IAW-Diskussionspapiere

5

A Microeconometric Characterisation of Household Consumption Using Quantile Regression

Niels Schulze
Gerd Ronning

April 2002

ISSN: 1617-5654

Institut für Angewandte Wirtschaftsforschung
Ob dem Himmelreich 1
72074 Tübingen

T: (0 70 71) 98 96-0
F: (0 70 71) 98 86-99
E-Mail: iaw@iaw.edu
Internet: www.iaw.edu

A MICROECONOMETRIC CHARACTERISATION OF HOUSEHOLD CONSUMPTION USING QUANTILE REGRESSION

Niels Schulze and Gerd Ronning
Eberhard-Karls-Universität Tübingen
Mohlstraße 36, D-72074 Tübingen

Key Words: econometric demand analysis, Engel curve, demand system, micro data
JEL Classification: C21, C49, D12

ABSTRACT

The paper uses micro cross-section data from the GfK consumer panel for econometric demand analysis of private households in Germany. Contrary to most research which considered "average" behavior we extend this approach to consumer behavior for different "intensities" of consumption. Our analytical tool is quantile regression which allows us to describe the conditional distribution for any quantile including the (conditional) median representing "average" behavior. As an illustrative example we use the demand for beer and wine. The paper shows quite distinct patterns regarding price and income effects for different goods which leads us to an extended characterization of household demand.

1 INTRODUCTION

Econometric demand analysis played an important role in the 70's and the 80's: Christensen, Jorgenson and Lau (1975) proposed the "Translog demand system" and Deaton and Muellbauer (1980) introduced an alternative specification termed "Almost Ideal Demand System" (AIDS). Both approaches based their empirical analysis on the aggregate time series data from a sample of households continuing the work started by Richard Stone who established the "Linear Expenditure System" quite a while earlier (Stone 1954). Many studies failed in trying to verify the constraints like symmetry and Engel and Cournot aggregation established by (static) microeconomic theory. Econometric issues arose from the fact that both translog and AIDS were formulated in terms of (dynamic) share equations; the implied problems have been solved only marginally. See Ronning (1992) for an overview.

On the other hand, surprisingly little work has been done in consumption analysis on the basis of individual cross-section (or panel) household data. This fact is even more worth mentioning when comparing it with a huge bulk of microeconometric studies involved in qualitative choice behavior initiated by Daniel McFadden who studied the structure of travel demand. Exceptions are, among others, the microeconometric studies based on the British family expenditure survey (see for example Atkinson, Gomulka and Stern (1990) and Blundell, Pashardes and Weber (1993)). Deaton (1997) is the most recent example for this kind of research who studies the consumption pattern in underdeveloped countries.

All research so far has concentrated on "average" behavior, i.e. on the expected value of the conditional distribution. Our paper is concerned with a more detailed description of characteristics of this (conditional) distribution; in particular we consider quantiles of consumption which means that we not only consider average behavior but also behavior of "extreme" consumers thereby installing a new characterization of consumer behavior.[1] Fox example it might well happen for some good that extreme consumers' demand elasticities differ in sign from the one shown by the average consumer: The study by Manning, Blumberg and Moulton (1995, p. 125) on the demand for alcohol tries to find out "...whether light and heavy drinkers have different price responses". In our paper we try to give an explanation for this varying demand structure which to our best knowledge so far has only be noted - in a quite different context - by the paper just mentioned. We use data from the German GfK consumer panel[2] for the year 1995. Special attention is given to the kind of temporal aggregation employed in order to include all purchases within this period.

The paper is organized as follows: In section 2 we establish or rather report some basic results from empirical demand analysis needed when interpreting later on our own empirical results.

[1]Former studies have tried to provide such information by evaluating the price elasticities at certain quantiles of the dependent variables. See, for example, Blundell et al (1993) table 3 or - as a most recent example - Newey (2001) table 3.

[2]GfK = Gesellschaft für Konsumforschung, Nuremberg/Germany.

Section 3 contains a short description of quantile regression first introduced by Koenker and Basset (1978). We also mention the main features of the software package used for quantile regression. Section 4 describes the data. In section 5 we report on our own results from which we move to some general statements regarding consumer behavior. Section 6 adds some concluding remarks.

2 ECONOMETRIC DEMAND ANALYSIS

The main concern of econometric demand analysis is with consumer's reactions to changes in prices and income[3] as described, for example, in Varian's textbook on microeconomic theory (Varian 1992 section 3.3). Two main approaches have been used:[4]

- analysis of Engel curves for certain goods or groups of goods, that is the relation between consumption and income for a certain good (group of goods). From this analysis the income elasticity may be deduced.
- estimation of demand systems (LES, Translog, AIDS, generalized Leontieff) using information on prices and quantities for all goods (groups of goods). Only this approach allows an adequate examination of substitution patterns between goods.

In the following we list some of the most important topics and problems arising from econometric demand analysis:

(a) Aggregate data versus micro data It has already been pointed out in the introduction that most of the work concerning estimation of demand systems has been done on the basis of (monthly or yearly) aggregate data for some population. Typically consumption shares for a moderate number of good categories have been analyzed over time. See for example, the pioneering papers by Christensen et al (1975) and Deaton and Muellbauer (1980a). Most of the studies failed to verify the restrictions postulated by (static) microeconomic theory.

[3]The table displays the possible cases. Normal goods are further separated into luxuries when income elasticity is greater than one and necessities if income elasticity is smaller than one. Two goods are called substitutes if the cross-price elasticity is positive and complements if it is negative.

		change of income μ	
		"normal good" $\partial q_j/\partial \mu > 0$	"inferior good" $\partial q_j/\partial \mu < 0$
change of price p_j	"ordinary good" $\partial q_j/\partial p_j < 0$	example: butter	example: margarine
	"Giffen good" $\partial q_j/\partial p_j > 0$		example: potatoes in Ireland, 19-th century

[4]See Ronning (1988) for an overview.

About a decade ago Richard Blundell and others (Blundell et al 1993) posed the question: "What do we learn on consumer demand patterns from micro data?" Their paper shows - at least for the data set used - that estimation on the basis of micro data much better fulfills the demand restrictions (Blundell et al 1993 p. 577), a fact which to our best knowledge has not been appropriately noted in the literature. On the other hand each model based on micro data has to relate its results to the macro level thereby facing the problem of aggregation. This aspect has been treated, too, in the paper by Blundell et al (1993).[5]

(b) Income versus expenditure (consumption versus purchase) Ideally consumption for all goods and services should be included into the analysis. However, usually only a subgroup has been considered so that it is unclear how "income" has been distributed between this subgroup and the remaining goods. Therefore the expenditures spent on this subgroup is used instead. Moreover, Keen (1986) has stressed the important distinction between consumption and purchase of a good: "Zero consumption" can only be defined via observed purchases. If an household buys a good infrequently, we speak of zero consumption although it is not clear whether the good really is not consumed. See Ronning (1988 p. 71) and Blundell et al (1993 p. 575).

(c) Price information Prices of most goods will not vary over individuals in a cross-section. This is an argument in favor of disregarding price effects in the analysis of Engel curves where typically cross-section data are used. On the other hand it complicates the estimation of price effects on consumption, especially for goods with regulated prices ("Preisbindung zweiter Hand", for example). The situation is improved when panel data are available. If groups of goods are used in microeconometric research, then prices have to be aggregated (see, for example, Blundell et al 1993). The data set used in this paper contains information for single households over one year indicating expenditures and quantities for each single purchase. However, prices have to be derived from these data. As we shall explain in section 4, prices have been aggregated over the whole year to make the econometric demand analysis possible since other variables, in particular income, are only given on a yearly basis.

Manning et al (1995) have pointed out that consumers of a certain good (and facing the same income and prices) may behave quite differently depending on the *intensity of consumption*. In their study on the demand for alcohol they find remarkable differences between "heavy" and "light" drinkers with respect to the own-price elasticity and income elasticity. For example, the same good is considered as "inferior" by some consumers and "normal" by others. Another example is given by Koenker and Hallock (2001) where different Engel curves are shown for "heavy" and "light" consumption implying varying income elasticities for these subgroups.[6] In both cases the

[5] Ronning and Zimmermann (1991) give an introduction to a series of papers in "ifo-Studien" on the relevance of microeconometric models for economic policy.

[6] The paper uses the original data from Engel's study on food expenditures.

method of quantile regression first proposed by Koenker and Basset (1978) is employed which will be presented in the next section.

3 QUANTILE REGRESSION

Given a random variable Y with right continuous distribution function $F_Y(a) = P(Y \leq a)$, the *quantile function* Q_Y can be defined by

$$\begin{aligned} [0,1] &\mapsto \mathbb{R} \\ Q_Y(\theta) &= F_Y^{-1}(\theta) = \inf \{a \mid F_Y(a) \geq \theta\} \end{aligned} \tag{1}$$

Similarly, taking a random sample $Y_1, Y_2, \ldots, Y_n$ with empirical distribution function $\hat{F}_Y(a) = \frac{1}{n} \#\{Y_i \leq a\}$, we define the *empirical quantile function*

$$\hat{Q}_Y(\theta) = \hat{F}_Y^{-1}(\theta) = \inf \{a \mid \hat{F}_Y(a) \geq \theta\} \tag{2}$$

Koenker and Bassett (1978) showed the equivalence to the following minimization problem[7]:

$$\begin{aligned} \hat{Q}_Y(\theta) &= \underset{a}{\operatorname{argmin}} \left\{ \sum_{i:Y_i \geq a} \theta |Y_i - a| + \sum_{i:Y_i < a} (1-\theta)|Y_i - a| \right\} \\ &= \underset{a}{\operatorname{argmin}} \sum_i \rho_\theta(Y_i - a) \end{aligned} \tag{3}$$

$$\text{with} \quad \rho_\theta(z) = \begin{cases} \theta z & : z \geq 0 \\ (\theta - 1)z & : z < 0 \end{cases} = (\theta - I(z<0))\, z$$

If the median ($\theta = 0.5$) is taken, (3) simplifies to the well-known expression

$$\widehat{\text{med}}_Y = \underset{a}{\operatorname{argmin}} \sum_i |Y_i - a| \tag{4}$$

Assuming that Y is linearly dependent on a vector of exogenous variables x, the *conditional quantile function* can be written as

$$\begin{aligned} Q_Y(\theta|x) &= \inf\{a \mid F_Y(a|x) \geq \theta\} \\ &= \sum_k x_k \beta_{\theta k} = x'\beta_\theta \end{aligned} \tag{5}$$

[7]See Koenker and Bassett (1978), page 38. Their own comment says "The case of the median (θ=1/2) is, of course, well known, but the general result has languished in the status of curiosm."

In analogy to (3), we finally obtain the regression quantiles by solving with respect to β_θ

$$\begin{aligned}\hat{\beta}_\theta &= \underset{\beta_\theta \in \mathbb{R}^k}{\operatorname{argmin}} \left\{ \sum_{i:Y_i \geq x'\beta_\theta} \theta |Y_i - x_i'\beta_\theta| + \sum_{i:Y_i < x'\beta_\theta} (1-\theta)|Y_i - x_i'\beta_\theta| \right\} \\ &= \underset{\beta_\theta}{\operatorname{argmin}} \sum_i \rho_\theta(Y_i - x_i'\beta_\theta) \end{aligned} \tag{6}$$

There does not exist a general closed solution to the minimization problem, but after some slight modifications it can easily be solved by simplex methods. Barrodale and Roberts (1974) developed an algorithm for the median case. Koenker and d'Orey (1987,1993) described an implementation for the general quantile problem with desirable properties for small to medium number of observations. Portnoy and Koenker (1987) showed that an alternative interior method published by Koenker and Park (1996) is competitive to least-squares estimation even for very large data sets.

Most of the computations conducted in this paper have been calculated with the software package STATA. Apart from the evaluation of the desired quantile coefficients, the program also allows to obtain appropriate confidence intervals by the means of bootstrapping.[8] Furthermore, some of our results have been verified by MATLAB routines based on the algorithms provided by David Hunter at the webpage `www.stat.psu.edu/~dhunter/qrmatlab/`.[9]

In the following, some important properties of the quantile estimation process are briefly described:[10]

- Quantile regression reveals information about the complete conditional distribution of the response variable. No constraints on the error term are imposed.
- The estimation is *robust* against outliers. In other words, every observation can be made arbitrarily big (or small) without changing the result unless it does not cross the estimated (hyper-)plane.
- Using censored figures (e.g. topcoded income data), censoring does not distort the outcome as long as the censoring point is not reached (e.g. analysis of lower income classes).
- In several cases (e.g. Cauchy or some mixed normal distribution), quantile regression is more efficient than least-squares estimation.

[8] For questions concerning the implementation as well as the utilized algorithms see Stata Corporation (2001).
[9] Another software archive for quantile regression is maintained by Roger Koenker and can be retrieved at `www.econ.uiuc.edu/~roger/research/rq/rq.html`
[10] A more comprehensive discussion can be found e.g. in Buchinsky (1998) or Koenker and Hallock (2001).

- Regression quantiles are equivariant to the following transformations:

$$\hat{\beta}(\theta, \lambda y, X) = \lambda\hat{\beta}(\theta, y, X) \qquad \lambda \in [0, \infty[\tag{7}$$
$$\hat{\beta}(\theta, -\lambda y, X) = \lambda\hat{\beta}(1-\theta, y, X) \qquad \lambda \in [0, \infty[\tag{8}$$
$$\hat{\beta}(\theta, y + X\gamma, X) = \hat{\beta}(\theta, y, X) + \gamma \qquad \gamma \in \mathbb{R}^k \tag{9}$$
$$\hat{\beta}(\theta, y, XA) = A^{-1}\hat{\beta}(\theta, y, X) \qquad A \text{ nonsingular} \tag{10}$$

  Furthermore, the quantile function is invariant to any monotone transformation of the dependant variable ($Q_{h(Y)}(\theta|x) = h(Q_Y(\theta|x))$).

- The method is asymptotically consistent:

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \quad \xrightarrow{n\to\infty} \quad N(0, \Lambda_\theta) \tag{11}$$

$$\text{with } \Lambda_\theta \;=\; \theta(1-\theta)\;(E[f_{u_\theta}(0|x_i)x_i x_i'])^{-1}\; E[x_i x_i']\;(E[f_{u_\theta}(0|x_i)x_i x_i'])^{-1}$$

  If the error term u is distributed independently of the covariates x ($f_{u_\theta}(0|x) = f_{u_\theta}(0)$), the variance-covariance-matrix simplifies to

$$\Lambda_\theta \;=\; \theta(1-\theta)\,\frac{(E[x_i x_i'])^{-1}}{f^2_{u_\theta}(0)} \tag{12}$$

4 THE DATA

Our empirical analysis is based on data from the *ConsumerScan* household panel maintained by "Gesellschaft für Konsumforschung" (GfK) since 1957. The panel currently consists of about 12,000 households constantly reporting their purchases of *Fast Moving Consumer Goods* on an individual buying basis. A subset of this data set is available for scientific use from Zentrum für Umfragen, Methoden und Analysen (ZUMA) at Mannheim.[11] However, this file is confined to the year 1995.

The ZUMA data set contains all 9,064 households continuously reporting their purchases in 1995. The products are divided into 81 categories, that is, no brand names are given. For each individual purchasing act the following information is collected: date, day of the week, subcategory chosen (within the goods group), type of retailer, product identification number, type of price (normal/special), total quantity, amount spent, time since last buying. Other specific characteristics of the products are given as well (for example, packaging). For some product categories only a subsample (4,426 respectively 4,638) of households has been reporting.

In table 1 which provides information for a number of groups this is indicated in the fourth column. Moreover, the table contains information about the total number of purchases from

[11] A detailed description of the provided data can be found in Papastefanou (2001).

Table 1: Purchasing data

ident.	product	purchasing acts	number of households	purchasing frequency (av.)	proportion of non-buyers (%)
04	detergents for dishes	34556	7780	4.44	14.17
08	milk	204339	*4185	48.83	5.45
12	pure coffee (roasted)	143194	8457	16.93	6.70
17	frozen food	230841	8175	28.24	9.81
22	fats	233124	*4617	50.49	0.45
33	beer	131245	7485	17.53	17.42
35	wine	27614	2954	9.35	67.41
46	lemonade	155447	7254	21.43	19.97
66	animal food	123133	3056	40.29	66.28
84	mineral water	174470	8414	20.74	7.17
91	pasta	53201	*4165	12.77	10.20
99	toilet paper	32435	*4039	8.03	8.74

Note: An asterisk indicates that only a subsample of the households is captured

which we estimated the average purchasing frequency: For example, households buy about 49 times milk and about 21 times mineral water during the year. Additionally the table reports the proportion of households not buying from a certain product group. This ranges from 0.45% not buying fats to 67.41% not buying wine and thereby illustrating the fact of "zero consumption" discussed in section 2.

Table 2: Distribution of income

net income	number	percentage	accumulated
up to 499 DM	24	0.26	0.26
500 DM - 999 DM	189	2.09	2.35
1000 DM - 1249 DM	312	3.44	5.79
1250 DM - 1499 DM	366	4.04	9.83
1500 DM - 1999 DM	890	9.82	19.65
2000 DM - 2499 DM	1235	13.63	33.27
2500 DM - 2999 DM	1233	13.60	46.88
3000 DM - 3499 DM	1235	13.63	60.50
3500 DM - 3999 DM	907	10.01	70.51
4000 DM - 4499 DM	852	9.40	79.91
4500 DM - 4999 DM	484	5.34	85.25
5000 DM - 5499 DM	474	5.23	90.48
5500 DM and more	829	9.15	99.62
not reported	34	0.38	100.00
total	9064	100.00	

Furthermore, some socioeconomic and demographic information is provided on a household basis. This information includes federal state, size of community, age of head of household, number of children in different age groups (up to 6, 7-14 and 15-18 years), income, occupational and educational status. Moreover, information about equipment of the household is provided. For some of the households these variables are missing. A major drawback of this data set is the fact that income and age are reported only in grouped form. Table 2 shows the number of missing values as well as the grouping for the variable household income which plays a central role in demand analysis. Finally, the GfK panel reports on attitudes of consumers. These attitudes concern e.g. nutritional, environmental and other aspects of daily life, but are not used in the study.

For our empirical analysis some aggregation of the data is required. This results from the fact that we have a detailed information on each purchase regarding quantity and amount spent from which we can deduce the price per unit by computing for each product

$$\text{price} \quad = \quad \frac{\text{amount spent}}{\text{quantity}} \quad .$$

However, we have - of course - only one observation regarding income for each household. Therefore we determine the yearly average price for each household by computing for each product

$$\text{average price} \quad = \quad \frac{\text{total amount spent by a household on this product within the year}}{\text{total quantity within the year}} \quad .$$

In the following we call this derived average price simply "the price". In order to illustrate the effect of our aggregation approach, we show in figure 1 the distribution of quantity, amount spent and price (per purchase) for the case of beer. Note the peaks at 10 and 20 liters whereas the price distribution is rather well behaved. The figures 2 and 3 then display the result of aggregation in two scatter diagrams (double linear scale and double log scale) for price and quantity where each point represents one household.

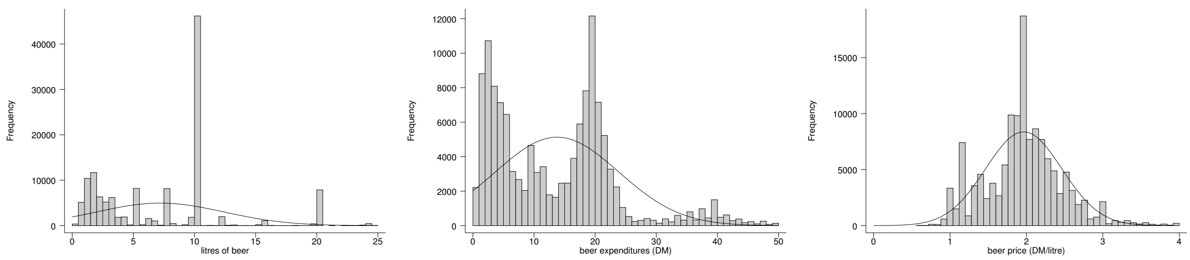

Figure 1: Beer consumption (quantity, expenditures and price)

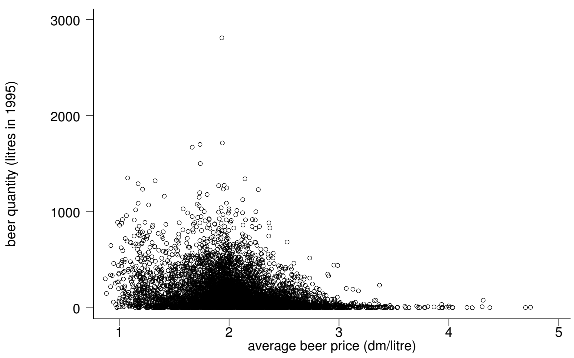

Figure 2: Aggregated beer data

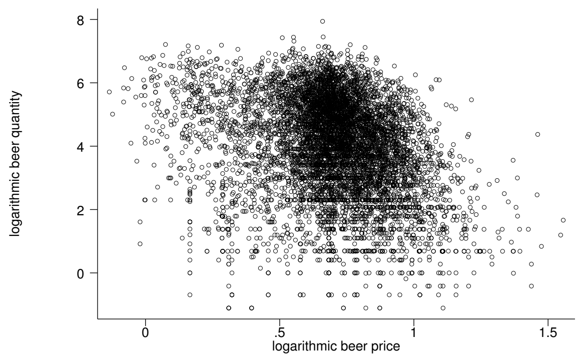

Figure 3: Aggregated beer data (logs)

5 EMPIRICAL RESULTS

In the following we present estimation results. We start by showing the least-squares estimates as a sort of benchmark which then are contrasted with outcomes from quantile regression. In the last subsection we will give some qualification of our estimation results with regard to possible bias due to potential endogeneity of regressors.

We will concentrate on the consumption of beer and wine since alcohol consumption has a particularly clear interpretation of the "intensity" of consumption. We also would like to contrast our results with those from Manning et al (1995). Later on in subsection 5.3 we add some results for other products trying to explain the varying consumption patterns for different goods more generally.

We use the simplest specification possible relating quantity and price by a log-linear model. This has the advantage that coefficients can be interpreted as elasticities. Income is available only in grouped form. We therefore first exploit this information as well as that from other discrete or grouped explanatory variables (age, household size). However, in order to obtain at least rough estimates for the income elasticity, we construct an artificial continuous income variable. The same is done for age. Details of our data transformations are presented in section 5.3.

5.1 Least-Squares Estimation

Table 3 shows the results from least-squares estimation for the consumption of beer. Besides the price of this good the impacts of income (grouped), age of head of household (grouped) and household size are considered by defining three sets of dummies. For each categorical variables the first category is omitted (household size one, income lower than 1000 DM and age less than 25). We note a rather pronounced price elasticity of -1,75% which is comparable in size to the estimated elasticities of alcohol in Blundell et al (1993 table 3) and Manning et al (1995 table 2). Household size matters much more than age or income when looking at the t-ratios. We note for later reference that income has an (albeit slight) significant effect at higher income classes.

Turning to the corresponding results for wine (see table 4) we obtain a quite different picture: the price elasticity is *positive* making wine a "Giffen" good which is at odds with a priori expectations. However, the estimate is not significantly different from zero. Household size is almost insignificant contrary to the results for beer. The income effect is nearly monotone, i.e. coefficients are greater for larger incomes. Age, too, has an impact on wine consumption.

Table 3: Least-squares regression (beer)

log. beer quantity	coef.	std. err.	t	$P > \|t\|$	95% conf. interval	
constant	3.35284	.1889605	17.74	0.000	2.982424	3.723256
log. av. price	-1.753606	.0786581	-22.29	0.000	-1.907799	-1.599414
hhsize = two	1.062992	.0522729	20.34	0.000	.9605224	1.165462
hhsize = three	1.213293	.0622057	19.50	0.000	1.091352	1.335234
hhsize = four	1.320731	.0689660	19.15	0.000	1.185538	1.455924
hhsize = five	1.330039	.0938747	14.17	0.000	1.146018	1.514060
hhsize >= six	1.357907	.1490717	9.11	0.000	1.065685	1.650130
income ∈ [1000,1249]	.0878681	.1583885	0.55	0.579	-.2226182	.3983544
income ∈ [1250,1499]	.0880574	.1531222	0.58	0.565	-.2121056	.3882203
income ∈ [1500,1999]	.2436030	.1359381	1.79	0.073	-.0228742	.5100802
income ∈ [2000,2499]	.1308066	.1341529	0.98	0.330	-.1321710	.3937842
income ∈ [2500,2999]	.3098026	.1353236	2.29	0.022	.0445299	.5750752
income ∈ [3000,3499]	.3405131	.1359922	2.50	0.012	.0739298	.6070963
income ∈ [3500,3999]	.3250298	.1392488	2.33	0.020	.0520627	.5979969
income ∈ [4000,4499]	.3633615	.1398491	2.60	0.009	.0892177	.6375053
income ∈ [4500,4999]	.3150890	.1476771	2.13	0.033	.0256000	.6045780
income ∈ [5000,5499]	.2552472	.1475486	1.73	0.084	-.0339899	.5444842
income >= 5500	.2976291	.1412991	2.11	0.035	.0206429	.5746153
age ∈ [25,29]	.3611248	.1597493	2.26	0.024	.0479710	.6742786
age ∈ [30,34]	.4054011	.1567852	2.59	0.010	.0980577	.7127445
age ∈ [35,39]	.5560247	.1573868	3.53	0.000	.2475020	.8645473
age ∈ [40,44]	.6279981	.1572959	3.99	0.000	.3196535	.9363426
age ∈ [45,49]	.7142163	.1578534	4.52	0.000	.4047788	1.023654
age ∈ [50,54]	.8675696	.1564659	5.54	0.000	.5608521	1.174287
age ∈ [55,59]	.8535417	.1536883	5.55	0.000	.5522691	1.154814
age ∈ [60,65]	.7362940	.1544137	4.77	0.000	.4335994	1.038989
age ∈ [65,69]	.5831481	.1545569	3.77	0.000	.2801727	.8861235
age >= 70	.4721374	.1534021	3.08	0.002	.1714259	.7728490

Table 4: Least-squares regression (wine)

log. wine quantity	coef.	std. err.	t	$P > \|t\|$	95% conf. interval	
constant	.8621149	.2742967	3.14	0.002	.3242790	1.399951
log. av. price	.1011763	.0542206	1.87	0.062	-.0051385	.2074911
hhsize = two	.2183280	.0779064	2.80	0.005	.0655705	.3710855
hhsize = three	.0995142	.0956803	1.04	0.298	-.0880939	.2871223
hhsize = four	.2701201	.1071512	2.52	0.012	.0600201	.4802202
hhsize = five	.1343788	.1470194	0.91	0.361	-.1538942	.4226518
hhsize >= six	.2052321	.2434585	0.84	0.399	-.2721368	.6826010
income ∈ [1000,1249]	.1044016	.224500	0.47	0.642	-.3357938	.5445971
income ∈ [1250,1499]	-.1895866	.2251126	-0.84	0.400	-.6309831	.2518100
income ∈ [1500,1999]	.3246291	.1906123	1.70	0.089	-.0491199	.6983782
income ∈ [2000,2499]	.3041583	.1862344	1.63	0.103	-.0610067	.6693233
income ∈ [2500,2999]	.3381186	.1900110	1.78	0.075	-.0344516	.7106887
income ∈ [3000,3499]	.5028804	.1919761	2.62	0.009	.1264571	.8793038
income ∈ [3500,3999]	.4895946	.1978534	2.47	0.013	.1016473	.8775419
income ∈ [4000,4499]	.5577149	.1964025	2.84	0.005	.1726125	.9428172
income ∈ [4500,4999]	.6424704	.2096058	3.07	0.002	.2314791	1.053462
income ∈ [5000,5499]	.8378914	.2125869	3.94	0.000	.4210548	1.254728
income >= 5500	.9920589	.1991890	4.98	0.000	.6014927	1.382625
age ∈ [25,29]	.3174385	.2453108	1.29	0.196	-.1635625	.7984394
age ∈ [30,34]	.4703598	.2413946	1.95	0.051	-.0029624	.943682
age ∈ [35,39]	.4929683	.2423548	2.03	0.042	.0177633	.9681732
age ∈ [40,44]	.6611684	.2403771	2.75	0.006	.1898414	1.132495
age ∈ [45,49]	.6735898	.2424984	2.78	0.006	.1981033	1.149076
age ∈ [50,54]	.8447658	.2408852	3.51	0.000	.3724424	1.317089
age ∈ [55,59]	.7958071	.2366303	3.36	0.001	.3318268	1.259787
age ∈ [60,64]	.6219815	.2366123	2.63	0.009	.1580364	1.085927
age ∈ [65,69]	.6377288	.2378547	2.68	0.007	.1713476	1.104110
age >= 70	.7447125	.2352835	3.17	0.002	.2833730	1.206052

5.2 Results from Quantile Regression

We now turn to the results from quantile regression which analyzes the conditional distribution to a greater extent. For this we compute quantile regressions for every integer quantile by the methods discussed in section 3. Again we include the explanatory variables price, household size, income and age as explanatory variables. For the consumption of beer this would result in 99 tables corresponding to table 3 for least-squares results. A better way of presenting the results is in form of graphics: Figure 4 displays the estimated price elasticities for all 99 quantiles. The 95% confidence bands from bootstrapped estimation errors are also shown as dotted lines. The same figure shows additionally the corresponding results for wine. We note at first sight the positive price elasticity of wine and a large negative elasticity for beer at the median (50% quantile) which is roughly comparable to the least-squares procedure given in section 5.1 and presented in this figure by horizontal dashed lines. Note that the confidence band regarding elasticities for wine is strictly positive for quantiles around 50% whereas the corresponding least-squares estimate was insignificant.

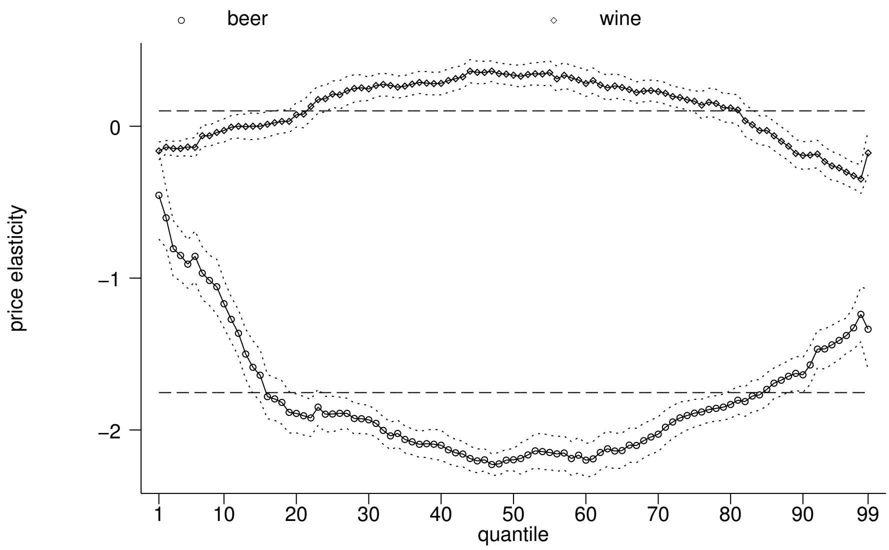

Figure 4: Comparison of price elasticity coefficients

Taking a closer look at the whole pattern reveals interesting findings. For beer, the price elasticity coefficient shows a pronounced U-shaped form, starting at values between -0.4 and -0.9 for small quantiles, peaking at -2.23 (47% quantile) and coming back to values around -1.3 for the largest quantiles. In other words, those consumers either purchasing a very little *or* a very high amount of beer are much less price sensitive than "average" consumers. These findings could be explained as follows: Maybe those purchasing only little do not care much about price *because* of their small amount whereas some of the heavy consumers may be partly addicted to alcohol and therefore as well less price sensitive. The presented results coincide in some way with the findings of Manning et al. (1995). They analyzed the relationship between alcohol consumption (not only beer) and regional average price by quantile regression and also reported a U-shaped price elasticity.

The corresponding results for wine however remind us that the U-shaped pattern is not typical for acoholic beverages: Besides the fact already noted of a *positive* price elasticity the quantile estimation outcome for wine shows a *reversed* U-shape. The elasticity is negative for quantiles smaller than 14% and bigger than 83%, but reaches values greater than +0.35 for the quantiles around 50%. This results could perhaps be explained by the much stronger dispersion of the average price paid for wine and the fact that wine is much more related to social status which may, for example, lead so-called "yuppies" to buy the more expensive wine whenever available.

Since the results so far (which have not been presented besides the price elasticities in figure 4) have not allowed us to estimate income elasticity, we now convert the grouped data back to artificial continuous variables. The details are given in table 5. Most importantly, income classes are now related to a certain income value thereby only approximating the variation between groups and disregarding the variation within groups. For example we assign the income of DM 2,750 to all households from the income interval 2,500 to 3,000. Moreover, we take the logarithms of these values in the estimations presented below. For age a similar procedure is adopted which allows us also to include age squared. For the household size we use just the integers as regressor variables. The estimation results are presented - for some selected quantiles - in tables 6 and 7. As one can see in figure 5, the coefficients for the price elasticity have not changed much compared to the first model. This may serve as an indicator for the suitability of our data manipulations. Moreover, figure 6 displays the estimated income elasticities for all 99 quantiles for both wine and beer.

Table 5: Pseudo continuous variables

hhsize	%	value	income	%	value	age	%	value
one	25.59	1	<= 999	2.36	700	<= 24	1.48	22
two	35.20	2	1000-1249	3.46	1125	25-29	6.37	27
three	18.44	3	1250-1499	4.05	1375	30-34	9.30	32
four	14.95	4	1500-1999	9.86	1750	35-39	9.72	37
five	4.46	5	2000-2499	13.68	2250	40-44	9.57	42
>= six	1.35	6	2500-2999	13.65	2750	45-49	8.29	47
			3000-3499	13.68	3250	50-54	8.64	52
			3500-3999	10.04	3750	55-59	11.32	57
			4000-4499	9.44	4250	60-64	10.12	62
			4500-4999	5.36	4750	65-69	10.68	67
			5000-5499	5.25	5250	>= 70	14.53	77
			>= 5500	9.18	6000			

Table 6: Regression results for beer (t-values in brackets)

log. beer quantity	least sq.	5% quant.	25% quant.	50% quant.	75% quant.	95% quant.
constant	-.6420831	-4.088360	-1.797473	-1.183578	.9832792	3.921834
	[-1.88]	[-4.43]	[-3.75]	[-3.19]	[2.41]	[9.04]
log. av. price	-1.902398	-.9192157	-2.105968	-2.367971	-2.122683	-1.529505
	[-23.83]	[-4.87]	[-15.67]	[-23.29]	[-27.72]	[-11.14]
log. income	.3875034	.4626996	.4378815	.5056042	.3160614	.0799454
	[9.39]	[3.45]	[6.89]	[9.83]	[5.94]	[1.28]
household size	.2660790	.2936658	.2946095	.2693489	.2452102	.1659373
	[15.23]	[7.50]	[11.50]	[12.06]	[10.52]	[7.53]
age	.0852777	.0679076	.0820750	.0866123	.0915853	.0818739
	[11.07]	[4.51]	[6.63]	[10.38]	[11.50]	[8.85]
squared age	-.0007800	-.0006507	-.0007563	-.0007851	-.0008163	-.0007303
	[-10.51]	[-4.78]	[-6.25]	[-9.72]	[-10.05]	[-8.09]

Table 7: Regression results for wine (t-values in brackets)

log. beer quantity	least sq.	5% quant.	25% quant.	50% quant.	75% quant.	95% quant.
constant	-3.727660 [-7.50]	-1.067146 [-2.48]	-3.421165 [-4.15]	-4.261983 [-8.07]	-3.949232 [-8.79]	-3.981429 [-5.32]
log. av. price	.1030630 [1.92]	-.1747103 [-3.14]	.1888987 [1.80]	.3705933 [5.43]	.1859693 [2.69]	-.3212998 [-4.53]
log. income	.5559961 [9.07]	.1218493 [2.67]	.4245471 [3.78]	.5511245 [9.20]	.6407591 [9.76]	.8672602 [10.76]
household size	.0405598 [1.51]	.0409576 [1.15]	.0561165 [1.32]	.0373011 [1.18]	.0455761 [1.88]	-.018845 [-0.37]
age	.0454465 [3.96]	.0068891 [0.61]	.0367241 [2.14]	.0553589 [4.72]	.0597191 [4.10]	.0620024 [2.97]
squared age	-.0003653 [-3.31]	-.0000657 [-0.61]	-.0003397 [-1.99]	-.0004768 [-4.40]	-.0004792 [-3.48]	-.0004107 [-1.98]

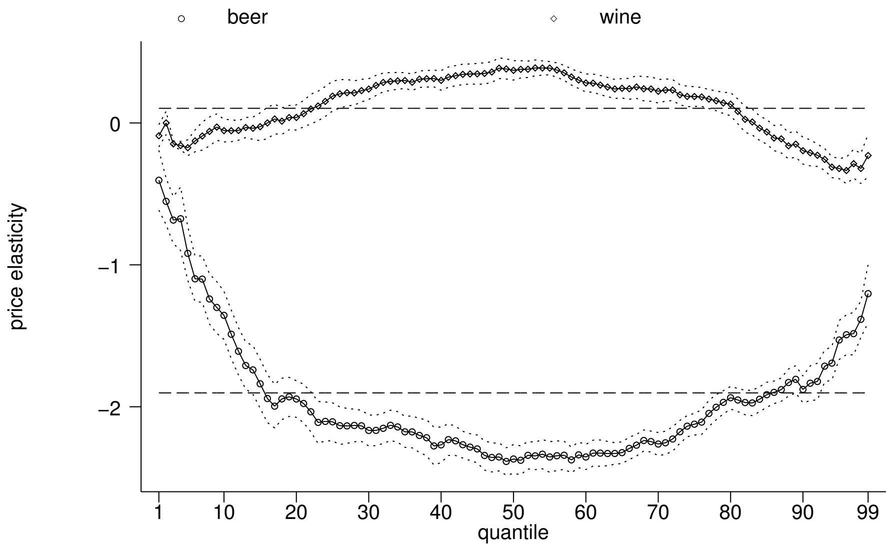

Figure 5: New price elasticities

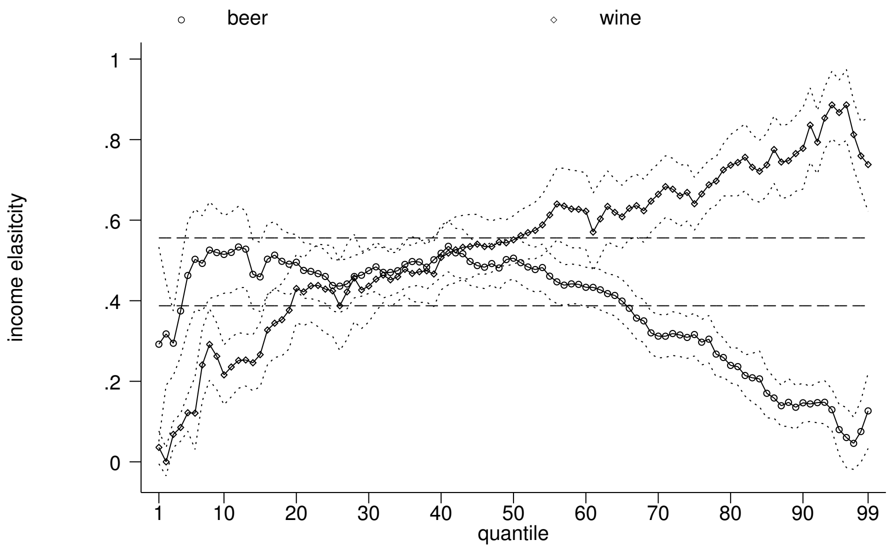

Figure 6: Income elasticities

Figure 6 shows that income elasticities behave quite differently for beer and wine. First we note that the least-squares coefficients are 0.3875 for beer and 0.5560 for wine, respectively. In other words, the positive effect of an income rise on the expected consumption is a bit higher for wine than for beer. Looking at the quantile values, a different picture can be stated. For beer, the elasticity is roughly constant for the quantiles lower than the median, but diminishes at the right tail of the distribution. As far as the impact on wine consumption is concerned, quite the opposite can be observed. The coefficient is negligible small at low quantiles and rises up to 0.8862 at the 96%-quantile. In conclusion, those households consuming only a small amount of beer are more income sensitive than those purchasing a higher quantity while this relationship is reversed for wine. Again this could be explained by the association with social status in the case of wine whereas beer is considered as a every-day good.

Finally, the influence of the household size is depicted in figure 7. It can be seen that the quantile coefficients do not differ much from the least-squares results for wine. The same applies to beer, only at large quantiles some smaller values can be observed.

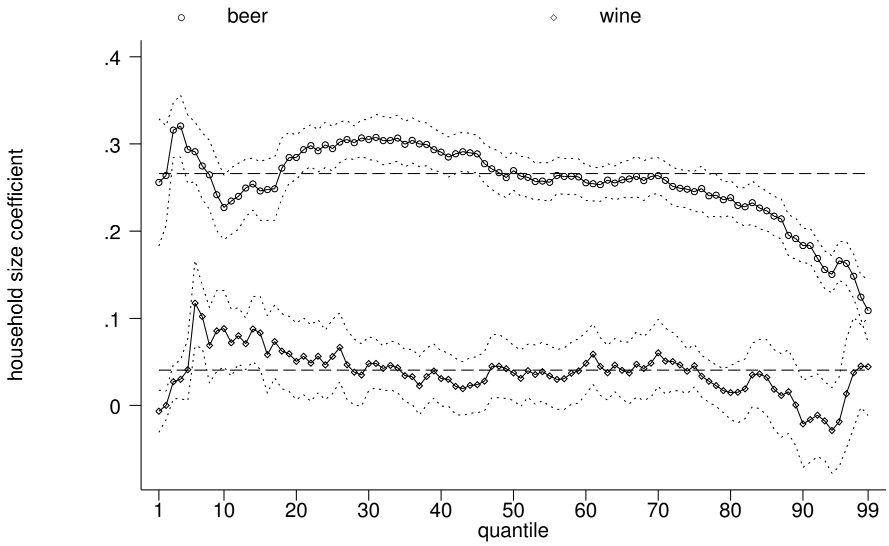

Figure 7: Effect of household size

5.3 Results for other goods

In this subsection we add results for other groups of goods. Again the estimated coefficients are presented in a graphical manner. The method of quantile regression here, too, enables us to reveal more differentiated and detailed results than from standard least-squares estimation. Figure 8 shows results for coefficients regarding price, income and household size for the following categories: frozen food (number 17), fats (22), lemonade (46), animal food (66) as well as mineral water (84). The different vertical scales should be noted.

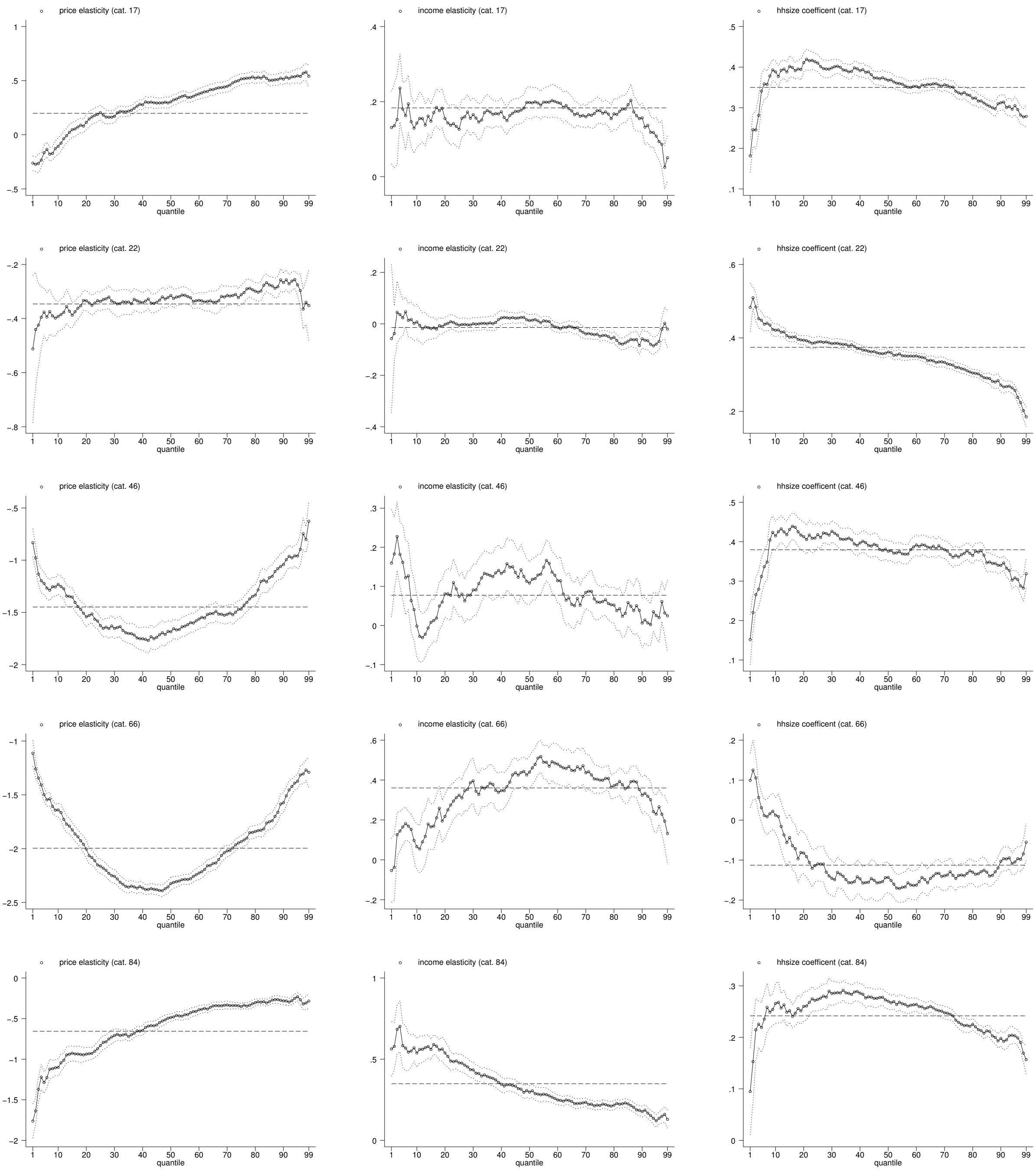

Figure 8: Further categories (price elasticity, income elasticity and household effect)

For the groups considered the following facts can be stated:

- Only for frozen food (first row) the price elasticity moves from negative to positive values for the larger proportion of quantiles. All other goods show normal price reactions over all quantiles.

- None of the goods shows a monotonically *decreasing* graph of price elasticities. A monotonically increasing pattern is given for frozen food and mineral water whereas lemonade and animal food show an U-shape as in the case of beer.

- Income elasticities are - with the exception of fats - always positive and much smaller in size than price elasticities. The income elasticity of 0.8 for wine (see figure 8) is by far the greatest value observed in our data set and the monotonically increasing graph for this good seems to be an exception. Mostly the income elasticities show now a pronounced pattern. None of the goods is a "luxury" one.

- Household size should have an impact on substitutional processes which cannot be observed from our results. For example, households with children will switch from expensive food to less expensive food. Therefore negative coefficients should be possible. However, in all cases considered the impact of household size is positive with the exception of animal food where a negative sign arises for "moderate" and "heavy" consumption. This may be a good example for such substitutional processes: If the household has children and also has animals, then for those households spending a lot of money for animals this will result in a decrease of consumption.

From these (not yet) "stylized facts" we conclude that attitudes of consumers seem to play an important role in the complete description of consumption patterns. We will try to exploit this information provided by the GfK data set on attitudes of consumers[12] to obtain a more profound picture of the different types of consumers.

5.4 Instrumental Variables Estimation

Our discussion of estimation results so far has not raised the question whether our estimates are biased due to non-exogeneity of regressors. Since we use the income of the household and not total expenditures for the goods considered, we maintain that this explanatory variable should be of no concern regarding biased estimation.[13] However, prices of single purchases are expected to be endogenous indeed since prices will have impact on the decision to buy. Whether this is still a problem for the aggregated prices (see section 4) is an open question. We therefore plan to extend our results to IV estimation. However, two severe problems arise which have hindered so far this approach:

- From the description of the data in section 4 it becomes clear that it is hard to define appropriate instruments. Our idea is to use the attitudes of consumers which however are available only as ordinal data.

[12] See section 4.
[13] See Blundell et al (1993 section C) for a discussion on the treatment of total expenditure being endogenous.

- IV estimation in quantile regression is a rather new topic.[14] However, Arias, Hallock and Sosa-Escudero (2001 section 5.1) propose an estimation method which should be easy to implement once the instrumental variables have been found.

Therefore we plan to report on this extension in a later version provided we can manage to find the appropriate instrumental variables.

6 CONCLUDING REMARKS

Our paper presents empirical results obtained from quantile regressions which indicate that there is much heterogeneity around the "average" consumer regarding reactions to prices and income. Some typical patterns have been obtained for different groups of goods which we try to characterize by different attitudes towards consumption of these goods. For example, beer consumption shows the greatest (negative) price elasticity for "moderate" drinkers whereas both "light" and "heavy" drinkers are less price sensitive. On the other hand, for wine consumption price elasticity is *positive* for moderate drinkers and negative for those with very large and very small demand although the price reactions are much smaller than for beer. We have argued that "yuppies" could be regarded as people who sometimes drink wine because of its status effect and therefore prefer the more expensive bottle. On the other hand, people who are used to drinking wine will of course react in a normal manner to price increases. The same price response is plausible for those drinking almost no wine. Contrary to this, beer is much more every-day consumption good (see table 1 for the proportion of non-buyers) and therefore all consumers of beer are rather price-sensitive. For "heavy drinkers" however[15] the problem of addiction makes them less price sensitive. For quite another reason "light" beer drinkers care less about the price.

We have not yet offered a complete picture of this more detailed description of consumption patterns. We plan to expand our analysis by including attitudes of households which are available from the data set used. This should help to test the hypothesis that both intensity of consumption and attitudes must be considered in order to obtain a complete characterization of demand.

REFERENCES

Abadie, A., Angrist, J. and Imbens, G. (2002). "Instrumental Variables Estimates of the Effects of Subsidized Training on the Quantiles of Trainee Earnings", *Econometrica 70*, 91-117.

[14] Some recent studies on this topic are provided by Chesher (2001a, 2001b) and Abadie, Angrist and Imbens (2002).

[15] Figure 2 in section 4 tells us that this means a yearly consumption of more than 500 litres per year.

Arias, O., Hallock K.F. and Sosa-Escudero, W. (2001). "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data", *Empirical Economics 26*, 7-40.

Atkinson, A.B., Gomulka, J. and Stern, N.H. (1989). "Spending on Alcohol: Evidence from the Family Expenditure Survey 1970-1983", *The Economic Journal 100*, 808-827.

Barrodale, I. and Roberts, F. (1974). "Solution of an overdetermined system of equations in the ℓ_1 norm", *Communications of the ACM 17*, 319-320.

Blundell, R.P., Pashardes, P. and Weber, G. (1993). "What Do We Learn On Consumer Demand Patterns From Micro Data?", *American Economic Review 83*, 570-97.

Buchinsky, M. (1998). "Recent Advances in Quantile Regression Models: A practical guide for empirical research", *Journal of Human Resources 33*, 88-126.

Chesher, A. (2001a). "Exogenous impact and conditional quantile functions", Mimeo, 15 August 2001.

Chesher, A. (2001b). "Quantile Driven Identification of Structural Derivatives.", Discussion Paper, Centre for Microdata Methods and Practice, 01 December 2001.

Christensen, L.R., Jorgenson, D.W., and Lau, L.J. (1975). "Transcendental Logarithmic Utility Functions", *American Economic Review 65*, 367-383.

Deaton, A. and Muellbauer, J. (1980a). "An Almost Ideal Demand System", *American Economic Review 70*, 312-326.

Deaton, A. and Muellbauer, J. (1980b). *Economics and Consumer Behavior*, Cambridge University Press.

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

Keen, M. (1986). "Zero Expenditures and the Estimation of Engel Curves", *Journal of Applied Econometrics 1*, 277-286.

Koenker, R. and Basset, G. (1978). "Regression Quantiles", *Econometrica 46*, 33-50.

Koenker, R. and d'Orey, V. (1987). "Computing Regression Quantiles", *Applied Statistics 36*, 383-93.

Koenker, R. and d'Orey, V. (1993). "A Remark on Computing Regression Quantiles", *Applied Statistics 43*, 410-14

Koenker, R. and Park, B. (1996). "An interior point algorithm for nonlinear quantile regression", *Journal of Econometrics 71*, 265-83.

Koenker, R. and Hallock, K.F. (2001). "Quantile Regression", *Journal of Economic Perspectives 15*, 143-156.

Manning, W.G., Blumberg, L. and Moulton, L.H. (1995). "The demand for alcohol: The differential response to price", *Journal of Health Economoics 14*, 123-148.

Papastefanou, G. (Ed.) (2001). "Social and Economic Research with Consumer Panel Data", *ZUMA-Nachrichten Spezial Band 7*, Mannheim.

Portnoy, S. and Koenker, R. (1997). "The Gaussian Hare and the Laplacian Tortoise: Computability of squared-error versus absolute-error estimators, with discussion", *Statistical Science 12*, 279-300.

Ronning, G. (1988). "Möglichkeiten und Grenzen der ökonometrischen Nachfrageanalyse", *Wirtschaftswissenschaftliches Studium 17*, 65-72.

Ronning, G. and Zimmermann, K.F. (1991a). "Mikroökonomie, Mikroökonometrie und Wirtschaftspolitik", Contribution to the special issue of "Ifo-Studien" entitled "Analyse von Mikrodaten als Basis wirtschaftspolitischer Entscheidungen", *Ifo-Studien 25*, 189-200.

Ronning, G. (1992). "Share Equations in Econometrics: A Story of Repression, Frustration and Dead Ends", *Statistical Papers 33*, 307-334.

Stata Corporation (2001). *Stata Reference Manual*, Stata Press.

Stone, R. (1954). "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand", *Economic Journal 64*, 511-527.

Varian, H.R. (1992). *Microeconomic Analysis*, 3rd edition, W.W. Norton & Company