

# **Public Policies and the Demand for Carbonated Soft Drinks: A Censored Quantile Regression Approach**

**Geir Gustavsen**

Paper prepared for presentation at the XI<sup>th</sup> Congress of the EAAE  
(European Association of Agricultural Economists),  
'The Future of Rural Europe in the Global Agri-Food System',  
Copenhagen, Denmark, August 24-27, 2005

Copyright 2005 by Geir Gustavsen. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

## **Public Policies and the Demand for Carbonated Soft Drinks: A Censored Quantile Regression Approach**

**Abstract:** Heavy consumption of soda may contribute to obesity, strokes, and cardiac problems. From a health perspective, the distribution of the consumption is at least as important as the mean. Censored as well as ordinary quantile regression techniques were used to estimate the demand for sugary soda based on household data from 1989 to 1999. It was found that heavy drinkers are more price- and expenditure-responsive than are light drinkers. The study shows that increasing the taxes on carbonated soft drinks will lead to a small reduction in consumption for small and moderate consumers and a huge reduction for heavy consumers.

**Keywords:** soda demand, quantile regression, taxes  
**JEL classification:** D12, I10

Heavy consumption of carbonated soft drinks may lead to excessive energy intake, contributing to obesity, strokes, and cardiac problems. These problems are increasing in the western world. In addition, soda consumption may contribute to dental caries and diabetes. The Norwegian per capita consumption of carbonated soft drinks is the third highest in the world. However, many Norwegians do not consume soda, indicating that a portion of the population consumes a larger quantity than recommended by health experts. Health experts recommend that no more than 10 percent of the energy intake should come from sugar, which corresponds to an amount of 35 to 40 grams for a child below six years, 45 to 55 grams for a schoolchild, 50 to 60 grams for an adult female, and about 70 grams for an adult male. In comparison, a 0.5 liter bottle of sugary soda normally contains about 50 grams of sugar. Although the mean soda consumption is of interest to producers in order to compute the total demand, it conveys less information to a nutrition expert. To examine the problem from a health perspective, it is important to take account of the whole distribution of the consumption. This is because there may be a greater pay-off from reducing the soda consumption of a heavy consumer than there is in the case of a low or moderate consumer. A person with heavy soda consumption will exceed the intake limit recommended by the experts, and is therefore more exposed to health problems.

This research has three main objectives. First, we will explore the purchase of soda in the whole conditional distribution, and find the factors that influence the demand. The mean effects estimated by limited dependent variable models may be satisfactory if the parameters are identical in the whole distribution. However, the effects are likely to be different for low-consumption households at the lower tail compared to persons with high consumption at the upper tail. Hence, we use a censored quantile regression approach. Second, we will examine whether price changes, which may be induced by tax changes or European Union (EU) membership, have different effects on low, moderate, and heavy soda consumers. Finally, we will model the demand for a censored good without relying on normality and identically distributed errors, two assumptions seldom fulfilled. The demand for censored goods is usually modeled with limited dependent variable models, but the consistency of these models is highly dependent on the normality and homoscedasticity of the error terms.

The next section introduces the empirical model. Then, the quantile regression and censored quantile regression techniques are presented. Next, the data are presented and the results from the quantile regressions are compared with the results from the symmetrically censored least squares (SCLS) model and the Tobit model. Finally, the price elasticities are used to calculate the effects of three different policy scenarios.

### **The Empirical Model**

As the purchase of sugary soda is censored, modeling the demand may best be done within a single equation context. Furthermore, using censored quantile regression, we cannot estimate a demand system with restrictions across the equations. Consequently, we specify Stone's logarithmic

demand function. For a discussion, see Deaton and Muellbauer (1980: 60-64). This function may be written as:

$$\ln q^h = \alpha + E \left[ \ln x^h - \sum_{j=1}^n w_{jt} \ln p_{jt} \right] + \sum_{j=1}^n e_j^* \ln p_{jt} \quad (1)$$

where  $q^h$  is household  $h$ 's per capita consumption of soda,  $x^h$  is total per capita expenditure on non-durables,  $w_{jt}$  is the average expenditure share on good  $j$  in the survey period  $t$ , and  $p_{jt}$  is the nominal price. The expenditure elasticity,  $E$ , the compensated price elasticity,  $e_j^*$ , and  $\alpha$  are the parameters to be estimated. Homogeneity in prices and total expenditure requires that  $\sum_j e_j^* = 0$ . Consequently, we may impose homogeneity by deflating the price variables in the term  $\sum_j e_j^* \ln p_{jt}$  with one of the

prices. The expression  $\sum_{j=1}^n w_{jt} \ln p_{jt}$  is Stone's price index. Moschini (1995) showed that this index is not invariant to changes in the units of measurement. To avoid this potentially serious problem, we used the (log of) consumer price index (CPI)<sup>1</sup>, which is a Laspeyres index and therefore invariant to changes in units of measurement (Moschini, 1995).

The constant term in equation (1) is expanded to include non-economic variables.  $A^h$  is the age of the head<sup>2</sup> of household  $h$ ,  $T_t$  is the two-week mean temperature in period  $t$ ,  $CH$  is a dummy variable for Christmas, and  $SC$  is a dummy variable taking account of the differences in demand before and after the introduction of the 0.5 liter plastic bottle with screw cap. Furthermore, the socioeconomic dummy and seasonal variables,  $Z^h$ , defined in table 2, and a stochastic error term,  $\varepsilon^h$ , are included. The model includes prices for two commodities only: sugary soda, and all other non-durables. Since expenditure on soda constitutes a marginal share of expenditures on non-durables, the prices for non-durables except for soda and the CPI are approximately equal. Consequently, homogeneity is imposed by deflating the soda price with the CPI. Then, the model to estimate becomes:

$$\ln q^h = \alpha_0 + \alpha_1 \ln A^h + \alpha_2 \ln T_t + \alpha_3 CH_t + \alpha_4 SC_t + \sum_{k=1}^K \beta_k Z_k^h + E \ln \frac{x^h}{CPI_t} + e^* \ln \frac{p_t}{CPI_t} + \varepsilon^h. \quad (2)$$

The compensated price elasticity,  $e^*$ , is approximately equal to the uncompensated price elasticity, because soda purchases constitute a very small share of the total consumption.

### Quantile Regression and Censored Quantile Regression

Both quantile regression and censored quantile regression are used in labor economics, but have rarely been used to study food consumption. Some exceptions are Manning (1995), who studied the demand for alcohol using quantile regression, and Variyam et al. (2002) and Variyam (2003), who study demand for nutrition using quantile regression. Steward et al. (2003) used censored quantile regression to study the effect of an income change on fruit and vegetable consumption in low-income households.

As discussed by Deaton (1997), quantile regression is most useful when the errors are heteroscedastic. Heteroscedasticity is frequently present in household expenditure data, meaning that the set of slope parameters of the quantile regressions will differ from each other as well as from the Ordinary Least Squares (OLS) parameters.

We say that a person consumes a product at the  $\theta^h$  quantile of a population if he or she consumes more of the product than the proportion  $\theta$  of the population does and less than the

<sup>1</sup> Our version of the CPI does not include durables.

<sup>2</sup> The head of the household is defined as the person who contributes most to the family economy.

proportion  $(1-\theta)$  consumes. Thus, half the households in a sample consume more than the median and half consume less. Similarly, 75 percent of the households consume less than the 0.75 quantile and 25 percent consume more. The unconditional quantile function is defined as the inverse of the cumulative distribution function.

Conditional quantile functions, or quantile regressions, define the conditional distribution of a dependent variable as a function of independent variables. If we have a relation as follows:

$$y_i = x_i' \beta + \varepsilon_i \quad (3)$$

where  $x_i$  is a vector of covariates and  $\varepsilon_i$  is a stochastic error term, the conditional expectation is  $E(y_i | x_i) = x_i' \beta$ , provided that  $E(\varepsilon_i | x_i) = 0$ . Likewise, the conditional quantile function  $Q_\theta(y_i | x_i) = x_i' \beta(\theta)$  if the  $\theta^h$  quantile of  $\varepsilon_i$  is zero. However, the coefficient vector  $\beta$  depends on the quantile  $\theta$ . Quantile regression, as introduced by Koenker and Basset (1978), is the solution to the following minimization problem:

$$\min_{\beta} \frac{1}{N} \left\{ \sum_{y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{y_i < x_i' \beta} (1-\theta) |y_i - x_i' \beta| \right\} \quad (4)$$

Given equation (4), no explicit expression exist for the estimators. Koenker and Basset (1978) showed that under some rather general conditions a unique solution of (4) exists. The minimization problem can be solved by linear programming (LP) techniques for the different quantiles of  $y$ . These methods are described in Koenker and D'Orey (1987) and Portnoy and Koenker (1997). When  $\theta = 0.5$ , the problem is minimizing the absolute value of the residuals, which is a median regression. By estimating different quantile regressions, it is possible to explore the entire shape of the conditional distribution of  $y$ , not just the mean, as in linear regressions. This implies that we can explicitly model the price and income reactions at different points in the conditional distribution of the demand function.

Quantile estimators are robust estimators, and are less influenced by outliers in the dependent variables than the least squares regression. When the error term is non-normal, quantile regression estimators may be more efficient than least squares estimators (Buchinsky, 1998). If the error terms are heteroscedastic, and the heteroscedasticity depends on the regressors, the estimated coefficients will have different values in the different quantile regressions. Potentially different solutions at distinct quantiles may be interpreted as differences in the response of the dependent variable to changes in the covariates at various points in the conditional distribution of the dependent variable. Quantile regressions are, like the OLS method, invariant to linear transformations.

Koenker and Basset (1982) introduced a formula for calculating the covariance matrix of the estimated parameters. However, in the Stata manual (StataCorp, 2001) it is argued that bootstrap methods give better estimates for the covariance matrix.

For a given set of prices, purchasing a product is partly a matter of income and partly a matter of taste. Zero observations are not necessarily the result of high prices or low incomes. When data is censored from below at zero, limited dependent variable models are often used. These models are dependent upon assumptions of normality and homoscedasticity in the error terms. Failure to fulfill these assumptions leads to inconsistent estimates of the parameters. Hurd (1979), Nelson (1981), and Arabmazar and Schmidt (1981) showed that estimating limited dependent variables with heteroscedasticity in the error terms leads to inconsistent parameter estimates. Goldberger (1983) and Arabmazar and Schmidt (1982) showed inconsistency because of non-normality in the error terms.

Powell (1984, 1986a) established that, under some weak regularity conditions, the censored quantile regression estimators are consistent and asymptotically normal, and that consistency of the estimators is independent of the distribution of the error terms. The only assumption is that the conditional quantile of the error term is zero:  $Q_\theta(\varepsilon_i | x_i' \beta) = 0$ .

One of the most useful properties of quantiles is that they are preserved under monotone transformations. For example, if we have a set of positive observations, and we take logarithms, the

median of the logarithm will be the logarithm of the median of the untransformed data. The censored regression model, where purchase is censored from below at zero, can be written as:

$$y_i = \max\{0, x_i'\beta + \varepsilon_i\}. \quad (5)$$

Because of the properties of the quantile function, the conditional quantile of this expression may be written as:

$$Q_\theta(y_i | x_i) = \max\{0, Q_\theta(x_i'\beta + \varepsilon_i | x_i)\} = \max(0, x_i'\beta) \quad (6)$$

when the conditional quantile of the error term is zero. Powell (1986a) shows that  $\beta$  can be consistently estimated by:

$$\min_{\beta} \frac{1}{N} \sum_{i=1}^n \rho_\theta \left[ y_i - \max\{0, x_i'\beta\} \right] \quad (7)$$

where  $\rho_\theta(\lambda) = [\theta - I(\lambda < 0)\lambda]$ .  $I$  is an indicator function which is equal to 1 when the expression is fulfilled and zero otherwise. For observations where  $x_i'\beta \leq 0$ ,  $\max(0, x_i'\beta) = 0$  and  $\rho$  is not a function of  $\beta$ . Hence, (7) is minimized using only those observations for which  $x_i'\beta > 0$ . Based on this fact, Buchinsky (1994) suggested an iterative LP algorithm in which the first quantile regression is run on all the observations, and the predicted values of  $x_i'\beta$  are calculated. These calculations are used to discard sample observations with negative predicted values. The quantile regression is then repeated on the truncated sample. The parameter estimates are used to recalculate  $x_i'\beta$  for the new sample, the negative values are discarded, and so on, until convergence. We have used this algorithm in combination with the qreg procedure in Stata.

The model estimated by quantile regression and censored quantile regression was compared with the model estimated by the SCLS method and the Tobit method. The SCLS estimation method proposed by Powell (1986b) is based on the ‘‘symmetric trimming’’ idea. If the true dependent variable is censored at zero and symmetrically distributed around  $x'\beta$ , we observe the dependent variable as asymmetrically distributed due to the censoring. However, symmetry can be restored by ‘‘symmetrically censoring’’ at  $2x'\beta$ . This is done below with the algorithm proposed in Johnston and DiNardo (1997). First, we estimate  $\beta$  using OLS on the original data. Then, we compute the predicted values. If the predicted value is negative, we set the observation to missing. If the predicted value of the dependent value is greater than twice the predicted value, we set the value of the dependent variable equal to  $2x_i'\beta$ . We then run OLS on these altered data. Finally, we repeat this procedure until convergence is achieved. The  $t$ -values were found by 100 bootstrap repetitions.

The Tobit model has the following likelihood function:

$$L = \prod_{y_i=0} \left[ 1 - \Phi \left( \frac{x_i'\beta}{\sigma} \right) \right] \cdot \prod_{y_i>0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \frac{(y_i - x_i'\beta)^2}{\sigma^2} \right],$$

where  $y$  is the left-side variable and  $x$  is the vector of right-side variables. To obtain estimates of the marginal effects that are comparable to the SCLS parameters, we have to multiply the parameter estimates with the probability of a positive outcome:  $\beta^* = \beta \Pr(y_i > 0)$ . We use the share with positive consumption, which is a consistent estimate of the probability.

## Data

The sample is obtained from the household expenditure surveys of Statistic Norway over the period from 1989 to 1999. Each year, between 1200 and 1400 households kept account of their

purchases over a two-week period. Thus, our total sample consists of about 14,000 observations. The households are evenly distributed throughout the year and throughout the country, so the data are representative. The surveys were conducted continuously, with new households participating every year, so our data consist of repeated cross-section samples. For food products, the quantities purchased and the corresponding expenditures are recorded. Table 1 shows the yearly per capita consumption of sugary carbonated soft drinks from 1989 to 1999. The years are in the first column. In the second column, the percentage of the sample with zero observations each year is presented. Then, the quantiles 0.25, 0.50, 0.75, 0.90, and 0.95 follow. The quantiles presented in the table are asymmetric to emphasize the high-consumption households. The mean values for each year follow the quantiles, and “Dis” is the yearly mean value of the disappearance data from the Breweries’ Association. We note that the mean value of the disappearance data is between 62 to 92 percent higher than the mean value in the survey data. One likely explanation for this difference is that many children do not report the whole quantity of soda purchase to their parents (who keep the accounts), and many adults forget to report the soda they buy at the gas station and similar places. “% Sug” is the share of the total carbonated soft drink sales that contain sugar.

The last row shows statistics from linear regressions, using year as the explanatory variable in each regression and the other columns as dependent variables. Trend is the parameter value, which measures the expected change in liters purchased from one year to another. We note that the share of the households that do not purchase sugary carbonated soft drinks is decreasing. The purchased quantity is increasing in all the quantiles, but the biggest increase is at the upper tail. All the trend parameters are significantly different from zero at the five percent level.

Table 1. Distribution of Annual per Capita Purchases of Sugary Carbonated Soft Drinks

Year	Zero%	Quantile					Mean	Dis	%Sug
		0.25	0.5	0.75	0.9	0.95			
1989	33	0	17	52	96	121	34		89
1990	35	0	16	49	87	124	33		83
1991	33	0	20	52	104	143	38		83
1992	24	1	26	62	100	135	41	73	82
1993	25	1	27	59	107	143	43	77	82
1994	21	7	33	72	121	156	49	95	85
1995	21	17	29	68	120	169	49	95	86
1996	20	9	39	75	126	181	54	96	86
1997	18	10	39	78	124	171	56	106	89
1998	19	8	36	74	117	163	51	101	89
1999	16	12	39	78	130	169	57	106	91
Trend (in liters)	-1.9	1.4	2.5	3.1	3.8	5.4	2.5	4.6	1.3

Note: The quantities are measured in liters per capita per year.

Dis = the mean value from the disappearance data.

% Sug = the percentage of sugary soda purchases in the total soda purchase.

While the expenditures are derived directly from the surveys, we used price variables derived from the CPI. Although we could have constructed unit prices, these would reflect quality as well as price variations. In addition, unit prices are missing for households that do not purchase any sugary soda. Because of these problems, we used the soda price sub index from the CPI as an explanatory variable. The CPI is a monthly Laspeyres index, where the sub indexes have fixed weights that are changed once a year according to the observed changes in budget shares. One problem with combining the survey data with the monthly price indices is that the survey period may involve two different months. We solved this problem in the following way. For the households keeping accounts within one month, we used the prices for that month. For the households keeping accounts in a period overlapping two months, we used a weighted average of the prices for the two months, using the number of days in the survey period in each month as weights.

To take account of the climatic conditions in Norway, with long winters and short summers, we introduce a temperature variable. We assume that when the temperature is above 15 degrees Celsius, people do more outdoor activities like sports, hiking, bathing, picnicking, and so on, thereby increasing the demand for soda. The temperature variable is constructed as the two-week mean temperature measured at the meteorological stations located in each of the six regions of Norway that are included in this study. These variables are linked to the households according to purchase time and place of abode. Further, we assume that temperatures below 15 degrees Celsius do not influence soda consumption. Therefore, the temperature variable has a value of one below 15 degrees Celsius, whereas above 15 degrees Celsius it has the value of the temperature.

Table 2. Average Values of Variables in Different Quantile Groups

Variable	Zero	Quantile Group					Mean
		0.25	0.50	0.75	0.90	1.0	
<u>Indexes</u>							
Soda consumption	0.0	0.0	0.7	1.9	3.6	7.4	1.9
Total expenditure	5.4	5.4	5.3	5.4	5.5	5.7	5.4
Price of soda	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Age (Year)	52.4	52.4	44.1	43.5	42.4	42.2	45.6
Temperature	2.0	2.0	2.1	2.4	2.7	2.8	2.3
<u>Dummy variables in %</u>							
Christmas	2.6	2.7	2.0	3.0	4.4	6.4	3.2
Screw cap	63.4	64.1	76.8	77.6	78.0	75.4	73.9
<u>Household type</u>							
One person	31.0	30.3	5.8	7.7	9.5	17.9	14.2
Couple without children	32.9	33.0	18.6	17.6	17.7	23.5	22.3
Couple with children	22.1	22.6	59.2	59.2	56.8	43.1	48.1
Single parent	3.5	3.6	5.2	4.8	4.7	4.9	4.6
Other household	10.4	10.5	11.2	10.7	11.5	10.5	10.9
<u>Region</u>							
Central East	20.5	20.6	20.4	18.2	18.0	20.2	19.5
Other East	26.6	26.4	26.1	27.5	30.7	31.1	27.7
South	14.4	14.6	14.7	15.0	13.7	12.6	14.4
West	17.7	17.7	18.2	19.1	18.0	16.8	18.1
Central	9.4	9.2	10.2	9.3	9.8	9.6	9.6
North	11.4	11.4	10.4	10.9	9.8	9.7	10.6
<u>Season</u>							
Winter	26.6	26.4	25.3	23.6	22.8	18.7	24.1
Spring	25.8	25.8	26.1	28.3	27.6	30.5	27.2
Summer	19.8	19.8	21.7	21.8	23.2	23.7	21.7
Fall	27.8	28.0	27.0	26.3	26.4	27.0	27.0

Table 2 shows the variables in categories corresponding to the quantile groups defined by the purchase of carbonated soft drinks. The quantile groups are defined according to the distribution of the dependent variable, measured by an index of per capita sugary soda expenditures divided by the soda price index. The “Zero” column shows the mean values for the households that did not purchase sugary soda in the survey period. The following five columns show the mean values for the quantile groups, and the last column gives the mean values of all the households. The 0.25 quantile group reports the mean values for the 25 percent of households with the lowest per capita sugary soda purchases, including the households in the “Zero” column. The 0.50 quantile group shows the mean values of the households having between 25 and 50 percent of the lowest sugary soda consumption, and so on. The “1” column shows the mean values for the 10 percent of households with the highest per capita consumption of sugary soda.

The first row in table 2 consists of the mean values of the dependent variable<sup>3</sup> in each quantile group. The next row shows the expenditure variable, which is the logarithm of the expenditure per capita deflated by the CPI. The third row lists the average soda price deflated with the CPI. The age of the head in each household and the temperature variable follow. The next variable is a Christmas dummy variable to account for the Christmas period. This variable has a value of one in the 26<sup>th</sup> two-week period and zero otherwise. In addition, we include a dummy variable to take account of the introduction of the 0.5 liter bottle with screw cap. Before 1992, soda was sold in small glass bottles containing just 0.33 liters of soda, with an iron cap. Thus, the likelihood of an open bottle being carried around was limited. This likelihood greatly increased after the introduction of the screw cap bottle. To model the combined effects of increased bottle size and the screw cap, we use a dummy variable taking a value of zero before 1992 and a value of one in 1992 and after. Finally, several dummy variables taking care of the household-specific characteristics, location, and time period are introduced.

We note that the expenditure variable is higher in the upper part of the distribution than at the mean. Next, the age of household heads declines gradually from the lower to the higher parts of the distribution. In addition, there are more households in the upper 10 percent during Christmas time, and there are fewer households consuming no sugary soda after the introduction of the new screw cap bottle than there were before. Further, one-person households are over-represented in the upper quantile groups, whereas couples with children are over-represented in the middle quantile groups.

## Results

Model (2) was estimated using Buchinsky's (1994) algorithm for censored quantile regression, implemented in Stata (StataCorp, 2001). From a health perspective, consumption of soda with sugar is of strong interest. The purchase of soda with sugar represents between 82 and 91 percent of the total soda purchase. We attempted to estimate a model involving all carbonated soft drinks – those with sugar and those with artificial sweetener. However, it turned out that the demand for soda with artificial sweetener was not very responsive to price. In addition, we obtained very unclear estimates for both total soda consumption and consumption of soda with artificial sweetener.

Table 3 shows the estimated parameters/marginal effects in five different quantile regressions, and the corresponding marginal effects of the SCLS and the Tobit models. In the 0.25-quantile regression, 26 percent of the observations were censored away. In the 0.5-, 0.75-, 0.90-, and 0.95-quantile regressions, the censoring did not have any effect, and the complete data sample was used. Consequently, we estimated the model simultaneously for these quantiles to take account of the possible correlation between the error terms. The marginal effects of the SCLS and the Tobit models are presented in the two rightmost columns.

The expenditure elasticity is significantly different from zero in all the quantiles, and it increases from 0.25 in the 0.25 quantile to 0.45 in the 0.95 quantile. The price elasticity is not significant in the lowest quantile, whereas at the median it is significant at the 10 percent level, and in all the other quantiles it is significant at the five percent level. The numerical value increases steadily from -0.62 in the 0.25 quantile to -1.60 in the 0.95 quantile. Age has a negative and significant effect in all the quantiles. Except for the lowest quantile, the effect is similar in all the quantiles. The temperature elasticity is about 0.06 in all the quantiles. This means that an increase in the two-week mean temperature from 18 to 19 degrees, which is an increase of 5.6 percent, will increase the demand for soda by 0.34 percent. Further, we can see that the introduction of new and larger bottles with screw caps increased consumption by between 8 and 11 percent. The consumption of carbonated sugary soft drinks shifts upward by about 30 percent in the two-week period around Christmas. Families with children is the reference household, the Central East region is the reference location, and winter is the reference quarter.  $R^2$  is low, which is common when cross-sectional data is used. In the last row, the number of observations for each quantile regression is printed.

We note that the comparable elasticities of the SCLS model are quite near the median in most cases, whereas the Tobit estimates are lower. In some cases, they are even lower than in the 0.25-quantile regression, indicating that the Tobit model is too restrictive.

<sup>3</sup> The dependent variable is in logarithmic form, after adding one to avoid  $\ln(0)$ . However, here it is shown untransformed



Table 3. Quantile Regression, SCLS and Tobit Estimates

Variable	Quantile					SCLS	Tobit
	0.25	0.50	0.75	0.90	0.95		
Total expenditure	0.25 (13.17)	0.31 (17.60)	0.38 (23.47)	0.43 (22.25)	0.45 (16.36)	0.31 (25.83)	0.27 (24.44)
Price of soda	-0.62 (-1.39)	-0.77 (-1.93)	-1.05 (-2.47)	-1.48 (-3.21)	-1.60 (-2.20)	-0.88 (-2.59)	-0.55 (-1.89)
Age	-0.16 (-4.80)	-0.37 (-11.47)	-0.38 (-12.14)	-0.35 (-9.00)	-0.32 (-7.49)	-0.35 (-11.67)	-0.33 (-18.39)
Temperature	0.06 (5.02)	0.07 (6.62)	0.06 (6.86)	0.06 (4.92)	0.06 (2.97)	0.06 (6.00)	0.05 (6.44)
Screw cap	0.11 (3.80)	0.11 (4.27)	0.10 (3.36)	0.10 (3.22)	0.08 (1.67)	0.11 (5.50)	0.10 (5.26)
Christmas	0.28 (6.01)	0.32 (5.57)	0.31 (5.84)	0.30 (6.05)	0.33 (5.02)	0.28 (7.00)	0.23 (7.25)
One person	-0.83 (-13.62)	-0.61 (-20.55)	-0.31 (-8.38)	-0.04 (-0.90)	0.09 (1.88)	-0.59 (-19.67)	-0.47 (-25.33)
Couple without children	-0.56 (-19.29)	-0.30 (-12.76)	-0.14 (-6.11)	-0.03 (-1.28)	-0.01 (-0.38)	-0.28 (-14.00)	-0.24 (-16.00)
Single parent	-0.14 (-4.04)	-0.16 (-4.65)	-0.06 (-1.32)	-0.04 (-0.90)	-0.01 (-0.10)	-0.14 (-4.67)	-0.12 (-4.86)
Other household	-0.23 (-8.66)	-0.06 (-2.05)	0.02 (0.73)	0.05 (1.64)	0.06 (1.62)	-0.05 (-2.50)	-0.05 (-2.68)
Other East	0.18 (6.76)	0.17 (6.75)	0.15 (6.30)	0.12 (4.42)	0.12 (3.17)	0.16 (8.00)	0.12 (7.52)
South	0.06 (2.11)	0.06 (1.94)	0.04 (1.29)	0.02 (0.67)	0.06 (1.26)	0.05 (2.50)	0.03 (1.49)
West	0.15 (5.43)	0.10 (3.98)	0.08 (3.44)	0.03 (1.11)	0.00 (0.07)	0.10 (5.00)	0.07 (4.16)
Central	0.16 (4.80)	0.12 (3.77)	0.11 (3.25)	0.08 (2.77)	0.02 (0.32)	0.13 (5.21)	0.09 (4.17)
North	0.13 (3.94)	0.09 (3.10)	0.05 (1.61)	0.02 (0.50)	0.05 (0.90)	0.09 (3.58)	0.05 (2.32)
Spring	0.07 (2.96)	0.08 (3.69)	0.10 (4.71)	0.11 (4.21)	0.10 (3.04)	0.08 (4.71)	0.07 (4.51)
Summer	0.04 (1.52)	0.01 (0.63)	0.03 (1.28)	0.06 (2.26)	0.05 (1.21)	0.03 (1.29)	0.03 (1.55)
Fall	-0.03 (-1.01)	-0.02 (-0.64)	-0.01 (-0.51)	0.01 (0.27)	-0.03 (-0.85)	-0.02 (1.20)	-0.02 (-0.97)
Constant	-0.50 (-3.21)	0.41 (2.61)	0.51 (3.53)	0.48 (2.60)	0.54 (2.48)	0.36 (2.93)	0.31 (3.21)
R <sup>2</sup>	0.03	0.08	0.06	0.07	0.08	0.21	0.06
# observations	10282	13985	13985	13985	13985	13985	13985

Note: The  $t$ -values are reported in parentheses.

Figure 1 presents the estimates for some of the most important of the quantile elasticities and the corresponding SCLS elasticities. For the expenditure elasticity, the price elasticity, and the age elasticity, we plot the different quantile regression results for 0.25, 0.50, 0.75, 0.90, and 0.95, with the solid curves representing the 90 percent confidence band. The dashed lines represent the SCLS estimates with the 90 percent confidence band. In all the panels, the quantile regression estimates lie at some points outside the confidence interval for the SCLS model, suggesting that the effects of these covariates are not constant across the conditional distribution of the dependent variable.

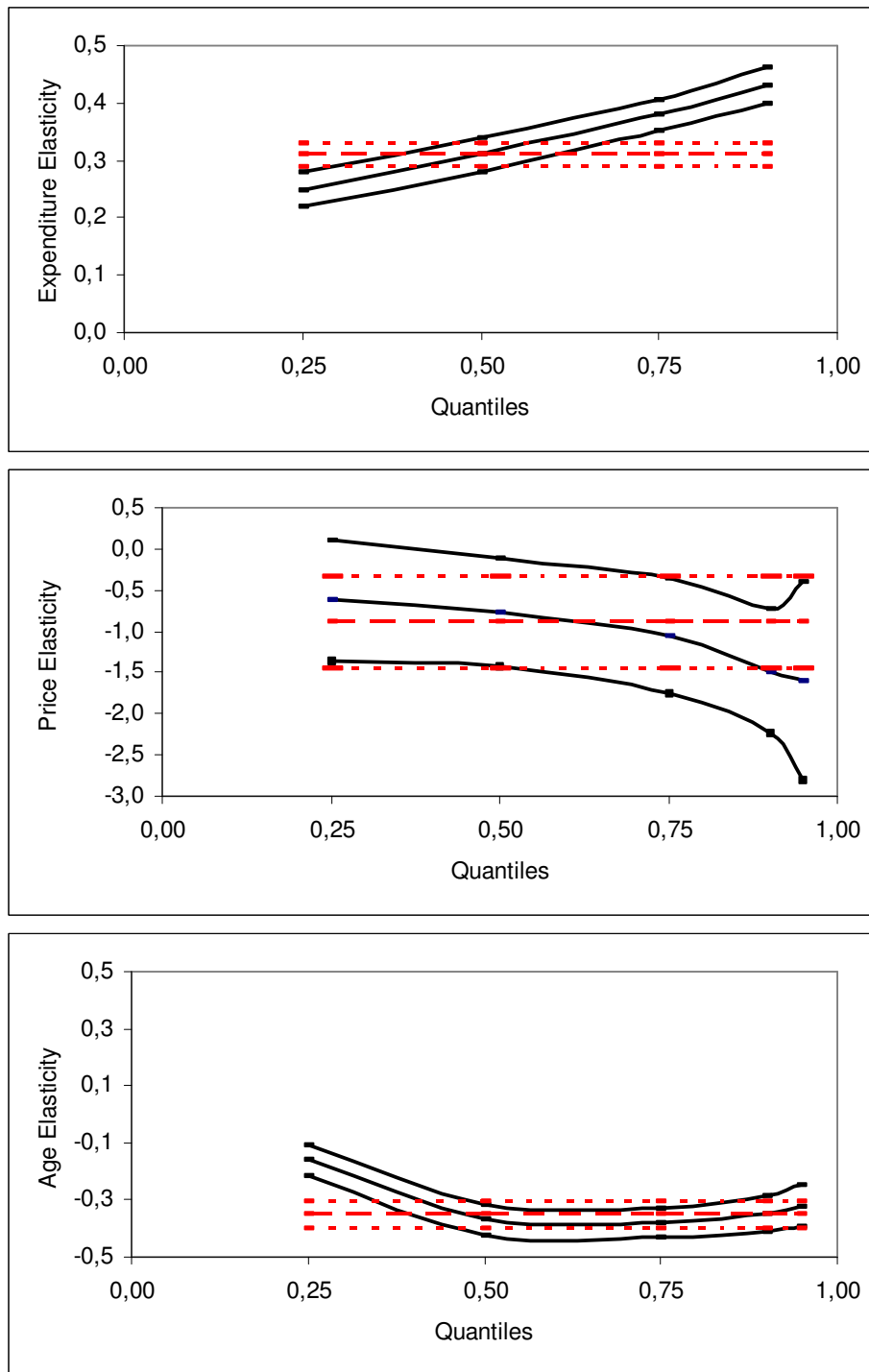


Figure 1. Quantile Regression SCLS Estimates with 90 Percent Confidence Intervals

Results from statistical tests for equality of coefficients across the estimated quantiles are presented in table 4. When one or both the quantile regressions are censored, different parts of the sample are used for estimation, and we cannot obtain the covariance between the regressions. In these cases, we calculate quasi  $t$ -statistics to test for equality between the coefficients. The quasi  $t$ -statistics ignore any covariance between the coefficients. The first three columns of table 4 give the quasi  $t$ -statistics for equality tests of the coefficients at the 0.25 quantile, with the coefficients at the 0.75,

0.90, and 0.95 quantiles. If the numerical value of the  $t$ -statistics is larger than 1.96, then equality is rejected at the five percent level of significance. As discussed above, censoring was not a problem at the 0.50, 0.75, 0.90, and 0.95 quantiles. Therefore, these equations were estimated simultaneously, and the covariance matrix between the coefficients was calculated by bootstrapping. In the last three columns of table 4, the  $t$ -statistics of tests for equality between coefficients at the 0.50, 0.75, 0.90, and 0.95 quantiles are reported.

Table 4. Tests for Equality of Coefficients across Quantiles

	$q_{25} = q_{75}$	$q_{25} = q_{90}$	$q_{25} = q_{95}$	$q_{50} = q_{90}$	$q_{50} = q_{95}$	$q_{75} = q_{95}$
Total Expenditure	-5.15*	-7.00*	-6.45*	5.80*	4.47*	2.29*
Price of soda	0.68	1.28	1.24	1.44	1.17	0.82
Age	5.09*	4.17*	2.86*	0.37	1.17	1.68
Temperature	-0.09	-0.13	0.14	0.37	0.49	0.22
Screw cap	0.26	0.24	0.68	0.24	0.67	0.56
Christmas	-0.45	-0.24	-0.64	0.32	0.24	0.36
One person	-7.86*	-11.67*	-12.50*	12.90*	12.92*	8.36*
Couple without children	-11.27*	-13.34*	-11.68*	8.74*	9.28*	4.99*
Single parent	-1.60	-1.91	-2.00*	2.72*	2.22*	0.86
Other household	-6.41*	-6.85*	-5.82*	3.06*	2.88*	1.27
Other East	0.68	1.38	1.14	1.47	1.11	0.69
South	0.65	0.96	0.13	0.91	0.00	0.44
West	1.84	2.92*	3.08*	2.03*	2.35*	2.13*
Central	1.16	1.74	2.51*	1.03	1.73	1.78
North	1.76	2.36*	1.48	2.05*	0.76	0.10
Spring	-0.85	-1.27	-0.75	1.16	0.50	0.10
Summer	0.31	-0.48	-0.17	1.62	0.84	0.50
Fall	-0.40	-0.89	0.06	0.79	0.35	0.51
Constant	-4.79*	-4.41*	-3.84*	0.36	0.51	0.14

Note: An asterix indicates significance at the five percent level.

The tests reject the  $H_0$  hypothesis of equality for all the expenditure elasticities. For the price elasticities, however, the  $H_0$  hypotheses are not rejected between any of the quantiles. Further, the tests suggest that the age elasticity is less in the 0.25 quantile than in the other quantiles. For the temperature, the tests suggest that the effect is similar in all parts of the distribution. This is also true for the effect of the introduction of larger bottles with screw caps, and for the effect of Christmas. The differences of single households (relative to couples with children) vary across the distribution. The same is true for couples without children and other households as compared with the reference group.

These tests indicate that the effects of many of the covariates are different in different parts of the conditional distribution of soda consumption. Hence, a quantile regression approach is warranted.

### The Effects of Public Policies

The demand for carbonated soft drinks containing sugar may continue to increase if nothing is done to prevent it. Unless younger people completely change their attitudes as they age, the negative age elasticity indicates that consumption will increase. The positive expenditure elasticity, together with the steadily growing real household income, will also contribute to growing consumption.

Public authorities have several options for influencing the demand for soda. First, they could ban the sale of soft drinks in schools. Furthermore, they could restrict school children from going outside the school area during school time. Second, as with smoking and drinking, information about the health aspects of soda consumption may be used to prevent further increases in consumption. Last, but not least, economic means may be used to reduce the demand for sugary drinks, either by influencing the income of the households and/or the prices of the products. The disadvantage of influencing

household income, for example by income taxes, is that it will have an effect on the consumption of all goods, healthy or unhealthy. Hence, it is better to use prices to influence the consumption.

In Norway, carbonated soft drinks are exposed to a production tax of NOK<sup>4</sup> 1.55 per liter. In addition, soft drinks have a value added tax (VAT) of 12 percent, which is the same as for other food products. Most non-food products have a 24 percent VAT. We will study three price scenarios for sugary carbonated soft drinks. In the first scenario, we use the elasticities from the quantile regression model and the SCLS model to calculate the effects of a doubling of the VAT. This means a price increase of 10.8 percent. In the second scenario, we calculate the impact of doubling the production tax as well as doubling the VAT. This corresponds to a price increase of 27.3 percent. In the third scenario, we study the effect of Swedish prices in Norway. According to Statistics Norway and Eurostat, the European purchase parity survey (Bruksås et al., 2001) shows that Swedish soda and juice prices are about 29.8 percent lower than Norwegian prices. However, the general price level is about 10.4 percent lower, and, correspondingly, the real soda price level is about 21.7 percent lower in Sweden than in Norway. We assume that Norwegian soda prices decrease down to the Swedish level, which may occur if Norway joins the EU. Table 5 shows the results from the three price scenarios in percentages and liters. Purchases in 1999 are used as a base level to calculate the changes in liters.

Table 5 shows that the percentage effects are largest in the upper quantiles. Furthermore, the changes in liters are even larger in the upper quantiles. If the objective is to reduce consumption among the heavy soda consumers, price changes seem to be an effective tool. A doubling of production tax and the VAT will reduce the consumption of the top five percent of soda consumers by approximately 44 percent, or 74 liters per year. The lowest soda consumers will reduce their consumption by 17 percent, or about two liters per year. The mean effects are calculated using the SCLS elasticities. They are between the median and the 0.75 quantile in all the scenarios, which is reasonable. To find the effects of a price change on the zero-consumption households, we estimated a binary logit model. The own-price parameter was very small and insignificant. Hence, we believe that price changes will not have any effect on the zero-consumption households.

Table 5. Predicted Annual Changes in Soda Purchases per Capita due to Price Changes

Policy Change	Quantile					SCLS
	0.25	0.50	0.75	0.90	0.95	
<b>Doubling of VAT for soda</b>						
Change in percent	-6.7	-8.3	-11.3	-16.0	-17.3	-9.5
Change in liters	-0.8	-3.2	-8.8	-20.8	-29.2	-5.1
<b>Doubling of VAT and production tax for soda</b>						
Change in percent	-16.9	-21.0	-28.7	-40.0	-43.7	-24.0
Change in liters	-2.0	-8.2	-22.4	-52.5	-73.8	-12.9
<b>Swedish prices in Norway</b>						
Change in percent	13.5	16.7	22.8	32.1	34.7	19.1
Change in liters	1.6	6.5	17.8	41.8	58.7	10.2

### Concluding Remarks

Our analysis investigates the demand for sugary carbonated soft drinks and how the authorities may influence consumption. Steady increases in consumption of soft drinks have been observed for many years. Until recently, studies have focused on average values, but because heavy consumption of sugary soft drinks contributes to obesity and other health issues, the focus should be on heavy consumption. Moderate or low consumption is of less concern.

The results show that many of the covariates have different effects in different parts of the conditional distribution, warranting a quantile regression approach. Heavy drinkers are more

<sup>4</sup> The exchange rate from the Central Bank of Norway is currently US\$1 = 6.27 NOK (January 25, 2005).

expenditure-responsive than light drinkers are, whereas age seems to be more important at and above the median than below it. While the expenditure effect is positive, the age effect is negative. This means that the trend towards increasing consumption of sugary soda will continue if young people do not drastically change their habits when they grow older. Steady growth in incomes and the consumption trend will almost surely continue, pushing soda consumption higher, with the highest growth in the upper quantiles.

High temperature increases consumption, and has a similar effect on sugary soda consumption in all the quantiles. Due to the change in the bottle type, from the 0.33 liter glass bottle with an iron cap to the 0.5 liter plastic bottle with a screw cap, the demand shifted upwards by about 10 percent in all quantiles.

The study shows that a doubling of the production tax and the value added tax will reduce the consumption of sugary soda by two liters per year for the moderate consumers and by 74 liters per year for those in the top five percent in terms of consumption.

## References

- Arabmazar, A. and Schmidt, P., (1981). Further evidence on the robustness of the Tobit estimator to heteroscedasticity. *Journal of Econometrics* 17: 253-258.
- Arabmazar, A. and Schmidt, P. (1982). An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica* 50: 1055-1063.
- Bruksås, N., Myran, K. and Svennebye, L.H. (2001). *Prisnivå på matvarer i de nordiske land, Tyskland og EU*. Statistics Norway.
- Buchinsky, M. (1994). Changes in the US wage structure 1963-1987 – Application of quantile regression. *Econometrica* 62: 405-458
- Buchinsky, M. (1998). Recent advances in quantile regression models – A practical guideline for empirical research. *The Journal of Human Resources* 33: 88-126.
- Deaton, A. (1997). *The Analysis of Household Surveys – A Microeconomic Approach to Development Policy*. The John Hopkins University Press, Baltimore and London.
- Deaton, A. and Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge University Press, New York.
- Goldberger, A.S. (1983). Abnormal Selection Bias, in: Karlin, S., Amemiya, T. and Goodman, L. (eds.), *Studies in Econometrics, Time Series and Multivariate Statistics*, New York, Academic Press, 67-84.
- Hurd, M. (1979). Estimation in truncated samples when there is heteroscedasticity. *Journal of Econometrics* 11: 247-258.
- Johnston, J. and DiNardo, J. (1997). *Econometric Methods*. New York: McGraw-Hill.
- Moschini, G. (1995). Units of measurement and the Stone index in demand system estimation. *American Journal of Agricultural Economics* 77: 63-68.
- Koenker R. and Bassett jr., G. (1978). Regression quantiles. *Econometrica* 46: 33-50.
- Koenker, R. and D'Orey, V. (1987). Computing regression quantiles. *Journal of the Royal Statistical Society, Applied Statistics* 36: 383-393.
- Manning, W.G., Blumberg, L. and Moulton, L.H. (1995). The demand for alcohol: The differential response to price. *Journal of Health Economics* 14: 123-148.
- Nelson, F.D. (1981). A test for misspecification in the censored normal model. *Econometrica* 49: 1317-1329.
- Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared error versus absolute-error estimators. *Statistical Science* 12: 279-300.
- Powell, J.L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25: 303-325.
- Powell, J.L. (1986a). Censored regression quantiles. *Journal of Econometrics* 32: 143-155.
- Powell, J.L. (1986b). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* 54: 1435-1460.
- Stewart, H., Blisard, N. and Jolliffe, D. (2003). Do income constraints inhibit spending on fruits and vegetables among low-income households? *Journal of Agricultural and Resource Economics* 28: 465-480.
- StataCorp (2001). *Stata Statistical Software: Release 7.0*. College Station, Texas: Stata Corporation.

- Variyam, J.N., Blaylock, J. and Smallwood, D. (2002). Characterizing the distribution of macronutrient intake among U.S. adults: A quantile regression approach. *American Journal of Agricultural Economics* 84: 454-466.
- Variyam, J.N. (2003). Double impact: Educational attainment and the macronutrient intake of US adults. in Chern. W.S. and Rickertsen, K. (eds.), *Health, Nutrition and Food Demand*. Wallingford: CABI Publishing, 53-72.