Zero Expenditures and Engel Curve Estimation

Timothy K.M. Beatty
Canada Research Chair
Food and Resource Economics Group
University of British Columbia
timothy.beatty@ubc.ca

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 23-26, 2006

Zero Expenditures and Engel Curve Estimation

Timothy K.M. Beatty[1]
Canada Research Chair
Food and Resource Economics Group
University of British Columbia
timothy.beatty@ubc.ca

*** DRAFT ***

PRELIMINARY AND INCOMPLETE

PLEASE DO NOT CITE WITHOUT EXPLICIT PERMISSION

Abstract
Zeroes expenditure represent a difficulty in the analysis of household survey data. Zero expenditures are the result of two phenomena: nonconsumption and infrequency of purchase. Distinguishing between these types of zeroes is difficult in the kinds of data of interest to agricultural economists. This paper proposes a novel approach that yields less biased estimates of latent expenditure when the cause of the zeroes is unknown.

Keywords
Zero Expenditure, Nonconsumption, Infrequency of Purchase, Quantile Regression.

Introduction
In many household expenditure surveys, respondents report zero expenditure on some commodities. These zero expenditures represent an important challenge for the analysis of household expenditure data. The difficulty arises because the factors that cause zero expenditures, have important implications for consistently recovering demand relationships. This paper explores the use of censored quantile regression to recover demand relationships without making strong (possibly incorrect) identifying assumptions concerning the nature of the zeros.

There are two frequently cited reasons for a household reporting no expenditure on a good, or in a category:

- Nonconsumption.
- Infrequency of purchase.

[1] I would like to thank the Canada Research Chair Program and the SSHRC Standard Research Grant program for financial support. I would like to thank Craig Mcintosh and Jay Shimshack for helpful comments. All remaining errors are my own.

Each of these explanations has very different implications for consistently estimating an Engel curve. What are the implications for making the wrong identifying assumption about their cause? Indeed, is there an approach that is robust to assumptions concerning the nature of the zeroes?

Zero expenditure on certain goods may be the utility maximizing solution to a household's choice problem. For example, there may be no set of relative prices that will induce a household to purchase tobacco products. In terms of predicting response to policy changes, it seems reasonable to expect that households, whose optimal level of expenditure for a given commodity is zero, are unlikely to change their behavior much for small changes in relative prices.

Infrequency of purchase occurs when the survey period is not long enough to capture expenditures on goods that a household is most likely purchasing. Many expenditure surveys consist of a limited period wherein a household may report no expenditure on a commodity that they have previously purchased, e.g. clothing expenditure may be zero but presumably this is not true in the long run. Problems associated with infrequency of purchase are not limited to zeroes. Households may begin the survey period with a large stock of a given commodity and as a result may only be observed purchasing a small quantity. In contrast to the corner solution zeroes described above, changes in relative prices will result in changes in expenditure amongst those households with zero recorded expenditure.

As the unit of observation, over time and commodities, becomes finer, it is progressively more difficult to distinguish between types of zeroes. This is likely to be particularly troublesome for the commodities of central interest to Agricultural Economists. For example, does observing a household with no meat purchases mean the household never purchases meat or simply did not purchase any during the sample period? In a large enough sample, both kinds of zeroes are likely to be present.

As the following quote from Meghir and Robin (1992) makes clear, from the econometrician's point of view, the nature of the observed zeroes is essentially a maintained hypothesis.

Meghir and Robin (1992)

> **On the type of data usually available in surveys, we believe that it is not possible to identify the nature of the observed zeros without prior information. Ultimately, the assumption that they are the result of nonconsumption or of infrequency is an identifying assumption.**

Beyond underscoring the importance of the identification assumption in dealing with zero expenditures, the main contribution of this paper is to apply a result from Powell (1986) concerning the consistency of censored regression quantile functions to the zeroes problem described above. I show that median demand is resistant to different types of zeroes where mean demand is not. In addition, I develop a novel means of estimating semiparametric quantile Engel curves. A number of other advantages to estimating conditional quantile functions are also explored.

The paper is structured as follows. We begin by examining the consequences of making the incorrect identifying assumption, i.e. assuming that corner solutions are in effect the result of infrequency of purchases. Subsequently, we propose a semiparametric means to consistently recover quantile Engel curves. These are compared to estimated Engel curves when the zero expenditures are either excluded or included in the model.

Zeroes

The fundamental question is how to relate observed expenditure on a commodity of interest to latent expenditure[2]. Let $e_{i,j}$ and $e^{*}_{i,j}$ denote the observed and latent expenditure of household *i* on good *j*.

Nonconsumption Zeroes

In the case of nonconsumption zeroes, the identifying assumption is that demand is censored from below at zero. In utility theoretic terms, this corresponds to a corner solution to the individual's utility maximization problem. Thus observed expenditure of household *i* on good *j*, $e_{i,j}$ is related to latent expenditure in the following way:

$$e_{i,j} = \max\left\{0, e^{*}_{i,j}\right\}. \tag{1}$$

Deaton and Irish (1984) were amongst the first to use generalizations of the tobit model to analyze the demand for commodities where nonconsumption seems a reasonable assumption, e.g. alcohol and tobacco. Heien and Wessells (1990), Shonkwiler and Yen (1999) and Perali and Chavas (2000) use similar approaches to model the demand for food. Note that if zeroes are excluded, we can consistently estimate latent expenditure conditional on positive expenditure. That is,

$$E\left[e_{i,j} \mid e_{i,j} > 0\right] = E\left[e^{*}_{i,j} \mid e_{i,j} > 0\right]. \tag{2}$$

Infrequency of purchase zeroes

If the causal explanation for observed zero expenditures is purchase infrequency then latent expenditures can be linked to observed expenditure in expectation.

$$E\left[e_{i,j}\right] = E\left[e^{*}_{i,j}\right] \tag{3}$$

The identifying assumption here is that if the observation period were long enough, latent and observed expenditures coincide. Let $P_{i,j}$ denote the probability of household *i* purchasing good *j* during the survey period. In this manner, observed expenditure can be decomposed:

---

[2] Note that this differs from some of the previous literature in that the latent variable is latent expenditure, not latent consumption. Intuitively, this can be thought of as an average of long-run expenditure.

$$E\left[e_{i,j}\right] = E\left[e_{i,j} \mid e_{i,j} > 0\right]P_{i,j} + E\left[e_{i,j} \mid e_{i,j} \leq 0\right]\left(1 - P_{i,j}\right). \quad (4)$$

Thus latent expenditure can be related to observed expenditure:

$$E\left[e_{i,j} \mid e_{i,j} > 0\right]P_{i,j} = E\left[e^{*}_{i,j}\right] \quad (5)$$

Equation (5) shows how latent expenditure can be recovered from observed expenditure when zeroes are the result of an infrequency of purchase situation, using the proportion of zero expenditures in the sample as an estimate of *P*. For example, if in a weeklong survey households purchase good *i* every other week, latent expenditure will be one half of observed expenditure. On important consequence of this model is that observations on households with zero expenditure are necessary to scale positive observed expenditure in order to recover the latent variable $e^{*}_{i,j}$.

The literature on infrequency of purchase, see for example Keen (1986) and Meghir and Robin (1992), models the situation where observed expenditure differs from latent expenditure when households are only observed for a limited period of time. If the household enters the period with a sufficiently large positive stock of the good, reported expenditure will be zero whereas latent expenditure will be positive. Casual empiricism suggests that a simple dichotomous purchase/no purchase model may mask the full richness of the infrequency of purchase problem. Infrequency of purchase results from durability of goods, fixed costs associated with shopping and the limited nature of the survey period. If a household enters the survey period with some positive stock of the good, or exits the survey period with positive stock of the good, observed expenditure will differ from latent expenditure, but not in the simple purchase/no purchase dichotomy that has been assumed in the literature.

Consequences of the identification assumption

What bias is introduced when a nonconsumption zero is incorrectly assumed to be an infrequency of purchase zero? As noted above, we can consistently recover estimates of latent consumption by scaling up observed consumption by the proportion of zeroes. Denote $\tilde{P} > P$ the incorrectly estimate of the proportion of infrequency of purchase zeroes. Then

$$E\left[e_{i,j} \mid e_{i,j} > 0\right]\tilde{P} \geq E\left[e^{*}_{i,j}\right]. \quad (6)$$

Thus, under the false assumption of infrequency of purchase, we will inappropriately scale up expenditure, yielding a biased overestimate of latent expenditure.

What occurs when an infrequency of purchase zero is incorrectly assumed to be a nonconsumption zero. If the econometrician omits the zero in an effort to compute the latent expenditure conditional on consumption, this will yield a biased underestimate of the latent expenditure conditional on consumption.

How will this affect the relationship between income and expenditure on the commodity of interest? For normal goods assuming infrequency of purchase will yield larger estimates of responsiveness than assuming nonconsumption. Intuitively, this results from the fact that if zeroes are the result of purchase infrequency, the entire sample is assumed to be responding to changes in income. In the case of nonconsumption, the impact of a change in income will largely confined to those who consume the good.

An Alternative Approach: Censored Quantile Regression

Above the difficulty of identifying the nature of the observed zeros without prior information and the attendant biases associated with making the wrong identifying assumption were made clear. Rather than base inferences on mean expenditure, we propose the use of quantile regressions, which are robust to the nature of the zeroes, to consistently recover the parameters of interest.

Quantiles are order statistics which divide a sample of observations on, say budget shares, into two groups: the $(1-\tau)^{th}$ proportion of the sample which has a larger budget share and the τ^{th} proportion of the sample which has a smaller budget share. Let $Q(e_j \mid \tau)$ be the τ^{th} quantile of expenditure on good *j*. If $\tau = 0.5$, $Q(e_j \mid 0.5)$ will be the median level of observed expenditure.

The intuition here is fairly straightforward. Quantiles, at which observed expenditure is positive, are not affected by truncation at zero in the same way as the mean. Assuming median expenditure is positive, the median is not affected by truncation at zero. Because of this, latent median expenditure will be equal to observed median expenditure, if the median latent expenditure is positive and zero otherwise.

In addition, infrequency of purchase zeroes will not affect estimates of median latent expenditure. In a similar manner, if the sample period were long enough, latent expenditure and observed expenditure would coincide. If this is the case, it follows immediately, that median latent expenditure is equal to median expenditure, $Q(e_j \mid 0.5) = Q(e_j^* \mid 0.5)$.

Quantile regressions have a number of other attributes that make them well suited to the analysis of household survey data. Davison (2003) describes quantiles as "resistant" statistics. That is, they are robust to outliers and to contamination. When working with large-scale household survey data this resistance is a major advantage (Deaton (1997)) as outliers and contamination seem to be the rule rather than the exception. Buchinsky (1998) and Chay and Powell (2001) provide examples of other applications in economics.

Estimation Approach

Quantile Regression

Quantile regression has proved extremely useful in the estimation of Engel curves as illustrated by Koenker and Hallock (2001) and Deaton (1997). In his estimates of Engel curves in Pakistan, Deaton finds differences in slopes for different regression quantiles. However, the functional forms imposed by these analyses may fail to capture more important differences between quantiles. Quantile regressions, as described proposed by Koenker and Bassett (1978), provide a means of estimating conditional quantile functions.

Powell (1984) shows that least absolute deviation regression can be used to obtain consistent estimates of the parameters of interest in the presence of censoring. This result was later extended to arbitrary quantiles by Powell (1986) and to arbitrary censoring by Honoré et al. (2002). In terms of **Error! Reference source not found.**, the estimation problem can be written,

$$Q_n\left(g(\cdot)\mid\tau\right)=\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(w_i-\max\left\{0,g_\tau\left(x_i\right)\right\}\right), \qquad (7)$$

where ρ_τ is the check function for the τ th quantile.

One of the shortcomings in previous applications of quantile regressions to expenditure data is that most studies impose linearity in the conditional quantile function. This may not be appropriate for estimating Engel curves. In a well known paper, Banks et al. (1997) find that models that do not allow sufficient curvature in the Engel curves will not accurately describe observed behavior. In order to relax the assumption of linearity, we will employ nonparametric methods and estimate the unknown function $g()$.

Regression Splines

Regression splines offer a simple way of estimating nonparametric and semiparametric models. Regression splines augment the basis functions of ordinary least squares regression models, often through the use of functions of the form $(x-\kappa_j)_+^p$, (the function is zero if log expenditure is below the knot point, $x-\kappa_j<0$, and $(x-\kappa_j)^p$ otherwise). Here κ_j refers to the j^{th} knot and *p* is the degree of the basis. In practice the augmented basis is chosen to be linear or quadratic (*p=1* or *p=2*). Much of the literature on regression splines has focused on algorithms for choosing the number and location of the knots.

An alternative, penalized regression splines (p-splines) was developed by Ruppert and Carroll (1997) [3]. Computationally, the p-spline approach simplify the process of fitting regression splines by choosing a (relatively) small number of knots, *K*, and shrinking the jumps at each knot towards zero through the use of a penalty function. P-splines solve a minimization problem of the form:

$$\min_{\beta,\gamma}\sum_{i=1}^{N}\left(y_i-\beta_0-\beta_1x_i-\beta_1x_i^2-\sum_{j=1}^{K}\gamma_j(x_i-\kappa_j)_+^2\right)^2, \qquad (8)$$

subject to a smoothness constraint of the form $\sum_{j=1}^{K}\gamma_j^2\le C$, where the choice of *C* will determine the smoothness of the fit for the variable or variables chosen to enter nonparametrically. Ruppert (2002) provides an algorithm for the choice and location of the knots in p-spline models.

[3] Eilers and Marx (1996) proposed a penalized regression model that uses a B-Spline extended basis.

I now describe penalized quantile regression splines and their estimation. I illustrate the insights that can be gained from this model by applying it to expenditure data for eight different commodities.

Penalized Quantile Regression Splines

To combine the resistance of censored quantile regression with the tractability of p-splines, I extend the basis functions in a manner similar to p-splines, but fit the extended basis function using a censored quantile regression, subject to a constraint on the magnitude of the parameters of the extended basis function. We term these penalized quantile regression splines (pq-splines). PQ-Splines extend quantile regression techniques, in a straightforward manner, allowing a subset of the explanatory variables to be modeled nonparametrically [4].

The resulting minimization problem can be written:

$$\min_{\beta,\gamma} \sum_{i=1}^{N} \rho_\tau \left(\max\left(0, y_i - \beta_0 - \beta_1 x_i - \beta_1 x_i^2 - \sum_{j=1}^{K} \gamma_j (x_i - \kappa_j)_+^2 \right) \right) \tag{9}$$

again subject to a roughness penalty, in this case of the form $\sum_{j=1}^{K} \left|\gamma_j\right| \leq C$. Using a roughness penalty of this form preserves the linear structure that makes computing estimates straightforward. Note that, if desired, additional economic restrictions on the slope and curvature can be imposed and tested in terms of constraints on the parameters β and γ.

Because (9) is written in terms of $p+K+1$ parameters, calculating the derivative is relatively easy compared to other nonparametric approaches. This makes penalized quantile regression splines particularly well suited to the analysis of consumer behavior where elasticities are often the object of interest. For the Working-Leser specification proposed above the τ^{th} quantile expenditure elasticity can be written:

$$\eta_k^\tau = 1 + f_k'(x)/w_k^\tau, \tag{10}$$

where for $p = 2$,

$$f_k'(x) = \hat{\beta}_1^\tau + 2\hat{\beta}_2^\tau x + \sum_{k=1}^{K} 2\hat{\gamma}_k^\tau \left(x - \kappa_j\right)_+ . \tag{11}$$

Note that these elasticities will not be well defined in the region where $w^\tau \approx 0$.

Implementation

Because (9) can be written as a linear programming problem estimating a pq-spline model is computationally straightforward. I rewrite (9) as:

[4] Note this differs from the piecewise linear Quantile Smoothing Splines proposed by Koenker et al. (1994). For a discussion on the differences between smoothing and regression splines see Ruppert et al. (2003).

$$\min_{\beta,\gamma} \sum_{i=1}^{N} \tau u_i + (1-\tau) v_t \quad \text{s.t.}$$
$$y_i = \beta_0 + \beta_1 x_i + \beta_1 x_i^2 + \sum_{j=1}^{K} \gamma_j (x_i - \kappa_j)_+^2 + u_i - v \quad \forall i$$
$$u_i \geq 0 \quad \forall i \qquad (12)$$
$$v_i \geq 0 \quad \forall i$$
$$\sum_{j=1}^{K} \left|\gamma_j\right| \leq C$$
where u_i and v_i are respectively the positive and negative residuals for observation *i*.

The degree of smoothing, in this case controlled by the value of *C*, is a key component of any nonparametric analysis. Here I exploit a result a key result of Koenker et al. (1994) concerning the knots. Because the roughness penalty can be written as a linear constraint, a finite number of knots will be active at the solution (in other words, at the basic solution some of the $\hat{\gamma}_j$ will be exactly zero). Denote the number of nonzero knots p_C. Koenker et al. (1994) argue that this is a plausible measure of the degrees of freedom of the fit and I adopt it here. Given an estimate of degrees of freedom, one can implement a number of data dependant criteria for selecting *C* (GCV, AIC, SIC, Mallows Cp). Following Machado (1993) and Koenker et al. (1994), I use the SIC[5],

$$\text{SIC}(p_C) = \log\left[N^{-1} \sum_{i=1}^{N} \rho_\tau \left(y_i - \beta_0 - \beta_1 x_i - \beta_1 x_i^2 - \sum_{j=1}^{K} \hat{\gamma}_j (x_i - \kappa_j)_+^p \right)\right] + \frac{1}{2} N^{-1} p_C \log N \text{ .} \qquad (13)$$

We search over a 40-point grid equally spaced over $\log_{10}(-4)\ldots\log_{10}(4)$ and choose the value of *C* that minimizes (13).

Several choices for *K*, the number of knots, were investigated. Choosing $K > 10$ did not significantly change the fit. Thus, for the purposes of what follows, *K* is set equal to 10. Results were robust to different choices of *K*. Following the recommendation of Ruppert et al. (2003) knots are placed according to:

$$\kappa_k = \left(\frac{k+1}{K+2}\right) \text{th sample quantile of the unique } \mathrm{x_i}. \qquad (14)$$

The model was implemented using the NuOpt add-on to S-Plus. In all cases convergence occurred extremely quickly (typically under 4 seconds on a 2.2 Ghz AMD Opteron processor). NuOpt uses an interior point algorithm based on the higher-order correction model proposed by Mehotra (1992) and Gondozio (1994).

Inference

Inference for the semi parametric model developed in this paper is somewhat difficult. We use a bootstrap procedure to circumvent these difficulties. Inference is based upon 2500 nonparametric bootstrap replication. Hahn (1998) shows that the bootstrap

[5] Alternative goodness of fit criteria were explored and yielded broadly similar results.

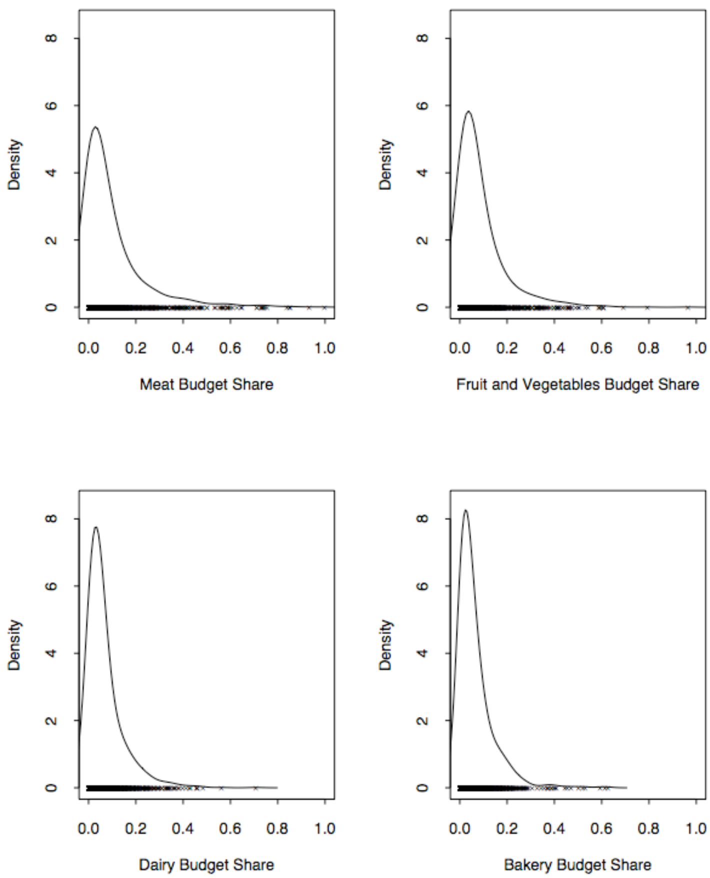
procedure is applicable for censored regression quantiles. Using Monte Carlo experiments, Buchinsky (1995) shows that this procedure performs relatively well.

Data

The data for our analysis is drawn from the 1996 Family Food Expenditure Survey (FOODEX). The Family Food Expenditure survey is a two-week survey conducted by Statistics Canada in which respondents keep a diary of all food expenditures. Over the course of a twelve-month period, slightly more than ten thousand households were surveyed. For the purposes of this paper, to avoid problems with heterogeneiy, we select a relatively homogeneous subsample of households: single person households where the head of household is under 65. focus on four commodities Meat, Fruits & Vegetables, Dairy and Bakery goods.

Figure 1 plots the distribution of the food share expenditures considered in this paper. The key feature of interest is the large number of households reporting zero expenditure for the commodities of interest. Again, are the households in the sample who fail to purchase meat during the survey period vegetarians, or do they simply own a chest freezer? Without further information, this is unknown and unknowable to the econometrician.

Figure 1 Data Summary



Results

Figures 2-5 and Table 2 summarize the main results of this paper. The first four figures show the results of estimating the censored quantile regression model described above to the FOODEX sample. Table 2 reports expenditure elastiticities and confidence intervals at the median of log-expenditure.

Engel Curves

Figures 2-5 show the estimated quantile Engel curves for the 4 commodities of interest: Meat, Fruits and Vegetables, Dairy and Bakery goods. A 90% pointwise confidence interval is also reported. For comparative purposes, we fit a pq spline model to the mean of the data where the zeroes have been included, as would be correct under a maintained hypothesis of infrequency of purchase zeroes. In addition, we fit a model to the mean of the positive observations, which yields a consistent estimate of positive latent expenditure under a maintained hypothesis of nonconsumption.

Note the Engel curves computed using the mean spline fits lie outside the pointwise 90% confidence interval for the median Engel curve over the bulk of the distribution of the log income. The 90% confidence interval is larger where the data is relatively sparse and smaller where the data is relatively dense.

In all cases, the median Engel curve is everywhere below the Engel curve estimated using the mean for all observations. Over most of the distribution of income, the two curves are significantly different from one another. This is with the theory developed earlier. The mean Engel curve estimated over all data is a consistent estimator of latent expenditure under the maintained assumption that all zeroes are due to infrequency of purchase, which is almost certainly not the case.

As one would expect the mean Engel curves computed for positive observations lies everywhere below the quantile Engel curves computed for all observations. One provides an unbiased measure of positive latent expenditure conditional on the assumption that zeroes are due to nonconsuption. This is compared to the quantile Engel curve which is an unconditional measure of latent expenditure.

One important note is that the median Engel curves are more resistant to outliers than the those which fit to the mean. Given that the distributions of the budget shares are right skewed as evidenced by figure 1, it should be generally the case that the median will be smaller than the mean.

Figure 2. Meat

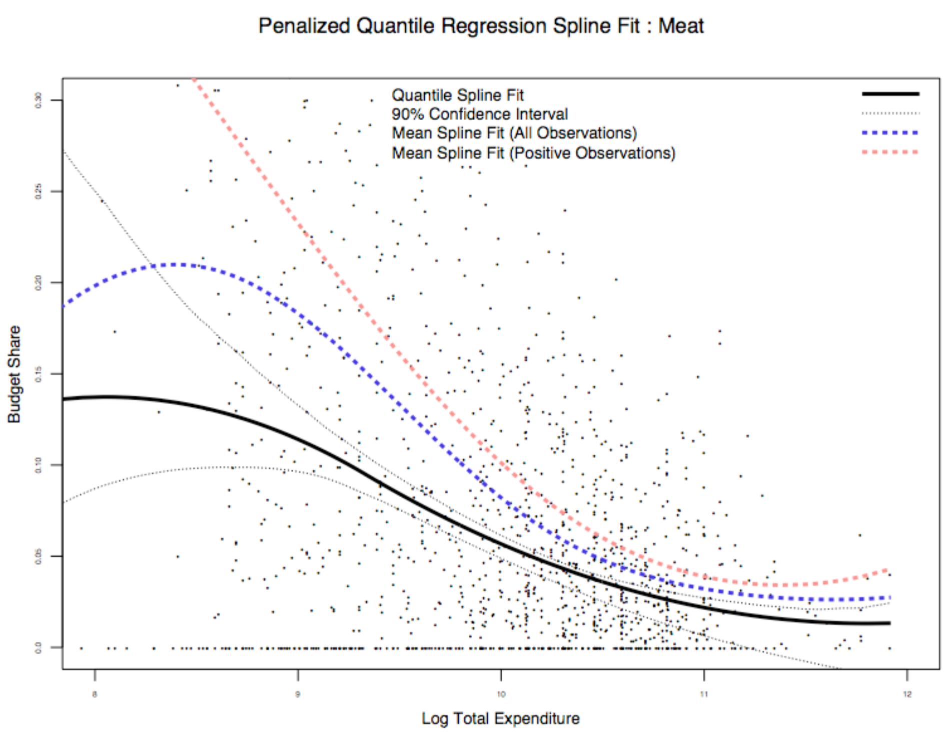


Figure 3. Fruits & Vegetables

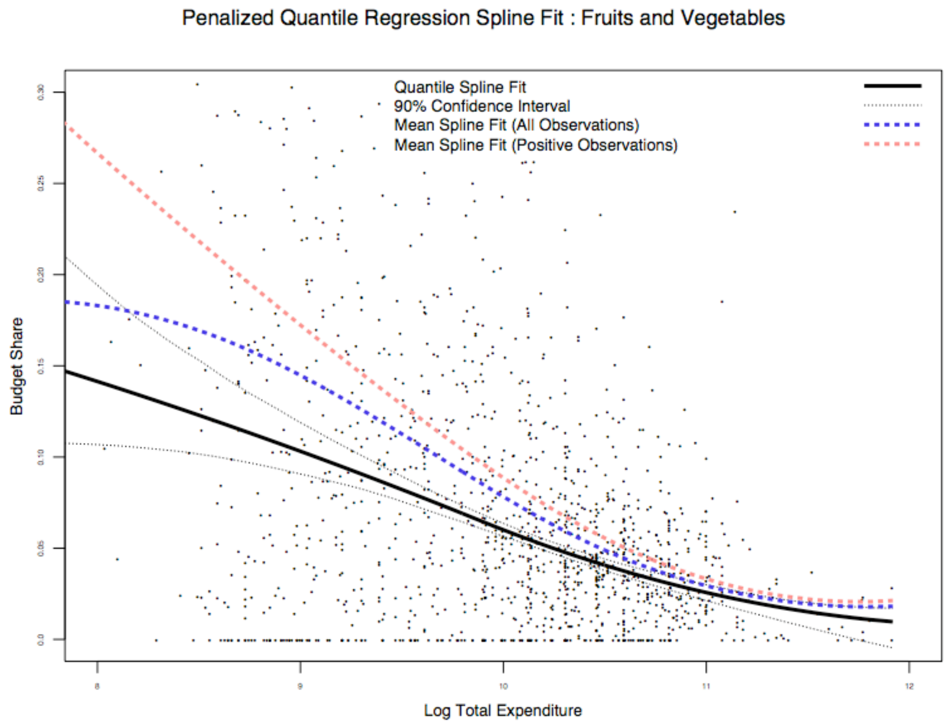


Figure 4. Dairy

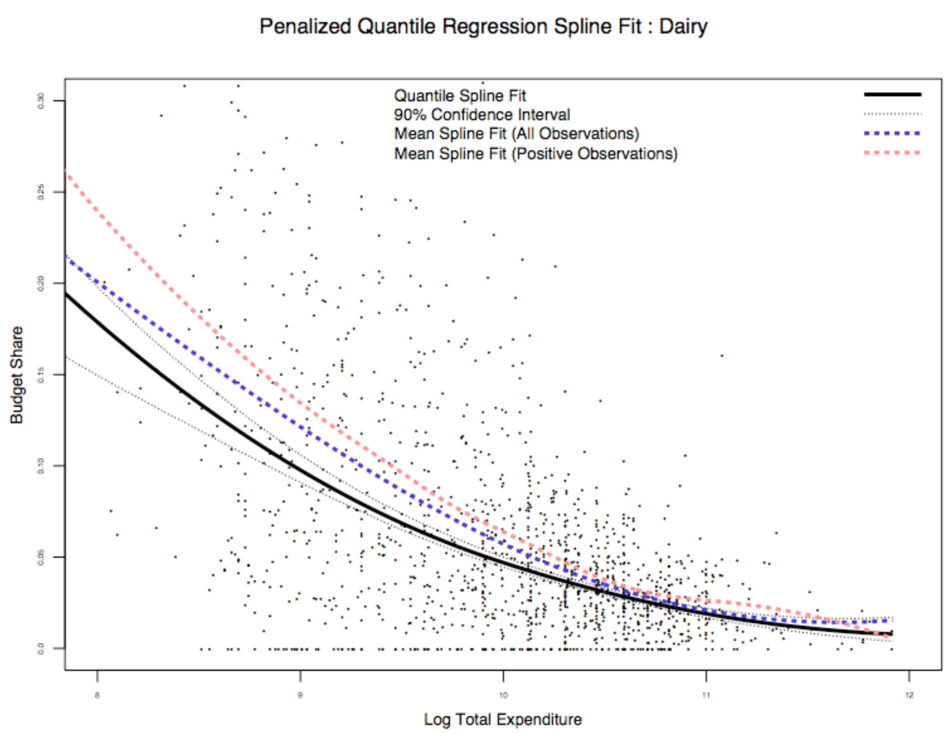
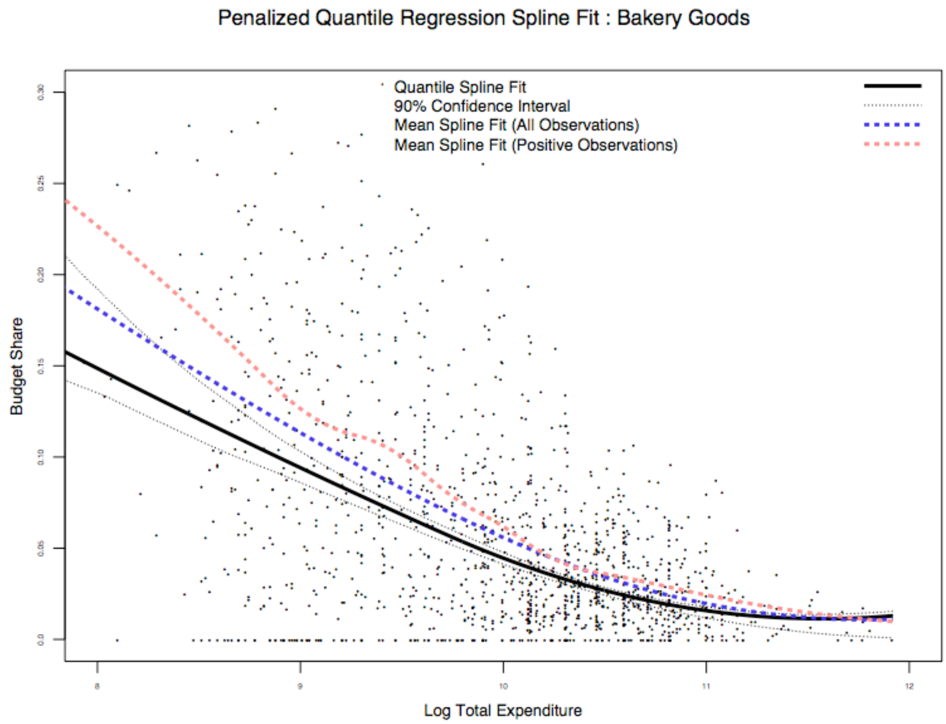


Figure 5. Bakery Goods



Conclusions & Shortcomings

This paper represents a first attempt to develop a more robust means recovering a less biased measure of latent expenditure in the presence of zeroes of unknown cause. Because this research represents a first attempt, we chose to estimate a simple Engel curve rather for a homogeneous population subgroup. An obvious extension is to estimate a more complex demand system.

REFERENCES:

Banks, J., R. W. Blundell and A. Lewbel (1997). "Quadratic Engel Curves and Consumer Demand." The Review of Economics and Statistics 79(4): 527-539.

Blundell, R., A. Duncan and K. Pendakur (1998). "Semiparametric Estimation and Consumer Demand." Journal of Applied Econometrics 13: 435-461.

Buchinsky, M. (1995). "Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study." Journal of Econometrics 68: 303-338.

Buchinsky, M. (1998). "Recent Advances in Quantile Regression Methods." Journal of Human Resources 33(1): 88-126.

Chay, K. L. and J. L. Powell (2001). "Semiparametric Censored Regression Models." Journal of Economic Perspectives 15(4): 29-42.

Davison, A. (2003). Statistical Models. Cambridge, Cambridge University Press.

Deaton, A. (1997). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Baltimore, Johns Hopkins.

Deaton, A. and M. Irish (1984). "Statistical Models for Zero Expenditures in Budget Shares." Journal of Public Economics 23: 59-80.

Eilers, P. H. C. and B. D. Marx (1996). "Flexible Smoothing with B-splines and Penalties." Statistical Science 11(2): 89-121.
Gondozio, J. (1994). Multiple Centrality Corrections in a Primal-Dual Method for Linear Programming. Technical Report. Geneva, H.E.C.
Hahn, J. (1998). "Bootstrapping quantile regression estimators." Econometric Theory 11: 105-121.
Heien, D. and C. R. Wessells (1990). "Demand Systems Estimation with Microdata: A Censored Regression Approach." Journal of Business Economics and Statistics 8(3): 365-371.
Honoré, B., S. Khan and J. L. Powell (2002). "Quantile regression under random censoring." Journal of Econometrics 109.
Keen, M. (1986). "Zero Expenditures and the Estimation of Engel Curves." Journal of Applied Econometrics 1(3): 277-286.
Koenker, R. and G. Bassett (1978). "Regression Quantiles." Econometrica 46: 33-50.
Koenker, R. and K. Hallock (2001). "Quantile Regression." Journal of Economic Perspectives 15(4): 143-156.
Koenker, R., P. Ng and S. Portnoy (1994). "Quantile Smoothing Splines." Biometrika 81(4): 673-680.
LaFrance, J. T. (1991). "When is Expenditure Exogenous in Separable Demand Models." Western Journal of Agricultural Economics 16: 1991.
Machado, J. A. F. (1993). "Robust Model Selection and and M-estimation." Econometric Theory 9: 478-493.
Meghir, C. and J.-M. Robin (1992). "Frequency of Purchase and the estimation of demand systems." Journal of Econometrics 53: 53-85.
Mehotra, S. (1992). "On the implementation of a Primal-Dual Interior Point Method." SIAM Journal of Optimization 2: 575-601.
Perali, F. and J.-P. Chavas (2000). "Estimation of Censored Demand Equations from Large Cross-Section Data." American Journal of Agricultural Economics 82(4): 1022-1037.
Powell, J. L. (1984). "Least absolute deviations for the censored regression model." Journal of Econometrics 25: 303-325.
Powell, J. L. (1986). "Censored Regression Quantiles." Journal of Econometrics 32: 143-155.
Ruppert, D. (2002). "Selecting the number of knots for penalized splines." Journal of Computational and Graphical Statistics 11: 735-757.
Ruppert, D. and R. J. Carroll (1997). Penalized Regression Splines. Working Paper: 30.
Ruppert, D., M. P. Wand and R. J. Carroll (2003). Semiparametric Regression. Cambridge, Cambridge University Press.
Shonkwiler, J. S. and S. T. Yen (1999). "Two-Step Estimation of a Censored System of Equations." American Journal of Agricultural Economics 81: 972-982.