Semiparametric Estimation of Regression Functions Under Shape Invariance Restrictions.

Ralf Wilke*

University College London

Department of Economics

London, November 2001

This Version: February 2002

Abstract

This paper considers the shape invariant modelling approach in semiparametric regression estimation. Nonparametric regression functions of similar shape are linked by parametric transformations with unknown parameters. A computationally convenient estimation procedure is suggested. Finite sample performance of this estimator is investigated by simulations. Consistency of the parameter estimates is shown. An application to consumer data illustrates the importance of this method for applied statistics. Estimation results indicate that the imposed shape invariance restrictions have empirical evidence in the semiparametric modelling of consumer demand.

Keywords: Shape invariant modelling, semiparametric regression, simulations, consistency, consumer data

*This paper has been presented in seminars at University College London and University of Pittsburgh. I am grateful to the participants for the comments. I am very grateful to Hidehiko Ichimura for his supervision. I am also grateful to Richard Blundell for making comments on a previous version of this paper, giving me access to his unpublished work and for providing me the data. The author is Marie Curie training research fellow at University College London, Department of Economics, and member of the Graduiertenkolleg "Allokationstheorie, Wirtschaftspolitik und kollektive Entscheidungen" of the Deutsche Forschungsgemeinschaft at Universität Dortmund (Germany), Department of Economics and Faculty of Statistics. Corresponding address: University College London, Department of Economics, Gower Street, WC1 E6BT London, UK, wilke@statistik.uni-dortmund.de, www.wilke.org/ralf

1 Introduction

Semiparametric estimation of regression functions has become an important tool for applied statistical analysis during the past two decades. This paper is a contribution to the so called "shape invariant modelling" approach. We identify some difficulties for this class of models and impose the necessary and the sufficient conditions in order to obtain consistent estimates. Moreover, a 4 step estimation procedure is defined which is computationally convincing. Simulations show that this estimator has a better performance in finite samples than former specifications. Consistency of the parameter estimates is derived. An application to consumer data justifies the importance of this method for applied research.

Let us briefly motivate this approach with an example from consumer theory. Blundell, Duncan and Pendakur (1998) investigate expenditure shares of couples with one child that are supposed to be related by parametric transformations to the expenditure shares of couples with two children. Figure 1 presents nonparametric estimates of the transport expenditure shares for these two groups using household data from the British Family Expenditure Survey. It is apparent that the two functions are similar in shape. Consumer theory suggests that the expenditure shares for the two groups are related by a horizontal and by a vertical shift with unknown parameters. The econometrician wants to identify the unknown functions as most accurate as possible and he wants to know the true values of the parameters.

More generally, the principle of shape invariant models is the following: Suppose there is a finite number of samples with unknown regression functions. These regression functions are assumed to be similar in shape and linked by transformations with unknown parameters. There are two aims for the researcher in this approach: First, the identification of the parameters and second, the pooling of the regression functions. The first point is interesting for the usual reasons. The idea of the second is to achieve a more accurate nonparametric pooling estimate of the regression function. This paper focuses on the first point. The second was already subject to deep analysis in Pinkse and Robinson (1995).

Mainly two theoretical articles are concerned with this class of semiparametric models: Härdle and Marron (1990) and Pinkse and Robinson (1995). The first paper provides a general framework for nonstochastic regressors and derives asymptotic properties of the estimators whereby the consistency proof is not convincing. The second paper considers the

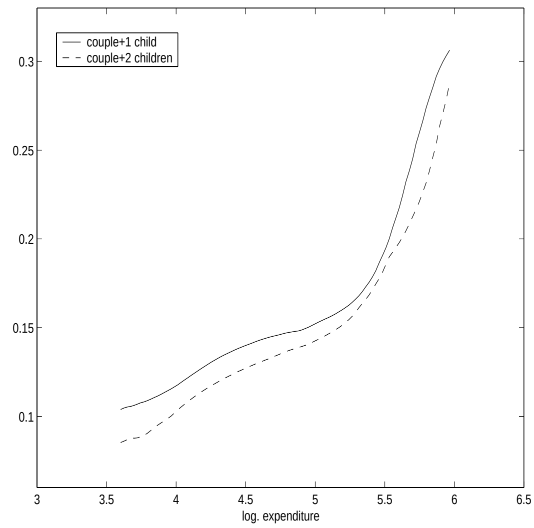

Figure 1: Nonparametric estimates of transport expenditure shares.

case of independent stochastic regressors and the case with a limited dependency between the stochastic regressors of the samples. The authors show $\sqrt{N}$ consistency of the parameter estimators. However, their chosen specification of the loss function is not convincing since it imposes a weak finite sample performance.

However, three general difficulties are involved in the shape invariant modelling approach: 1. The general estimation method is defined in such a way that it minimizes a loss function over a multi dimensional parameter space. The loss mainly consists of the distances between the nonparametric regression estimates. The researcher has carefully to select an appropriate algorithm in order to avoid exploding computational effort. 2. It might be the case that the true parameters of a horizontal shift have a value such that the two samples are indeed not comparable. 3. The shape of the unknown functions has to be restricted in order to ensure the consistency of the parameter estimates. The purpose of this paper is to tackle these problems such that this class of estimators can become more popular in applied research.

The paper is organized as follows: Section 2 presents the model, defines a 4 step estimation and provides an intuitive discussion of the above mentioned three difficulties. Section 3 investigates these findings with the help of Monte Carlo studies. Moreover, an estimator using the Härdle and Marron specification is compared to an estimator of the Pinkse and Robinson specification. Explanations for the different behavior of the two specifications are also provided. Section 4 imposes the necessary and the sufficient conditions on the model such that the parameter estimates are indeed consistent. Section 5 presents an application to consumer data.

2 The Model

Consider two samples $(Y_i, X_i)_{i=1,\ldots,N}$ and $(Z_i, W_i)_{i=1,\ldots,N}$ of size N. The samples sizes might also be different without affecting the following analysis. Suppose

$$\begin{aligned} Y_i &= m_0(X_i) + U_i \\ Z_i &= m_1(W_i) + V_i,\ i = 1, \ldots, N \end{aligned}$$

with $E[U_i|X_j] = E[V_i|W_j] = 0$ almost surely for all i, j. U_i and V_i have finite second moments and the pairs (U_i, V_i) are mutually independent. $X_i \in \mathcal{X}_1$ and $W_i \in \mathcal{W}$ are random variables with i.i.d realizations on compact sets with continuous marginal distributions $\inf_{x\in\mathcal{X}_1} f_x(x) > 0$ and $\inf_{w\in\mathcal{W}} f_w(w) > 0$. Suppose the unknown functions m_0 and m_1 are twice differentiable. Let m_0 and m_1 and its first two derivatives be uniformly continuous and bounded over its supports. Furthermore a_0, b_0 and μ_0 are unknown parameters in the interior of open subsets in $\mathbb{R}$. The following equation is supposed to hold:

$$m_1(x) = a_0 + b_0 m_0\big(T_{\mu_0}(x)\big), \tag{1}$$

where T is an invertible parametric transformation with $T_{\mu_0}^{-1}(W_i) \in \mathcal{X}_2$ and $T_\mu(X_i) \in \mathcal{W}_\mu$. In other words there exist parametric links with unknown parameters between the unknown functions m_0 and m_1. Let us denote $\hat{m}_1(x)$ and $\hat{m}_\mu(x) = \hat{m}_0(T_\mu(x))$ the nonparametric estimates of $m_1(x)$ and $m_\mu(x) = m_0(T_\mu(x))$ respectively. This model setup is similar to one of the models defined in Pinkse and Robinson (1995).

Since we intent to analyze a problem with a simple structure, we suppose in the following $T_\mu(x) = T_c(x) = x - c$. Model (1) becomes now:

$$m_1(x) = a_0 + b_0 m_0(x - c_0), \tag{2}$$

where $c_0 \in C \subset \mathbb{R}$. Accordingly, let denote $\mathcal{W}_c = \mathcal{W}_\mu$ and $m_c(x) = m_0(x - c)$.

Pinkse and Robinson (1995) The definition of this estimator is essentially based on the Nadaraya-Watson estimator. Define:

$$\begin{aligned} \hat{m}_1(x) &= \hat{r}(x)/\hat{f}(x) \text{ and} \\ \hat{m}_c(x) &= \hat{r}_c(x)/\hat{f}_c(x), \end{aligned}$$

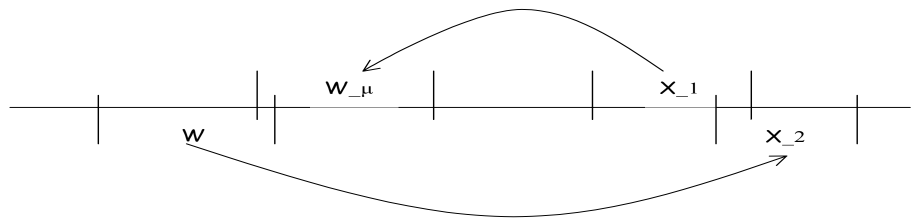

Figure 2: Relationship between $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{W}$ and $\mathcal{W}_\mu$.

where

$$\hat{r}(x) \quad = \quad \frac{1}{Nh_N}\sum_i K\left(\frac{x-X_i}{h_N}\right)Y_i$$

and

$$\hat{f}(x) \quad = \quad \frac{1}{Nh_N}\sum_i K\left(\frac{x-X_i}{h_N}\right),$$

where $K(*)$ is a nonnegative symmetric Kernel function and $h_N > 0$ denotes the bandwidth which is a function of N. Finding the parameter estimates corresponds to minimizing the loss function

$$L_N(a,b,c) = \int \left[\hat{f}(x)\hat{r}_c(x) - a\hat{f}(x)\hat{f}_c(x) - b\hat{f}_c(x)\hat{r}(x)\right]^2 w(x)dx$$

with respect to the parameters, where $w(x)$ is a nonnegative weight function. It is later shown that this specification of the loss function imposes two essential weaknesses for the estimation: First, the loss is zero whenever the marginal distributions are zero. Second, due to the multiplicative writing of the elements $\hat{r}$, $\hat{r}_c$, $\hat{f}$ and $\hat{f}_c$, the finite sample bias for this specification is greater in comparison to using the fractions $\hat{r}/\hat{f}$ and $\hat{r}_c/\hat{f}_c$. See Section 3 for a detailed discussion.

Härdle and Marron (1990) Suppose $\hat{m}_1(x)$ and $\hat{m}_c(x)$ are nonparametric estimates of $m(x)$ and $m_c(x)$ respectively. The parameters are estimated by minimizing the loss function

$$L_N(a,b,c) = \int \left[\hat{m}_1(x) - a - b\hat{m}_c(x)\right]^2 w(x)dx, \tag{3}$$

where $w(x)$ is a known nonnegative weight function. This loss function is also minimized whenever the marginal distributions are zero.

More generally, let us outline the difficulties that are involved in the above defined model:

- **Computation Problem:** The loss function is to be minimized numerically on a multidimensional parameter space. In practice this is done with compact parameter spaces. This requires a lot of computational effort.
- **Support Problem:** If the supports of X_i and $W_i + c_0$ are disjoint compact sets, the function $m_1\big(W_i + c_0\big)$ cannot be compared to $m_0(X_i)$ since their nonparametric estimates are evaluated on different supports even as $N \to \infty$.
- **Identification Problem:** The unknown function m_0 has to follow some shape restrictions otherwise the parameters cannot be identified.

Computation problem Suppose for instance that $\mathcal{X}_1 \cap \mathcal{X}_2$ is non empty. Let us now introduce an alternative formulation for the loss function criterion as given in (2) and (3). A four step estimator is defined for this purpose:

1. Estimate m_0 and m_1 on their support using a nonparametric estimator.
2. Define $R_c = \big(1 \; \hat{m}_c(x)\big)$. The least squares estimator for a and b, given c is defined as
$$\begin{pmatrix} \hat{a}_c \\ \hat{b}_c \end{pmatrix} = (R_c' R_c)^{-1} R_c' \hat{m}_1(x)$$
3. Estimate c by minimizing the loss function
$$\begin{aligned} L_N(c) &= \frac{\int \mathbb{I}_{\{x \in \mathcal{W} \cap \mathcal{W}_c\}} \big[\hat{m}_1(x) - \hat{a}_c - \hat{b}_c \hat{m}_c(x)\big]^2 w(x)dx}{\int \mathbb{I}_{\{x \in \mathcal{W} \cap \mathcal{W}_c\}} w(x)dx} \\ &= \frac{\int_{\mathcal{W} \cap \mathcal{W}_c} \big[\hat{m}_1(x) - \hat{a}_c - \hat{b}_c \hat{m}_c(x)\big]^2 w(x)dx}{\int_{\mathcal{W} \cap \mathcal{W}_c} w(x)dx} \qquad \text{(HM)} \end{aligned} \tag{4}$$
where the integral is now restricted to the intersection of $\mathcal{W}$ and $\mathcal{W}_c$ since this is the sole part where both samples are comparable. The denominator is required for weighting purposes. We have to compensate for the fact that the size of $\mathcal{W} \cap \mathcal{W}_c$ depends on c.
4. $\hat{a} = \hat{a}_{\hat{c}}$ and $\hat{b} = \hat{b}_{\hat{c}}$.

This estimator is to be referred as the HM 4 step estimator. Instead of minimizing (4) one could also use the Pinkse and Robinson specification for the third step:

$$L_N(c) = \frac{\int_{\mathcal{W}\cap\mathcal{W}_c}\left[\hat{f}(x)\hat{r}_c(x) - \hat{a}_c\hat{f}(x)\hat{f}_c(x) - \hat{b}_c\hat{f}_c(x)\hat{r}(x)\right]^2 w(x)dx}{\int_{\mathcal{W}\cap\mathcal{W}_c} w(x)dx} \quad \text{(PR)}. \tag{5}$$

This specification is to be referred as the PR 4 step estimation.

By breaking up the loss function minimization into two parts, the numerical minimization reduces to a one dimensional problem which has the following advantages:

- Minimization with respect to a and b on a unbounded parameter space with low computational effort.
- Minimization of L reduces to a one dimensional problem. Allows for graphical analysis.
- If the grid on C is carefully selected, the unknown functions have only to be estimated once.

This formulation of the estimation procedure induces therefore low computational effort.

Support problem We require some restrictions on the parameter space in order to ensure that the two samples are comparable.

Proposition 1 *If $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$, $m_0(x)$ and $m_0(w + c_0)$ are observed on disjoint support and hence, they cannot be compared. Then a, b, c are not identifiable. Pooling of the two samples does not improve the accuracy of the nonparametric estimate of m_0.*

In practice we therefore have to ensure that c_0 is located in a suitable parameter space with respect to $\mathcal{X}_1$ and $\mathcal{W}$.

An example is given in Figure 3: $\mathcal{X}_1 \cap \mathcal{X}_2 = [5, 12]$. Accordingly, $\mathcal{W} \cap \mathcal{W}_{c_0} = [0, 7]$. If $|c_0| \geq 12$, the functions are observed on different supports.

Identification problem The identification of the unknown parameters in model (2) is not yet ensured. The loss function under the above conditions needs not to have a unique global minimum at the true parameter values. This paragraph describes intuitively the main identification conditions which are formally derived in the proof of consistency. In particular, we have to impose some shape restrictions for the unknown function m_0. These conditions are violated if:

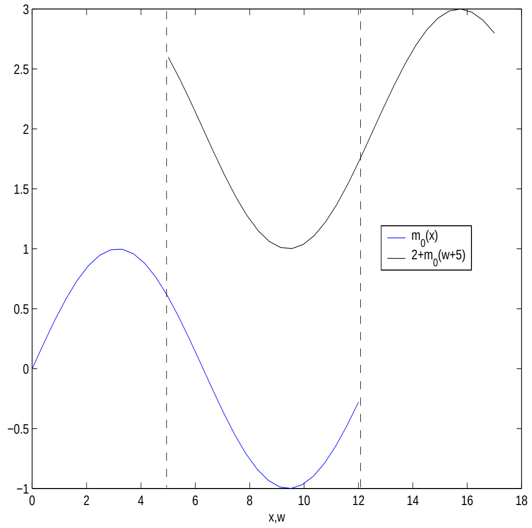

Figure 3: Intersection of observed support: $y = sin(0.5x)$, $z = 2 + sin\big(0.5(w + 5)\big)$, $\mathcal{X}_1 = [0, 12]$, $\mathcal{W} = [0, 12]$ and $\mathcal{X}_2 = [5, 17]$.

1. The unknown function $m_0(x)$ belongs to the class of linear functions.

2. The unknown function $m_0(x)$ is cycling, i.e.

$$\exists c \in C \text{ such that for all } x - c \in \mathcal{W} \cap \mathcal{W}_c,\ m_0(x) = m_0(x - c).$$

The first difficulty makes it impossible to identify a and c. The loss functions (4) and (5) are constant in this case, i.e. $L(c) = L$:

Proposition 2 *If $m_0(x)$ belongs to the class of linear functions, $L(c)$ is constant and therefore does not posses a local minimum since the sufficient condition $\partial_c^2 L(c) > 0$ does not hold. The parameters a and c cannot be identified. Nevertheless, a pooling estimate might yield a more accurate estimate of the unknown function.*

The second difficulty implies that (4) and (5) do not have a unique minimum on the support of c but there is a multiple set of global minima. Therefore c cannot be identified.

Proposition 3 *If m_0 is cycling on $\mathcal{W} \cap \mathcal{W}_c$, the parameter c cannot be identified.*

Figure 4 presents an example using a cycling sine function. In this case there are three minima of the loss function on C.

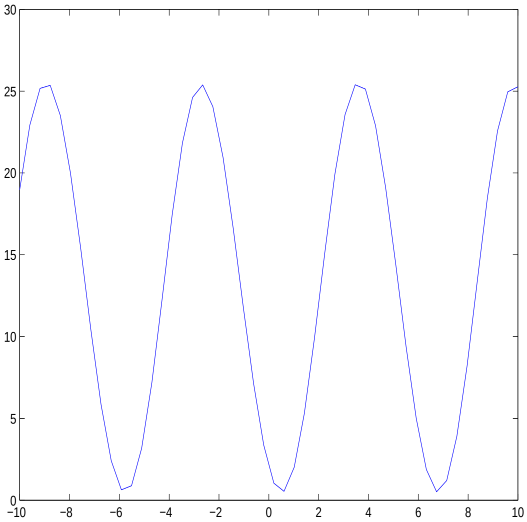

Figure 4: Multiple minima of the loss function: $y = sin(0.5x)$, $z = 5 + 0.5sin(0.5(x - c))$, $C = [-10, 10]$, $c_0 = 0.5$.

However, the smaller is the intersection of $\mathcal{X}_1$ and $\mathcal{X}_2$ the more unlikely the non-linearity condition holds because we have imposed some smoothness conditions on the unknown functions. This might lead to the following complication: The nonlinear parts of $m_{c_0}(x)$ drop out of the support and a and c are not longer identifiable.

Proposition 4 *If the intersection of $\mathcal{X}_1$ and $\mathcal{X}_2$ is too small, the identification of the parameters might be impossible even as $N \to \infty$.*

This difficulty should have relevance in applications. It is therefore reasonable to restrict C such that the intersection of $\mathcal{W}$ and $\mathcal{W}_c$ is not too small. However, even if the parameters are identifiable, the convergence rate of the parametric estimator is lower than $\sqrt{N}$ since many of the observations cannot be used for the estimation.

3 Simulations

Let us now investigate the finite sample performance of the above defined 4 step estimator using the HM specification as given in (4) and the PR specification as given in (5). Moreover, the semiparametric estimators are compared to a parametric estimator. It turns out that the semiparametric estimators differ. Explanations for these differences are provided afterwards.

	HM4SE	PR4SE	HM4SE	PR4SE
a	5.2328(1.2211)	6.8020(6.7376)	4.8844(0.3122)	4.1837(4.0385)
b	2.1716(7.9016)	-0.4570(31.2005)	0.2633(10.4807)	-1.1374(19.3405)
c	0.2398(2.2289)	0.4324(12.6926)	0.9527(10.5263)	-1.6030(12.8688)

Table 1: Mean parameter estimates of the first (left) and of the second (right) Monte Carlo experiment; (variances in brackets)

Let λ denote the Lebesgue measure defined on the smallest σ-algebra containing all open sets in $\mathbb{R}$. For simplicity suppose $\lambda(\mathcal{X}_1) = \lambda(\mathcal{W})$ in this section. Suppose also $\mathcal{X}_1 = \mathcal{W}$ and $\lambda(\mathcal{X}_1 \cap \mathcal{X}_2) \geq \lambda(\mathcal{X}_1)/2$. The latter condition implies $c \in [-\lambda(\mathcal{X}_1), \lambda(\mathcal{X}_1)]$. As a consequence of Proposition 4 we restrict C such that $c \in [-\lambda(\mathcal{X}_1)/2, \lambda(\mathcal{X}_1)/2]$. Therefore, C is properly defined.

Monte Carlo Study Two Monte Carlo series shall help investigating the properties of both estimators. The following model is used:

$$\begin{aligned} m_1(x) &= 5 + 3sin\big(0.5(x - c_0)\big) \\ m_0(x) &= sin(0.5x), \end{aligned}$$

$X_i, W_i \sim U(0, 10)$, $U_i, V_i \sim N(0, 1)$, $N = 200$, 1000 simulations. The two experiments only differ due to the value of c_0, where we use $c_0 = 0$ in the first experiment and $c_0 = 4$ in the second Monte Carlo experiment. The model setup up is interesting because the estimators have to detect a unique minimum of the loss function in the first experiment and two minima in the second experiment.

Figure 5 and 6 show the mean loss functions in c for the parametric estimator, the HM 4 step estimator and the PR 4 step estimator. Note that the loss functions have different scalings and can therefore only compared in relative shape. Table 1 presents the mean parameter estimates of the two experiments.

The results of the simulations can be summarized as follows: The HM 4 step estimator detects any minimum of the loss functions. This is in contrast to the PR 4 step estimator which performs badly in the second experiment since it does not detect one of the minima. Moreover, from Table 1 it is apparent that the HM 4 step estimator is superior to the PR 4 step estimator under the imposed model specification.

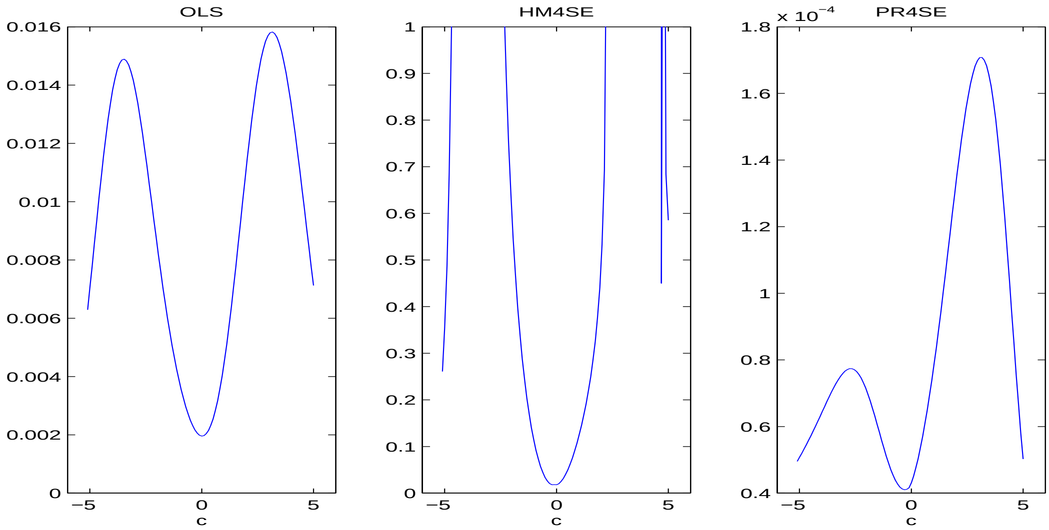

Figure 5: Mean conditional Loss functions $L(c|a,b)$ of the first Monte Carlo Series ($c_0 = 0$): a) parametric b)HM 4 step estimator c) PR 4 step estimator.

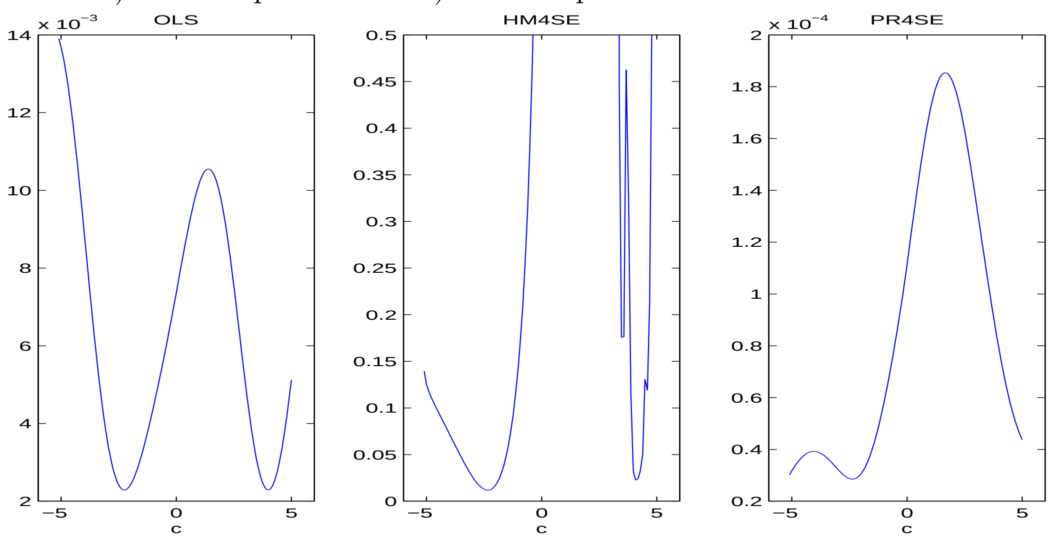

Figure 6: Mean conditional Loss functions $L(c|a,b)$ of the second Monte Carlo Series ($c_0 = 4$): a) parametric b)HM 4 step estimator c) PR 4 step estimator.

A variation of C should therefore lead to a significant shift or change in shape of the distribution of $\hat{c}$ as estimated by one of the above estimators. Histograms, as given in Figure 7 and 8 support this guess for the PR 4 step estimator. A Scientist who applies these estimators to data might be faced to such a situation. In this case a graphical analysis of the loss function is a very convenient way to check whether there exists a unique global minimum.

The next paragraph discusses why the two estimators behave different.

On the differences between the HM and the PR specification The differences between the two specifications are due to two effects:

1. different distributions of the errors (Variance effect)
2. proportionality of the bias (Bias effect)

1. Variance effect: Suppose that in both specifications we use the Nadaraya-Watson estimator:

$$\begin{aligned} \hat{r}(x) &= r(x) + \epsilon_r(x),\ \hat{r}_c(x) = r_c(x) + \epsilon_{r_c}(x) \\ \hat{f}(x) &= f(x) + \epsilon_f(x),\ \hat{f}_c(x) = f_c(x) + \epsilon_{r_c}(x), \end{aligned}$$

where $\epsilon_l(x)$ are random variables. These pointwise errors depend on the marginal distributions, the bandwidths and the unknown regression functions. In HM 4 step estimation we minimize

$$\frac{r_c(x) + \epsilon_{r_c}(x)}{f_c(x) + \epsilon_{f_c(x)}} - a - b\frac{r(x) + \epsilon_r(x)}{f(x) + \epsilon_f(x)}$$

and the PR 4 step estimator minimizes

$$\begin{aligned} r_c f + r_c\epsilon_f + f\epsilon_{r_c} + \epsilon_{r_c}\epsilon_f &- a\big[f_c f + f_c\epsilon_f + f\epsilon_{f_c} + \epsilon_f\epsilon_{f_c}\big] \\ &- b\big[r f_c + r\epsilon_{f_c} + f_c\epsilon_r + \epsilon_r\epsilon_{f_c}\big], \end{aligned}$$

where we write $f(x) = f$ etc..

The variance effect becomes clear when considering a simplified case: Suppose $\epsilon_f = \epsilon_{f_c} = 0$, i.e. the marginal distributions are known. The minimization problems become:

$$\frac{r_c(x) + \epsilon_{r_c}(x)}{f_c(x)} - a - b\frac{r(x) + \epsilon_r(x)}{f(x)}$$

for the HM specification and

$$r_c(x)f(x) + f\epsilon_{r_c}(x) - af_c(x)f(x) - b\big[r(x)f_c(x) + f_c(x)\epsilon_r(x)\big]$$

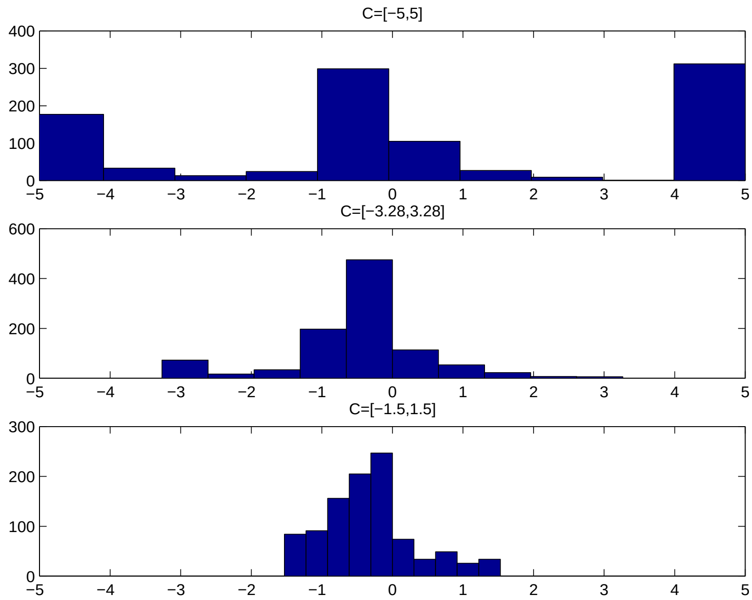

Figure 7: Three histograms for the distribution of $\hat{c}$ obtained with the Pinkse-Robinson 4 step estimator using different supports of c. First Monte Carlo series ($c_0 = 0$).

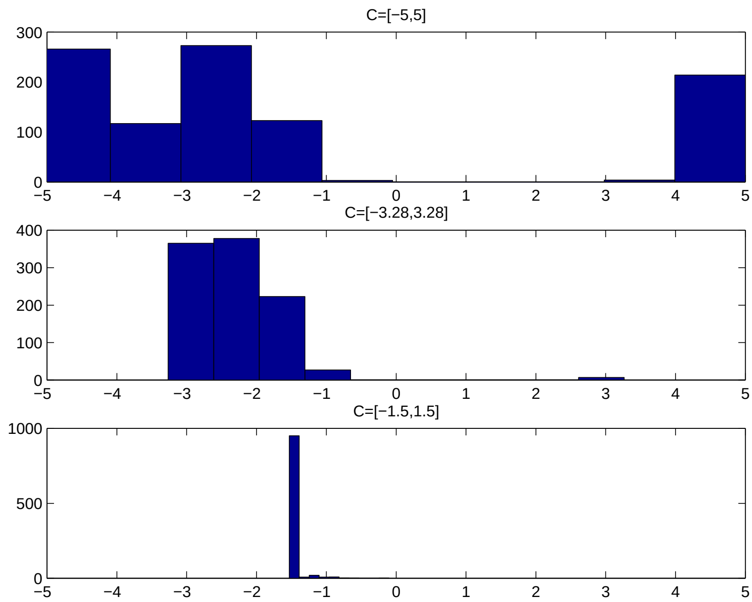

Figure 8: Three histograms for the distribution of $\hat{c}$ obtained with the PR 4 step estimator using different supports of c. Second Monte Carlo series ($c_0 = 4$).

for the PR specification. It is clear that if $f_c(x) = f(x)$, both estimators are the same. Otherwise it is important to point out that their error distributions differ. The variance of the HM 4 step estimator is larger whenever f_c and f are smaller than one. Otherwise it is smaller.

2. Bias effect: The second point becomes clear when rewriting the problem:

$$\begin{aligned}\hat{r}(x) &= r(x)\xi_r(x),\ \hat{r}_c(x) = r_c(x)\xi_{r_c}(x) \\ \hat{f}(x) &= f(x)\xi_f(x),\ \hat{f}_c(x) = f_c(x)\xi_{r_c}(x)\end{aligned}$$

and for the HM specification we obtain accordingly

$$\frac{r_c(x)\xi_{r_c}(x)}{f_c(x)\xi_{f_c}(x)} - a - b\frac{r(x)\xi_r(x)}{f(x)\xi_f(x)}.$$

$\xi_r(x)$ and $\xi_f(x)$ are unequal to one whenever the corresponding estimates are biased. From Figure 9 it is moreover apparent that $\xi_f(x)$ and $\xi_r(x)$ are very similar functions. Therefore, their ratio deviates less from one than each of the functions itself. A part of the pointwise bias is therefore ruled out by the division. Rewriting the estimator in the Pinkse and Robinson style causes a loss of this nice property. Estimators using the specification

$$f(x)\xi_f(x)r_c(x) - af_c(x)\xi_{f_c}(x)f(x)\xi_f(x) - br(x)\xi_r(x)f_c(x)\xi_{f_c}(x)$$

therefore behave worse in the case of small samples in particular at the boundaries where the bias is supposed to be large. This effect becomes stronger due to the multiplicative structure. As a consequence the estimates are more affected by the bias of $\hat{f}$, $\hat{f}_c$, $\hat{r}$ and $\hat{r}_c$.

We conclude that there is a trade-off and it depends on the specific situation which estimator is preferable. In small samples the second point should clearly dominate the first one, since the systematic bias has more evidence. The PR specification should therefore not be applied in these cases. The simulations ($N = 200$) impressively support these findings. In the second experiment ($c_0 = 4$) the overlapping support at $c_0 = 4$ is small. Since the two nonparametric estimates are supposed to be more biased at the boundaries, we expect the same for the estimates of the unknown functions on a large subset of $\mathcal{W} \cap \mathcal{W}_{c_0}$. As a consequence of the above findings, the estimator using the PR specification is not able to detect the second minimum of the loss function (Figure 6c).

4 Asymptotic Properties

This section derives asymptotic properties for the model defined in Section 2. For this purpose we use a modified Härdle and Marron loss function as given in (3) that incorporates

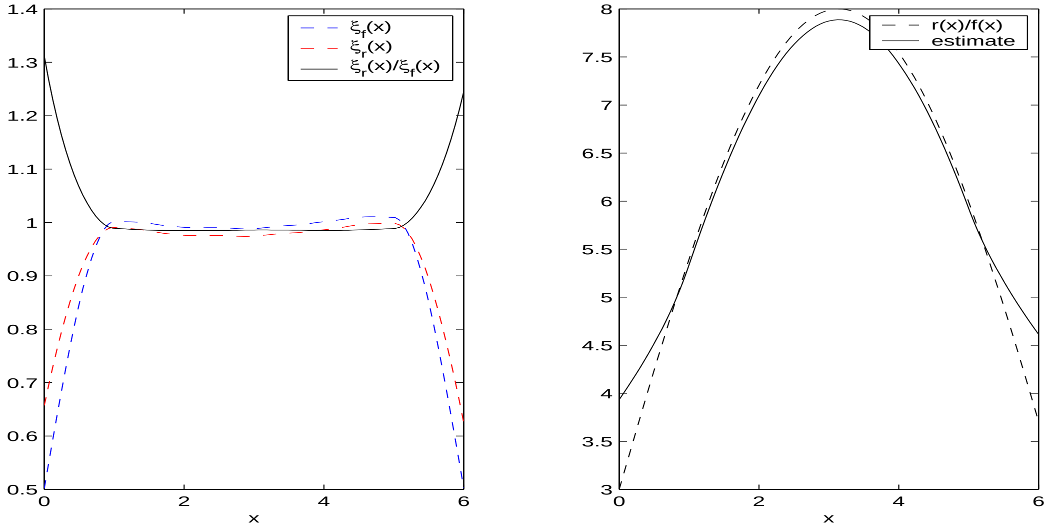

Figure 9: Proportionality of the bias: $N = 2000$, $y = 3 + 5sin(0.5x)$, $x \sim U[0, 6]$, $h = 0.5$, mean of 1000 samples

the intuitive findings of Section 2.

Consistency Härdle and Marron (1990) assume that the loss function is convex around the true parameter values. We are going to derive here the necessary and sufficient conditions on the shape of the unknown function m_0 such that the loss function has indeed a unique minimum.

Denote by $\hat{m}(x)$ the nonparametric estimate of m_0 evaluated at $X_i = x$. Accordingly, we have $x - c \in \mathcal{W}_c$ for all $c \in C$. Let be $\{x_t - c\} = \{x - c | x - c \in \mathcal{W}\}$ for all $c \in C$. Define $T_c = \text{card}\{x - c | x - c \in \mathcal{W}\}$. Note that $T_c \leq N$ and T_c weakly increases in N.

Assumption 1 *c_0 is an interior point of C, where C is such that for all $c \in C$: $T_c \geq 3$.*

$T_c > 0$ solves the support problem. $T_c \geq 3$ is required for the identifiability of a, b and c. Define a sequence $t = 1, \ldots, T_c$ of evaluation points $w_t^c \in \mathcal{W}$ such that for a given c: $\{w_t^c\}_{t=1,\ldots,T_c} = \{x_t - c\}_{t=1,\ldots,T_c}$. Denote $\{\hat{m}_1(w_t^c)\}_{t=1,\ldots,T_c} = \{\hat{m}_1(x_t - c_0)\}_{t=1,\ldots,T_c}$ as the nonparametric estimates of m_1 evaluated at w_t^c. Moreover, denote $\hat{m}_0(x - c)$ as the nonparametric estimate of m_0 evaluated at x and horizontally shifted to $x - c$ for all x and c. The loss function (4) can then be rewritten as:

$$L_N(a, b, c) = \sum_{t=1}^{T_c} \left[\hat{m}_1(x_t - c_0) - a - b\hat{m}_0(x_t - c)\right]^2 / T_c. \qquad (6)$$

Intuitively, the loss per evaluation point is minimized. Note that this function depends on N due to the nonparametric estimates and T_c. The following two assumption are necessary for the identifiability and have already been discussed in Section 2.

Assumption 2 *$m_0(x_t - c)$ is not cycling on $\mathcal{W} \cap \mathcal{W}_c$, i.e. there does not exists $c \neq c_0$ such that $m_0(x_t - c) = m_0(x_t - c_0)$ for all $x_t - c \in \mathcal{W} \cap \mathcal{W}_c$.*

Assumption 3 *$m_0(x_t - c)$ is nonlinear on $\mathcal{W} \cap \mathcal{W}_c$ for all c, i.e.*

$$(1 \ m(x_t - c) \ m'(x_t - c))$$

are linearly independent on $\mathcal{W} \cap \mathcal{W}_c$ for all c.

Assumptions 1-3 ensure the necessary conditions for the consistency of the parameter estimates.

The nonparametric estimates for m_0 and m_1 can be written as

$$\hat{m}_1(x_t - c_0) = a_0 + b_0 m_0(x_t - c_0) + \epsilon_1(w_t^c, N) \tag{7}$$

$$\hat{m}_0(x_t - c) = m_0(x_t - c) + \epsilon_0(x_t - c, N). \tag{8}$$

for $t = 1, \ldots, T_c$ given $c \in C$.

Assumption 4 *$\epsilon_0(x, N)$ and $\epsilon_1(w, N)$ converge to 0 in probability uniformly in x and w, i.e.*

$$lim_{N\to\infty} P\big[sup_{x\in\mathcal{X}_1} |\epsilon_0(x, N)| < \delta\big] = 1 \ for \ any \ \delta > 0$$
$$lim_{N\to\infty} P\big[sup_{w\in\mathcal{W}} |\epsilon_1(w, N)| < \delta\big] = 1 \ for \ any \ \delta > 0.$$

This assumption can be for example justified for the class of Kernel estimators by the following theorem:

Theorem 2.1 Nadaraya (1989), p.122 *The Kernel estimators of the regression functions are uniformly strongly consistent, i.e.*

$$sup_{x\in\mathcal{X}_1} |\hat{m}_0(x) - m_0(x)| \to 0 \quad a.s.$$
$$sup_{w\in\mathcal{W}} |\hat{m}_1(w) - m_1(w)| \to 0 \quad a.s.$$

if the following conditions on the bandwidth and on the Kernel function hold:
$0 < h_N \rightarrow 0$ *as* $N \rightarrow \infty$ *and*

$$\sum_{N=1}^{\infty} exp(-\gamma N h_N^2) < \infty \quad \textit{for any } \gamma > 0.$$

$K(x)$ *is a kernel function which satisfies:*

$$\begin{aligned}
&sup_{-\infty<x<\infty}\left|K(x)\right| < \infty \\
&lim_{|x|\rightarrow\infty}\left|x\right|K(x) = 0 \\
&K(x) = K(-x) \\
&x^2 K(x) dx \in L_1(-\infty, \infty) \\
&K(x) \quad \textit{is a function with bounded variation on } \mathcal{X}_1 \textit{ and } \mathcal{W}.
\end{aligned}$$

However, there is a broad class of nonparametric estimators satisfying Assumption 4, e.g. local polynomials and splines.

Let us now state the theorem of this paragraph which says that the parameter estimates $\hat{a}$, $\hat{b}$ and $\hat{c}$ are weakly consistent under appropriate regularity conditions:

Theorem 1 *Under Assumptions 1-4, a root of Model (6) is consistent, i.e.*

$$lim_{N\rightarrow\infty} P \left[inf_{a,b,c \in \hat{\mathcal{B}}} \left(\begin{pmatrix} a \\ b \\ c \end{pmatrix} - \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} \right)' \left(\begin{pmatrix} a \\ b \\ c \end{pmatrix} - \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} \right) > \epsilon \right] = 0 \textit{ for any } \epsilon > 0$$

where $\hat{\mathcal{B}}$ *is the set of roots. Moreover, the set of roots consists of one single element.*

Proof: Appendix 1.

Asymptotic Normality Asymptotic normality has already been shown by Härdle and Marron (1990) and Pinkse and Robinson (1995) for their frameworks. Both show that despite the lower convergence rate of the nonparametric estimates, the rate $\sqrt{N}$ for the parametric estimates can be achieved. Whether it is indeed achieved mainly depends on the convergence rate of the nonparametric estimator.

Using the loss function specification as given in (6), this does not hold in general since $T_c \leq N$. If $\mathcal{X}_1$ and $\mathcal{X}_2$ are small enough then T_{c_0} tends to be much smaller than N as

N becomes large. In this case it is hard to believe that the parameter estimates converge at rate $\sqrt{N}$ since the number of observations that count for the comparison of the two samples is much smaller. It is therefore more reasonable that the convergence rate of the parameters depends on the probability that the event $X_i - c_0 \in \mathcal{W}$ and $W_i \in \mathcal{W}_{c_0}$ occurs. This probability depends on the proportion of each marginal distribution that is assigned to $\mathcal{W} \cap \mathcal{W}_{c_0}$:

$$\int_{\mathcal{W}\cap\mathcal{W}_{c_0}} f_j(x)dx, \quad j = x, w,$$

where X_i and W_i are independent. A later version of this paper is supposed to present more details in form of a simulation study and in form of a theorem.

Note that Pinkse and Robinson and Härdle and Marron specify their loss functions in such a way that it takes into account N realizations and not only the observations in $\mathcal{W}\cap\mathcal{W}_c$.

5 Application

This section is devoted to an application of the HM 4 step estimator to consumer data. We mainly follow Blundell, Duncan and Pendakur (1998) who use an estimator of the Pinkse and Robinson specification in order to estimate unknown expenditure shares under shape invariance restrictions. It should therefore be of interest to investigate how the HM 4 step estimator behaves in comparison. We use the same cross section samples of the British Family Expenditure Survey (FES I) for this purpose. Afterwards the estimation is done for samples (FES II) which are also used in Blundell, Chen and Kristensen (2001).

Blundell, Duncan and Pendakur estimate expenditure shares for several commodities using an extended semiparametric specification as given in the model of Section 2. The parametric shifts are now related to observable household characteristics like the number of children in a household. Accordingly, they compare couples with one child to couples with two children. The expenditure shares for the two groups are linked by the following model:

$$m_1(x) = a + m_0(x - c),$$

where x is the log-expenditure of a household. Kernel density estimates of the samples are shown in Figure 10. Blundell, Duncan and Pendakur choose this specification because the commonly used partially linear model is ruled out by economic theory since in this case the unknown function m_0 has to be linear. For further details see Lemma 3.1 and Lemma 3.2

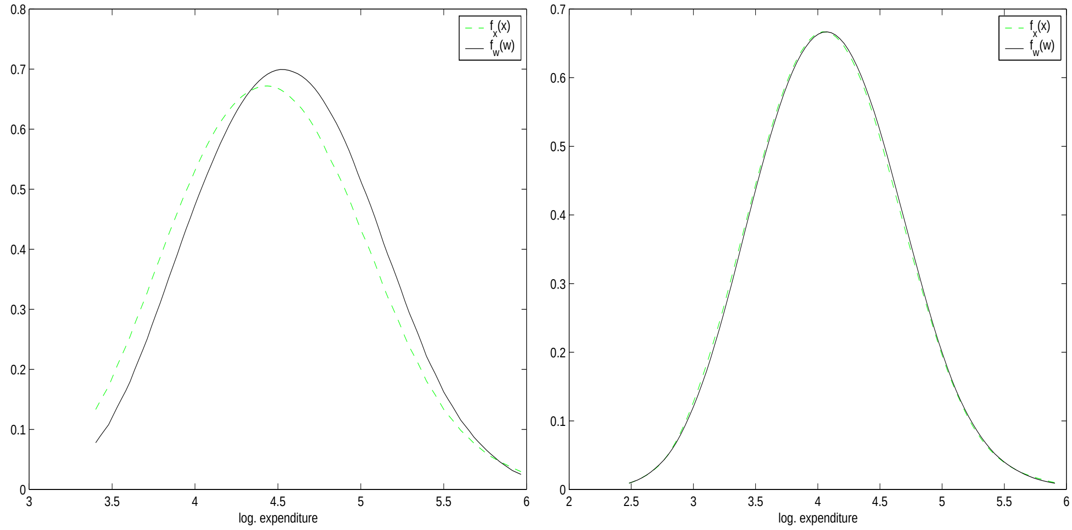

Figure 10: FES I: Kernel estimates of the marginal distribution of log expenditure a) FES I b) FES II

in their paper or Blundell, Browning and Crawford (1997).

Results for the FES I sample using the HM 4 step estimator are presented in Figures 16-21 in Appendix 2. For the nonparametric estimation we use a local linear smoother with either a constant or a variable bandwidth. The bandwidths are obtained with an iterative plug-in method as described for example in Fan and Gijbels (1995). At a glance, these Figures indicate that for most of the commodities this specification is appropriate. When looking at the corresponding loss functions this opinion has to be revised since in many cases the shape of the loss function indicate that the identification conditions for the parameters are not given. For example in the case of food, the hypothesis cannot be rejected that expenditure shares are linear. In this cas the parameter estimates are inconsistent, since the loss function does not possess a unique minimum. Similar reasoning applies for some of the other commodities.

Since the partially linear model is ruled out by economic theory, Blundell, Duncan and Pendakur consider a model under shape invariance restrictions, the so called Extended Partially Linear Model (EPLM), which is given by

$$m_1^{(j)}(x) = a^{(j)} + m_0^{(j)}(x - c) \text{ for } j = 1, \ldots, J,$$

where J is the number of equations (commodities). The loss function (4) becomes in this

	fixed bandwidth	variable bandwidth
expenditure shares	$\hat{a}^{(j)}$	
food	−0.0292	0.0776
fuel	−0.0176	0.0140
clothing	0.0209	−0.0293
alcohol	−0.0009	−0.0137
transport	0.0149	−0.0376
other goods	0.0125	−0.0162
$\hat{c}$	0.3926	−0.3402

Table 2: EPLM, FES I

case:

$$L_N(a,c) = \sum_{j=1}^{J} \int_{\mathcal{W}^{(j)} \cap \mathcal{W}_c^{(j)}} \frac{\left[\hat{m}_1^{(j)}(x) - a^{(j)} - \hat{m}_0^{(j)}(x-c)\right]^2 w(x)dx}{\int_{\mathcal{W}^{(j)} \cap \mathcal{W}_c^{(j)}} w(x)dx}$$

The horizontal shift is supposed to be the same for all commodities. This specification appears crucial for FES I data since $\hat{c}^{(j)}$ varies across the single equation estimates (Figures 16-21). Estimates of the EPLM confirm these doubts concerning the specification: $\hat{c}$ reacts sensitive on the choice of the bandwidth and the exclusion of irrelevant information (food expenditure share).

From Figures 11 and 12 it is apparent that the loss function tends to have two minima, one around $c = 0.5$ and the other around -0.4. The parameter c is the log of the so called equivalence scale. Negative values of c do not have a reasonable economic interpretation since this would imply $exp(c) < 1$. However, the global minimum is in most of the cases located at $\hat{c} < 0$. Parameter estimates for the EPLM are given in Table 2. In contrast to our findings, Blundell, Duncan and Pendakur obtain $\hat{c} = 0.259$ using the Pinkse and Robinson specification and restricting the space C to $[0, 1]$. As we have seen in Section 3, the finite sample performance of this specification is weaker and might end up in a larger bias of the parameter estimates. Our specification using the full system and using a fixed bandwidth ($\hat{c} = 0.3926$) is the closest to their specification. However, it uses the local linear smoother instead of the Nadaraya-Watson estimator.

FES II is a sample of the British Family Expenditure Survey as well. In comparison to FES I the composition is different: We now compare couples without children to couples

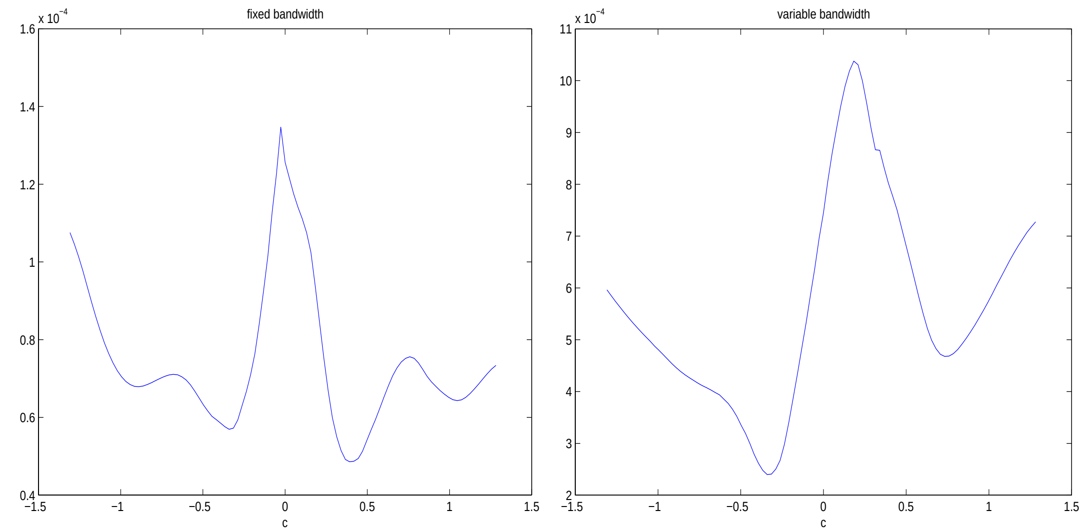

Figure 11: FES I: Loss function of EPLM

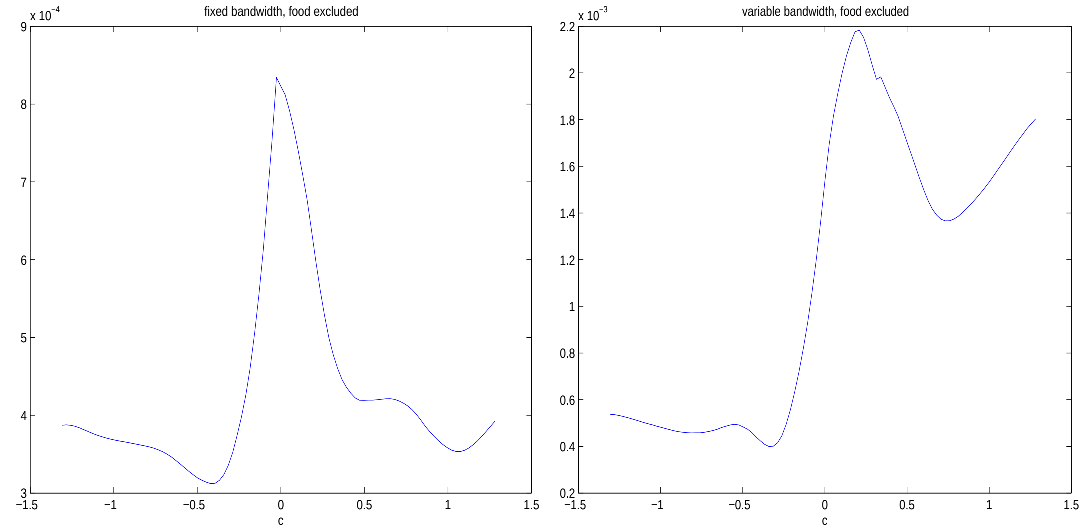

Figure 12: FES I: Loss function of EPLM, J=5 (food excluded).

	fixed bandwidth	variable bandwidth
expenditure shares	$\hat{a}^{(j)}$	
alcohol	-0.0200	-0.0178
catering	-0.0040	-0.0036
clothing	-0.0029	0.0067
food	-0.0065	-0.0191
personal goods and services	0.0027	0.0030
leisure goods	0.0137	0.0158
travel	-0.0065	-0.0122
$\hat{c}$	0.4606	0.5593

Table 3: EPLM, FES II

with one or two children. Estimates for the EPLM are presented in Figures 13 and 14. The corresponding loss functions behave smoothly and possess a unique minimum in the interior of C, see Figure 15. The model specification seems to be appropriate in this case. The horizontal shifts in Figures 13 and 14 seem to be reasonable and the parameter estimates (Table 3) have reasonable economic intuition. The estimated equivalence scale is positive.

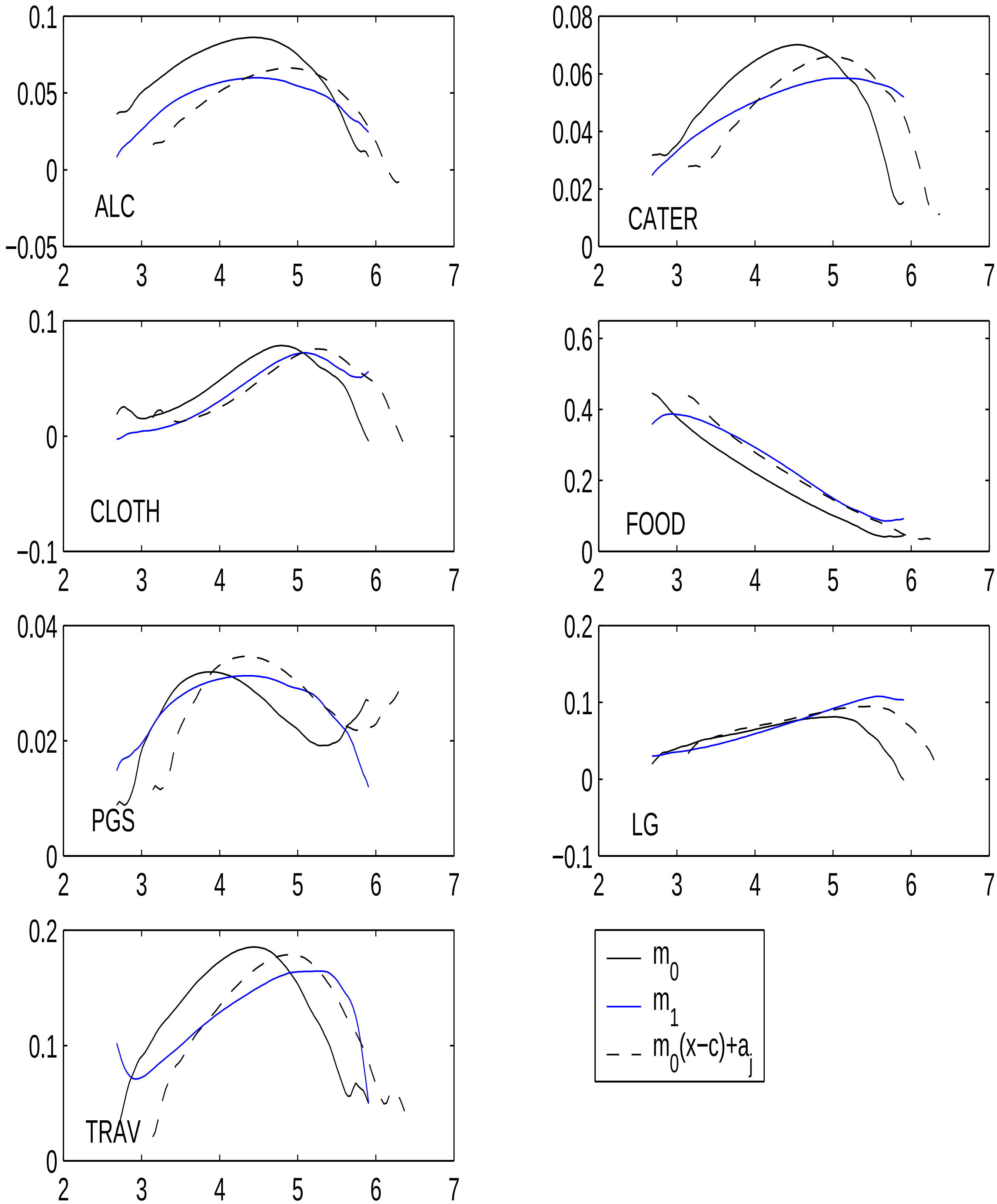

Figure 13: FES II, EPLM, fixed bandwidth

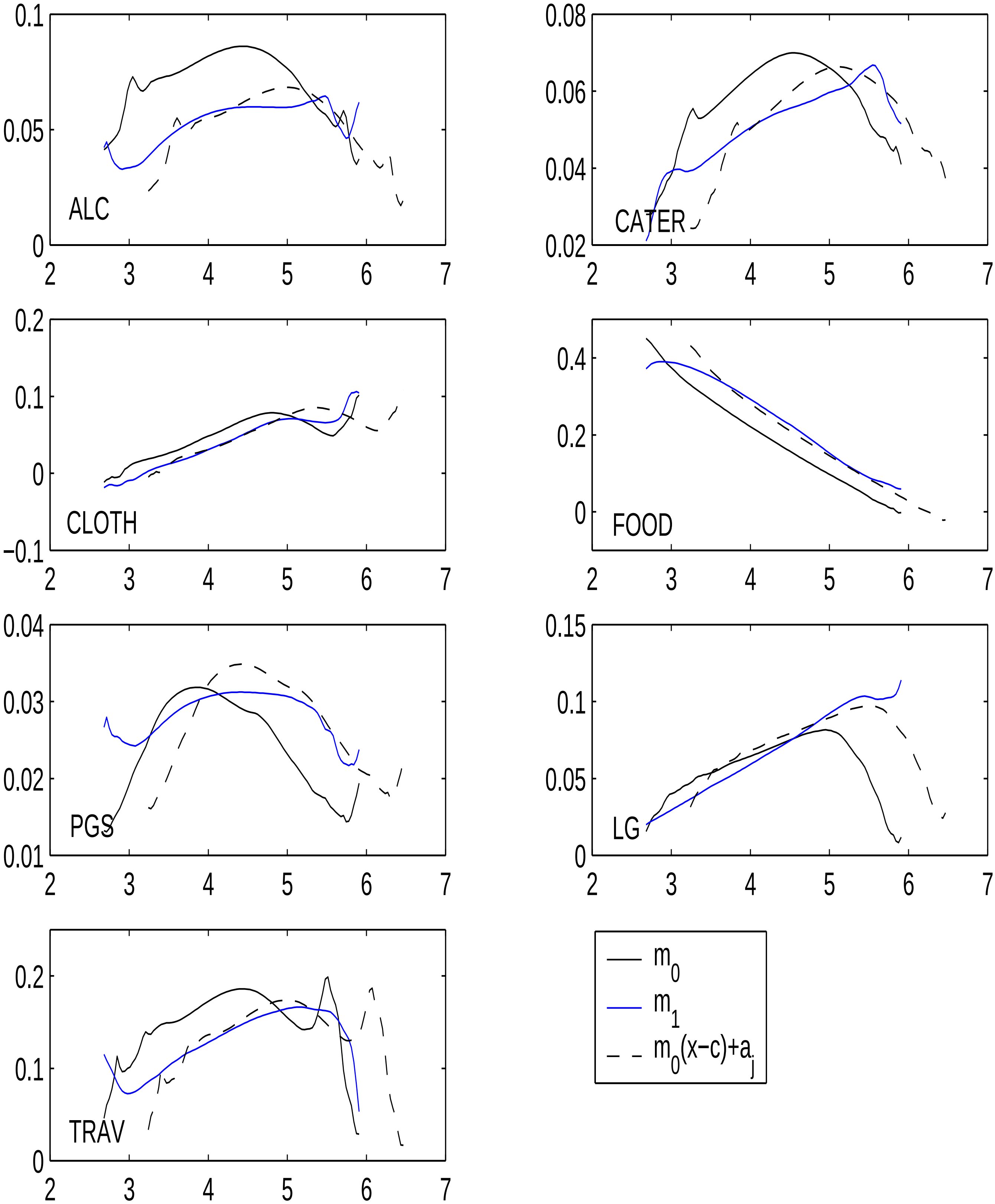

Figure 14: FES II, EPLM, variable bandwidth

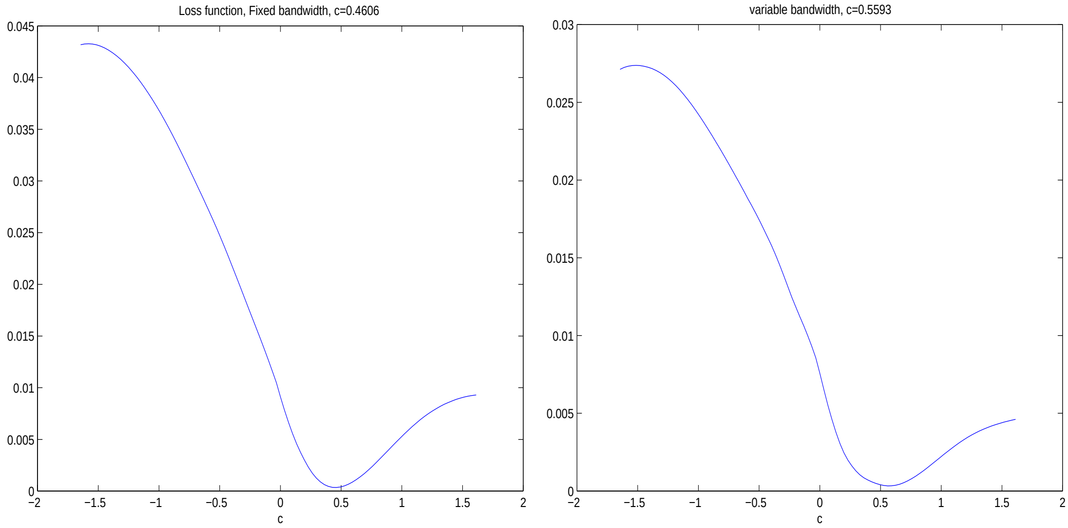

Figure 15: FES II, EPLM, Loss functions

Appendix 1: Proof of Theorem

Proof of Theorem 1 According to Theorem 4.3.1 in Amemiya (1985) we have to check:

1. The parameter space is an open subset in $\mathbb{R}^3$. The true value is an interior point of this set.

2. The objective function $L_N(\hat{m}_0, \hat{m}_1, a, b, c)$ is a measurable function of $\hat{m}_0$ and $\hat{m}_1$, continuous in a, b, c uniformly in N. The partial derivatives of L_N with respect to the parameters exist and are continuous in an open neighborhood of (a_0, b_0, c_0).

3. There exists an open neighborhood of (a_0, b_0, c_0) such that $L_N(a, b, c)$ converges to a nonstochastic function $L(a, b, c)$ in probability uniformly in (a, b, c).

4. $\text{plim} L_N(a, b, c)$=0 at (a_0, b_0, c_0) and greater than zero elsewhere.

The first and the second condition are clearly satisfied due to the model specification. The third and the fourth condition can be written as

$$3. \qquad \text{plim} L_N(a, b, c) = L(a, b, c)$$

and

$$4. \qquad \frac{\partial L(a, b, c)}{\partial (a, b, c)}\bigg|_{(a_0, b_0, c_0)} = 0.$$

3. Combining (6),(7) and (8) yields

$$\begin{aligned}
L_N(a,b,c) &= \sum_{t=1}^{T_c} T_c^{-1}\big[a_0 + b_0 m_0(x_t - c_0) + \epsilon_1(x_t - c_0, N) - a - b m_0(x_t - c) - b_0\epsilon_0(x_t - c, N)\big]^2 \\
&= \sum_{t=1}^{T_c} T_c^{-1}\big[a_0 + b_0 m_0(x_t - c_0) - a - b m_0(x_t - c)\big]^2 \\
&\quad + \sum_{t=1}^{T_c} T_c^{-1}\big[\epsilon_1(x_t - c_0, N) - b\epsilon_0(x_t - c, N)\big]^2 \\
&\quad + 2\sum_{t=1}^{T_c} T_c^{-1}\big[a_0 + b_0 m_0(x_t - c_0) - a - b m_0(x_t - c)\big]\big[\epsilon_1(x_t - c_0, N) - b(\epsilon_0(x_t - c, N))\big] \\
&= A_1 + A_2 + A_3
\end{aligned}$$

By the Slutsky Theorem it suffices to show that the plim of A_1, A_2 and A_3 respectively exists.

plim A_3 can be derived by using the fact that $\epsilon_0(x_t, N)$ and $\epsilon_1(x_t - c_0, N)$ converge to zero in probability uniformly:

$$\begin{aligned}
\text{plim}\ \sup_{b,c}\Big|T_c^{-1}\sum_t b\epsilon_0(x_t - c, N)\Big| &= \sup_b\Big|b\,\text{plim}T_c^{-1}\sum_t \epsilon_0(x_t, N)\Big| \\
&\leq \sup_b\Big|b\,\text{plim}\ \sup_{x_t\in\mathcal{X}_1}|\epsilon_0(x_t, N)|\Big| \\
&= 0
\end{aligned}$$

and using the fact that

$$\sup_{\substack{a,b,c \\ x\in\mathcal{X}_1}} \big|a_0 + b_0 m_0(x - c_0) - a - b m_0(x - c)\big| < \infty.$$

Hence plim $A_3 = 0$. Repeated application of the Slutsky Theorem to A_2 yields plim $A_2 = 0$. plim A_1 can be derived using the fact that X_i are i.i.d. and

$$\sup_{a_1,a_2,b_1,b_2,c_1,c_2}\big|E\big[a_1 + b_1 m_0(x - c_1)\big)\big(a_2 - b_2 m_0(x - c_2)\big]\big| < \infty.$$

for all $c_1, c_2 \in C$. Applying Theorem 4.2.1 and Theorem 3.2.6 of Amemiya (1985) yields

$$\text{plim}A_1 = \frac{E\Big[\big(a_0 + b_0 m_0(x - c_0) - a - b m_0(x - c)\big)^2|x - c \in \mathcal{W}\Big]}{\int_{\underline{x}(c)}^{\overline{x}(c)} f_x(x)dx}.$$

where the integration bounds are such that

$$\begin{aligned}
\int_{\underline{x}(c)}^{\overline{x}(c)} f_x(x)dx &= F(\overline{x}(c)) - F(\underline{x}(c)) \\
&= \text{Prob}(X_i \in \mathcal{X}_1 | X_i - c \in \mathcal{W}).
\end{aligned}$$

4. To be shown: The limit loss function plim$L_N(a,b,c)$ =plimA_1 has a unique minimum in a_0, b_0, c_0, i.e.

$$\text{plim}\frac{\partial L_N(a,b,c)}{\partial(a,b,c)}\bigg|_{(a_0,b_0,c_0)} = 0$$

We have to check the necessary and the sufficient conditions.

The first order conditions are:

$$\begin{aligned}
\partial_a \text{plim} A_1(a,b,c) &= -\frac{2E\big[a_0 - a + b_0 m_0(x-c_0) - b m_0(x-c)|x-c \in \mathcal{W}\big]}{\big[F(\overline{x}(c)) - F(\underline{x}(c))\big]} = 0 && (9)\\
\partial_b \text{plim} A_1(a,b,c) &= -\frac{2E\big[m_0(x-c)\big(a_0 - a + b_0 m_0(x-c_0) - b m_0(x-c)\big)|x-c \in \mathcal{W}\big]}{\big[F(\overline{x}(c)) - F(\underline{x}(c))\big]}\\
&= 0 && (10)\\
\partial_c \text{plim} A_1(a,b,c) &= \frac{2E\big[b m_0'(x-c)\big(a_0 - a + b_0 m_0(x-c_0) - b m_0(x-c)|x-c \in \mathcal{W}\big)\big]}{\big[F(\overline{x}(c)) - F(\underline{x}(c))\big]}\\
&\quad -\frac{E\big[\big(a_0 - a + b_0 m_0(x-c_0) - b m_0(x-c)\big)^2|x-c \in \mathcal{W}\big]}{\big[F(\overline{x}(c)) - F(\underline{x}(c))\big]^2}\\
&\quad \times\big[\overline{x}'(c) f_x(\overline{x}(c)) - \underline{x}'(c) f_x(\underline{x}(c))\big] && (11)\\
&= 0
\end{aligned}$$

From (9) and (10) we obtain

$$\begin{aligned}
\hat{a} &= a_0 + E\big[b_0 m_0(x-c_0) - b m_0(x-c)|x-c \in \mathcal{W}\big] && (12)\\
\hat{b} &= \frac{E\big[m_0(x-c)\big(a_0 - a + b_0 m_0(x-c_0)\big)|x-c \in \mathcal{W}\big]}{E\big[m_0(x-c)^2|x-c \in \mathcal{W}\big]}
\end{aligned}$$

Substituting for a yields:

$$\hat{b} = \frac{cov\big(b_0 m_0(x-c_0), m_0(x-c)|x-c \in \mathcal{W}\big)}{var\,(m_0(x-c)|x-c \in \mathcal{W})} \qquad (13)$$

The condition given by equation (11) is stronger than required. We need to show that the loss function is zero at the true parameter values and greater than zero elsewhere. We know that the denominator is greater than zero and less or equal to one. It is therefore enough to show that the nominator of the loss function is only zero at the true parameter values. We can therefore substitute (11) by

$$2E\big[b m_0'(x-c)\big(a_0 - a + b_0 m_0(x-c_0) - b m_0(x-c)\big)|x-c \in \mathcal{W}\big] = 0 \qquad (14)$$

Using (12) and (13) to substitute for a and b in (14) yields

$$\begin{aligned}
0 &= E\Bigg[\frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})} m_0'(x-c) \\
&\quad \times\Bigg(\frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})} E\left[m_0(x-c)|x-c\in\mathcal{W}\right] \\
&\quad -b_0 E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right] + b_0 m_0(x-c_0) \\
&\quad\quad -\frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})} m_0(x-c)\Bigg)|x-c\in\mathcal{W}\Bigg] \\
&= \frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})} \\
&\quad \times\Bigg(b_0 E\left[m_0'(x-c) m_0(x-c_0)|x-c\in\mathcal{W}\right] \\
&\quad -b_0 E\left[m_0'(x-c)|x-c\in\mathcal{W}\right] E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right] \\
&\quad +\frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})} \\
&\quad\quad \times\big(E\left[m_0'(x-c)|x-c\in\mathcal{W}\right] E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right] \\
&\quad\quad -E\left[m_0'(x-c) m_0(x-c)|x-c\in\mathcal{W}\right]\big)\Bigg) \\
&= \frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})}\Bigg(b_0 cov\,(m_0'(x-c), m_0(x-c_0)|x-c\in\mathcal{W}) \\
&\quad -\frac{cov\,(b_0 m_0(x-c_0), m_0(x-c)|x-c\in\mathcal{W})}{var\,(m_0(x-c)|x-c\in\mathcal{W})} cov\,(m_0'(x-c), m_0(x-c)|x-c\in\mathcal{W})\Bigg)
\end{aligned}$$

Assumption 3 ensures that

$$\begin{aligned}
cov\,(m_0'(x-c), m_0(x-c)|x-c\in\mathcal{W}) &\neq 0 \quad \text{and} \\
cov\,(m_0'(x-c), m_0(x-c_0)|x-c\in\mathcal{W}) &\neq 0 \quad \text{for all } c\in C.
\end{aligned}$$

Assumptions 2 ensures that the equality only holds at $c = c_0$.

For the sufficient conditions we need to analyze the second order conditions. Denote

$$H_{11} = 1/2\partial_a^2|_{a=a_0} E\left[(a_0 + b_0 m_0(x-c_0) - a - b m_0(x-c))^2\,|x-c\in\mathcal{W}\right]$$

and H_{kl} accordingly. It is easy to show that the Hessian $\mathbf{H}$ is symmetric at (a_0, b_0, c_0). The sufficient conditions for having a minimum of the nominator of the loss function are:

1. H_{11}, H_{22} and $H_{33} > 0$
2. $H_{11}H_{22} - H_{12}^2 > 0$

3. $\det\mathbf{H} > 0$

The elements of the Hessian are:

$$\begin{aligned}
H_{11} &= 1 \\
H_{22} &= E\left[m_0(x-c_0)^2|x-c\in\mathcal{W}\right] \\
H_{33} &= b_0^2 E\left[m_0'(x-c_0)^2|x-c\in\mathcal{W}\right] \\
H_{12} &= E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right] \\
H_{13} &= -b_0 E\left[m_0'(x-c_0)|x-c\in\mathcal{W}\right] \\
H_{23} &= -b_0 E\left[m_0'(x-c_0)m_0(x-c_0)|x-c\in\mathcal{W}\right]
\end{aligned}$$

Condition 1 is clearly satisfied. It is to be shown that the other two conditions also hold.

Condition 2 holds, since

$$E\left[m_0(x-c_0)^2|x-c\in\mathcal{W}\right] > \left(E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right]\right)^2$$

due to the Cauchy-Schwartz inequality.

Condition 3 requires

$$\begin{aligned}
0 < {} & E\left[m_0(x-c_0)^2|x-c\in\mathcal{W}\right] b_0^2 E\left[m_0'(x-c_0)^2|x-c\in\mathcal{W}\right] \\
& +b_0^2\left(E\left[m_0'(x-c_0)m_0(x-c_0)|x-c\in\mathcal{W}\right]\right)^2 \\
& +E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right] b_0^2 E\left[m_0'(x-c_0)|x-c\in\mathcal{W}\right] E\left[m_0'(x-c_0)m_0(x-c_0)|x-c\in\mathcal{W}\right] \\
& +E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right] b_0^2 E\left[m_0'(x-c_0)|x-c\in\mathcal{W}\right] E\left[m_0'(x-c_0)m_0(x-c_0)|x-c\in\mathcal{W}\right] \\
& -\left(E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right]\right)^2 b_0^2 E\left[m_0'(x-c_0)^2|x-c\in\mathcal{W}\right] \\
& -E\left[m_0(x-c_0)^2|x-c\in\mathcal{W}\right] b_0^2\left(E\left[m_0'(x-c_0)|x-c\in\mathcal{W}\right]\right)^2
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
& 2\left(E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right]\right)^2\left(E\left[m_0'(x-c_0)|x-c\in\mathcal{W}\right]\right)^2 \\
& +E\left[m_0(x-c_0)^2|x-c\in\mathcal{W}\right] E\left[m_0'(x-c_0)^2|x-c\in\mathcal{W}\right] \\
& +\left(E\left[m_0'(x-c_0)m_0(x-c_0)|x-c\in\mathcal{W}\right]\right)^2 \\
> {} & \left(E\left[m_0(x-c_0)|x-c\in\mathcal{W}\right]\right)^2 E\left[m_0'(x-c_0)^2|x-c\in\mathcal{W}\right] \\
& +\left(E\left[m_0'(x-c_0)|x-c\in\mathcal{W}\right]\right)^2 E\left[m_0(x-c_0)^2|x-c\in\mathcal{W}\right].
\end{aligned}$$

The inequality can be shown by an application of the Cauchy- Schwarz inequality to the second and the third term of the left hand side. ■

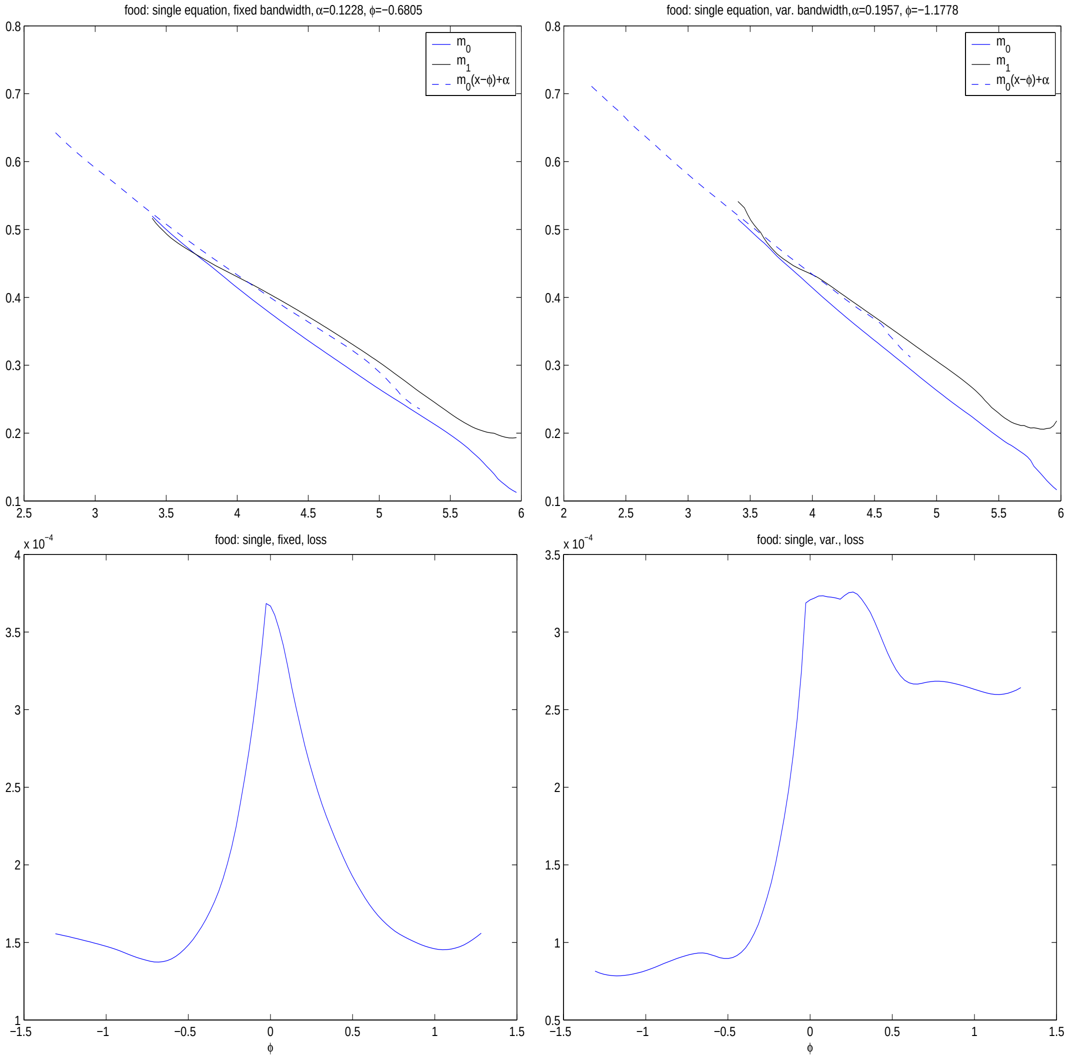

Figure 16: FES I: Food expenditure share.

Appendix 2: Estimation results

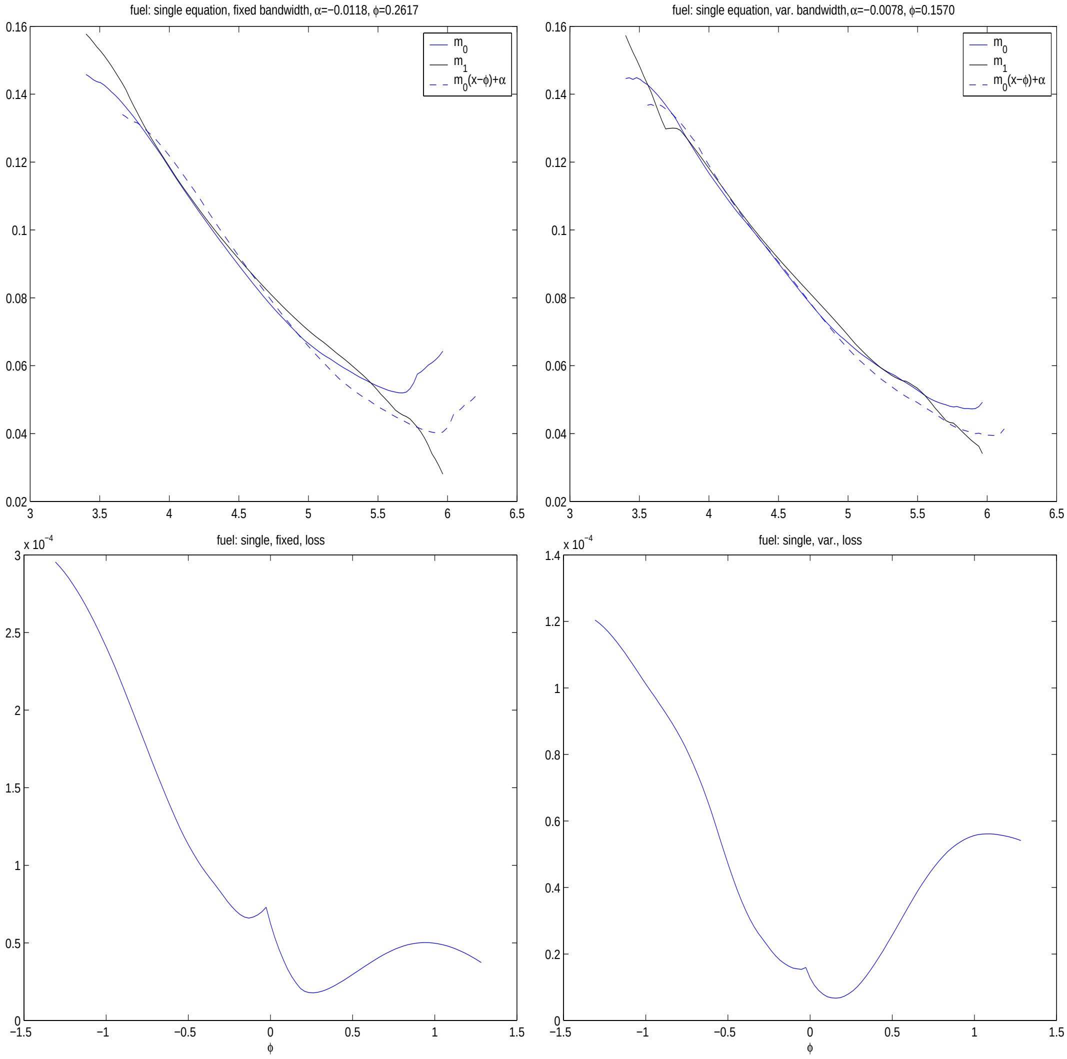

Figure 17: FES I: Fuel expenditure share.

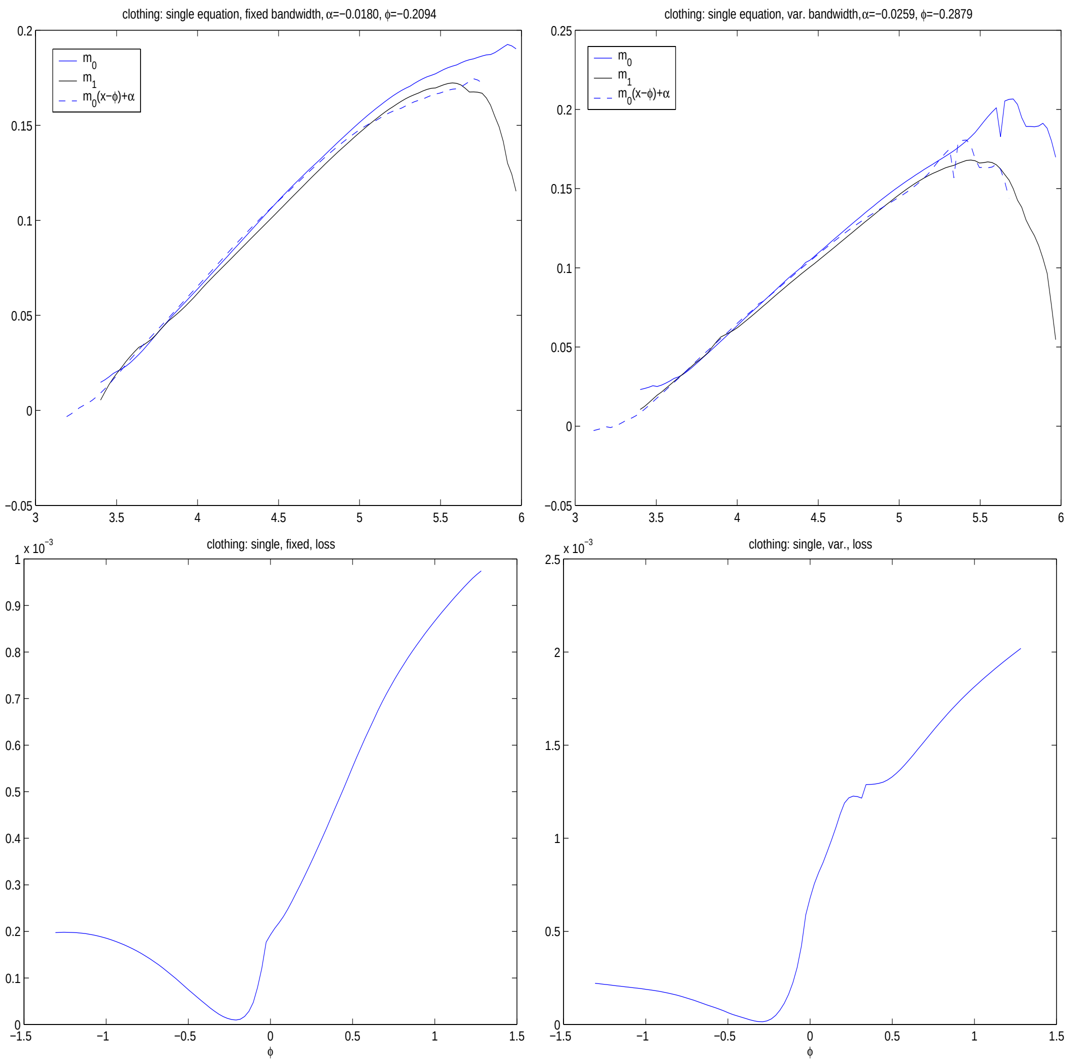

Figure 18: FES I: Clothing expenditure share.

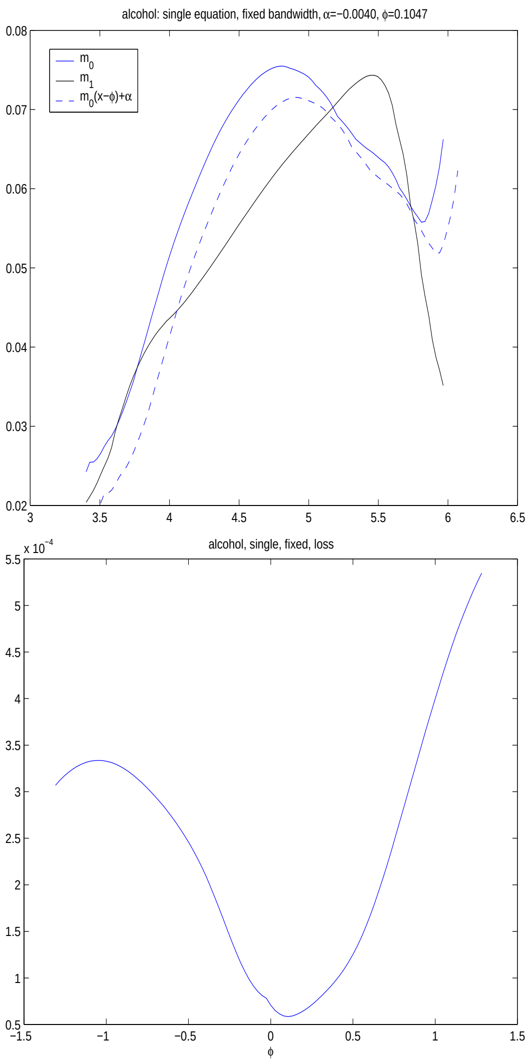

Figure 19: FES I: Alcohol expenditure share.

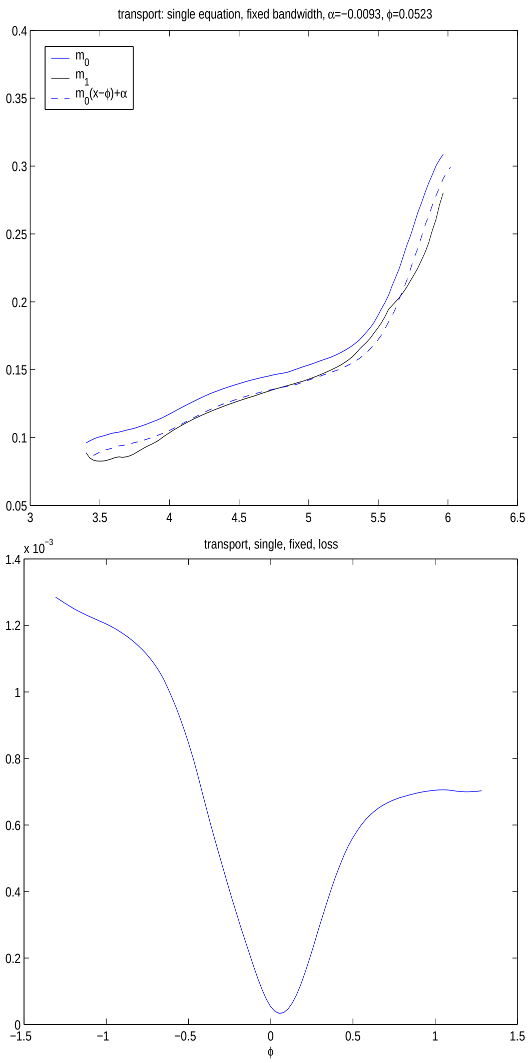

Figure 20: FES I: Transport expenditure share.

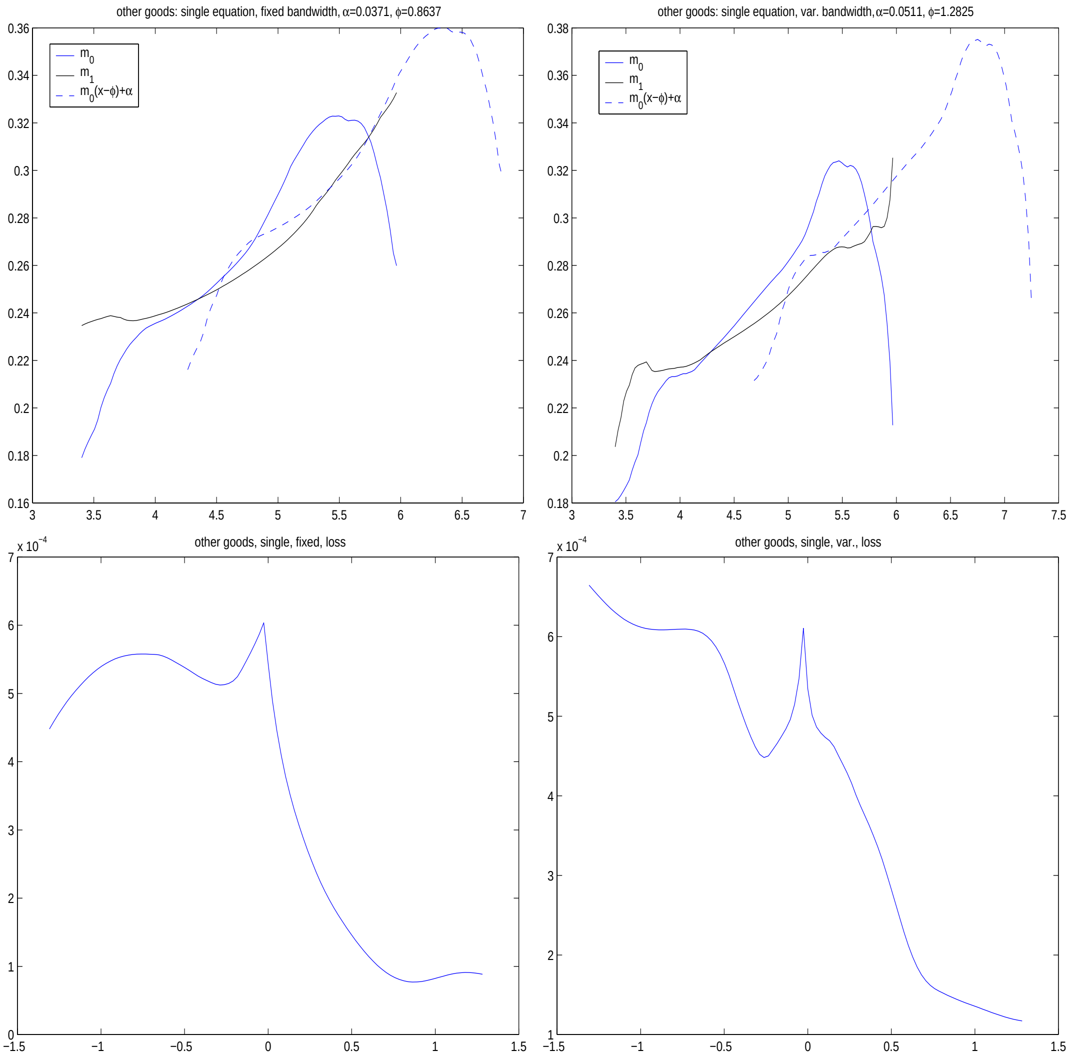

Figure 21: FES I: Other goods expenditure share.

References

[1] Amemiya, T. (1985). Advanced Econometrics. Oxford: Basil Blackwell.

[2] Blundell, R., Browning, M. and Crawford, I. (1997). Nonparametric Engel Curves and Revealed Preferences. *IFS Working Paper* W97/14.

[3] Blundell, R., Chen, X. and Kristensen, D. (2001). Semiparametric Shape Invariant Engel Curves with Endogeneous Expenditure. *Mimeo* University College London.

[4] Blundell, R., Duncan, A. and Penkadur (1998). Semiparametric Estimation and Consumer Demand. *Journal of Applied Econometrics*, Vol.13, No.5,pp. 1-30

[5] Fan,J. and Gijbels, I. (1995). Local Polynomial Modelling and Its Applications. London: Chapman and Hall.

[6] Härdle, W. and Marron, J.S. (1990). Semiparametric Comparison of Regression Functions. *Annals of Statistics*, 18, 63-89.

[7] Nadaraya, E. (1989). Nonparametric Estimation of Probability Densities and Regression Curves. *Mathematics and Its Applications.* Soviet Series, translated by S.Kotz Dordrecht: Kluwer Academic Publishers

[8] Pinkse, C. and Robinson, P.M. (1995). Pooling Nonparametric Estimates of Regression Functions with a Similar Shape. *Advances in Econometrics and Quantitative Economics, eds. G.Maddala, P.Phillips and T.N.Srinvinsan*, pp. 172-195.

[9] Wilke, R. (2000). Semiparametric Estimation of Consumer Demand. *Discussion Papers in Economics No. 2000-07*, University of Dortmund.