Bachelorarbeit

Microeconometric Estimation of the Demand for Financial Assets Using Quantile Regression

Niklas Schmitz

Contents

List of Figures

List of Tables

1. Introduction

Engel curves have been a standard method to examine how changes in income affect the demand for certain goods since their first use in the 19th century (Engel and Kneip, 1996, p. 187). While the main areas of application have often been commodity groups such as food, clothing or housing, the income-demand relationship for financial assets has not yet been extensively explored using Engel curves.

While there is a rich literature on household saving behaviour aiming to identify its several explanatory factors, much of the analysis has dealt with the channels through which individual factors might affect saving behaviour or total wealth accumulation (for example, see Seguino and Floro, 2003; Guiso, Sapienza and Zingales, 2008; Bucher-Koenen and Lamla, 2014; Almenberg and Dreber, 2015). Without a doubt, these studies have contributed to a deeper understanding of the mechanisms explaining the demand for financial assets. However, an analysis of the determining factors, particularly the income-demand nexus, using this work's approach has – to the best of my knowledge – not yet been presented.

A profound comprehension of this link is valuable in terms of both economic theory and policy. Income is a key determinant of saving in standard saving models and results from an empirical examination of the income-demand nexus may thus allow for a better assessment of the structure of these models (for example, see Browning and Lusardi (1996)).

In terms of economic policy, household saving contributes strongly to overall saving in an economy and thus impacts economic growth through the investment channel. A more profound understanding of the determinants of household saving behaviour could thus improve the capability of public authorities to influence economic performance. Additionally, households' saving decisions affect their ability to smooth consumption over the business cycle or compensate unexpected income losses (Seguino and Floro, 2003, p. 149). Retirement income is also determined by financial behaviour, and especially due to the multi-pillar public pension system in Germany, households are required to at least partly accumulate sufficient wealth on their own to finance their expenses during retirement (Bucher-Koenen and Lusardi, 2011, pp. 565 f.). From a policy perspective, it could thus be of interest to identify what determines weak demand for financial assets as these can be expected to build an integral part of the saving strategy of many households. Therefore, an analysis of different demand quantiles is desirable.

For this reason, this work will present quantile regression as a useful addition to the economist's apparatus of statistical tools for the research field of demand analysis. In contrast to the Ordinary Least Squares (OLS) approach, quantile regression, which was originally introduced by Koenker and Basset (1978), aims to estimate coefficients on the explanatory factors at the quantiles of the reponse variable[1]. It is thus possible to leave the assumption of constant coefficients across quantiles behind, which would be imposed on the model by OLS. By deploying this technique, the paper aims to contribute to the existing literature on household demand for financial assets by answering the following two questions: How is household demand for financial assets correlated with income and additional factors such as wealth, age, gender or education? Do the effects of these determinants vary among quantiles of the demand distribution?

On a more technical level, this work shall also demonstrate the benefits of quantile regression compared to OLS. Koenker and Basset (1982, p. 43) questioned if the conditional expectation function as approximated by OLS provides an accurate description of the relationship among variables and concluded that this holds only "within the confines of extremely restrictive parametric models". Quantile regression may provide additional insights about the relationship between variables, and this is especially true when analyzing distributions with large dispersion, outliers, and heteroscedastic error terms.

Irrespective of the estimation technique, the question as to which factors ultimately determine household saving behaviour and how the precise relation between income and demand can be characterized is very complex and a myriad of factors could be taken into account. As a consequence, this work confines itself to a mere description of the observed correlations. While causal interpretations will be suggested at some points, a definitive statement about causal relationships will not be provided. In this regard, the focus of this paper is very close to the intention of the original work by Ernst Engel from 1857, who tried to *describe* observed economic regularities in contrast to *explaining* household behaviour (Chakrabarty and Hildenbrand, 2011, p. 290).

Another important distinction has to be made: In contrast to much of the existing literature on household investment decisions, I do not examine the determinants of total net assets, but the factors that influence net purchases of these assets within a given time frame. Therefore, this work focuses on

[1] It is important to note that quantile regression allows to examine if the effect of independent variables differs across quantiles of the dependent variable, not quantiles of the independent variable. Thus, when speaking of "quantiles" I refer to quantiles of the *dependent* variable unless stated otherwise.

changes in net assets rather than the overall value of a portfolio.
The remainder of this thesis is structured as follows. Section 2 provides an overview of the literature on Engel curves and some applications of quantile regression to demand analysis. Theoretical considerations are discussed in Section 3. Section 4 describes the main features of the German Income and Consumption Survey datasets. Section 5 presents the estimation results and an interpretation of the findings. Section 6 concludes.

2. Literature Review

Engel curves have been a research field of vast interest since their first notion in the mid-19th century. The following review will give a compressed overview on some of the main lines of development in this area with a particular focus on the choice of functional forms and applications of quantile regression in demand analysis. It should by no means be understood as an all-encompassing treatment of the studies on demand analysis.
In 1857, Ernst Engel published his analysis on the consumption patterns of Belgian working class families, in which he aimed to describe the observed relationship between income and food expenditure. His main finding was that expenditure on food increases in both income and family size, but that the share of income spend on food decreases as income rises (Perthel, 1975, p. 211). This phenomenon became known as Engel's Law and constituted the foundation onto which a rich literature on demand analysis has developed.
Given the focus of Engel's work, most of the research in the subsequent decades focused on analyzing the income-demand nexus for commodities such as food or clothing. In 1935, Allen and Bowley connected demand analysis with utility theory (Lewbel and Houthakker, 2008, p. 847). This framework allowed to transition from a mere description of observed patterns to an interpretation based on economic theory. Their results suggested that demand for commodities could be described by a linear or parabolic regression line. Farrell's (1954) estimation of Engel curves for motorcars in the U.S. demonstrated the differences in terms of utility theory between consumption goods and durables. He argued that, in contrast to consumption goods, consumer utility does not arise from consuming, but owning durable goods like cars. He came to the conclusion that durables have much in common with investment in this respect and "perhaps the best procedure in theoretical work is to regard it as such" (Farrell, 1954, p. 171).

While Engel could not rely on statistical methods such as regression analysis as they were not developed yet, much of the research in the 20th century regarding Engel curves focused on finding a theoretical model that provides a well-fitting regression line. A well-known contribution to the choice of the functional form was presented by Prais and Houthakker in 1955 who attempted to identify functional forms that could characterize the income-demand relationship for certain goods (Deaton, 1986, p. 1799). While they found that food Engel curves were best described by a semi-log form, the income-demand relationship for other commodity groups was shown to rather follow a double-log form (Forsyth, 1960, p. 368)[2].

The work by Working (1943) and Leser (1963) yielded a specification known as the Working-Leser form. Working studied the distribution of expenditure among families in the U.S. in the 1930s and their dependance on several factors such as income level, occupational class and type of community. He found that there are differences in expenditure patterns depending on these characteristics, but overall, his work confirmed Engel's Law by stating that the budget share of food expenditure can be closely approximated by $\frac{F}{T} = \alpha - \beta \cdot log(T)$, where F denotes expenditure on food, T total expenditure and α and β are parameters to be estimated (Working, 1943, p. 45). This form also indicates that Working chose to use total expenditure as a proxy for income. This decision is mainly based on the view that a specification using total expenditure will be more stable over time since changes in income will strongly affect the proportion of income saved (Working, 1943, p. 56). Thus, decreases in income might lead to a decreasing percentage of income spent on food, while the share of food in total expenditure remains constant.

Based on these results, Leser (1963) tried to generalize the specification from Working (1943) to find a model that provides accurate estimation results for most commodity groups. He compared several two-parameter specifications based on the requirements that the mathematical form should have its foundation in utility theory, be valid for all (or at least a wide range of) positive values of total expenditure and provide plausible estimates of income elasticities (Leser, 1963, p. 694). Additionally, Engel curves should fulfill the criterion of additivity, which means that expected total expenditure can be calculated as the sum of expected expenditure on individual commodity groups conditional on a certain level of income, i.e. budget shares need to add up to unity (see also Hausman, Newey and Powell, 1995, p.218). While he found that

[2] It should be noted, however, that their theoretical approach did not remain uncriticized, for example see Deaton and Muellbauer (1980, pp. 196 ff.).

the Working-Leser form as suggested by Working (1943) provided better fit than other specifications, he noted that the fit could be improved by using a three-parameter Engel function of the form $\frac{F}{M} = \alpha - \beta \cdot log(M) + \frac{\gamma}{M}$, where M denotes income (Leser, 1963, p. 699)[3]. This functional form is able to capture a more complex relationship between income and demand and thus presents an improvement over previous, simpler functions.
Aitchison and Brown (1954) proposed yet a different functional form. Since consumption for some goods might not increase above a certain satiety level, they developed an Engel curve function that is able to capture this property of consumer behaviour. The result is a sigmod (S-shaped) Engel curve that reflects demand for luxury goods at low income levels and demand for necessities at high income levels. They found that this functional form provides a good approximation of the income-demand relationship for working-class expenditures for food, clothing, fuel and rent from 1937 to 1938.
While the Engel curve functions presented above might provide a good fit in various instances, they all require the researcher to define the functional form a priori. Since economic theory does not provide information on a correct functional form, the estimated coefficients may be inconsistent in case of a misspecified parametric model (Engel and Kneip, 1996, p. 188)[4]. Nonparametric regression offers a solution to this problem. Instead of specifying a model to describe the data, the technique derives a regression line from an analysis of the data. Bierens and Pott-Buter (1990) demonstrated the advantage of this method in the area of applied demand analysis. While they acknowledged that this approach does not rest upon economic theory, they also argued that previous research had relied on very strict assumptions regarding household behaviour, mainly that all households face an identical utility function. Since their dataset did not contain enough information to distinguish the shape of their household Engel curves from simple linear specifications, they concluded that the widely used nonlinear Engel curves might not be realistic (Bierens and Pott-Buter, 1990, p. 175).
Banks, Blundell and Lewbel (1997) used nonparametric regression to show that the Working-Leser form cannot capture the curvature that was evident from the consumption patterns of British households. However, linearity of Engel curves could not be rejected for food and fuel. For the remaining commodity

[3] One may argue that this specification introduces the problem of multicollinearity. Leser (1963, p. 700) acknowledges this by stating that the estimation of parameters b and c is "not of primary concern".
[4] Hausman, Newey and Powell (1995, p. 218) note that additivity of Engel curves is a criterion imposed by economic theory, but this restriction is typically enforced in the data.

groups "clothing", "alcohol" and "other" they found evidence that the budget share w_i can be estimated by an Engel curve function including the square of the logarithm of total expenditure z: $w_i = \alpha + \beta \cdot log(z) + \gamma \cdot log(z)^2 + \varepsilon_i$ (Hausman, Newey and Powell, 1995, p. 219; Banks, Blundell and Lewbel, 1997, p. 537). This form was also suggested by Hausman, Newey and Powell (1995), who additionally found that incorporating the cube of the logarithm of expenditure yields even more accurate results when estimating Engel curves based on data from the Consumer Expenditure Survey.

Irrespective of the estimation procedure, the aforementioned studies all focused on describing the income-demand nexus through the expected value of demand conditional on income. While this approach is very common in regression analysis, quantile regression may provide some additional information about the distribution and has thus also been applied in the field of demand analysis. Koenker and Bassett (1982) used Engel's data on Belgian workmen's families to demonstrate their method of estimating conditional quantile functions.

Deaton (1997) emphasizes the benefits of quantile regression in the analysis of household survey data, which is mainly motivated by the issue of heteroscedasticity. His application to food Engel curves for Pakistan reveals an increase in the dispersion of budget shares for wealthy households, a property that had not previously been apparent from OLS regression (Deaton, 1997, pp. 72 f.). Koenker and Hallock (2001, p. 147) demonstrated that Engel's original analysis of families' food expenditure could be improved through the use of quantile regression. Due to outliers and the asymmetry of the conditional density, the least squares estimator seems to fail to capture household behaviour accurately, especially for low income households.

Other applications in the field of demand analysis include Hendricks and Koenker (1992), who used quantile regression to analyze the determinants of the demand for electricity. Since electricity demand fluctuates throughout the day it is particularly interesting to identify the determinants causing differences between times of high and low demand. Therefore, estimating multiple regression lines to describe quantiles of demand seems advantageous over OLS regression.

Manning, Blumberg and Moulton (1995) used quantile regression to examine differences in the effect of price changes on drinking behaviour among light, moderate and heavy drinkers. Their results indicated that taxes on alcohol may reduce consumption. However, heavy drinkers are significantly less price-sensitive than moderate drinkers, and thus taxes on alcohol may fail to reduce drinking within the main target group (Manning, Blumberg and Moulton,

1995, p. 141).
Overall, this review on some key publications in the field of Engel curve estimations in particular and demand analysis more generally should have demonstrated that the use of quantile regression may yield deeper knowledge regarding the determinants of household consumption behaviour. Additionally, the field is characterized by a strong relation between economic theory and econometric models. Deaton (1986, p. 1768) came to the conclusion that "it is not possible to study applied demand analysis without keeping statistics and economic theory simultaneously in view". Therefore, this work proceeds by dealing with the main theoretical considerations for regression analysis before providing the economic rationale for the choice of variables.

3. Theory

This section starts by covering the basic theory of the quantile regression approach. Subsequently, it uses these insights to consider some typical OLS assumptions and explain similarities as well as differences in comparison to quantile regression. Lastly, it discusses the choice of variables.

3.1. The Quantile Regression Model

Using averages to describe the distribution of a random variable is well-established in econometrics. OLS regression is a typical way to estimate the expected value of a dependent variable Y_i conditional on a set of independent variables, X_i. The true conditional expected value of the random variable Y_i, $E[Y_i|X_i]$, is described through the conditional expectation function (CEF) (Angrist and Pischke, 2009, p. 30). However, OLS regression might only answer the question of what happens to averages when independent variables change. It cannot capture features such as an increasing spread of the distribution.
For example, income could have a varying impact on households' demand for financial assets: For those households that demand relatively few financial assets, increases in income could lead to rather small increases (or even decreases in case of dissaving) in the quantity demanded. In contrast, households already investing heavily in financial assets might invest a large share of an increase in income. Quantile regression embodies a useful method to describe properties

of this kind[5].
Analogously to the conditional expectation function in case of linear regression, the basis of quantile regression is the conditional quantile function (CQF)[6]:

$$Q_\tau(Y_i|X_i) = {F_Y}^{-1}(\tau|X_i). \tag{1}$$

$F_Y(y|X_i)$ denotes the distribution for Y_i at y conditional on a vector of regressors X_i and $Q_\tau(Y_i|X_i)$ is the τth quantile of Y_i conditional on X_i. Thus, $\tau = .5$ yields a description of the median, whereas $\tau = .25$ yields the lower quartile, or, more generally, $P[Y_i \leq Q_\tau(Y_i|X_i)|X_i] = \tau$ (Angrist and Pischke, 2009, p. 270).
While the CEF minimizes the expected value of squared deviations from Y_i, the CQF solves a similar minimization problem (Angrist and Pischke, 2009, p. 271):

$$Q_\tau(Y_i|X_i) = \underset{q(X)}{\arg\min}\, E[\rho_\tau(Y_i - q(X_i))], \tag{2}$$

where $\rho_\tau(u)$ is an asymmetric loss function (also called the "check function") and defined as

$$\rho_\tau(u) = 1(u > 0) \cdot \tau|u| + 1(u \leq 0) \cdot (1 - \tau)|u|. \tag{3}$$

The τth quantile of Y_i conditional on X_i is the value $q(X)$ that minimizes the expected value of the check function ρ_τ. Since positive and negative deviations of the estimated value from Y_i shall not be weighted equally as in the case of the CEF[7], the check function is used as a means to weight deviations in different directions with different factors depending on τ. For high values of τ, positive deviations from $q(X_i)$ $(u > 0)$ are assigned large weights τ whereas negative deviations $(u \leq 0)$ are only weighted with $1-\tau$. Therefore, $Q_\tau(Y_i|X_i)$ is relatively high to minimize the expected value of the check function. Analogously, lower values of τ yield lower conditional quantiles. Note that the check function minimizes the sum of *absolute values* of the errors conditional on τ instead of the sum of squared errors as in the case of OLS and is thus less sensitive to outliers.

[5] Davino, Furno and Vistocco (2013, pp. 38 ff.) note that conditional quantiles can also be calculated via OLS by "adding/subtracting the corresponding standardized normal quantiles from the OLS slope estimate". However, this only holds if errors follow a normal distribution. In contrast, quantile regression allows calculation of the conditional quantiles irrespective of the error distribution.
[6] Note that the following holds only for continuous random variables. A more general definition is given in Angrist and Pischke (2009, p. 270).
[7] Except for the median, of course, where $\tau = .5$ ensures that an equal proportion of observations is below and above the median (Koenker and Hallock, 2001, p. 145).

For reasons of simplicity, Angrist and Pischke (2009, p. 271) suppose to substitute a linear model $X_i'b$ for $q(X_i)$. This yields the following[8]:

$$\beta_\tau \equiv \arg\min_b E[\rho_\tau(Y_i - X_i'b)]. \tag{4}$$

This minimization problem can be solved easily using computer statistics software, which results in the quantile regression estimator $\hat{\beta}_\tau$. While this assumes the CQF is linear in its parameters, quantile regression does not rely on this assumption. Even without assuming a specific form for the unknown CQF, quantile regression provides the best linear approximation to the true CQF as shown by Angrist, Chernozhukov and Fernández-Val (2006). This approximation property is similar to that of OLS, which provides the best linear approximation to the CEF even in case of misspecification (Angrist and Pischke, 2009, p. 38).

For the interpretation of estimated coefficients, two cases of quantile treatment effects should be distinguished (Angrist and Pischke, 2009, pp. 273 ff.; Davino, Furno and Vistocco, 2013, p. 39): In case of homoscedasticity, the quantile regression coefficients are constant among quantiles. Changes in the regressors X_i only result in a *location shift*, i.e. increases in the intercept. This can be formalized as follows. Suppose the random variable Y_i is normally distributed with $E[Y_i|X_i] = X_i'\beta$ and the constant variance of the error term $\varepsilon_i \equiv Y_i - X_i'\beta$ is ${\sigma_\varepsilon}^2$:

$$Y_i \sim N(X_i'\beta, {\sigma_\varepsilon}^2). \tag{5}$$

From this, one can show that

$$Q_\tau(Y_i|X_i) = X_i'\beta + \sigma_\varepsilon \Phi^{-1}(\tau), \tag{6}$$

where Φ^{-1} denotes the inverse of the standard normal conditional distribution function. It is apparent that the coefficient β does not depend on τ. Differences among the quantiles arise solely from changes in the intercept $\sigma_\varepsilon \Phi^{-1}(\tau)$. The conditional distribution of Y_i is thus equally spread for all X_i.

In contrast, heteroscedasticity implies varying effects of the regressors on the dependent variable depending on the quantile τ. The variance of the error term is not constant, but instead a function of X_i:

$$Y_i \sim N(X_i'\beta, \sigma^2(X_i)). \tag{7}$$

[8] X' is used to denote the transposed form of X.

Be $\sigma^2(X_i) = (\lambda' X_i)^2$ and λ a vector of positive coefficients such that $\lambda' X_i > 0$. Then, the CQF can be written as

$$Q_\tau(Y_i|X_i) = X_i'\beta + (\lambda' X_i)\Phi^{-1}(\tau) = X_i'[\beta + \lambda\Phi^{-1}(\tau)]. \quad (8)$$

In this case, $\beta_\tau = \beta + \lambda\Phi^{-1}(\tau)$, which implies the effect of changes in X_i depends on τ. This allows to capture changes in the scatter of the dependent variable as independent variables vary in their values, which is called a *scale shift* (Koenker, 2005, p. 29). Note that both changes in the intercept and the coefficients dependent on τ are expressed by the above equation if the constant has been moved into X_i. This is called a *location-scale shift* model (Koenker, 2005, p. 29). One of the advantages of quantile regression over OLS becomes apparent when one considers that the latter always implicitly assumes a location shift, while the former additionally allows for a location-scale (or just a scale) shift.

In summary, quantile regression allows to deal with heteroscedasticity that might be present in the conditional distributions. It enables to examine if changes in the regressors are correlated with differing changes in the independent variable depending on its quantile.

3.2. Model Assumptions

There are some typical assumptions that are made to allow for causal interpretation of OLS regression results. Following Greene (1993, pp. 171 f.), Stock and Watson (2015, pp. 245 f.) and von Auer (2016, pp. 166 f.), some assumptions often made for OLS regression will be presented and compared to those of quantile regression:

Assumption 1. No relevant variables are missing and no irrelevant variables are included. Omission of relevant variables causes biased estimates and useless hypothesis tests, while the inclusion of irrelevant variables leads to inefficient, albeit unbiased parameter estimates and imprecise hypothesis tests (von Auer, 2016, pp. 305 ff.).

This basic principle remains true for quantile regression. All, but only relevant variables should (theoretically) be included in the regression specification.

Assumption 2. There is a linear relationship between the dependent variable and the parameters. Therefore, the functional form is specified as $y = X'\beta + \varepsilon$. Implicitly, parameters are assumed to be constant across all quantiles of the dependent variable, which has been described as a location-shift model above.

Assumption 2 need not hold in case of quantile regression. One key advantage of this estimation procedure is that it allows for coefficients to vary across quantiles of the dependent variable, which can result in a location-scale model (see above). Nevertheless, the response variable is still assumed to be a linear function of the parameters.

Assumption 3. The expected value of the error term ε_i conditional on the dependent variables $X_{1i}, ..., X_{ki}$ is zero: $E(\varepsilon_i|X_{1i}, ..., X_{ki}) = 0$. From this follows that the omitted factors contained in the error term are uncorrelated with the k regressors: $Cov[X_{ji}, \varepsilon_i] = Cov[X_{ji}, E(\varepsilon_i|X_{ji})] = 0, j = 1, ..., k$ (this uses the law of iterated expectations and a proof can be found in Angrist and Pischke (2009, p. 32)). Therefore, the OLS estimators are not systematically biased and Y_i can be expected to lie on the estimated regression line on average.

Quantile regression also aims at minimizing the expected deviation of the regression line from the true values. However, it should be noted that both procedures vary significantly in terms of the loss function. OLS minimizes equally weighted squared errors, while quantile regression uses an asymmetric loss function to minimize absolute deviations.

Assumption 4. $(X_{1i}, ..., X_{ki}, Y_i), i = 1, ..., n$ are independently and identically distributed (i.i.d.) random variables. This is true if the data are obtained by random sampling. Assuming an identical distribution rules out heteroscedasticity. Note that heteroscedasticity does not prohibit OLS regression as estimators remain unbiased. However, OLS becomes inefficient and standard errors should be corrected in this case (Deaton, 1997, p. 71).

While OLS remains feasible under heteroscedastic error terms, quantile regression provides a more flexibe framework to analyze a dataset of this kind. Varying effects of covariates across quantiles become apparent, which constitutes an advantage over OLS. Still, heteroscedasticity-robust standard errors should also be computed for quantile regression estimates (Koenker, 2005, p. 74).

Assumption 5. Large outliers are unlikely. Since OLS estimators can be sensitive to values outside the typical range of the data, large outliers can bias the estimation results. They often occur due to measurement or data entry errors and should be either corrected or dropped from the sample (Stock and Watson, 2015, p. 173 f.).

Quantile regression is less sensitive to outliers than OLS due to its distinct loss function. Therefore, this assumption can be dropped.

Assumption 6. The regressors are not perfectly multicollinear. This assumption prohibits a situation in which one regressor can be expressed as a perfect linear function of one or multiple other regressors. Multicollinearity would make OLS estimation impossible: The coefficient on a regressor measures the impact of a change in this regressor on the regressand holding the values of all other regressors constant, but in case of multicollinearity the value of at least on other regressor must change simultaneously (Stock and Watson, 2015, p. 246).

Perfect multicollinearity also poses a problem for quantile regression, but as it is typically caused by misspecification of the model, it can be avoided through careful construction of the regression model. Note that problems may still arise from imperfect multicollinearity although not violating Assumption 6 (Stock and Watson, 2015, pp. 251 f.).

3.3. Discussion of Variables

The question remains as to which variables should be included in the regression models. Income will be one of the regressors as the income-demand nexus is of particular interest in this work. Apart from this rather obvious choice, there are several other covariates which will be briefly discussed in the following to justify the choice of model specifications.

Wealth. It seems reasonable to believe that the level of wealth is correlated with absolute changes in wealth and one may thus argue that some measure of it should be incorporated. Households having accumulated high wealth in the past may be likely to continue with this habit. However, once entering retirement, well-endowed households can be expected to show higher divestment as suggested by life-cycle models (Browning and Crossley, 2001).

Age. Standard life-cycle models suggest that households accumulate wealth throughout their working life and dissave in retirement (Browning and Crossley, 2001, p. 14). Therefore, one could expect age to have a varying effect on net demand for financial assets. Until retirement, net demand should increase on average due to the retirement motive, while it turns negative with retirement (Cagetti, 2003, p. 339). The importance of including age and wealth as variables when examining households' financial decisions has also been stressed by Campbell (2006, p. 1555).

Gender. Seguino and Floro (2003, p. 149) have argued that gender differences in saving motives are observable. If households pool their incomes, bargaining about the allocation to saving and consumption could arise, and the result will be affected by the bargaining power of male and female negotiators. Negotiation power is determined by the outside option, and main income earners can thus be expected to have stronger negotiation power on average. Therefore, the gender of the household head might affect the demand for financial assets. Additionally, gender gaps in financial literacy or risk attitude may cause differences in saving behaviour (Dohmen et al., 2011; Halko, Kaustia and Alanko, 2012; Bucher-Koenen et al., 2017).

Marital Status. The potential gender gap in saving behaviour may be diminished in cases where the female main income earner is married. One could assume that there is specialization among the partners where men typically take care of household financial decisions (Bucher-Koenen et al., 2017, p. 270). The effect of gender on demand for financial assets may thus become insignificant when controlling for marital status.

Education. Education could serve as a proxy for financial literacy. Bucher-Koenen and Lusardi (2011, p. 572) presented results indicating a high correlation between education and financial literacy. Financial literacy may then affect overall saving behaviour as more financially literate individuals can be expected to show higher saving during the worklife (c.p.) and consequently higher divestment during retirement (Lusardi, 2009, p. 123). A measure of education should also be included since it may partly explain the effect of gender on the demand for financial assets through gender differences in financial literacy (Bucher-Koenen et al., 2017, p. 266).

Region. Given the long-term separation of Germany into the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR) and the vast social, political and economic differences between these two regions, one might suppose that the demand for financial assets could vary between East and West German households. While the FRG was characterized by free access to capital markets, the command economy and lack of well-developed capital markets in the GDR shaped its inhabitants' attitude toward the financial system as well as their willingness to educate themselves in financial matters (Bucher-Koenen and Lamla, 2014, p. 2).

Profession. People working in the financial sector could be more open toward investment in financial assets (in particular those with higher complexity and risk) as they could be expected to be more financially literate and may have higher trust in the financial system.

Self-employment. Saving behaviour could differ between the self-employed and those who are permanently employed. Campbell (2006, p. 1570) argues that self-employed individuals could be willing to take less risks when investing their money since they are already exposed to substantial business risk. This could well affect overall saving as unexpected income losses may be particularly strong in case of self-employment. Saving could thus be above-average as a precautionary measure (c.p.).

Household Size. Larger households generally face higher living costs, but can also benefit from economies of scale as per-capita living expenses are typically lower for larger families than for single households. However, per-capita income may also decrease as household size increases since usually there live at most two income earners in one household, which means each additional household member adds no income to the household budget. These effects may compensate each other, and in case household size is a significant determinant of net demand for financial assets, the sign of the coefficient will reveal which effect prevails.

4. Data

The analysis in this paper is based on the Sample Survey of Income and Expenditure (EVS) 2003 and 2013. The survey is conducted by the Federal Statistical Office of Germany (henceforth "FSO") in cooperation with the federal states' statistical offices. Every five years, German households[9] are asked to voluntarily participate in the survey to answer detailed questions on basic household demographics as well as their earnings and pattern of consumption. This chapter discusses the data source by first providing an overview of the data collection and preparation process. Afterwards, variables to describe household characteristics as well as descriptive statistics are discussed. Lastly, a quick assessment of the data quality is presented.

4.1. Data collection and preparation

The results of the EVS are not based on a random, but on a quota sample. The FSO (2017, p. 18) argues that random sampling would require a way larger initial sample given low response rates and the risk of underrepresentation of certain socioeconomic groups (for example, those with exceptionally high income). Therefore, households with particular characteristics are deliberately

[9] The EVS only asks private households at their primary residences to participate in the study, communal accommodation and secondary residences are not included (FSO, 2005, p. 10; FSO, 2017, p. 19).

targeted as potential survey participants in accordance with a quotation plan to ensure the sample is representative of the German population. Originally, 74,600 (EVS 2003) and 79,287 (EVS 2013) households were sent questionnaires, and due to nonresponse 53,432 (2003) and 53,490 (2013) of these were returned by the respondents (FSO, 2005, p. 46; FSO, 2017, p. 24). Based on these samples, the Microcensus 2010 was used as a reference statistic for the quotation and the Microcensuses 2012 and 2013 for extrapolation in the EVS 2013, whereas the Microcensuses from 2000, 2002 and 2003, respectively, were used for the EVS 2003 (FSO, 2005, p. 26; FSO, 2017, p. 18). This ensures representativeness of the sample regarding the predefined characteristics for quotation, which include type of household, social status of the household head and monthly net household income (FSO, 2005, p. 31; FSO, 2017, p. 22).

The EVS consists of four parts. First, all participants are asked to provide general information on their household demographics such as household size, age of household participants or occupation on the 1st of January. Additionally, information on financial and non-financial assets per the 1st of January are requested. Subsequently, an equal share of participants for each quarter of the year answer more detailed questions on their income sources and expenditure during the three-month period. Additionally, 20% of participants are required to complete a detailed questionnaire on their consumption of food, alcohol and tobacco over the period of one month (FSO, 2005, p. 27; FSO, 2017, p.20).

The datasets used in this work consist of 80% subsamples that have been randomly drawn from the original samples of the EVS 2003 and 2013, respectively (FSO, 2005, p. 41; FSO, 2017, p. 46). For reasons of anonymity, the five highest and lowest values for some attributes are only given as the arithmetic mean of the respective values. Additionally, a random error of 1% is employed on the remaining values in the lowest and highest deciles of each characteristic (FSO, 2005, p.42; FSO, 2017, p. 47)[10].

Due to differences in the encoding of variables, some observations were removed from the datasets: Household size is given as a numerical value, but in the EVS 2013 (EVS 2003) a value of eight (a value of nine) is used to indicate a household with eight (nine) *or more* household members. Thus, households with eight or more members have been removed from the sample to improve interpretability of the coefficient. This applied to only 16 households in the EVS 2016 and 41 households in the EVS 2003. Additionally, 13 (EVS 2003)

[10] Further adjustments were made by the FSO to ensure anonymity of survey participants. However, these will not be discussed in further detail as they are irrelevant for the upcoming analyses.

and 7 (EVS 2013) households with non-positive disposable incomes were removed from the samples. Negative incomes can probably be explained by high payments of tax arrears that occurred during the reporting quarter. While it is still possible that some of the positive household incomes are also biased by special effects of this kind, removing those households with zero or below-zero income seems to improve representability of the datasets[11]. In accordance with the practice of the FSO, households with a monthly income of above 18,000€ were removed due to the low number of observations[12] (FSO, 2005, p. 21; FSO, 2017, p. 17).

4.2. Variables and descriptive statistics

After the adjustments described in the previous subsection, 42,680 observations remained for the 2003 dataset and 42,756 observations for the 2013 dataset. Comprehensive descriptive statistics are given in Tables 2 and 3. For a detailed overview on the definition of certain variables see Table 1.
Demand is defined as monthly net demand for financial assets in €, i.e. money invested in certain assets less divestment of these asset positions. Monetary assets include saving accounts, call money, fixed deposits, building loan contracts, bonds, stocks, investment funds and other securities. Demand for financial assets only covers purchases directly made by the household and does not include capital-forming benefits. In the EVS 2003, average demand (arithmetic mean) for financial assets is 210€ per month. However, values are dispersed over a wide range of around 130,000€. Average demand in 2013 is considerably lower at 138€ per household per month.
Average monthly household income in 2003 is at 3,535€, where *Income* is defined as disposable income in 1000€ as reported in the EVS. For 2013, households reported a slightly higher monthly income on average. Note that the range of incomes is affected by the adjustments made to the datasets (see the previous section).
Average household wealth has also increased from 2003 to 2013. *Wealth* refers to wealth in those financial assets for which demand is estimated less consumer loans and is also given in 1000€[13]. Therefore, negative values for wealth can

[11] In addition, removing these households is necessary to be able to use the logarithm in some regression specifications.
[12] Note that the FSO uses net income to determine the cut-off value while disposable income is used as an income measure in this work.
[13] Note that for the EVS 2013, the FSO suggests to also subtract education loans to calculate net wealth (FSO, 2014, p. 8). However, no information on these loans is included in the

be reported, and minimum wealth lies below -200,000€ in both datasets.
Individual variables are chosen for the main income earner as stated by the household on the EVS questionnaires. *Female* is a dummy indicating whether the main income earner is female (1) or not (0). *Age* gives the main income earner's age in years and varies between 20 and 85 years for both datasets. *Married* equals unity if the main income earner is married, while all other relationship statuses are indicated by a value of zero. While 62% of household heads had reported to be married in 2003, this rate dropped to 52% by 2013. *University* is a dummy indicating if the main income earner holds a university degree[14].
East is a binary variable indicating if the household head is currently residing in East Germany. This rate remained relatively constant at around 20%. Note that the definitions of "East Germany" differed between both datasets[15]. To obtain conformance, *East* was set to equal one if a survey participant reported to reside in Saxony, Saxony-Anhalt, Thuringia, Brandenburg or Mecklenburg-Vorpommern. Residents of Berlin are not considered as East Germans in this work due to the unique historical role of the city.
Self indicates whether a respondent is self-employed (1) or not (0). Around five percent of the participants reported to be self-employed in 2003, and this rate decreased to four percent in the 2013 dataset, which equals approximately 1600 survey participants. *Profession* is used to indicate if the main income earner of a household works in a business- or finance-related field, more precisely credit institutions, insurance companies or other financial services providers. *Size* represents the number of household members. German households have slightly above two members on average, and average household size is lower for the 2013 dataset.

4.3. Data quality

The main goal of the EVS is to provide a representative picture of the living conditions in Germany. As it is the basis of many decisions regarding family, business cycle, and tax policy, several means are deployed to ensure high data

2003 dataset, and thus they had to be omitted from the calculation.

[14] Unfortunately, other dummies to assess education could not be included. In particular, dummies indicating each main income earner's high school diploma had to be omitted since the EVS 2003 did not provide sufficient information.

[15] While the EVS 2003 defined East Berlin as part of the eastern states of Germany, the EVS 2013 viewed the entire state of Berlin as part of East Germany.

quality (FSO, 2017, p. 5). The FSO (2005, pp. 36; 2017, pp. 33 ff.) tests every completed questionnaire for obvious mistakes and implausible statements. In addition, *budgeting* is used to compare total expenditure and total income: In theory, both values should be close to equal since expenditure also comprises saving. If the so-called *statistical difference* between both measures was too large, the respective households were contacted to resolve potential mistakes. In case large differences remained inexplicable, the households could be removed from the sample. To ensure a realistic representation of the income situation and consumption patterns, household questionnaires were allocated over the length of a year as this corrects for saisonal effects and sporadic events such as holidays (FSO, 2005, p. 27; FSO, 2017, p. 20).

Despite these measures one should keep in mind that the accuracy of survey results ultimately relies on the willingness of survey participants to give realistic answers. Additionally, Deaton (1997, p. 27) noted that recall errors may occur even for short reporting periods, and this may impair data quality.

While quotation aims to create a representative sample, some socioeconomic groups remain underrepresented: Homeless people are not included in the sample as well as households with a monthly net income of above 18,000€ (FSO, 2005, p. 14; FSO, 2016, p. 3). More generally, quota samples may suffer from bias regarding those characteristics that are not used for quotation. In the EVS 2013, households with heads of foreign nationality as well as households with heads that are over 80 years old are underrepresented. Additionally, the sample is biased toward households with higher education, most likely since participation in the survey requires a certain level of affinity for writing and calculating (FSO, 2017, p. 40). In addition, comparisons to other data sources indicate an under-reporting of income from investment and self-employment which could be explained through the high complexity and sensibility of questions regarding investment income as well as the inconstancy of these income streams. Furthermore, households may have problems to correctly identify personal drawings (FSO, 2017, p. 41). It should also be noted that households might tend to under-report their net wealth as this may be considered highly private information.

Anonymization may bias household data since some positions had to be adjusted to prevent households from being distinctly identifiable. While aggregates over households are not (significantly) affected by anonymization, data on individual households is likely affected in its accuracy. This demands some caution in interpreting estimation results on a household level, in particular for very small groups within the subsample.

Since this work analyses both the EVS 2003 and 2013, comparability of the results obtained from the datasets is desirable. The FSO (2016, p. 9) states there are no significant differences between the EVS 2003 and 2013 with regard to methodology. Importantly, the definitions of components of financial assets are identical (FSO, 2004, p. 8; FSO, 2014, p. 8)[16]. The results presented in both surveys are thus considered comparable and one should not expect differences in estimation results to be caused by a disparity in sample selection or survey design.
In summary, there are some limitations with regard to data quality. Nevertheless, the samples offer a large number of observations and detailed, relatively accurate information on household income, consumption and saving, which form the data basis of the estimation results being presented in the following.

5. Estimation Results

To analyze the effects of income and other covariates on net demand for financial assets, I first present OLS regression result. Quantile regression results are addressed subsequently, followed by a discussion of factors limiting the significance of results.

5.1. OLS Regression Results

Three different models were estimated via OLS with respect to the presumed effect of income on demand. The first model assumes a linear relationship between income and demand. This will be called the Linear Model in this section[17]. While this presumes a very simple relationship of demand and income, linearity of Engel curves could not be rejected for some commodity groups by Banks, Blundell and Lewbel (1997) using nonparametric regression. I also examine if there is a quadratic effect of disposable household income on demand (Quadratic Model). This would imply that income increases have a particularly strong effect on demand for those with already high income, which seems natural under the assumption of a decreasing marginal propensity to consume.

[16] There are some exceptions, but only for individual assets. The aggregate of monetary assets as used in this work covers identical asset classes for both samples.

[17] Note that all regression models estimated in this work are linear in their parameters. However, the term “linear” is used here to describe that *Income* only has an exponent of one. Analogous logic applies to the other models.

Lastly, a model similar to Engel curves proposed by Hausman, Newey and Powell (1995) and Blanks, Blundell and Lewbel (1997) including both the logarithm and the squared logarithm of income (Logarithmic Model) was tested as this has been found to provide an accurate description of the income-demand nexus for some commodity groups.

For each of these models, three different specifications are reported in Tables 5, 6 and 7 for 2003 and in Tables 8, 9 and 10 for 2013: Version (1) only includes those variables that proved to be most significant, while (2) tests all explanatory variables discussed above. Variant (3) examines if the effect of gender on demand changes when accounting for potentially beneficial effects of marriage through an interaction term. Income is given in 1000€ to improve interpretability of coefficients. The same applies to *Wealth*.

Income is found to be correlated with higher net demand irrespective of the number of included variables or the dataset in the Linear Model: An increase in monthly disposable household income by 1000€ is associated with an increment in net demand of around 200€. However, OLS suggests the effect of income on demand is unlikely to be linear. Including a quadratic term lets the coefficient on *Income* become insignificant, and the overall effect of income appears to be weaker for lower levels of household income than suggested by the Linear Model. However, at the highest income levels (i.e. close to $18,000$€ per month), increases of income by 1000€ are associated with demand increases of over 300€[18]. The Logarithmic Model is able to capture the most flexible relationship between income and demand, nameley of the form $\frac{\partial Demand}{\partial Income} = \beta_1 \cdot \frac{1}{Income} + \beta_2 \cdot (log(Income)+1) + \beta_3 \cdot log(Income) \cdot (log(Income)+2)$. OLS results for this model contrast previous results in that they indicate a decreasing marginal effect of income on demand, which even becomes negative at income levels of around 3380€ or 2150€ for the 2003 and 2013 data, respectively. See below for an assessment of the validity of these results.

Interestingly, wealth increases appear to be associated with decreases in demand in 2003, but with increasing demand in 2013. However, coefficients are insignificant across all models, which prohibits a meaningful interpretation of this finding. This poses a surprising result as Campbell (2006, pp. 1555 and 1565) designates wealth as one of the main determinants in household financial behaviour. Similar results have been found by van Rooij, Lusardi and Alessie (2011) and Halko, Kaustia and Alanko (2012). One possible explanation could be that the effect of wealth might be negative for those with low (negative)

[18] Note that results for 2013 might suggest there is a negative relationship between income and demand for very low income. However, coefficients on *Income* are not significant, and thus not too much emphasis should be put on this result.

net demand and positive for those with high net demand, thus not making the coefficient significantly different from zero when using OLS. If that was true, we should see statistically significant estimates for *Wealth* when using quantile regression.

Age is found to have a negative and statistically significant effect on net demand for results from the EVS 2013, while the effect is not significant at the 10% level for data from 2003, except in the Linear Model. Negative age effects on average could indicate that there is a considerable share of people in retirement that dissave to finance their everyday consumption. It remains unclear why the effect is insignificant for the 2003 data, but life-cycle theory of consumption and saving suggests that saving is positive during the worklife and decreases (and generally becomes negative) when households enter retirement (for example, see Browning and Crossley, 2001). While the average effect may be insignificant, there could be a positive relation between age and demand until retirement and a clearly negative relationship afterwards. Therefore, an analysis using dummy variables for different age groups could provide revealing insights. Since the focus of this work lies on the income-demand nexus, however, this will not be further investigated.

The effect of gender on overall household saving is not statistically significant. Including the interaction term to account for beneficial effects of marriage increased the coefficients in all but one case. However, these are also found to be not significant. While there is a rich literature exploring the gender gap in saving behaviour, it seems these differences rather appear on the level of individual asset classes. Halko, Kaustia and Alanko (2012, p. 67) demonstrated that gender effects can be observed in the demand for stocks (conditional on participating in the stock market) and persist when controlling for several factors[19]. For example, differences in investment knowledge apparently do not explain the gender gap fully and it remains constant over the life cycle (Halko, Kaustia and Alanko, 2012, pp. 67 and 75). Even women who are single, divorced or widowed and thus have to take care of their financial matters single-handedly show lower levels of financial literacy than their male counterparts (Bucher-Koenen et al., 2017, p. 271). Risk attitude has also been shown to be correlated with gender in several studies (see, for example, Halko, Kaustia and Alanko, 2012; Dohmen et al., 2011). Pooling several asset classes with different risk profiles together, however, seems to mitigate the effect of gender.

Marriage is found to have a negative correlation with income: monthly net

[19] In contrast, there is evidence suggesting that gender differences in stock market *participation* cannot be explained by gender itself, but by other covariates (Almenberg and Dreber, 2015, p. 142; Halko, Kaustia and Alanko, 2012, p 67).

demand for financial assets decreases by up to 180€ if the household head stated to be married. The effect is generally shown to be more pronounced in the 2013 dataset, particularly in the Linear Model.

Having a university degree is associated with lower net demand for financial assets across all model specifications. However, the effect and its statistical significance are weaker in 2013.

Regional differences in saving behaviour also become apparent from the regression results. Households residing in East Germany show higher asset demand on average for both periods and across different models. It remains unclear from the OLS results if higher average net demand is mainly caused by less dissaving during retirement or higher net demand across all quantiles. One could assume that East Germans who had to accumulate wealth during the times of the GDR and are retired now have less potential to divest due to lower wealth levels. This could increase average net demand and thus present an incomplete picture of regional differences in the demand for financial assets. Examining the effects of *East* at different quantiles should be useful in this regard.

While working in a finance-related field or being self-employed apparently do not have a significant effect on net asset demand, OLS suggests that an increase in the number of household members by one decreases net asset demand by around 100€[20]. This supports the hypothesis that larger households show lower net demand because of higher living expenses and lower income per capita.

The question remains as to which of the models provides the best description of the income-demand nexus. The following analysis focuses on specification (1) of the respective models as this includes those factors that have been found to be most significant. Due to the wide spread of the demand distribution, adjusted R^2 is low for all model specifications. While the Linear Model shows slightly lower explanatory power according to this measure, the Quadratic and Logarithmic Model do not increase the explained variance considerably. This result is somewhat disappointing as it questions the existence of a strong relationship between income and net demand for financial assets. A wide range of potential explanatory factors has already been included, but the variation household in saving decisions still cannot be explained to a large extent.

To test the models for misspecification, the RESET test by Ramsey (1969) has been applied to assess if the explanatory power of the models can be improved by including squared, cubic or quartic fitted values (von Auer, 2016, p. 345 f.). If the null hypothesis that coefficients on these additional factors are all zero is

[20] In fact, the effect varies between -76€ and -142€.

rejected, this indicates misspecification of the model. Given the large datasets and wide dispersion of demand, the test results indicate misspecification for all of the models (see Tables 11 and 12).
Differences between the models become apparent when considering potential problems arising from multicollinearity. Obviously, perfect multicollinearity is not an issue in any of the specifications, but imperfect multicollinearity can still cause coefficients to be imprecisely estimated (Stock and Watson, 2015, p. 251). To assess the impact of imperfect mutlicollinearity on the precision of estimates, Variance Inflation Factors (VIF) are calculated for each of the three models. The VIF is a number indicating the increase in variance of the slope estimate due to multicollinearity. As a rule of thumb, a variance inflation factor above 10 is often considered an indication of potential problems in the estimate's precision (Stine, 1995, pp. 53 f.). Variance Inflation Factors are reported in Tables 13 and 14. Values for the Linear and Quadratic Model are considered unproblematic[21]. However, the Logarithmic Model shows very high values for two of the income terms. Since the income-demand nexus is the main focus of this work, multicollinearity may impair the validity of these coefficient estimates. Consequently, the Logarithmic Model has been excluded from further investigation.
Since a major motivation for the use of quantile regression is the existence of heteroscedasticity, the White test (White, 1980) has been applied to test for varying variance of the error terms. The test has been conducted for the Quadratic Model only as this model is already more flexible than the Linear Model, and its results indicate the presence of heteroscedasticity (see Table 15). While statistical inference is not constrained if heteroscedasticity-robust standard errors are used, OLS still fails to provide a description of the varying spread of the demand distribution.
At this point it should also be noted that OLS results could be affected by outliers and removing some of these values may improve the OLS results. However, they remained included in the data to ensure comparability to the results of quantile regression as it is one of the main advantages of the latter to be less sensitive to outliers.
In summary, the presumption that average effects might not provide a precise description of the relationship between some explanatory variables and demand due to the large scatter of the data and the issue of heteroscedasticity are all indications that the use of quantile regression may be beneficial to provide a

[21] The cutoff value of 10 is by no means undisputed. For example, O'Brien (2007) argues that even higher values for the VIF (for example, a VIF of 40) do not need to put regression results into question.

more comprehensive description of the demand distribution. Its results will be presented in the following subsection.

5.2. Quantile Regression Results

Since the Logarithmic Model may suffer from imperfect multicollinearity, quantile regression has only been performed for the Linear and the Quadratic Model. Results for the 10th, 25th, 50th, 75th and 90th percentile are reported in Tables 16 to 19[22].

The Linear Model indicates that the relation between income and demand is negative at the 10% quantile. Net demand at this quantile is generally negative (see Table 4), and this means those who are dissaving tend to dissave more if their income is higher. For higher quantiles, the effect of *Income* becomes strongly positive, and income increases of 1000€ are associated with increases of the 90% quantile of demand for financial assets of over 400€ in the 2003 data. A similar pattern is observable for the 2013 dataset, although the effects of income are slightly weaker. Heteroscedasticity-robust standard errors indicate statistical significance of the coefficient estimates at the 1% level across all but one quantile: The quantile at $\tau = 0.25$ is special in the sense that it equals zero across all income levels (see Table 4). Since much higher quantiles are also zero (e.g. at the 35% level), coefficient estimates are zero at this quantile for the 2003 data and not statistically significant[23]. Similar effects of coefficients close to zero are found for the 2013 dataset, albeit less pronounced.

The results for the Quadratic Model propose a similar income-demand relation as already found in the linear case. At low quantiles, income increases lower the demand quantile, while the effect of income becomes positive at higher quantiles, and demand increases at the 90th percentile are around $130 + 60 \cdot Income$ in 2013 and even higher in 2003. Therefore, quantile regression seems to provide revealing insights in comparison to OLS since the income-demand nexus can now be described more accurately. High-income households do not necessarily demand more financial assets than lower-income peers, but the effect

[22] Note that for the estimation of standard errors, a method accounting for non-identical distribution of errors has been used that calculates local density estimates (Koenker, 2018, p. 88). In some cases, these estimates were negative and thus automatically set to zero. While this does not impair coefficient estimates, a large number of negative density estimates may be an indication for misspecification. However, this issue will not be discussed in any furhter detail within the scope of this work.

[23] A closer look at the estimation results shows that coefficient estimates are not exactly, but very close to zero. Given the number of decimal points reported in Table 16, however, estimates appear indistinguishable from zero.

depends on the quantile of the demand distribution. At low τ, income increases would be associated with a decrease of the quantile. This could have various reasons. Given the fact that household data was collected over one quarter, dissaving to finance high one-time expenses such as cars or vacation trips could yield negative net demand. Dissaving during retirement would also be associated with negative values for net asset demand, and in both cases higher-income households could be expected to show stronger dissaving. Analogously, high net demand for financial assets can also be inferred to be higher for those households with high disposable income since these generally have more potential to save.

For the 2003 dataset, most coefficients at the lower quartile are again not significantly different from zero and similar, but attenuated effects are apparent for the 2013 data. This property also becomes apparent when looking at the illustration of estimation results in Figures 1 to 4. For each of the models and datasets, quantile regression estimates are depicted conditional on τ. Black dots indicate the respective coefficient estimates. Since there is a finite number of estimates, the solid black lines connecting the dots are only approximations for the real pattern of the conditional quantile function (see Davino, Furno and Vistocco, 2013, p. 51). The shaded gray areas represent 90% pointwise confidence bands. OLS estimates are given by the solid red lines, while dotted red lines are used to indicate borders of the 90% confidence intervals (Koenker and Hallock, 2001, p. 149).

From Figures 1 and 4, the special behaviour of coefficient estimates at the 25% quantile also becomes apparent: For some coefficients, the estimates jump to a value of approximately zero, before returning to values close to their previous level. One can also clearly see the increasing effect of income on demand depending on the quantile, which casts doubt on the OLS results.

In addition, the graphs allow for a comparison of conditional median estimates and estimates for the conditional mean effect. Under the assumption of a symmetric distribution, mean and median should be (approximately) equal. However, estimates are considerably apart at $\tau = 0.5$, where OLS seems to indicate a stronger positive effect of income on demand than described by quantile regression. This suggests there may be some high-income households showing very large demand for financial assets and pulling up the estimate for the conditional mean. Additionally, outliers may disturb OLS results and thus cause a discrepancy between median and mean estimates.

The results for other coefficients also contrast the previous findings from OLS. While no significant effect of *Wealth* on *Demand* was found using the Least

Squares approach, quantile regression reveals the effect of wealth may be significant and increasing with τ. Again, higher wealth is correlated with lower net demand at low quantiles, but with higher demand at high quantiles. This effect is very similar for the Linear and the Quadratic Model and both datasets. The impact of marriage is also found to be varying across quantiles, with a particularly strong effect at the 10th percentile. Results based on the EVS 2003 imply that households with married household heads dissave around 180€ more than their peers, and this effect is found to be even stronger in 2013. For the other quantiles, the effect is considerably less severe, although it remains negative.

The effect of education on net financial asset demand remains surprising even under the use of quantile regression. Across all quantiles, the coefficient on *University* is reported below zero. While the effect is particularly strong for the lowest quantile under consideration, the suggested negative impact of higher educational levels on net savings in monetary assets at higher quantiles remains inexplicable. Additionally, differences between the 2003 and 2013 results cannot be explained. However, coefficient estimates are often insignificant for the 2013 data, and the effect of education on savings in financial assets may thus be considered negligible.

The 10% demand quantile is higher for East German households. While this effect diminishes for high quantiles in the 2013 dataset, it increases based on the 2003 data. Apparently, there was a East-West gap in dissaving in 2013, and one might conclude that West German households show higher potential to divest their accumulated wealth, which had already been conjectured because of limited access to capital markets in the GDR.

The negative effect of increases in household size on net asset demand found by OLS is corroborated by the results from quantile regression. However, the effect appears to be constant in neither the Linear nor the Quadratic Model for any of the two observation periods. In fact, large households sizes tend to be associated with higher net demand at the 10th percentile, but the coefficients decrease subsequently. At the 90% quantile, an increase in household size by one member is accompanied by a decrease of demand at this quantile of up to 130€.

An assessment of the explanatory power of the models can be conducted similarly to OLS. However, Koenker and Machado (1999) proposed the use of $R^1(\tau)$ instead of adjusted R^2. While adjusted R^2 is a global measure comparing the unweighted sum of squared residuals of an unrestricted versus a restricted model (typically the intercept-only model), $R^1(\tau)$ is a local measure depending on the quantile τ of the dependent variable. The measure can thus be denoted as

$$R^1(\tau) = 1 - \frac{\min_b \sum \rho_\tau(y_i - x_i'b)}{\min_{b_1} \sum \rho_\tau(y_i - x_{1i}'b_1)}, \tag{9}$$

where ρ_τ denotes the check function at τ, b is a vector of coefficients and $x_{1i}'b_1$ are the fitted values of the intercept-only model (Koenker and Machado, 1999, p. 1297). As for adjusted R^2, $R^1(\tau)$ ranges between 0 and 1, where higher values indicate stronger explanatory power. Results for the measure are reported in the quantile regression tables (Tables 16 to 19) and indicate that the Linear and Quadratic Model explain a similar share of the variation in the demand distribution. This holds true for both datasets and across quantiles. However, $R^1(\tau)$ is highest for the upper quartile and the 90th percentile, while lowest values are observed for the median estimate.

While Figures 1 to 4 already suggest that the coefficients vary across quantiles as the quantile regression estimates lie outside the 90% confidence intervals of OLS estimates for at least some quantiles, the results of two formal testing procedures will be presented in the following to further support this hypothesis (see Tables 20 to 22 for detailed test results). First, a Wald test is conducted to test the equality of slope coefficients at the 10th, 25th, 50th, 75th and 90th percentile. The hypothesis is strongly rejected, indicating that the effect of income on demand in fact depends on the quantile of demand under consideration. Rejection of the equality of slopes is analogous to the rejection of a location-shift model, and an approach applied to the field of quantile regression by Koenker and Xiao (2002) can be used to test this hypothesis for a finite number of equally spaced values of τ in a given interval (Koenker and Xiao, 2002, pp. 1597 ff.; Koenker, 2005, p. 103). Test results reject the location-shift model.

However, Figures 1 to 4 also question that the relationship between demand and the covariates is described appropriately by a location-scale model. In this case, one would expect all coefficients to have similar shapes when expressed as functions of τ, which is not apparent (Koenker, 2005, p. 103). The test proposed by Koenker and Xiao (2002) can also be used to the location-scale model and the results show that this model is also rejected, therefore indicating that the covariates do not only change the location and scale of the conditional distribution of *Demand*, but also its shape (Koenker and Xiao, 2002, p. 1584). This feature cannot be captured by any of the models presented above and shall not be discussed in any more detail here as it seems to exceed the scope of this work. Instead, the following subsection will quickly deal with some further limitations.

5.3. Limitations

Regarding the choice of model specifications it should be mentioned that there is no obvious relationship between income and demand apparent from the data. Economic theory also does not suggest a specific parametric form. While there have been some considerations regarding the choice of parameters, one could still argue that the choice of regression models seems rather arbitrary and is merely based on the attempt to fit a regression line through a scatterplot. Therefore, nonparametric regression might be advantageous as it does not assume any specific parametric form, which can be considered a fundamental benefit of this technique (Bierens and Pott-Butter, 1990, p. 124).
Explanatory power of OLS is very low as measured by adjusted R^2 despite a large number of included variables. An improvement would have been not to use proxies for factors such as financial literacy, but variables directly measuring certain characteristics. Additionally, some other factors could not be included that might affect household net asset demand. For example, behavioural factors such as the attitude toward the financial system, the reaction to variations in the macroeconomic environment and the overall motivation to relinquish current consumption in favor of saving for purposes of retirement income or consumption smoothing might serve as additional explanatory variables. Alas, no information of this kind was available in the data. While these factors could explain inter-household variation within the same time period, intertemporal differences in demand behaviour could be based on macroeconomic factors such as the current labour market situation, recent developments on financial markets and market sentiment.
Given these constraints, one might consider a combination of different data sources to take account of the aforementioned notions. In addition, the comparison of only two datasets is insufficient to ascertain long-term trends in net asset demand and the role of its explanatory factors. Using a larger number of periods might prove promising in this regard. Furthermore, the EVS does not include information on households' cash holdings and check account balances. Bearing the rather conservative investment behaviour of German households in mind, a considerable share of their wealth may be parked in these alternative monetary assets and could hence distort the results.
Ultimately, the quality of the estimation output is dependent on the quality of

the input, and some issues regarding data quality have already been covered in section 4.3.

6. Conclusion

This work aimed to contribute to the existing literature in the intersection of Engel curve analysis and household finance by providing a description of the interrelation between income as well as other covariates and demand of German households for financial assets.

Given the varying spread of the conditional demand distribution, OLS regression was unable to provide an accurate approximation of the income-demand relationship. Explanatory power of all models was low despite a considerable number of independent variables such as wealth, educational level, marital status, region or household size. The income-demand relationship was found to be rather weak in terms of explained variance.

This demonstrated that the dynamics behind household saving behaviour are hard to capture by simple regression equations as the processes determining saving decisions are very complex. Quantile regression allowed to provide a more comprehensive picture of the demand distribution as variations across quantiles could be observed. The implicit OLS assumption of constant coefficients across quantiles of the demand distribution could be rejected, and negative income effects were found for negative values of net demand, while higher income is associated with higher demand for higher quantiles. Therefore, quantile regression has proven to be a useful technique to deal with widely spread distributions, especially those featuring heteroscedasticity and large outliers.

Within the scope of this work, I focused on a mere description of the connection between net demand for financial assets and its explanatory factors. Given the significance of a profound understanding of the determinants of household financial decisions, future research could explore the underlying causal relationships. Consideration of additional variables such as psychological or macroeconomic factors may be promising. Further, the inclusion of data on cash holdings and check account balances might improve explanatory power of the models. Additionally, a comparison of different income sources could be interesting. For example, it could be examined if labour and investment income differ in their effect on demand for financial assets. Distinguishing different income sources has been a practice for several decades and could be of particular interest with regard to household saving behaviour over the business cycle since labour income streams may be considered relatively stable, while investment income might not (Ando and Modigliani, 1963, p. 73). Moreover, time series analysis could yield insights regarding long-term trends in household financial decisions that have not become apparent from the comparison of only two survey periods.

Ultimately, a different estimation technique may also be considered. Since there is no clear guidance from economic theory as to which parametric form might provide the best fit to the data, nonparametric estimation methods could be advantageous. Bierens and Pott-Buter (1990, p. 124) argue that the choice of a parametric model is usually made with regard to "tractability rather than on the basis of a priori knowledge of the true functional form" and may thus be imprecise or wrong. Koenker (2005, p. 222) noted that nonparametric quantile regression may provide a very flexible framework for certain occasions, and given the results in this work further improvements could be achieved through its use.

References

Aitchison, J. and Brown, J.A.C (1954) 'A Synthesis of Engel Curve Theory', *The Review of Economic Studies*, 22(1), pp. 35–46.

Almenberg, J. and Dreber, A. (2015) 'Gender, stock market participation and financial literacy', *Economics Letters*, 137(C), pp. 140–142.

Ando, A. and Modigliani, F. (1963) 'The "Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests', *The American Economic Review*, 53(1), pp. 55–84.

Angrist, J.D. and Pischke, J.-S. (2009) *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006) 'Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure', *Econometrica*, 74(2), pp. 539–563.

Banks, J., Blundell, R. and Lewbel, A. (1997) 'Quadratic Engel Curves and Consumer Demand', *The Review of Economics and Statistics*, 79(4), pp. 527–539.

Bierens, H.J. and Pott-Buter, H.A. (1990) 'Specification of Household Engel Curves by Nonparametric Regression', *Econometric Reviews*, 9(2), pp. 123–184.

Browning, M. and Crossley, T.F. (2001) 'The Life-Cycle Model of Consumption and Saving', *Journal of Economic Perspectives*, 15(3), pp. 3–22.

Browning, M. and Lusardi, A. (1996) 'Household Saving: Micro Theories and Micro Facts', *Journal of Economic Literature*, 34(4), pp. 1797–1855.

Bucher-Koenen, T. and Lamla, B. (2014) *The Long Shadow of Socialism: On East-West German Differences in Financial Literacy*. Munich: Munich Center for the Economics of Aging. MEA Discussion Papers 282–2014.

Bucher-Koenen, T. and Lusardi, A. (2011) 'Financial Literacy and Retirement Planning in Germany', *Journal of Pension Economics and Finance*, 10(4), pp. 565–584.

Bucher-Koenen, T., Lusardi, A., Alessie, R. and van Rooij, M. (2017) 'How Financially Literate Are Women? An Overview and New Insights', *The Journal of Consumer Affairs*, 51(2), pp. 255–283.

Cagetti, M. (2003) 'Wealth Accumulation over the Live Cycle and Precautionary Savings', *Journal of Business Economics Statistics*, 21(3), pp. 339–353.

Campbell, J.Y. (2006) 'Household Finance', *The Journal of Finance*, 61(4), pp. 1553–1604.

Chakrabarty, M. and Hildenbrand, W. (2011) 'Engel's Law Reconsidered', *Journal of Mathematical Economics*, 47(3), pp. 289–299.

Davino, C., Furno, M. and Vistocco, D. (2013) *Quantile Regression: Theory and Applications*. Chichester: Wiley.

Deaton, A. (1986) 'Demand Analysis', in Griliches, Z. and Intriligator, M.D. (eds.) *Handbook of Econometrics, Volume III*. New York, NY: Elsevier, pp. 1768–1839.

Deaton, A. (1997) *The Analysis of Household Surveys – A Microeconometric Approach to Development Policy*. Baltimore, MD: Published for the World Bank by John Hopkins University Press.

Deaton, A. and Muellbauer, J. (1980) *Economics and consumer behaviour*. Cambridge: Cambridge University Press.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G.G. (2011) 'Individual Risk Attitudes: Measurement, Determinants, and Behavioural Consequences', *Journal of the European Economic Association*, 9(3), pp. 522–550.

Engel, J. and Kneip, A. (1996) 'Recent Approaches to Estimating Engel Curves', *Journal of Economics*, 63(2), pp. 187–212.

Farrell, M.J. (1954) 'The Demand for Motor-Cars in the United States', *Journal of the Royal Statistical Society. Series A (General)*, 117(2), pp. 171–201.

Federal Statistical Office (2004) *Einkommens- und Verbrauchsstichprobe – Geldvermögenbestände und Konsumentenkreditschulden privater Haushalte.* Wiesbaden: Statistisches Bundesamt. Fachserie 15 Heft 2.

Federal Statistical Office (2005) *Einkommens- und Verbrauchsstichprobe – Aufgabe, Methode und Durchführung der EVS*. Wiesbaden: Statistisches Bundesamt. Fachserie 15 Heft 7.

Federal Statistical Office (2014) *Einkommens- und Verbrauchsstichprobe – Geld- und Immobilienvermögen sowie Schulden privater Haushalte*. Wiesbaden: Statistisches Bundesamt. Fachserie 15 Heft 2.

Federal Statistical Office (2016) *Einkommens- und Verbrauchsstichprobe – Qualitätsbericht*. Wiesbaden: Statistisches Bundesamt.

Federal Statistical Office (2017) *Einkommens- und Verbrauchsstichprobe – Aufgabe, Methode und Durchführung*. Wiesbaden: Statistisches Bundesamt. Fachserie 15 Heft 7.

Forsyth, F.G. (1960) 'The Relationship between Family Size and Family Expenditure', *Journal of the Royal Statistical Society. Series A (General)*, 123(4), pp. 367–397.

Greene, W.H. (1993) *Econometric Analysis*. 2nd edn. Englewood Cliffs, NJ: Prentice Hall.

Guiso, L., Sapienza, P. and Zingales, L. (2008) 'Trusting the Stock Market', *The Journal of Finance*, 53(6), pp. 2557–2600.

Halko, M.-L., Kaustia, M. and Alanko, E. (2012) 'The Gender Effect in Risky Asset Holdings', *Journal of Economic Behaviour Organization*, 83(1), pp. 66–81.

Hausman, J.A., Newey, W.K. and Powell, J.L. (1995) 'Nonlinear Errors in Variables Estimation of some Engel Curves', *Journal of Econometrics*, 65(1), pp. 205–233.

Hendricks, W. and Koenker, R. (1992) 'Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity', *Journal of the American Statistical Association*, 87(417), pp. 58–68.

Koenker, R. (2005) *Quantile Regression*. New York, NY: Cambridge University Press.

Koenker, R. (2018) *Package 'quantreg'*. Available at: https://cran.r-project.org/web/packages/quantreg/quantreg.pdf (Accessed: 02/07/2018).

Koenker, R. and Basset, G. (1978) 'Regression Quantiles', *Econometrica*, 46(1), pp. 33–50.

Koenker, R. and Basset, G. (1982) 'Robust Tests for Heteroscedasticity Based on Regression Quantiles', *Econometrica*, 50(1), pp. 43–61.

Koenker, R. and Hallock, K.F. (2001) 'Quantile Regression', *Journal of Economic Perspectives*, 15(4), pp. 143–156.

Koenker, R. and Machado, J.A.F. (1999) 'Goodness of Fit and Related Inference Processes for Quantile Regression', *Journal of the American Statistical Association*, 94(448), pp. 1296–1310.

Koenker, R. and Xiao, Z. (2002) 'Inference on the Quantile Regression Process', *Econometrica*, 70(4), pp. 1583–1612.

Leser, C.E.V. (1963) 'Forms of Engel Functions', *Econometrica*, 31(4), pp. 694–703.

Lewbel, A. and Houthakker, H.S. (2008) 'Engel Curve', in Durlauf, S. and Blume, L.E. (eds.) *The New Palgrave Dictionary of Economics*. 2dn edn. London: Palgrave Macmillan, pp. 847–850.

Lusardi, A. (2009) 'Household Savings Behaviour in the United States:

The Role of Literacy, Information, and Financial Education Programs', in Foote, C.L., Goette, L. and Meier, S. (eds.) *Policymaking Insights from Behavioural Economics*. Boston, MA: Federal Reserve Bank of Boston, pp. 109–149.

Manning, W.G., Blumberg, L. and Moulton, L.H (1995) 'The Demand for Alcohol: The Differential Response to Price', *Journal of Health Economics*, 14(2), pp. 123–148.

O'Brien, R. (2007) 'A Caution Regarding Rules of Thumb for Variance Inflation Factors', *Quality & Quantity*, 41(5), pp. 673–690.

Perthel, D. (1975) 'Engel's Law Revisited', *International Statistical Review*, 43(2), pp. 211–218.

Ramsey, J.B. (1969) 'Tests for Specification Errors in Classical Linear Least-squares Regression Analysis', *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2), pp. 350–371.

Stine, R.A. (1995) 'Graphical Interpretation of Variance Inflation Factors', *The American Statistician*, 49(1), pp. 53–56.

Seguino, S. and Floro, M.S. (2003) 'Does Gender have any Effect on Aggregate Saving? An empirical analysis', *International Review of Applied Economics*, 17(2), pp. 147–166.

Stock, J.H. and Watson, M.W. (2015) *Introduction to Econometrics*. 3rd edn. New York: Prentice Hall.

Van Rooij, M., Lusardi, A. and Alessie, R. (2011) 'Financial Literacy and Stock Market Participation', *Journal of Financial Economics*, 101(2), pp. 449–472.

Von Auer, L. (2016) *Ökonometrie*. 7th edn. Heidelberg: Springer.

White, H. (1980) 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica*, 48(4), pp. 817–838.

Working, H. (1943) 'Statistical Laws of Family Expenditure', *Journal of the American Statistical Association*, 38(221), pp. 43–56.

A. Variable Definitions

Table 1: Variable Definitions

Variable	Definition	Components
Demand	Monthly net household demand for financial assets in 1000€.	Includes investment in savings accounts and other investment products such as call money or fixed deposits, buildings loan contracts, bonds, stocks, investment funds and other securities and company shares. From this, withdrawals from savings accounts as well as term deposit accounts, call money accounts or building loan contracts and divestment of stocks and other securities and company shares were subtracted.
Income	Monthly disposable household income in 1000€.	As defined in the EVS.
Wealth	Net household wealth in respective assets as stated per the 1st of January on the EVS questionnaire.	Includes building savings deposits, savings deposits and other investment products such as call money or fixed deposits, stocks, bonds, investment funds, other securities and company shares. Outstanding consumer credit debt was subtracted.

B. Descriptive Statistics

Table 2: Descriptive Statistics - 2003

Statistic	N	Mean	St. Dev.	Min	Max
Demand	42,680	210.11	2,174.76	−59,737	69,648
Income	42,680	3,534.90	2,138.57	8	17,829
Wealth	42,680	33,204.19	73,397.22	−255,191	2,614,081
Female	42,680	0.32	0.47	0	1
Age	42,680	50.41	14.63	20	85
Married	42,680	0.62	0.49	0	1
University	42,680	0.18	0.38	0	1
East	42,680	0.20	0.40	0	1
Self	42,680	0.05	0.22	0	1
Profession	42,680	0.04	0.19	0	1
Size	42,680	2.42	1.22	1	7

Table 3: Descriptive Statistics - 2013

Statistic	N	Mean	St. Dev.	Min	Max
Demand	42,756	138.04	2,928.21	−92,278	156,243
Income	42,756	3,633.91	2,273.75	20	17,944
Wealth	42,756	40,634.85	95,525.63	−240,149	2,657,426
Female	42,756	0.38	0.49	0	1
Age	42,756	53.30	15.67	20	85
Married	42,756	0.52	0.50	0	1
University	42,756	0.15	0.35	0	1
East	42,756	0.20	0.40	0	1
Self	42,756	0.04	0.19	0	1
Profession	42,756	0.04	0.19	0	1
Size	42,756	2.10	1.09	1	7

Table 4: Unconditional Demand Quantiles - 2003 and 2013

	2003	2013
5%	-1,216	-1,427
10%	-493	-643
15%	-183	-303
20%	0	-100
25%	0	0
30%	0	0
35%	0	0
40%	10	0
45%	50	0
50%	83	30
55%	125	55
60%	175	100
65%	250	152
70%	333	225
75%	450	330
80%	616	471
85%	833	670
90%	1,187	1,039
95%	1,899	1,826

Note: All values in €.
25% Demand quantiles conditional on income are zero across the income groups of width 1,000€ from 0€ to 18,000€.

C. OLS Regression Results

Table 5: Linear Model OLS Results - 2003

	Dependent variable: Demand		
	(1)	(2)	(3)
Income	215.1880***	217.4800***	217.3313***
	(11.5252)	(11.6230)	(11.6253)
Wealth	−0.8264	−0.8782	−0.8798
	(0.7699)	(0.7847)	(0.7845)
Age		1.5593*	1.5621*
		(0.8050)	(0.8162)
Female		15.7678	
		(21.6315)	
$Married * Female$			2.5505
			(36.7797)
Married	−104.5117***	−119.8821***	−127.1728***
	(27.9858)	(29.1984)	(29.7016)
University	−135.1057***	−138.3703***	−138.2822***
	(35.8669)	(36.3161)	(36.3404)
East	110.9427***	108.4622***	110.1070***
	(21.0417)	(20.9574)	(21.2149)
Profession		−145.6703**	−145.3677**
		(66.0410)	(66.0151)
Self		13.6155	12.9021
		(68.7830)	(68.8265)
Size	−111.2502***	−101.8882***	−102.2073***
	(13.6004)	(13.0631)	(13.2054)
Constant	−186.3480***	−283.6833***	−273.4739***
	(27.3297)	(46.9155)	(44.5190)
Observations	42,680	42,680	42,680
R^2	0.0303	0.0306	0.0306
Adjusted R^2	0.0302	0.0304	0.0303
Residual Std. Error	2,141.7050	2,141.5010	2,141.5100
	(df = 42673)	(df = 42669)	(df = 42669)
F Statistic	222.2880***	134.6098***	134.5731***
	(df = 6; 42673)	(df = 10; 42669)	(df = 10; 42669)
Notes:	Heteroscedasticity-robust standard errors in parentheses. Significance levels: *p<0.1, **p<0.05, ***p<0.01.		

Table 6: Quadratic Model OLS Results - 2003

	Dependent variable: Demand		
	(1)	(2)	(3)
Income	29.9251	33.7254	33.5281
	(26.8425)	(27.1685)	(27.2497)
$Income^2$	16.2386***	16.0713***	16.0832***
	(2.6840)	(2.7057)	(2.7095)
Wealth	−0.8570	−0.8971	−0.8966
	(0.7735)	(0.7888)	(0.7886)
Age		1.2783	1.3204
		(0.8107)	(0.8208)
Female		−0.0732	
		(21.3959)	
$Married * Female$			15.5647
			(36.8803)
Married	−29.9607	−49.7943*	−52.6467*
	(27.5103)	(29.2965)	(29.3303)
University	−113.2806***	−114.4950***	−114.6220***
	(35.2761)	(35.7400)	(35.7600)
East	80.1275***	79.7418***	78.0098***
	(20.1184)	(20.1367)	(20.3265)
Profession		−104.7250	−105.0012
		(66.0289)	(66.0162)
Self		−20.5598	−20.4255
		(69.3455)	(69.3810)
Size	−84.4967***	−76.7815***	−76.1749***
	(13.5943)	(13.0029)	(13.1689)
Constant	83.3194**	8.5661	6.2941
	(38.6987)	(58.5441)	(56.8013)
Observations	42,680	42,680	42,680
R^2	0.0343	0.0344	0.0344
Adjusted R^2	0.0341	0.0342	0.0342
Residual Std. Error	2,137.3400	2,137.2650	2,137.2610
	(df = 42672)	(df = 42668)	(df = 42668)
F Statistic	216.3759***	138.3408***	138.3552***
	(df = 7; 42672)	(df = 11; 42668)	(df = 11; 42668)

Notes: Heteroscedasticity-robust standard errors in parentheses.
Significance levels: *p<0.1, **p<0.05, ***p<0.01.

Table 7: Logarithmic Model OLS Results - 2003

	Dependent variable: Demand		
	(1)	(2)	(3)
$Log(Income)$	507.9952***	503.5453***	503.1330***
	(89.0554)	(89.1776)	(89.0963)
$Log(Income) * Income$	−179.2043***	−174.8055***	−174.7869***
	(49.5188)	(49.9495)	(49.9409)
$Log(Income)^2 * Income$	92.4644***	91.0521***	91.0517***
	(18.5475)	(18.7006)	(18.6999)
$Wealth$	−0.8644	−0.8959	−0.8955
	(0.7737)	(0.7891)	(0.7890)
Age		0.9921	1.0256
		(0.8135)	(0.8229)
$Female$		0.2849	
		(21.3792)	
$Married * Female$			12.3046
			(36.8054)
$Married$	−49.2722*	−65.1272**	−67.5047**
	(27.2315)	(29.1615)	(29.2218)
$University$	−107.7244***	−109.0104***	−109.1144***
	(35.1358)	(35.6059)	(35.6263)
$East$	79.1560***	78.5643***	77.2356***
	(20.0838)	(20.0999)	(20.2882)
$Profession$		−106.9696	−107.1764
		(65.9893)	(65.9807)
$Self$		−21.1391	−21.0506
		(69.3538)	(69.3892)
$Size$	−85.6680***	−79.6007***	−79.1252***
	(13.5860)	(12.9669)	(13.1239)
$Constant$	65.8105***	11.6920	10.0447
	(22.9627)	(42.8096)	(40.9851)
Observations	42,680	42,680	42,680
R^2	0.0347	0.0348	0.0348
Adjusted R^2	0.0345	0.0346	0.0346
Residual Std. Error	2,136.9080	2,136.8580	2,136.8550
	(df = 42671)	(df = 42667)	(df = 42667)
F Statistic	191.6901***	128.3005***	128.3087***
	(df = 8; 42671)	(df = 12; 42667)	(df = 12; 42667)
Notes:	Heteroscedasticity-robust standard errors in parentheses. Significance levels: *p<0.1, **p<0.05, ***p<0.01.		

Table 8: Linear Model OLS Results - 2013

	Dependent variable: Demand		
	(1)	(2)	(3)
Income	198.0530***	198.5608***	198.5466***
	(18.1600)	(18.8197)	(18.6542)
Wealth	0.1454	0.1969	0.1968
	(0.5163)	(0.5208)	(0.5204)
Age		−2.0687**	−2.0649*
		(1.0432)	(1.1042)
Female		0.9364	
		(33.8032)	
Married ∗ Female			1.5942
			(67.8483)
Married	−187.4946***	−160.0566***	−160.7836***
	(40.1189)	(47.5785)	(51.8343)
University	−90.7774*	−93.3735*	−93.3857*
	(53.4298)	(53.7666)	(53.5370)
East	97.7126***	96.3499***	96.3005***
	(22.7837)	(22.5059)	(23.5251)
Profession		−103.2219	−103.2197
		(65.9595)	(65.7498)
Self		−88.0124	−88.0543
		(95.0112)	(95.1076)
Size	−130.0395***	−142.4589***	−142.4188***
	(20.8349)	(20.4311)	(20.9252)
Constant	−223.4476***	−97.7366	−97.3588
	(34.4697)	(83.8040)	(78.0409)
Observations	42,756	42,756	42,756
R^2	0.0168	0.0170	0.0170
Adjusted R^2	0.0167	0.0168	0.0168
Residual Std. Error	2,903.6420	2,903.5410	2,903.5410
	(df = 42749)	(df = 42745)	(df = 42745)
F Statistic	122.0767***	73.9468***	73.9468***
	(df = 6; 42749)	(df = 10; 42745)	(df = 10; 42745)
Notes:	Heteroscedasticity-robust standard errors in parentheses. Significance levels: *p<0.1, **p<0.05, ***p<0.01.		

Table 9: Quadratic Model OLS Results - 2013

	Dependent variable: Demand		
	(1)	(2)	(3)
Income	−60.8054	−63.3890	−62.8423
	(39.7017)	(39.0668)	(39.3709)
Income^2	22.7603***	22.9592***	22.9316***
	(4.6170)	(4.5846)	(4.6049)
Wealth	0.1681	0.2291	0.2325
	(0.5168)	(0.5214)	(0.5211)
Age		−2.4837**	−2.4456**
		(1.0195)	(1.0820)
Female		−17.7580	
		(32.5466)	
$Married * Female$			14.2512
			(68.0003)
Married	−105.0186***	−77.6573*	−74.2045
	(35.0179)	(43.0588)	(45.4155)
University	−69.5753	−71.2319	−71.6489
	(52.2863)	(52.6064)	(52.4065)
East	67.9607***	68.1540***	65.3829***
	(21.4999)	(21.5527)	(22.4169)
Profession		−67.8696	−69.2221
		(64.2818)	(64.1784)
Self		−126.3417	−125.4002
		(95.4866)	(95.6384)
Size	−88.4728***	−103.2661***	−102.2462***
	(17.9860)	(18.7245)	(19.2081)
Constant	170.7915**	337.5950***	322.6814***
	(69.2979)	(82.0337)	(82.4786)
Observations	42,756	42,756	42,756
R^2	0.0214	0.0216	0.0216
Adjusted R^2	0.0212	0.0214	0.0214
Residual Std. Error	2,896.9440	2,896.7570	2,896.7650
	(df = 42748)	(df = 42744)	(df = 42744)
F Statistic	133.5373***	85.8534***	85.8330***
	(df = 7; 42748)	(df = 11; 42744)	(df = 11; 42744)

Notes: Heteroscedasticity-robust standard errors in parentheses.
Significance levels: *p<0.1, **p<0.05, ***p<0.01.

Table 10: Logarithmic Model OLS Results - 2013

	Dependent variable: Demand		
	(1)	(2)	(3)
$Log(Income)$	601.9021***	617.4207***	615.5106***
	(166.5172)	(165.8057)	(166.1463)
$Log(Income) * Income$	−287.4538***	−295.8248***	−294.5099***
	(89.9065)	(89.2472)	(89.5460)
$Log(Income)^2 * Income$	132.3145***	135.2077***	134.7696***
	(34.0633)	(33.8287)	(33.9383)
$Wealth$	0.1740	0.2419	0.2459
	(0.5167)	(0.5214)	(0.5211)
Age		−2.8007***	−2.7683***
		(1.0087)	(1.0716)
$Female$		−23.0256	
		(32.3018)	
$Married * Female$			11.0903
			(68.0363)
$Married$	−116.5369***	−86.7173**	−80.5322*
	(35.7619)	(43.5564)	(45.8397)
$University$	−63.9337	−65.9109	−66.3779
	(51.8145)	(52.1698)	(51.9677)
$East$	61.4313***	61.7194***	58.7840***
	(21.3863)	(21.4740)	(22.3060)
$Profession$		−70.4482	−71.9918
		(64.3735)	(64.2734)
$Self$		−126.3791	−125.1630
		(95.4986)	(95.6477)
$Size$	−87.1429***	−104.0811***	−103.0719***
	(17.9611)	(18.7573)	(19.2489)
$Constant$	49.2677*	233.7977***	216.5871***
	(26.6371)	(66.3841)	(63.7202)
Observations	42,756	42,756	42,756
R^2	0.0217	0.0220	0.0220
Adjusted R^2	0.0216	0.0217	0.0217
Residual Std. Error	2,896.4640	2,896.2180	2,896.2330
	(df = 42747)	(df = 42743)	(df = 42743)
F Statistic	118.7800***	80.1384***	80.1005***
	(df = 8; 42747)	(df = 12; 42743)	(df = 12; 42743)

Notes: Heteroscedasticity-robust standard errors in parentheses. Significance levels: *p<0.1, **p<0.05, ***p<0.01.

D. OLS Regression Test Results

Table 11: RESET test results - 2003

	Tested Model: 2nd, 3rd and 4th powers of fitted response.		
	Linear	Quadratic	Logarithmic
RESET	101.52	26.61	19.67
df1	3	3	3
df2	42,670	42,669	42,668
p-value	< 0.001	< 0.001	< 0.001

Table 12: RESET test results - 2013

	Tested Model: 2nd, 3rd and 4th powers of fitted response.		
	Linear	Quadratic	Logarithmic
RESET	137.78	28.25	18.12
df1	3	3	3
df2	42,746	42,745	42,744
p-value	< 0.001	< 0.001	< 0.001

Table 13: Variance Inflation Factors - 2003

	Dependent variable: Demand		
	Linear	Quadratic	Logarithmic
$Income$	1.55	9.91	
$Income^2$		8.18	
$Log(Income)$			18.14
$Log(Income) * Income$			230.05
$Log(Income)^2 * Income$			145.65
$Wealth$	1.11	1.11	1.11
$Married$	1.58	1.65	1.69
$University$	1.09	1.09	1.09
$East$	1.05	1.06	1.06
$Size$	1.65	1.71	1.71

Table 14: Variance Inflation Factors - 2013

	Dependent variable: Demand		
	Linear	Quadratic	Logarithmic
$Income$	1.59	10.47	
$Income^2$		8.72	
$Log(Income)$			19.33
$Log(Income) * Income$			266.27
$Log(Income)^2 * Income$			168.84
$Wealth$	1.13	1.13	1.13
$Married$	1.61	1.65	1.66
$University$	1.04	1.04	1.04
$East$	1.03	1.04	1.04
$Size$	1.74	1.79	1.79

Table 15: White test results - 2003 and 2013

Test results for the Quadratic Model.
Tested Model: 2nd, 3rd and 4th powers of fitted response.

	2003	2013
BP	665.66	482.58
df	31	31
p-value	< 0.001	< 0.001

E. Quantile Regression Results

Figure 1: Quantile Regression Results for the Linear Model - 2003

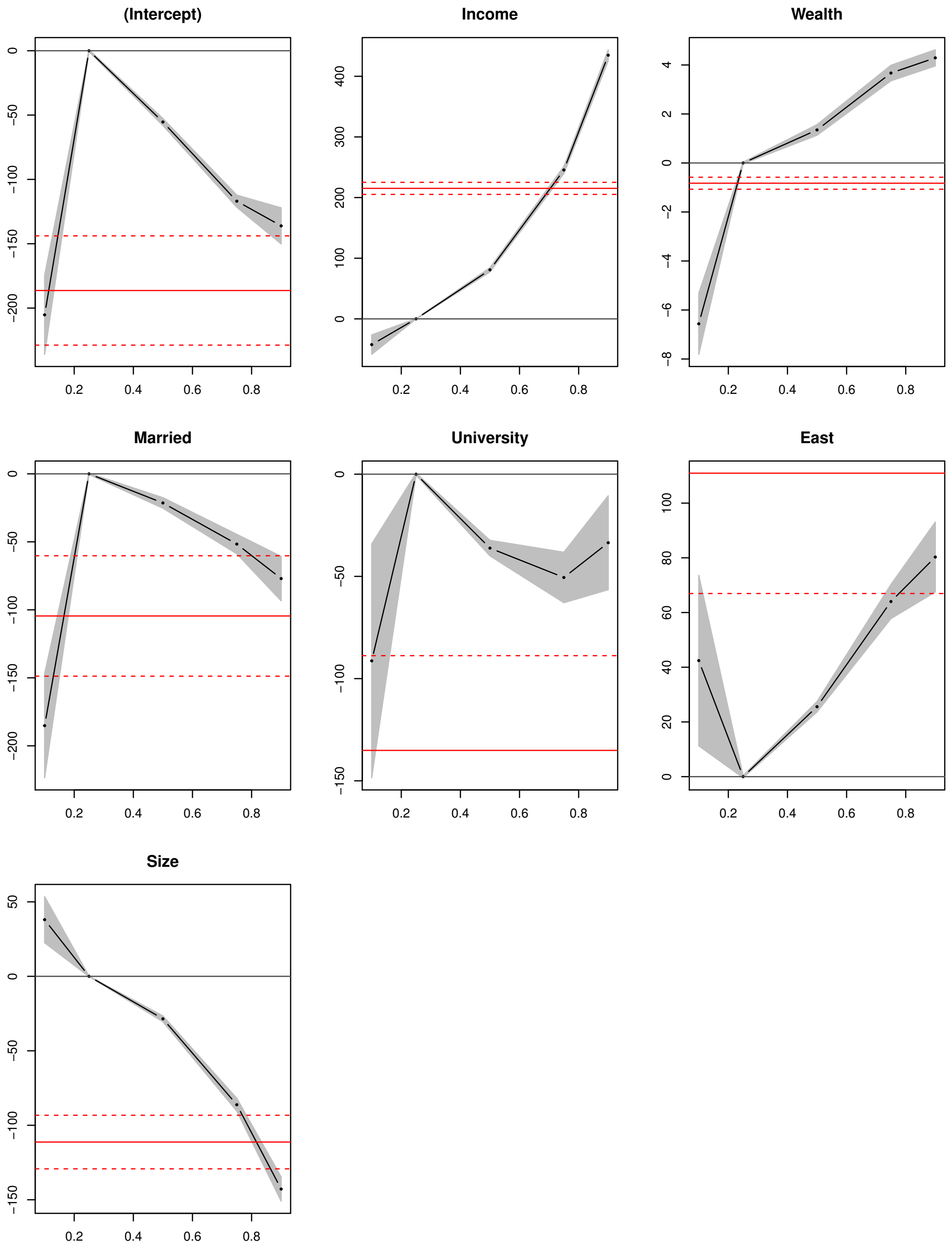

Table 16: Linear Model Quantile Regression Results - 2003

	Dependent variable: Demand				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Income	-42.4569^{***}	0.0000	80.8757^{***}	245.4869^{***}	434.6962^{***}
	(9.7269)	(0.1615)	(2.1483)	(3.6052)	(5.4859)
Wealth	-6.5635^{***}	0.0000	1.3481^{***}	3.6705^{***}	4.2903^{***}
	(0.7574)	(0.0203)	(0.1321)	(0.1929)	(0.2004)
Married	-185.1710^{***}	0.0000	-21.3936^{***}	-51.6556^{***}	-77.0177^{***}
	(23.2778)	(0.2923)	(2.3650)	(4.3590)	(9.9470)
University	-91.3147^{***}	0.0000	-36.1683^{***}	-50.5348^{***}	-33.5392^{**}
	(34.7575)	(0.2916)	(2.3164)	(7.5038)	(13.9643)
East	42.4299^{**}	0.0000	25.5654^{***}	64.0182^{***}	80.2839^{***}
	(18.9543)	(0.2432)	(1.1629)	(3.7870)	(7.7641)
Size	38.0201^{***}	0.0000	-28.5097^{***}	-86.1931^{***}	-142.7995^{***}
	(9.4217)	(0.0908)	(1.1843)	(2.6220)	(4.9411)
Constant	-205.2774^{***}	0.0000	-55.3671^{***}	-116.9293^{***}	-136.1022^{***}
	(18.6446)	(0.4571)	(1.6552)	(2.8666)	(8.5251)
Observations	42,680	42,680	42,680	42,680	42,680
R^1	0.1285	0.0965	0.0862	0.1998	0.3386

Notes: Heteroscedasticity-robust standard errors in parentheses.
Significance levels: $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

Figure 2: Quantile Regression Results for the Quadratic Model - 2003

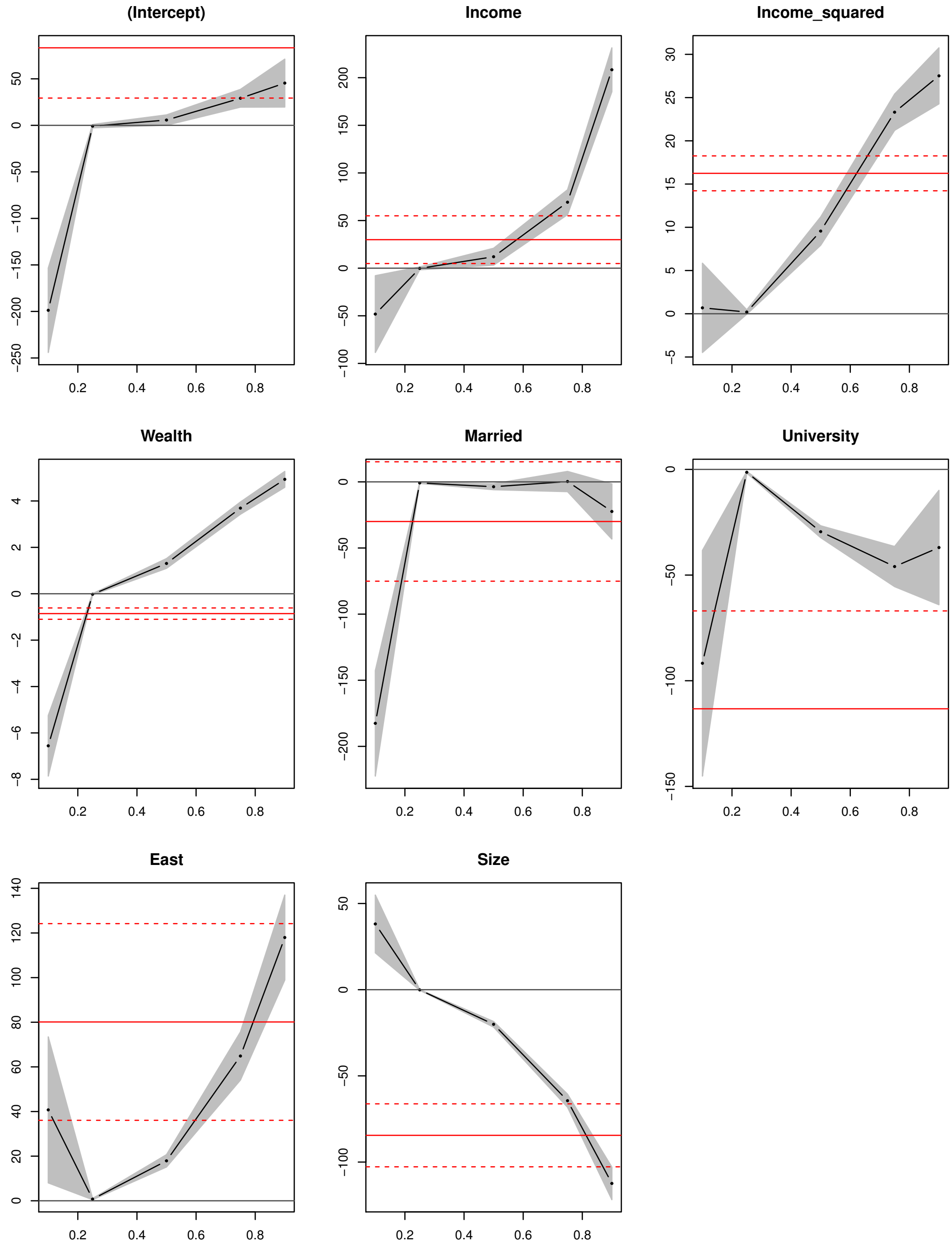

Table 17: Quadratic Model Quantile Regression Results - 2003

	Dependent variable: Demand				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Income	-48.2093^{**}	-0.2071	12.0092^{**}	69.2617^{***}	208.3472^{***}
	(24.5148)	(0.9072)	(3.3582)	(7.935)	(14.0010)
Income^2	0.6840	0.2011	9.5597^{***}	23.3090^{***}	27.5144^{***}
	(3.1393)	(0.1537)	(0.9954)	(1.2742)	(1.9703)
Wealth	-6.5558^{***}	-0.0230	1.3058^{***}	3.6909^{***}	4.9316^{***}
	(0.7939)	(0.0167)	(0.1285)	(0.1604)	(0.2078)
Married	-182.5672^{***}	-0.8006^{***}	-3.6829^{***}	0.2618	-22.3769^{*}
	(24.2610)	(0.2288)	(1.3648)	(4.6462)	(12.6866)
University	-91.7330^{***}	-1.3476^{***}	-29.4815^{***}	-45.9666^{***}	-36.9636^{**}
	(32.4610)	(0.2447)	(1.7120)	(5.7737)	(16.4581)
East	40.7349^{**}	0.7533	17.9105^{***}	64.8706^{***}	117.9676^{***}
	(19.8890)	(0.1638)	(1.6806)	(6.5461)	(11.5727)
Size	38.1284^{***}	-0.1321	-20.0824^{***}	-64.4042^{***}	-112.3982^{***}
	(10.2003)	(0.1131)	(0.9871)	(2.3084)	(5.7396)
Constant	-198.7535^{***}	-0.8599	5.6981^{*}	29.1047^{***}	45.4114^{***}
	(27.5768)	(1.1090)	(3.3582)	(5.7975)	(15.6530)
Observations	42,680	42,680	42,680	42,680	42,680
R^1	0.1285	0.0965	0.0889	0.2117	0.3512

Notes: Heteroscedasticity-robust standard errors in parentheses.
Significance levels: $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

Figure 3: Quantile Regression Results for the Linear Model - 2013

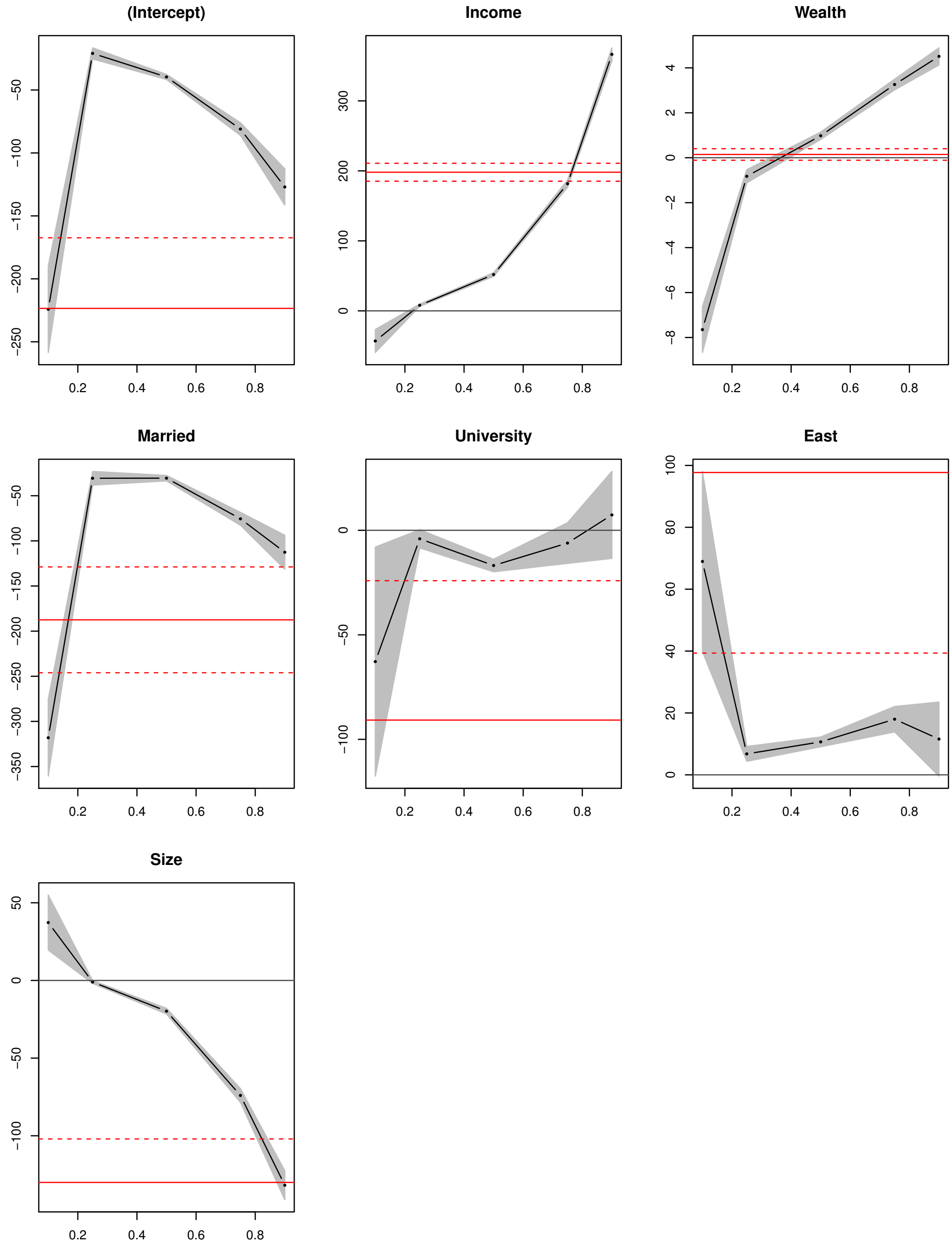

Table 18: Linear Model Quantile Regression Results - 2013

	Dependent variable: Demand				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Income	-43.0233^{***}	7.9841^{***}	51.8807^{***}	181.6065^{***}	366.4144^{***}
	(10.0111)	(1.0026)	(1.5912)	(3.1942)	(5.7119)
Wealth	-7.6496^{***}	-0.8256^{***}	0.9782^{***}	3.2587^{***}	4.5108^{***}
	(0.6193)	(0.1785)	(0.1002)	(0.1455)	(0.2319)
Married	-318.1069^{***}	-30.6364^{***}	-30.5445^{***}	-75.5545^{***}	-112.5795^{***}
	(25.8347)	(4.6072)	(1.8386)	(4.3393)	(11.4506)
University	-62.7927^{*}	-4.0566	-16.8340	-6.1004	7.3653
	(33.3042)	(2.6464)	(1.8299)	(5.9075)	(12.6494)
East	68.9589^{**}	6.7701^{***}	10.6558^{***}	17.9790^{***}	11.5645
	(17.6920)	(1.4595)	(0.9596)	(2.5011)	(7.2772)
Size	37.2803^{***}	-1.0180^{*}	-19.7872^{***}	-74.0561^{***}	-131.8651^{***}
	(10.7739)	(0.5659)	(1.1041)	(2.6904)	(5.6584)
Constant	$-224.1640{*}{*}{*}$	-20.9358^{***}	-39.6435^{***}	-81.0358^{***}	-127.0291^{***}
	(20.8743)	(2.6983)	(1.2180)	(3.0870)	(8.7579)
Observations	42,756	42,756	42,756	42,756	42,756
R^1	0.0496	0.0017	0.0238	0.1289	0.2264

Notes: Heteroscedasticity-robust standard errors in parentheses.
Significance levels: *p<0.1, **p<0.05, ***p<0.01.

Figure 4: Quantile Regression Results for the Quadratic Model - 2013

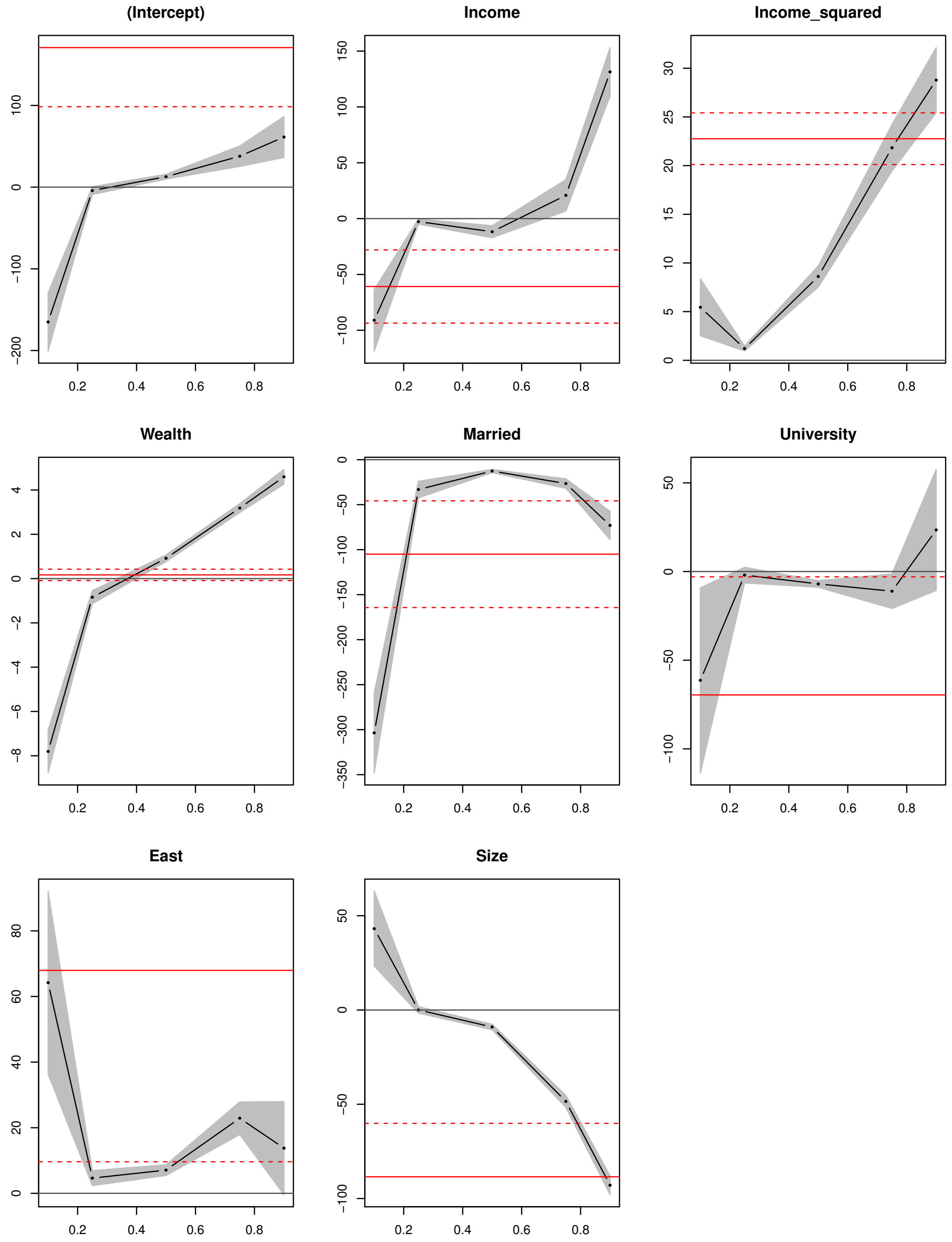

Table 19: Quadratic Model Quantile Regression Results - 2013

	Dependent variable: Demand				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Income	−91.0340***	−2.6850**	−11.7425***	20.8239**	131.3735***
	(16.7320)	(1.3915)	(3.2400)	(8.5219)	(13.2180)
Income^2	5.4428***	1.2093***	8.6108***	21.8323***	28.7834***
	(1.7772)	(0.1543)	(0.6629)	(1.4511)	(2.0290)
Wealth	−7.8054***	−0.8415***	0.9175***	3.1847***	4.5960***
	(0.5865)	(0.1795)	(0.0921)	(0.1228)	(0.2024)
Married	−303.5627***	−33.1467***	−12.6063***	−26.5318***	−73.0464***
	(26.9875)	(5.5854)	(1.0825)	(3.1647)	(9.5627)
University	−61.3532*	−2.0247	−6.9886***	−11.0967*	23.4544**
	(31.7025)	(2.6834)	(1.1394)	(5.9233)	(20.7406)
East	64.2202**	4.6391***	7.0804***	22.9005***	13.7726***
	(16.9385)	(1.3897)	(0.9885)	(3.0181)	(8.6128)
Size	43.1525***	0.0400	−9.0159***	−48.4709***	−92.9330***
	(12.0701)	(1.032)	(0.8813)	(1.9321)	(3.0384)
Constant	−165.0431***	−4.2471	12.7860***	37.8503***	61.3080***
	(21.6218)	(2.8974)	(1.7657)	(7.6086)	(15.2223)
Observations	42,756	42,756	42,756	42,756	42,756
R^1	0.0499	0.0018	0.0274	0.1409	0.2407

Notes: Heteroscedasticity-robust standard errors in parentheses.
Significance levels: *p<0.1, **p<0.05, ***p<0.01.

F. Quantile Regression Test Results

Table 20: Analysis of Deviance - Wald test results - 2003 and 2013

Joint test of Equality of Slopes for τ in (0.1, 0.25, 0.5, 0.75, 0.9).				
	2003		2013	
	Linear	Quadratic	Linear	Quadratic
df	24	28	24	28
Residual df	$213,376$	$213,372$	$213,756$	$213,752$
F-value	693.63	582.84	383.62	398.22
p-value	< 0.001	< 0.001	< 0.001	< 0.001

Table 21: Khmaladze test results - 2003

Tests of the location shift and location-scale shift hypothesis for τ in (0.10, 0.11, 0.12, ..., 0.90).				
	Location Shift		Location-Scale Shift	
	Linear	Quadratic	Linear	Quadratic
Joint Test Statistics	1772.8800	2767.0070	435.4033	1436.3860
Individual Test Statistics				
Income	353.7306	23.0290	29.2839	16.9431
$Income^2$		67.0452		93.6514
Wealth	131.9381	264.1151	111.5655	25.4085
Married	104.9619	149.9893	60.0939	11.0608
University	64.3712	54.7772	26.3784	45.8335
East	117.7892	289.1679	22.6292	148.4530
Size	282.8741	204.1872	53.5133	60.3340
Notes:	As noted in Koenker and Xiao (2002, pp. 1606 f.), critical values for the univariate sub-hypotheses can be reported to test individual “contribution” to the joint significance. Following Koenker (2005, pp. 317 f.), significance levels for these individual sub-hypotheses are 2.102 at the 5% level and 2.640 at the 1% level. For the joint test, the critical value is 8.559 at the 1% level in case of the Linear Model and 9.573 for the Quadratic Model.			

Table 22: Khmaladze test results - 2013

Tests of the location shift and location-scale shift hypothesis for τ in (0.10, 0.11, 0.12, ..., 0.90).				
	Location Shift		Location-Scale Shift	
	Linear	Quadratic	Linear	Quadratic
Joint Test Statistics	1105.3970	1650.3940	1818.3530	1503.7120
Individual Test Statistics				
Income	443.4880	14.4218	308.3486	6.1141
$Income^2$		198.7173		141.0950
Wealth	374.1641	69.0199	113.8959	8.5687
Married	106.2415	80.1315	27.4019	14.3220
University	34.6644	25.4381	9.1119	10.9742
East	56.8169	155.5205	8.3193	80.8863
Size	142.8763	108.5680	107.3163	38.4250

Notes: As noted in Koenker and Xiao (2002, pp. 1606 f.), critical values for the univariate sub-hypotheses can be reported to test individual "contribution" to the joint significance. Following Koenker (2005, pp. 317 f.), significance levels for these individual sub-hypotheses are 2.102 at the 5% level and 2.640 at the 1% level. For the joint test, the critical value is 8.559 at the 1% level in case of the Linear Model and 9.573 for the Quadratic Model.

Eidesstattliche Erklärung

….