RC
EA

THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 22-07

PANAYIOTA LYSSIOTOU
University of Cyprus

PANOS PASHARDES
University of Cyprus

THANASIS STENGOS
University of Guelph
and
Rimini Center for Economic Analysis

"DEMOGRAPHIC VS EXPENDITURE FLEXIBILITY IN ENGEL CURVES"

Demographic vs Expenditure Flexibility in Engel Curves

Panayiota Lyssiotou

University of Cyprus, POBox 20537, 1678 Nicosia, Cyprus (p.lyssiotou@ucy.ac.cy)

Panos Pashardes

University of Cyprus, POBox 20537, 1678 Nicosia, Cyprus (p.pashardes@ucy.ac.cy)

Thanasis Stengos[*]

University of Guelph, Guelph, Ontario, Canada N1G 2W1
and Rimini Center for Economic Analysis (tstengos@.uoguelph.ca)

Abstract

We consider the effects of demographic and expenditure variables on consumer demand in a system of Engel curves using a smooth coefficient semiparametric model where the expenditure effects on the budget shares vary nonparametrically with demographic variables such as the age of head and number of children in the household. Our findings, based on UK micro data, suggest that with a smooth coefficient semiparametric model there is no need for nonlinear logarithmic expenditure effects in the budget shares. Furthermore, we find evidence of a trade-off between demographic and expenditure effects in Engel curves and that a rank-2 system of Engel curves where the logarithmic expenditure effects are allowed to vary with demographic characteristics either nonparametrically or as a third degree polynomial function cannot be rejected against a rank-3 (quadratic logarithmic) model. The implications on household behaviour and welfare are also examined.

---

1 Introduction

There has been considerable interest in econometrics in developing models that incorporate both parametric and nonparametric components. The regression relationship is usually expressed in terms of variables that enter parametrically because they are known to interact with the dependent variable in a specific way, and those that enter nonparametrically since there is no a-priori reason for them to follow a particular pattern.

In the context of Engel curves most empirical studies allow demographic and other household characteristics to enter parametrically and expenditure to enter nonparametrically. This is because knowledge of the expenditure effect on demand is of primary concern in analysing consumer behaviour and welfare and the use of nonparametric methods is part of the effort to find appropriate ways of modelling this effect in empirical applications. In particular, investigators have been interested in the rank of demand system, the maximum dimension of the expenditure space spanned by Engel curves, given that demand systems with a rank above three cannot be integrable (Gorman 1981, Lewbel 1991).

This paper is motivated by the empirical finding that while the rank of Engel curves tends to be above three in studies that do not control for demographic heterogeneity in preferences (Grodal and Hildenbrand 1992, Hildenbrand 1994 and Kneip 1994), it becomes three when either (a) a relatively small, demographically homogeneous subset of data is used (Lewbel 1991, Banks, Blundell and Lewbel 1997, Donald 1997 and Hausman, Newey and Powell 1995); or (b) the effects of demographic heterogeneity are removed using a semiparametric estimation technique (Lyssiotou, Pashardes and Stengos 1999). These findings suggest that the effects of demographic characteristics on consumer demand may not be as uncomplicated as commonly assumed and a search for an appropriate Engel curve specification may have to place as much emphasis on avoiding a-priori restrictions on demographic effects as it does on the expenditure effects.

To examine this issue, we employ a smooth coefficient model combining parametric and nonparametric components: the large number of household characteristics typically captured by binary variables (gender, housing location and tenure, occupation etc.) enter linearly while important demographic variables such as the age of the household head and/or the number of children, are allowed to directly affect the coefficients of the model. Thus, unlike the linear semiparametric model, the smooth coefficient approach yields expenditure effects that vary with consumer characteristics. In fact, it is a special case of a varying coefficient model offering a particular way of combining parametric and nonparametric components, see Fan and Zhang (1999), Cai, Fan and Li (2000) and Cai, Fan and Yao (2000) and Zhang et al (2002). Li et al. (2002) used such a model to estimate the production function of the non-metal mineral industry in China.

The framework of analysis in the paper is a system of Engel curves that vary with a vector of exogenous characteristics z_h and the logarithm of the budget level y_h. For a given good, the observed Engel curve (budget share) w for household h can be written as,

$$w(z_h, y_h) = a(z_h) + d(z_h)f(y_h) + v_h \tag{1}$$

where $a(z_h)$, $d(z_h)$ and $f(y_h)$ are unknown functions. We shall use (1) to discuss and investigate empirically:

(i) appropriate empirical specifications for the $a(z_h)$, $d(z_h)$ and $f(y_h)$ functions,

(ii) the extent to which z_h can be reparameterised into y_h effects and vice versa, and

(iii) the behavioural and welfare implications of such reparameterisations.

Various approaches to searching for an empirically adequate specification of Engel curves can be seen as special cases of (1). For example, the restriction $d(z_h) = 1$, yields the Additively Separable Regression model used by Lyssiotou, Pashardes and Stengos (2002). If we also assume that $a(z_h) = z_h'\alpha$, where z_h is of dimension $p \times 1$ and a an unknown parameter of dimension $p \times 1$, we obtain the semiparametric partially linear

model (Engle et al. 1986, Robinson 1988 and Stock 1989),

$$w_h = z_h{}'\alpha + f(y_h) + u_h. \tag{2}$$

This model has been used in the literature to recover the accurate specification of the Engel curve relation (Blundell, Duncan and Pendakur 1998).

The smooth coefficient semiparametric model considered in this paper is also a special case of (1). Unlike (2), however, is parametric in $f(y_h)$ and nonparametric in $a(z_h)$ and $d(z_h)$,

$$w_h = a(z_h) + y_h d(z_h) + \epsilon_h \tag{3}$$

We use (3) to investigate whether the presence of unrestricted demographic effects in the empirical specification bears on the effects of y_h. We also use this model as a benchmark to test the empirical performance of the Engel curves of the Quadratic Logarithmic model (Lewbel 1990, Banks, Blundell and Lewbel 1997), a rank-3 demand system generally found through (2) to be an empirically adequate parametric specification.

Comparison of results obtained from the linear and smooth coefficient semiparametric models can yield conclusions about the extent to which demographic flexibility can help reduce expenditure flexibility but cannot tell us how (2) and (3) perform relative to (1), the specification in which they are nested. To find how either higher demographic or higher expenditure flexibility fare against a model that has both we assume that $a(z_h)$, $d(z_h)$ and $f(y_h)$ in (1) have polynomial functional forms and examine how a decrease in demographic flexibility (lowering the degree of the $a(z_h)$ and $d(z_h)$, polynomials) compares with a decrease in expenditure flexibility (lowering the degree of $f(y_h)$) in terms of fitting the data. This parametric tests can also help investigate whether higher order terms in $a(z_h)$ and $d(z_h)$ can be reparameterised into higher order terms in $f(y_h)$, and vice versa; and if so which parameterisation is statistically more adequate on empirical grounds.

The next section presents the basic structure of a smooth coefficient semiparametric model as applied to Engel curves and reports empirical results obtained from the ap-

plication of this model to UK individual household data drawn from the 1980 Family Expenditure Survey. Section 3 describes the parametric model used in our analysis and also reports results obtained from the same data set. Section 4 concludes the paper.

2   Demographic vs expenditure nonparametric effects

As in the previous section, w_h denotes the dependent variable, z_h be $p \times 1$ and y_h a one-dimensional variable. We write (3) as

$$
\begin{aligned}
w_h &= (1, y_h) \begin{bmatrix} \alpha(z_h) \\ d(z_h) \end{bmatrix} + \varepsilon_h \\
&= Y_h^T \xi(z_h) + \varepsilon_h
\end{aligned} \tag{4}
$$

where $\xi(z_h) = (\alpha(z_h), d(z_h))^T$ is a smooth but unknown function of z_h. One can then estimate $\xi(z_h)$ at each point z using a local least squares approach, where

$$
\begin{aligned}
\widehat{\xi}(z) &= [(n\theta^p)^{-1} \sum_{h=1}^{n} Y_h Y_h^T K(\frac{z_h - z}{\theta})]^{-1} \{(n\theta^p)^{-1} \sum_{h=1}^{n} Y_h w_h K(\frac{z_h - z}{\theta})\} \tag{5} \\
&= [D_n(z)]^{-1} A_n(z)
\end{aligned}
$$

and $D_n(z) = (n\theta^p)^{-1} \sum_{h=1}^{n} Y_h Y_h^T K(\frac{z_h - z}{\theta})$, $A_n(z) = (n\theta^p)^{-1} \sum_{h=1}^{n} Y_h w_h K(\frac{z_h - z}{\theta})$, is a kernel function and $\theta = \theta_n$ is the smoothing parameter for sample size n.

The intuition behind the above local least-squares estimator is straightforward. Let us assume that z is a scalar and $K(.)$ is a uniform kernel. In this case the expression for $\widehat{\xi}(z)$ becomes

$$
\widehat{\xi}(z) = [\sum_{|z_h - z| \leq b} Y_h Y_h^T]^{-1} \sum_{|z_h - z| \leq b} Y_h w_h. \tag{6}
$$

In this case $\widehat{\xi}(z)$ is simply a least squares estimator obtained by regressing w_h on Y_h using the observations of (Y_h, w_h) that their corresponding z_h is close to z ($|z_h - z| \leq \theta$). Since $\xi(z)$ is a smooth function of z, $|\xi(z_h) - \xi(z)|$ is small when $|z_h - z|$ is small. The condition that nb^p is large ensures that we have sufficient observations within the interval $|z_j - z| \leq b$ when $\xi(z_h)$ is close to $\xi(z)$. Therefore, under the conditions that

$b \to 0$ and $nb^p \to \infty$, one can show that the local least squares regression of w_h on Y_h provides a consistent estimate of $\xi(z)$. In general it can be shown that

$$\sqrt{n\theta^p}(\widehat{\xi}(z) - \xi(z)) \to N(0, \Omega)$$

where Ω can be consistently estimated. For more details see Fan and Zhang (1999) and Li et al (2002).

We apply this method to estimate a system of Engel curves consisting of six categories of non-durable goods: food, alcohol, fuel, clothing, other goods and services. The data are drawn from the 1980 UK Family Expenditure Survey and include all two-adult households whose head is under retirement age and not self-employed. The size of the sample is 1305 observations. For each household, in addition to expenditure on each of the six categories of goods above, we have information about a large number of demographic and other characteristics, such as the family size and age composition, the sex, employment and economic position of members, the geographical location and housing tenure, the ownership of cars, domestic appliances and other durables etc[1]. Specifically we use a model that is a simple generalization of equation (4), where we allow for some of the demographic characteristics (in particular the ones that are described by dummy variables) to enter the regression function separately. We denote these variables as z_{1h}, whereas the demographic characteristic of interest, denoted by z_{2h}, is allowed to enter the regression function nonparametrically as an argument of the coefficient $\xi(.)$ in equation (4).

The regression function that we estimate is written as

$$
\begin{aligned}
w_h &= z_{1h}^T \gamma + (1, y_h, y_h^2) \begin{bmatrix} \alpha(z_{2h}) \\ \beta(z_{2h}) \\ \lambda(z_{2h}) \end{bmatrix} + \varepsilon_h \\
&= z_{1h}^T \gamma + Y_h^T \xi(z_{2h}) + \varepsilon_h
\end{aligned}
\tag{7}
$$

[1] The Family Expenditure Survey data has been extensively used in the literature and to conserve space we do not report summary statistics. These are available from the authors. To check the robustness of our results we also tried a sample of similar size from 1985. The results are overall qualitatively similar and they are also not reported for space conservation purposes.

where $\xi(z_{2h}) = (\alpha(z_{2h}), d(z_{2h}))^T$ where $d(z_{2h}) = (\beta(z_{2h}), \lambda(z_{2h}))$ and $Y_h^T = (1, y_h, y_h^2)^T$. The z_{2h} variables that we use include the age of household head and the number of children; hence $p = 1$. Equation (7) is a more general version of equation (4), since it includes the linear part consisting of the z_1 variables. We use the methodology of Zhang et al (2002) to estimate the linear part and the smooth coefficient function part in two stages. We find that either of these two characteristics gives us very similar results and they are sufficient to capture by themselves any nonlinearity in expenditure effects due to heterogeneity that arise in the specification of the Engel curves that we examine[2].

The diagrams in Figures 1-4 show how $\alpha(z_{2h})$, $d(z_{2h})$ $\lambda(z_{2h})$ in (7) vary with the age of the household head for food, alcohol, fuel and clothing. The function $\alpha(z_{2h})$, is represented by the continuous line, $d(z_{2h})$ by the long-segments discontinuous line and $\lambda(z_{2h})$ by the short-segments discontinuous line. The striking feature of these diagrams is that $\lambda(z_{2h})$ are very close to zero throughout the age range in all the diagrams, whereas $d(z_{2h})$ and $\lambda(z_{2h})$ exhibit substantial variation over age.

The results in Figures 1-4 suggest that when the age effects on consumer demand are allowed to enter nonparametrically, a rank-2 system of Engel curves may not be statistically inadequate. To investigate this finding further, we compare the predicted budget shares at different log-expenditure levels obtained from the smooth coefficient model (7), with the results obtained from the linear semiparametric model, as defined by (2).[3] The results of this comparison are shown in the diagrams of Figures 5-8. The continuous line represents the predictions of the rank-2 smooth coefficient model and the discontinuous line the predictions corresponding to the linear semiparametric model. As seen from the diagrams, the predictions from the two models are close to each other except for the extremes of log-expenditure distribution: for food, fuel and clothing

---

[2] We have also done the analysis by instrumenting y_h using all the variables in vectors z_{1h} and z_{2h} and the logarithm of household income and its powers (up to the 4rth one). The results were very similar to the uninstrumented case that we report below and are omitted to conserve space.

[3] For the instrumented version of this model we apply the estimation method of Lyssiotou et al. (1999) and Blundell et al. (1998) based on an augmented version of the partial linear model with the unknown function of the residuals from the instrument regression.

the two models differ at the lower end of the log expenditure distribution, whereas for alcohol they differ at the upper end of the log expenditure distribution.

One conclusion emerging from the diagrams of Figures 5-8 is that for a wide range of expenditure, it makes no difference in terms of prediction whether one uses a linear in logarithmic system of Engel curves where the log expenditure effects vary nonparametrically with age or a system of Engel curves that are parametrically linear in demographic characteristics and nonparametric in log expenditure. This trade off between demographic and expenditure flexibility in the system of Engel curves under consideration is investigated further using a parametric approach in the next section.

However, the differences in predictions obtained from two models at the extremes of log expenditure distribution raises the question whether these differences are statistically significant and if so which of the two models is more appropriate. To answer this question we use the one-sited t-statistic described in the Appendix to test the rank-3 quadratic logarithmic system of Engel curves given by equation (13), found elsewhere to be a statistically adequate empirical model, against the smooth coefficient model (7) considered in this paper. The value of this statistic is calculated at 3.7651, whereas its critical value at 5% significance level is 1.1051. This suggests that the rank-3 quadratic logarithmic model is rejected against the rank-2 smooth coefficient model.

3 Demographic vs expenditure effects in parametric models

The analysis in the previous section shows that when the effects of demographic variables enter Engel curves nonparametrically, there is no need for quadratic logarithmic expenditure effects in these curves. This suggests that it may be possible to also do away with the same effects in a parametric specification of Engel curves if the demographic effects enter in a functionally flexible manner.

To investigate this point we allow the $a(z_h)$, $d(z_h)$ and $f(y_h)$ functions in (1) to have

the following polynomial:

$$a(z_h) = \sum_{m=0}^{M} a_m z_{2h}^m, \;\; d(z_h) = \sum_{k=0}^{K} d_k z_{2h}^k, \;\text{and}\; f(y_h) = \sum_{r=1}^{R} f_{r-1} y_h^{r-1} \tag{8}$$

where z_{2h} are as defined in the previous section.

Then, for $M = 1$, $K = 1$ and $R = 2$ we have the rank-2 budget share equation system corresponding to the Almost Ideal (AI) demand system of Deaton and Muellbauer (1980)

$$w(z_h, y_h) = \delta_o + \delta_1 z_{2h} + (\beta_o + \beta_1 z_{2h}) y_h \tag{9}$$

where $\delta_o = a_o + d_o f_o$, $\delta_1 = a_1 + f_o d_1$, $\beta_o = f_1 d_o$, and $\beta_1 = f_1 d_1$. If $K = 0$, then $\beta_1 = 0$ and changes in the budget level have a uniform effect on demand across households with different characteristics z_{2h}.

Also, for $M = 1$, $K = 1$ and $R = 3$ we have the rank-3 budget share equation system corresponding to the Quadratic Almost Ideal Demand System (QUAIDS) of Banks, Blundell and Lewbel (1997)

$$w(z_h, y_h) = \delta_o + \delta_1 z_{2h} + (\beta_o + \beta_1 z_{2h}) y_h + (\lambda_o + \lambda_1 z_{2h}) y_h^2 \tag{10}$$

where $\lambda_o = f_2 d_o$, and $\lambda_1 = f_2 d_1$.

In the empirical analysis below we investigate these and more complicated systems resulting for choosing higher values for M, K and R with a view to finding whether there is competition between greater demographic flexibility (larger values for M and K) and log expenditure flexibility (larger values for R). Indeed, this is a likely phenomenon because, as is well known, z_{2h} and y_h tend to be correlated in the data, e.g. the household budget increases with the number of children in the family.

To see this let $E(y_h z_h) = \gamma_{zy} E(y_h^2)$ where γ_{zy} is the regression coefficient part of z_{2h} on y_h. Then (9) can be written as

$$E(w) = \delta_o + \delta_1 E(z_{2h}) + \beta_o E(y_h) + \beta_1^* E(y_h^2) \tag{11}$$

where $\beta_1^* = \gamma_{zy}\beta_1$ making the demand system appear as a rank-3. Given the significance of β_1, one may not reject the empirical hypothesis that (11) is a rank-3 model if γ_{zy} is significant.

The opposite, however, is also true, e.g. the rank of the demand system can be understated if log expenditure is parameterised as demographic flexibility, e.g. replacing $E(y_h z_h) = \gamma_{yz}E(z_h^2)$ where γ_{yz} is the regression coefficient part of y_h on z_h. Then (10) can be written as a rank-2 demand system

$$E(w) = \delta_o + \delta_1 E(z_{2h}) + \beta_o E(y_h) + \beta_1' E(z_{2h}^2) + \lambda_o E(y_h^2) + \lambda_1' E(z_{2h}^2) \qquad (12)$$

where $\beta_1' = \beta_1 \gamma_{yz}$ and $\lambda_1' = \lambda_1 \gamma_{y^2 z}$. In this case, when $\lambda_o = 0$ the demand system can be mistaken for a rank-2 in empirical application.

To investigate the arguments above on empirical grounds, we define z_{2h} to be the age of the household head and estimate the system of Engel curves (1) where the functions $a(z_h)$, $d(z_h)$ and $f(y_h)$ take the form given by (8) assuming $M = 3$, $K = 3$, and $R = 3$, i.e. a rank-3 specification with up to cubic demographic effects everywhere. Also, we condition on a number of other household characteristics, besides the age of household head, included in vector z_{1h} and found by previous studies to significantly affect demand behavior.[4]. We then use this specification as our maintained hypothesis to test alternative assumptions about the degrees of the polynomials (8). Table 1 reports the p-values corresponding these tests. As seen from the p-values corresponding to the rank-2 demand system, a system of Engel curves with linear logarithmic expenditure effects cannot be rejected against the rank-3 model at the 5% significance level if these effects are allowed to vary with; (i) age and its square $(K = 2)$ and the intercept of the budget share equations also include the same variables $(M = 2)$; or (ii) age, its square and its cubic $(K = 3)$ and the intercept of the budget share equations includes age $(M = 1)$.

---

[4]One could instrument y_h by using the methodology of Hausman et al (1995). However, in this version the results presented are without instruments.

It is clear from the results reported in Table 1 that age effects on the budget shares can be reparameterised in logarithmic expenditure effects, and vice versa. For example, a rank-2 model with up to cubic age effects in the intercept and linear logarithmic expenditure effects varying with age and age square ($R = 2$, $M = 3$, $K = 2$) has the same p-value when tested against the null, as a rank-3 model with linear age effects in the intercept and linear and quadratic logarithmic expenditure effects varying with age and age square ($R = 3$, $M = 1$, $K = 2$).

To demonstrate the observational equivalence between these two models we have plotted the predicted budget shares against log expenditure and age of household head: throughout the range of the log expenditure and age variables the predictions from the 'age-flexible' and 'expenditure-flexible' model are remarkably close to each other, as they are to the full model. Furthermore, the budget elasticities calculated from the parameter estimates of the 'expenditure-flexible' model with those calculated from the parameter estimates of the 'age-flexible' model do not exhibit substantial differences, indicating the similarity in behaviour implied by the two models.[5]

The welfare implications of the observational equivalence between age and expenditure flexibility in Engel curves is illustrated in the diagram of Figure 9, where each line shows differences in the true cost of living index between households with head aged 45 and households with head 25 years old, over a period of seventeen years (1970-86). The differences represented by the discontinuous line are calculated using the parameter estimates obtained from the expenditure-flexible model; whereas the differences represented by the discontinuous line are calculated using the parameter estimates obtained from the age-flexible model.

It is clear from the diagram of Figure 9 that both models suggest that inflation over the period under consideration was not neutral to the age of household head. The expenditure-flexible model suggests that during the period prior to 1980 price increases

---

were lower and in subsequent years higher for goods purchased by households with head aged 45 compared to the price increases faced by households with head aged 25. The age-flexible model, however, suggests that during the period prior to 1980 price increases were the same for both households while in subsequent years goods purchased by households with head aged 45 increase slower in price than goods purchased by households with head aged 25.

4   Conclusion

This paper investigates two prominent features of micro-level Engel curves expressed in budget share form: nonlinearity and heterogeneity of the logarithmic expenditure effects. This investigation is performed using nonparametric and parametric methods.

For the nonparametric analysis we use a smooth coefficient semiparametric model, a member of the varying coefficient models. In this model the nonparametric components relate to the effects of demographic characteristics (age and children), unlike the linear semiparametric model where the nonparametric component corresponds to the effects of logarithmic expenditure. The empirical results based on UK household data for six commodities show that with a smooth coefficient semiparametric model there is no need for nonlinear logarithmic expenditure effects in the budget shares. This, together with the finding that the smooth coefficient and linear semiparametric models yield similar predictions over a wide expenditure range suggests that there can be a trade-off between allowing more flexibility in the demographic and expenditure effects in Engel curves. However, at the extremes of expenditure distribution the two models suggest considerably different consumer behaviour. Notably, the rank-3 quadratic logarithmic system of Engel curves, where the logarithmic expenditure effects are not allowed to vary with demographic characteristics, is rejected against the rank-2 smooth coefficient model. Similar results were obtained by Calvet and Comon (2000) using a different methodology. Specifically, they used a semiparametric model of consumer demand that

imposes no restrictions on the joint distribution of expenditure and preferences and found evidence in favor of a diminished income effect, and a rank-2 demand system, due to high correlation between income and tastes.

Further investigation of the extent to which there is a trade off between demographic and expenditure effects in Engel curves and the implications of these for the empirical specification of Engel curves is conducted using a parametric model where these effects are allowed to enter in a polynomial form. The results reaffirm the findings of the nonparametric investigation in the sense that a system of linear logarithmic cannot be rejected against a quadratic logarithmic system of Engel curves when the expenditure effects are allowed to interact with a cubic age function. The empirical results from the parametric analysis also suggest that age effects on the budget shares can be reparameterised in logarithmic expenditure effects, and vice versa.

We investigate the implications of this reparameterisation using the budget elasticities and true cost of living indices calculated from the parameter estimates of two observationally equivalent (non-rejectable against the null) empirical specifications, one 'expenditure-flexible' and one 'age-flexible'. In the case of budget elasticities, we do not find notable differences between the two specifications, except for clothing at low and high budget levels. Therefore, in terms of consumer behaviour it makes no difference whether an 'expenditure-flexible' or an 'age-flexible' model is used, as long as no particular concern is attached to demand for clothing at the extremes of income distribution. However, the estimated true cost of living indices differ substantially between the two models: both suggest inflation effects varying with the age of household head, but the pattern of this variation is different when an 'expenditure-flexible' rather than an 'age-flexible' model is used. Given the observational equivalence of the two models, these inflation effects should be treated as lower and upper bounds as should the budget elasticities obtained from the two models.

Appendix

We will present below a test statistic that was used by Li et al. (2001). In our implementation we will use a bootstrap version of this test. Let y_i denote the dependent variable, and let x_i be $p \times 1$ and z_i be $q \times 1$ vectors of exogenous variables. Consider the following linear model

$$
\begin{aligned}
w_h &= z_1^T \gamma(z_i) + \alpha_0(z_{2h}) + y_h^T d_0(z_{2h}) + y_h^{2T} \lambda_0(z_{2h}) + \varepsilon_h \\
&= z_1^T \gamma(z_i) + Y_h^T \xi_0(z_{2h}) + \varepsilon_h
\end{aligned} \tag{13}
$$

where $\xi_0(z_{2h}) = (\alpha_0(z_{2h}), d_0(z_{2h}), \lambda_0(z_{2h}))^T$ is a smooth known function of z, denoted by the subscript 0.

We can test the adequacy of (13), the H_0, against the smooth coefficient semiparametric alternative (7) using the following test statistic.

$$
\begin{aligned}
\widehat{I}_n &= \frac{1}{n^2 \theta^p} \sum_h \sum_{h \neq s} Y_h^T (w_h - Y_h^T \widehat{\xi}_0(z_{2h})) Y_s (w_s - Y_s^T \widehat{\xi}_0(z_{2s})) K\left(\frac{z_{2s} - z_s}{\theta}\right) \\
&= \frac{1}{n^2 \theta^p} \sum_h \sum_{h \neq s} Y_h^T Y_s \widehat{\varepsilon}_h \widehat{\varepsilon}_s K\left(\frac{z_{2s} - z_s}{\theta}\right)
\end{aligned}
$$

where $\widehat{\varepsilon}_h$ denotes the residual from parametric estimation (under H_0). It can be shown that under H_0, $J_n = n\theta^{p/2} \widehat{I}_n / \widehat{\sigma}_0 \longrightarrow N(0,1)$, where $\widehat{\sigma}_0$ is a consistent estimator of the variance of $\widehat{I}_n$, see Li et al (2001). It can be shown that the test statistic is a consistent test for testing H_0 (equation (13)) against H_1 (equation (7)). We use a bootstrap version of the above test statistic, since boostrapping improves the size performance of kernel based tests for functional form, see Zheng (1996) and Li and Wang (1998).

References

Banks J., Blundell R, Lewbel A (1997) Quadratic engel curves and consumer demand. Review of Economics and Statistics 4:527-539

Blundell R, Duncan A, Pendakur K (1998) Semiparametric estimation and consumer demand. Journal of Applied Econometrics 13:435-461

Cai Z, Fan J, Li R (2000) Efficient estimation and inferences for varying-coefficient models. Journal of American Statistical Association 95:888-902

Cai Z, Fan J, Yao Q (2000) Functional coefficient regression models for nonlinear time series models. Journal of American Statistical Association 95:941-956

Calvet EL, Comon E (2000) Behavioral heterogeneity and the income effect. Discussion Paper No. 1892, Harvard Institute of Economic Research

Deaton A, Muellbauer J (1980) An almost ideal demand system. American Economic Review 70:312-326.

Donald SG (1997) Inference Concerning the number of factors in a multivariate non-parametric relationship. Econometrica 65:103-132

Engle RF, Grange CWJ, Rice J, Weiss A (1986) Semiparametric estimates of the relationship between weather and electricity sales. Journal of the American Statistical Association 81:310-320

Fan J, Zhang W (1999) Statistical estimation in varying-coefficient models. The Annals of Statistics 27:1491-1518

Gorman T (1981) Some engel curves. In: Deaton A (ed.) Essays in the theory and measurement of consumer behaviour in honor of Richard Stone. Cambridge University Press

Grodal B, Hildenbrand W. (1992) Cross section engel curves, expenditure distributions and the law of demand. In: Phlips L, Taylor LD (eds) Essays in honor of H. S. Houthakker. Kluwer, Dordrecht, pp. 85-120

Hausman J, Newey W, Powell J (1995) Nonlinear errors in variables: estimation of some engel curves. Journal of Econometrics 65:203-233

Hildenbrand W (1994) Nonparametric estimation of common regressors for similar curve data. The Annals of Statistics 3:1386-1427

Kneip A (1994) Market Demand: Theory and Empirical Evidence. Princeton University Press, Princeton NJ

Li Q, Wang S (1998) A Simple consistent bootstrap test for a parametric regression functional form. Journal of Econometrics 87:145-165

Li Q, Huang C, Li D, Fu T (2002) Semiparametric smooth coefficient models. Journal of Business and Economic Statistics 20:412-422

Lewbel A (1991) The rank of demand systems: theory and nonparametric estimation.

Econometrica 59:711-730

Lewbel A. (1990) Full rank demand systems. International Economic Review 31:289-300

Lyssiotou P, Pashardes P, Stengos T (1999) Preference heterogeneity and the rank of demand systems. Journal of Business and Economic Statistics 17:248-252

Lyssiotou P, Pashardes P, Stengos T (1999) Testing the rank of engel curves with endogenous expenditure. Economics Letters 64:61-65

Lyssiotou P, Pashardes P, Stengos T (2002) Age effects on consumer demand: an additive partially linear regression model. Canadian Journal of Economics 35:153-165

Robinson P (1988) Root-n-consistent semiparametric regression. Econometrica 56:931-954

Stock JH (1989) Nonparametric policy analysis. Journal of American Statistical Association 84: 567-575

Zheng JX (1996) A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75:263-289

Zhang W, Lee, S-Y, Song X (2002) Local polynomial fitting in semivarying coefficient model. Journal of Multivariate Analysis 82:166-188

Table 1:

P-values of various specifications tested against a rank-3
model with up to cubic demographic effects

Polynomial degree		P-value	
M	K	Rank-2	Rank-3
3	3	0.2577	
3	2	0.3158	0.8773
3	1	0.0318	0.0968
2	3	0.3521	0.8152
2	2	0.1899	0.5176
2	1	0.0218	0.0642
1	3	0.0617	0.0458
1	2	0.0356	0.3060
1	1	0.0107	0.0221
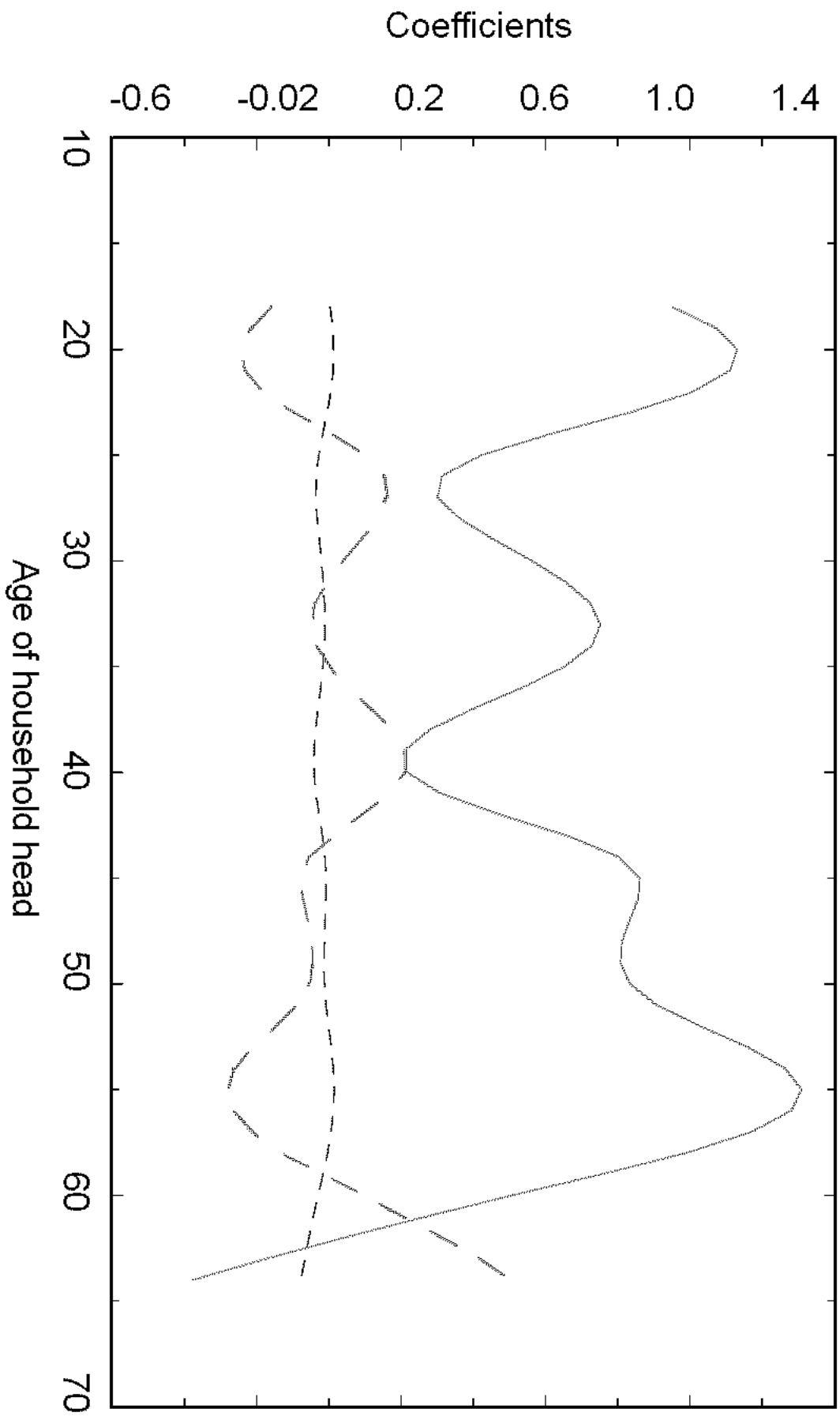
Figure 1: Smooth coefficients vs household age: food

Figure 2: Smooth coefficients vs household age: alcohol
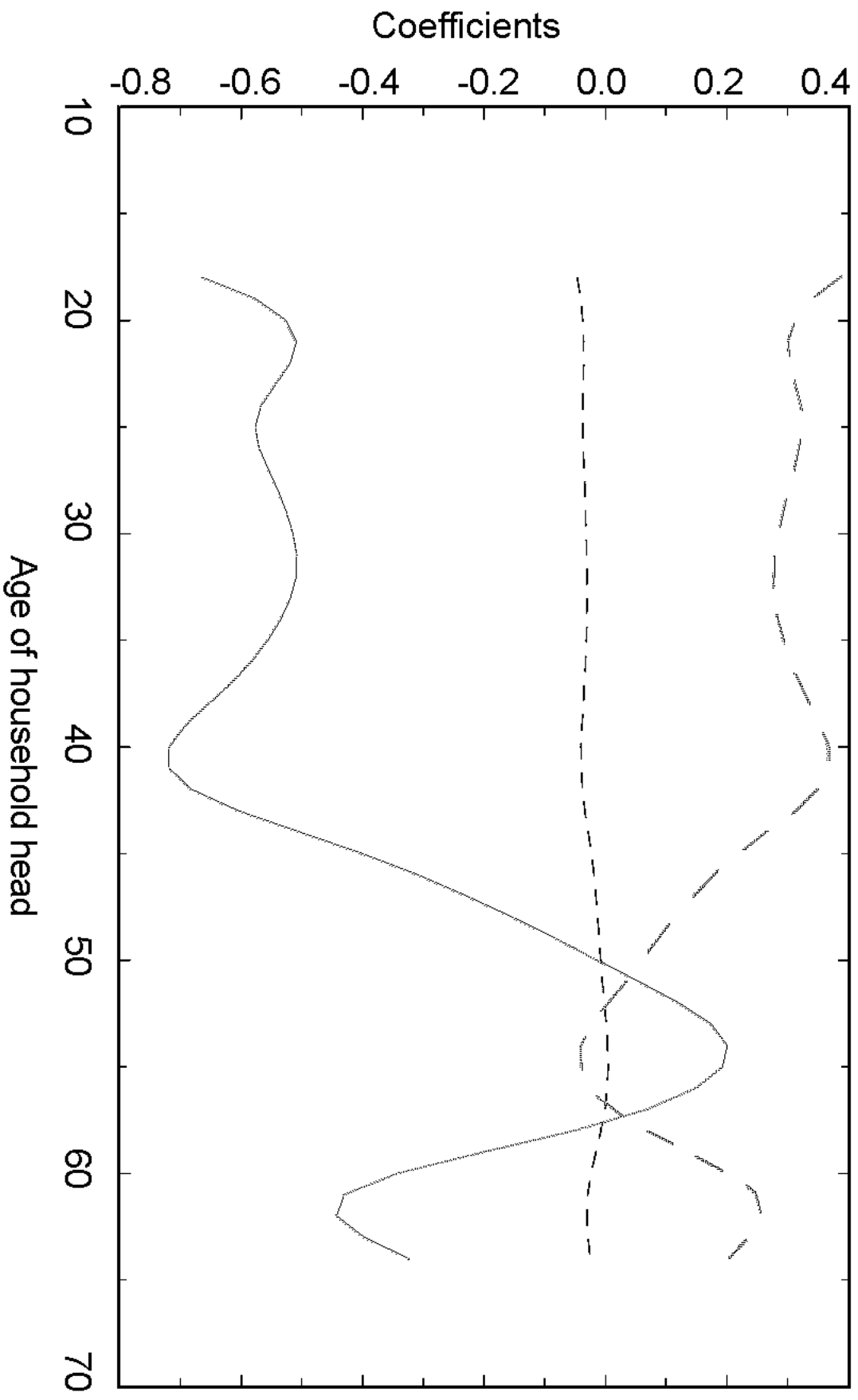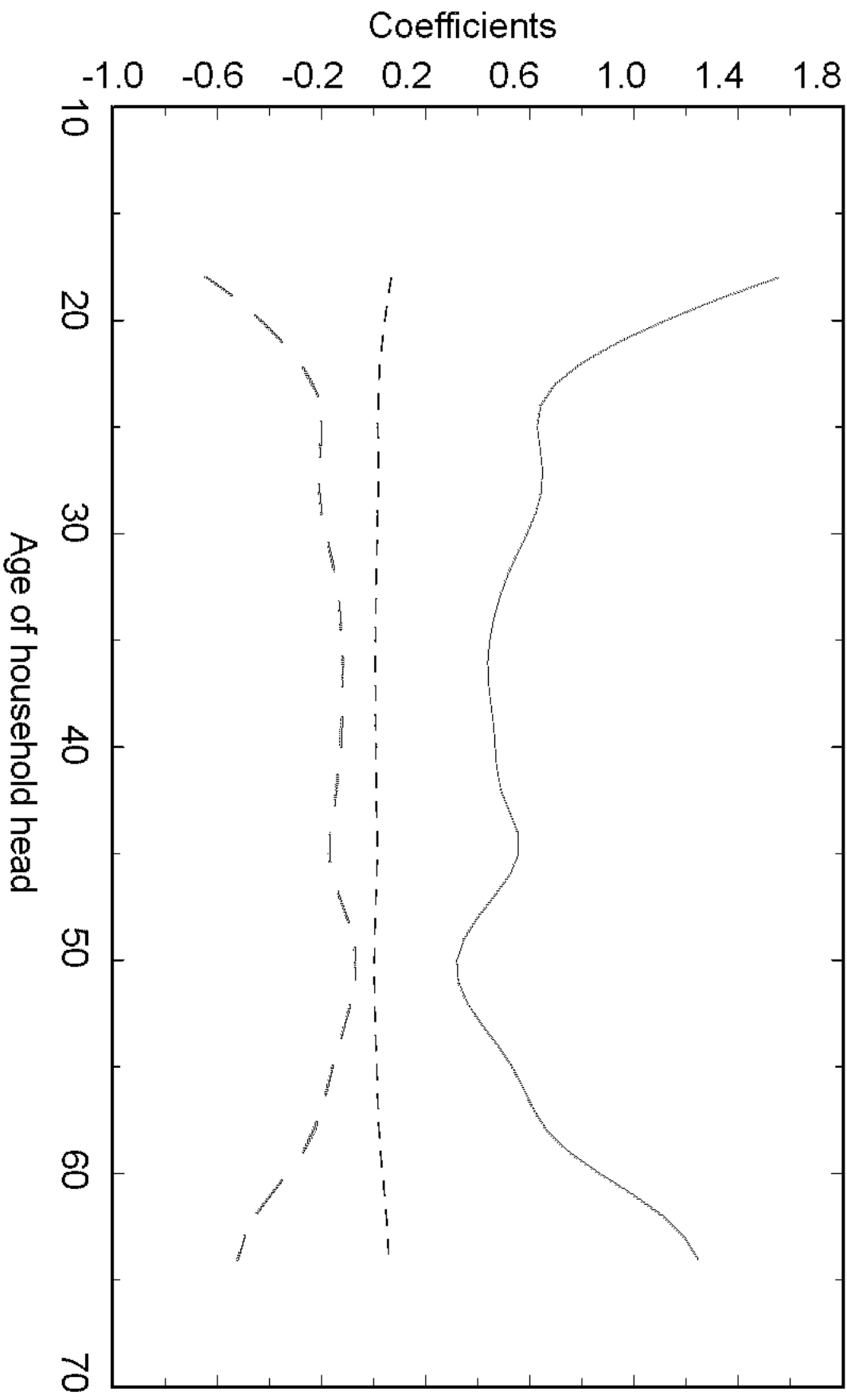
Figure 3: Smooth coefficients vs semiparametric: fuel
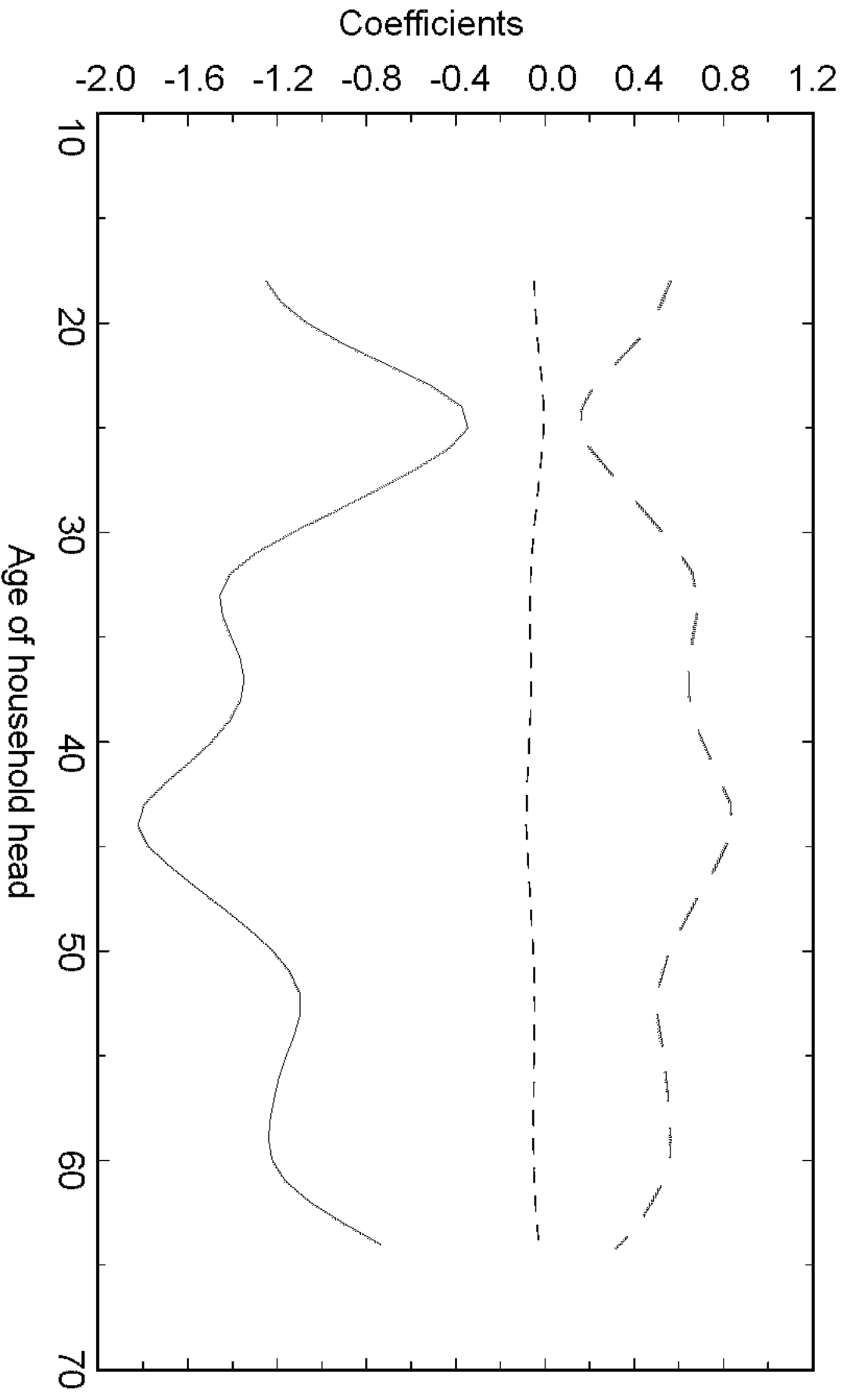
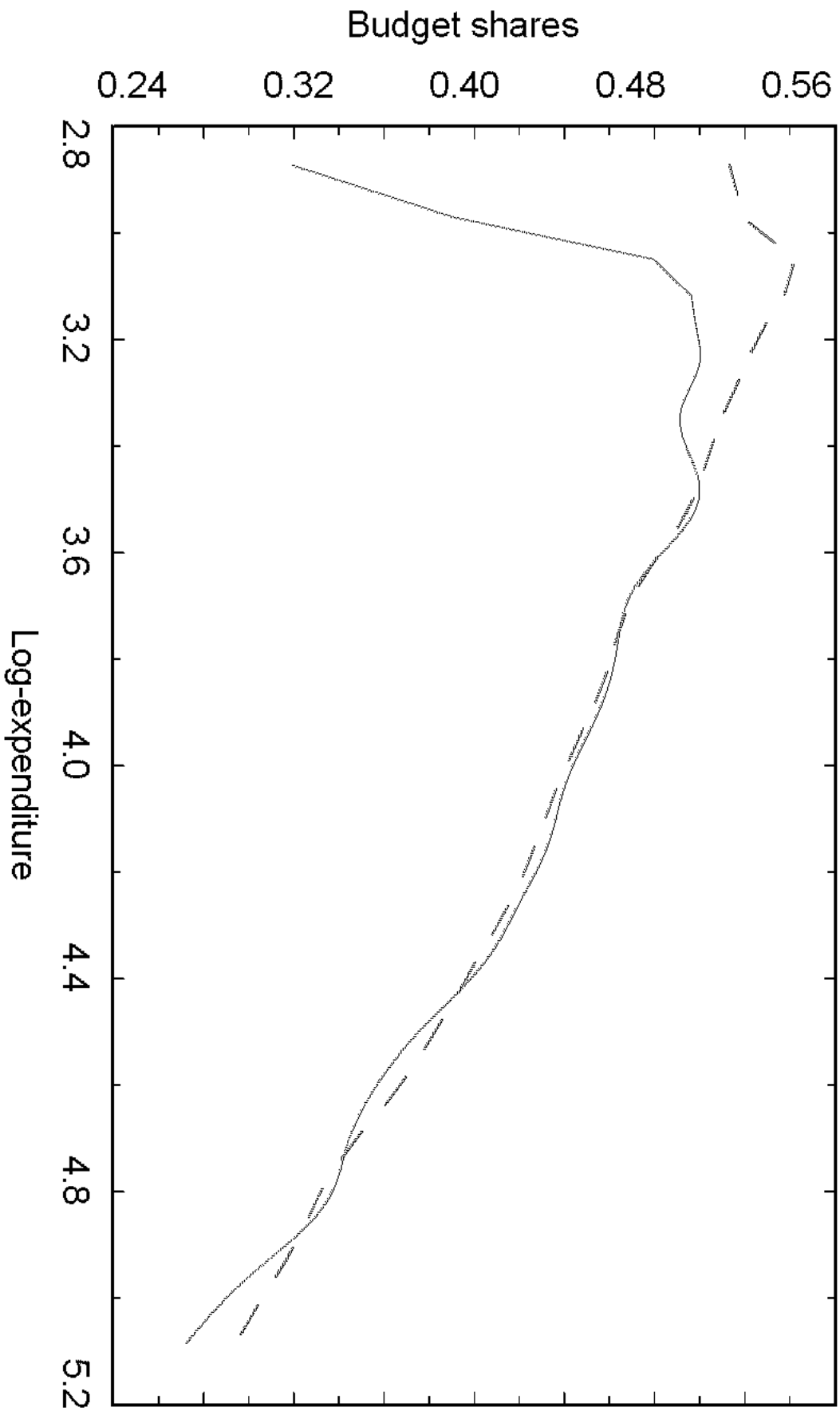Figure 4: Smooth coefficients vs household age: clothing

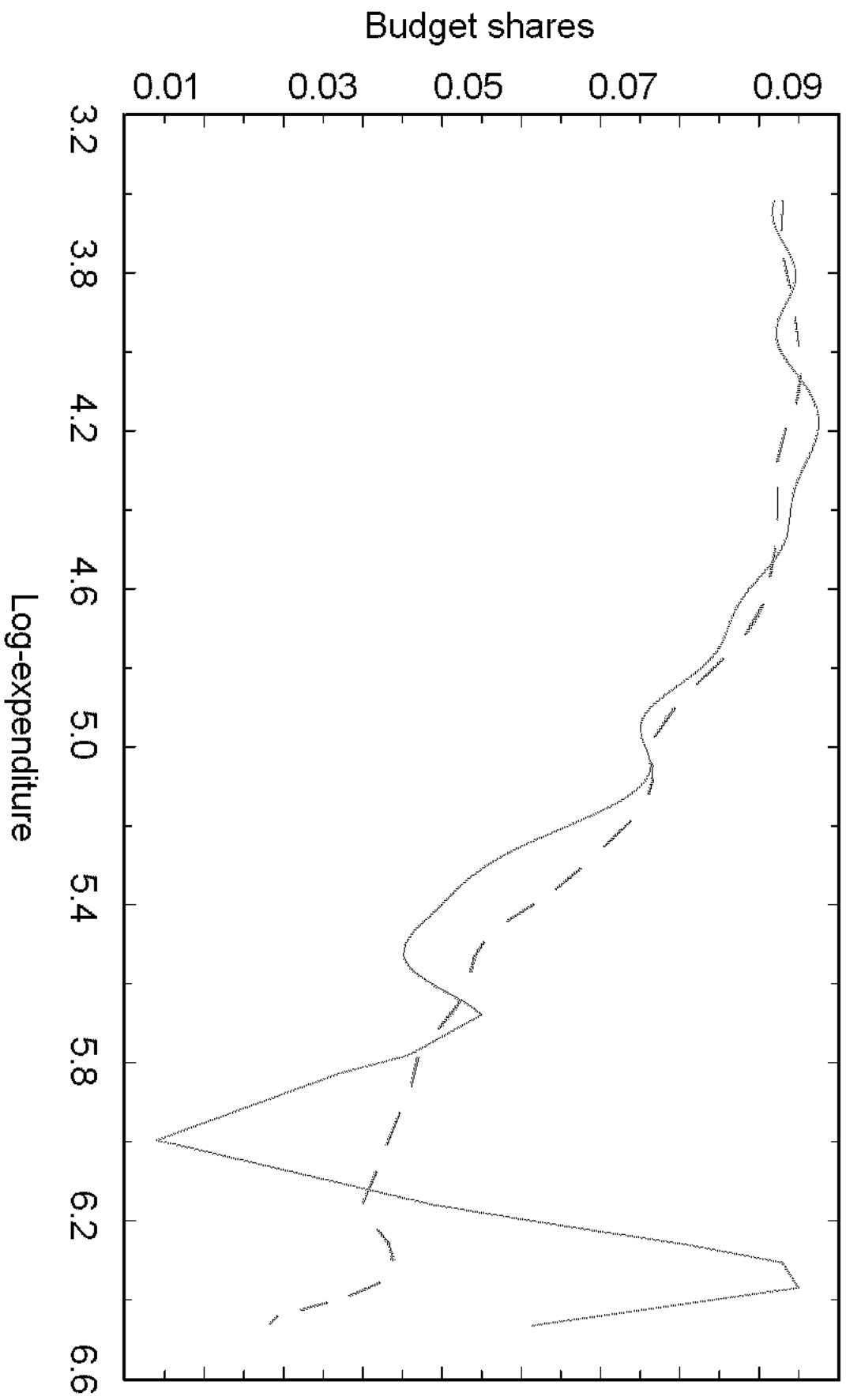Figure 5: Smooth coefficients and standard semiparametric: food

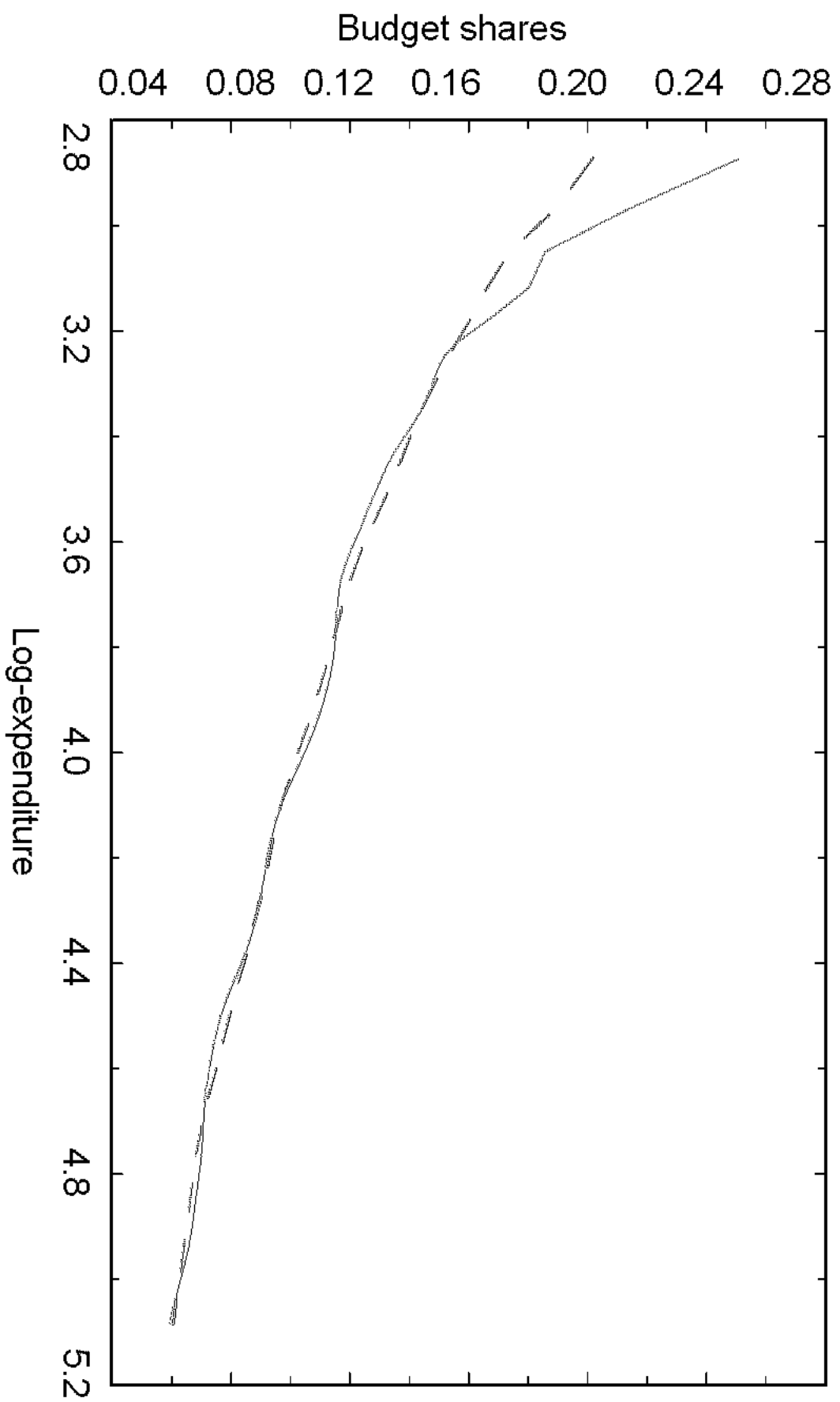Figure 6: Smooth coefficient and standard semiparametric: alcohol

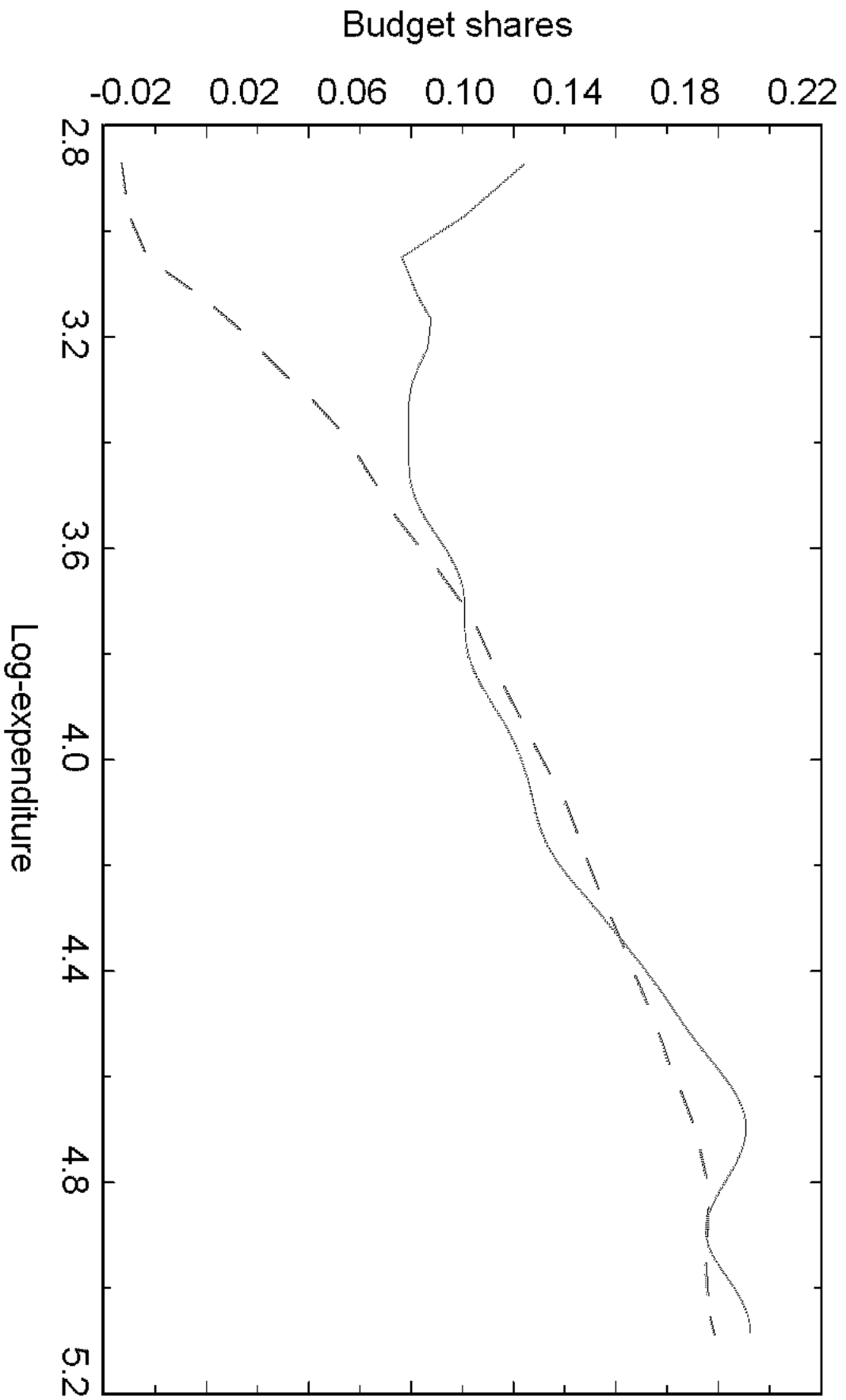Figure 7: Smooth coefficients and standard semiparametric: fuel

Figure 8: Smooth coefficients and semiparametric: clothing

Figure 9: Cost of living differences between households with different age of head