Discussion Paper No. 03-64

Semiparametric Estimation of Regression Functions Under Shape Invariance Restrictions

Ralf A. Wilke

ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

Centre for European Economic Research

Discussion Paper No. 03-64

Semiparametric Estimation of Regression Functions Under Shape Invariance Restrictions

Ralf A. Wilke

Download this ZEW Discussion Paper from our ftp server:

ftp://ftp.zew.de/pub/zew-docs/dp/dp0364.pdf

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non technical summary This paper considers the shape invariant semiparametric regression model in which nonparametric functions of similar shape are linked by parametric transformations with unknown parameters. Existing contributions do not provide clear guidance to the applied researcher in terms of model identification, numerical implementation and finite sample performance of the estimators. This paper extensively discusses the identification issue. Furthermore, it introduces a new estimator which has desirable theoretical and practical properties: i) $\sqrt{N}$- consistency of the parameter estimates is proved ii) the suggested implementation is computationally convenient iii) its finite sample performance is superior to former specifications. iv) a small application to British consumer data illustrates the importance of this method for applied statistics. The estimation results indicate that the imposed shape invariance restrictions have empirical evidence in the semiparametric modelling of consumer demand. This kind of models can be used for the consistent estimation of equivalence scales for social security systems.

Semiparametric Estimation of Regression Functions Under Shape Invariance Restrictions.

Ralf A. Wilke*

ZEW Mannheim, L7.1, 68161 Mannheim, Germany

October 2003

Abstract

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

Keywords: shape invariant modelling, semiparametric regression, simulations, large sample properties, consumer data

JEL: C14, C31, D12

*E-mail address: wilke@zew.de

1 Introduction

Semiparametric estimation 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, where unknown functionals are linked by parametric transformations. The general estimation frameworks of Härdle and Marron (1990) and Pinkse and Robinson (1995) provide conditions which ensure the identifiability of the parameters. However, these conditions are difficult to scrutinize in applications. Moreover, for particular model specifications it is not evident what particular identification conditions are required. This paper considers a simplified model which is of interest for the estimation of consumption based equivalence scales. We identify the difficulties which might occur in applied analysis. Necessary and sufficient conditions are derived in order to ensure identifiability of the parameters. An new estimator estimator is suggested. Consistency and asymptotic normality of the parameter estimates are proved. The new estimator requires less computational effort, it is convenient to implement and has better finite sample properties. Simulations compare different specifications of the estimator for finite samples. A illustrative application to consumption based equivalence scales estimation demonstrates the importance of this method for applied research.

Let us briefly state some motivation for the application of shape invariant models by considering the semiparametric estimation of equivalence scales. Blundell, Duncan and Pendakur (1998) investigate expenditure shares of couples with one child that are supposed to be related by horizontal and vertical shift to the expenditure shares of couples with two children. The horizontal shift is directly related to the equivalence scale between the two groups of households. Equivalence scales are of major interest for welfare policies because they provide information about how to equalize the income between demographic groups of households. Figure 1 presents the nonparametric estimates of the transport expenditure shares for the two groups using the same household data from the British Family Expenditure Survey as in Blundell, Duncan and Pendakur (1998). It is apparent that the two functions are similar in shape and that they might be related by constant shift. The econometrician wants to identify the unknown functions as accurately as possible and wants to know the true values of the parameters.

More generally, suppose there is a finite number of samples with unknown regression functions. These regression functions are assumed to be similar in shape. In fact they are linked by known transformation functions with unknown parameters. There are two aims for the researcher in this approach: first, the identification of the parameters and second, the exact pooling of the data. 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 unknown regression function. This paper focuses on the first point. The second is already subject to deep analysis in Pinkse and Robinson (1995).

Härdle and Marron (1990) suggest a general framework for nonstochastic regressors, whereby the identification problem is not convincingly solved: instead of deriving precise regularity conditions on the model, they simply assume a nicely shaped objective function. The same criticism applies to Pinkse and Robinson (1995). They consider the case of independent stochastic regressors and the case of limited dependency between the stochastic regressors of the samples. $\sqrt{N}$ consistency of the parameter estimates is proved in both of the papers. The applied researcher cannot find in none of the papers an example with a full set of identification conditions, not even for a simple model specification. The estimators of the two papers are related but not the same. Simulations in this paper indicate that the specification of the estimator affects the finite sample performance. Moreover, it seems that the Pinkse and Robinson specification performs weaker in small samples.

Two general difficulties for the shape invariant modelling approach are identified in this paper: 1. Precise identification conditions are derived for a specific parametric transformation function. The purpose of this paper is to tackle these problems such that this class of estimators can become more popular in applied research. 2. Because of the complex structure of the estimation problem, the researcher has to carefully select an appropriate algorithm in order to avoid exploding computational effort.

The paper is organized as follows: Section 2 presents the model, informally introduces the necessary and the sufficient conditions for identification, suggests a new 4 step estimation procedure and provides an intuitive discussion of the above mentioned difficulties. Section 3 considers the large sample properties of the estimator. The consistency proof applies to a wide range of models and nonparametric estimators. Asymptotic normality of the parameter estimates is shown for the Nadaraya-Watson estimator as nonparametric regression estimator. Section 4 presents results of Monte Carlo experiments in order to investigate the finite sample behavior. Section 5 presents an illustrative application to consumer data.

2 The Model

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

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

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 fourth moments and the pairs (U_i, V_i) are mutually independent. $X_i \in \mathcal{X}_1$ and $W_i \in \mathcal{W}$ are i.i.d. random variables with realizations on compact sets with twice continuously differentiable densities f_x and f_w. Let the densities satisfy $\inf_{x\in\mathcal{X}_1} f_x(x) > 0$ 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 their 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:

$$m_1(x) = a_0 + b_0 m_0(x - c_0), \tag{1}$$

where $c_0 \in \mathcal{C} \subset \mathbb{R}$, $W_i + c_0 \in \mathcal{X}_2$ and $(X_i - c) \in \mathcal{W}_c$. In other words there exist horizontal and vertical translations with unknown parameters between the unknown functions m_0 and m_1. This model setup is similar to one of the models defined in Härdle and Marron (1990) and Pinkse and Robinson (1995) but here we restrict the model to the case of linear and constant translations of m_0 and x respectively. This is because the objective of this paper is to point out and to solve the main difficulties for the identification and estimation for this specific class of models. The example of the introduction applies to the considered class of translations. We provide directly applicable solutions for applied researchers. The spirit of these solutions would carry over to more general frameworks but more technicalities would cause a loss of intuition. Let us denote $\hat{m}_1(x)$ and $\hat{m}_c(x) = \hat{m}_0(x - c)$ the nonparametric estimates of $m_1(x)$ and $m_c(x) = m_0(x - c)$ respectively.

The Objective Function In order to estimate the unknown parameters (a_0, b_0, c_0) we now suggest minimizing the loss function

$$\begin{aligned} L_N(a, b, c) &= \frac{\int \mathbb{I}_{\{x\in\mathcal{W}\cap\mathcal{W}_c\}} \big[m_1(x) - a - bm_c(x)\big]^2 w(x)dx}{\int \mathbb{I}_{\{x\in\mathcal{W}\cap\mathcal{W}_c\}} f_x(x)w(x)dx} \\ &= \frac{\int_{\mathcal{W}\cap\mathcal{W}_c} \big[m_1(x) - a - bm_c(x)\big]^2 w(x)dx}{\int_{\mathcal{W}\cap\mathcal{W}_c} f_x(x)w(x)dx} \end{aligned} \tag{2}$$

in (a, b, c). The function w is a nonnegative weight function which may help in improving the efficiency of the estimation. The loss function L is constructed such that it minimizes the distance between m_0 and m_1 over the space of possible translations of m_0 for any point of $x \in \mathcal{W} \cap \mathcal{W}_c$. This set is compact. Restricting the integration bounds is necessary in order to ensure that we only consider points where we may obtain consistent estimates of m_c and m_1. The denominator is required to ensure that the objective function does not become close to zero as the intersection of $\mathcal{W} \cap \mathcal{W}_c$ becomes small. This is not required for identification but it improves the finite sample properties of the estimations.

Let us now outline the difficulties that are involved in identifying and computing the true parameters:

- **Identification:** If the supports of X_i and $W_i + c_0$ are disjoint, i.e. $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$,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. Furthermore, the unknown function m_0 has to follow some shape restrictions (non linear and non cycling) otherwise the parameters cannot be identified.

- **Computation:** 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. This paper introduces a convenient 4 step estimation procedure.

Identification We need to distinguish between two issues: restrictions on the space of the unknown parameter c and restrictions on the shape of the unknown function m_0.

We require some restrictions on the parameter space in order to ensure that the nonparametric estimates of 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.*

In applications 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 2: $\mathcal{X}_1 \cap \mathcal{X}_2 = [5, 12]$. Accordingly, $\mathcal{W} \cap \mathcal{W}_{c_0} = [0, 7]$. If $|c_0| \geq 12$, the functions are observed on different supports.

The identification of the unknown parameters in model (2) is still not yet ensured. The loss function under the above conditions might not have a unique global minimum at the true parameter values. We now introduce intuitively the identification restriction for the shape of m_0 which are formally required in the consistency proof. In particular, we have to impose some shape restrictions for the unknown function m_0. These conditions are violated if:

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 (3) and (4) 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.*

The second difficulty implies that (3) and (4) 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. Nevertheless, pooling of the data would still be possible. Pooling estimates of the unknown function can improve the accuracy of the nonparametric estimate of m_0.

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

Figure 3 presents an example using a cycling sine function. In this case there are three minima of the loss function on C.

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 if the samples are large.*

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. If we would allow for more general transformation functions $T_c(x)$ this issue issue would become even more problematic because the more flexible the transformation, the more likely the shape restrictions are violated, in particular if $\mathcal{W} \cap \mathcal{W}_c$ is small. Thus, in applications the loss function might be systematically minimized if $\mathcal{W} \cap \mathcal{W}_c$ is small. However, even if the parameters are identifiable, the variance of the estimates would become large since many of the observations cannot be used for the estimation. This is formally shown in section 3 when deriving the limit distribution of the parameter estimates.

Computation 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 (5). A four step estimator is defined for this purpose:

1. Estimate m_0 and m_1 on their support using a nonparametric estimator.

2. The least squares estimator for a and b, given a c is defined as

$$\min_{a,b} \int_{\mathcal{W} \cap \mathcal{W}_c} \left(\hat{m}_1(x) - a - b\hat{m}_c(x)\right)^2 w(x)dx,$$

   whereby it reduces to OLS in applications if $w(x) = 1$. More efficient estimates may be obtained by using information about var$\hat{m}_1(x)$ and var$\hat{m}_c(x)$ for the construction of $w(x)$.

3. Estimate c by minimizing the conditional loss function

$$L_N(c) \quad = \quad \frac{\int_{\mathcal{W} \cap \mathcal{W}_c} \left[\hat{m}_1(x) - \hat{a}_c - \hat{b}_c\hat{m}_c(x)\right]^2 w(x)dx}{\int_{\mathcal{W} \cap \mathcal{W}_c} \hat{f}_x(x)w(x)dx} \quad \text{(HM)} \qquad (3)$$

   where the integral is now restricted to the intersection of $\mathcal{W}$ and $\mathcal{W}_c$ since this is the set 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. The denominator improves the finite sample performance and does not affect asymptotic properties.

4. $\hat{a} = \hat{a}_{\hat{c}}$ and $\hat{b} = \hat{b}_{\hat{c}}$.

This estimator is to be referred to as the HM 4 step estimator. Instead of minimizing (3) one could also use the following specification for the third step:

$$L_N(c) = \frac{\int_{\mathcal{W} \cap \mathcal{W}_c} \left[\hat{f}(x)\hat{r}_c(x) - \hat{a}_c\hat{f}(x)\hat{f}_c(x) - \hat{b}_c\hat{f}_c(x)\hat{r}(x)\right]^2 w(x)dx}{\int_{\mathcal{W} \cap \mathcal{W}_c} \hat{f}_x(x)w(x)dx} \quad \text{(PR)}, \qquad (4)$$

where we assume that the nonparametric regression estimators may be written as fractions:

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

This specification is to be referred to as the PR 4 step estimation, because it is similar to the specification of the Pinkse and Robinson (1995) estimator.

By breaking up the minimization of the loss function into two parts, the numerical problem reduces to one dimension with the following advantages:

- Minimization with respect to a and b on a unbounded parameter space with low computational effort, e.g. least squares.
- 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 and not for every grid search step for c.

Therefore, this formulation of the estimation procedure induces feasible computational effort. Before deriving the large sample properties of this estimator let us briefly compare it to the estimators given by Härdle and Marron (1990) and Pinkse and Robinson (1995).

Pinkse and Robinson (1995) Their loss function is given by

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

which has to be minimized with respect to the parameters, where $w(x)$ is again a weight function. In fact this specification of the loss function has some weaknesses: first, the loss function is not minimized at the true parameter values, whenever there does not exist x such that $\hat{f}(x) > 0$ and $\hat{f}_{c_0} > 0$. Therefore, Pinkse and Robinson need to impose conditions on the model which may imply in applications that one of the samples of x and w need to have a support ranging from $-\infty$ to ∞. Their identification conditions are difficult to scrutinize if we leave the world of econometric theory and turn to specific applications. Second, by leaving out the denominator of equation (4), the loss function tends to zero whenever $\mathcal{W} \cap \mathcal{W}_c$ is small. For the large sample sample properties, i.e. $N \to \infty$, this does not matter but in finite samples the minimum of their loss function is naturally attained for values of c such that $\mathcal{W} \cap \mathcal{W}_c$ is small. Third, due to the multiplicative writing of $\hat{r}$, $\hat{r}_c$, $\hat{f}$ and $\hat{f}_c$ (instead of $\hat{r}/\hat{f}$ etc.), the finite sample properties for the parameter estimates using this specification seem to be weaker instead of using the fractions $\hat{r}/\hat{f}$ and $\hat{r}_c/\hat{f}_c$. See Section 4 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

$$L_N(a, b, c) = \int \left[\hat{m}_1(x) - a - b\hat{m}_c(x)\right]^2 w(x)dx \tag{5}$$

with respect to a, b and c, where Härdle and Marron suggest that $w(x)$ is a known nonnegative weight function which is zero whenever one of the densities $\hat{f}$ and $\hat{f}_c$ is zero. This avoids that the loss function is always zero if $\mathcal{W} \cap \mathcal{W}_c$ is empty. By leaving out the denominator of equation (4), the same problems occur in finite samples as in the Pinkse and Robinson (1995) framework.

3 Asymptotic Properties

This section presents asymptotic properties for the estimator defined in Section 2. For this purpose we use a modified Härdle and Marron loss function as given in (3) that incorporates the intuitive findings of Section 2 concerning identification. Consistency of this estimator is shown. This result is not restricted to the case of regression functions. It holds for a wide range of nonparametric functionals, e.g. density functions, and nonparametric estimators, e.g. local polynomial smoothers. Asymptotic normality of this estimator is shown for the case of regression functions when using the Nadaraya-Watson estimator as nonparametric regression estimator. However, it might be that the limit distribution is not affected if we use a similar estimator like a local linear smoother.

Consistency Härdle and Marron (1990) very generally assume that the loss function is convex around the true parameter values. In the light of section 2 we derive here the necessary and sufficient conditions on the shape of the unknown function m_0 and the parameter space $c \in \mathcal{C}$ such that the loss function has indeed a unique minimum.

Denote $\hat{m}(x_t)$ as the nonparametric estimate of m_0 evaluated at $X_i = x_t$. Define $T_c = \text{card}\{x_t - c | x_t - c \in \mathcal{W}\}_{i=1,\dots,N}$. Accordingly, we have $x_t - c \in \mathcal{W}_c$ for all $c \in C$. Let $\{x_t - c\}_{i=1,\dots,T_c} = \{X_i - c | X_i - c \in \mathcal{W}\}_{i=1,\dots,N}$ for all $c \in C$. Note that $T_c \leq N$. Note also that T_c weakly increases in N.

Assumption 1 *c_0 is an interior point of $\mathcal{C}$, where $\mathcal{C}$ is such that $\mathcal{W} \cap \mathcal{W}_c$ is non empty for all c.*

In terms of the empirical model this implies that $\mathcal{C}$ has to be such that for all $c \in \mathcal{C}$: $T_c \geq 3$. $T_c > 0$ 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 $\{\hat{m}_1(w_t^c)\}_{t=1,\ldots,T_c} = \{\hat{m}_1(x_t)\}_{t=1,\ldots,T_{c_0}}$ as the nonparametric estimates of m_1 evaluated at w_t^c. Moreover, denote $\hat{m}_0(x_t - c)$ as the nonparametric estimate of m_0 evaluated at x_t and horizontally shifted to $x_t - c$ for all x and c. The loss function (3) can then be rewritten as:

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

Intuitively, the loss per evaluation point is minimized. Note that this function depends on N due to the nonparametric estimates and T_c.

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 identifiability of the unknown parameters.

The nonparametric estimates for m_0 and m_1 may be written as

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

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

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.*

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

This assumption can be justified for the class of Kernel estimators by Theorem 2.1 of Nadaraya (1989). Let us now state the theorem about consistency of the parameter estimates:

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

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

where $\hat{\mathcal{B}}$ is the set of roots.

Proof: see 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. This paragraph derives the limit distribution for the parameter estimates of model (3). The result only applies to regression functions as nonparametric functions and the Nadaraya-Watson estimator as nonparametric estimator. It is shown that the denominator does not affect the finite sample properties.

Assumption 5 *The nonparametric regression estimator is the Nadaraya-Watson estimator.*

This assumption looks rather specific. It is done in for convenience in the proof of the theorem below. Intuitively, the results should not change when using a related nonparametric estimator like a local linear smoother. This is because the crucial requirement that the nonparametric estimates converge uniformly at a fast enough rate can also be achieved under similar conditions. Newey (1994) derives results for the limit distribution of semiparametric estimators which endorse this suspicion.

Assumption 6 *The Kernel function $K(x)$ is differentiable of order 2, the derivatives of order 2 are bounded, $K(x)$ is zero outside a bounded set, $\int K(x)dx = 1$ and for $j = 1, \ldots, 4$, $\int K(x)[x^j]dx = 0$.*

Higher order kernels are used for bias reduction of the regression function estimates and their derivatives.

Assumption 7 *$E[|y|^4] < \infty$, $E[|z|^4] < \infty$, $E[|y|^4|x]f_0(x)$ and $E[|z|^4|w]f_w(w)$ are bounded.*

Assumptions 6 and 7 replace Assumption 4. In contrast to Assumption 4 they are not only required for uniform convergence of the nonparametric estimators of the regression function and its first derivative. They also ensure that the convergence rate of $\hat{m}_0$ and $\hat{m}'_0$ is at least $N^{1/4}$ and that $\sup_{x \in \mathcal{W}_c}|m''_0(x) - E\hat{m}''_0(x)|$ is stochastically bounded.

Theorem 2 *Let the following conditions on the bandwidth h_N hold:*

$$\begin{aligned} Nh_N^6/(lnN)^2 &\rightarrow \infty \\ Nh_N^8 &\rightarrow 0. \end{aligned}$$

Then, under Assumptions 1-3 and 5-7 a consistent root of of model (3) satisfies

$$\sqrt{N}\begin{pmatrix} \hat{a} - a_0 \\ \hat{b} - b_0 \\ \hat{c} - c_0 \end{pmatrix} \xrightarrow{D} N(0, \boldsymbol{Q}^{-1}\boldsymbol{\Sigma}\boldsymbol{Q}^{-1}),$$

where

$$\boldsymbol{Q} = E\left[\begin{pmatrix} 1 & m_0(x-c_0) & -bm'_0(x-c_0) \\ m_0(x-c_0) & m_0(x-c_0)^2 & -b_0m'_0(x-c_0)m_0(x-c_0) \\ -bm'_0(x-c_0) & -b_0m'_0(x-c_0)m_0(x-c_0) & b_0^2m'_0(x-c_0)^2 \end{pmatrix}\middle| x - c_0 \in \mathcal{W}\right]$$

and

$$\begin{aligned} \boldsymbol{\Sigma} \;\; = \;\; & E\left[\left(\frac{1}{f_w(x)^2}(z - m_1(x))^2 + \frac{b_0^2}{f_x(x-c_0)^2}(y - m_0(x-c_0))^2\right)\right. \\ & \left. \times \begin{pmatrix} 1 & m_0(x-c_0) & -b_0m'_0(x-c_0) \\ m_0(x-c_0) & m_0(x-c_0)^2 & -b_0m'_0(x-c_0)m_0(x-c_0) \\ -b_0m'_0(x-c_0) & -b_0m'_0(x-c_0)m_0(x-c_0) & b_0^2m'_0(x-c_0)^2 \end{pmatrix}\middle| x - c_0 \in \mathcal{W}\right]. \end{aligned}$$

Proof: see Appendix 1.

There are two properties of the limit distribution which are worth to mention: the limit distribution is not affected by the denominator of the loss function (3) which is intuitive because it is just added to the model in order to improve the finite sample performance. It can be seen in the proof that the denominator is eliminated by the usual "sandwich" formula. In contrast, the probability that an observation of each sample falls into $\mathcal{W} \cap \mathcal{W}_{c_0}$ affects the variance matrix of the estimator. It explodes as this probability for one of the samples goes to zero. In this case most of the observations of the respective sample cannot be used for the estimation of the parameters and therefore the estimation becomes more unreliable. However, we do not observe that the convergence rate is affected by these probabilities.

4 Simulations

Let us now investigate the finite sample performance of the 4 step estimator defined above using the HM specification as given in (3) and the PR specification as given in (4). A non-linear least squares estimators is used as a parametric benchmark. It turns out that the results for the two semiparametric estimators differ. Explanations for these differences are provided afterwards.

Let λ denote the Lebesgue measure defined on the Borel-σ-algebra. 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 $c \in [-\lambda(\mathcal{X}_1), \lambda(\mathcal{X}_1)]$. Due to 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 Experiments Two Monte Carlo series shall help to investigate 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 in the value of c_0, where we use $c_0 = 0$ in the first Monte Carlo study and $c_0 = 4$ in the second. 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 4 and 5 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 HM 4 step estimator detects any minimum of the loss functions in contrast to the PR 4 step estimator. The latter 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 of the first experiment. The results of the second experiment are only presented for completeness.

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 6 and 7 support this guess for the PR 4 step estimator. A researcher 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. A specification test might also be constructed using some information about the shape of the objective function.

On the finite sample performance of the HM and the PR specification This paragraph cannot provide a complete formal treatment of the question of interest but it points out two points which (among other things) cause the differences between the two specifications:

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 the integrated square of

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

and the PR 4 step estimator minimizes the integrated square of

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

where we write $f(x) = f$ etc.. What is called variance effect becomes clear when considering a simplified case. Suppose $\epsilon_f = \epsilon_{f_c} = 0$, i.e. the marginal distributions are known. The heart of the minimization problem becomes:

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

for the HM specification and

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

for the PR specification. It is clear that both estimators are the same if $f_c(x) = f(x)$. Otherwise it is important to point out that their error distributions differ. Standard least squares theory tells us that the variance of the HM 4 step estimator is larger whenever f_c and f are less than one. Otherwise it is smaller in this specific case. This simple example serves just as an illustration that each specification of the estimator may perform better.

2. Bias effect: This point becomes clear when rewriting the problem into

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

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 $\xi_f(x)$ differ from one whenever the corresponding estimates are biased. Figure 8 presents some mean $\xi_r(x)$ and $\xi_f(x)$ that are obtained by averaging 1000 monte carlo samples. It is 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 their 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. They are in particular weak if the data is not trimmed in order to eliminate the boundaries. Thus, the final parameter 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 between. Which estimator is preferable depends on the specific situation. In small samples the second point should clearly dominate the first one, since the systematic bias is more evident. The PR specification should therefore not be applied in such 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 assumed to be more biased at the boundaries, we expect the same for the estimates of the unknown functions on a large subset of $\mathcal{W} \cap \mathcal{W}_{c_0}$. In addition, we should take into account that the PR specification is limited to the Nadaraya-Watson estimator, whereby the HM specification works with different nonparametric estimators. In the simulations we use the local linear smoother for the HM specification. This and the above findings are probably the main reasons why the HM specification performs better in the above experiments. (see Figure 5c).

5 Application

This section is devoted to an illustrative application of the HM 4 step estimator to British consumer data. We mainly follow Blundell, Duncan and Pendakur (1998) who use an estimator of the Pinkse and Robinson specification in order to estimate the unknown translation parameters between the expenditure shares of two demographic groups of households. It should therefore be of interest to investigate how the HM 4 step estimator behaves in comparison. Two different data are used for this purpose.

1. **FES I** is the same cross section data of the British Family Expenditure Survey (FES) 1980-1982 as in Blundell et al. (1998). The samples contain couples with one child and couples with two children, respectively.

2. **FES II** is the same cross section data of the FES 1982 as in Blundell, Chen and Kristensen (2001). The samples contain a subsample of childless couples and couples with one or two children.

Further description of the data can be found in the relevant papers. A detailed application of the estimator to German consumer can be found in Wilke (2003).

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 of total expenditure of a household. In contrast to the model of section 2 we now have different sample sizes for the two sample (X_i, Y_i) and (W_i, Z_i). In FES we have $N_x = 594$ and $N_w = 925$, whereby in FES II we have $N_x = 895$ and $N_w = 2456$. This would suggest to introduce an optimal weighting scheme for the comparison of the unknown curves as for example derived in Pinkse and Robinson (1995). Since this is not subject to analysis here, a weighting scheme is not introduced.

Kernel density estimates of the sample distributions using fixed bandwidth are shown in Figure 9. Interestingly, the densities of the FES I sample seem to be related by parametric shifts, too. Since the consistency result also hold for density comparisons, we apply the estimator

$$L_N(a,c) = \frac{\int_{\mathcal{W}\cap\mathcal{W}_c} \left[\hat{f}_1(x) - a - \hat{f}_0(x-c)\right]^2 dx}{\int_{\mathcal{W}\cap\mathcal{W}_c} \hat{f}_x(x)dx}$$

to the densities of the log of total expenditure. Figure 10 and Table 2 show the resulting transformation and the corresponding loss function. Note that Theorem 2 does not hold in this case. Variances of the parameter estimates are therefore not available.

Back to regression estimation, we apply the HM4SE to FES. Appendix 2 presents the estimation results for the different commodities. 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 indicates that the identification conditions for the parameters are not satisfied. For example in the case of food, the hypothesis that expenditure shares are linear cannot be rejected (Blundell et al., 1998). In this case the parameter estimates are inconsistent, since the loss function does not possess a unique minimum. Similar reasoning applies to some of the other commodities.

Blundell, Duncan and Pendakur choose this semiparametric specification because the commonly used partially linear model is ruled out by economic theory. For further details see Lemma 3.1 and Lemma 3.2 in their paper or Blundell, Browning and Crawford (2003). Blundell, Duncan and Pendakur consider a system of unknown regression functions 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). In this case the loss function (3) becomes:

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

The unknown horizontal shift is supposed to be the same for all commodities. All parameters have therefore to be estimated simultaneously. 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: $\hat{c}$ is very sensitive to 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 3. 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 yield 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 here the local linear smoother instead of the Nadaraya-Watson estimator. Since the estimation results indicate that the model is not very well supported by the data, it is of interest to see what happens when using different data.

Estimates for the FES II data 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 4) have reasonable economic intuition. The estimated equivalence scale is positive and suggest that the additional costs of a child are in the range of 58 percent (fixed bandwidth) to 75 percent (variable bandwidth) of the household income of the reference group of households. The reported variances are based on the asymptotic theory of Section 3. Most of the parameter estimates $\hat{a}$ are not significant. Large variances for the parameter estimates are in accordance with the results of the small sample simulations in section 4. However, the main interest of this application would be to estimate c.

Tables

	first experiment		*second experiment*	
	HM4SE	PR4SE	HM4SE	PR4SE
$\hat{a}$	5.2328 (1.2211)	6.8020 (6.7376)	4.8844 (0.3122)	4.1837 (4.0385)
$\hat{b}$	2.1716 (7.9016)	−0.4570 (31.2005)	0.2633 (10.4807)	−1.1374 (19.3405)
$\hat{c}$	0.2398 (2.2289)	0.4324 (12.6926)	0.9527 (10.5263)	−1.6030 (12.8688)

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

	fixed bandwidth
marginal distribution	
$\hat{a}$	0.0053
$\hat{c}$	0.0724

Table 2: FES I, transformation of densities

	fixed bandwidth	variable bandwidth
expenditure shares	$\hat{a}^{(j)}$	
food	−0.0292 (0.2423)	0.0776 (0.5014)
fuel	−0.0176 (0.0336)	0.0140 (0.0437)
clothing	0.0209 (0.1238)	−0.0293 (0.2410)
alcohol	−0.0009 (0.0520)	−0.0137 (0.0562)
transport	0.0149 (0.1502)	−0.0376 (0.2798)
other goods	0.0125 (0.1039)	−0.0162 (0.1280)
$\hat{c}$	0.3926 (0.0086)	−0.3402 (0.0409)

Table 3: Estimation results for the EPLM using FES I; variances in brackets.

	fixed bandwidth	variable bandwidth
expenditure shares	$\hat{a}^{(j)}$	
alcohol	−0.0200 (0.3961)	−0.0178 (0.2849)
catering	−0.0040 (0.0546)	−0.0036 (0.0578)
clothing	−0.0029 (0.1115)	0.0067 (0.1971)
food	−0.0065 (1.5455)	−0.0191 (1.7819)
personal goods and services	0.0027 (0.1065)	0.0030 (0.0856)
leisure goods	0.0137 (0.1145)	0.0158 (0.1144)
travel	−0.0065 (0.3632)	−0.0122 (0.3872)
$\hat{c}$	0.4606 (0.0139)	0.5593 (0.0159)

Table 4: Estimation results for the EPLM using FES II; variances in brackets.

Figures

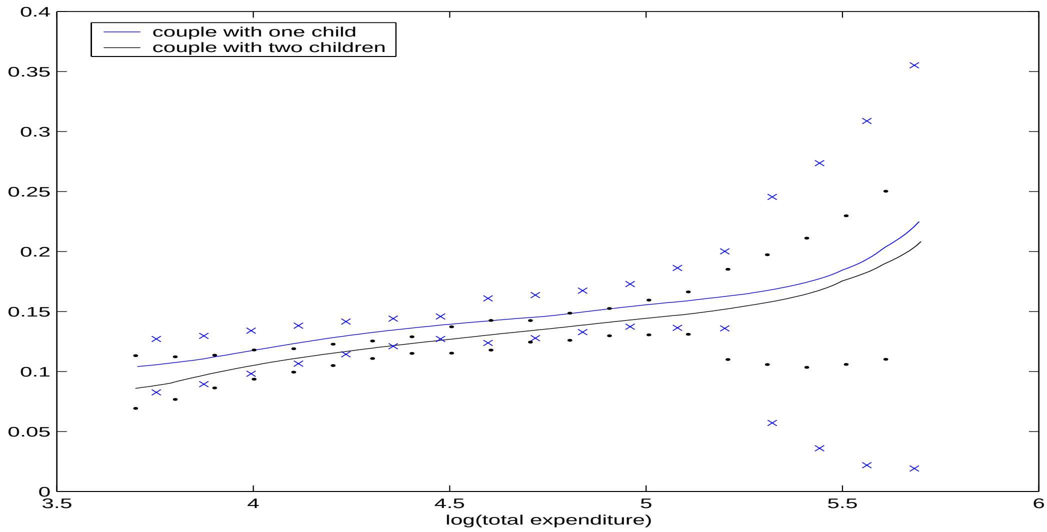

Figure 1: Nonparametric estimates of transport expenditure shares with 95% wild bootstrap confidence bands using the data of Blundell, Duncan and Pendakur (1998).

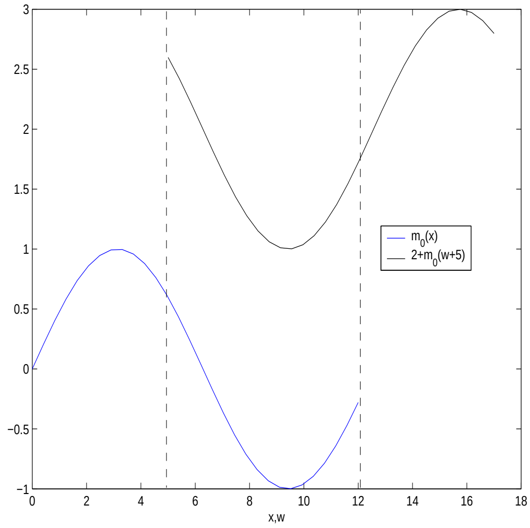

Figure 2: Intersection of observed support: $y = sin(0.5x)$, $z = 2 + sin\big(0.5(w+5)\big)$, $\mathcal{X}_1 = [0, 12]$, $\mathcal{W} = [0, 12]$ and $\mathcal{X}_2 = [5, 17]$.

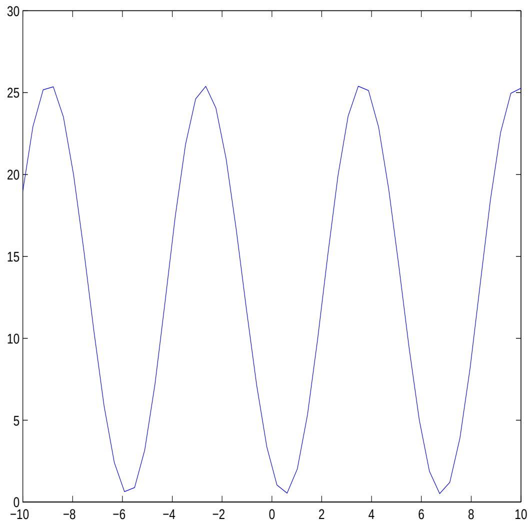

Figure 3: Multiple minima of the loss function: $y = sin(0.5x)$, $z = 5 + 0.5sin(0.5(x - c))$, $C = [-10, 10]$, $c_0 = 0.5$.

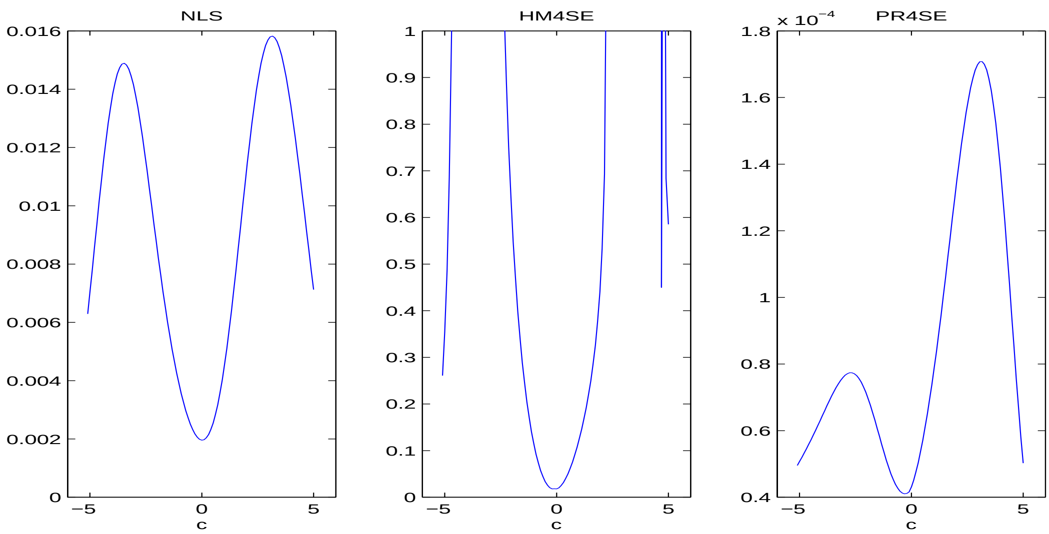

Figure 4: Mean conditional Loss functions $L(c|a,b)$ of the first Monte Carlo Series ($c_0 = 0$): a) parametric b)HM 4 step estimator c) PR 4 step estimator.

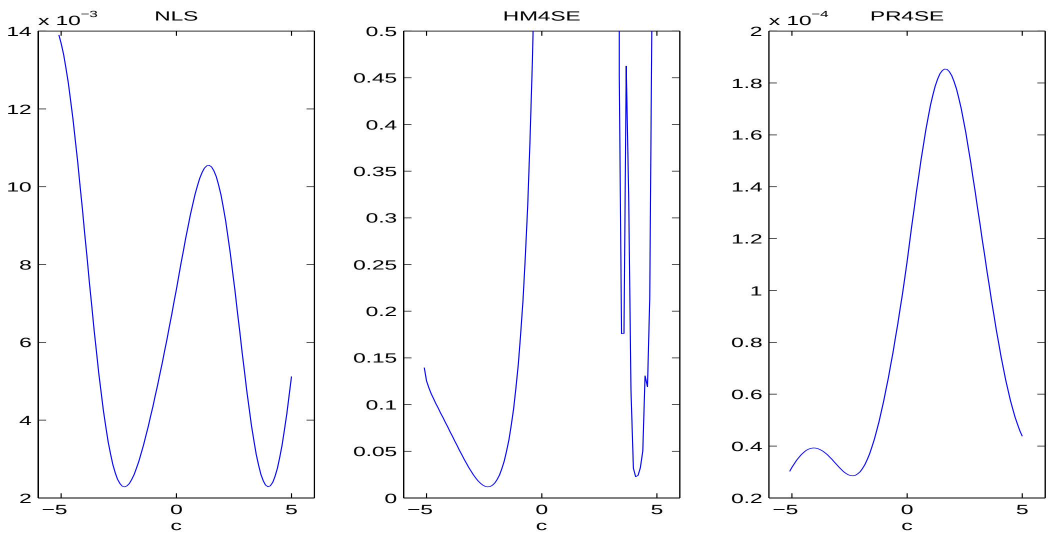

Figure 5: Mean conditional Loss functions $L(c|a,b)$ of the second Monte Carlo Series ($c_0 = 4$): a) parametric b)HM 4 step estimator c) PR 4 step estimator.

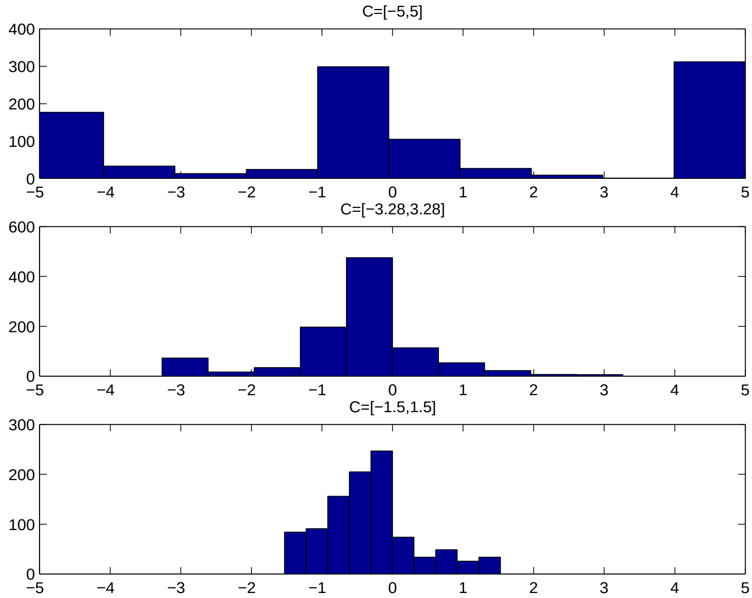

Figure 6: Three histograms for the distribution of $\hat{c}$ obtained with the Pinkse-Robinson 4 step estimator using different supports of c. First Monte Carlo series ($c_0 = 0$).

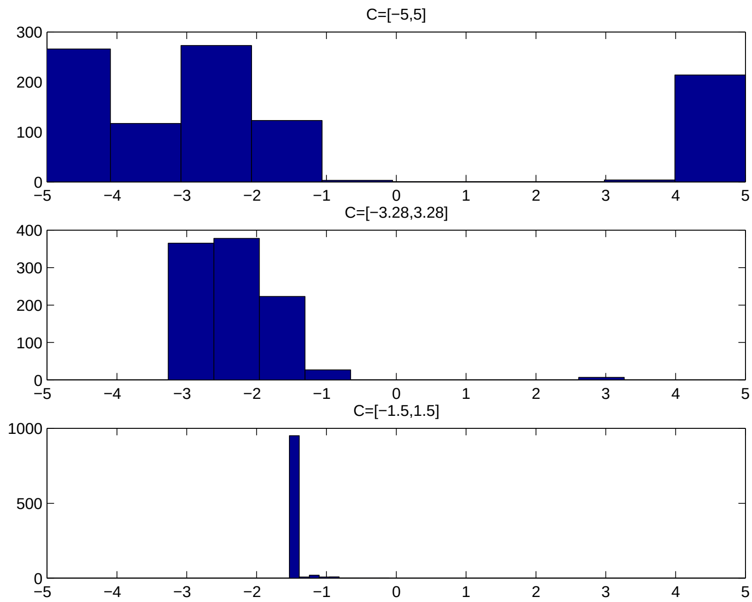

Figure 7: Three histograms for the distribution of $\hat{c}$ obtained with the PR 4 step estimator using different supports of c. Second Monte Carlo series ($c_0 = 4$).

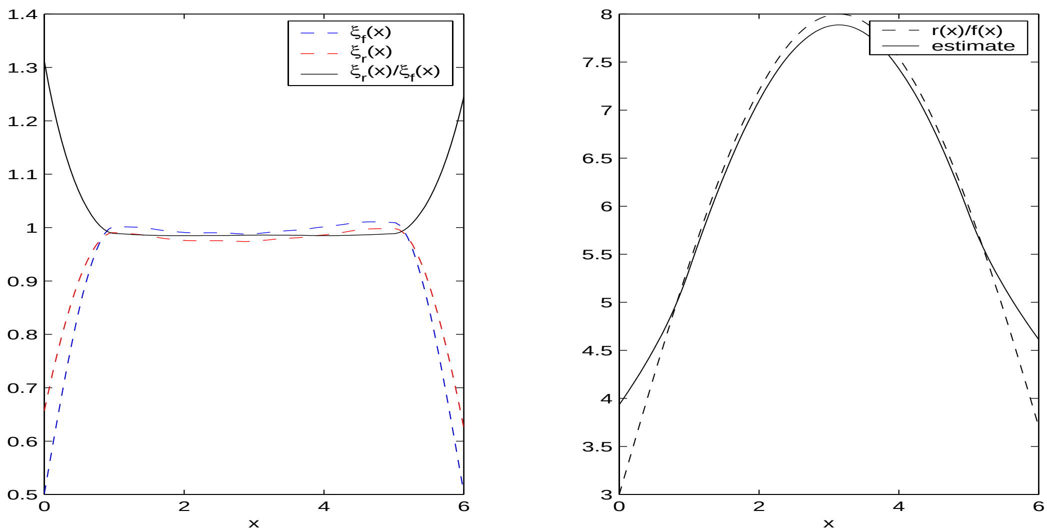

Figure 8: Proportionality of the bias: $N = 2000$, $y = 3 + 5sin(0.5x)$, $x \sim U[0, 6]$, $h = 0.5$, $\xi_f(x)$ and $\xi_r(x)$ are means of 1000 samples

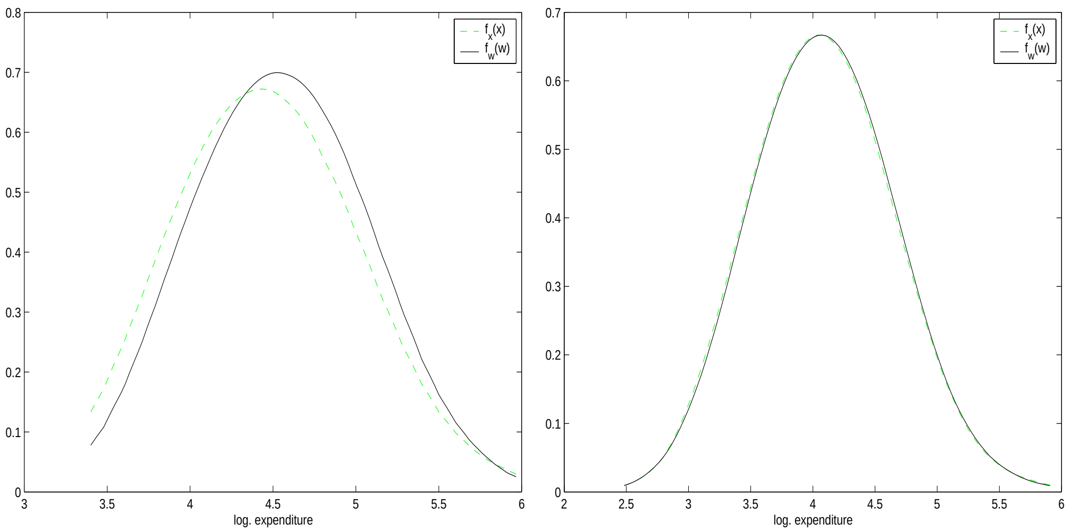

Figure 9: FES I: Kernel estimates of the densities of log expenditure a) FES I b) FES II

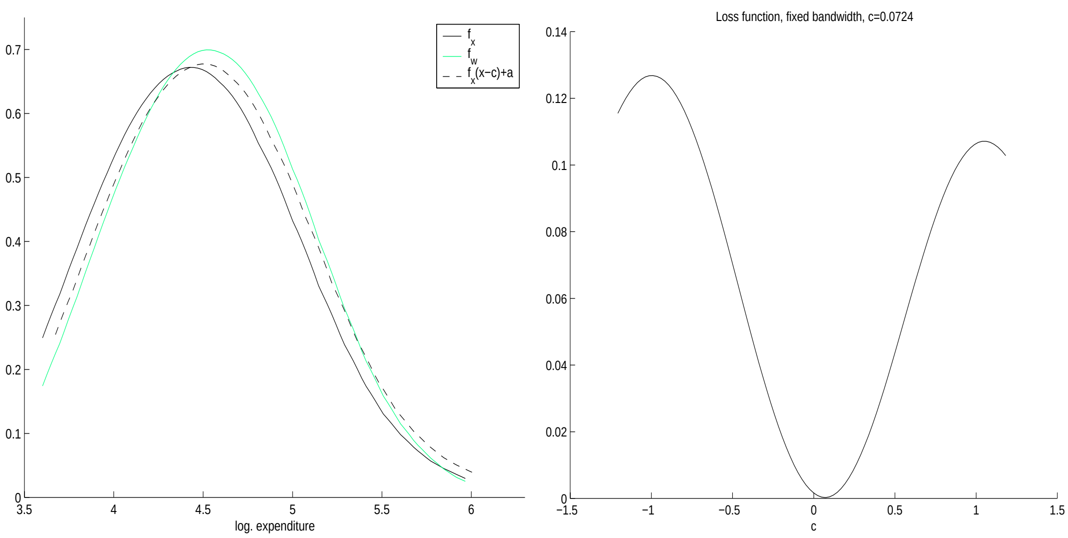

Figure 10: FES I: Pooling of densities a) loss function in c b) (un-)transformed densities

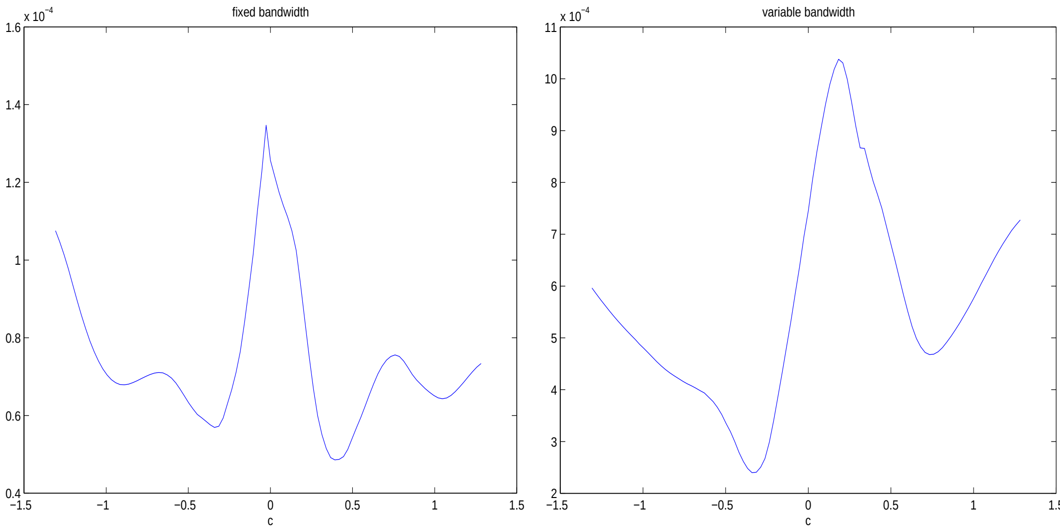

Figure 11: FES I: Loss function of EPLM

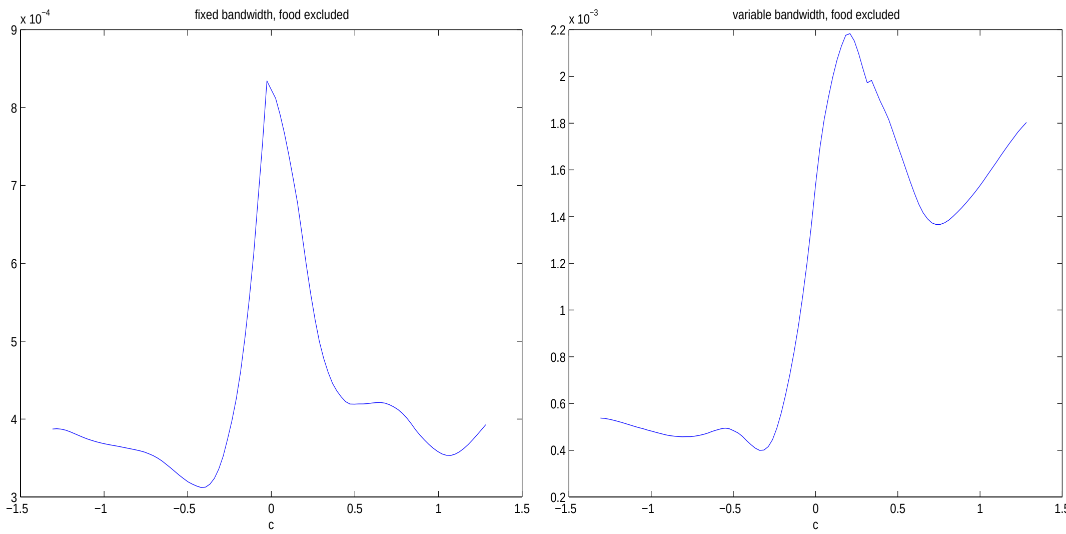

Figure 12: FES I: Loss function of EPLM, J=5 (food excluded).

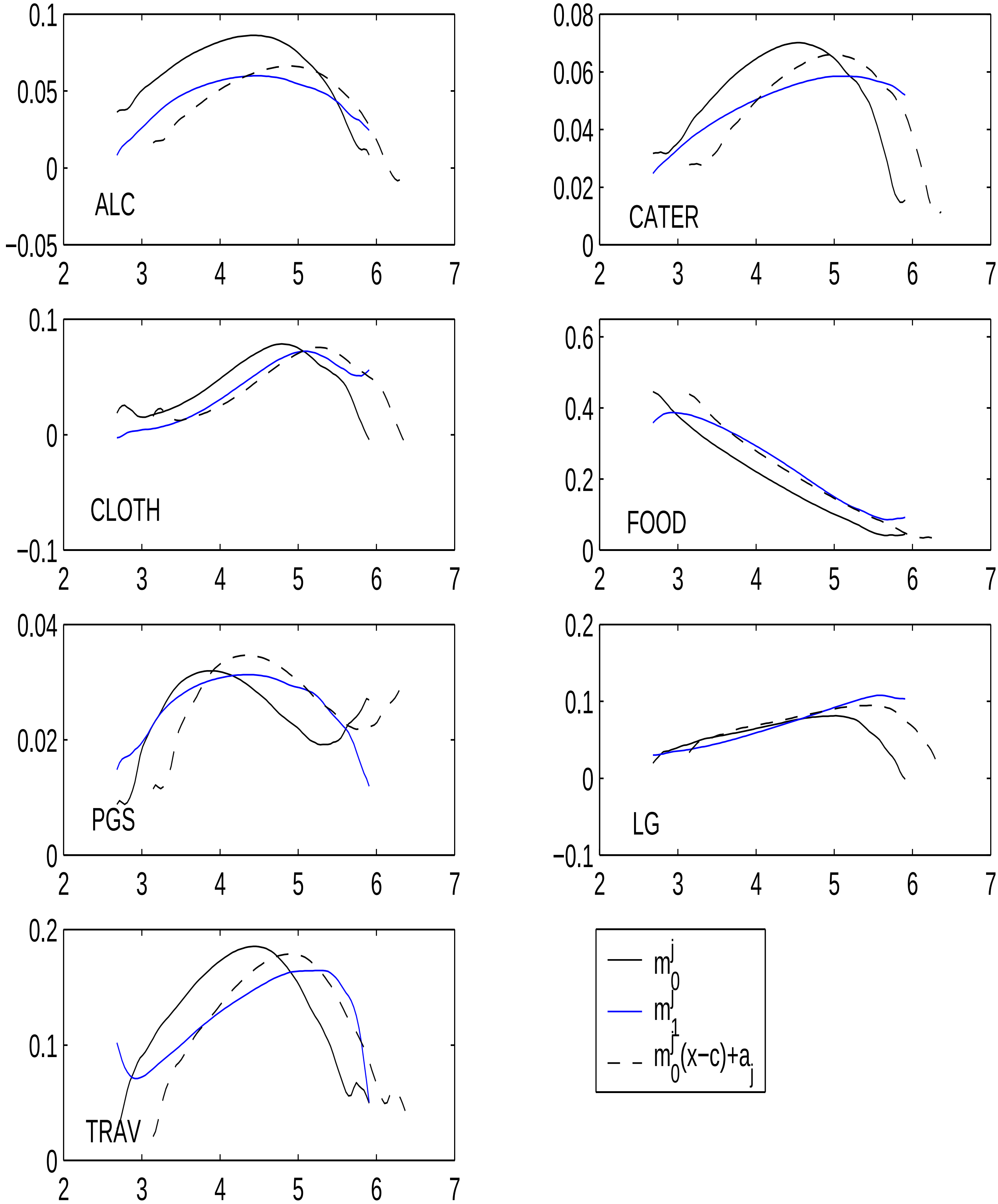

Figure 13: FES II, EPLM, fixed bandwidth

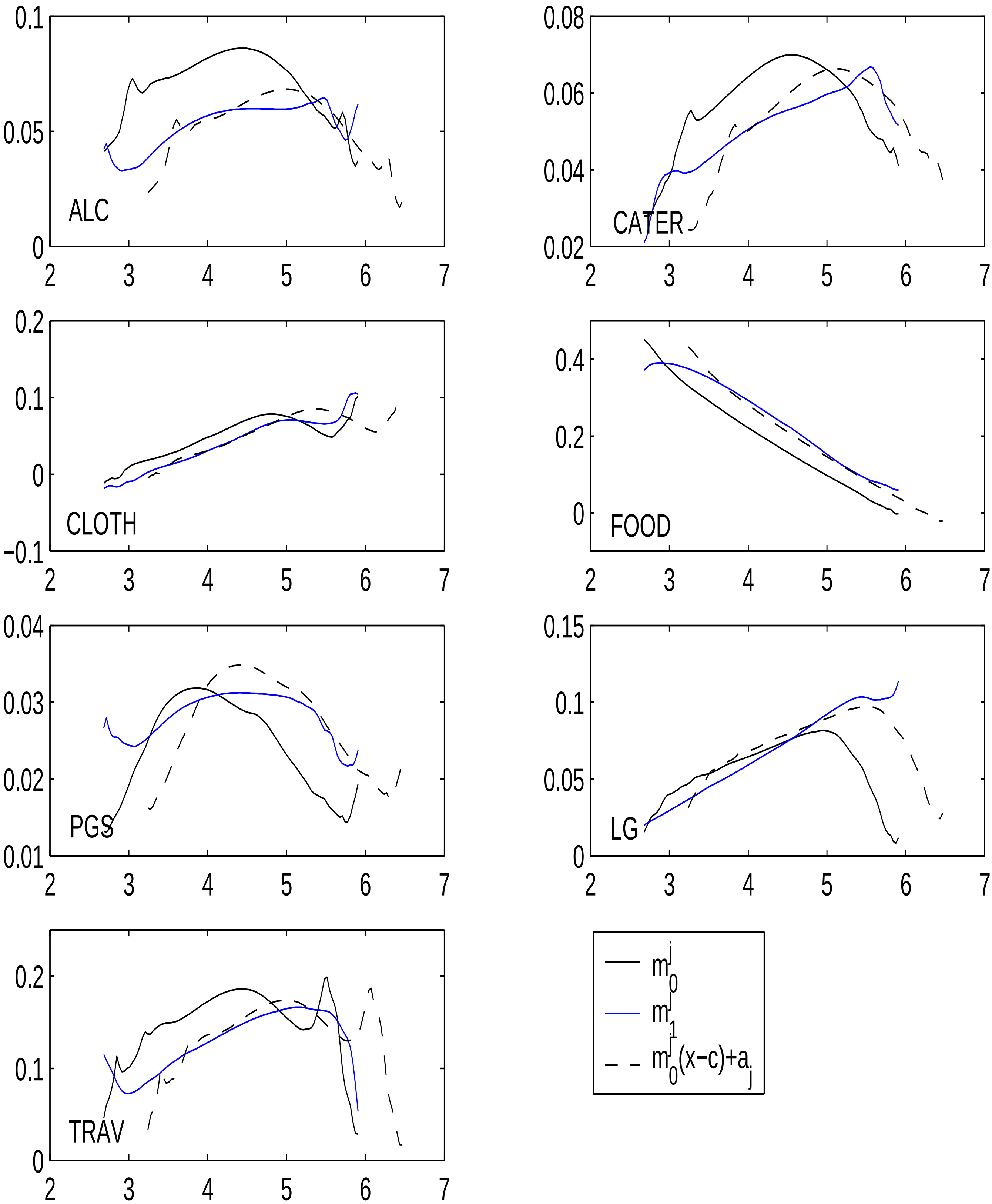

Figure 14: FES II, EPLM, variable bandwidth

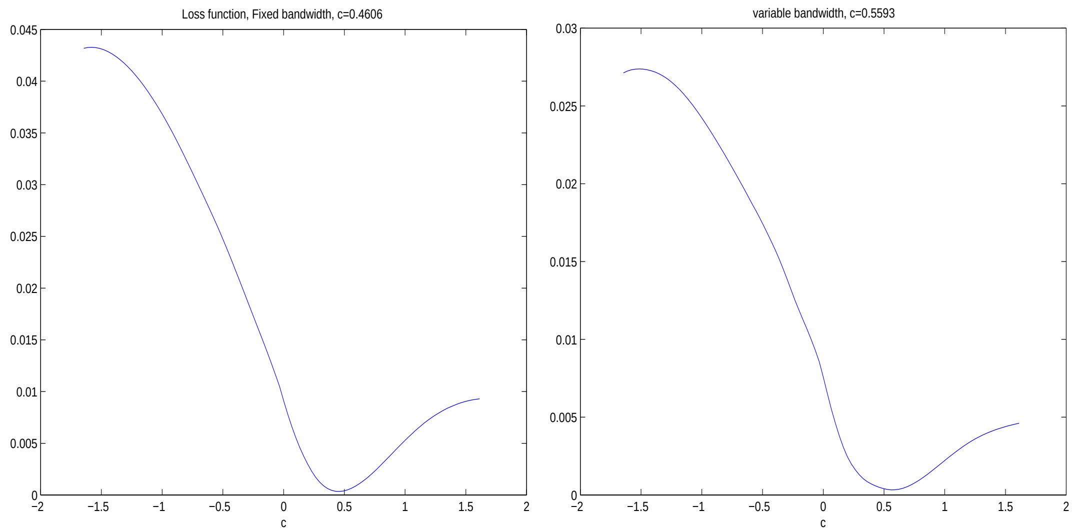

Figure 15: FES II, EPLM, Loss functions

Appendix 1: Proof of Theorems

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. 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

$$\begin{array}{rl} 3. & \text{plim} L_N(a,b,c) = L(a,b,c) \\ \text{and} & \\ 4. & \left.\dfrac{\partial L(a,b,c)}{\partial(a,b,c)}\right|_{(a_0,b_0,c_0)} = 0. \end{array}$$

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 exist.

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:

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

and using the fact that

$$\sup_{\substack{a,b,c \\ x\in\mathcal{X}_1}} \big|a_0 + b_0 m_0(x - c_0) - a - bm_0(x - c)\big| < \infty.$$

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

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

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

$$\text{plim}A_1 = \frac{E\big[\big(a_0 + b_0 m_0(x - c_0) - a - bm_0(x - c)\big)^2|x - c \in \mathcal{W}\big]}{\int_{\underline{x}(c)}^{\overline{x}(c)} f_x(x)dx}.$$

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 probability limit of the loss function, i.e. plim$L_N(a, b, c)$ =plimA_1, has a unique minimum at a_0, b_0, c_0, i.e.

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

We have to check the necessary and the sufficient conditions.

The first order conditions are:

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

From (9) and (10) we obtain

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

Substituting for a yields:

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

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 than or equal to one. It is therefore enough to show that the numerator 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

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

Assumption 3 ensures that

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

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

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

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 numerator 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. ■

Proof of Theorem 2: The derivative of the objective function at the minimum is a moment condition. This moment condition can be decomposed. Based on the Slutsky theorem we show that some of the elements converge in probability and that the score term converges in distribution. The latter is shown using the results of semiparametric two step estimators given by Newey and McFadden (1994).

Observe first that Assumptions 1-3, 5-7 and the appropriate bandwidth choice ensure the existence of a consistent root. Observe also, that the loss function (3) can be rewritten as:

$$L_N(\theta) = \frac{1}{\sum_{i=1}^N \mathbb{I}_{X_i - c \in \mathcal{W}}} \sum_{i=1}^N \mathbb{I}_{X_i - c \in \mathcal{W}} \big(\hat{m}_1(w_i^c) - a - b\,(\hat{m}_0(X_i - c))\big)^2$$

where $\theta = (a, b, c)'$ and $m_1(x) = a_0 + b_0 m_0(x - c_0)$. This loss function implies that the nonparametric function m_0 is evaluated at X_i and shifted by the value c to w_i^c, i.e. $w_i^c = x_i - c$. The nonparametric function m_1 is evaluated at w_i^c. Note that in most cases $w_i^c \neq W_i$. This ensures that we can compare the shifted estimate of m_0 to the estimate of m_1. According to the consistency proof, let denote $\hat{m}_1(w_i^c) = \hat{m}_1(X_i)$.

A first order Taylor expansion of $\partial_\theta L_N(\theta)/\partial_\theta|_{\theta_0}$ yields:

$$\sqrt{N}(\hat{\theta} - \theta_0) \quad = \quad - \left[\frac{\partial L_N(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta^*} \right]^{-1} \sqrt{N} \frac{\partial L_N(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0},$$

where θ^* lies between θ_0 and $\hat{\theta}$ and denote $\theta_0 = (a_0, b_0, c_0)'$. We have to show that

1. the score term at the true parameter values
$$\sqrt{N} \frac{\partial L_N(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}$$
converges in distribution to a mean zero normal random variable.

2. the Hessian
$$\frac{\partial L_N(\theta)}{\partial\theta\partial\theta'} = \sum_i \mathbf{H}_i(\theta^*)$$
converges in probability uniformly to a nonsingular matrix $\mathbf{Q}$.

Introduce the following notations:
the score term may be written as

$$\begin{aligned} \sqrt{N} \frac{\partial L_N(\theta)}{\partial\theta} \quad &= \quad \sqrt{N} \sum_{i=1}^N q_i(\theta, \hat{m}_0, \hat{m}_1) \\ &= \quad \frac{1/\sqrt{N} \sum_i q_{1i}(\theta, \hat{m}_0, \hat{m}_1)}{1/N \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}}, \end{aligned}$$

where

$$q_i(\theta, \hat{m}_0, \hat{m}_1) = \frac{q_{1i}(\theta, \hat{m}_0, \hat{m}_1)}{\sum_{i=1}^{N} \mathbb{1}_{X_i - c \in \mathcal{W}}}$$

with the 3×1 vector

$$\begin{aligned} q_{1i}(\theta, \hat{m}_0, \hat{m}_1) &= -2\mathbb{1}_{X_i - c \in \mathcal{W}} \begin{pmatrix} \hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c) \\ \hat{m}_0(X_i - c)\left[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c)\right] \\ b\hat{m}_0'(X_i - c)[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c)] \end{pmatrix} \\ &\quad -\mathbb{1}_{X_i - c \in \mathcal{W}} \begin{pmatrix} 0 \\ 0 \\ [\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c)]^2 \left[\frac{\frac{\partial}{\partial c}\sum_i \mathbb{1}_{X_i - c \in \mathcal{W}}}{\sum_i \mathbb{1}_{X_i - c \in \mathcal{W}}}\right] \end{pmatrix}. \end{aligned}$$

The following relations hold:

$$0 < \sum_i \mathbb{1}_{X_i - c \in \mathcal{W}_c \cap \mathcal{W}} \leq N$$

and

$$\frac{1}{N} \sum_i \mathbb{1}_{X_i - c \in \mathcal{W}} \xrightarrow{P} \text{Prob}(x - c \in \mathcal{W}) \text{ a.s.}$$

Therefore, $N^{-1} \sum_i \mathbb{1}_{X_i - c \in \mathcal{W}}$ converges almost surely to a twice continuously differentiable function. Then, the denominator of

$$\sqrt{N} \sum_i q_i(\theta, \hat{m}_0, \hat{m}_1)$$

converges almost surely to a twice continuously differentiable function. Moreover, we have

$$q_i(\theta_0, m_0, m_1) = 0.$$

ad **1.**

By the Slutsky Theorem we have to show that the numerator of the score term, i.e.

$$\frac{1}{\sqrt{N}} \sum_i q_{1i}(\theta, \hat{m}_0, \hat{m}_1)$$

converges in distribution. Using the linearization

$$\hat{\beta}\hat{\alpha} - \beta\alpha = \alpha(\hat{\beta} - \beta) + (\hat{\alpha} - \alpha)\hat{\beta}$$

we obtain

$$\begin{aligned}
& q_{1i}(\theta, \hat{m}_0, \hat{m}_1) - q_{1i}(\theta, m_0, m_1) \\
= \; & 2\left[\hat{m}_1(X_i) - m_1(X_i) - b[\hat{m}_0(X_i - c) - m_0(X_i - c)]\right] \begin{pmatrix} -1 \\ -m_0(X_i - c) \\ bm_0'(X_i - c) \end{pmatrix} \\
& +2[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c)] \begin{pmatrix} 0 \\ -\hat{m}_0(X_i - c)] + m_0(X_i - c) \\ b[\hat{m}_0'(X_i - c) - m_0'(X_i - c)] \end{pmatrix} \\
& -\left([\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c)]^2 - \left[[m_1(X_i) - a - bm_0(X_i - c)]^2\right]^2 \right) \begin{pmatrix} 0 \\ 0 \\ \left[\frac{\frac{\partial}{\partial c}\sum_i \mathbb{1}_{X_i - c \in \mathcal{W}}}{\sum_i \mathbb{1}_{X_i - c \in \mathcal{W}}}\right] \end{pmatrix} \\
= \; & A_i + B_i + C_i
\end{aligned}$$

To be shown:

I

$$\frac{1}{\sqrt{N}} \sum \mathbb{1}_{X_i - c \in \mathcal{W}} A_i \xrightarrow{D} N(0, \Sigma)$$

II

$$\frac{1}{\sqrt{N}} \sum_i \mathbb{1}_{X_i - c \in \mathcal{W}} B_i \xrightarrow{P} 0$$

uniformly in x and

$$\frac{1}{\sqrt{N}} \sum_i \mathbb{1}_{X_i - c \in \mathcal{W}} C_i \xrightarrow{P} 0$$

uniformly in x.

ad **III**

We show this using the framework of Newey and McFadden Theorem 8.11. For this reason we introduce the following notations:

$$
\begin{aligned}
A_i &= \frac{1}{2}\left(g_i(\theta,\hat{\gamma}) - g_i(\theta,\gamma)\right) \\
\gamma &= (\gamma_1,\gamma_2,\gamma_3,\gamma_4)' \\
m_1 &= \gamma_2/\gamma_1 \\
m_0 &= \gamma_4/\gamma_3
\end{aligned}
$$

Note, that we use here the information that the Nadaraya-Watson estimator can be written as a fraction. Then,

$$
\begin{aligned}
& g_i(\theta,\hat{\gamma}) - g_i(\theta,\gamma) \\
= \;& \mathbb{1}_{X_i - c\in\mathcal{W}}\left(\frac{\hat{\gamma}_2(X_i)}{\hat{\gamma}_1(X_i)} - \frac{\gamma_2(X_i)}{\gamma_1(X_i)} - b\left(\frac{\hat{\gamma}_4(X_i-c)}{\hat{\gamma}_3(X_i-c)} - \frac{\gamma_4(X_i-c)}{\gamma_3(X_i-c)}\right)\right)\begin{pmatrix} -1 \\ -m_0(X_i-c) \\ bm_0'(X_i-c). \end{pmatrix}
\end{aligned}
$$

Linearize in $\hat{\gamma}_2/\hat{\gamma}_1$ and $\hat{\gamma}_4/\hat{\gamma}_3$ around the true values by

$$
\frac{\hat{\alpha}}{\hat{\beta}} - \frac{\alpha}{\beta} = \beta^{-1}\left[\hat{\alpha} - \alpha - \frac{\alpha}{\beta}(\hat{\beta}-\beta)\right]
$$

in order to obtain a function that is linear in the error of the nonparametric estimates:

$$
\begin{aligned}
& G(x,\theta,\hat{\gamma}-\gamma) \\
= \;& \mathbb{1}_{x-c\in\mathcal{W}}\left(\frac{1}{\gamma_1}\left[(\hat{\gamma}_2-\gamma_2) - \frac{\gamma_2}{\gamma_1}(\hat{\gamma}_1-\gamma_1)\right] - \frac{b}{\gamma_3}\left[(\hat{\gamma}_4-\gamma_4) - \frac{\gamma_4}{\gamma_3}(\hat{\gamma}_3-\gamma_3)\right]\right)\begin{pmatrix} -1 \\ -m_0(x-c) \\ bm_0(x-c) \end{pmatrix} \\
= \;& \left(\mathbb{1}_{x-c\in\mathcal{W}}\left[\frac{1}{\gamma_1}[-m_1(x),1], \frac{b}{\gamma_3}[m_0(x-c),-1]\right)\right]\begin{pmatrix} \hat{\gamma}_1-\gamma_1 \\ \hat{\gamma}_2-\gamma_2 \\ \hat{\gamma}_3-\gamma_3 \\ \hat{\gamma}_4-\gamma_4 \end{pmatrix}\right)\begin{pmatrix} -1 \\ -m_0(x-c) \\ bm_0(x-c) \end{pmatrix}
\end{aligned}
$$

Define the norm $\|\gamma\| = \max_{0\le l\le 1}\sup_x|\partial^l\gamma(x)/\partial_x^l|$ as a Sobolev norm. We have now to check the four conditions of Theorem 8.11:

(i) We show this using Lemma 8.10 of Newey and McFadden (1994). The conditions for the application of this Lemma are satisfied by the model setup and Assumptions 5-7. Hence, we have

$$
\sqrt{N}\|\hat{\gamma}-\gamma\|^2 \xrightarrow{P} 0.
$$

(ii)

$$
\|G(x,\theta,\gamma)\| \le c(x,\theta)\|\gamma\|
$$

holds since γ_1 and γ_3 are bounded away from zero on their bounded supports, $|b| < \infty$ and $\sup_x |m_0(x)|$ and $\sup_x |m_1(x)| < \infty$.
(iii) There is $v(x)$ with $\int G(x,\theta,\gamma)dF(x) = \int v(x)\gamma(x)dx$ for all $\|\gamma\|$. Choose

$$v(x) = \mathbb{I}_{x-c\in\mathcal{W}} \begin{pmatrix} -\left[\frac{1}{\gamma_1}[-m_1(x),1], \frac{b}{\gamma_3}[m_0(x-c),-1])\right] \\ -m_0(x-c)\left[\frac{1}{\gamma_1}[-m_1(x),1], \frac{b}{\gamma_3}[m_0(x-c),-1])\right] \\ bm_0(x-c)\left[\frac{1}{\gamma_1}[-m_1(x),1], \frac{b}{\gamma_3}[m_0(x-c),-1])\right] \end{pmatrix}$$

(iv) $v(x)$ is continuous except at the two points where the indicator function switches. Moreover, it is bounded for all x.

Then,

$$\begin{aligned} \delta(x,y,z) &= [1,z,1,y]v(x) - E([1,z,1,y]v(x)) \\ &= \mathbb{I}_{x-c\in\mathcal{W}}\Bigg(\frac{1}{\gamma_1}(z-m_1(x)) + \frac{b}{\gamma_3}(y-m_0(x-c)) \\ &\qquad -E\left[\frac{1}{\gamma_1}(z-m_1(x)) + \frac{b}{\gamma_3}(y-m_0(x-c))\right]\Bigg)\begin{pmatrix} -1 \\ -m_0(x-c) \\ bm_0'(x-c)\end{pmatrix} \end{aligned}$$

Noting that $E(U_i|X_i) = E(V_i|W_i) = 0$ for all i and j, mutually independence and $g(x,\theta_0,\gamma) = 0$ yields

$$\begin{aligned} &\text{var}(g(x,\theta_0,\gamma) + \delta(x,y,z)) = E[\delta(z)^2] \\ &= E\left(\mathbb{I}_{x-c\in\mathcal{W}}\left[\frac{1}{\gamma_1}(z-m_1(x)) + \frac{b}{\gamma_3}(y-m_0(x-c))\right]^2 \begin{pmatrix} -1 \\ -m_0(x-c) \\ bm_0'(x-c)\end{pmatrix}\begin{pmatrix} -1 \\ -m_0(x-c) \\ bm_0'(x-c)\end{pmatrix}'\right). \end{aligned}$$

Let denote $E[\delta(x,y,z)^2]|_{\theta=\theta_0} = E[\delta_0(x,y,z)^2]$, then

$$\begin{aligned} &E[\delta_0(x,y,z)^2] \\ &= E\Bigg[\left(\frac{1}{f_w(x)^2}(z-m_1(x))^2 + \frac{b_0^2}{f_x(x-c_0)^2}(y-m_0(x-c_0))^2\right) \\ &\times \left.\begin{pmatrix} 1 & m_0(x-c_0) & -b_0m_0'(x-c_0) \\ m_0(x-c_0) & m_0(x-c_0)^2 & -b_0m_0'(x-c_0)m_0(x-c_0) \\ -b_0m_0'(x-c_0) & -b_0m_0'(x-c_0)m_0(x-c_0) & b_0^2m_0'(x-c_0)^2 \end{pmatrix}\right| x-c_0\in\mathcal{W}\Bigg] \end{aligned}$$

Hence, application of Theorem 8.11 of Newey and McFadden (1994) yields

$$\sqrt{N}\sum_i q_i(\theta)|_{\theta=\theta_0} \xrightarrow{D} N\left(0, \frac{4}{\text{Prob}(x-c_0\in\mathcal{W})^2}\mathbf{\Sigma}\right),$$

where

$$\boldsymbol{\Sigma} = E[\delta_0(x,y,z)^2]$$

ad **II**

By the Slutsky Theorem it suffices to show for each i

$$\mathbb{1}_{X_i - c \in \mathcal{W} \cap \mathcal{W}_c} B_i \xrightarrow{P} 0$$

uniformly in x. First note again that we have $\hat{\theta} \xrightarrow{P} \theta_0$. Thus, the problem is shown if the following holds

$$\begin{aligned}
& \mathbb{1}_{X_i - c \in \mathcal{W}} \left(\hat{m}_1(X_i) - a_0 - b_0 \hat{m}_0(X_i - c_0)\right) \begin{pmatrix} 0 \\ -\hat{m}_0(X_i - c_0) + m_0(X_i - c_0) \\ +b_0 \hat{m}_0'(X_i - c_0) - b_0 m_0'(X_i - c_0) \end{pmatrix} \\
\leq\ & \mathbb{1}_{X_i - c \in \mathcal{W}} \left(|\hat{m}_1(X_i) - m_1(X_i)| + |m_1(X_i) - a_0 - b_0 \hat{m}_0(X_i - c_0)|\right) \\
& \begin{pmatrix} 0 \\ -\hat{m}_0(X_i - c_0) + m_0(X_i - c_0) \\ +b_0(\hat{m}_0'(X_i - c_0) - m_0'(X_i - c_0)) \end{pmatrix} \\
=\ & o_p(N^{-1/2})
\end{aligned}$$

Therefore, we have to verify two conditions:

(i)

$$\left[|\hat{m}_1(X_i) - m_1(X_i)| + |b_0(m_0(X_i - c_0) - \hat{m}_0(X_i - c_0))|\right] (m_0(X_i - c_0) - \hat{m}_0(X_i - c_0)) = o_p(N^{-1/2})$$

(ii)

$$\left[|\hat{m}_1(X_i) - m_1(X_i)| + |b_0(m_0(X_i - c_0) - \hat{m}_0(X_i - c_0))|\right] b_0 \left(m_0'(X_i - c_0) - \hat{m}_0'(X_i - c_0)\right) = o_p(N^{-1/2})$$

Noting that by Lemma 8.10 Newey McFadden we have that the nonparametric estimates and its derivatives converge at rate $N^{1/4}$, we obtain

$$\sqrt{N}\|\hat{\gamma} - \gamma\|^2 \;=\; N^{1/4}\|\hat{\gamma} - \gamma\| N^{1/4}\|\hat{\gamma} - \gamma\| \xrightarrow{P} 0$$

and therefore (i) and (ii) hold.

A similar reasoning applies to the proof of

$$\mathbb{1}_{X_i - c \in \mathcal{W}} C_i \xrightarrow{P} 0$$

uniformly in x for all i and taking into account that one of the squared terms is zero at the true parameter values.

ad **2.**

This step can be proved by using Theorem 4.1.5 Amemiya (1985) or Theorem 8.12 Newey and McFadden (1994). We use here the first one. The sum of the symmetric Hessians is given by a 3×3 matrix with entries

$$\begin{aligned}
\sum_i \mathbf{H}_{i(aa)}(\theta) &= \frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \\
\sum_i \mathbf{H}_{i(ab)}(\theta) &= \frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \hat{m}_0(X_i - c) \\
\sum_i \mathbf{H}_{i(ac)}(\theta) &= -\frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \Bigg(\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} b \hat{m}_0'(X_i - c) \\
&\quad - \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} [\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c)] \left[\frac{\frac{\partial}{\partial_c} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \right] \Bigg) \\
\sum_i \mathbf{H}_{i(bb)}(\theta) &= \frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \hat{m}_0(X_i - c)^2 \\
\sum_i \mathbf{H}_{i(bc)}(\theta) &= \frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \Bigg(\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \hat{m}_0'(X_i - c) \left[\hat{m}_1(X_i) - a - 2b\hat{m}_0(X_i - c) \right] \\
&\quad - \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} [\hat{m}_1(X_i) - a - b\hat{m}_1(X_i - c)] \hat{m}_0(X_i - c) \left[\frac{\frac{\partial}{\partial_c} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \right] \Bigg) \\
\sum_i \mathbf{H}_{i(cc)}(\theta) &= -\frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \Bigg(\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} b \hat{m}_0''(X_i - c) \left[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c) \right] \\
&\quad - \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} b^2 \hat{m}_0'(X_i - c)^2 \Bigg) \\
&\quad - \frac{2}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \times \mathbb{I}_{X_i - c \in \mathcal{W}}} \\
&\quad \times \Bigg(\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} b \hat{m}_0'(X_i - c) \left[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c) \right]^2 \left[\frac{\frac{\partial}{\partial_c} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \right] \\
&\quad + \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} b \hat{m}_0'(X_i - c) \left[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c) \right] \frac{\partial}{\partial_c} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \\
&\quad + \frac{1}{2} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \left[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c) \right]^2 \frac{\partial^2}{\partial c^2} \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \\
&\quad - \sum_i \mathbb{I}_{X_i - c \in \mathcal{W}} \left[\hat{m}_1(X_i) - a - b\hat{m}_0(X_i - c) \right]^2 \left[\frac{\frac{\partial}{\partial_c} \sum_i \mathbb{I}^2_{X_i - c \in \mathcal{W}}}{\sum_i \mathbb{I}_{X_i - c \in \mathcal{W}}} \right] \Bigg)
\end{aligned}$$

Using the information $\theta^* \xrightarrow{P} \theta$, $\|\hat{m}_0 - m_0\| \xrightarrow{P} 0$ and $\|\hat{m}_1 - m_1\| \xrightarrow{P} 0$ and $|E\hat{m}_0'' - m_0''|$ is stochastically bounded (Lemma 8.9, Newey and McFadden, 1994) we can apply Theorem 4.1.5 Amemiya (1985) if the limit functions are continuous at the true parameter values. Indeed, we have by the Slutsky Theorem

$$
\begin{aligned}
& \operatorname{plim} \sum_i \mathbf{H}_i(\theta^*) \\
= \; & 2 \begin{pmatrix}
1 & \frac{E[m_0(x-c_0)|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})} & -\frac{E[bm_0'(x-c_0)|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})} \\
\frac{E[m_0(x-c_0)|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})} & \frac{E[m_0(x-c_0)^2|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c \in \mathcal{W})} & -\frac{E[b_0 m_0'(x-c_0) m_0(x-c_0)|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})} \\
-\frac{E[bm_0'(x-c_0)|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})} & -\frac{E[b_0 m_0'(x-c_0) m_0(x-c_0)|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})} & \frac{E[b_0^2 m_0'(x-c_0)^2|x-c_0 \in \mathcal{W}]}{\text{Prob}(x-c_0 \in \mathcal{W})}
\end{pmatrix} \\
= \; & \frac{2}{\text{Prob}(x - c_0 \in \mathcal{W})} \mathbf{Q},
\end{aligned}
$$

where the convergence in probability is uniformly in x and the positive definite matrix $\mathbf{Q}$ (see sufficient condition in consistency proof) consists of continuous functions. Note that all remainders in $\mathbf{H}$ converge to zero in probability uniformly at the true parameter values. ■

Appendix 2: Estimation results

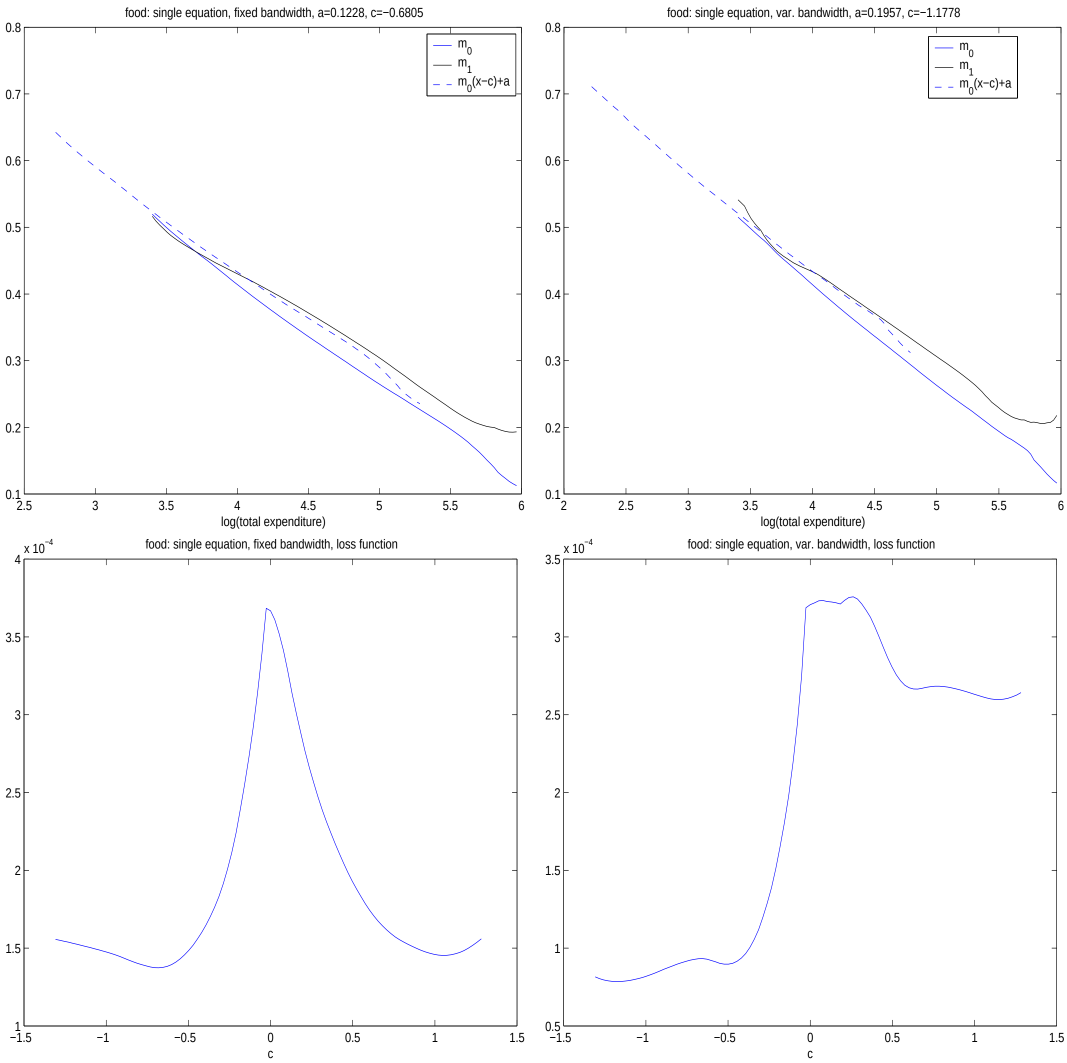

Figure 16: FES I: Food expenditure share.

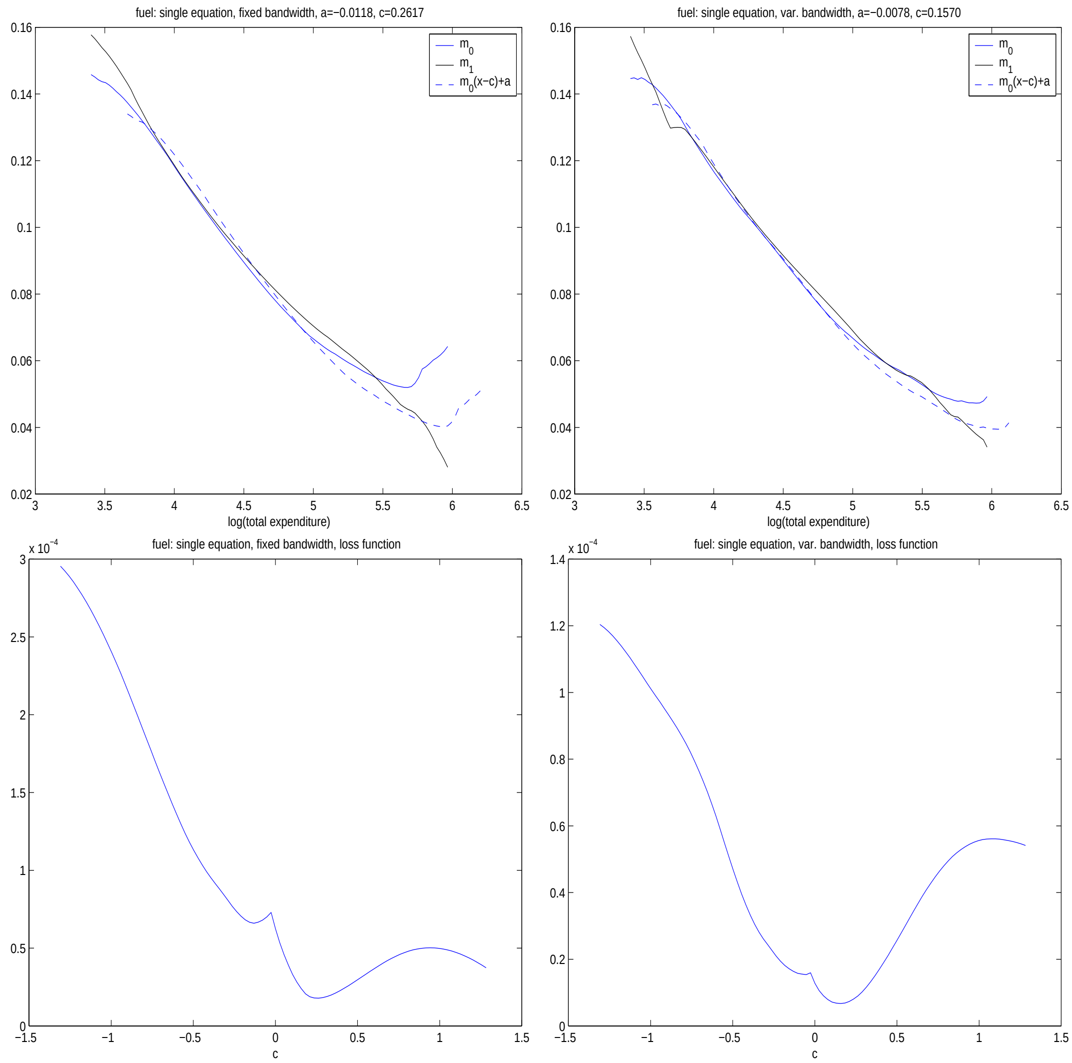

Figure 17: FES I: Fuel expenditure share.

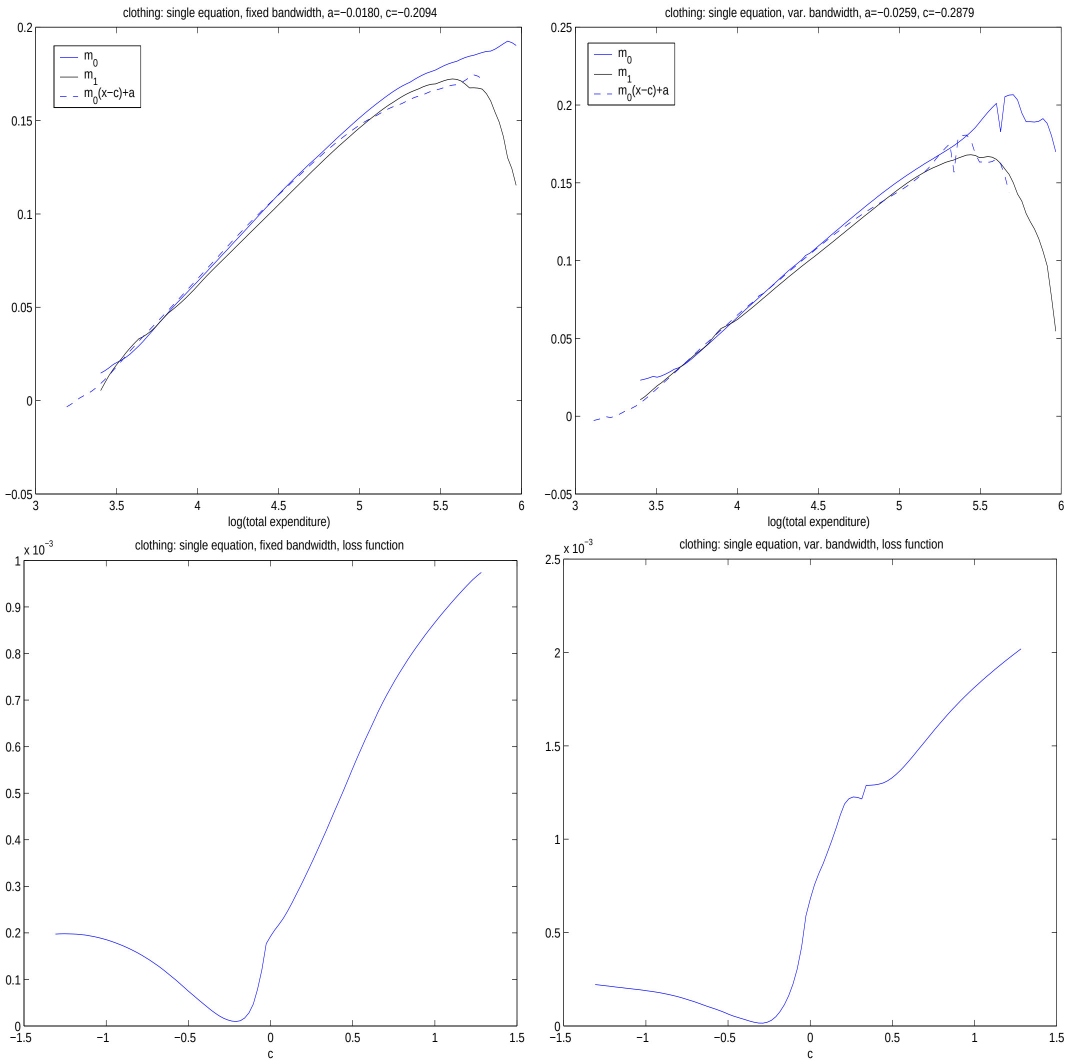

Figure 18: FES I: Clothing expenditure share.

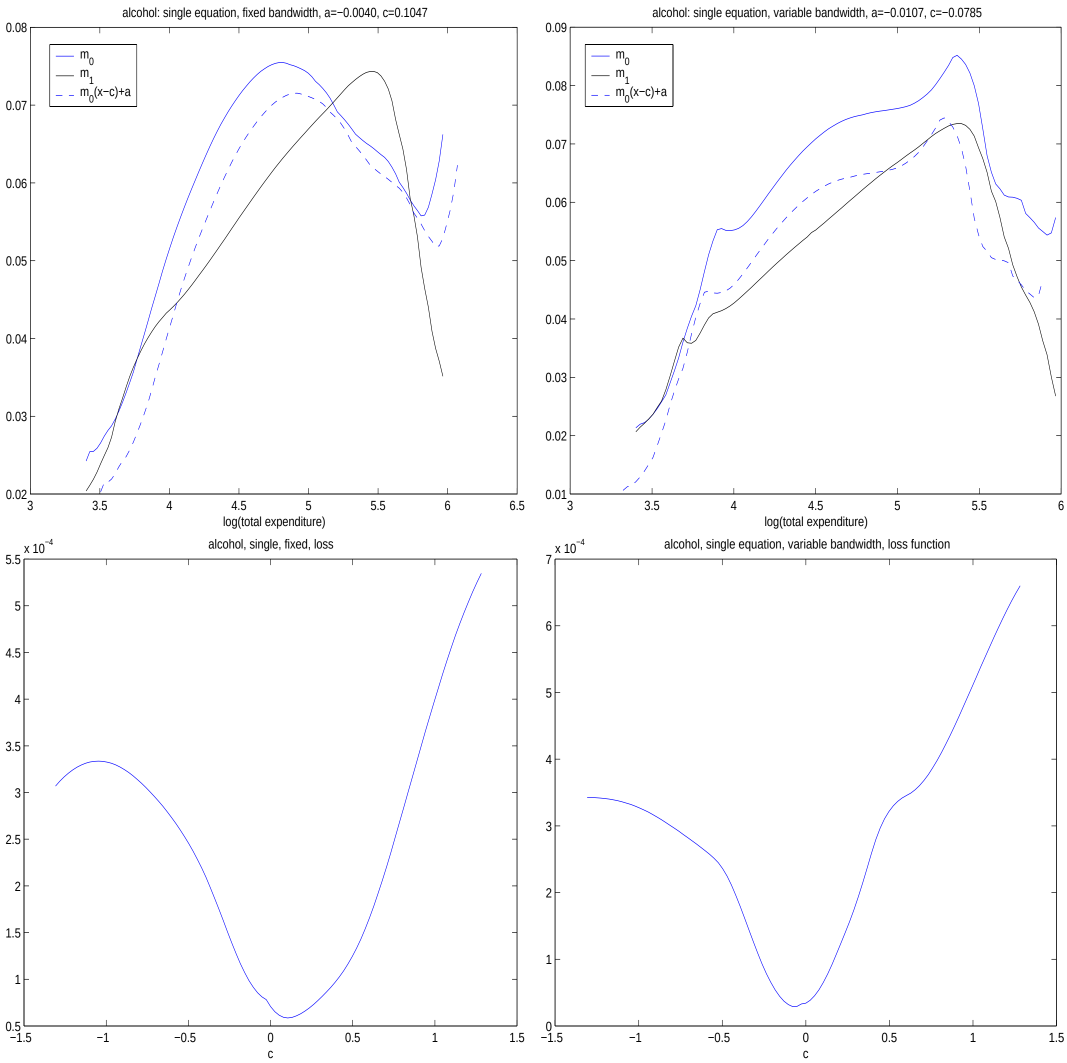

Figure 19: FES I: Alcohol expenditure share.

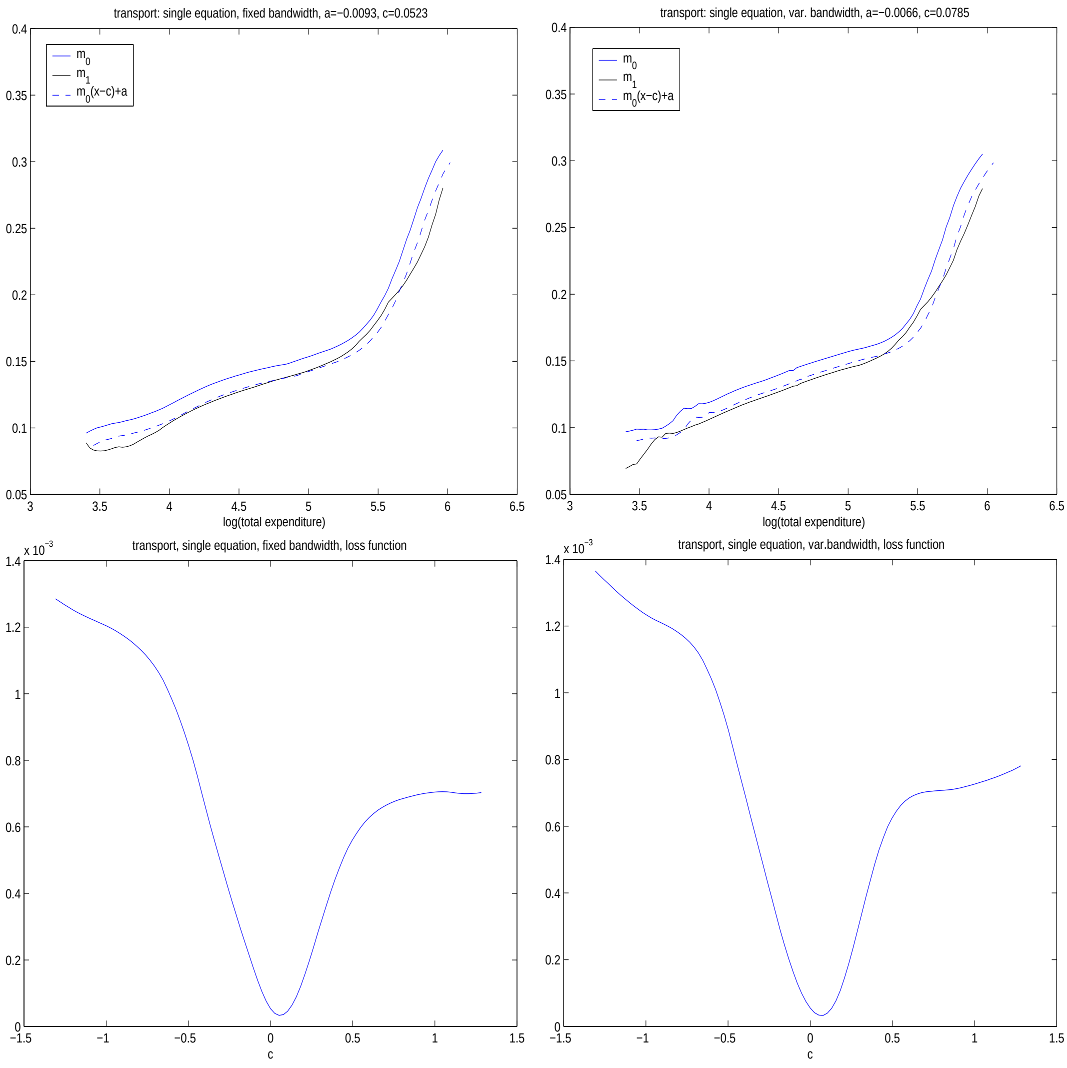

Figure 20: FES I: Transport expenditure share.

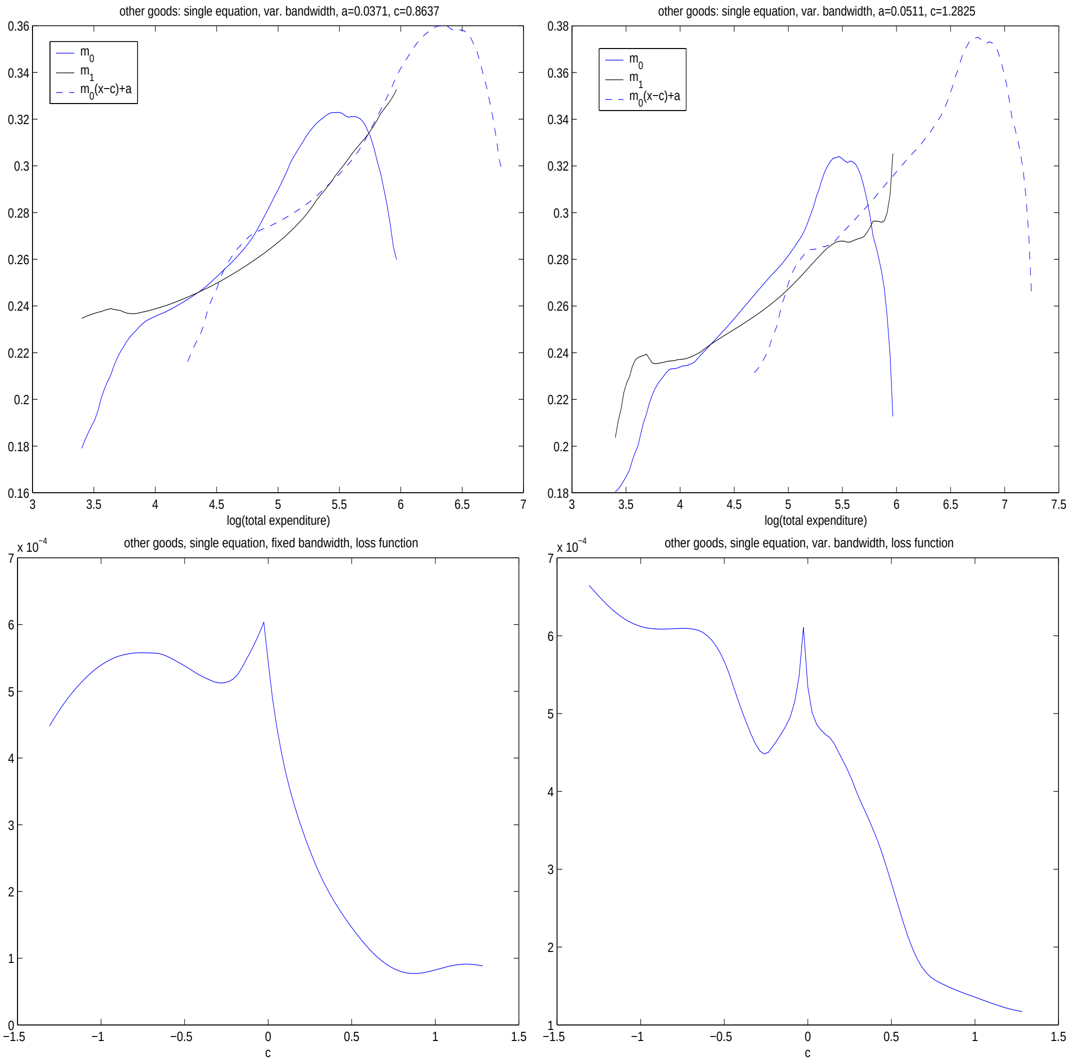

Figure 21: FES I: Other goods expenditure share.

Acknowledgements

I am very grateful to Hidehiko Ichimura for his supervision during my stay at University College London. 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. Thanks are also due to Francois Laisney and the seminar participants at UCL, Pittsburgh, Dortmund, ZEW Mannheim and ESEM 2002 for their comments. Financial support of the European Union (Marie-Curie training fellowship at University College London, Department of Economics) and of the German Research Foundation (Ph.D. scholarship at Dortmund University, Department of Economics) is gratefully acknowledged.

References

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

[2] Blundell, R., Browning, M. and Crawford, I., 2003. Nonparametric Engel Curves and Revealed Preferences. *Econometrica* Vol. 71, No.1, 205-240.

[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, 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] Newey, W.K., 1994. The asymptotic Variance of Semiparametric Estimators. *Econometrica* Vol. 62, No.6, 1349-1382.

[9] Newey, W.K. and McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics Vol.IV, eds. R.F. Engle and D.L. McFadden*, 2113-2245.

[10] 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*, 172-195.

[11] Wilke, R., 2003. Semiparametric Estimation of Consumption Based Equivalence Scales for Germany. *Mimeo*, ZEW Mannheim.