food aggregates.

The paper proceeds as follows. First, I review the theory of unit values. I then describe the data that will be used to explore the scope of the unit value problem. The analysis begins with a presentation of a simple graphical illustration of the nature of the problem. This is followed by an approach, which dates back to Prais and Houthakker (1955) , to estimate the expenditure elasticity of quality for 196 different commodity aggregates. Given that the commodities are far more detailed than would typically be encountered in most demand system work, I aggregate from detailed to broad food aggregates and reestimate the quality demand equations to examine the extent to which aggregation is important. I then generalize the model to allow for a nonlinear expenditure elasticity of quality. The final section concludes.

Theory of Unit Values

I begin by reviewing Deaton's framework for thinking about unit values. Necessarily, any discussion of the issue of unit values draws heavily on the seminal work by Deaton (1988, 1997), the present paper is no exception.

The unit of analysis is a household, indexed by *i*, in a geographic cluster, indexed by *c*. The vector of prices for the goods that make up a given food aggregate *G* is denoted p_G. Following previous work, I assume that relative prices are fixed within each food aggregate, such that I can write $p_G = \pi_G \tilde{p}_G$, where π_G is a scalar linear homogeneous price level for food aggregate *G* in cluster *C* and $\tilde{p}_G$ represents the within group/cluster relative structure of prices.

Denote the total quantity of food aggregate *G* purchased Q_G and for ease of notation assume that the goods that make up food aggregate *G* are in commensurate units (e.g. physical quantities). As a result, I can write $Q_G = \sum_{g \in G} q_g$. In other words, the total quantity of cheese consumed is the sum of the quantity of cheddar cheese, grated cheese, processed cheese etc… Total expenditure on food aggregate *G*, for a household in cluster *c*, can be written $x_G = \sum_{g \in G} p_g q_g$.

Finally the unit value for food aggregate *G*, can be written $V_G = x_G / Q_G$. Simple substitutions yields

$$V_G = \pi_G \left(\frac{\sum_{g \in G} \tilde{p}_g q_g}{\sum_{g \in G} q_g} \right) \tag{1}$$

Equation (1) shows that the unit value can be thought of as the product of the general price level for food aggregate π_G and an expensiveness index. This expensiveness index is a measure of, and is henceforth referred to as, quality and denoted υ. Here quality measures an average price per unit of measurement. For example, if the category were beef, quality would measure an average price per kilogram above the overall price level in the geographic cluster. Finally taking logarithms of both sides yields

$$\ln V_G = \ln \pi_G + \ln \upsilon_G . \tag{2}$$

Data

To investigate the scope of the unit value problem, I use the 2001 Family Food Expenditure Survey (FOODEX) conducted by Statistics Canada. The FOODEX is a national diary survey. In addition to a set of demographic variables, households record expenditure and quantity information on 196 distinct food categories over a two-week period. The survey was conducted in five different regions (British Columbia, Prairies, Ontario, Quebec and Atlantic) over the course of 2001. An important feature of the 2001 survey is that to enhance the accuracy of reporting and diminish measurement error, respondents were asked to provide receipts for all purchases. The resulting quantity and expenditure information, in contrast to a pure recall survey, is considered highly accurate.

As with all expenditure surveys, a number of households were excluded, due to concerns about data validity. In particular, I exclude households that did not complete the entire two-week period, households who purchased only a single food item and household who reported spending more than 80% of income on food. Excluding these households yields a useable sample of 4622 households.

Empirical Cumulative Distribution Functions

I begin with a simple graphical analysis of the distribution of unit values a relatively small number of food aggregates. Differences between unit values and prices will only be problematic for recovering information about demand behavior if these differences are correlated with variables of interest. To see whether this is the case, Figures 2-4 plot the empirical cumulative distribution functions of six food aggregates (2% Milk, Eggs, Chicken, Potatoes, Lettuce and Bread) for three well-known demand shifters, the presence of children, the age of the head of household and total household income. If unit values and prices coincide, or alternatively if their difference is uncorrelated with demand shifters of interest, these empirical distribution functions should not differ in a statistically significant or economically important way.

Figure 2 plots the empirical distribution of unit values for two adult households with and without children for six representative commodities.

Figure 2. The Effect of Children on Unit Value

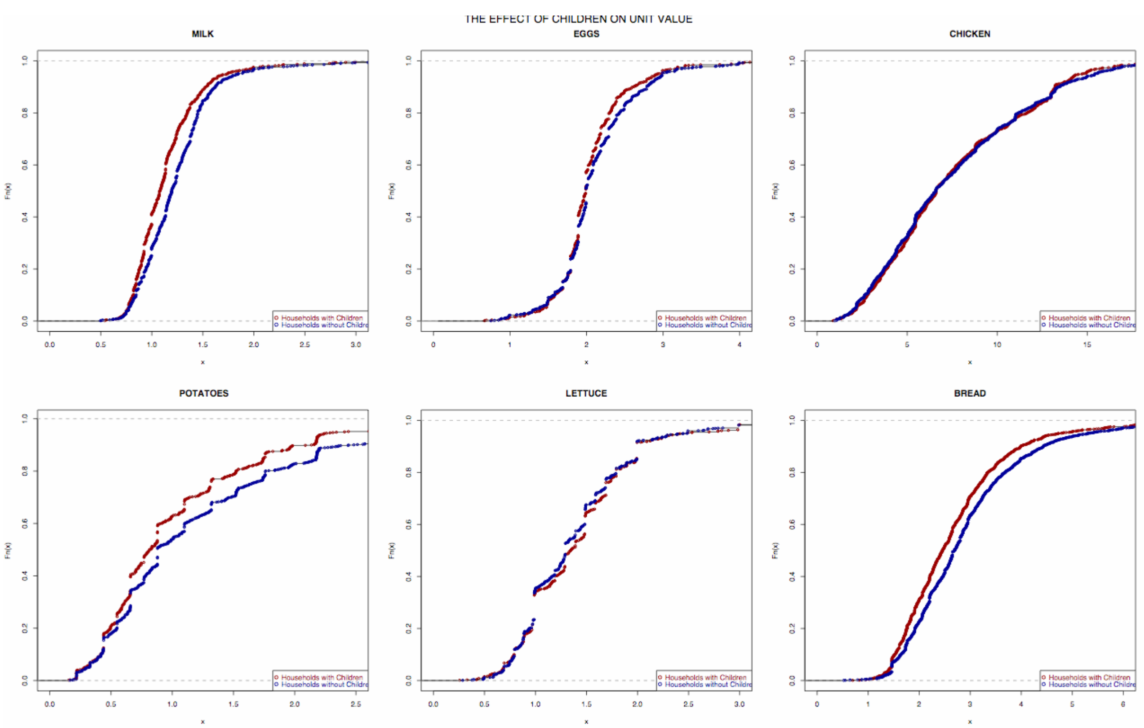


Figure 2 is interesting in several respects. First note that for four of the six goods under consideration (2% milk, eggs, potatoes and bread), households with children pay less for these aggregates than households without children. For chicken and lettuce the effect is not statistically significant at any conventional level. Table 1 reports the t-statistics and p-values for simple difference in means tests.

Table 1. Effect of Children on Unit Values

Effect of Children		
Commodity	t-Statistic	P-Value
Milk	5.41	0.00
Eggs	2.51	0.01
Chicken	0.05	0.96
Potatoes	4.47	0.00
Lettuce	0.9	0.36
Bread	5.56	0.00

Figure 3 plots the empirical distribution of unit values for households with no children where the head of household is over 65 as compared to households with no children where the head of household is under 65. Again, if unit values are to be treated as prices and assumed exogenous to the household, age and reported unit values should not be correlated.

Figure 3. The Effect of Age on Unit Value

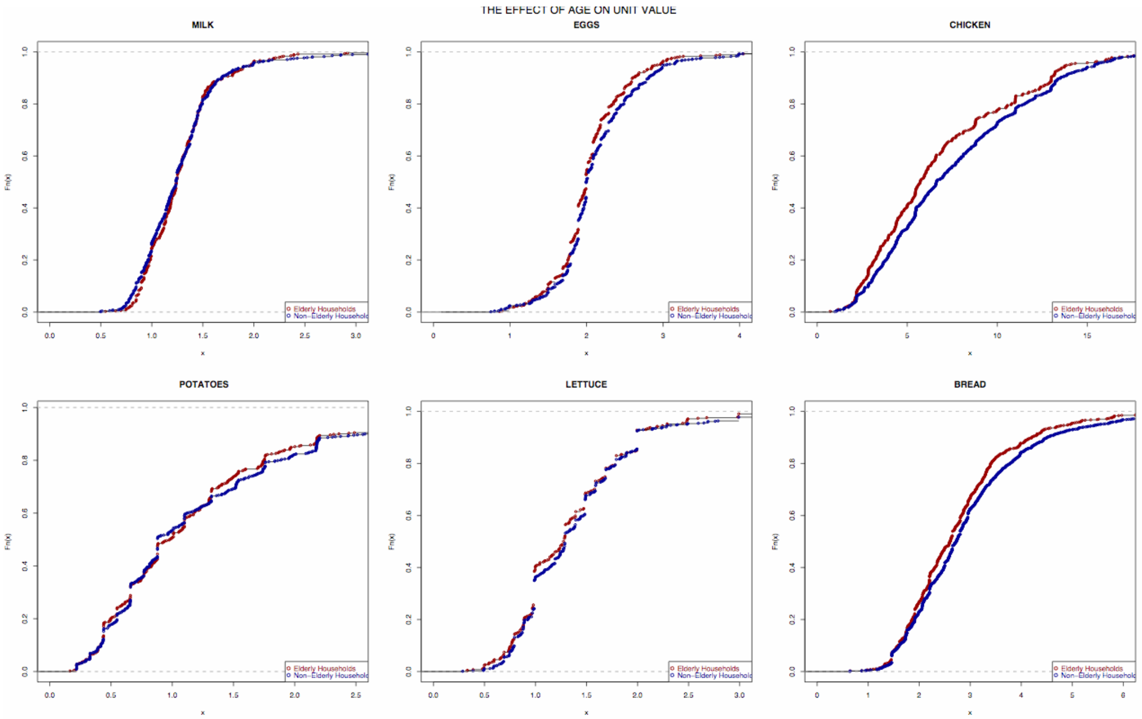


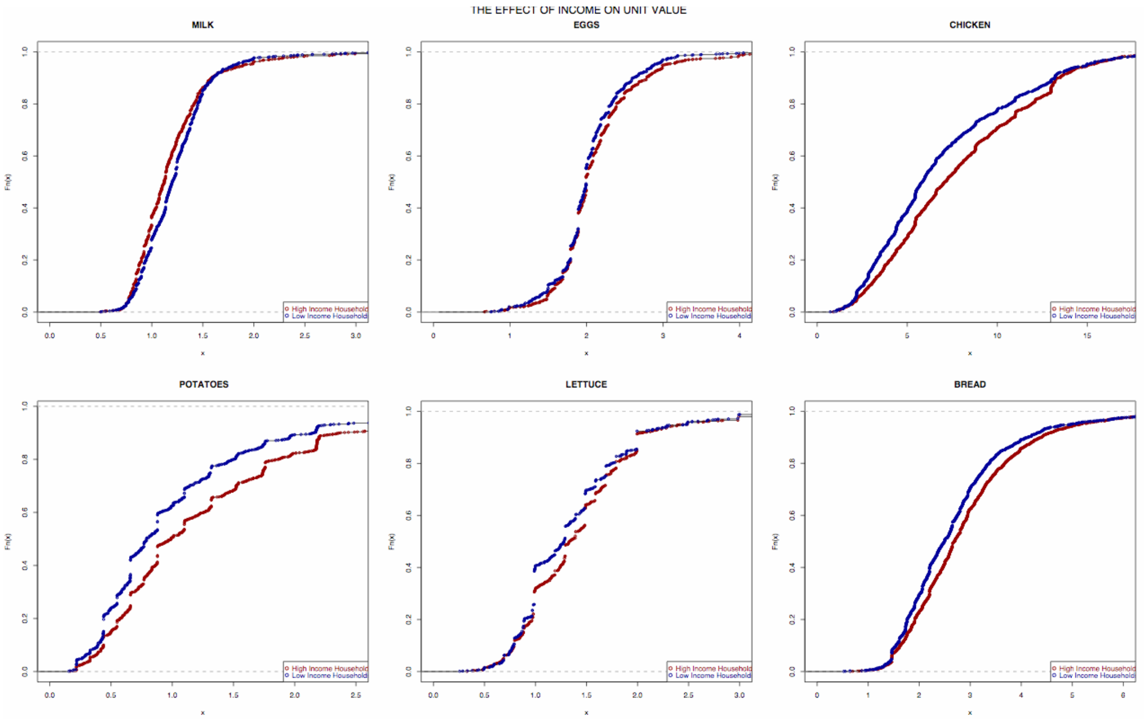
Figure 3 shows us that for three of the six commodities under consideration there is evidence that unit values differ in a statistically significant fashion between elderly and non-elderly households. In particular, elderly households report significantly lower unit values for eggs, chicken and bread. Table 2 reports t-statistics and p-values for a simple test of equality of means for the commodities of interest.

Table 2 The Effect of Age on Unit Value

Effect of Age		
Commodity	T-Stat	P-Value
Milk	0.29	0.77
Eggs	2.49	0.01
Chicken	2.95	0.00
Potatoes	0.98	0.33
Lettuce	1.47	0.15
Bread	3.58	0.00

Figure 4, plots the empirical cumulative distribution function for high and low income households, where high or low is defined as being above or below median income respectively.

Figure 4. The Effect of Income on Unit Value



Income (note this is total household income, not expenditure) seems to have a large influence on the unit values reported by households in our sample. For all six food aggregates, there is a clear visual difference between the empirical CDFs. For eggs, chicken, potatoes, lettuce and bread, low-income households report significantly lower unit values for these goods than high-income households. Intriguingly, for 2% milk, high-income households report lower unit values. Table 3 shows this difference is statistically significant for all goods.

Table 3. Effect of Income on Unit Value

Effect of Income		
Commodity	T-Stat	P-Value
Milk	2.23	0.03
Eggs	3.84	0.00
Chicken	5.2	0.00
Potatoes	5.46	0.00
Lettuce	2.94	0.00
Bread	5.01	0.00

Estimating the Demand for Quality

The results of the previous section provide preliminary evidence that unit values and prices differ in important ways and that simply treating unit values as prices may yield faulty inference. I now turn to a more systematic study of the scope of the unit value problem. Using an approach that dates back to Prais and Houthakker (1955) I estimate the expenditure elasticity of quantity for a large number of food aggregates. If the expenditure elasticity of quality is significantly different from zero, then prices and unit values also differ in a statistically significant manner.

For a given household, in a given cluster, for a given food aggregate I rewrite (2), suppressing subscripts.

$$\ln V = \ln \pi + \ln \upsilon . \tag{3}$$

The logarithm of the unit value is equal to the logarithm of the general price level plus the logarithm of quality. If quality is not important (e.g , $\ln \upsilon = 0$) then unit values and prices will coincide. Following Deaton (1997) , quality can be modeled as a function of the log of total food expenditure $\ln(X)$ and other demand shifters *S*.

$$\ln(\upsilon) = \alpha + \beta \ln X + \sum_{j=1}^{J} \theta_j S_j + \varepsilon \tag{4}$$

For the purposes of the current exercise, the main variable of interest will be ß, which is as the expenditure elasticity of quality. One important caveat is that prices (π) are not observed. A standard assumption in the unit value literature is that relative prices for a given food aggregate do not vary within a given cluster. In the context of the current exercise, I define a cluster to be a geographic region in a given quarter. Cluster dummies control for the effects of relative prices within a cluster and permit us to identify the coefficient on the logarithm of total food expenditure, e.g. the expenditure elasticity of quality from (4). Finally, I use the log of total household size and the share of individuals in a household in four age categories (under 15, 15-24,25-65 and over 65) as demand shifters *S*.

Substituting (4) into (3), yields an estimable equation

$$\ln V = \alpha + \beta \ln X + \sum_{j=1}^{J} \theta_j S_j + \sum_{c=1}^{C-1} \delta_c D_c + \varepsilon , \tag{5}$$

where $\sum_{c=1}^{C-1} \delta_c D_c$ are the cluster dummies.

Finally, there may be reason to believe that suggests that the log of total food expenditure is at least potentially endogenous in (5). Several studies, notably LaFrance (1991, (1993) and Dhar, Chavas and Gould (2003) , found that failing to control for expenditure endogeneity can lead to misleading inferences[1]. For this reason, I instrument $\ln X$ using the log of total household income. Equation (5) is estimated using two stage least squares for all 196 food aggregates in the FOODEX. Table 4 summarizes the results of estimating (5) for a typical food aggregate, in this case bread. For ease of presentation, the appendix contains the estimated quality elasticity of expenditure for all 196 food aggregates.

Table 4 summarizes the results of estimating (5) on the unit value of bread. It is indicative of the results obtained for other goods.

Table 4. Quality Demand Equation for Bread

N=5945

R2=0.1265

Variable	Estimate	Std. Err.	T-Stat	P-Value
Log of Household Size	-0.078	0.015	-5.38	0.00
Share of Household over 65	-0.016	0.013	-1.18	0.24
Share of Household between 15 and 24	-0.037	0.024	-1.52	0.13
Share of Household under 15	-0.113	0.030	-3.82	0.00
Log of Total Food Expenditure	0.094	0.008	11.99	0.00
		Cluster Dummies		
Constant	0.450	0.042	10.79	0.00

The key variable of interest is the coefficient on the log of total food expenditure and represents the expenditure elasticity of quality. For bread, the expenditure elasticity of quality is 0.094 and is significantly different from zero at all conventional significance levels. The log of household size is negative and significantly different from zero, suggesting that on average larger households purchase lower quality bread. The effect of

[1] To see whether total food expenditure was endogenous in this context, I used a variant of the Durbin-Wu-Hausman test suggested by Davidson and MacKinnon (2004) p. 340. In an overwhelming majority of cases, exogeneity of total food expenditure was rejected.

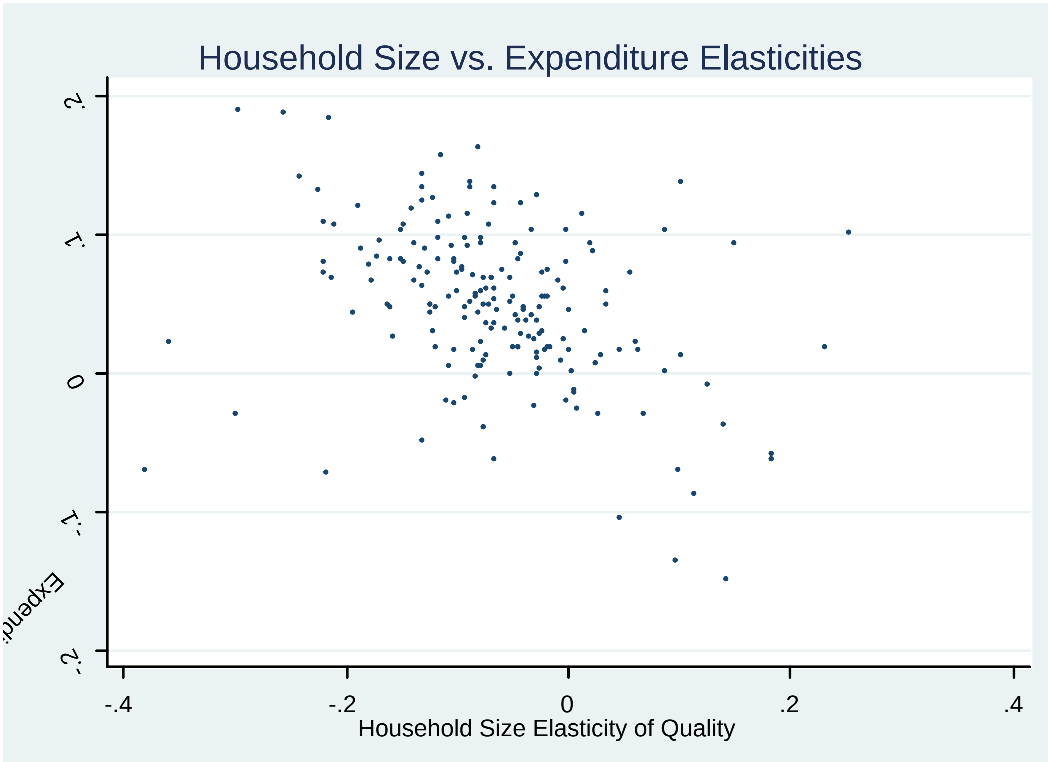
household composition is negative, relative to the omitted share of adults between the ages of 25 and 65. This effect is statistically significant only for the share of a household under the age of fifteen.

Appendix A contains the results of estimating (5) for all 196 food aggregates in the FOODEX. I will now briefly summarize the main findings. The transaction weighted mean of expenditure elasticity is relatively modest, 0.055 and the standard error is .0040. However the mean conceals considerable variation. Point estimates of the expenditure elasticities of quality range from between -0.393 ("Other poultry meat and offal") and 0.444 ("Cured fish"). Both the maximum and minimum are statistically significant at conventional levels. 116 food aggregates are statistically significantly different from zero at the 10% level and 96 food aggregates are statistically significantly different from zero at the more conservative 5% level. In other words, for the relatively detailed commodities considered in this paper (far more detailed than would typically be used in almost any demand system estimation), and where one might expect quality issues to be mitigated, approximately half have statistically significant quality elasticities. If I restrict attention to those food aggregates with more than 300 recorded transactions (approximately 75% of the 196 food aggregates), roughly two thirds of the remaining food aggregates have expenditure elasticities of quality that are significant at the 10% level. Interestingly, 29 food aggregates have negative quality elasticities, although only 6 of these are significant at the 10% level[2].

Finally, I compare the expenditure elasticity of quality with the effects of the logarithm of household size. Figure 5 plots the point estimates for these two components for all food aggregates. The relationship between food expenditure and household size is clearly negative. Food aggregates with large expenditure elasticities of quality have small household size elasticities of quality and vice-versa. The correlation coefficient is -0.61. The implication is that increasing household size has a similar effect to reducing household food expenditure. This is consistent with the results from Figure 2.

[2] "Other poultry meat and offal", "Other cured meat", "Gum", "Other coffee", "Lard" and "Infant formula".

Figure 5.



<u>Aggregating over Commodities</u>

In the previous section, expenditure elasticities of quality were shown to be significant for slightly more than half of the commodities considered. A reasonable criticism of the preceding analysis is that the food aggregates considered are far more detailed than those typically included in most demand studies. To see whether the effect of quality averages out or becomes more important when commodities are aggregated further, I construct several broader food aggregates and reestimate the quality demand equation (5). Table 5 lists the composition of the broad food aggregates and Table 6 summarizes the results of estimating (5) on these broad food aggregates.

Table 5. Composition of Aggregate Commodities

BROAD AGGREGATE	Commodities Aggregated
MILK	"Cream (excluding sour cream)", "Fluid whole milk", "Low-fat milk (2%)", "Low-fat milk (1%)", "Fluid skim milk".
CHEESE	"Cheddar cheese", "Grated cheese", "Process cheese", "Cottage cheese", "Other cheese".
BEEF	"Hip cuts (excluding shank cuts)", "Loin cuts", "Rib cuts", "Chuck cuts (excluding shank cuts)", "Stewing beef", "Ground beef (including patties)", "Other beef (including shank cuts)".
FRESH FRUIT	"Apples", "Bananas and plantains", "Grapefruit", "Grapes", "Lemons and limes", "Melons", "Oranges and other citrus fruit", "Peaches and nectarines", "Pears", "Plums", "Other tropical fruit", "Strawberries" ,"Other fresh fruit".
FRESH VEGETABLES	"Green or wax beans", "Broccoli", "Cabbage", "Carrots", "Cauliflower", "Celery", "Corn", "Cucumbers", "Lettuce", "Mushrooms", "Onions", "Peppers", "Potatoes", "Radishes", "Spinach", "Tomatoes", "Turnips and rutabagas", "Other seed and gourd vegetables", "Other root vegetables", "Other leaf and stalk vegetables".

Table 6. Expenditure Elasticity of Quality for Aggregated Commodities

Aggregate	Estimate	Std. Err.	T-Statistic	P-Value
Milk	0.070	0.009	7.85	0.00
Cheese	0.058	0.011	5.24	0.00
Beef	0.092	0.014	6.77	0.00
Fresh Fruits	0.127	0.010	12.65	0.00
Fresh Vegetables	0.102	0.011	9.44	0.00

For each of these broad food aggregates, the expenditure elasticity of is significantly different from zero at all conventional levels. This is true even when the expenditure elasticities of the components were not significantly different from zero at the 5% level. Consider the case of Milk. Of the five commodities that make up milk, only "Cream (excluding sour cream)" had a quality elasticity that was significantly different from zero. However the resulting aggregate elasticity is statistically significant and economically important. Given our definition of quality, (1), this result is not surprising. If the unit value problem is due to unobserved quality heterogeneity, aggregating over a large number of commodities is unlikely to ameliorate the issue.

Semiparametric Approach

I now relax assumption that quality is l,a linear function of total food expenditure. Indeed, there is no reason to believe that the expenditure elasticity of quality is constant over the range of the log of total food expenditure. To this end, I generalize Deaton's approach and allow the effect of total food expenditure to enter nonparametrically. Specifically, I rewrite the demand for quality υ as

$$\upsilon = f\left(\ln X\right) + \sum_{j=1}^{J} \theta_j S_j + \varepsilon . \qquad (6)$$

As before, the demand shifters S capture the demographic structure of the household. Substituting (6) into (3) and incorporating the cluster-level dummies yields

$$\ln V = \alpha + f\left(\ln X\right) + \sum_{j=1}^{J} \theta_j S_j + \sum_{c=1}^{C-1} \delta_c D_c + \varepsilon . \qquad (7)$$

Note that in particular, I am interested in $\partial f(\cdot) / \partial \ln X$, which is the nonparametric expenditure elasticity of consumption (analogous to ß in equation (5)). I now provide an overview of the estimation strategy.

Estimation Technique

There are a number of ways to estimate models of the form (7). I employ a parsimonious approach known as penalized regression splines (p-splines) that is relatively common in the statistical literature, but is somewhat less well known in econometrics. In its present form, this approach was first proposed by Eilers and Marx (1996) and Ruppert and Carroll (1997) [3].

The smooth function $f\left(\cdot\right)$ can be written using a cubic radial basis spline. The cubic degree radial basis spline model (sometimes called a thin plate spline) for the logarithm of total food expenditure, for household *i* can be written

$$f\left(\ln X_i\right) = \gamma_0 + \gamma_1 \ln X_i + \sum_{k=1}^{K} \mu_k \left|\ln X_i - \kappa_k\right|_+^3 , \qquad (8)$$

where, $\kappa_1 < \kappa_2 < \ldots < \kappa_K$, denote the knot points and the functions $\left|\ln X_i - \kappa_k\right|_+^3$ are the cube of the absolute value of the difference between a value of the log of total food expenditure and a given knot point. Following the recommendation of Ruppert et al. (2003) the number of knots is chosen according to $K = \min(0.25 \times \text{ number of unique } X_i, 35)$ and are evenly spaced over the range of $\ln X$.

[3] For a textbook length treatment of this approach see Ruppert, Wand and Carroll (2003) .

Recasting the estimation problem in matrix form, write the vector of unit values $\mathbf{v} = \left[V_1 \ldots V_N\right]^{\mathrm{T}}$, define the design matrices

$$\begin{aligned}\mathbf{X} &= \left[1, \ln X\right]_{1 \le i \le N} \\ \mathbf{Z} &= \left[\left|\ln X_i - \kappa_1\right|^3, \ldots, \left|\ln X_i - \kappa_K\right|^3\right]_{1 \le i \le N}\end{aligned}, \tag{9}$$

coefficient vectors $\mathbf{g} = [\gamma_0, \ldots, \gamma_p]^{\mathrm{T}}$, $\mathbf{m} = [\mu_1, \ldots, \mu_K]^{\mathrm{T}}$ and error term $\mathbf{e} = \left[\varepsilon_1 \ldots \varepsilon_N\right]^{\mathrm{T}}$. The estimation problem can be concisely written as

$$\mathbf{v} = \mathbf{Xg} + \mathbf{Zm} + \mathbf{e}\,. \tag{10}$$

Note that if one wanted, equation (10) can be fit using ordinary least squares. However, this can result in overfitting the component being modeled nonparametrically. In order to avoid this, the influence of the extended basis function $\mathbf{Z}$ needs to be constrained in some way. Following Ruppert et al. (2003) , based on earlier work by Robinson (1991) and Brumback, Ruppert and Wand (1999) ,this is accomplished by writing $\mu_k \sim N\left(0, \sigma_\mu^2\right)\ \forall k$. In other words, by modeling the parameters on the extended basis function as random with mean zero and finite variance. The result is a fit where the degree of smoothness is a function of σ_μ^2. Note that ordinary least squares is the special case where the variance term, σ_μ^2, is infinite.

More formally, given (10) assuming

$$\mathrm{E}\begin{pmatrix}\mathbf{m} \\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}0 \\ 0\end{pmatrix}, \tag{11}$$

and

$$\mathrm{Cov}\begin{pmatrix}\mathbf{m} \\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\sigma_\mu^2 \mathbf{I} & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{I}\end{pmatrix}, \tag{12}$$

the log likelihood function can be written

$$\ell\left(\mathbf{g}, \Sigma\right) = -\frac{1}{2}\left(n \log(2\pi) + \log\left|\Sigma\right| + \left(\mathbf{v} - \mathbf{Xg}\right)^{\mathrm{T}} \Sigma^{-1} \left(\mathbf{v} - \mathbf{Xg}\right)\right), \tag{13}$$

where $\Sigma = \mathrm{Cov}(\mathbf{v}) = \sigma_\mu^2 \mathbf{Z}\mathbf{Z}^{\mathrm{T}} + \sigma_\varepsilon^2 \mathbf{I}$.

Incorporating additional parametric covariates is simply a matter of appending additional columns to the $\mathbf{X}$ matrix and adding the corresponding parameters to the vector **g**. E.g. $\tilde{\mathbf{X}} = [\mathbf{X} \mid \mathbf{S} \mid \mathbf{D}]$ and $\tilde{\mathbf{g}} = [\mathbf{g} \mid \theta_1 \ldots \theta_J \mid \delta_1 \ldots \delta_{C-1}]$, where as before, **S** is a matrix of demographic variables, **D** is a matrix of cluster dummies and their parameters are θ_j and δ_c respectively. Note that (7) trivially nests (5) and as a result one can use simple likelihood ratio tests to see whether the added complexity of (10) is supported by the data.

Recall that for the semiparametric model, the expenditure elasticity of quality is $\partial f(\ln X) / \partial \ln X$. The nonparametric estimate of the expenditure elasticity of quality is obtained by taking the derivative of (8) with respect to $\ln X$ and evaluating the result at the estimates obtained from maximizing (13). It can be written,

$$\frac{\partial f\left(\ln X_i\right)}{\partial \ln X_i} = \gamma_1 + \sum_{k=1}^{K} 3\mu_k \left(\ln X_i - \kappa_k\right)\left|\ln X_i - \kappa_k\right| . \tag{14}$$

Controlling for endogeneous total food expenditure, in the sense that in equation (8) $E\left(\varepsilon \mid \ln X\right) \neq 0$, is only slightly more difficult in the current semiparametric approach. I follow Blundell, Duncan and Pendakur (1998) and use a simple two-step procedure due to Holly and Sargan (1982) and later generalized by Newey, Powell and Francis (1999) . First I regress the log of total food expenditure on the log of total income (y)

$$\ln X = \lambda \ln y + \omega , \tag{15}$$

where $E(\omega \mid y) = 0$. Then appending an estimate of ω, to the model (8) controls and provides a test for the endogeneity of total food expenditure.

Written in this way the estimation of the nonparametric component is cast in the context of a simple linear random effects model and estimated via restricted maximum likelihood using mixed effects software (e.g. SAS PROC MIXED or S-Plus/R nlme) [4]. In short, the p-spline model described above can be written as an additive mixed model. In the current application, the model is fit using R (R Development Core Team (2006)) and the nlme module (Pinheiro et al. (2006))[5].

---

[4] Ngo and Wand (2004) provides examples.

[5] A textbook length treatment of this type of model and of the nlme software is available in Pinheiro and Bates (2000) .

Semiparametric Results

For ease of exposition, I report the results of estimating (10) for the same six commodities considered in the first section (2% Milk, Eggs, Chicken, Potatoes, Bread and Lettuce). Figures 6-11 summarize these results. For each commodity, I plot the estimate, the derivative and a 2 standard deviation wide point wise confidence band. I begin by discussing results for 2% Milk, Bread and Lettuce, where the semiparametric model offers clear advantages. I then discuss the results for Eggs, Chicken and Potatoes, where the semiparametric model does not offer any gains.

Figure 6 reports the estimates of (10) when applied to the food aggregate 2% Milk, Figure 7 reports the estimates of (10) when applied to the food aggregate Eggs and Figure 8 reports the estimates of (10) when applies to the food aggregate Lettuce.

Figure 6. Semiparametric Expenditure Elasticity of 2% Milk

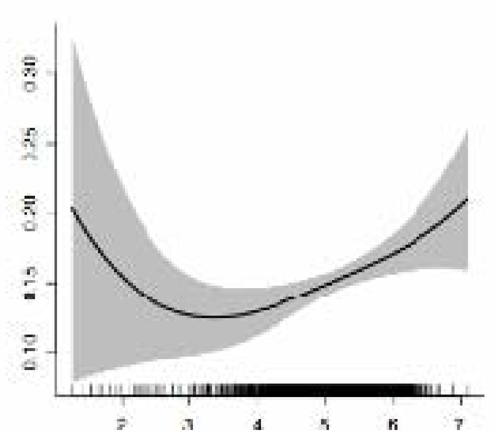
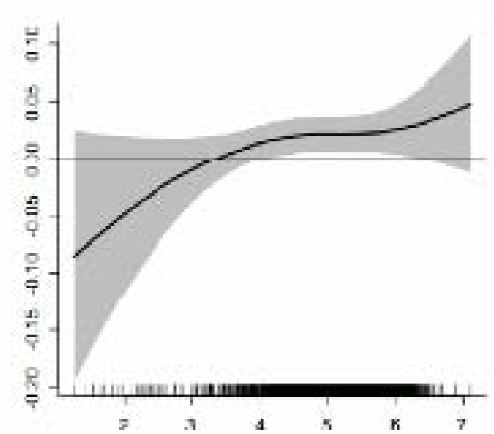
2% Milk	
$\hat{f}(\ln(X))$	$\partial\hat{f}(\ln X)/\partial\ln X$
	

Figure 7. Semiparametric Expenditure Elasticity of Bread

Bread	
$\hat{f}(\ln(X))$	$\partial\hat{f}(\ln X)/\partial\ln X$

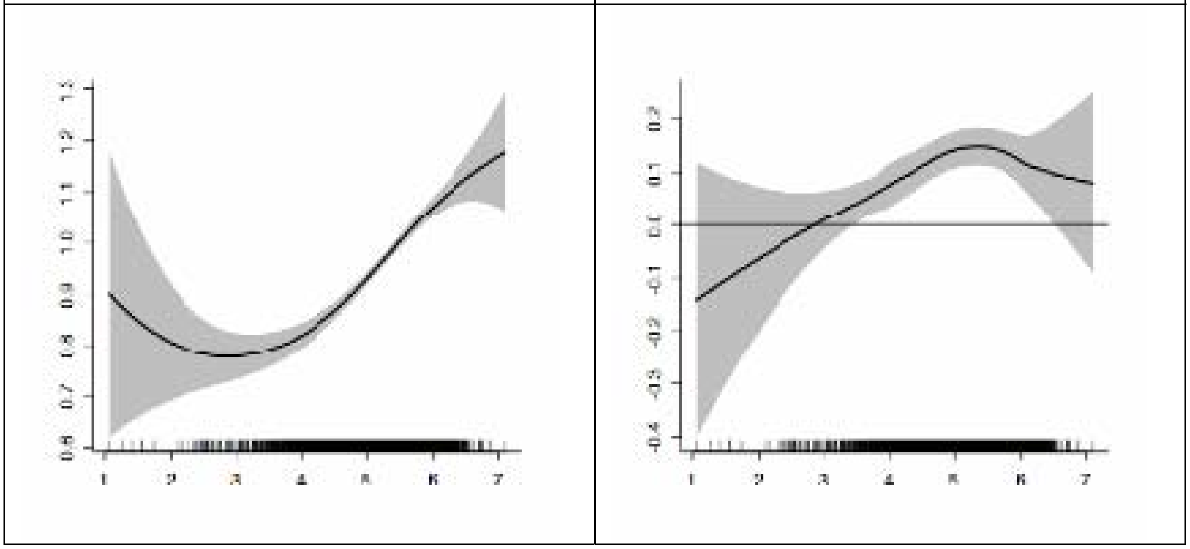
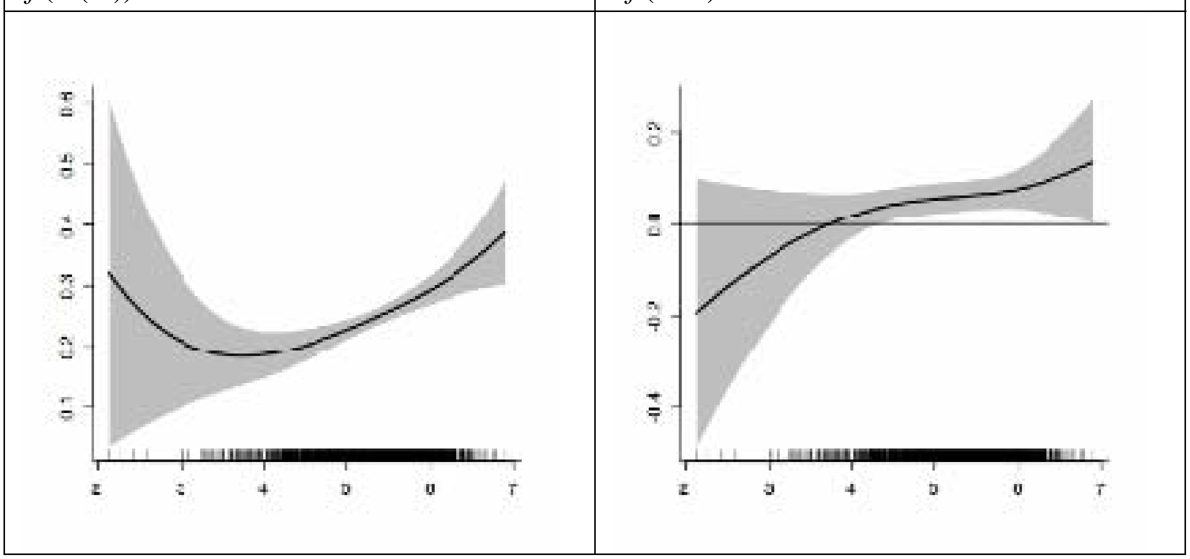


Figure 8. Semiparametric Expenditure Elasticity of Lettuce

Lettuce	
$\hat{f}(\ln(X))$	$\partial\hat{f}(\ln X)/\partial\ln X$



In the left panel, $\hat{f}(\ln(X))$ displays some curvature for smaller values of $\ln X$. Over the bulk of the data, the effect appears to be increasing in a somewhat linear fashion for all three food aggregates. $\partial\hat{f}(\ln X)/\partial\ln X$, the nonparametric expenditure elasticity of quality is significantly different from zero over at least some subset of the range of the log of total food expenditure. Note that estimated quality elasticity from (5) for 2% Milk is 0.008 and is not significantly different from zero. Thus at least for the case of milk, the semiparametric model explains variation in the data that the parametric model cannot. For bread and lettuce, the estimated expenditure elasticity of quality is increasing over the range of the data. The linear model (5) does not provide an adequate description of the data, at least for these commodities.

I consider a group of commodities where the semiparametric offers no advantage. Figure 9 reports the estimates of (10) when applied to the food aggregate Eggs, Figure 10 reports the estimates of (10) when applied to the food aggregate Chicken and Figure 11 reports the estimates of (10) when applies to the food aggregate Potatoes.

Figure 9. Semiparametric Expenditure Elasticity of Eggs

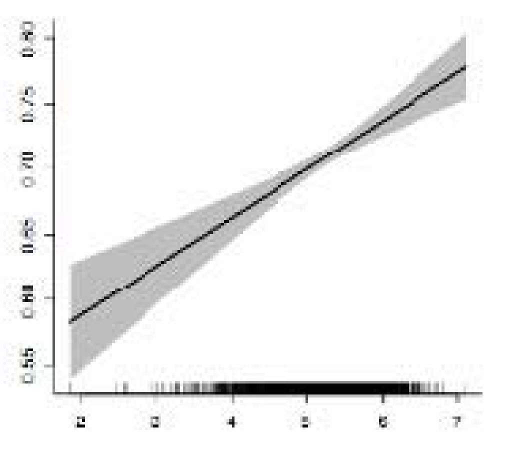
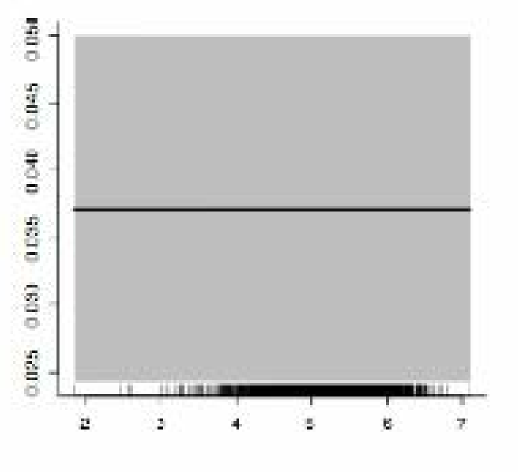
Eggs	
$\hat{f}(\ln(X))$	$\partial\hat{f}(\ln X)/\partial\ln X$
	

Figure 10. Semiparametric Expenditure Elasticity of Chicken

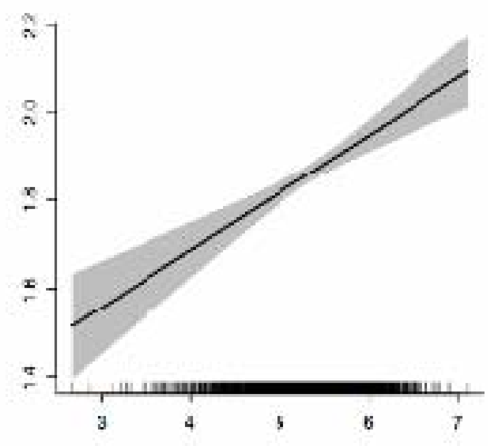
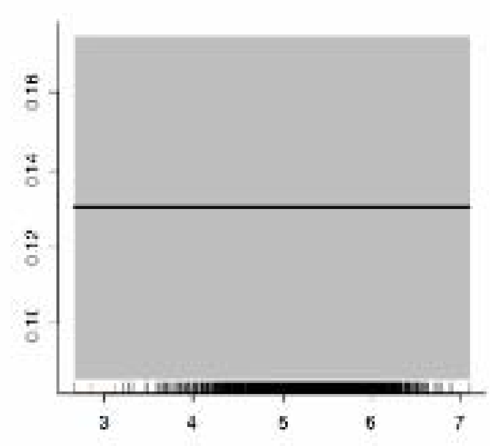
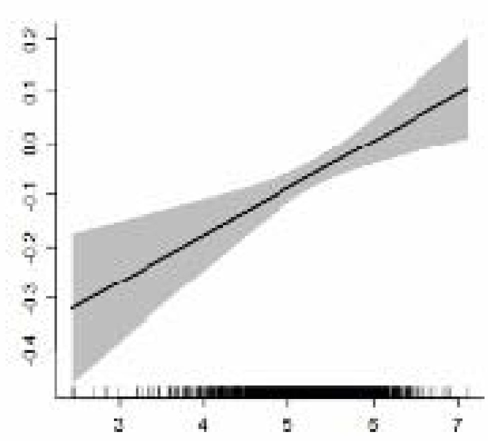
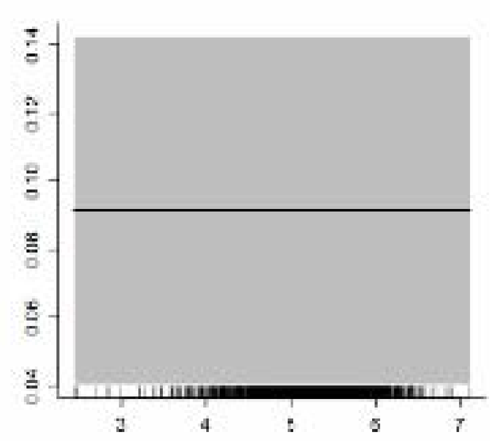
Chicken	
$\hat{f}(\ln(X))$	$\partial\hat{f}(\ln X)/\partial \ln X$
	

Figure 11. Semiparametric Expenditure Elasticity of Potatoes

Potatoes	
$\hat{f}(\ln(X))$	$\partial\hat{f}(\ln X)/\partial \ln X$
	

The results for Eggs, Chicken and Potatoes are broadly similar. For this group of commodities the semiparametric methodology offers little advantage. Figure 9 reports the estimates of (9) when applied to the food aggregate Eggs, Figure 10 reports the estimates of (9) when applied to the food aggregate Chicken and Figure 11 reports the estimates of (9) when applies to the food aggregate Potatoes. As before these figures plot $\hat{f}(\ln X)$ in the left panel and $\partial\hat{f}(\ln X)/\partial \ln X$ in the right panel**.** In contrast to the previous results, the estimate of $\hat{f}(\ln X)$ is virtually linear for these food aggregates. As a result $\partial\hat{f}(\ln X)/\partial \ln x$, the nonparametric expenditure elasticity of quality is flat. For comparative purposes, note that estimated quality elasticity from (5) is 0.035 for eggs, 0.13 for chicken and 0.072 for potatoes and is significantly different from zero in all cases. In all cases, the estimates from the linear model (5) lies within the 2 standard deviation confidence band. Thus, in contrast to the results above, in the case of eggs, chicken and potatoes, the semiparametric model offers no real advantage over the standard linear model, but does not result in misleading inferences.

Recall the semiparametric model (7) nests the linear model (5). For half of goods considered the departures from linearity are statistically significant. This would suggest that in general, the added computational complexity associated with the semiparametric approach may be worthwhile.

Conclusion

This paper represents an attempt to assess the magnitude of the unit value problem. I estimate a quality demand equation for a wide variety of food aggregates. The results suggest that the unit value problem is ubiquitous. Specifically, I find that the expenditure elasticity of quality is statistically significant in almost half of the cases considered. This is true even thought the commodities considered are relatively detailed and from the point of view of an outside observer appear to be relatively homogeneous.

I then aggregate from the detailed food categories in the 2001 FOODEX to see whether aggregating over a wider range of commodities exacerbates or mitigates the quality bias. As expected, simply aggregating over commodities does not ameliorate the problem. Given expenditure and price endogeneity is present both at the household and at more aggregate levels (see Dhar et al. (2003)) and is present for detailed as well as aggregated commodities, it is difficult to imagine a situation in which unit values should be treated as prices.

Finally, I extend the standard quality demand model to a semiparametric framework. The semiparametric specification offers greater explanatory power for some commodities. Notably, the expenditure elasticity of quality may be significantly different from zero only over a range of total food expenditure. Again this argues against simply using unit values as prices.

References

Blundell, R., A. Duncan and K. Pendakur. 1998. Semiparametric estimation and consumer demand. *Journal of Applied Econometrics* 13: 435-461.

Brumback, B., D. Ruppert and M. P. Wand. 1999. Comment on 'Variables selection and function estimation in additive nonparametric regression using data-based prior' by Shively, Kuoh and Wood. *Journal of the American Statistical Association* 94: 794-797.

Cox, T. L. and M. K. Wohlgenant. 1986. Prices and quality effects in cross-sectional demand analysis. *American Journal of Agricultural Economics* 68(4): 908-919.

Davidson, R. and J. G. MacKinnon. 2004. *Econometric Theory and Methods*. New York: Oxford University Press.

Deaton, A. 1997. *The Analysis of Household Sruveys: A Microeconometric Approach to Development Policy*. Baltimore: The Johns Hopkins University Press.

Dhar, T., J. P. Chavas and B. Gould. 2003. An empirical assessment of endogeneity issues in demand analysis for differentiated products. *American Journal of Agricultural Economics* 85(3): 605-617.

Dong, D. and H. M. Kaiser. 2005. Coupon redemption and its effect on household cheese purchases. *American Journal of Agricultural Economics* 87(3): 689-702.

Dong, D., J. S. Shonkwiler and O. Capps. 1998. Estimation of demand functions using cross-sectional household data: The problem revisited. *American Journal of Agricultural Economics* 80(3): 466-473.

Eilers, P. H. C. and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2): 89-121.

Holly, A. and J. Sargan. 1982. Testing for exogeneity in a limited information framework. Cahiers de Recherches Economiques, No. 8204, Universite de Lausanne.

Houthakker, H. 1952. "Compensated changes in quanitities and qualities consumed". *Review of Economic Studies* 19: 155-164.

LaFrance, J. T. 1991. When is expenditure exogenous in separable demand models. *Western Journal of Agricultural Economics* 16: 1991.

LaFrance, J. T. 1993. Weak separability in applied welfare analysis. *American Journal of Agricultural Economics* 75: 770-775.

Nelson, J. A. 1991. Quality variation and quantity aggregation in consumer demand for food. *American Journal of Agricultural Economics* 73(4): 1204-1212.

Newey, W. K., J. L. Powell and V. Francis. 1999. Semiparametric estimation of triangula simultaneous equation models. *Econometrica* 67(3): 565-603.

Ngo, L. and M. P. Wand. 2004. Smoothing with mixed model software. *Journal of Statistical Software* 9(1).

Pinheiro, J. and D. Bates. 2000. *Mixed Effects Models in S and S-PLUS*. New York: Springer.

Pinheiro, J., D. Bates, S. Debroy and D. Sarkar. 2006. nlme: Linear and nonlinear mixed effects models.

Prais, S. J. and H. Houthakker. 1955. *The Analysis of Family Budgets*. Cambridge: Cambridge University Press.

Robinson, G. K. 1991. That BLUP is a good thing: The estimation of random effects. *Statistical Science* 6: 15-51.

Ruppert, D. and R. Carroll. 1997. Penalized regression splines. Cornell University.

Ruppert, D., M. P. Wand and R. J. Carroll. 2003. *Semiparametric Regression*. Cambridge: Cambridge University Press.

Theil, H. 1952. Qualities, prices and budget enquiries. *Review of Economic Studies* 19: 129-147.

Appendix A

This table reports the expenditure elasticites of quality for each of 196 distinct food aggregates in the 2001 FOODEX. For each commodity, I report the number of times it was transacted (e.g. I observe the same household on multiple purchase occasions, degrees of freedom have been appropriately corrected), the expenditure elasticity of quality, the standard error, a T-Statistic and the resulting P-Values.

Food Aggregate	N	Estimate	Std. Err.	T-Stat	P-Value
Hip cuts (excluding shank cuts)	1176	0.000	0.021	0.014	0.49
Loin cuts	394	0.128	0.043	2.967	0.00
Rib cuts	398	0.050	0.047	1.052	0.15
Chuck cuts (excluding shank cuts)	247	0.019	0.051	0.373	0.35
Stewing beef	248	0.002	0.029	0.058	0.48
Ground beef (including patties)	2242	0.037	0.014	2.598	0.00
Other beef (including shank cuts)	94	-0.071	0.120	0.591	0.28
Leg cuts (excluding hocks)	131	0.019	0.082	0.227	0.41
Loin cuts	1396	0.048	0.022	2.150	0.02
Belly cuts	113	0.131	0.067	1.959	0.03
Shoulder cuts (excluding hocks)	151	0.138	0.059	2.348	0.01
Other pork (including hocks)	174	-0.019	0.062	0.307	0.38
Chicken (including fowl)	2787	0.126	0.024	5.245	0.00
Turkey	431	-0.049	0.076	0.648	0.26
Other poultry meat and offal	40	-0.394	0.211	1.870	0.03
Veal	208	0.104	0.066	1.573	0.06
Liver	120	0.068	0.084	0.810	0.21
Other offal	87	0.055	0.193	0.285	0.39
Lamb and mutton - fresh or frozen	155	0.219	0.086	2.548	0.01
Bacon	981	0.058	0.019	3.096	0.00
Ham (excluding cooked ham)	647	0.019	0.037	0.522	0.30
Other cured meat	195	-0.149	0.067	2.223	0.01
Uncooked sausage	745	0.048	0.030	1.573	0.06
Bologna	633	-0.029	0.035	0.837	0.20
Wieners	1038	0.056	0.028	1.972	0.02
Other cooked/cured sausage	1346	0.109	0.024	4.551	0.00
Cooked (boiled) ham	1377	0.023	0.022	1.086	0.14
Other ready-cooked meat	1815	0.053	0.025	2.118	0.02
Other meat preparations	532	0.056	0.037	1.523	0.06
Hams - canned	139	-0.030	0.088	0.342	0.37
Other canned meat and meat preparations	468	0.038	0.055	0.685	0.25
Cod	62	-0.009	0.099	0.087	0.47
Flounder and sole	154	0.037	0.058	0.646	0.26
Haddock	100	0.005	0.048	0.116	0.45
Salmon	397	0.045	0.047	0.953	0.17
Other sea fish	359	0.118	0.050	2.380	0.01
Freshwater fish	97	-0.070	0.188	0.375	0.35
Pre-cooked frozen fish portions	223	0.080	0.057	1.386	0.08

Cured fish	117	0.449	0.176	2.558	0.01
Salmon	308	0.093	0.048	1.926	0.03
Tuna	656	0.184	0.036	5.097	0.00
Other canned fish	191	0.103	0.081	1.281	0.10
Shrimps and prawns	345	0.056	0.045	1.241	0.11
Shellfish - other	372	0.139	0.059	2.349	0.01
Cream (excluding sour cream)	1079	0.069	0.026	2.705	0.00
Low-fat milk (1%)	1693	0.016	0.011	1.407	0.08
Fluid whole milk	1162	0.005	0.010	0.553	0.29
Low-fat milk (2%)	3406	0.008	0.007	1.172	0.12
Fluid skim milk	952	0.013	0.015	0.864	0.19
Specialty milk products	103	0.063	0.039	1.630	0.05
Yogurt	2007	0.045	0.016	2.756	0.00
Butter	1347	0.016	0.010	1.662	0.05
Cheddar cheese	1927	0.019	0.011	1.639	0.05
Grated cheese	316	0.016	0.041	0.388	0.35
Process cheese	1253	0.027	0.018	1.477	0.07
Cottage cheese	562	0.016	0.032	0.516	0.30
Other cheese	2258	0.047	0.016	2.865	0.00
Condensed or evaporated milk	424	0.134	0.034	3.974	0.00
Ice cream and ice milk	1162	0.098	0.029	3.394	0.00
Ice cream and ice milk novelties	325	-0.019	0.061	0.314	0.38
Frozen yogurt	75	0.188	0.145	1.303	0.10
Eggs	2948	0.035	0.006	5.527	0.00
Dairy products - other	1763	0.067	0.032	2.110	0.02
Bread	5945	0.094	0.009	10.126	0.00
Unsweetened rolls and buns	3427	0.059	0.018	3.356	0.00
Crackers and crisp breads	1711	0.082	0.024	3.448	0.00
Cookies and sweet biscuits	2657	0.075	0.015	4.958	0.00
Doughnuts	368	0.058	0.047	1.249	0.11
Yeast-raised sweet goods	288	0.031	0.051	0.616	0.27
Dessert pies, cakes and other pastries	2115	0.081	0.019	4.222	0.00
Muffins	499	0.031	0.041	0.748	0.23
Other bakery products	1283	0.008	0.028	0.271	0.39
Canned pasta products	415	0.121	0.035	3.471	0.00
Dry or fresh pasta	1762	0.125	0.028	4.439	0.00
Pasta mixes	848	0.006	0.039	0.144	0.44
Rice (including mixes)	876	0.073	0.044	1.672	0.05
Flour	433	-0.017	0.056	0.303	0.38
Other grains, unmilled or milled	281	-0.104	0.083	1.252	0.11
Breakfast cereal	2381	0.043	0.015	2.761	0.00
Cake and other flour-based mixes	621	0.030	0.038	0.797	0.21
Cereal-based snack foods	1733	0.050	0.019	2.549	0.01
Other cereal products	81	0.093	0.073	1.271	0.10
Apples	2582	0.072	0.015	4.637	0.00
Bananas and plantains	4164	0.015	0.009	1.673	0.05
Grapefruit	451	0.094	0.049	1.919	0.03

Grapes	1432	0.076	0.022	3.423	0.00
Lemons and limes	585	0.060	0.030	2.002	0.02
Melons	935	0.103	0.031	3.275	0.00
Oranges and other citrus fruit	1905	0.055	0.017	3.322	0.00
Peaches and nectarines	645	0.050	0.036	1.390	0.08
Pears	881	0.048	0.027	1.770	0.04
Plums	449	0.072	0.035	2.059	0.02
Other tropical fruit	1137	0.093	0.030	3.124	0.00
Strawberries	544	0.088	0.038	2.326	0.01
Other fresh fruit	522	0.123	0.051	2.425	0.01
Frozen fruit	106	-0.021	0.085	0.246	0.40
Dried or other preserved fruit	518	0.133	0.044	3.034	0.00
Raisins	194	0.019	0.061	0.308	0.38
Other dried/preserved fruit (excluding canned)	359	0.163	0.053	3.068	0.00
Apple juice	797	0.082	0.023	3.538	0.00
Grapefruit juice	164	0.017	0.038	0.441	0.33
Orange juice	1353	0.061	0.017	3.568	0.00
Other fruit juice	1929	0.028	0.020	1.412	0.08
Orange juice	493	0.014	0.024	0.579	0.28
Other fruit juice	613	-0.023	0.028	0.829	0.20
Peaches	227	0.069	0.042	1.653	0.05
Pineapple	320	0.072	0.030	2.421	0.01
Mixed fruit	444	-0.015	0.033	0.443	0.33
Other canned fruit	438	-0.026	0.051	0.504	0.31
Jam, jelly and other preserves	635	0.078	0.028	2.806	0.00
Fruit pie fillings	115	-0.038	0.055	0.685	0.25
Unshelled nuts	382	0.082	0.061	1.345	0.09
Shelled peanuts	216	0.013	0.057	0.229	0.41
Other shelled nuts	530	0.141	0.054	2.597	0.00
Green or wax beans	455	0.113	0.041	2.721	0.00
Broccoli	1314	0.114	0.022	5.175	0.00
Cabbage	453	0.047	0.037	1.257	0.10
Carrots	2019	0.092	0.026	3.568	0.00
Cauliflower	617	0.107	0.030	3.580	0.00
Celery	1139	0.103	0.023	4.396	0.00
Corn	365	0.024	0.043	0.555	0.29
Cucumbers	1571	0.072	0.023	3.127	0.00
Lettuce	2641	0.061	0.014	4.492	0.00
Mushrooms	1561	0.051	0.016	3.171	0.00
Onions	2350	0.033	0.035	0.940	0.17
Peppers	1815	0.090	0.027	3.332	0.00
Potatoes	2214	0.072	0.026	2.762	0.00
Radishes	358	0.044	0.082	0.536	0.30
Spinach	344	0.055	0.050	1.111	0.13
Tomatoes	2902	0.082	0.020	4.193	0.00
Turnips and rutabagas	518	0.022	0.032	0.690	0.25
Other leaf and stalk vegetables	1101	0.285	0.047	6.037	0.00

Other seed and gourd vegetables	740	0.107	0.044	2.440	0.01
Other root vegetables	808	0.082	0.054	1.507	0.07
Corn	178	0.093	0.055	1.707	0.04
Peas	212	0.009	0.056	0.153	0.44
Potato products	931	0.067	0.030	2.242	0.01
Other frozen vegetables	546	0.051	0.039	1.325	0.09
Potato products - dried	139	-0.069	0.084	0.823	0.21
Other vegetables - dried	205	0.098	0.171	0.574	0.28
Green or wax beans	338	0.024	0.034	0.726	0.23
Baked beans	400	0.090	0.029	3.108	0.00
Other beans	404	0.068	0.043	1.588	0.06
Corn	648	0.028	0.026	1.107	0.13
Mushrooms and truffles	349	0.098	0.035	2.799	0.00
Peas	344	0.050	0.031	1.591	0.06
Tomatoes (including paste)	971	0.004	0.033	0.132	0.45
Other canned vegetables	416	0.124	0.057	2.186	0.01
Tomato juice	325	0.096	0.043	2.229	0.01
Other canned vegetable juice	471	0.081	0.031	2.600	0.00
Pickles (including olives)	810	0.039	0.038	1.030	0.15
Ketchup	539	-0.013	0.025	0.512	0.30
Other sauces and sauces mixes	2252	0.068	0.028	2.412	0.01
Mayonnaise and salad dressings	1444	0.001	0.022	0.058	0.48
Other condiments (including vinegar)	727	0.133	0.064	2.086	0.02
Spices	636	0.143	0.088	1.620	0.05
Sugar	967	0.026	0.027	0.950	0.17
Syrups and molasses	351	0.201	0.053	3.783	0.00
Gum	617	-0.057	0.042	1.373	0.09
Chocolate bars	1105	0.010	0.021	0.483	0.31
Other chocolate confections	649	0.075	0.043	1.761	0.04
Sugar candy	1354	0.017	0.035	0.492	0.31
Other sugar confections	1443	0.084	0.030	2.790	0.00
Other sugar preparations	255	0.109	0.064	1.698	0.05
Roasted or ground coffee	707	0.107	0.041	2.617	0.00
Other coffee	749	-0.063	0.036	1.754	0.04
Tea	711	0.044	0.058	0.769	0.22
Margarine	1607	0.041	0.020	2.022	0.02
Shortening	117	0.019	0.032	0.585	0.28
Lard	84	-0.135	0.081	1.669	0.05
Cooking/salad oil	617	0.081	0.077	1.056	0.15
Canned soup	2109	0.077	0.019	4.045	0.00
Dried soup	858	0.116	0.052	2.218	0.01
Canned infant or junior foods	138	0.049	0.046	1.078	0.14
Infant cereals and biscuits	59	0.023	0.113	0.201	0.42
Infant formula	84	-0.270	0.163	1.655	0.05
Pre-cooked frozen dinners	689	0.074	0.027	2.735	0.00
Dessert pies, cakes, other pastries	584	0.087	0.030	2.882	0.00
Frozen meat or poultry pies	140	0.031	0.094	0.331	0.37

Other pre-cooked food preparations	1601	0.046	0.021	2.161	0.02
Materials for food preparations	704	0.062	0.076	0.809	0.21
Honey	211	0.019	0.041	0.466	0.32
Peanut butter	573	0.049	0.029	1.684	0.05
Dairy product substitutes	550	-0.062	0.058	1.072	0.14
Flavouring extracts and essences	62	0.226	0.289	0.781	0.22
Flavouring powders and crystals	354	0.101	0.133	0.760	0.22
Food seasonings (including salt)	764	-0.030	0.099	0.300	0.38
Jelly powders	216	0.158	0.082	1.937	0.03
Prepared dessert powders	227	-0.087	0.082	1.065	0.14
Potato chips and similar products	1983	0.037	0.017	2.156	0.02
Food drink powders	233	0.190	0.103	1.847	0.03
Canned puddings and custards	546	0.046	0.035	1.316	0.09
All other food preparations	2567	0.069	0.023	2.964	0.00
Carbonated beverages	3533	-0.003	0.017	0.175	0.43
Fruit drinks	938	-0.001	0.030	0.020	0.49
Other non-alcoholic beverages	1186	-0.039	0.043	0.894	0.19