Censored Quantile Regression and Purchases of Ice Cream

Geir Wæhler Gustavsen[1], Dean Jolliffe[2] and Kyrre Rickertsen[3]

[1] Senior Research Economist, Norwegian Agricultural Economics Research Institute, P.O. Box 8024 Dep., 0030 Oslo, Norway (phone: +47 2236 7228; fax: +47 2236 7299; e-mail: geir.gustavsen@nilf.no)

[2] Senior Research Economist, Economic Research Service, United States Department of Agriculture. 1800 M Street. Washington DC, 20036, USA. (phone: +1 202 694 5430; fax: +1 694 5663; e-mail: jolliffe@ers.usda.gov)

[3] Professor, Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Aas, Norway (phone: +47 6496 5679; fax: +47 6496 5701; e-mail: kyrre.rickertsen@umb.no) and Senior Research Economist, Norwegian Agricultural Economics Research Institute

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, Florida, July 27-29, 2008

Censored Quantile Regression and Purchases of Ice Cream

The effects on purchases of ice cream of increasing the value added tax (VAT) for less healthy foods and removing the VAT for healthy foods are estimated. The effects on high- and low-purchasing households are estimated by using quantile regressions. Many households did not purchase ice cream and censored quantile regressions are estimated by a recently developed algorithm, which is simple, robust, and performs well near the censoring point. High-purchasing households will reduce their annual per capita purchases with 1.8 kilograms corresponding to an annual reduction of more than half a kilogram of body weight.

Keywords: Censored quantile regression, ice cream, obesity, purchase, taxes.

According to the World Health Organization (2005), the Norwegian obesity rate is low compared with most other European countries. However, it is increasing, and about 18% of 40-year old people have a body mass index (BMI) above 30 and are therefore defined as obese. Common health problems associated with obesity include type 2 diabetes, cardiac diseases, stroke, certain types of cancer, and muscular and joint diseases (Departementene 2007: 8, 108).

Contributing to increased obesity is a high intake of sugar and fat. The World Health Organization (2003) recommends that no more than 10% of a person's energy intake should come from sugar, and less than 10% from saturated fat. Ice cream is high in both sugar and saturated fat, and contributes in a large amount to both calories and fat intake. For example, among US adults, ice cream is ranked fifth among foods that contribute to saturated fat intake

(Cotton et al. 2004) and among Danish children, ice cream is ranked fourth among foods that contribute to sugar intake (Danish Academy of Technical Sciences 2007: 45).

From a public health perspective, the consumption of sugar and fat should be reduced, and a tax on ice cream can potentially reduce this intake. A tax can be implemented by increasing the value added tax (VAT). The current VAT rate for food is 14% while it is 25% for most nonfood items. We will investigate the effects on purchases of ice cream of jointly increasing the current VAT for ice cream and other less healthy food from 14% to 25% and removing the VAT for healthy food.

The risks of obesity are highest in households having a high intake of ice cream. Hence, the distribution of consumption across households is at least as important as the mean consumption, and the effects of a VAT increase on households with different levels of ice cream purchases are investigated. Our sample has 20,550 observations on household purchases from 1986 to 2001, and the distribution of purchases is shown in Table 1. Our data excludes purchases away from home. The table shows the average percentage of households reporting zero purchases of ice cream in each two-week survey period, the distribution of annual per capita purchases in kilograms, and the reported annual mean purchases.[1]

Quantiles are points on the cumulative distribution function of a random variable. If we calculate the per capita purchases of every household in the sample and sort the households according to their purchases, these unconditional quantiles divide the data into subsets. A household that purchases at the θ quantile buys more per capita than does the proportion, θ, of households and less than the proportion $(1 - \theta)$. Thus, the 0.70 quantile of purchases indicates that 70% of households buy less than this amount and 30% buy more. The 0.50 quantile column presents median purchases.

In 2001, 52% of the households did not purchase ice cream during the survey period. Furthermore, 60% of the households purchased less than 0.7 kilograms, 90% of the households purchased less than 13 kilograms, and mean purchase was 3.7 kilograms. Since many households did not purchase ice cream, the data are censored. Tobit models are frequently used to correct for censoring, and as a benchmark, we estimate the mean effects by using a Tobit model.[2] However, because the marginal effects are likely to differ at the lower and higher conditional quantiles as compared to the conditional mean, Tobit estimates may be inaccurate for low- and high-consuming households. Furthermore, unlike the Tobit estimator, the censored quantile regression (CQR) estimator is a consistent estimator in the presence of heteroskedasticity or nonnormally distributed errors (Powell 1986).

The CQR estimator is usually estimated either by using an algorithm proposed by Buchinsky (1994) or an algorithm proposed by Fitzenberger (1997). But, as Fitzenberger (1997) noted, these algorithms perform poorly when a large proportion of the data is censored (as is the case with our data on ice cream expenditures). To overcome this problem, we use a recently developed three-step algorithm proposed by Chernozhukov and Hong (2002), which is simple, robust, and performs well near the censoring point.

This paper contributes to the literature in three ways. First, studies of food consumption using CQRs include Stewart, Blisard, and Jolliffe (2003); and Gustavsen and Rickertsen (2006), however, none of these studies used the improved algorithm of Chernozhukov and Hong (2002). Second, ice cream purchases have, to our knowledge, not previously been studied and the results are of intrinsic interest. Third, the estimation results offer policy relevant results on the likely effects on purchases of subsidizing healthy and taxing less healthy food.

Table 1 about here

Empirical Model

We estimate the equation[3]

$$(1)\qquad \begin{aligned} Q^h &= \beta_0 + \sum_{i=1}^{4}\beta_i \ln\left(\frac{P_{it}}{P_{5t}}\right) + \beta_5 \ln\left(\frac{EXP^h}{P_{5t}}\right) + \beta_6 \left\{\ln\left(\frac{EXP^h}{P_{5t}}\right)\right\}^2 \\ &+ \beta_7 \ln AGE^h + \beta_8 \ln T_t + \beta_9 \ln Temp_t + \sum_{j=1}^{J}\gamma_j D_j^h + \varepsilon^h \end{aligned}$$

where Q^h is household h's per capita purchases of ice cream; P_{it} is the price of good i in survey period t (good 1 is ice cream, good 2 is candy, good 3 is potato chips, good 4 is fresh fruits, and good 5 is other non-durables and services); EXP^h is total per capita expenditure on non-durables and services; AGE^h is the age of the head of the household;[4] T_t is an annual trend variable taking the value of 1 in 1986 and 16 in 2001; $Temp_t$ is temperature; D_j^h are dummy variables representing region, season, and household type; and ε^h is an error term. Further description of the variables is provided in Table 2, and the data are discussed in more detail below. The total expenditure elasticity for the θ quantile is calculated as

$$(2)\qquad E_\theta = \frac{1}{\hat{Q}}(\hat{\beta}_5 + 2\hat{\beta}_6\overline{EXP})\cdot \Pr(Q>0)$$

where $\overline{EXP}$ is the mean of the variable $\ln\left(EXP^h/P_{5t}\right)$ in the sample, $\hat{Q}$ is the mean of the positive predicted ice cream purchases in the θ CQR, and $\Pr(Q>0)$ is the probability of purchasing. The own-price elasticity is calculated as

$$(3)\qquad e_\theta = \frac{\hat{\beta}_1}{\hat{Q}}\cdot \Pr(Q>0)\,.$$

The elasticities with respect to cross prices, age, trend, and temperature are calculated in similar ways. However, the elasticities must be interpreted with caution. As explained in Buchinsky (1998), it does not necessarily follow that a household in the θ quantile before a price or income change will remain in that quantile after the change. Such effects are not incorporated in the estimated elasticities we report.

Quantile Regression, Censored Quantile Regression, and Survey Weighting

A linear regression model defines the conditional mean of the dependent variable, *y*, as a linear function of a vector of explanatory variables, *x*; that is,

(4) $$E(y_i \mid x_i) = x_i'\beta .$$

Correspondingly, quantile regressions (QRs) define the conditional quantiles of the dependent variable as a function of the explanatory variables. QRs enable us to describe the entire conditional distribution of the dependent variable given the explanatory variables. In our case, the dependent variable is purchases of ice cream. The QR model, as introduced by Koenker and Bassett (1978), can be written as

(5) $$Q_\theta(y_i \mid x_i) = x_i'\beta_\theta$$

where $Q_\theta(y_i \mid x_i)$ denotes the θ conditional quantile of y_i and the conditional quantile of the error term is zero. The QR estimator of β_θ is found by solving the problem

(6) $$\min_{\beta_\theta} \frac{1}{N}\left\{ \sum_{y_i \geq x_i'\beta_\theta} \theta \mid y_i - x_i'\beta_\theta \mid + \sum_{y_i < x_i'\beta_\theta} (1-\theta) \mid y_i - x_i'\beta_\theta \mid \right\}.$$

This minimization problem can be solved by linear programming for the different quantiles of the dependent variable as described in, for example, Koenker (2005). When θ = 0.5, the problem is reduced to minimizing the sum of the absolute deviations of the error terms, which results in the least absolute deviation estimator.

Heteroscedasticity is often associated with cross-sectional data. According to Deaton (1997), QR is useful in the presence of heteroscedasticity. If the heteroscedasticity depends on the regressors, the estimated slope parameters differ in different quantiles. However, when the distribution of the errors is homoskedastic, the estimated slope parameters from QR and ordinary least squares (OLS) are identical and only the intercepts differ (Deaton 1997). When the distribution of the errors is symmetrical, the OLS intercepts and those of the median regression are also identical. Buchinsky (1998) notes two more advantages of the QR model. First, when the error terms are not normally distributed, the QR estimator may be more efficient than the OLS estimator. Second, the QR parameter estimates are relatively robust to outliers because, in contrast to the OLS, it is not the magnitude of the dependent variable that matters but on which side of the estimated hyperplane is the observation. In the case of a positive residual, the dependent variable can be increased towards infinity without altering the solution. For a negative residual, the solution will be the same if the dependent variable is decreased towards minus infinity.

Many households did not purchase ice cream during the survey period and so the data are censored at zero. A standard procedure for correcting for zero censoring is to use a Tobit model, as discussed in, for example, Amemiya (1984). The Tobit model can be written as

$$(7) \qquad y_i = \max\left\{0, x_i'\beta_\theta + \varepsilon_i\right\}.$$

If the errors are not normally distributed and homoskedastic, the Tobit estimator is biased and inconsistent (Greene 2000). Powell (1986) showed that, under some weak regularity conditions, a class of the CQR estimator is consistent whatever the distribution of the error term and, furthermore, is asymptotically normal. When the conditional quantile of the error term is zero, a CQR model of purchases that are censored at zero can be written as

(8) $$Q_\theta(y_i \mid x_i) = \max\left\{0, Q_\theta(x_i'\beta_\theta + \varepsilon_{\theta i} \mid x_i)\right\} = \max\left\{0, x_i'\beta_\theta\right\}.$$

The CQR estimator of β_θ suggested by Powell (1986) is found by solving

(9) $$\min_{\beta_\theta} \frac{1}{N}\sum_{i=1}^{N}\left[\left\{\theta - I\left(y_i < \max\left\{0, x_i'\beta_\theta\right\}\right)\right\}\left(y_i - \max\left\{0, x_i'\beta_\theta\right\}\right)\right]$$

where I is an indicator function taking the value of unity when the expression holds and zero otherwise. Equation (9) may be estimated by an iterative algorithm proposed by Buchinsky (1994), or a programming algorithm by Fitzenberger (1997). But, as Fitzenberger noted, these algorithms perform poorly when a large proportion of the data is censored. The reason for the poor performance is that there are convergence problems when searching for a global minimum near the censoring point. We have used the three-step algorithm proposed by Chernozhukov and Hong (2002), which does not have such problems. This algorithm is simple, robust, and performs well near the censoring point. This procedure selects a sub sample by a separation restriction that is put on the censoring probability, and estimates the model twice by quantile regression. The goal of the first estimation is to find an appropriate sub sample, and the purpose of the second estimation is to make the estimator efficient. In our model, with censoring at zero, the algorithm is described in the following three step procedure.

Step 1: Estimate a probability model on the sample: $\Pr(y_i > 0 \mid x_i) = F(x_i'\gamma) + \varepsilon_i$ Use the probability model to select the sub sample $J_0 = \{i : x_i'\hat{\gamma} > 1 - \theta + c\}$ where c is a trimming constant between 0 and 1. The goal of step 1 is to chose a subset of observations where $\Pr(y_i > 0 \mid x_i) > 1 - \theta$, that is, where the quantile line $x_i'\beta_\theta$ is above the censoring point. In our case, we estimated the Logit model in step 1.

Step 2. Obtain the initial estimator, $\hat{\beta}_\theta^0$, by ordinary QR on the sample J_0. It is shown by Chernozhukov and Hong (2002) that this step gives a consistent but inefficient estimator. Use the initial estimator to select the sample $J_1 = \left\{x_i'\hat{\beta}_\theta^0 > 0\right\}$ to be used in step 3.

Step 3: Estimate the model by ordinary QR on the sample J_1. Chernozhukov and Hong (2002) show that this step gives a consistent and efficient estimate of $\hat{\beta}_\theta$.

To choose the trimming constant, *c*, Chernozhukov and Hong (2002) suggest using the minimum value of the Powell objective function in (9). In addition, they recommend as a robustness diagnostic to check whether $J_0 \subset J_1$. If a large proportion of observations in J_0 is not in J_1, then one should revise the trimming constant and possibly also the separation models or the conditional quantile models in question. We used the following procedure to choose *c*: In the second step, we estimated QRs for *c* taking the values of 0.01, 0.02, 0.03,…. The *c*-value that minimized the value of the objective function in (9) was chosen after substituting $\hat{\beta}_\theta$ for β_θ. To avoid using a too small sample in the second step, we imposed that $\frac{\# J_1}{\# J_0} > 0.66$. To ensure robustness, we imposed that, $\frac{\#\{J_0 \not\subset J_1\}}{\# J_1} < 0.1$.

The standard errors of the parameter estimates are obtained by using the CQR bootstrapping procedure of Bilias, Chen and Ying (2000). This algorithm uses the predicted values of the CQR to select the bootstrap sample, and they show that the distribution of the CQR bootstrap estimator converges to the CQR estimator. We have implemented the CQR algorithm and the bootstrap procedure in Stata (StataCorp. 2007) and our program is developed from the “qreg” command.

To take account of non-response in the surveys, we used the probability weights constructed by Statistics Norway. Lipsitz et al. (1997) showed that if the probability of participating in the survey is π_i, then using $x_i^* = \pi_i^{-1} x_i$ for the explanatory variables and $y_i^* = \pi_i^{-1} y_i$ for the dependent variable will yield unbiased parameter estimates in the QRs. Hence, we use x_i^* and y_i^* in the estimation of the CQRs . The Tobit model was estimated using x_i and y_i, and weighting the log-likelihood function with π_i^{-1}.

Data and Price Construction

The data were obtained from the household expenditure surveys of Statistics Norway over the 1986–2001 period and are described in Statistics Norway (1996). In the surveys, the country is divided into sampling areas corresponding to the more than 400 counties of Norway. These sampling areas are grouped in 109 strata, and a sample area is randomly drawn from each stratum. Sampling areas are drawn with a probability proportional to the number of persons living in the area. Next, persons are randomly drawn from the 109 sampling areas such that by design the sample is self-weighted (the need for weights stems from non-response). When a person is drawn, the household of that person is included. Finally, these households are randomly drawn to record their expenditures in one of the 26 two-week survey periods of the year. Each year 2,200 persons are initially drawn. The non-response rate varies between 33% and 52% and our total sample consists of 20,550 cross-sectional observations.

For food and beverage products, the quantities purchased and the corresponding expenditures are recorded, and these values can be used to calculate unit values. However, unit values are affected by quality differences. Such quality differences include the brand of ice cream, the size of the package of ice cream, and the place of purchase. Furthermore, unit values are missing for households not purchasing ice cream in the survey period. Therefore, we constructed qualityadjusted prices following Cox and Wohlgenant (1986). This method was also used by Park and Capps (1997) and Kuchler, Tegene, and Harris (2005).

First, the unit value of ice cream was calculated by dividing expenditure by quantity (in kilograms) for each household with a positive purchase of ice cream in the survey period. To avoid outliers, in each year we replaced the unit values above the 0.99 quantile with the 0.99 quantile unit value, and the unit values below the 0.01 quantile with the 0.01 quantile unit value.

Second, we estimated a linear regression using the calculated unit values for ice cream as dependent variable and total expenditure, age of household, size of household, type of household, and dummy variables representing year, quarter, and region as independent variables. The parameter estimates of this regression are shown in Table A1 in the appendix.

Third, household h's quality-corrected price, $\hat{p}^h$, is constructed as

$$(10) \qquad \hat{p}^h = \hat{\alpha} + \sum_{t=1987}^{2001} \hat{\beta}_t D_t^h + \sum_{j=1}^{3} \hat{\gamma}_j Q_j^h + \sum_{k=1}^{5} \hat{\delta}_k R_k^h + \hat{\varepsilon}^h$$

where D_t are yearly dummy variables, Q_j are quarterly dummy variables, R_k are regional dummy variables, and $\hat{\varepsilon}^h$ is the residual term for household h. The parameter values $\hat{\alpha}, \hat{\beta}, \hat{\gamma},$ and $\hat{\delta}$ and their associated t-values are reported in Table A1. The quality-corrected prices of households without purchases are set to the value of equation (10) excluding the residual term. These prices are the average of the quality corrected prices as regards year, quarter, and region. This procedure was also used to compute quality-corrected prices for potato chips, candy, and fresh fruits. The consumer price index for non-durables and services was used as a proxy variable for the price of the group other non-durables and services.

Table 2 shows the distributions of the dependent and the explanatory variables across the unconditional quantile groups defined according to the distribution of per capita purchases of ice cream. The column headed “Zero” shows the mean values of the variables for households that did not purchase ice cream in the survey period. The subsequent five columns show the mean values for five quantile groups. The final column reports the mean values for all households. The 0.50 quantile column reports the mean values for households with the lowest 50% of purchases of ice cream and includes households that did not purchase ice cream. The 0.60 quantile column shows the mean values for households with between the lowest 50% and 60% of purchases, and so on. We will refer to households in the 0.50 and 0.60 quantiles as light ice cream eaters, households in the 0.70 and 0.80 quantiles as moderate ice cream eaters, and households in the 0.90 quantile as heavy ice cream eaters.

The first row reports average purchases of ice cream. They vary between zero and 446 grams (slightly more than a pint of ice cream) per person in each two-week period. The subsequent rows report indexes of total (log) expenditures on non-durables and services, squared (log) expenditures, and (log) prices, all deflated by the consumer price index for non-durables and services[5]. These indexes vary little between quantiles. To take account of the effect of warm weather on ice cream purchases, we introduced a temperature variable. We set the temperature variable to one for survey periods with mean temperatures below 15 degrees Celsius and to (log of) the mean temperature for the remaining survey periods. We use the temperature as measured at a meteorological station located in each of the six regions. The temperature is linked to each household according to the household’s survey period and location. As expected, the proportion of heavy ice cream eaters increased during periods with high temperature. The mean values of the (log of) age of the head of the household and the (log of) trend variable are also shown in the table. Information on other potentially important household characteristics, such as education and ethnic origin, is not included in the surveys.

Finally, mean values of dummy variables defining region, degree of urbanization, season, and household type are reported. The dummy variable for the Christmas season is set to 1 in the 26th two-week period of each year and zero otherwise. The mean values of the dummy variables represent the percentages of the total number of households belonging to the respective categories. The reference household lives in the Central East region, in a non-major city, is surveyed during winter but not Christmas, and comprises a couple with children. Non-consumers and light ice cream eaters are strongly represented among one-person households and couples without children, moderate and heavy ice cream eaters are strongly represented among couples with children. We also note that many households are heavy ice cream eaters during spring, summer, and Christmas.

Table 2 about here

Results

Table 3 reports the marginal effects of the Logit model used in the first step, the CQR model, and the Tobit model.[6] The Logit marginal effects are the estimated coefficients multiplied by the probability density function, and the CQR and Tobit marginal effects are the estimated coefficients multiplied by the probability of a positive purchase. The corresponding *t*-values are reported in parentheses. As discussed above, the (inconsistent) estimates of the Tobit model are included for comparison.

The effect of the total expenditure variable is negative and significant, and the effect of the squared total expenditure variable is positive and significant except for heavy ice cream eaters. In the 0.90 quantile, the expenditure effects are insignificant. The effect of the price of ice cream is negative and significant in all the quantiles, and the magnitude is increasing with the quantiles. Candy prices have insignificant effects except for the 0.80 quantile. The effects of potato chips prices are positive and significant over the entire conditional distribution, indicating that chips are a substitute for ice cream. Fruit prices have also positive and significant effects over the entire distribution indicating that for many households fresh fruits are a good substitute for ice cream.

The Tobit estimates differ substantially from the CQR estimates. The Tobit estimates for the expenditure variables are significant, but mostly have the opposite sign of the CQR estimates. The own-price effects are lower in the Tobit than in the CQR model. Contrary to the highly significant cross-price effects for potato chips and fresh fruits, the corresponding Tobit cross-price effects are insignificant. These differences demonstrate the usefulness of the CQR approach.

Most of the dummy variables are significant at the 5% level. The reference region is the Central East. Purchases in most of the other regions are lower in all quantiles. Purchases in the three major cities and rural areas are lower than in the non-major cities. As expected, more ice cream is purchased in spring and summer than in fall and winter. The effect of summer is also partly captured by the significantly positive effect of the temperature variable discussed below. The effects of household composition, as compared with the reference household, are not very different across quantiles. In all quantiles, ice cream purchases are higher for families consisting of couples with children. The pseudo R^2 values vary from 0.06 in the 0.50 quantile to 0.31 in the 0.90 quantile[7].

The elasticities of the continuous variables are reported in Table 4. Using bootstrap estimates of the variance-covariance matrix, we find that most of the differences in the estimated elasticities are statistically significant. The total expenditure elasticity increases from 0.01 in the lowest to 0.28 in the highest quantile and it is significant in all the quantiles, except for the lowest. The own-price elasticity changes from –2.40 in lowest to –1.18 in the highest quantile. The cross-price elasticity between ice cream and candy is low and insignificant, except for in the 0.80 quantile, where it is negative and significant. The cross-price elasticities between ice cream and potato chips and ice cream and fresh fruits, however, are positive and significant in the whole distribution, suggesting that decreasing chips or fruit prices contribute to lower ice cream purchases.

Age of the head of household has a significant and negative effect in all the quantiles, suggesting that households purchase less ice cream as the head ages. The obesity rate is highest among younger people (Departementene 2007: 108) so this result is not surprising. In the 0.50 quantile, a 1% increase in age of the head leads to a 0.24% decrease in purchases of ice cream. The effect of age is largest among light and moderate ice cream eaters and the Tobit model mispredicts the effect for these households. The trend is significant and negative in all the quantiles, and strongest among the light ice cream eaters. The relative effect of temperature is also strongest among light and moderate ice cream eaters. A 1% increase in temperature above 15 centigrade results in a 0.40% increase in the purchases of ice cream at the lowest quantile.

In most cases, the elasticities differ across the quantiles, and frequently the elasticities of the Tobit model are different from the elasticities estimated by the CQR model. These differences are illustrated in Figure 1 showing the distribution of the own-price and total expenditure elasticities across the quantiles. The CQR and Tobit elasticities and their 90% confidence intervals are shown. For the lower quantiles, the CQR total expenditure elasticities are well outside the confidence intervals of the Tobit elasticity. The CQR own-price elasticities change substantially across the quantiles and are mostly outside the confidence intervals of the corresponding Tobit elasticity. Given different effects of price changes, we will investigate the effects of increasing the VAT among light, moderate, and heavy ice cream eaters.

Table 3 about here

Table 4 about here

Figure 1 about here

The Effects of An Increase in the Value Added Tax of Ice Cream

We calculated the average predicted purchases along each of the conditional quantiles with and without a VAT increase from 14% to 25% for ice cream, candy, and potato chips and a removal of the VAT for fresh fruits. This corresponds to a price increase of 9.6% for the less healthy food and a price reduction of 14% for the healthy food.

The law of iterated expectations ensures that the mean of the conditional means of the dependent variable is equal to the unconditional mean of the dependent variable. This law implies that the unconditional mean of the dependent variable can be predicted using the conditional means. Unfortunately, there exists no equivalent law of iterated quantiles and calculating the mean of the unconditional quantile directly from the means of the conditional quantile is only an approximation. Gustavsen and Rickertsen (2006) used this approximation and predicted the effects of a VAT change using the elasticities on the unconditional quantiles of the purchases. However, in our case, we estimated the conditional median, and the value of the unconditional median is zero, so we selected a different approach

We simulated model (1) with the marginal effects presented in Table 3 over the whole sample for each of the conditional quantiles 0.50, 0.60, 0.70, 0.80, and 0.90. i.e. we inserted the values of the covariates in model (1), multiplied by the estimated parameters and the probabilities and found the predicted per capita purchase for each household. Consequently, we predicted the purchases of ice cream along each conditional quantile, not the quantiles of the unconditional purchases. Some of the households with positive purchases before the price changes, were predicted to have negative purchases after the changes. The purchases for those households were set to zero.

Table 5 shows the annual per capita average predicted purchases of ice cream before and after the VAT changes. In the lowest quantile, the average purchase decreases from 1.4 to 1.0 kilograms, or by 28.0%. Most of the reduction is due to reduced purchases (27%) and only a small reduction (1.0%) is due to people who stop purchasing ice cream altogether. In the 0.70 quantile, the average purchase is reduced from 4.2 to 3.3 kilograms, a 22.2% decrease. The reduction is mainly due to peoples' reduction in purchases (21.5%), but to some extent people also stop purchasing ice cream (0.7%). In the 0.90 quantile, the purchase is reduced from 13.6 to 11.7 kilograms, a 13.6% reduction. Again the reduction is mainly due to reduced purchases (13.4%). Over the whole distribution, the percentage decrease is greatest among light ice cream eaters, but the decrease in terms of kilograms is largest among heavy ice cream eaters. The Tobit model predicts only a 5.2% decrease in the average purchase, which corresponds to 0.1 kg a year.

To investigate the effects of reduced consumption on body weight, we converted the reduced purchases to annual changes in body weight using conversion factors published by the Swedish Food Administration (National Food Administration 2007). One kilogram of ice cream with a 15% fat content contains about 2,220 kilocalories (kcal). Body fat contains about 20% water, and $9{,}000 \cdot 0.8 = 7{,}200$ kcal are required to gain one kilogram body weight. Using these values and assuming that the reduced ice cream purchases are not replaced by purchases of other foods or beverages, the suggested VAT changes will lead to more than half a kilogram annual reduction in body weight among people in the 0.90 quantile. In the lower quantiles, the annual reduction will be in the range of 0.1 to 0.4 kilograms. Taking into account that obesity builds up over a long period of time the accumulated weight increase may be notable among heavy ice cream eaters.

Table 5 about here

Conclusions

The consumption of ice cream is problematic from a public health perspective. Ice cream is high in both sugar and saturated fats and consumption may contribute to obesity with associated health problems. While decreasing ice cream consumption among those who consume a small amount of ice cream is not likely to have much health benefit, the public health benefit could be quite different for those who are consuming large quantities. When there are different marginal effects from changing a covariate among persons purchasing much and persons purchasing little, a quantile regression approach is warranted. Ice cream purchases are highly censored, and we used the algorithm for CQRs developed by Chernozhukov and Hong (2000). This algorithm is simple, robust, performs well near the censoring point, and results in estimates that differ substantially from the Tobit estimates.

Our results suggest that a policy of increasing the VAT for less healthy food items and removing the VAT for healthy food items will have largest percentage impact among light ice cream eaters, but the absolute impact will be highest among heavy ice cream eaters. Heavy ice cream eaters will reduce their annual purchases by about 1.8 kilograms. If these purchases are not replaced with purchases of other foods, this reduction corresponds to an annual reduction of more than half a kilogram of body weight. Over a ten year period the accumulated effect could be more than five kilograms of body weight suggesting that price interventions may be an efficient policy tool in reducing the growth of obesity.

Notes

[1] In our household expenditure survey data, the per capita purchases of each household are multiplied by 26 to approximate annual per capita consumption.

[2] Zero purchases may arise for several reasons. First, in survey data with short recall periods, zero purchases may arise because of infrequency of purchases. Second, some households may not purchase ice cream because they do not like the product. Third, zero purchases may arise because of standard corner solutions that result in a Tobit model. We do not have information that enables us to identify the main reason for the reported zero purchases in our sample, and we have chosen to use the Tobit model as our benchmark model.

[3] A demand-system framework could incorporate substitution effects between ice cream and, for example, candy, potato chips, and fresh fruits. However, across-equation restrictions such as symmetry cannot be imposed neither on quantile regression (QR) nor on CQR models. Therefore, a system framework is less meaningful and a single-equation model is estimated. An area of future research is to extend the QR framework to a system of demand equations.

[4] The head of the household is defined as the household member with the highest income. Note, however, that household income data are unavailable.

[5] For households surveyed on days in two different months, we used a weighted average of the consumer price index for those two months. The number of survey days in each month were used as weights.

[6] As discussed above, the CQR estimator has usually been estimated either by using the algorithm proposed by Buchinsky (1994) or the algorithm proposed by Fitzenberger (1997). These algorithms are expected to perform poorly when are large proportion of the data is censored. For comparison, we estimated equation (9) using Buchinsky's algorithm. As expected, many of the estimated parameters changed substantially. Especially in the lower quantiles, the absolute values

of the parameter estimates increased. One possible explanation for these differences is that the solutions estimated by Buchinsky's algorithm are local and not global minimum points.

[7] The pseudo R^2 value is calculated as

$$1-\frac{\text{sum of weighted deviations about estimated quantile}}{\text{sum of weighted deviations about raw quantile}},$$

i.e., the numerator is calculated by using equation (6) inserting the estimated parameter values and the denominator is calculated by using equation (6) inserting the intercept calculated by an intercept only model.

References

Amemiya, T. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 24:3–61.

Bilias, Y., S. Chen, and Z. Ying. 2000. "Simple Resampling Methods for Censored Regression Quantiles." *Journal of Econometrics* 99:373-86.

Buchinsky, M. 1994. "Changes in the US Wage Structure 1963–1987: Application of Quantile Regression." *Econometrica* 62:405–58.

Buchinsky, M. 1998. "Recent Advances in Quantile Regression Models – A Practical Guideline for Empirical Research." *Journal of Human Resources* 33:88–126.

Cotton, P.A., A.F. Subar, J.E. Friday, and A. Cook. 2004. "Dietary Sources of Nutrients among US Adults, 1994 to 1996." *Journal of American Dietetic Association* 104:921-30.

Chernozhukov, V., and H. Hong. 2002. "Three-Step Censored Quantile Regression and Extramarital Affairs." *Journal of American Statistical Association* 97:872-82.

Cox, T.L., and M.K. Wohlgenant. 1986. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68:909-19.

Danish Academy of Technical Sciences (2007). "Economic Incentives in the Nutritional Policy"

(Økonomiske virkemidler i ernæringsplitikken). Report from Danish Academy of Technical Sciences. Available at: http://www.atv.dk/c/C1-ATV-publikationer-omtaler-pdf-filer/C133-Okonomiske-virkemidler.pdf (accessed January 30 2008).

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

Departementene. 2007. "A Recipe for a Healthier Diet" (Oppskrift for et sunnere kosthold). Available at: http://www.regjeringen.no/upload/kilde/hod/prm/2007/0006/ddd/pdfv/304657-kosthold.pdf (accessed July 5, 2007.

Fitzenberger, B. 1997. "A Guide to Censored Quantile Regression." In Maddala, G.S., and C.R. Rao (Eds.) *Handbook of Statistics* 15, pp. 405-437. Elsevier, Amsterdam.

Gustavsen, G.W., and K. Rickertsen K. 2006. "A Censored Quantile Regression Analysis of Vegetable Demand: The Effects of Changes in Prices and Total Expenditure." *Canadian Journal of Agricultural Economics* 54:631–45.

Greene, W.H. 2000. *Econometric Analysis*. 4th ed. Prentice-Hall Inc., Upper Saddle River, NJ.

Koenker, R. 2005. *Quantile Regression*. Cambridge University Press, New York.

Koenker, R., and G. Bassett Jr. 1978. "Regression Quantiles." *Econometrica* 46: 33–50.

Kuchler, F., A. Tegene, and J.M. Harris. 2005. "Taxing Snack Foods: Manipulating Diet Quality or Financing Information Programs?" *Review of Agricultural Economics* 27: 4-20.

Lipsitz, S.R., G.M Fitzmaurice, G. Molenberghs, and L.P. Zhao. 1997. "Quantile Regression Methods Longitudinal Data with Drop-Outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus." *Applied Statistics* 46:463-76.

National Food Administration. 2007. Available at: http://192.121.81.11/livsmedelsok/sok.aspx?lang=2 (accessed February 26 2008).

Park, J.L., and O. Capps. 1997. "Demand for Prepared Meals by US Households." *American Journal of Agricultural Economics* 79:814-24.

Powell, J.L. 1986. "Censored Regression Quantiles." *Journal of Econometrics* 32: 143–155.

StataCorp. 2007. *Stata Statistical Software*, Release 10. College Station, TX.

Statistics Norway. 1996. "Survey of Consumer Expenditure 1992–1994." Official Statistics of Norway. Oslo–Kongsvinger.

Stewart, H., N. Blisard, and D. Jolliffe. 2003. "Do Income Constraints Inhibit Spending on Fruits and Vegetables among Low-Income Households?" *Journal of Agricultural and Resource Economics* 28: 465–80.

World Health Organization. 2003. "Diet, Nutrition and the Prevention of Chronic Diseases." WHO Technical Report Series 916. Available at: http://www.fao.org/docrep/005/AC911E/AC911E00.HTM (accessed February 20 2008).

World Health Organization. 2005. "The Challenge of Obesity in the WHO European Region." Fact sheet EURO13/05. Copenhagen.

Table 1. Distribution of Annual Per Capita Ice Cream Purchases

Year	Zero purchases	Quantile 0.50	0.60	0.70	0.80	0.90	Mean
1986	64	0.0	0.0	2.0	6.5	13.0	3.8
1987	62	0.0	0.0	2.6	6.5	13.0	3.8
1988	57	0.0	0.9	3.8	7.8	13.0	4.2
1989	58	0.0	0.8	3.6	7.3	14.3	4.6
1990	58	0.0	0.6	2.7	6.5	13.0	4.2
1991	52	0.0	1.6	4.3	8.5	15.6	5.1
1992	50	0.0	2.1	4.7	10.0	18.2	5.3
1993	55	0.0	1.2	3.3	7.4	14.6	4.6
1994	55	0.0	1.1	4.9	9.1	16.1	5.4
1995	60	0.0	0.2	3.2	8.7	16.6	5.0
1996	59	0.0	0.3	3.3	8.1	15.6	4.7
1997	54	0.0	1.3	5.2	9.8	18.7	5.3
1998	56	0.0	0.7	3.8	8.4	17.1	5.3
1999	52	0.0	1.0	3.9	7.8	15.6	4.7
2000	45	0.2	1.3	4.0	8.5	14.7	4.4
2001	52	0.0	0.7	2.6	6.8	13.0	3.7

Notes: The unconditional quantile and mean values are weighted with sampling weights to account for non-response. The column "Zero purchases" reports the average percentage of households reporting zero purchases of ice cream in each two-week survey period.

Table 2. Mean Values by Quantile

Variable	Zero	Quantile 0.50	0.60	0.70	0.80	0.90	Mean
Ice cream purchases[a]	0.0	0.1	35.5	128.0	262.4	446.4	180.2
Indexes (log of)							
Total expenditure	2.2	2.2	2.1	2.2	2.1	2.2	2.2
Total expenditure squared	5.0	5.0	4.9	5.0	4.8	5.2	5.0
Price of ice cream	–1.3	–1.3	–1.1	–1.1	–1.5	–1.6	–1.4
Price of candy	–0.9	–0.9	–0.9	–0.9	–0.9	–0.9	–0.9
Price of chips	–1.4	–1.4	–1.4	–1.4	–1.4	–1.4	–1.4
Price of fresh fruits	–2.8	–2.8	–2.8	–2.8	–2.8	–2.8	–2.8
Other variables							
Age (log of years)	3.8	3.8	3.7	3.7	3.7	3.8	3.8
Trend (log of)[b]	1.8	1.8	2.0	1.9	1.9	1.9	1.9
Temperature[c]	11.4	11.3	20.6	30.0	26.2	36.0	21.3
Dummy variables (%)							
Region							
Central East	19.3	19.2	18.8	20.6	19.2	19.8	19.6
Rest of East	29.2	29.2	27.8	23.8	26.2	27.0	27.6
South	12.5	12.5	14.8	15.6	14.9	16.4	14.2
West	17.3	17.3	16.9	18.5	20.0	18.6	17.8
Central	10.4	10.5	10.5	9.4	9.2	8.9	9.9
North	11.3	11.3	11.1	12.0	10.5	9.2	10.8
Urbanization							
Rural area	23.4	23.5	20.7	20.0	21.3	21.1	21.9
Non-major city	58.7	58.8	62.3	61.8	61.7	61.0	60.3
Major city	17.8	17.8	17.0	18.2	17.0	17.9	17.8
Season							
Winter	31.3	31.2	17.4	16.8	19.8	16.3	23.9
Spring	19.8	19.7	29.6	37.4	31.2	35.2	27.2
Summer	16.4	16.4	26.2	26.9	25.4	27.2	21.8
Fall	32.6	32.6	26.8	18.9	23.7	21.2	27.1
Christmas	3.7	3.8	1.7	2.5	4.1	3.9	3.4
Household type							
One person	22.4	22.0	4.7	8.7	4.7	6.1	15.0
Couple without children	27.3	27.1	16.9	14.5	12.7	27.1	22.8
Couple with children	34.2	34.8	60.9	61.1	66.8	52.6	46.4
Single parent	4.8	4.8	5.2	4.8	4.5	3.9	4.6
Other households	11.3	11.4	12.3	10.8	11.3	10.2	11.0

Notes: [a]Ice cream purchases are multiplied by 1000 (i.e., a conversion from kilograms to grams).
[b]Trend is a variable that takes the value of 1 in 1986, 2 in 1987,…16 in 2001.
[c]Temp is defined as 0 for temperatures below 15 C° and the log of C° when temperature is equal to or above 15 C°. The indexes for Temp are multiplied by 100 in the table.

Table 3. Marginal Effects for the Logit , CQR and Tobit Models

Variable	Logit	Quantile 0.50	0.60	0.70	0.80	0.90	Tobit
Total expenditure	2.74	−0.53	−0.34	−0.32	−1.10	0.32	1.37
	(6.28)	(−7.46)	(3.62)	(−1.96)	(−3.77)	(0.55)	(7.21)
Total expenditure sq	−0.28	0.12	0.12	0.21	0.42	0.28	−0.16
	(−2.88)	(7.23)	(5.04)	(4.79)	(5.70)	(0.99)	(−3.90)
Price of ice cream	−2.19	−2.26	−3.08	−4.41	−5.41	−6.46	−2.12
	(−12.18)	(−28.61)	(−34.26)	(−39.36)	(−39.76)	(−40.63)	(−28.54)
Price of candy	−0.41	−0.01	−0.03	−0.08	−0.24	−0.40	−0.28
	(−2.25)	(−0.66)	(−0.76)	(−1.33)	(−2.25)	(−1.67)	(−3.52)
Price of chips	0.30	0.17	0.33	0.45	0.61	0.85	0.09
	(1.96)	(7.82)	(9.65)	(10.16)	(7.13)	(3.99)	(1.18)
Price of fresh fruits	−0.05	0.30	0.45	0.74	0.86	1.33	0.10
	(0.26)	(11.96)	(12.83)	(14.33)	(8.26)	(5.80)	(1.18)
Age	−0.12	−0.23	−0.35	−0.57	−0.39	−0.34	0.07
	(−0.79)	(−12.11)	(−12.64)	(−12.86)	(−5.76)	(−1.94)	(1.01)
Trend	0.34	−0.16	−0.22	−0.28	−0.32	−0.39	−0.05
	(5.44)	(−16.77)	(−14.53)	(−12.50)	(−6.67)	(−6.19)	(−2.04)
Temperature	0.60	0.38	0.55	0.56	0.58	0.58	0.26
	(7.96)	(7.56)	(14.75)	(11.41)	(8.77)	(5.19)	(9.21)
Rest of East	−0.29	−0.13	−0.14	−0.16	−0.33	−0.59	−0.15
	(−1.89)	(−6.50)	(−4.48)	(−3.28)	(−3.39)	(−2.75)	(−2.17)
South	0.37	−0.15	−0.10	−0.03	−0.10	0.23	0.12
	(2.08)	(−5.69)	(−2.20)	(−0.52)	(−0.72)	(0.90)	(1.46)
West	0.11	−0.10	−0.13	−0.08	−0.03	0.07	0.04
	(0.67)	(−4.13)	(−3.74)	(−1.58)	(−0.26)	(0.29)	(0.59)
Central	0.16	−0.04	0.01	−0.06	− 0.01	−0.13	0.05
	(0.88)	(−1.79)	(0.19)	(−1.07)	(−0.08)	(0.54)	(0.66)
North	−0.07	−0.13	−0.19	−0.25	−0.38	−0.33	−0.05
	(−0.37)	(−5.50)	(−5.17)	(−4.44)	(−3.28)	(−1.27)	(−0.55)
Rural area	−0.25	−0.07	−0.10	−0.19	−0.29	−0.35	−0.13
	(−2.14)	(−1.49)	(−4.77)	(−5.59)	(−4.98)	(−2.27)	(−2.68)
Major city	0.01	−0.04	−0.06	−0.12	−0.35	−0.21	0.02
	(0.09)	(−1.90)	(−1.74)	(−2.38)	(−3.65)	(−0.89)	(0.22)
Spring	3.48	0.83	1.71	2.75	3.77	3.81	1.69
	(28.38)	(25.93)	(30.48)	(34.76)	(31.66)	(18.87)	(23.42)
Summer	2.81	0.66	1.07	1.99	2.80	2.99	1.35
	(19.87)	(24.59)	(27.31)	(31.03)	(23.74)	(13.64)	(18.44)
Fall	0.45	0.01	−0.04	0.01	0.06	−0.02	0.18
	(3.43)	(0.54)	(−2.21)	(0.45)	(1.30)	(−0.11)	(2.77)
Christmas	0.78	0.18	0.35	0.61	0.77	0.93	0.31
	(2.94)	(3.96)	(5.82)	(6.67)	(3.30)	(1.59)	(2.26)
One person	−4.16	−0.34	−0.55	−1.15	−1.06	−0.39	−1.14
	(−43.44)	(−11.47)	(−14.50)	(−19.78)	(−9.81)	(−1.38)	(−18.47)
Couple without children	−2.46	−0.20	−0.28	−0.34	−0.42	−0.35	−0.62
	(−26.93)	(−9.09)	(−9.59)	(−7.10)	(−4.94)	(−2.10)	(−14.09)
Single parent	−1.65	−0.35	−0.58	−0.91	−0.92	−1.01	−0.51
	(−10.83)	(−12.43)	(−12.93)	(−15.62)	(−9.83)	(−4.89)	(−7.21)
Other households	−1.43	−0.30	−0.47	−0.51	−0.62	−0.65	−0.44
	(−11.99)	(−14.43)	(−16.75)	(−11.09)	(−8.06)	(−4.25)	(−8.88)
Pseudo R^2	0.14	0.06	0.11	0.19	0.22	0.31	0.13

Sample size (in the last step)	18,062	18,560	18,525	19,813	20,141
Trimming constant	0.10	0.15	0.05	0.11	0.26

Notes: All the marginal effects are multiplied by 10. The *t*-values are reported the in parentheses. The CQR and Tobit marginal effects are equal to the estimated parameters multiplied by the probability of purchasing ice cream. The following dummy variables define the reference alternative and are excluded from the table: Central East, Non-major city, Winter, and Couple with children.

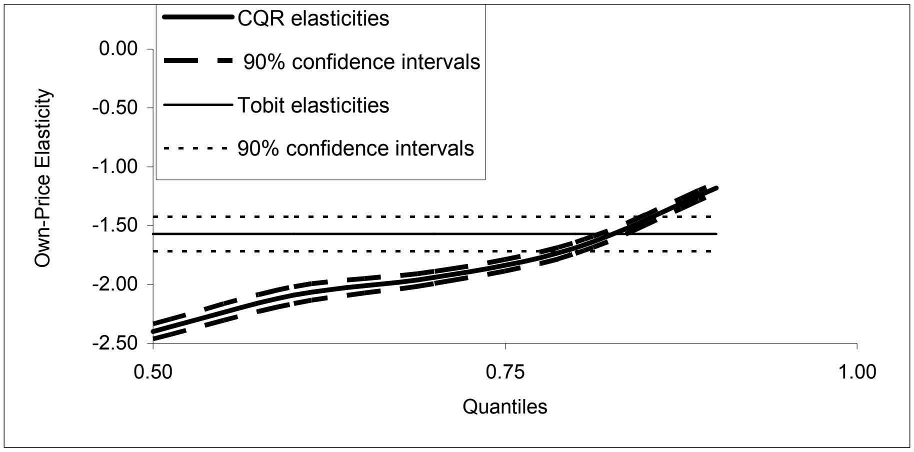
Table 4. Estimated Elasticities

Elasticity	Quantile 0.50	0.60	0.70	0.80	0.90	Tobit
Total expenditure	0.01	0.11	0.25	0.22	0.28	0.28
	(0.61)	(9.26)	(16.08)	(10.75)	(9.49)	(15.68)
Price of ice cream	–2.40	–2.09	– 1.94	–1.69	–1.18	–0.86
	(–61.96)	(–48.40)	(–61.42)	(–57.43)	(–46.98)	(–38.19)
Price of candy	–0.02	–0.02	–0.03	–0.08	–0.07	–0.11
	(–1.00)	(–0.82)	(–1.35)	(–2.28)	(–1.67)	(–3.56)
Price of chips	0.18	0.22	0.20	0.19	0.15	0.04
	(11.45)	(10.65)	(10.06)	(7.10)	(3.96)	(1.18)
Price of fresh fruits	0.32	0.30	0.33	0.27	0.24	–0.04
	(17.84)	(15.05)	(14.53)	(8.59)	(5.80)	(–1.18)
Price of other goods	1.91	1.47	1.19	1.08	0.57	1.26
	(26.48)	(19.52)	(13.99)	(8.06)	(25.22)	(13.23)
Age	–0.24	–0.24	–0.25	–0.12	–0.06	0.03
	(–18.79)	(–14.61)	(–13.35)	(–5.69)	(–1.92)	(1.02)
Trend	–0.17	–0.15	–0.12	–0.10	–0.07	–0.02
	(–27.37)	(–17.02)	(–13.02)	(–6.73)	(–4.13)	(2.05)
Temperature	0.40	0.37	0.25	0.18	0.10	0.10
	(11.26)	(15.12)	(11.37)	(8.82)	(5.11)	(9.60)

Note: *t*-values are reported in parentheses.

Table 5. Predicted Effects of Hypothetical VAT Changes

	Quantile					
	0.50	0.60	0.70	0.80	0.90	Tobit
Predicted purchase before VAT changes (kg)	1.4	2.6	4.2	7.3	13.6	2.2
Predicted purchase after VAT changes (kg)	1.0	2.0	3.3	5.9	11.7	2.1
Difference in percentage	– 28.0	– 23.3	– 22.2	– 18.7	– 13.6	– 5.2
Difference in kg ice cream	– 0.4	– 0.6	– 0.9	– 1.4	– 1.8	– 0.1
Difference in kg body weight	-0.1	-0.2	-0.3	-0.4	-0.6	-0.0



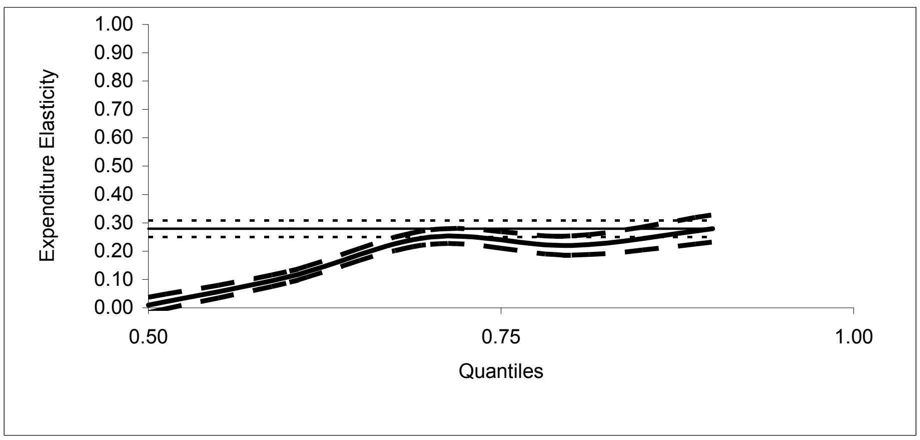


Figure 1. Censored quantile regression and Tobit elasticities with 90% confidence intervals

Appendix

Table A1. Unit Value Regression Results for Ice Cream

Variable	Parameter	*t*-value
Total expenditure	0.00	1.08
Total expenditure squared	–0.00	–1.21
Age	–0.05	–0.33
Age squared	–0.00	–1.37
Number of persons	–2.11	–5.14
One person	–4.52	–2.50
Couple without children	–6.20	–4.84
Single parent	0.55	0.32
Other households	–1.03	–0.88
Rural area	–0.70	–0.80
Major city	1.17	1.16
Spring	16.64	17.34
Summer	17.71	17.68
Fall	2.24	2.14
Rest of East	–1.45	–1.35
South	–3.84	–3.22
West	–2.88	–2.66
Central	–0.63	–0.49
North	0.31	0.24
Year		
1987	3.28	1.71
1988	13.45	7.14
1989	15.48	7.75
1990	19.46	9.87
1991	20.53	10.84
1992	18.93	10.40
1993	19.78	10.66
1994	11.48	6.19
1995	7.19	3.76
1996	11.57	6.14
1997	12.93	6.84
1998	17.35	8.91
1999	19.23	10.08
2000	23.04	12.06
2001	24.15	12.30
Constant	46.00	11.07
R^2	0.11	