Essays on Economic and Econometric Modelling of Behavioral Heterogeneity in Demand Theory

Ralf Wilke

Universität Dortmund
Fachbereich für Wirtschafts- und Sozialwissenschaften
DFG Graduiertenkolleg
Allokationstheorie, Wirtschaftspolitik und kollektive Entscheidungen
Dissertation

Essays on Economic and Econometric Modelling of Behavioral Heterogeneity in Demand Theory

Ralf Andreas Wilke

Dortmund, 4.2002

Vorwort

Die Modellierung des individuellen Verhaltens der Subjekte einer Ökonomie ist in der Regel das Fundament der Wirtschaftstheorie, auf dem Aussagen abgeleitet werden. Der entsprechende Modellierungsansatz wird durch Vermutungen an jenes Verhalten begründet. Ergebnisse sind daher meistens das Resultat von Annahmen. Sie sind demnach nicht allgemein gültig, sondern halten nur in dem vorgegebenen Modellrahmen. Es ist in der Regel leicht zu zeigen, dass diese Ergebnisse nicht erzielt werden, wenn entsprechende Bedingungen verletzt sind. Ausserdem ist es meistens nicht schwierig zu sehen, dass Vermutungen über individuelles Verhalten in der Realität bestenfalls für eine Gruppe von Individuen oder Wirtschaftssubjekten erfüllt sind, jedoch fast sicher nicht für alle Individuen einer Population. Damit stellt sich die Frage, auf welche Berechtigung sich die Wirtschaftstheorie stützt, diese Vorgehensweise weiterzuführen.

Es soll und kann nicht das Ziel der vorliegenden Dissertation sein, eine Antwort auf diese Fragestellung zu liefern. Vielmehr sollte sie jeder Wissenschaftler selbst beantworten, seine Ziele definieren und entsprechende Methoden wählen oder erfinden, um diese zu erreichen. Deshalb sollte ein Wissenschaftler in zumindest den folgenden drei Kriterien über gewisse Fähigkeiten verfügen: Kritik, Innovation, Technik.

Diese Dissertation liefert vielmehr einige Beiträge zu einer interessanten Vorgehensweise, das angedeutete Dilemma der Wirtschaftstheorie zumindest in Teilen zu überwinden. Die zugrundeliegende Idee ist dabei nicht, das Verhalten jedes Individuums zu spezifizieren, sondern vielmehr die Bevölkerung als Ganzes zu betrachten und aufgrund von einer gewissen Heterogenität in der Bevölkerung Gesetzmässigkeiten für das durchschnittliche Verhalten abzuleiten. Bei der Modellierung spielen z.B. die Verteilung des verfügbaren Einkommens von Haushalten oder Unterschiede im Verhalten eine wichtige Rolle. Die Dissertation liefert Beiträge zur Methodik der Wirtschaftstheorie und der Ökonometrie, die einen Schritt ermöglicht hin zu angewandteren Problemstellungen. Der Rahmen dieser Beiträge ist die Nachfragetheorie. Er wurde gewählt, da es in diesem Bereich bereits eine Reihe von Forschungen gibt, die diese Art der Modellierung eingeführt und bearbeitet haben. Das Gebiet ist darüber hinaus interessant, da es hinreichend grosse Datensätze gibt, die komplexe ökonometrische Untersuchungen erlauben.

Diese Arbeit gliedert sich in zwei Teile. Der erste Teil umfasst drei Beiträge zur Wirtschaftstheorie. Der erste Beitrag fasst bestehende Ergebnisse der Nachfragetheorie zusammen. Dieser Überblick erstreckt sich von klassischen, auf Nutzenmaximierungsprinzipien beruhenden Ansätzen bis hin zu den neueren verteilungstheoretischen Ansätzen. Der zweite Beitrag untersucht, wie die Aggregation von beliebigen Bevölkerungen strukturelle Eigenschaften der aggregierten Nachfrage erzeugen kann. Dabei stellt sich in dem gewählten Rahmen heraus, dass durch die Aufteilung der Bevölkerung in homogene Teilbevölkerungen strukturelle Eigenschaften der aggregierten Nachfrage verloren gehen können. Im dritten Beitrag zur Wirtschaftstheorie werden in einem allgemeinen Modellrahmen einige Eigenschaften für die Verteilung von individuellem Verhalten und des Einkommens von Haushalten hergeleitet, sodass die durchschnittliche Nachfrage das *Law of Demand* erfüllt. Ausserdem werden die in der Literatur eingeführten Konzepte und Definition von Verhaltensheterogenität verglichen und es wird untersucht, welche Arten von Verhaltensheterogenität wie zu interpretieren sind. Ein neues Konzept der Verhaltensunterschiede wird eingeführt und so definiert, dass es messbar ist.

Der zweite Teil der Arbeit beschäftigt sich in zwei Beiträgen mit der ökonometrischen Modellierung von Verhaltensheterogenität. Es sollen Nachfragesysteme modelliert werden, die auf der einen Seite genügend Flexibilität besitzen, ein breites Spektrum an individuellem Verhalten zuzulassen, aber auf der anderen Seite nicht zu grosse Datensätze benötigen, um konkrete und präzise Aussagen zu treffen. Dieser Trade-Off wird mit Hilfe von semiparametrischen Schätzern durchbrochen, deren nichtparametrischer Teil genügend Flexibilität und deren parametrischer Teil eine hohe Konvergenzrate bietet. Im ersten Beitrag dieses Teils werden zwei Schätzer für Ausgabenanteilsfunktionen und Engel-Kurven vorgeschlagen, die nicht das konkrete Verhalten von Haushalten, sondern nur systematische Unterschiede zwischen homogenen Gruppen von Haushalten parametrisieren. Simulationen untersuchen die Leistungsfähigkeit dieser Schätzer in endlichen Stichproben unter richtigen Spezifikationen und unter Fehlspezifikationen. Im zweiten Beitrag wird ein weiterer Schätzer für die Schätzung von Engel-Kurven eingeführt, der mit bereits in der Literatur existierenden Schätzern vergleichbar ist. Simulationen bescheinigen dem neuen Schätzer eine überlegene Leistungsfähigkeit im Vergleich zu den existierenden Ansätzen. Die Bedingungen für die Konsistenz des Schätzer werden hergeleitet. Darüber hinaus

wird der Schätzer auf Konsumdaten angewendet. Die Ergebnisse zeigen die Relevanz dieser Methode für die angewandte Mikroökonometrie.

Diese Arbeit wurde im Rahmen des DFG Graduiertenkollegs Allokationstheorie, Wirtschaftspolitik und kollektive Entscheidungen an den Universitäten Dortmund und Bochum erstellt. Ein Teil der Beiträge wurde in den Workshops des Graduiertenkolleges an der Universität Dortmund, dem Workshop über Ökonomien mit heterogenen Agenten an der Universiteit Maastricht und dem Ökonometrie Seminar an der University of Pittsburgh vorgestellt. Ich bedanke mich bei allen Teilnehmern für ihre Kommentare. Insbesondere danke ich Marina Bauer, Dinko Dimitrov, Luis Gonzalez, Christian Kleiber, Walter Krämer, Wolfgang Leininger und Thomas Sparla für detaillierte Hinweise und Diskussionen. Ein Teil der Arbeit wurde im Rahmen eines Marie Curie Training Research Fellowships am University College London, Department of Economics, durchgeführt. Ich bedanke mich bei den Teilnehmern des Metrics Lunch Seminars und des Student Lunch Seminars für Ihre Kommentare. Besonderer Dank gilt hier Richard Blundell und Hidehiko Ichimura für die Betreuung eines Teiles der Arbeit und letzterem ausserdem für das intensive Training in Asymptotik. Ausserdem bedanke ich mich bei Ben Groom und Katrin Voss für die sprachliche Korrektur meiner Arbeit.

Inhaltsverzeichnis

Essay 1

A Selective Survey of Demand Theory

April 22, 2002

Abstract

This essay surveys recent approaches to the explanation of how the shape of mean demand is determined. The main focus is to address the following questions: Is it necessary to impose specific behavior of the households in order to achieve regularities of mean demand or is it sufficient to know that a certain degree of behavioral heterogeneity implies structural properties of mean demand, such as the Law of Demand? Firstly, results of classic, utility based, demand theory are presented. Secondly, newer approaches are shown which are essentially based on the fact that households behave heterogeneously. It is shown that both approaches may serve as a justification for structural properties of mean demand.

1 Introduction

Demand theory is one of the most investigated fields of economic theory. It has been of particular interest to find a realistic explanation for having a mean demand function satisfying some nice mathematical properties. Apart from the modelling of specific household behavior there are also approaches considering heterogeneity in behavior and characteristics as a source of generating structural properties of mean demand. This paper briefly reviews these two approaches to demand theory. The underlying mathematical concepts and assumptions are presented. The proofs can be found in the cited references.

The classic approach in economic theory is mainly based on a utility maximizing representative consumer. This ensures a convex mathematical problem both at the micro and at the macro level. Standard methods like Lagrange, Kuhn-Tucker and Euler can be applied either at the micro or at the macro level with an identical structure of the economic model. Solutions to these constrained optimization problems yield explicit values. Although very convenient, this kind of modelling is nevertheless only a rough approximation of the truth, since it neglects the heterogeneity of the population. In contrast, there are economists who do not assume explicit behavior at the micro level. They instead try to explain structural properties for aggregated demand by the behavioral heterogeneity of the population.

The paper is organized as follows: In Subsection 1.1, an economy is mathematically defined, which will be the basis for further analysis. Section 2 is concerned with utility based demand theory, and Section 3 shows that mean demand might also be influenced by behavioral heterogeneity. Section 4 summarizes, provides a brief critique and presents some ideas for future research.

1.1 The Economy

Suppose there are n households indexed by $i = 1, \ldots, n$ with M observable characteristics indexed by $m = 1, \ldots, M$. There are K goods indexed $k = 1, \ldots, K$ with prices $p_k \in \mathbb{R}_{++}$. Denote $p \in \mathbb{R}^{\mathrm{K}}_{++}$ as the vector of prices. It could also include interest rates and other economic variables, but we do not consider this case, as is common in theoretical analysis.

Assumption 1 *Each household i possesses a demand function $f_k(p, a_i)$ for each good k, where $f \in \mathcal{F}$ and $a_i = (a_{1i}, \ldots, a_{Mi}) \in \mathcal{A}^M \subset \mathbb{R}^{\mathrm{M}}$, i.e. $f_k : \mathbb{R}^{\mathrm{K}}_{++} \times \mathcal{A}^{\mathrm{M}} \mapsto \mathbb{R}_+$.*

The household characteristics a_{ji} include for example, the income, the number of children or cars owned by the household as well as the age of the household head. This assumption ensures a unique mapping of the individual behavior. Therefore, the households do not possess demand correspondences. The space $\mathcal{F}$ is the space of admissible demand functions. Different functional forms of the demand functions across the households are due to unobservable heterogeneity. Differences in the functional form of household demand functions could be interpreted as different demand behavior. Therefore, a population with somewhat different behavior at the micro level possesses at certain degree of behavioral heterogeneity.

For analysis, let us work with the expenditure share $w_k(p, a_i)$ and the consumption expenditure $c_k(p, a_i)$ of household i for good k:

$$\begin{aligned} c_k(p, a_i) &= p_k f_k(p, a_i) \\ w_k(p, a_i) &= c_k(p, a_i)/\text{income of household } i, \end{aligned}$$

where $w \in \mathcal{W}$ and $c \in \mathcal{C}$, the spaces of admissible functions f and w. I would like to emphasize that I did not restrict by any assumption the individual behavior, except for the existence of the demand function. Therefore, the spaces $\mathcal{F}$, $\mathcal{W}$ and $\mathcal{C}$ are only restricted by the uniqueness of the mappings.

In this paper I also treat the topic of aggregation over individuals. Aggregate demand or mean demand is defined by

$$F_k(p, a) := 1/n \sum_{i=1}^{n} f_k(p, a_i).$$

Similar definitions could be given for the mean expenditure share and the mean consumption expenditures. To simplify the notation in further analysis, I will sometimes omit the index i, indicating individual values.

2 Utility Maximization

Suppose that all households make their consumption decisions depending solely on the household's income, which is called $x \in \mathbb{R}_+$ for further analysis, and the price system p. Suppose also that all households have rational and continuous preferences, i.e. there are (direct) utility functions $u(q)$ for each household which depend positively on the quantity q of consumed goods and are concave in their argument.[1] The expenditure function $c(p, u)$ is the unique

[1]For further analysis, I omit the index i.

solution to the following optimization problem:

$$\min_q pq \text{ s.t. } u(q) \geq \bar{u},$$

where $\bar{u}$ is a reservation utility. Notice that the expenditure function is not identical to the consumption expenditure share. Of course, the so-called budget identity, $c(p,u) = x$, holds under these assumptions. Therefore, indirect utility functions $v(p,x) = \bar{u}$ exist.

With "Shepard's Lemma", one obtains the Hicksian demand function $h_k(p,u)$

$$\partial c(p,u)/\partial p_k = h_k(p,u) = q_k$$

and with "Roy's Identity" the (Marshallian) demand function

$$f_k(p,x) = -\frac{\partial v(p,x)/\partial p_k}{\partial v(p,x)/\partial x},$$

where $f \in \mathcal{F}_R$, the space of demand functions generated by transitive preferences. Transitivity of household preferences corresponds to rationality.

The expenditure share, $w_k(p,x)$, which in this case is equal to the budget share, is obtained by

$$\frac{p_k q_k}{x} = \frac{p_k}{c(p,u)} \frac{\partial c(p,u)}{\partial p_k} = \frac{\partial \ln x_k}{\partial \ln p_k} = w_k(p,x).$$

For the derivation of these results, see Deaton (1986) or a microeconomics textbook like Mas-Collel et al. (1995).

Gorman (1981) proves that budget shares generated by rational preferences, and therefore solutions to the differential equation above, necessarily have the specific functional form

$$w_k(p,x) = \sum_{r \in \mathbb{R}} \gamma_{kr} \phi_r(\ln x).$$

In other words the space of expenditure shares of rationally behaving households, $\mathcal{W}_R$, is restricted in some ways. Gorman shows, that the maximum rank of γ_{kr} is not larger than three and the functions $\phi_r(.)$ follow some further restrictions. Clearly $\mathcal{W}_R \subset \mathcal{W}$.

In the classic utility maximization framework, one can choose arbitrary functional forms for $u(q)$ that are continuous and increasing in its arguments. Engel curves[2], which are

[2]Engel curves were introduced by Ernst Engel in 1895. They show the relationship between consumption expenditure and household income. Nowadays, some economists call the relationship between expenditure share and income an Engel curve, too. The shape of these two different functions is obviously the same.

consistent with the Gorman conditions and which have some empirical evidence, are given by a popular specific form:

$$w_k(p,x) = A_k(p) + B_k(p)\mathrm{ln}x + C_k(p)g(x) \tag{1}$$

where $A_k(p)$, $B_k(p)$, $C_k(p)$ and $g(x)$ are differentiable functions. In applied analysis one often sets $g(x) = \sum_{m=1}^{M} \gamma_m(p)(\mathrm{ln}x)^m$.

Let me present some specific forms of these functions, which have been used in applied analysis:

- **PIGLOG:** In 1943 and 1963, Working and Leser suggested the following functional form

$$c_k(p,x) = \alpha_k x + \beta_k x(\mathrm{ln}x), \tag{2}$$

where α_k and β_k are parameters. This functional form is contained in the class of price-independent generalized logarithmic forms (PIGLOG):

$$w_k(p,x) = A_k(p) + B_k(p)\mathrm{ln}x = \frac{\partial \mathrm{ln}c(p,u)}{\partial \mathrm{ln}p_k} \tag{3}$$

where $c(u,p) = x$. The general solution to this differential equation is

$$\mathrm{ln}c(p,u) = u\mathrm{ln}B_k(p) + (1-u)\mathrm{ln}A_k(p)$$

where u varies with the households, depending upon wealth: "very poor (u=0) and very rich (u=1) respectively" (Deaton, 1986, p. 1775, l. 13).

- **PIGL:** A budget share like

$$w_k(p,x) = A_k(p) + B_k(p)\Big((1-x)^{-\alpha}\Big)\alpha^{-1} = \frac{\partial \mathrm{ln}c(p,u)}{\partial \mathrm{ln}p_k}$$

is of the price independent generalized linear form (PIGL). The general solution to this differential equation is

$$c(p,u)^{\alpha} = u(B_k(p))^{\alpha} + (1-u)(A_k(p))^{\alpha}$$

where PIGL becomes PIGLOG for $\alpha = 0$. For details, see Deaton-Muellbauer (1980).

- **GL:** The most general suggested specification is of the generalized linear form (GL):

$$w_k(p,x) = A_k(p) + B_{ki}(p) + C_k(p)g(p,x)$$

where $\sum_k C_k(p) = \sum_k B_{ki}(p) = \sum_i B_{ki}(p) = 0$ and $\sum_k A_k(p) = 1$. A PIGL budget share is contained in the GL class. In this special case, set $g(p,x) = (1 - x^{-\alpha})^{\alpha^{-1}}$, i.e. $g(p,x)$ is price independent[3].

From these three classes of Engel curves, it is possible to derive some parametric forms of demand functions, which can be empirically scrutinized:

1. Deaton and Muellbauer (1980) took budget shares of the PIGLOG form and specified $\ln A_k(p)$ and $\ln B_k(p)$ such that

$$\ln c(p,u) = a_0 + \sum_l \alpha_l \ln p_l + \ln\frac{1}{2}\sum_l \sum_j \gamma^*_{lj} \ln p_k \ln p_j + u\beta_0 \Pi_l p_l^{\beta_l}$$

where α_s, β_s and γ^*_{st} are parameters such that $c(p,u)$ is consistent with utility maximization theory, i.e. it is linear homogeneous in p. After some calculations one obtains the almost ideal demand system (AIDS):

$$w_k(p,x) = \alpha_k + \sum_j \gamma_{kj} \ln p_j + \beta_k \ln(x/P), \tag{4}$$

where P is a price index and $\gamma_{kj} = \frac{1}{2}(\gamma^*_{kj} + \gamma^*_{jk})$. Notice that this system has rank two. For further details, see Deaton-Muellbauer (1980).

2. A quadratic extension of the AIDS is given by Blundell, Pashardes and Weber (1993) and Banks, Blundell and Lewbel (1997). The latter show that exactly aggregable utility derived demand systems of rank three must have $g(x) = (\ln x)^2$ in (1) and $B_k(p)$ and $C_k(p)$ cannot both be price independent. Hence their quadratic almost ideal demand system (QUAIDS) is based on rank three Engel curves of the form

$$w_k(p,x) = A_k(p) + B_k(p)\ln x + C_k(p)(\ln x)^2.$$

Using information on the functional form of the indirect utility function and using specific translog forms for the unknown functions, similar to Deaton-Muellbauer (1980), they obtain the QUAIDS, which is

$$w_k(p,x) = \alpha_k + \sum_j \gamma_{kj} \ln p_j + \beta_k \ln(x/P_1) + \frac{\lambda_k}{P_2}\Big(\ln(x/P_1)\Big)^2 \tag{5}$$

[3] As an extension, the function $g_i(p,x)$ could also depend on an individual behavior parameter. For simplicity, I have set this constant equal to one. For further details, see Deaton-Muellbauer (1980).

where α_s, β_s and γ_{st} are parameters as in the AIDS, P_1 and P_2 form indices and λ_k is a new parameter, due to the third rank of the demand system.

Up to now I have derived and pointed out two different demand systems, given by (4) and (5), both of which are consistent with utility maximization theory and commonly used in applied analysis. Now I will switch to the aggregation framework of demand systems.
It is easy to aggregate these kinds of demand systems, as I will show for the QUAIDS. Taking into account that i is the index for households $i = 1, \ldots, n$, (5) then becomes

$$w_k^i(p, x_i) = \alpha_{ki} + \sum_j \gamma_{kj} \ln p_j + \beta_{ki} \ln(x_i/P_1) + \frac{\lambda_{ki}}{P_2} \Big(\ln(x_i/P_1) \Big)^2 .$$

The mean expenditure share for good k as defined by

$$W_k(p, x) = \frac{1}{n} \sum_i w_k^i(p, x_i)$$

can then be written as

$$W_k(p, x) = \mathrm{mean}_i \alpha_{ki} + \sum_j \gamma_{kj} \ln p_j + \mathrm{mean}_i \Big(\beta_{ki} \ln(x_i/P_1) \Big) + \mathrm{mean}_i \left(\frac{\lambda_{ki}}{P_2} \Big(\ln(x_i/P_1) \Big)^2 \right)$$

which has apparently the same structural form as the individual shares. The reason for this is its linearity in the household specific parameters. Blundell et al. (1993) extend this approach by suggesting to write individual behavior parameters α_{ki}, β_{ki} and λ_{ki} as a linear combination of alternative household characteristics a_i, like age and sex. The approach of Banks et al. (1997) points in the same direction. For further details, see their papers. In a recent survey, Blundell and Stoker (2000) are concerned with impacts of the household income distribution on aggregate demand. They show that this kind of heterogeneity affects the shape of rank two and rank three demand systems.

3 Behavioral Heterogeneity

Let us now focus on the second theoretical approach, the modelling of behavioral heterogeneity, which is quite different to what we have seen in Section 2. For one thing, cancel the assumptions I made in Section 2. In the following section, I do not state any assumption about individual behavior of households. I will work with a much wider class of admissible demand and, therefore, expenditure share functions. Distributions of individual characteristics and heterogeneous behavior will play an important role. It is shown that structural properties of aggregate demand can be induced by behavioral heterogeneity and not only by imposing some regularity conditions on household behavior. This section is based on two articles: Kneip (1999) and Hildenbrand and Kneip (1999).

Kneip (1999) Kneip considers a continuum of households with expenditure shares $w \in \mathcal{W}$ and household characteristics $a \in \mathcal{A}^M$. Firstly, he states some technical assumptions:

Assumption 2 *There exists a continuous density $\mu(a) : \mathcal{A}^M \mapsto [0,1]$ for the distribution of individual characteristics a.*

This assumption is restrictive in one way because some of the observable characteristics, for example age, are discrete numbers. To simplify the theoretical analysis, we assume the densities are continuous. This does not decrease the generality of the economy. Note that Kneip only considers a single individual characteristic, $M = 1$, namely disposable household income. I therefore here consider an extended framework with more than one individual characteristic.

Assumption 3 *1. For some $\gamma \in \mathbb{R}_+$ and some subspace $I \subset [0,\gamma]^K$, the space of admissible expenditure share functions is a subset of the set $\mathcal{V}(I)$ of all functions from $\mathbb{R}^{\mathrm{K}}_+ \times \mathbb{R}_+$ to I. There exists a Lebesgue measure on I.*

2. The space $\mathcal{W}$ is large enough such that for any $w \in \mathcal{W}$ and all $\Delta \in \mathbb{R}^{\mathrm{K}}_{++}$, $\Theta \in \mathbb{R}^{\mathrm{M}}_+$ the function v satisfying

$$v(p,x) = w(\Delta * p, \Theta * a) \text{ for all } p \in \mathbb{R}^{\mathrm{K}}_{++} \text{ and } \mathrm{a} \in \mathcal{A}^{\mathrm{M}},$$

is also an element of $\mathcal{W}$. Furthermore, for all p and a the set $\{w(p,a) \in I | w \in \mathcal{W}\}$ is Lebesgue measureable.

These conditions are of a technical nature. They do not affect an economist's point of view. In addition to the previous assumptions we need a precise definition of a probability space for ν.

Assumption 4 *1. The distribution ν is a probability measure on the σ-algebra $\mathcal{A}_{\mathcal{W}}$ of $\mathcal{W}$.*

2. For every continuous function $v : I \mapsto \mathrm{I\!R}$ the integral $\int_I v(z)\nu_{p,a}(dz)$ is differentiable with respect to p and a.

The first condition makes use of $\mathcal{A}_{\mathcal{W}}$, the smallest σ-algebra of $\mathcal{W}$, containing all sets of the form $A = \{w \in \mathcal{W} | w(p,a) \in J_{p,a}\}$, where $\{J_{p,a}\}_{p,a \in \mathrm{I\!R}^{\mathrm{K}}_{++} \times \mathcal{A}^{\mathrm{M}}}$.
As a consequence of these assumptions, the following integral exists

$$W_\nu(p,a) = \int_{\mathcal{W} \times \mathcal{A}^M} w(p,a) d\nu \mu(a) da.$$

In other words, one has a convenient tool with which to write the mean expenditure share.

Example 1 *Suppose that the budget identity holds for all households. The space I then restricts to $\{z \in \mathrm{I\!R}^{\mathrm{K}}_{++} | \sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{z_k} = 1\}$. Realizations on I are observable, hence also the distribution $\nu_{p,a}$ of the z_k's. Keeping that in mind, the above integral for fixed p, a becomes*

$$W_\nu(p,a) = \int_{\{z|z \in I\}} z\nu_{p,a}(dz) = \int_{\{z|z \in I\}} z\phi_{p,a}(z)dz$$

where for fixed p, a, $\nu_{p,a}$ is the k-dimensional distribution of z on I with the density $\phi_{p,a}(z)$.

For further analysis I also need a precise formulation of demand behavior.

Definition 1 *Consider two different households i and j. They possess a similar demand behavior, if $sup_{p,a} \| w^i(p,a) - w^j(p,a) \|_2$ is small, where $\| * \|_2$ is the Euclidean distance.*

Of course, similar demand behavior does not imply similar demand. With this definition Kneip skilfully uses the relative nature of share functions. For some examples, see Kneip (1999). Thus, a flat distribution of ν on $\mathcal{W}$ can be considered as behavioral heterogeneity. A uniform distribution can be interpreted as extreme heterogeneity. In other words, extreme heterogeneity then induces equal probability for equal sized sets. One can express this by the following relationship, which should hold for all Borel sets $J \subset I$:

$$\nu_{p,a}(J) = \nu_{\Delta * p, \Theta * a}(J)$$

for fixed p and x, where the right hand side is a distance preserving transformation of the left hand side with $(\Delta, \Theta) \in \mathrm{IR}^{\mathrm{K}}_{++} \times \mathcal{A}^{\mathrm{M}}$. For details, see Kneip (1999).
It is expected that the equation above does not hold, but $\nu_{p,a}(J) \approx \nu_{\Delta * p, \Theta * a}(J)$ for $(\Delta, \Theta) \approx (\mathbf{1}, \mathbf{1})$ could be reasonable. In addition, one can equivalently say that for any measurable subset $J \subset I$ the partial derivatives $\partial_{\Delta_r} \nu_{\Delta * p, \Theta * a}(J)|_{(\Delta,\Theta)=(\mathbf{1},\mathbf{1})}$ are close to zero, where $r = 1, \ldots, K, K+1, \ldots, K+M$ and $\Delta_{K+1} = \Theta_1, \ldots, \Delta_{K+M} = \Theta_M$.

Definition 2 *Let $\mathcal{C}(I, [0, \gamma])$ denote the space of all continuous functions from I into $[0, \gamma]$. A measure of the structural stability of mean demand is*

$$h(v) = max_r sup_{p,a} h_{p,a;r}(v),$$

the coefficient of sensitivity, where

$$h_{p,a;r}(v) = sup_{v \in \mathcal{C}(I,[0,\gamma])} \left| \partial_{\Delta_r} \left(\int_I v(z) \nu_{\Delta * p, \Theta * a}(dz) \right) |_{(\Delta,\Theta)=(\mathbf{1},\mathbf{1})} \right|.$$

The coefficient is small in the case of behavioral heterogeneity and in the case of invariant demand behavior. This is identified as structural stability.

Definition 3 *Mean demand is structurally stable, if*

- *the household expenditure shares are independent of household characteristics or prices respectively, or*

- *the population is very heterogeneous in its behavior.*

In addition, the coefficient of sensitivity becomes smaller with a combination of these two points. For an alternative definition, see Hildenbrand and Kneip (1997) and Hildenbrand and Kneip (1996). They consider a population over time and conclude that mean demand is structurally stable if the distributions of household characteristics are invariable over time. Kneip (1999) proves that a small coefficient of sensitivity implies certain structural properties of mean demand:

Proposition 1 *The Jacobian matrix of mean demand with respect to prices is negative definite for all prices $p \in \mathrm{IR}^{\mathrm{K}}_{++}$ if the coefficient of sensitivity $h(\nu)$ is sufficiently small and if there is a constant $c > 0$ such that $\int_{\mathcal{W}} w_k(p, a) d\nu \geq c$.*

Therefore, the *Law of Demand* holds for mean demand. In addition, Kneip shows negative semi-definiteness of the Slutsky substitution matrix of aggregate demand and positive definiteness of the Slutsky income matrix, hence the *Slutsky decomposition* for utility derived demand theory holds for the mean demand, if and only if the coefficient is small enough. We obtain rational mean behavior by a sufficient degree of structural stability.

Example 2 *Cobb-Douglas. Suppose that the household demand functions are given by* $f_k^i(p, x_i) = \beta_k^i x_i/p$*, where* x_i *denotes the income of each household and* $\sum_k \beta_k^i = 1$ *for all* i*. There is equivalence of the distribution* $\nu_{p,a}$ *and the distribution of* β*, since* $w_k^i(p, x_i) = \beta_k^i$*, and therefore* $h(\nu) = 0$*.*

More generally, Kneip shows that small derivatives of all household budget share functions imply a small coefficient of sensitivity $h(\nu)$.

Hildenbrand and Kneip (1999) The authors derive an index that measures the degree of behavioral heterogeneity of a population. In addition, they show that structural properties of mean demand, such as negative definiteness of its Jacobian matrix, are generated by heterogeneity in the households' behavior. In contrast to former approaches, i.e. Hildenbrand (1993) and Kneip (1999), they choose a decomposition of mean demand that requires less restrictive conditions on individual behavior than the Slutsky decomposition.

Suppose that the population H consists of $i = 1, \ldots, n$ households. Each household possesses a demand function f^i depending on income $x_i > 0$ and prices p, i.e. $f_k(p, a_i) := f_k^i(p, x_i)$. The specific behavior of a household is therefore completely described by the household's characteristics (f^i, x_i). The demand function of a household for good k given prices p and income x_i is $f_k^i(p, x_i)$, where $k = 1, \ldots, K$ and $p \in (0, \infty)^K$. Let X denote mean income of the population H

$$X = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Accordingly, the mean expenditure share for good k is defined as:

$$W_k(p) := \frac{p_k}{X} F_k(p).$$

Now, let us define $S_{kl}(p)$, the rate of change of $W_k(p)$ with respect to a percentage change of the price p_l as

$$S_{kl}(p) := \partial_\lambda W_k(p_1, \ldots, \lambda p_l, \ldots, p_K)|_{\lambda=1} = p_l \partial_{p_l} W_k(p)$$

and accordingly for $w_k^i(p, x_i)$ one obtains

$$s_{kl}^i(p, x_i) := \partial_\lambda w_k^i(p_1, \ldots, \lambda p_l, \ldots, p_K, x_i)|_{\lambda=1} = p_l \partial_{p_l} w_k^i(p, x_i).$$

Moreover, the upper bound for $|s_{kl}^i(p,x_i)|$ with respect to prices is given by

$$d_{kl}^i := \sup_p |s_{kl}^i(p,x_i)| = \sup_p p_l |\partial_{p_l} w_k^i(p,x_i)|$$

for all i.

The following three assumptions on the budget share functions are required in order to derive the main results (see Hildenbrand and Kneip, 1999):

Assumption 5 $0 \leq w_k^i(p,x_i) \leq 1$

Assumption 6 *$w_k^i(p,x_i)$ is continuously differentiable in p and x_i and $d_{kl}^i < \infty$ for all i and k.*

Assumption 7 *For all k,l, and $\bar{p} \in (0,\infty)^K$, the derivative of the function $p_l \mapsto w_k(\bar{p}_1,\ldots,p_l,\ldots,\bar{p}_K)$ changes its sign at most m times, where m is a positive integer.*

The domain of prices for which $p_l|\partial_{p_l} w_k^i(p,x_i)| \geq \epsilon d_{kl}^i$ is defined as

$$A_{kl}^{\epsilon}(w^i,x_i) := \{p \in (0,\infty)^K |\ p_l|\partial_{p_l} w_k^i(p,x_i)| \geq \epsilon d_{kl}^i\}$$

where $\epsilon \in [0,1]$.

In order to derive the measure of the degree of behavioral heterogeneity, let us clarify what Hildenbrand and Kneip (1999) mean by behavioral heterogeneity: $A_{kl}^{\epsilon}(w^i,x_i)$ are located in different regions of the price system in $(0,\infty)^K$ for different i. Accordingly, they define an intersection ratio

$$I_{kl}^{\epsilon}(p) := \frac{1}{n}\#\{i \in H | p \in A_{kl}^{\epsilon}(w^i,x_i)\}$$

which indicates whether the sets $A_{kl}^{\epsilon}(w^i,x_i)$ differ in i. Note that $I_{kl}^{\epsilon}(p)$ is a decreasing step function in ϵ. A high degree of behavioral heterogeneity implies a low intersection ratio. Then $\int_0^1 I_{kl}^{\epsilon}(p)d\epsilon$ is close to zero. Taking this into account, let us now define the degree of behavioral heterogeneity of the household population:

Definition 4 *The degree of behaviorial heterogeneity of a population H is measured by the index of heterogeneity $\gamma(H)$, where*

$$\gamma(H) := inf_{k,l,p}\gamma_{kl}(p) = 1 - sup_{k,l,p}\int_0^1 I_{kl}^{\epsilon}(p)d\epsilon.$$

Note that $0 \leq \gamma(H) \leq 1 - \frac{1}{n} < 1$.

Hildenbrand and Kneip (1999) show that a sufficiently small index of heterogeneity implies the law of demand for mean demand, although this property was not assumed for household demands:

Proposition 2 *A sufficiently high degree of behavioral heterogeneity implies negative diagonal dominance of the Jacobian matrix of mean demand with respect to prices if there is a constant $c \geq 0$ such that $d^i_{kl} \leq c$ for all k, l, i.*

Since small derivatives of the household expenditure shares also imply the law of demand, the following proposition can be proved:

Proposition 3 *The smaller the variability of the household expenditure shares, i.e. d^i_{kl} is small for all i, the smaller the degree of behavioral heterogeneity $\gamma(H)$ of the population H has to be in order to have negative diagonal dominance of the Jacobian matrix of mean demand with respect to prices.*

In other words both components work in the same direction. This combination should also be reasonable for a real economy. Nevertheless let us now illustrate the two extreme cases with the help of two examples:

Example 3 *Cobb-Douglas. Suppose again that all household demand functions have the form $f^i_k(p, x_i) = \beta^i_k x_i/p$, where $\sum_k \beta^i_k = 1$ for all i. Then $d^i_{kl} = 0$ and clearly $\gamma(H) = 0$.*

Example 4 *Suppose we have Grandmont's (1992) economy, in which the budget share functions are of the form $w^\alpha(p, x) = w(\alpha * p, x)$ for some $\alpha \in (0, \infty)^K$ where $\alpha * p = (\alpha_1 p_1, \ldots, \alpha_K p_K)$. In this economy $d_{kl} = d^\alpha_{kl}$ for $\alpha \in (0, \infty)^l$ holds. Obviously, the Grandmont economy is a special case in which heterogeneity is solely expressed by the distribution of the parameter α. Hildenbrand and Kneip (1999) have shown that $I^\epsilon_{kl}(p) = \nu\{B^\epsilon_{kl}(w^1, x) - log\ p\}$, where ν denotes the distribution of $log\ \alpha$ and $B^\epsilon_{kl}(w^1, x) := \{log\ p \in \mathrm{IR}^{\mathrm{K}} |\ |\partial_{log\ \mathrm{p_l}} \mathrm{w^1_k(p, x)}| \geq \epsilon \mathrm{d_{kl}}\}$. Hence, if the distribution ν is sufficiently uniformly spread, the intersection frequency $I^\epsilon_{kl}(p) := \nu^\epsilon_{kl}(p)$ becomes arbitrarily small. Then $\gamma(H)$ is close to 1.*

4 Summary and Criticism

Two different approaches to the modelling of mean demand have been surveyed. The classic, utility derived, approach uses well known results of microeconomics. There is a wide range of contributions in the literature dealing with this topic. In applied analysis, the resulting demand systems are estimable with parametric statistical methods. This has often been done, in many cases the systems fitted the data well. However, there are methodological problems imbedded in this approach. To state restrictive assumptions on the behavior of the households is difficult to justify. Moreover, it seems unrealistic that demand decisions of households are made independently of individual characteristics, except income. As a matter of fact there are some articles allowing for individual components but these approaches are artificial because they only exploit the freedom given by Gorman's Engel curve specification, for some differentiable functions. Nevertheless, the fit of aggregate data for Gorman form Engel curves can be considered a stylized fact. In contrast this fit at the macro level does not allow welfare analysis at the micro level as in Banks et al. (1997). Analyses of this kind and expressions of the form "*richer people have higher utility than poorer people*", as in Deaton (1986) show the methodological dilemma of the classic approach at the micro level. Nevertheless, the classical approach can be used as the theoretical foundation for empirical analysis at the macro level. However, semiparametric estimation seems to be a good compromise between an adequate rate of convergence of the estimators and a moderate risk of misspecification.

The second approach presented allows for more generality at the micro level. Here, it is not as necessary to describe completely each individual in order to explain structures of aggregate demand, as it is in the first approach. Indeed, it is possible to achieve compatible results for mean demand, e.g. if the sensitivity coefficient or the index of heterogeneity is small, without making restrictive assumptions on the behavior of the households, e.g. restrictions on the class of feasible expenditure share functions. As a matter of fact, it is difficult to scrutinize how structurally stable mean demand is. It is impossible to find out empirically how large the index of heterogeneity is for a given population. Moreover, behavioral heterogeneity is not uniquely defined by the authors.

There are only a few contributions treating the empirical evidence of this approach, e.g. Hildenbrand and Kneip (1996, 1999). The econometric modelling of behavioral heterogeneity is still in the first steps of development. Due to an increase of computational power during the past few years and significant progress in non- and semiparametric statistics, there are devices and tools now available which can play a key role in econometric modelling of behavioral heterogeneity in demand theory.

Finally, I conclude that from a theoretical point of view, there is no contradiction in the results of the two approaches presented. The statistician should use the advantages of both in order to build a model to be estimated.

References

[1] Banks, J. and Blundell, R. and Lewbel, A. (1997). Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics*, Vol. LXXIX (No. 4): 527–539.

[2] Blundell, R. and Pashardes, P. and Weber, G. (1993). What do we Learn about Consumer Demand Patterns from Micro Data. *American Economic Review*, Vol. 83 (No. 3): 570–597.

[3] Blundell, R. and Stoker, T. (2000). Models of Aggregate Economic Relationships that account for Heterogeneity. *Handbook of Econometrics V*, North-Holland, Amsterdam.

[4] Deaton, A (1986). Demand Analysis. *Handbook of Econometrics III*, North-Holland, Amsterdam.

[5] Deaton, A and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, Vol. 70: 312-336.

[6] Grandmont, J.M. (1992). Transformation of the Commodity Space, Behavioral Heterogeneity, and the Aggregation Problem. *Journal of Economic Theory*, 57(1): 1–35.

[7] Gorman, W. (1981). Some Engel Curves. In: *The Theory and Measurement of Consumer Behavior*, Angus Deaton (Ed.), Cambridge University Press, Cambridge.

[8] Hildenbrand, W. (1993). *Market Demand*. Princeton University Press, Princeton.

[9] Hildenbrand, W. and Kneip, A. (1996). Modelling Aggregate Consumption Expenditure and Income Distribution Effects. SFB 303 Discussion Paper No. A-510, Universität Bonn.

[10] Hildenbrand, W. and Kneip, A. (1999). Demand Aggregation under Structural Stability. *Journal of Mathematical Economics*, 31: 81-109.

[11] Hildenbrand, W. and Kneip, A. (1999). On Behavioral Heterogeneity. SFB 303 Discussion Paper No. A-589, Universität Bonn.

[12] Kneip, A. (1999). Behavioral Heterogeneity and Structural Properties of Aggregate Demand. *Journal of Mathematical Economics*, 31: 49-74.

[13] Leser C.E.V. (1963). Forms of Engel-Functions. *Econometrica*, Vol. 31: 694-703.

[14] Mas Colell, A., Whinston, Green (1995). *Microeconomic Theory.* Oxford University Press, Oxford.

[15] Working, H. (1943). Statistical Laws of Family Expenditure. *Journal of the American Statistical Association*, Vol. 38: 43-56.

Essay 2

A Note on Behavioral Heterogeneity and Aggregation

April 22, 2002

Abstract

The purpose of this note is to investigate how aggregation of households affects the variation of the index of heterogeneity as recently defined in Hildenbrand and Kneip (1999). We show the degree of behavioral heterogeneity of an entire population is at least as high as the smallest degree of behavioral heterogeneity of some disjoint subpopulation. We further derive conditions under which aggregation increases the degree of behavioral heterogeneity. Finally, we show that aggregation always weakly increases the degree of behavioral heterogeneity. Therefore we offer a theoretical framework that helps to answer the question of how the structural properties of aggregate demand are obtained due to aggregation.

1 Introduction

In a recent paper Hildenbrand and Kneip (1999) develop an index that measures the degree of behavioral heterogeneity of a population. It is well known that structural properties of aggregate demand, such as negative definiteness of the Jacobian matrix, are generated by heterogeneity in household's behavior. In particular, Hildenbrand and Kneip (1999) show that a sufficiently large index implies the law of demand for mean demand. In contrast to former approaches, i.e. Hildenbrand (1993) and Kneip (1999), these authors have chosen a decomposition of aggregate demand that requires less restrictive conditions on individual behavior than the Slutsky decomposition. In addition, their work is a generalization of Grandmont's (1992) economy.

The aim of this note is to derive further properties of the index of heterogeneity. We mainly focus on the question of what happens to the structural properties of mean demand if we aggregate subpopulations, i.e. we investigate the impact of aggregation on the degree of behavioral heterogeneity. We present three results: Firstly, aggregation never decreases the degree of behavioral heterogeneity. In other words, the degree of behavioral heterogeneity at the aggregate level is higher than the lowest degree of behavioral heterogeneity of arbitrary disjoint subpopulation. Secondly, we show that the degree of behavioral heterogeneity at the aggregate level can be either greater or smaller than the maximal degree of heterogeneity of all subpopulations. Since we intend to investigate the impacts of behavioral heterogeneity on the structural properties of aggregate demand, we derive conditions for generating heterogeneity due to aggregation. Thirdly, and finally, we show that aggregation always weakly generates heterogeneity. In other words, aggregation leads to a higher degree of behavioral heterogeneity at the aggregate level when compared to the weighted average of arbitrary disjoint subpopulations.

Notation We use the same notation as in Hildenbrand and Kneip (1999).

Suppose that every household $h \in H$ with income $x^h > 0$ has a demand function f^h. The specific behavior of a household is completely described by the household's characteristics (f^h, x^h). The demand function of a household for good i given prices p and x^h is $f_i^h(p, x^h)$, where $i = 1, \ldots, l$ and $p \in (0, \infty)^l$. Let X denote mean income of the population and let

$$F(p) := \frac{1}{\#H} \sum_{h \in H} f^h(p, x^h) \in \mathbb{R}^l_+$$

denote mean demand. The household expenditure share is defined as

$$w_i^h(p, x^h) := \frac{p_i}{x^h} f_i^h(p, x^h)$$

and analogously we obtain the mean expenditure share for good i:

$$W_i(p) := \frac{p_i}{X} F_i(p).$$

Now, consider the rate of change of $W_i(p)$ and $w_i^h(p, x^h)$ with respect to a percentage change of the price p_j which is defined as

$$S_{ij}(p) := \partial_\lambda W_i(p_1, \ldots, \lambda p_j, \ldots, p_l)|_{\lambda=1} = p_j \partial_{p_j} W_i(p)$$

and

$$s_{ij}^h(p, x^h) := \partial_\lambda w_i^h(p_1, \ldots, \lambda p_j, \ldots, p_l, x^h)|_{\lambda=1} = p_j \partial_{p_j} w_i^h(p, x^h).$$

Moreover, the upper bound for $|s_{ij}^h(p, x^h)|$ with respect to prices is given by

$$d_{ij}^h := \sup_p |s_{ij}^h(p, x^h)| = \sup_p p_j |\partial_{p_j} w_i^h(p, x^h)|$$

for all $h \in H$.

The following three assumptions on the budget share functions are required to derive the following results, see Hildenbrand and Kneip (1999):

Assumption 1 $0 \leq w_i(p, x) \leq 1$

Assumption 2 *$w(p, x)$ is continuously differentiable in p and x and $d_{ij}^h < \infty$ for all $h \in H$.*

Assumption 3 *For all i, j, and $\bar{p} \in (0, \infty)^l$, the derivative of the function $p_j \mapsto w_i(\bar{p}_1, \ldots, p_j, \ldots, \bar{p}_l)$ changes its sign at most m times, where m is a positive integer.*

The domain of prices for which $p_j |\partial_{p_j} w_i^h(p, x^h)| \geq \epsilon d_{ij}^h$ is defined as

$$A_{ij}^\epsilon(w^h, x^h) := \{p \in (0, \infty)^l |\; p_j |\partial_{p_j} w_i^h(p, x^h)| \geq \epsilon d_{ij}^h\}$$

where $\epsilon \in [0, 1]$.

In order to derive the definition of the degree of behavioral heterogeneity, let us clarify what it means in the framework of Hildenbrand and Kneip (1999): $A_{ij}^\epsilon(w^h, x^h)$ are located in different regions of the price system in $(0,\infty)^l$ for different h. Accordingly, they define an intersection ratio

$$I_{ij}^{\epsilon}(p) := \frac{1}{\#H}\#\{h \in H | p \in A_{ij}^{\epsilon}(w^h, x^h)\}$$

which indicates whether the sets $A_{ij}^{\epsilon}(w^h, x^h)$ differ in h. Note that $I_{ij}^{\epsilon}(p)$ is a decreasing step function in ϵ. A high degree of behavioral heterogeneity implies a low intersection ratio. Then $\int_0^1 I_{ij}^{\epsilon}(p)d\epsilon$ is close to zero. Taking this into account we can define the degree of behavioral heterogeneity of the household population H as

$$\gamma(H) := \inf_{i,j,p}\gamma_{ij}(p) = 1 - \sup_{i,j,p}\int_0^1 I_{ij}^{\epsilon}(p)d\epsilon.$$

Note that $0 \leq \gamma(H) \leq 1 - \frac{1}{\#H} < 1$.

In addition, define $G_{ij}^{\epsilon}(p) := 1 - I_{ij}^{\epsilon}(p)$. Then

$$G_{ij}^{\epsilon}(p) = \frac{1}{\#H}\#\{h \in H || s_{ij}^h(p, x^h)| < \epsilon d_{ij}^h\}$$

is the cumulative distribution function of $|s_{ij}^h(p, x^h)|/d_{ij}^h$.

2 Behavioral Heterogeneity and Aggregation

In this section, we derive some properties of the index of heterogeneity, while taking into account the aggregation over subpopulations.

Aggregation Over Subpopulations. Consider $m = 1, \ldots, k$ nonempty subpopulations of H with $\dot{\bigcup}_{m=1}^k H^m = H$, where $\dot{\bigcup}$ denotes the union of disjoint sets. The disjoint subpopulations are allowed to be of arbitrary size and arbitrary composition.

Definition 1 *Aggregation reduces heterogeneity as measured by γ, if*

$$\gamma(H) < inf_m \gamma(H^m)$$

is true.

Definition 2 *Aggregation increases heterogeneity as measured by γ, if*

$$\gamma(H) \geq sup_m \gamma(H^m)$$

is true.

Proposition 1 *Aggregation cannot reduce the degree of behavioral heterogeneity as measured by* γ*, i.e. for every* H *and* $\{H^m\}_{m=1,\ldots,k}$ *such that* $\dot{\bigcup}_{m=1}^{k} H^m = H$ *it follows*

$$\gamma(H) \geq inf_m \gamma(H^m).$$

Proof. It suffices to prove the proposition for $k = 2$, since $H_k \cup \left(\dot{\bigcup}_{m=1}^{k-1} H^m\right) = H$. Suppose we have two subpopulations m and n and assume without loss of generality that $\gamma(H^m) \leq \gamma(H^n)$. Then,

$$\begin{aligned} I_{ij}^{\epsilon}(p) &= \frac{1}{\#H^m + \#H^n} \#\{h \in H^m \cup H^n | p \in A_{ij}^{\epsilon}(w^h, x^h)\} \\ &= \frac{1}{\#H^m + \#H^n} \left(\#H^m I_{ij}^{\epsilon m}(p) + \#H^n I_{ij}^{\epsilon n}(p)\right) \\ &\leq \sup\{I_{ij}^{\epsilon m}(p), I_{ij}^{\epsilon n}(p)\} \end{aligned} \tag{1}$$

for $\epsilon \in [0,1]$, where

$$I_{ij}^{\epsilon m}(p) := \frac{1}{\#H^m} \#\{h \in H^m | p \in A_{ij}^{\epsilon}(w^h, x^h)\}.$$

Applying this inequality gives

$$\begin{aligned} 1 - \gamma_{ij}(p) = \int_0^1 I_{ij}^{\epsilon}(p) d\epsilon &\leq \int_0^1 \sup\{I_{ij}^{\epsilon m}(p), I_{ij}^{\epsilon n}(p)\} d\epsilon \leq 1 - \inf\{\gamma_{ij}^m(p), \gamma_{ij}^n(p)\} \\ \Leftrightarrow \gamma_{ij}(p) &\geq \inf\{\gamma_{ij}^m(p), \gamma_{ij}^n(p)\} \end{aligned}$$

for all i, j and $p \in (0, \infty)^l$, which proves Proposition 1. ■

Note, however, that we can neither infer from Proposition 1 that an expansion of the population by an additional household does not lead to a decrease in the index of heterogeneity nor that $\gamma(H) < \sup_m \gamma(H^m)$ holds. More generally, we ask whether $\gamma(H) \geq \sup_m \gamma(H^m)$ may occur. We show that this inequality does not hold in general: because of equation (1) one can infer $I_{ij}^{\epsilon}(p) \geq \inf\{I_{ij}^{\epsilon m}(p), I_{ij}^{\epsilon n}(p)\}$ (see Figure 1). Therefore

$$1 - \gamma_{ij}(p) = \int_0^1 I_{ij}^{\epsilon}(p) d\epsilon \geq \int_0^1 \inf\{I_{ij}^{\epsilon m}(p), I_{ij}^{\epsilon n}(p)\} d\epsilon \leq 1 - \sup\{\gamma_{ij}^m(p), \gamma_{ij}^n(p)\},$$

which is not a unique relation.

The next propositions shed light on this point. Proposition 2 looks at an expansion of the original population by an additional household, while Proposition 3 considers the general case when aggregating subpopulations of arbitrary size.

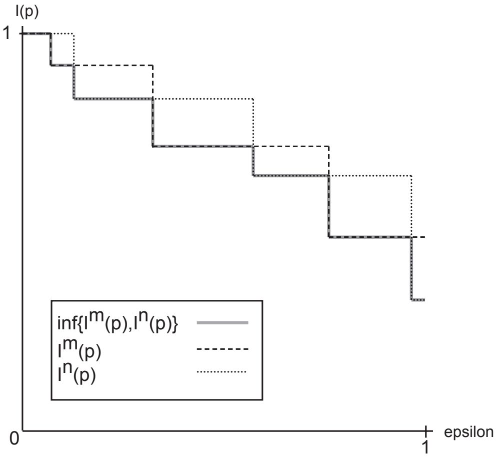

Figure 1: $\int_0^1 \inf\{I_{ij}^{\epsilon m}(p), I_{ij}^{\epsilon n}(p)\} d\epsilon \leq 1 - \sup\{\gamma_{ij}^m(p), \gamma_{ij}^n(p)\}$.

The Intruder's Influence. We are looking at the degree of heterogeneity while expanding the original population H by one additional household. Let $H^+ = H \cup H^1$ where H^1 consists of one household only, the intruder. Let $\mathbb{I}$ denote the set of (i, j, p) such that $\gamma(H) = 1 - \int_0^1 I_{ij}^\epsilon(p) d\epsilon$. Note that $\mathbb{I}$ may contain of more than one element.

Proposition 2 *Increasing the size of the population by one additional household leads to $\gamma(H^+) \geq \gamma(H)$ if*

$$c_{ij}^1(p) := \frac{p_j |\partial_{p_j} w_i^1(p, x^1)|}{d_{ij}^1} \leq \int_0^1 I_{ij}^\epsilon(p) d\epsilon \quad \textit{for all } (i, j, p) \in \mathbb{I}$$

and

$$\int_0^1 I_{ij}^\epsilon(p) d\epsilon \geq \frac{1}{(1 + \#H)} + \int_0^1 I_{\tilde{i}\tilde{j}}^\epsilon(\tilde{p}) d\epsilon \quad \textit{for all } (\tilde{i}, \tilde{j}, \tilde{p}) \in C\mathbb{I},$$

where $C\mathbb{I}$ denotes the complementary set of $\mathbb{I}$.

The first condition ensures that $1 - \int_0^1 I_{ij}^{\epsilon +}(p) d\epsilon \geq 1 - \int_0^1 I_{ij}^\epsilon(p) d\epsilon$ for all $(i, j, p) \in \mathbb{I}$. The intuition is that the inequality is more likely to be satisfied if the original population is homogeneous or if $c_{ij}^1(p)$ is small, meaning that the variability of the intruder's budget share is small at $(i, j, p) \in \mathbb{I}$. The second condition ensures firstly that $\gamma(H^+) = 1 - \int_0^1 I_{ij}^{\epsilon +}(p) d\epsilon \leq 1 - \int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon +}(\tilde{p}) d\epsilon$ for all $(\tilde{i}, \tilde{j}, \tilde{p}) \in C\mathbb{I}$ and secondly that at least one original element of $(i, j, p) \in \mathbb{I}$ remains in this set after the expansion of the population, i.e. we have the maximal area under the step function $I_{ij}^\epsilon(p)$. Obviously, this condition is likely to be satisfied for large populations. However, the condition is stronger than required, because if for all

$(\tilde{i},\tilde{j},\tilde{p}) \in C\mathbb{I} : c^1_{\tilde{i}\tilde{j}}(\tilde{p}) \leq \int_0^1 I^\epsilon_{\tilde{i}\tilde{j}}(\tilde{p})d\epsilon$, then $\int_0^1 I^{\epsilon+}_{\tilde{i}\tilde{j}}(\tilde{p})d\epsilon \leq \int_0^1 I^\epsilon_{\tilde{i}\tilde{j}}(\tilde{p})d\epsilon$. In those cases we do not require the second condition. However, we use the stronger version of the second condition.

The first condition is also satisfied if

$$c^1_{ij}(p) \leq \inf_{h\in H} \frac{p_j|\partial_{p_j} w^h_i(p,x^h)|}{d^h_{ij}},$$

meaning that the intruder needs to have less relative variability of the budget share for $(i,j,p) \in \mathbb{I}$ than every household of the original population. In fact, this condition is stronger than the first one.

Let us prove Proposition 2 and illustrate it with the help of three examples.

Proof. The inequality $\gamma(H^+) \geq \gamma(H)$ corresponds to $\sup_{i,j,p} \int_0^1 I^{\epsilon+}_{ij}(p) \leq \sup_{i,j,p} \int_0^1 I^\epsilon_{ij}(p)$. In the Part A we prove the proposition for the strong version of the first condition. In Part B we show the general result.

A Since $G^\epsilon_{ij}(p)$ is the cumulative distribution function of $\frac{|s_i(p,x^h)|}{d^h_{ij}}$, we have for all $(i,j,p) \in \mathbb{I}$ that $G^{\epsilon+}_{ij}(p) \geq G^\epsilon_{ij}(p)$ for $\epsilon \in [0,1]$, if $c^1_{ij}(p) \leq \inf_{h\in H} \frac{p_j|\partial_{p_j} w^h_i(p,x^h)|}{d^h_{ij}}$ and therefore $I^{\epsilon+}_{ij}(p) - I^\epsilon_{ij}(p) \leq 0$. Using the properties of first order stochastic dominance (Lemma 2 in the appendix) leads to $\gamma^+_{ij}(p) \geq \gamma_{ij}(p)$ for all $(i,j,p) \in \mathbb{I}$. In order to ensure $\int_0^1 I^\epsilon_{ij}(p)d\epsilon \geq \int_0^1 I^{\epsilon+}_{\tilde{i}\tilde{j}}(\tilde{p})d\epsilon$ for all $(i,j,p) \in \mathbb{I}$ and all $(\tilde{i},\tilde{j},\tilde{p}) \in C\mathbb{I}$ we need the second condition, since $\frac{1}{(1+\#H)} \geq \sup_{i,j,p}\left(\int_0^1 I^{\epsilon+}_{ij}(p)d\epsilon - \int_0^1 I^\epsilon_{ij}(p)d\epsilon\right)$ for all $(i,j,p) \in \mathbb{I} \cup C\mathbb{I}$.

B One can show that

$$I^{\epsilon+}_{ij}(p) - I^\epsilon_{ij}(p) = \begin{cases} \frac{1}{1+\#H}\left(1 - I^\epsilon_{ij}(p)\right) & \text{if } \frac{|w^1_i(p,x^1)|}{d^1_{ij}} \geq \epsilon \\ \frac{-I^\epsilon_{ij}(p)}{1+\#H} & \text{otherwise} \end{cases}$$

In order to obtain

$$\int_0^1 I^{\epsilon+}_{ij}(p) - I^\epsilon_{ij}(p)d\epsilon \leq 0,$$

we need

$$\frac{1}{1+\#H}\left(\int_0^{c^1_{ij}(p)} 1d\epsilon - \int_0^{c^1_{ij}(p)} I^\epsilon_{ij}d\epsilon - \int_{c^1_{ij}(p)}^1 I^\epsilon_{ij}(p)d\epsilon\right) \leq 0$$

and therefore

$$\frac{1}{1+\#H}\left(c^1_{ij}(p) - \int_0^1 I^\epsilon_{ij}(p)d\epsilon\right) \leq 0$$

for all $(i,j,p) \in I\!I$ such that $\gamma(H) = 1 - \int_0^1 I_{ij}^{\epsilon}(p)d\epsilon$. Using $\int_0^1 I_{ij}^{\epsilon}(p)d\epsilon - \int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon}(\tilde{p})d\epsilon \geq \frac{1}{(1+\#H)}$ for all $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I$ proves the proposition. ■

Example 1 *The intruder has a Cobb-Douglas demand function, i.e.* $d_{ij}^1 = 0 \leq d_{ij}^h$. *One can infer* $\#\{h \in H^1 | p \in A_{ij}^{\epsilon}(w^h, x^h)\} = 1$ *for* $\epsilon \in [0,1]$. *Therefore*

$$\begin{aligned} I_{ij}^{\epsilon+}(p) - I_{ij}^{\epsilon}(p) &= \frac{1 + \#\{h \in H | p \in A_{ij}^{\epsilon}(w^h, x^h)\}}{1 + \#H} - \frac{\#\{h \in H | p \in A_{ij}^{\epsilon}(w^h, x^h)\}}{\#H} \\ &= \frac{1}{1+\#H}(1 - I_{ij}^{\epsilon}(p)) \\ &\geq 0 \quad \textit{for } \epsilon \in [0,1]\ ,\ p \in (0,\infty)^l \textit{ and all } i,j. \end{aligned}$$

Note that the areas below $I_{ij}^{\epsilon+}(p)$ *and* $I_{ij}^{\epsilon}(p)$ *are equal in size if* $d_{ij}^h = 0\ \forall h \in H$, *or if there exists* $\tilde{p}_j : \tilde{p}_j |\partial_{\tilde{p}_j} w_i^h(p, x^h)| = d_{ij}^h\ \forall h \in H$. *Lemma 2 in the appendix leads to* $\gamma(H^+) \leq \gamma(H)$.

Example 2 *The intruder has a demand function such that* $w^1(p, x^1) \neq w^h(p, x^h) = w(p)$ $\forall h \in H$. *Thus, for* $(i,j,p) \in I\!I$, *we have* $I_{ij}^{\epsilon}(p) = 1$, *which leads to*

$$I_{ij}^{\epsilon+}(p) - I_{ij}^{\epsilon}(p) \leq 0 \quad \textit{for } \epsilon \in [0,1],$$

where the areas below $I_{ij}^{\epsilon+}(p)$ *and* $I_{ij}^{\epsilon}(p)$ *are equal in size if there exists* $\tilde{p}_j$ *such that* $\tilde{p}_j |\partial_{\tilde{p}_j} w^1(p, x^1)| = d_{ij}^1$ *and* $\tilde{p}_j |\partial_{\tilde{p}_j} w(p)| = d_{ij}$. *By Lemma 2 we obtain* $\gamma(H^+) \geq \gamma(H) = 0$. *Note that condition 2 is redundant, since* $\gamma(H)$ *assumes its minimum.*

Example 3 *Suppose the intruder's demand function is such that* $d_{ij}^1 > d_{ij}^h = d_{ij} = 0$, *i.e. the population* H *consists only of households with Cobb-Douglas demand functions. Then we have*

$$1 = I_{ij}^{\epsilon}(p) \geq I_{ij}^{\epsilon+}(p) \quad \textit{for } \epsilon \in [0,1].$$

Note that we have equality for $c_{ij}^1(p) \geq \epsilon$. *By definition of* d_{ij}^1, *there exists at least one* $\tilde{p}$ *such that* $\tilde{p}_j |\partial_{\tilde{p}_j} w^1(p, x^1)| = d_{ij}^1$, *hence we have* $I_{ij}^{\epsilon+}(\tilde{p}) = 1$ *for* $\epsilon \in [0,1]$ *and thus* $\gamma(H^+) = \gamma(H) = 0$. *Condition 2 is redundant since* $\gamma(H)$ *has its minimal value. Note that further expansions of* H^+ *to* H^{++} *lead to* $\gamma(H^{++}) > \gamma(H)$ *if and only if there exists no* $\tilde{p} : \tilde{p}_j |\partial \tilde{p}_j w_i^h(\tilde{p}, x^h)| = d_{ij}^h$ *for all* $h \in H^{++}$.

Increasing Heterogeneity Due to Aggregation Let us generalize Proposition 2. Suppose $H = \dot{\bigcup}_{m=1}^{k} H^m$ and let $\sup_m \gamma(H^m) =: \gamma(H^n)$. In addition, let $I\!I$ be the set of (i,j,p) such that $\gamma(H^n) = 1 - \int_0^1 I_{ij}^{\epsilon n}(p) d\epsilon$.

Proposition 3 *Aggregation increases the degree of behavioral heterogeneity as measured by* γ, *i.e.* $\gamma(H) \geq sup_m \gamma(H^m)$, *if the following conditions hold:*

$$\int_0^1 I_{ij}^{\epsilon m}(p) d\epsilon \leq \int_0^1 I_{ij}^{\epsilon n}(p) d\epsilon$$

for all $(i,j,p) \in I\!I$ *and* $m = 1, \ldots, k$ *and*

$$\int_0^1 I_{ij}^{\epsilon n}(p) d\epsilon - \int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon n}(\tilde{p}) d\epsilon \geq \frac{\#H - \#H^n}{\#H}$$

for all $(i,j,p) \in I\!I$ *and for all* $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I$ *such that*

$$\int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon \geq \int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon n}(\tilde{p}) d\epsilon.$$

Proof. The proof follows the same reasoning as in the proof of Proposition 2. The first condition implies

$$\int_0^1 I_{ij}^{\epsilon}(p) d\epsilon \leq \int_0^1 I_{ij}^{\epsilon n}(p) d\epsilon$$

for all $(i,j,p) \in I\!I$. Let $CI\!I = \cup_{i=1}^2 CI\!I_i$, where $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I_1$ if $\int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon \leq I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon \leq \int_0^1 I_{ij}^{\epsilon n}(p) d\epsilon$ and $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I_2$ if $\int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon \geq I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon \leq \int_0^1 I_{ij}^{\epsilon n}(p) d\epsilon$. Therefore we have

$$\int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon}(\tilde{p}) d\epsilon \leq \int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon$$

for all $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I_1$ and

$$\int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon}(\tilde{p}) d\epsilon \geq \int_0^1 I_{\tilde{i}\tilde{j}}^{\epsilon m}(\tilde{p}) d\epsilon$$

for all $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I_2$. Thus, the second condition ensures $\int_0^1 I_{ij}^{\epsilon}(p) d\epsilon \geq I_{\tilde{i}\tilde{j}}^{\epsilon}(\tilde{p}) d\epsilon$ for all $(i,j,p) \in I\!I$ and all $(\tilde{i},\tilde{j},\tilde{p}) \in CI\!I_2$, since $\sup_{i,j,p} \left(\int_0^1 I_{ij}^{\epsilon}(p) - I_{ij}^{\epsilon m}(p) d\epsilon \right) \leq \frac{\#H - \#H^m}{\#H}$ for all m, i, j and $p \in (0,\infty)^l$. Hence the set of (i,j,p) such that $\gamma(H) = 1 - \int_0^1 I_{ij}^{\epsilon} d\epsilon$ might consist of elements of $I\!I$, $CI\!I_1$ and $CI\!I_2$. ■

The first condition of Proposition 3 implies that for all $(i,j,p) \in I\!I$ the heterogeneity of subpopulation n has to be the lowest. The second condition says that for subpopulation n the largest expanding area below the step function has to be smaller than the largest diminishing area minus the largest possible size of variation. The second condition is more likely to be satisfied if $\#H^n$ is large compared to the rest of the population.

The following examples might help in illustrating Proposition 3.

Example 4 *Suppose we have two homogeneous subpopulations H^1 and H^2, where H^1 consists only of households with Cobb-Douglas demand functions, i.e. for all $h \in H^1$ we have $d^h_{ij} = 0$, and for all $h \in H^2$ we have $w^h(p, x^h) = w(p)$, i.e. $d^h_{ij} = d_{ij} \geq 0$. Therefore, we have $\gamma(H^1) = \gamma(H^2) = 0$ and, as is shown in Example 3, we obtain $\gamma(H) = 0$.*

Example 5 *Suppose we have two homogeneous subpopulations $m = 1, 2$, where H^m consists of households with demand functions such that $w^h(p, x^h) = w^m(p)$. Therefore for all $h \in H^m$ we have $d^h_{ij} = d^m_{ij} \geq 0$. We conclude $\gamma(H) > 0$ if $I\!I^1 \cap I\!I^2 = \emptyset$, and $\gamma(H) = 0$ otherwise, where $I\!I^m$ denotes the set of (i, j, p) such that $\gamma(H^m) = 1 - \int_0^1 I^{\epsilon m}_{ij}(p)d\epsilon$.*

Example 6 *Suppose we have the Grandmont economy in which the budget share functions are to be taken of the form $w^\alpha(p, x) = w(\alpha * p, x)$ for some $\alpha \in (0, \infty)^l$, where $\alpha * p = (\alpha_1 p_1, \ldots, \alpha_l p_l)$. It is well known that $d_{ij} = d^\alpha_{ij}$ for $\alpha \in (0, \infty)^l$. Hildenbrand and Kneip (1999) have shown that $I^\epsilon_{ij}(p) = \nu\{B^\epsilon_{ij}(w^1, x) - log\ p\}$, where ν denotes the distribution of $log\ \alpha$ and $B^\epsilon_{ij}(w^1, x) := \{log\ p \in \mathrm{I\!R}^l \mid |\partial_{log\ \mathrm{p_j}} \mathrm{w^1_i(p, x)}| \geq \epsilon \mathrm{d_{ij}}\}$. Hence, if the distribution ν is sufficiently uniformly spread, the intersection frequency $I^\epsilon_{ij}(p) := \nu^\epsilon_{ij}(p)$ becomes arbitrarily small. Obviously, the Grandmont economy is a special case in which heterogeneity is solely expressed by the distribution of the parameter α. Let ν^m denote the distribution of $log\ \alpha$ for subpopulation m. Then, the distribution of the entire population is a mixture distribution, i.e. we have*

$$log\ \alpha \sim \sum_m \frac{\#H^m}{\#H} \nu^m$$

for all m. Then, $\gamma(H) \geq sup_m \gamma(H^m)$ corresponds to

$$sup_{i,j,p} \int_0^1 \nu^\epsilon_{ij}(p)d\epsilon \leq inf_m sup_{i,j,p} \int_0^1 \nu^{\epsilon m}_{ij}(p)d\epsilon.$$

Consequently, the conditions of Proposition 3 immediately apply. Using this concept, increasing heterogeneity due to aggregation corresponds to a mixed distribution ν that is more uniformly spread than every ν^m.

Weakly Increasing Heterogeneity. Now, we look at a weaker definition of increasing heterogeneity. Since Proposition 2 and Proposition 3 involve complicated conditions, this may allow for more intuitive results. We use a concept that compares the degree of heterogeneity on average.

Definition 3 *Aggregation weakly increases heterogeneity, as measured by* γ*, if*

$$\gamma(H) \geq \sum_{m=1}^{k} \frac{\#H^m}{\#H} \gamma(H^m)$$

is true.

Before presenting the result by a proposition, we state a lemma.

Lemma 1 *For all* i, j *and* $p \in (0, \infty)$*,* $\gamma_{ij}(p)$ *is an element of a convex set. The lower bound is* $inf_m \gamma_{ij}^m(p)$ *and the upper bound is* $sup_m \gamma_{ij}^m(p)$*, where all nonempty subpopulations* H^m *are disjoint and* $H = \dot{\bigcup}_{m=1}^{k} H^m$ *for every positive integer* $k \leq \#H$*.*

Proof. We have to prove that $\gamma_{ij}(p) \in [\inf_m \gamma_{ij}^m(p), \sup_m \gamma_{ij}^m(p)]$ for all i, j and $p \in (0, \infty)$. For a fixed ϵ, one can infer from the definition of $I_{ij}^{\epsilon}(p)$ that

$$I_{ij}^{\epsilon}(p) = \sum_{m=1}^{k} \frac{\#H^m}{\#H} I_{ij}^{\epsilon m}(p),$$

which is a convex combination of $I_{ij}^{\epsilon m}(p)$ over $m = 1, \ldots, k$. By rearranging, it follows immediately that

$$\gamma_{ij}(p) := 1 - \int_0^1 I_{ij}^{\epsilon}(p) d\epsilon = 1 - \sum_{m=1}^{k} \frac{\#H^m}{\#H} \int_0^1 I_{ij}^{\epsilon m}(p) d\epsilon,$$

which is evidently a convex combination of $1 - \sup_m \int_0^1 I_{ij}^{\epsilon m}(p) d\epsilon$ and $1 - \inf_m \int_0^1 I_{ij}^{\epsilon m}(p) d\epsilon$. ■

Now we ask whether $\gamma(H) \geq \sum_{m=1}^{k} \frac{\#H^m}{\#H} \gamma(H^m)$ holds. Preliminarily, this inequality is likely to be satisfied if all $\gamma(H^m)$ are very small, which corresponds to very homogeneous subpopulations, or if $\gamma(H)$ is close to one. The next proposition provides an unambiguous answer.

Proposition 4 *Aggregation weakly generates heterogeneity as measured by* γ*.*

Before we provide some intuition, let us prove the proposition.

Proof.

$$\begin{aligned}
\inf_{i,j,p}\gamma_{ij}(p) &\geq 1-\sup_{i,j,p}\sum_{m=1}^{k}\frac{\#H^m}{\#H}\int_0^1 I_{ij}^{\epsilon m}(p)d\epsilon \\
&\geq 1-\sum_{m=1}^{k}\frac{\#H^m}{\#H}\sup_{i,j,p}\int_0^1 I_{ij}^{\epsilon m}(p)d\epsilon \\
&= \sum_{m=1}^{k}\frac{\#H^m}{\#H}\left(1-\sup_{i,j,p}\int_0^1 I_{ij}^{\epsilon m}(p)d\epsilon\right) \\
&= \sum_{m=1}^{k}\frac{\#H^m}{\#H}\inf_{i,j,p}\gamma_{ij}^m(p)
\end{aligned}$$

We remark that the first inequality is due to Lemma 1. ■

Intuitively, $\gamma(H)$ is the smallest weighted average over all $\gamma_{ij}^m(p)$ with respect to (i,j,p) due to $\gamma(H) := \inf_{i,j,p}\gamma_{ij}(p)$, while $\sum_{m=1}^{k}\frac{\#H^m}{\#H}\gamma(H^m)$ is the weighted average over $\inf_{i,j,p}\gamma_{ij}^m(p)$. Note, the fact that Proposition 4 includes Proposition 1 as a weak increase in heterogeneity rules out a decrease in heterogeneity as defined in Definition 1.

One can infer that the separation of the population into homogeneous subgroups affects the structural properties of mean demand. Then we have on average less behavioral heterogeneity and we therefore may lose the monotonicity property of mean demand.

Proposition 1 and Proposition 4 are in accordance with results from Kneip (1999). He proves that the coefficient of sensitivity, a measure of structural stability of a population, does not decrease on average when aggregating subpopulations. In general, a small coefficient of sensitivity corresponds either to high behavioral heterogeneity of the population or to low variability in the demand behavior of households.

3 Conclusion

We investigate the effects of aggregation on the degree of behavioral heterogeneity. It is shown that aggregation never reduces the degree of behavioral heterogeneity. It may be the case, however, that aggregation generates heterogeneity. We derive sufficient conditions for generating heterogeneity due to aggregation and we show that aggregation weakly generates heterogeneity. We conclude that restricting attention to homogeneous subgroups of households does not allow one to capture the impacts of behavioral heterogeneity on aggregate values, such as mean demand.

Though we obtain rather general theoretical results, aggregation rules for empirical analysis are hardly derived. One has to be aware of two major determinants of the structural properties of mean demand: firstly, heterogeneity in the demand behavior of households and secondly, the invariability in demand behavior of households, i.e. d^h_{ij} is close to zero. The second determinant is treated neither in this paper nor in Hildenbrand and Kneip (1999). A skillful combination of both may yield superior results in order to obtain an aggregation rule for empirical analysis. This can be seen as a proposal for future research.

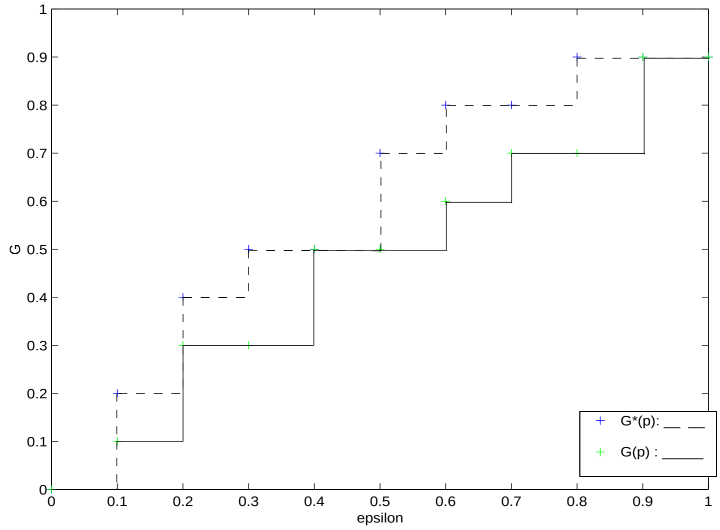

Figure 2: $G^{\epsilon}_{ij}(p)$ first order stochastically dominates $G^{\epsilon*}_{ij}(p)$.

Appendix

Lemma 2 *Consider two populations* H *and* H^* *such that* $I^{\epsilon*}_{ij}(p) - I^{\epsilon}_{ij}(p) \leq 0$ *for* $\epsilon \in [0,1]$. *Then we have*

$$\gamma^*_{ij}(p) - \gamma_{ij}(p) \geq 0.$$

Proof. For $\epsilon \in [0,1]$ we have

$$I^{\epsilon*}_{ij}(p) \leq I^{\epsilon}_{ij}(p) \ \Leftrightarrow \ G^{\epsilon*}_{ij}(p) \geq G^{\epsilon}_{ij}(p).$$

We know that $G^{\epsilon}_{ij}(p) = 1 - I^{\epsilon}_{ij}(p)$ is the cumulative distribution function of $\frac{|s^h_{ij}(p,x^h)|}{d^h_{ij}}$, so $G^{\epsilon}_{ij}(p)$ first order stochastically dominates $G^{\epsilon*}_{ij}(p)$. By definition of first order stochastic dominance one yields

$$\int \epsilon dG^{\epsilon}_{ij}(p) \geq \int \epsilon dG^{\epsilon*}_{ij}(p)$$

which is equivalent to

$$1 - \gamma_{ij}(p) \geq 1 - \gamma^*_{ij}(p).$$

If this inequality holds for all p, i, j, one obtains $\gamma(H^*) \geq \gamma(H)$ by definition of $\gamma(H)$. ■

References

[1] Grandmont, J.M. (1992). Transformation of the commodity space, behavioral heterogeneity, and the aggregation problem. *Journal of Economic Theory*, 57(1): 1–35.

[2] Hildenbrand, W. (1993). *Market Demand.* Princeton University Press, Princeton.

[3] Hildenbrand, W. and Kneip, A. (1999). On Behavioral Heterogeneity. SFB 303 Discussion Paper No. A-589, Universität Bonn.

[4] Kneip, A. (1999). Behavioral Heterogeneity and Structural Properties of Aggregate Demand. *Journal of Mathematical Economics*, 31: 49-74.

Essay 3

Behavioral Heterogeneity and the Law of Demand

April 22, 2002

Abstract

This paper focuses on the question how the shape of aggregate demand is affected by behavioral heterogeneity of a population. It is shown that the *Law of Demand* for aggregate demand holds under some conditions on the joint distribution of household behavior and disposable household income. Previous findings of Villemeur (2000b), who argues that behavioral complementarities induce the Law of demand for aggregate demand, are supported. In particular, it is shown that the joint distribution of disposable household income and not only the household behavior determines the shape of aggregate demand. Moreover, a new definition and a measure of behavioral differences is introduced and compared to existing concepts of behavioral heterogeneity as given in Grandmont (1992), Kneip (1999) and Hildenbrand and Kneip (1999). It is shown that extreme behavioral differences within a population imply the Law of demand and that this new concept overcomes some weaknesses of the previous approaches. In terms of aggregation, the measure of behavioral differences possesses the desirable properties of a heterogeneity measure (Wilke 2000).

1 Introduction

The modelling of an economy by a representative household cannot be considered as sophisticated. It is often done because of its simplicity. In general, the question arises whether such an idealization is crucial or not. It is not crucial if the average behavior of a population does not depend on distributional aspects or if the population indeed consists of homogeneous households. Otherwise information about heterogeneity has to be used. Kirman (1992) surveys the problem why the reduction of an heterogeneously behaving population to a rationally behaving representative household should fail. He concludes:

> This reduction of the behavior of a group of heterogeneous agents *even if they are all themselves utility maximizers*, is not simply an analytical convenience as often explained, but is both unjustified and leads to conclusions which are usually misleading and often wrong.

This is a strong conclusion and good news for scientists seeking to improve methods in economic theory. However, improving these methods is not an easy task, the development of fully general methods particularly so. This paper only treats a specific part of economic theory that is the modelling of aggregate demand for a heterogeneous population. The following questions are subject to analysis:

1. Does heterogeneity in the behavior of a population induce structural properties of aggregate demand, such as the Law of Demand? Does the shape of the distribution of household characteristics like disposable income have an impact?

2. What is a reasonable definition of behavioral heterogeneity? What definitions have been introduced in the past?

3. How can behavioral heterogeneity be measured? What kinds of measures already exist?

The answer to the first question is a clear yes. This question has already been treated and answered in several contributions. Among those are Grandmont (1992), Kneip (1999), Hildenbrand and Kneip (1999), Villemeur (2000b), Maret (2001) and Giraud and Maret (2001). All of these authors find that extreme behavioral heterogeneity (in various definitions) induces the Law of demand for aggregate demand, even if this property need not hold for any household of the population. The purpose of this paper is to derive exact conditions on the distribution of behavior and household characteristics of the population such that aggregate demand becomes regular without assuming the same for each household. Intuitively, the obtained conditions coincide with the so called "balancing effect" or complementarity of behavior as introduced by Villemeur (2000b). That is, given arbitrarily behaving households, aggregation can smooth individual behavior to aggregate behavior that fulfills some regularities, like monotonicity. In this paper it is shown that some kind of behavioral heterogeneity indeed leads to this effect. It also becomes clear that the amount of disposable income matters. A separation of the aggregate expenditure share is derived such that the scaling effect of the amount of disposable income is isolated from the effect of average demand behavior. In terms of affecting the shape of aggregate demand, it turns out that the behavior of households with higher income is weighted at least as high as the behavior of households with lower income.

In order to answer the second question there is some need for discussion. It should be obvious that behavioral heterogeneity means that households have different demand functions. This is something upon which all scientists have probably agreed. But there is even more which is widely accepted in the literature: extreme behavioral heterogeneity is considered to be present when each behavior occurs with the same probability. This can be described, for example, by a uniform distribution over a space of parameters if demand functions only differ in some parameters. It can also be described by a uniform distribution over the space of admissible demand functions, where admissible means that the space of functions is restricted due to some assumptions or regularity conditions. This viewpoint makes sense but has a major disadvantage: it is almost impossible to observe it empirically. Moreover, this very abstract concept is difficult to model and gives some freedom in terms of modelling to the scientist. Consequently, it is not uniquely defined in the literature. The aforementioned authors have introduced several definitions of behavioral heterogeneity in different frameworks of demand theory. Nevertheless, all succeeded in showing that extreme behavioral heterogeneity induces the Law of demand. Villemeur (1998),(1999) and (2000a) lays substantial criticism on some of these articles. He argues that by the construction of the definitions of behavioral heterogeneity some kinds of heterogeneity are ruled out. Moreover, he shows that in some of the above frameworks a very heterogeneous population in fact corresponds to one in which there are many very similar and regularly behaving households. In this case it is not surprising that aggregate demand inherits the same property. Recently, Maret (2001) and Giraud and Maret (2001) overcome his criticism by extending Kneip's (1999) definition of behavioral heterogeneity. The aforementioned authors base their definitions of behavioral heterogeneity on distributions of parameters or on distributions in the space of admissible expenditure share. This paper introduces a new definition of behavioral heterogeneity that is based on the images of household expenditure shares. More precisely, households are considered to be heterogeneous if there is a large distance between the images of their expenditure shares for the same commodity. If the images are close to each other they are considered to be homogenous in behavior. This concept will be referred to as 'behavioral differences'.

The answer to the third question is as follows: Some measures of behavioral heterogeneity have already been introduced in the past. For example Hildenbrand and Kneip (1999) define an "index of heterogeneity" which ranges from zero to one. The value one is obtained if the population is extremely heterogenous. The value zero can be obtained, for example, if all households possess the same demand functions. Kneip (1999) introduce a "coefficient of sensitivity" with similar properties. Both show that if the index or the coefficient approaches a certain value, the Law of demand holds for aggregate demand. This paper presents a new measure according to the concept of behavioral differences described above. It turns out that extreme behavioral differences in a population induce the Law of demand for aggregate demand. The new measure is therefore similar to the other measures even if the underlying concepts are very different. Thereafter, the direction to which the degree of behavioral differences is affected due to the aggregation of arbitrary disjoint subpopulations is determined. Again, the derived properties are similar to the properties of the index of heterogeneity and the coefficient of sensitivity.

To summarize, this paper is structured as follows: In the second part, the question of the conditions under what the Law of demand holds is analyzed and such conditions are derived. Commonly used definitions of behavioral heterogeneity are surveyed in the third part. In addition, a new definition is introduced and compared to the other concepts. Moreover existing measures of behavioral heterogeneity are surveyed. A new measure of behavioral differences is introduced. The fourth and last part investigates some properties of the measure of behavioral differences that are due to the aggregation of subpopulations.

2 Aggregate Demand and the Law of demand

Consider an economy with n households indexed by $i = 1, \ldots, n$. Each household i possesses a demand function $f^i(p, x^i) : \mathbb{R}_+^K \times \mathbb{R}_+ \mapsto \mathbb{R}_+^K$, where $p \in [0, \infty]^K$ denotes the vector of prices for the K commodities and $x^i \in \mathbb{R}_+$ is the disposable income of household i. The expenditure share of household i is given by $w^i(p, x^i) = p * f^i(p, x^i)/x^i$, where $p * f = (p_1 f_1, p_2 f_2, \ldots)$.

Aggregate demand is given by

$$F(p) = \sum_{i=1}^{n} f^i(p, x^i) = \sum_{i=1}^{n} x^i w^i(p, x^i)./p,$$

where $w./p = (w_1/p_1, w_2/p_2, \ldots)$ and the aggregate expenditure share is defined as

$$W(p) = p * F(p)/X = \sum_{i=1}^{n} x^i w^i(p, x^i)/X,$$

where $X = \sum_{i=1}^{n} x^i$. We state now two assumptions on the household expenditure shares:

Assumption 1 *The functions $w^i(p, x^i)$ are continuously differentiable in p and x^i for all i. Furthermore,*

$$sup_p \partial_p w^i(p, x^i)$$

is a finite matrix for $i = 1, \ldots, n$.

Assumption 2 $0 \leq w^i(p, x^i) \leq 1$.

Assumption 1 and 2 are of technical nature and are not restrictive from an economist's viewpoint. The so called 'Law of demand' has frequently been the subject of theoretical and applied analysis and is therefore not presented in detail here. However, the following remarks are made about the definition and about some of the main properties :

Remark 1 *The Law of demand for household demand holds if for two price vectors p and q the inequality $(p - q)'(F(p) - F(q)) \leq 0$ holds.*

Remark 2 *The Law of demand for household demand holds if the Jacobian matrix with respect to prices, $\partial_p f^i(p, x)$, is negative semi-definite.*

Remark 3 *The Law of demand for household demand holds if the household possesses a Cobb-Douglas demand function. In this case $w_k^i(p, x) = w_k^i$.*

Observe also that matrices with a negative dominant diagonal are also negative semi-definite. For further analysis we use this stronger negative dominant diagonal criterion due to its simplicity. In this case it is easy to derive the following useful results:

Lemma 1 *The Jacobian matrix of household demand has a negative dominant diagonal if*

$$\begin{aligned} p_k \partial_{p_k} w_k^i(p, x^i) - w_k^i(p, x^i) &< 0 \text{ and} \quad (1) \\ \left| p_k \partial_{p_k} w_k^i(p, x^i) - w_k^i(p, x^i) \right| &\geq p_k \sum_{l \neq k} \left| \partial_{p_l} w_k^i(p, x^i) \right| \quad (2) \end{aligned}$$

for $k = 1, \ldots, K$.

Lemma 2 *The Jacobian matrix of aggregate demand has a negative dominant diagonal if*

$$\begin{aligned} \sum_{i=1}^{n} \frac{x^i \partial_{p_k} w_k^i(p, x^i)}{p_k} - \sum_{i=1}^{n} \frac{x^i w_k^i(p, x^i)}{p_k^2} &< 0 \text{ and} \quad (3) \\ \left| \sum_{i=1}^{n} \frac{x^i \partial_{p_k} w_k^i(p, x^i)}{p_k} - \sum_{i=1}^{n} \frac{x^i w_k^i(p, x^i)}{p_k^2} \right| &\geq \sum_{l \neq k} \left| \sum_{i=1}^{n} \frac{x^i}{p_k} \partial_{p_l} w_k^i(p, x^i) \right| \quad (4) \end{aligned}$$

for $k = 1, \ldots, K$.

From Lemma 1 and Lemma 2 it is apparent that the less sensitive the household expenditure shares are to changes of the prices, the more likely is the Law of demand to hold. The next paragraphs shall illustrate the forces which may make the Law of demand for aggregate demand hold.

Aggregate Demand is Cobb-Douglas. Suppose aggregate demand is Cobb-Douglas, i.e. $W(p) = W$. Then

$$\partial W(p) = \sum_{i=1}^{n} x^i \partial_p w^i(p, x^i) = \mathbf{0}_K,$$

where $\mathbf{0}_K$ is a $K \times K$ matrix of zeros. This can be rewritten as:

$$\begin{aligned} \sum_{i=1}^{n} x^i \partial_{p_k} w_k^i(p, x^i) &= 0 \\ \sum_{i=1}^{n} x^i \partial_{p_l} w_k^i(p, x^i) &= 0 \text{ for } l \neq k. \end{aligned}$$

In other words, aggregation smoothes heterogenous household expenditure shares to a Cobb-Douglas aggregate expenditure share. Household expenditure shares may have either positive or negative partial derivatives with respect to prices. Let us now distinguish two cases:

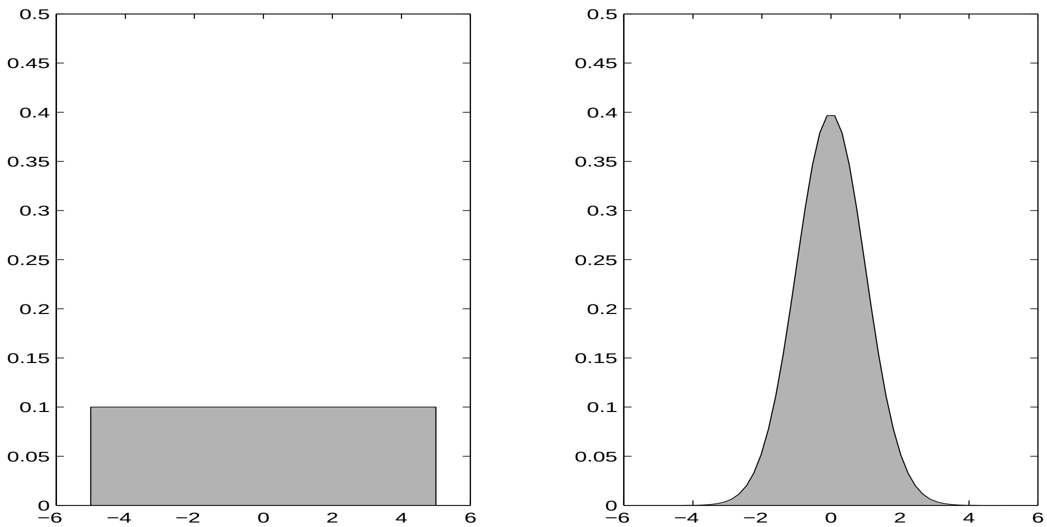

Figure 1: Uniform and truncated Standard Normal distribution of $\partial_{p_k} w_k^i(p, x)$ over $[-5, 5]$.

1. All households have the same disposable income:

Proposition 1 *If aggregate demand is Cobb-Douglas and if $x^i = x$ for all i, then the distribution of $\partial_{p_l} w_k^i(p, x)$ over all households i has mean 0 for $k, l = 1, \dots, K$.*

Note that this distribution need not to be symmetric. Since the partial derivatives of $w^i(p, x)$ with respect to prices are bounded, there exists a constant c, such that $c \geq |\partial_{p_l} w_k^i(p, x)|$. Figure 1 shows two popular parametric distributions which would fit the requirements, where $c = 5$. However, there is no reason why the empirical distribution could be approximated by a parametric distribution.

2. Households have different incomes:

Proposition 2 *If aggregate demand is Cobb-Douglas then the distribution of $x^i \partial_{p_l} w_k^i(p, x^i)$ over all households i has mean 0 for $k, l = 1, \dots, K$.*

Note that the distribution of $\partial_{p_l} w_k^i(p, x^i)$ and the distribution of $x^i \partial_{p_l} w_k^i(p, x^i)$ differ, because the latter is multiplied by a sequence of constants.

The two propositions reflect formally what Villemeur (2000b) introduces when he argues that only a specific kind of behavioral heterogeneity makes aggregate demand regular. He introduces the concept of strong complementarity of the household behavior:

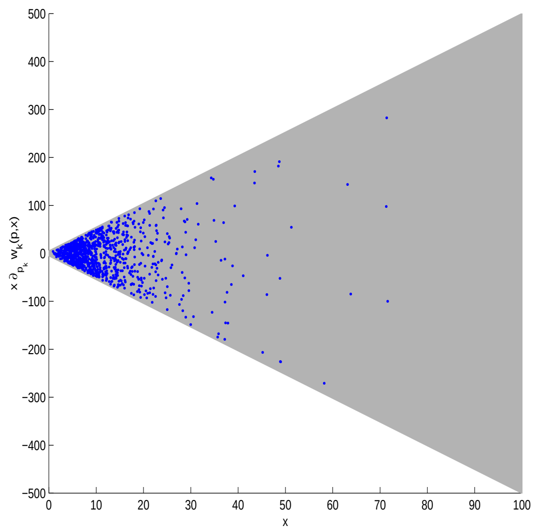

Figure 2: Illustration of a possible joint distribution of x and $x\partial_{p_k} w_k(p,x)$. The marginal distributions of $\partial_{p_k} w_k(p,x)$ and $x\partial_{p_k} w_k(p,x)$ are different.

Definition 1 *(Villemeur 2000b) The households $(w^i, x^i)_{i=1,\ldots,n}$ are said to display strong behavioral complementarities if their aggregate expenditure share is constant in prices, i.e.* $W(p) = c$.

Strong complimentary household behavior smoothes the aggregate expenditure share to a constant. Indeed, the conditions for strong behavioral complementarity are given in Proposition 1 and Proposition 2.

An interesting fact can be inferred from Proposition 2. It is easy to see that the forces which make the Law of demand hold are driven not only by the composition of the household expenditure shares within the economy, but also by the shape of the income distribution. We therefore have to consider the joint empirical distribution of the income x^i and the partial derivatives of $w_k^i(p, x^i)$ with respect to prices. Figure 2 shows the joint distribution for a population of 1000 households. The marginal distribution of $\partial_{p_k} w_k(p,x)$ is the uniform distribution over $[-5, 5]$ and the household incomes are log normally distributed. The shaded area illustrates the support of the joint distribution.

General Case Let us now consider the general case when $W(p)$ is a function of prices. We obtained some conditions in Lemma 2 that we are now reformulating in order to obtain some more convenient expressions. This convenience will be mainly based on the fact that we approximate empirical distributions by continuous distribution functions. We state two additional assumptions:

Assumption 3 *The empirical income distribution* $G_n(x) = n^{-1}\sum_i 1\!\!I_{x^i \leq x}$ *converges uniformly to a continuous cumulative distribution function* $G(x)$ *with density* $g_x(x)$ *and mean value* μ *as* n *becomes large.*

This assumption has several implications. Firstly, the uniform convergence allows us to substitute empirical frequencies by nicely behaved distribution functions. Secondly, continuity of the distribution is required for technical purposes. Moreover, it is precluded that there is one single household that dominates the whole economy in terms of disposable income.

Assumption 4 *The economy is large, i.e.* n *is large, such that* $sup_x|G_n(x) - G(x)| < \epsilon$ *for any* $\epsilon > 0$.

Let us now define some functions which assign values to the diagonal elements of the Jacobian of household demand with respect to prices.

Definition 2 *Let* $a_k^i(p, x^i) := p_k \partial_{p_k} w_k^i(p, x^i) - w_k^i(p, x^i)$ *be the diagonal value function for household* i *and let* $a_{k,n}(p, x) := \sum_{i=1}^n 1\!\!I_{x^i=x} a_k^i(p, x^i)$ *be the population diagonal value function at a given level of income* x*, where* $1\!\!I$ *denotes the indicator function.*

In addition, we use accordingly a definition for the off-diagonal elements:

Definition 3 *Let* $b_{kl}^i(p, x^i) := p_k \partial_{p_l} w_k^i(p, x^i)$ *be the off-diagonal value function of household* i *and* $b_{kl,n}(p, x) = \sum_{i=1}^n 1\!\!I_{x^i=x} p_k \partial_{p_l} w_k^i(p, x^i)$ *the population off-diagonal value function at a given level of income* x.

Let us now state two assumptions on the functions defined above, which can be considered as smoothness conditions across the income levels for a sufficiently large economy.

Assumption 5 $a_{k,n}(p, x)$ *converges uniformly to a finite function* $a_k(p, x)$ *that is continuous in* p *and* x *for all* k.

Assumption 6 $b_{kl,n}(p, x)$ *converges uniformly to a finite function* $b_{kl}(p, x)$ *that is continuous in* p *and* x *for all* k, l.

Assumptions 5 and 6 can be justified by a law of large numbers. As already mentioned in the previous paragraph, the shape of the functions a_k and b_{kl} is driven by two impacts: first, a_k^i and b_{kl}^i reflect the behavior of households and second, the sum over the indicator functions is an empirical frequency and therefore connected to the empirical income distribution. Using assumptions 3 and 4 we obtain a very useful separation for these two impacts:

Lemma 3 *The influence of household behavior and the income distribution on the structure of aggregate demand can be separated:*

$$a_k(p,x) = n\tilde{a}_k(p,x)g_x(x),$$

where

$$\tilde{a}_k(p,x) = lim_{n\to\infty}\frac{a_{k,n}(p,x)}{\sum_i 1\!\!I_{x^i=x}}.$$

Define $\tilde{a}_k(p,x) = 0$ whenever $g_x(x) = 0$. Note that $\tilde{a}_k(p,x)$ is continuous in p and x due to assumptions 3, 4 and 5. The same separation can be used for $b_{kl}(p,x)$. The question now arises, what is the exact meaning of $\tilde{a}_k$ and $\tilde{b}_{kl}$? Suppose there are $j = 1,\ldots,n_j$ households with income $x^j = x$. Then

$$\begin{aligned}\tilde{a}_k(p,x) &= \lim_{n\to\infty}\frac{\sum_i 1\!\!I_{x^i=x}\left(\partial_{p_k}w_k^i(p,x) - w_k^i(p,x)\right)}{\sum_i 1\!\!I_{x^i=x}} \\ &= \lim_{n\to\infty}\frac{\sum_{j=1}^{n_j}\left(\partial_{p_k}w_k^j(p,x) - w_k^j(p,x)\right)}{n_j}\end{aligned}$$

and an accordingly defined $\tilde{b}_{kl}(p,x)$ is the mean behavior of households with income x at prices p. Assumptions 3,4,5 and 6 imply that these two functions vary smoothly with changes in prices and across the income levels.

Let us now use the above results to define functions which determine the values of the diagonal and off-diagonal elements for the Jacobian of aggregate demand with respect to prices:

Definition 4 *Let*

$$\begin{aligned}A_k(p) &:= \int_{\mathbb{R}_+}\frac{x}{X}a_k(p,x)dx \\ &= n\int_{\mathbb{R}_+}\frac{x}{X}\tilde{a}_k(p,x)g_x(x)dx\end{aligned}$$

be the aggregate diagonal value function for $k = 1,\ldots,K$.

Definition 5 *Let*

$$\begin{aligned}B_{kl}(p) &:= \int_{\mathbb{R}_+}\frac{x}{X}b_{kl}(p,x)dx \\ &= n\int_{\mathbb{R}_+}\frac{x}{X}\tilde{b}_{kl}(p,x)g_x(x)dx \text{ for } k,l = 1,\ldots,K\end{aligned}$$

be the aggregate off-diagonal value function.

We can now formulate the main result of this section:

Proposition 3 *Under assumptions 1-6 the Law of demand for aggregate demand of a heterogenous population* $(f^i, x^i)_{i=1,\ldots,n}$ *holds if* $A_k(p)$ *and* $B_{kl}(p)$ *satisfy*

$$A_k(p) < 0 \text{ for } k = 1, \ldots, K$$

and

$$|A_k(p)| \geq \sum_{l \neq k} |B_{kl}(p)| \text{ for } k = 1, \ldots, K.$$

Proof. The first condition ensures negative diagonal elements of the Jacobian matrix of aggregate demand as they are given in Lemma 2. The second condition can be rewritten as

$$\int_{R_+} \left[|\tilde{a}_k(p,x)| - \sum_{l \neq k} \left| \tilde{b}_{kl}(p,x) \right| \right] g_x(x)dx \leq 0 \text{ for } k = 1, \ldots, K,$$

which is equivalent to the off-diagonal condition of Lemma 2. ■

According to the definition of strong behavioral complementarity, Villemeur (2000b) also considers the general case:

Definition 6 *(Villemeur 2000b) The households* $(w^i, x^i)_{i=1,\ldots,n}$ *are said to display behavioral complementarities if the aggregate expenditure share* $W(p)$ *verifies the Law of demand.*

Proposition 3 presents the conditions for behavioral complementarity.

The following example might help in illustrating how the forces work which make the Law of demand for aggregate demand hold.

Example 1 *Suppose there are* $i = 1, \ldots, n$ *households and two commodities, i.e.* $K = 2$. *From Proposition 3 we know that the Law of demand holds if the following inequalities are true:*

$$\begin{aligned} \int_{\mathbb{R}_+} \frac{nx}{X} \tilde{a}_k(p,x) g_x(x) dx &< 0 \text{ and} \\ \left| \int_{\mathbb{R}_+} \frac{nx}{X} \tilde{a}_k(p,x) g_x(x) dx \right| &\geq \left| \int_{\mathbb{R}_+} \frac{nx}{X} \tilde{b}_{kl}(p,x) g_x(x) dx \right| \text{ for } k,l = 1,2, \ k \neq l \end{aligned}$$

Using integration by parts, the first inequality can be rewritten as:

$$\mu \tilde{a}_k(p, \infty) < \int_{\mathbb{R}_+} \left[\int_0^x 1 - G(\tau) d\tau \right] \partial_x \tilde{a}_k(p,x) dx,$$

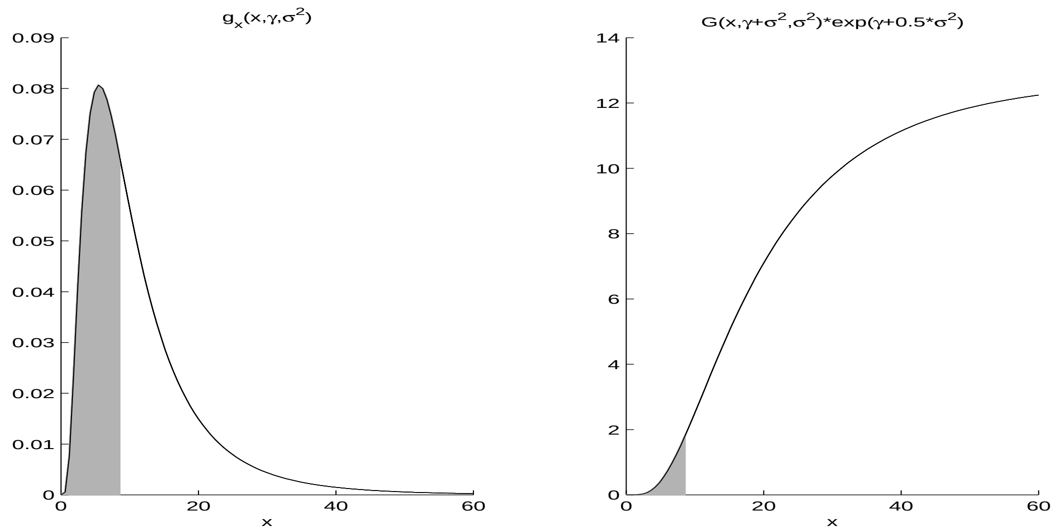

Figure 3: Logarithmic normal distribution of x with $\gamma = 2.2$ and $\sigma^2 = 0.7$ (left); corresponding incomplete moment function (right).

where again integration by parts yields $\int_{\mathbb{R}_+} 1 - G(x)dx = \mu$, *the mean household income. The left hand side is a constant and the right hand side contains the incomplete first moment function of the underlying income distribution. The second inequality holds if*

$$\left|\mu\tilde{a}_k(p,\infty) - \int_{\mathbb{R}_+}\left[\int_0^x 1 - G(\tau)d\tau\right]\partial_x\tilde{a}_k(p,x)dx\right|$$
$$\geq \left|\mu\tilde{b}_{kl}(p,\infty) - \int_{\mathbb{R}_+}\left[\int_0^x 1 - G(\tau)d\tau\right]\partial_x\tilde{b}_{kl}(p,x)dx\right|,$$

where $\mu\tilde{a}_k(p,\infty)$ *and* $\mu\tilde{b}_{kl}(p,\infty)$ *are equal to* 0, *since* $lim_{x\to\infty}g_x(x) = 0$ *induces* $lim_{x\to\infty}\tilde{a}_k(p,x) = \tilde{b}_{kl}(p,x) = 0$ *for all* $k \neq l$.

Suppose in addition that the empirical income distribution can be approximated by a logarithmic normal distribution with parameters γ *and* σ^2. *Note that in this case* $\mu = exp\{\gamma + 0.5\sigma^2\}$ *and following Butler and McDonald (1989) one can show that*

$$\int_0^x \left(1 - G(\tau;\gamma,\sigma^2)\right) d\tau = \mu G(x;\gamma+\sigma^2,\sigma^2)$$

holds. Figure 3 plots the density of the income distribution for empirically evident parameter values σ *and* μ, *and the corresponding incomplete first moment function. The incomplete first moment function is positive and nondecreasing from* 0 *to* μ *and can be interpreted as a weight function that assigns low weights to relatively poor households and large weights to relatively rich households. Therefore, the behavior of richer households is weighted at least* *as much as that of poorer households. From Figure 3 it is apparent that for a large fraction of the population the incomplete first moment function is close to zero. For a given level of income* $x = 8.5$*, we obtain* $G(8.5, 2.2, 0.7) = 0.47$ *and* $G(8.5, 2.9, 0.7)exp\{2.55\} = 1.77$*. This means that to almost* 50% *of the population has been assigned a weight less than or equal to* 1.8*. To see this look at the shaded areas in Figure 1. The richest households are assigned a weight up to a value of* 12*. Note that* $\partial_x \tilde{a}_k(p, x)$ *and* $\partial_x \tilde{b}_{kl}(p, x)$ *converge to* 0 *for* x *large enough (*> 60*) since* $g_x(x)$ *approaches* 0*. For* x *small enough we cannot infer mathematical properties for* $\partial_x \tilde{a}_k$ *and* $\partial_x \tilde{b}_{kl}(p, x)$ *from the above assumptions. Nevertheless, one might ask whether the partial derivatives are larger for low incomes in order to have some counter effect to the low income weights for poorer households. Intuitively, there is no reason for this to happen since we are considering the mean behavior of households with the same income.*

This example deepens the analysis of how both the aggregated demand behavior, expressed by the functions a_k and b_k, and the shape of the income distribution are the main determinants of the structure of aggregate demand. Moreover, it turns out that the behavior of poorer households plays a less important role than the one of richer households. This is in accordance with what we already found out discussing Propositions 1 and 2. Note that this effect is due to the fact that we are using expenditure shares as opposed to demand functions. If the latter were used this effect would be ruled out since demand is measured in absolute values.

This section has demonstrated that a somewhat heterogenous population may have an aggregate demand function for which the Law of demand holds. Propositions 1-3 show the conditions required for the distribution of household heterogeneity. Nevertheless, it it is not yet clear, whether heterogeneity in general or just a specific kind of heterogeneity causes the Law of demand to hold. The purpose of the following sections is to present existing definitions of behavioral heterogeneity and to introduce a new one.

3 Behavioral Heterogeneity: Definitions and Impacts

Considering the recent literature, behavioral heterogeneity has been defined in various ways. The contributions of Grandmont (1992), Kneip (1999) and Hildenbrand and Kneip (1999) are all concerned with the same topic but each use different concepts. Since the definition of behavioral heterogeneity is essential for further analysis we engage this problem in this section. After a review of existing definitions, a new definition is introduced. Moreover the impacts of behavioral heterogeneity on the shape of aggregate demand are presented using the framework of the previous section.

Something upon which all scientists may agree is that in the case of extreme behavioral homogeneity all households possess the same demand function. Finding a commonly accepted definition of behavioral heterogeneity is more difficult. The definition of behavioral heterogeneity is widely based on the dispersion of household characteristics. Grandmont (1992), Kneip (1999) and Hildenbrand and Kneip (1999) use this concept in order to show that if the households behaved extremely heterogeneously, the Law of demand for aggregate demand holds, even if none of the households behaves accordingly. The reasoning for using the dispersion of characteristics for measuring behavioral heterogeneity is explained for example by Villemeur (2000b): increasing dispersion of household characteristics means that we have less a priori information about the type of a household which is randomly chosen from the population. In other words low entropy means behavioral heterogeneity. As pointed out by Villemeur, this concept makes sense when considering compact spaces of parameters that differ across the households and which have some influence on the demand behavior of the households. If the space of parameters was not bounded, and even not observable, the reasoning becomes more complicated since the formulation can easily become an artificial construction. Intuitively, one can also refer to the dispersion over the space of feasible behavior. This definition is also based on the distribution on the parameter spaces of demand or even in the space of feasible expenditure shares. Figure 4 illustrates both concepts: on the top we see a distribution in the space of admissible demand functions $\mathcal{W}$. Extreme behavioral heterogeneity corresponds in this case to all subsets of $\mathcal{W}$ having same probability. More specifically, in the middle we see a distribution of parameters α. If all households had the same demand function and only differ in this parameter, one can say that the flatter the distribution of α the more heterogeneously the population behaves.

Alternatively, one might also use the distance of the images of household expenditure

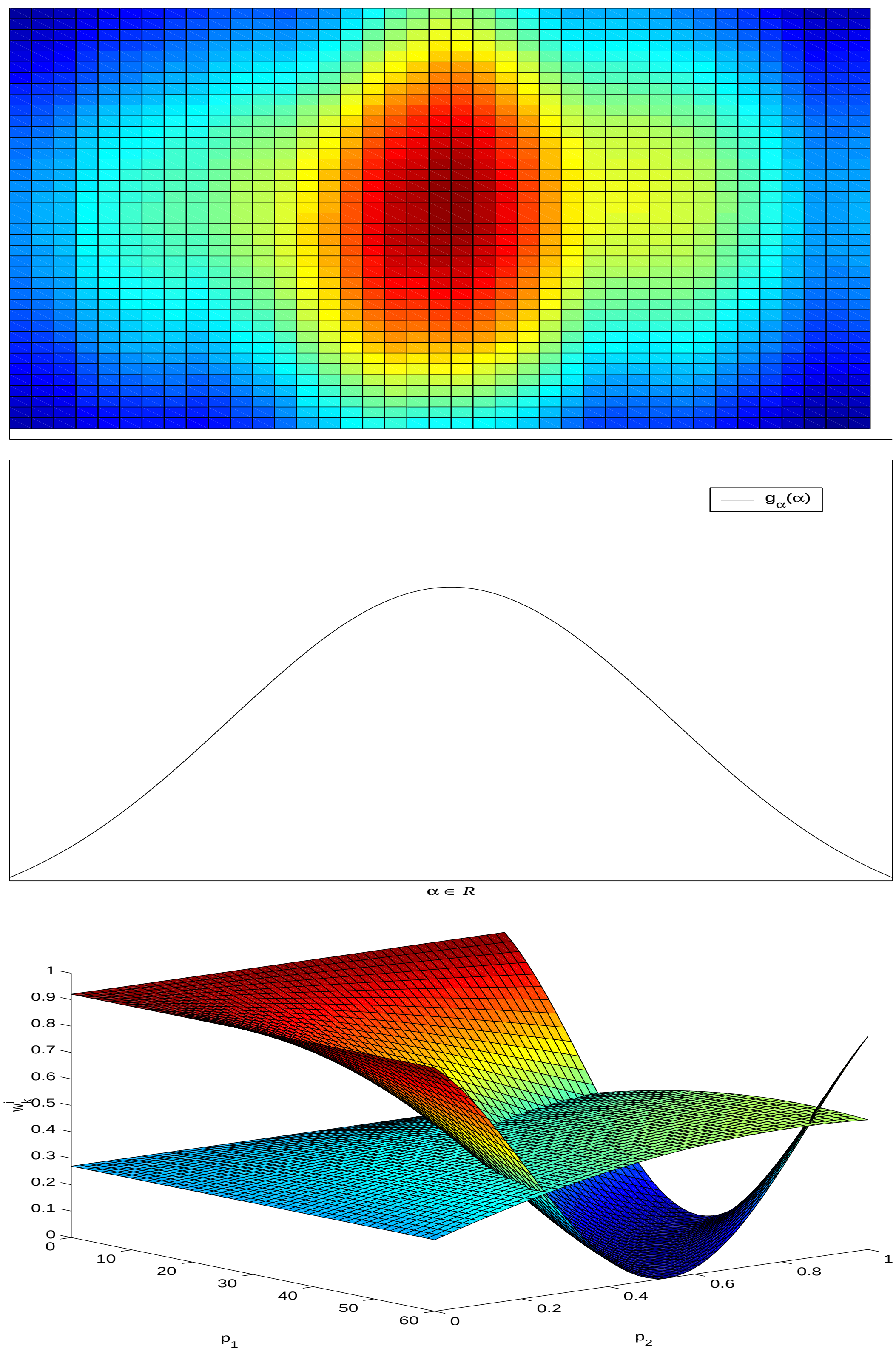

Figure 4: The definition of behavioral heterogeneity can be based on different concepts: distribution in the space of admissible expenditure shares $\mathcal{W}$ (top), distribution of parameters (middle), distances between the images of household expenditure shares (bottom)

shares in order to measure how different households behave. Increasing distance across the households in the population would then correspond to more behavioral differences. Figure 4 (bottom) illustrates two hyperplanes spanned in a $p_1 \times p_2$ space representing the images of two household expenditure shares. This new concept of behavioral differences is presented in detail and a measure of behavioral differences is defined. This new measure has similar properties compared to the degree of behavioral heterogeneity as recently introduced by Hildenbrand and Kneip (1999) but it overcomes two weaknesses: first, the intuition is clearer and second, it is able to distinguish between households with different Cobb-Douglas demand functions.

Before presenting and introducing the definitions of behavioral heterogeneity it is briefly justified why these concepts are based on the relative demand of household i for commodity k. As noted by Kneip (1999) it might be the case that

$$f_k(p, x) - f_k(p, x_0) \to \infty$$

and

$$w_k(p, x) - w_k(p, x_0) = 0$$

for $x \to \infty$ and $x_0 \to 0$. Hence, the use of the expenditure shares rules out the scaling effect of the household income that is present in the demand functions.

Distribution of Parameters: Grandmont (1992) Grandmont uses a parametric concept of behavioral heterogeneity. He assumes that the demand functions of households with income $x^i = x$ only differ in some parameters:

$$f^\alpha(p, x) = e^\alpha * \zeta(e^\alpha * p, x)$$

where $\alpha \in \mathbb{R}^K$ denotes a vector of parameters, which characterizes the household specific behavior.

Assumption 7 *The empirical distribution function of α converges uniformly to a twice boundedly differentiable distribution function with density $g_\alpha(\alpha)$.*

In this model the definition of behavioral heterogeneity is based on concept of entropy. Grandmont introduces accordingly:

Definition 7 *A population is the more heterogenous in its behavior the flatter is its marginal distribution of parameters, $g_\alpha(\alpha)$.*

Heterogeneity is measured by the flatness of the distribution of α: $\int_{\mathbb{R}} |\partial_{\alpha_k} g_\alpha(\alpha)|$.

Assumption 8 *$f^\alpha(p, x)$ is homogeneous of degree 0 in p and x.*

Assumption 9 *$f^\alpha(p, x)$ satisfies Walras's law, i.e. $p * f^\alpha(p, x) = x$.*

Assumption 10 *Aggregate demand $F(p)$ is bounded from below for $p \in \mathbb{R}^K$: $\exists c \in \mathbb{R}^K_{++}$, such that $F(p) \geq c$.*

Under these assumptions Grandmont proves the following proposition:

Proposition 4 *(Grandmont 1992) Extreme behavioral heterogeneity, i.e. $\sup_k \int_{\mathbb{R}} |\partial_{\alpha_k} g_\alpha(\alpha)| d\alpha = 0$, causes the Law of demand to hold for aggregate demand.*

The proof follows from the fact that he shows:

$$\left| \int_{\mathbb{R}^K} \partial_{p_l} w_k(e^\alpha * p, x) g_\alpha(\alpha) d\alpha \right| \leq \frac{\sup_{p,x} |w_k(e^\alpha * p, x)|}{p_l} \left(\int_{\mathbb{R}^K} |\partial_{\alpha_l} g_\alpha(\alpha)| d\alpha \right) \tag{5}$$

for all k, l. Using Lemma 2, extreme behavioral heterogeneity in this model therefore induces the off-diagonal elements of the Jacobian of aggregate demand to be zero. The first term of he diagonal elements also vanishes and the remaining second term is strictly positive due to assumption 10. The Jacobian of aggregate demand is therefore negative diagonal dominant, having a negative real part of its largest characteristic value.

Villemeur (1998) shows that household demand in this setting becomes more insensitive with respect to price changes the larger is $|\alpha_k|$. He also points out that an extremely flat distribution, i.e. a uniform distribution, of α on $\mathbb{R}^K$ implies that the mass of a compact set on the support of α converges to 0. Therefore, the mass of this distributions is allocated to asymptotically large values of α. In this case most of the households in this model are insensitive, i.e. Cobb-Douglas-like, to price changes. Extreme behavioral heterogeneity therefore corresponds to having a population consisting of Cobb-Douglas like behaving households.

Distribution of behavior: Kneip (1999) In contrast to the former approach, Kneip considers the distribution of a continuum of functions $w \in \mathcal{W}$, where $\mathcal{W}$ is the space of admissible expenditure shares. He defines a probabilistic framework resulting in a well defined probability measure ν on the smallest σ-Algebra of the space of admissible expenditure shares $\mathcal{W}$. Details of this measure theory based approach are not presented here. Based on the model introduced in Section 2 we consider an economy with a discrete set of households.

Kneip analyzes aggregate demand conditional on a given level of income x:

$$F(p,x) = xp^{-1} * \sum_{j=1}^{n_j} w^j(p,x)$$

He proves weak sensitivity of the sum, i.e. it has a small partial derivative with respect to a price, induces structure and definiteness on the Jacobian of aggregate demand. In order to measure the sensitivity of the above sum, Kneip introduces:

Definition 8 *Let ν denote the empirical distribution of $w^i \in \mathcal{W}$. The coefficient of sensitivity*

$$h(\nu) := max_{k,l}\{sup_{p,x}|\partial_{p_l} \sum_{j=1}^{n_j} w_k^j(p,x)|\}$$

measures the sensitivity of the aggregate expenditure share to price changes.

Definition 9 *Aggregate demand is defined as structurally stable whenever the coefficient of sensitivity is small enough.*

It is shown that structural stability of aggregate demand is caused by invariability of the household expenditure shares with respect to prices changes, meaning that all households behave as if Cobb-Douglas. Moreover, Kneip postulated that structural stability is also caused by extreme behavioral heterogeneity, where he defines the latter as:

Definition 10 *Extreme behavioral heterogeneity of a population is equivalent to all subsets of $\mathcal{W}$ of a given and equal size having the same probability.*

This concept intuitively corresponds to 'pointwise' heterogeneity. Each feasible behavior occurs with the same probability. This can be formalized as follows: Given a subset $J \in [0,1]^K$, the image space of admissible expenditure shares, we have a high amount of heterogeneity if for all J

$$\nu\big\{w^i \in \mathcal{W} | w(p,x) \in J\big\} \approx \nu\big\{w^i \in \mathcal{W} | w(q,y) \in J\big\}$$

holds, where q and y are transformed p and x.

Kneip shows under assumptions 1-3 and 10:

Proposition 5 *(Kneip 1999) Structural stability induces a negative definite Jacobian of aggregate demand.*

The proof in our framework follows immediately from Lemma 2 with the same reasoning as in the Grandmont model. Clearly these conditions are stronger than required since the weighting of income is not considered since $F(p,x)$ has a negative dominant diagonal for all given x. This means in particular that the functions $b_{kl}(p,x)$ are close to zero and $a_k(p,x) < 0$ for all k, l, p and x.

It is not very easy to construct an example which intuitively points out the meaning of 'pointwise' heterogeneity since its definition is quite abstract. It is difficult to imagine the impact of a uniform distribution over the space of feasible expenditure shares.

Example 2 *Consider an economy in which all households have the same income $x^i = x$. Suppose for instance that extreme behavioral heterogeneity as defined above induces a uniform distribution of $w_k^i(p,x)$ on $[0,1]$ for $k = 1,2$ and all $p \in [0,1]^K$. Then, for all k, $W_k(p) = \sum_{i=1}^{n} w_k^i(p,x)$ converges uniformly in p to 0.5 as n becomes large. Aggregate demand is therefore Cobb-Douglas with $\sum_{k=1}^{2} W_k(p) = 1$. More generally, define $\tilde{p}_k = (p_1, \ldots, p_{k-1}, \Delta + p_k, p_{k+1}, \ldots, p_K)$, where $\Delta \in \mathbb{R}_+$. Then*

$$\sum_{i=1}^{n} w_k^i(p, x^i) = \sum_{i=1}^{n} w_k^i(\tilde{p}_l, x) \text{ for } k, l = 1, \ldots, K$$

follows. Therefore, by Proposition 1 the Law of demand for aggregate demand holds since

$$\begin{aligned} 0 &= lim_{\Delta \to 0} \Delta p_k \sum_{i=1}^{n} \frac{w_k^i(\tilde{p}_l, x) - w_k^i(p, x)}{\Delta} \\ &= \Delta p_k \sum_{i=1}^{n} \partial_{p_l} w_k^i(p, x) \text{ for } k, l = 1, \ldots, K \end{aligned}$$

and $x^i = x$ for all i. The same reasoning holds if the empirical distribution of $w_k^i(p, x^i)$ on $[0,1]$ is a uniform distribution independent of the prices and the income. This example can be considered to be related to Becker's (1962) pioneering findings. He shows that if a uniform distribution of $w_k^i(p,x)$ on the budget line is not affected by price changes, then aggregate demand is declining.

Kneip's approach has been subject to criticism. Villemeur (1999) argues that a uniform distribution over an infinite 'space of behavior' gives all the weight to its boundary elements. He concludes that behavioral heterogeneity in this setup '*reduces dramatically the set of admissible expenditure share functions*', whereby it does not necessarily follow that this set only consists of Cobb-Douglas like households as in the Grandmont model. Maret (2001) and Giraud and Maret (2001) show that Kneip's result can also be achieved when $\mathcal{W}$ is a compact set. In this case the mass of the distribution is not allocated on the boundary and therefore it is not anymore subject to Villemeur's criticism. Moreover, they show that extreme behavioral heterogeneity corresponds to strong behavioral heterogeneity as given in Definition 1.

Distribution of partial derivatives: Hildenbrand and Kneip (1999) Hildenbrand and Kneip introduce an index, $\gamma \in [0, 1)$, that measures the degree of behavioral heterogeneity of the population. They show under assumptions 1-3 that the Law of demand for aggregate Demand holds if the demand behavior of the population is sufficiently heterogenous. Roughly speaking, their definition of behavioral heterogeneity is based on the distribution of the partial derivatives of household expenditure shares with respect to prices.

For $\lambda > 0$ define

$$S_{kl}(p) := \partial_\lambda W_k(p_1, \dots, \lambda p_l, \dots, p_K)|_{\lambda=1} = p_l \partial_{p_l} W_k(p)$$

as the rate of change of $W_k(p)$ with respect to a percentage change of the price p_l. Accordingly one obtains for each household

$$s^i_{kl}(p, x^i) := \partial_\lambda w^i_k(p_1, \dots, \lambda p_l, \dots, p_K, x^i)|_{\lambda=1} = p_l \partial_{p_l} w^i_k(p, x^i).$$

Moreover, the upper bound for $|s^i_{kl}(p, x^i)|$ with respect to prices is given by

$$d^i_{kl} := \sup_p |s^i_{kl}(p, x^i)| = \sup_p p_l |\partial_{p_l} w^i_k(p, x^i)|.$$

Let $\delta \in [0, 1]$. The domain of prices in which $p_l|\partial_{p_l} w^i_k(p, x^i)| \geq \delta d^i_{kl}$ is defined as

$$A^\delta_{kl}(w^i, x^i) := \{p \in (0, \infty)^K |\; p_l|\partial_{p_l} w^i_k(p, x^i)| \geq \delta d^i_{kl}\}.$$

In order to define the measure for the degree of behavioral heterogeneity of a population, it is important to clarify what Hildenbrand and Kneip (1999) mean by behavioral heterogeneity: $A^\delta_{kl}(w^i, x^i)$ are located in different regions of the price system in $(0, \infty)^K$ across the households i. Accordingly, they define an intersection ratio

$$I^\delta_{kl}(p) := \frac{1}{n} \text{card}\{i | p \in A^\delta_{kl}(w^i, x^i)\}$$

which indicates by how much the sets $A^\delta_{kl}(w^i, x^i)$ differ across the households i. Note that $I^\delta_{kl}(p)$ is a decreasing step function in δ. A high degree of behavioral heterogeneity implies a low intersection ratio. Then $\int_0^1 I^\delta_{kl}(p) d\delta$ is close to zero. Taking this into account, let us now define the degree of behavioral heterogeneity of the household population:

Definition 11 *The degree of behaviorial heterogeneity of a population is measured by the index of heterogeneity γ, where*

$$\gamma := inf_{k,l,p}\gamma_{kl}(p) = 1 - sup_{k,l,p}\int_0^1 I_{kl}^{\delta}(p)d\delta.$$

Note that $0 \leq \gamma \leq 1 - \frac{1}{n} < 1$.

Hildenbrand and Kneip (1999) show that a sufficiently small index of heterogeneity implies the Law of demand for mean demand, although this property was not assumed for household demands:

Proposition 6 *A sufficiently high degree of behavioral heterogeneity implies negative diagonal dominance of the Jacobian matrix of mean demand with respect to prices if there is a constant $c \geq 0$ such that $d_{kl}^i \leq c$ for all k, l, i.*

Villemeur (2000a) has shown that in the Hildenbrand and Kneip framework, extreme behavioral heterogeneity corresponds to a population of households that are mostly insensitive to price changes but in which for each household i, there exists a small area of prices at which i's demand is sensitive to price changes. Their definition of behavioral heterogeneity induces $\gamma = 0$ if all households have the same or a Cobb-Douglas demand function. Clearly, in this case there is no heterogeneity in the distribution of the partial derivatives across the households. Their example of extreme behavioral heterogeneity is the Grandmont economy with a flat distribution of α. As already noted, in this case the main part of the population behaves Cobb-Douglas-like for given prices.

Behavioral differences In what follows, the concept of behavioral heterogeneity is based on the distance of the images of household expenditure shares. In contrast to the framework defined in Section 2 the price system is now normalized to $[0,1]^K$. The distance between the expenditure shares with given income $x^i = x$ is measured on its support, which is the price system. For this purpose we use the Euclidian distance:

Definition 12 *The Euclidian distance between the images of the functions $w_k^i(p,x)$ and $w_k^j(p,x)$ is defined as*

$$\|w_k^i(p|x) - w_k^j(p|x)\|_2 := \left[\int_{(0,1]^K} \left(w_k^i(p,x) - w_k^j(p,x)\right)^2 dp\right]^{1/2}.$$

Note that we do not consider $p = 0$ since by definition of the expenditure shares: $w(0, x) = 0$ for all households. The definition of behavioral homogeneity and behavioral heterogeneity that is now introduced is not based on the concept of entropy and therefore differs from the foregoing approaches. Behavioral heterogeneity is now based on how 'different' the households behave. Let us begin with

Definition 13

$$\|w_k^i(p|x) - w_k^j(p|x)\|_2 =: d_k^{ij}(x)$$

is a pairwise measure of behavioral differences between household i and household j.

Note that due to Assumption 2 we have $\|w_k^i(p|x) - w_k^j(p|x)\|_2 \in [0, 1]$ for all i, j and k since $p \in [0, 1]^K$. The pairwise measure is the distance of two hyperplanes spanned over the $(0, 1]^K$ space (see Figure 4, bottom). The hyperplanes are the images of the household expenditure shares. See Figure 4. A small distance might be interpreted as behavioral homogeneity since in this case the images are almost identical.

Definition 14 *Two households i and j are said to be $c-$homogenous in their demand behavior for commodity k if*

$$d_k^{ij}(x) \leq c,$$

where $c \in [0, 1]$.

Accordingly, a large distance might be interpreted as behavioral heterogeneity:

Definition 15 *Two households i and j are said to be $c-$heterogeneous in their demand behavior for commodity k if*

$$d_k^{ij}(x) > c,$$

where $c \in [0, 1]$.

According to this pairwise measure $d_k^{ij}(x)$ we say that two households i and j are extremely homogenous in their demand behavior if they possess the same expenditure shares, i.e.

$$\sup_k \|w_k^i(p, x) - w_k^j(p, x)\|_2 = \inf_k \|w_k^i(p, x) - w_k^j(p, x)\|_2 = 0 \text{ for all } p \text{ and given } x.$$

Furthermore, two households i and j with the same income x are said to be extremely different in their demand behavior for commodity k if

$$d_k^{ij}(x) = 1.$$

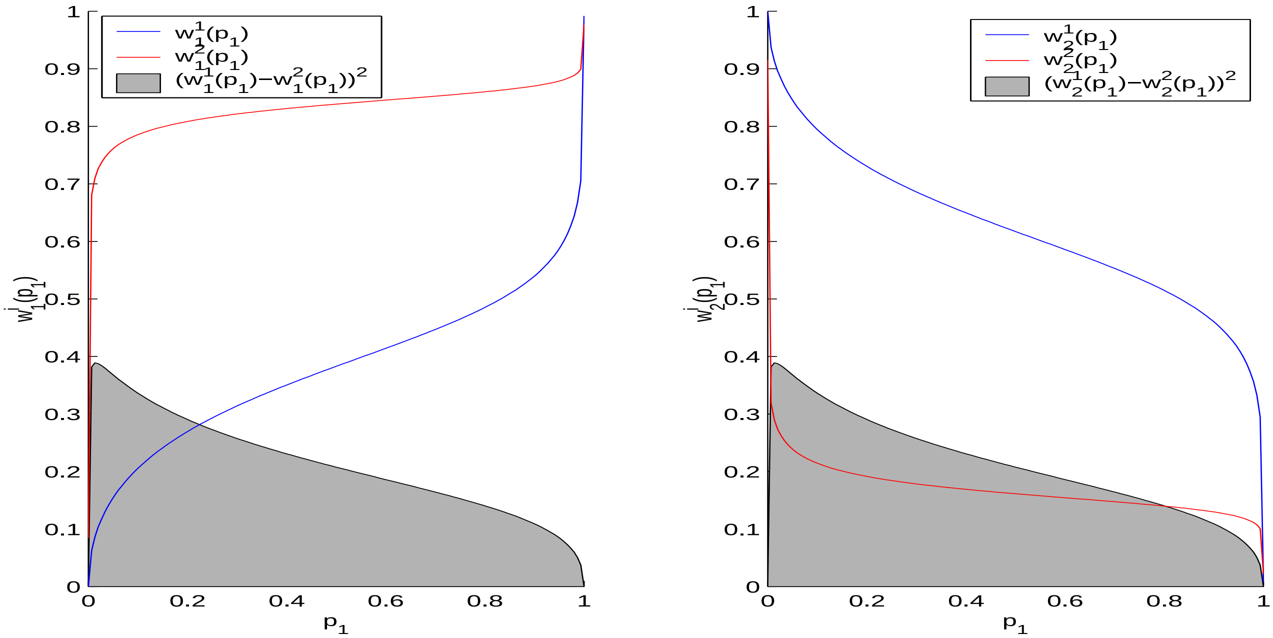

Figure 5: CES demand with $\eta = [0.4\ 0.9]$ and $\sigma = [0.5\ 0.8]$; the shaded area corresponds to the difference between the functions.

Example 3 *Suppose $K = 2$ and $n = 2$ and without loss of generality $\|p\| = 1$. The households have CES demand functions with expenditure shares*

$$w_k(p_1, p_2) = \frac{\eta^\sigma p_k^{1-\sigma}}{\eta^\sigma p_1^{1-\sigma} + (1-\eta)^\sigma p_2^{1-\sigma}} \text{ for } k = 1, 2,$$

where $0 \leq \eta \leq 1$ and $\sigma > 0$ denotes parameters describing the household behavior. Note that the expenditure shares are independent of income x and that $p_2 = \sqrt{1 - p_1^2}$. Therefore the argument space of the functions $w_k^i(p_1)$ boils down to dimension one. Figure 2 shows a particular case for specific parameter values. Note that the two subplots are reflections due to Walras' Law, i.e. $w_1^i(p_1) + w_2^i(p_1) = 1$ for $p_1 \in [0, 1]$.

Since we are interested in the behavioral differences of the whole population, we need to introduce another measure. A first step toward this purpose is to define an aggregated $d_k(x)$:

$$d_k(x) := \sum_j \sum_{i \leq j} d_k^{ij}(x) \text{ for } i, j = 1, \ldots, n.$$

Note that d_k cannot be used as a measure since it does not possess the properties of a measure. It is not difficult to construct a case such that $d_k(x) \to \infty$ as $n \to \infty$. Therefore, we have to relate it to the size of the population. This is done by the following normalization:

$$\phi_k(x) = \frac{d_k(x)}{D},$$

where D denotes the upper bound of $d_k(x)$, i.e. $d_k(x) \in [0, D]$. This upper bound depends on n and whether n is even or odd. It is given by

$$\begin{aligned} D_e &= \sum_{\substack{t=1 \\ t \text{ odd}}}^{n-1} t = \left(\frac{n}{2}\right)^2 \text{ for n even and} \\ D_o &= \sum_{\substack{t=0 \\ t \text{ even}}}^{n-1} t = \frac{1}{4}(n^2 - 1) \text{ for } n \text{ odd.} \end{aligned}$$

Hence, $0 \leq \phi_k(x) \leq 1$.

Lemma 4 *As $n \to \infty$ the impact of distinguishing between n even and n odd converges to zero. In this case D_o can be substituted by D_e.*

Proof. Suppose for simplicity $d_k(x) = c$. We have

$$\lim_{n\to\infty} \left[\frac{c}{D_e} - \frac{c}{D_o} \right] = 0^+, \tag{6}$$

where 0^+ means convergence from above. However, note that

$$\lim_{n\to\infty}(D_e - D_o) = 1/4.$$

Substituting D_o by D_e still ensures $\phi_k(x) \in [0, 1]$. This need not hold when substituting in the opposite direction since the convergence in (6) is from above. ■

Definition 16 *The degree of behavioral differences within a population for given household income x is measured by*

$$\phi(x) = inf_k \phi_k(x).$$

The intuition behind the degree of behavioral differences is the following: it relates the aggregated distance of the household expenditure shares within the population to its maximally theoretically attainable level. According to this definition, extreme behavioral differences within a population can solely be attained if there are two equal sized groups of households which have Cobb-Douglas demand functions one with $w_k^i(p, x) = 1$, and the other with $w_k^i(p, x) = 0$. These functions can therefore be considered as a kind of boundary behavior. The following proposition follows immediately from Remark 3 and taking into account that a sum of constants is also a constant:

Proposition 7 *$inf_x \phi(x) = 1$ induces the Law of demand for aggregate demand.*

	Extreme behavioral heterogeneity	Households posses the same demand function	Households posses Cobb-Douglas demand
		value of the index	
Coefficient of sensitivity	0	$0 \leq h < \infty$	0
Degree of behavioral heterogeneity	1	0	0
Degree of behavioral differences	1	0	$0 \leq \phi \leq 1$

Table 1: Comparison of the different concepts for the modelling of behavioral heterogeneity. Note that there is no index in the Grandmont approach.

A maximal $\phi(x)$ for all x corresponds to a population of Cobb-Douglas households. This property is obviously stronger than required. In order to make the Law of demand for aggregate demand hold it should therefore be sufficient to have a high degree of behavioral differences. However, differences in the Cobb-Douglas demand functions across the households now matter. This is in contrast to the approaches of Kneip (1999) and Hildenbrand and Kneip (1999). Meaning that the degree of behavioral differences can still be between zero and one whenever all households possess Cobb-Douglas demand functions. It is only zero if all households with the same income x possess the same expenditure share for at least one commodity k.

Let us conclude this section by briefly summarizing the main properties of the foregoing indices which result from different modelling concepts. Table 1 provides an overview. The coefficient of sensitivity does not distinguish between extreme behavioral heterogeneity and the case when all households have the same Cobb-Douglas demand function. The degree of behavioral heterogeneity overcomes this weakness and is bounded. However, it is always zero if the households posses Cobb-Douglas demand functions even with different parameters. This is in contrast to the degree of behavioral differences which is only zero if all households have the same demand function for at least one good given an income.

For future research it might be interesting to consider the following[1]: Suppose we have a given distribution of households. Is it possible to derive thresholds for the indices, such that

[1]This point has been suggested by Wolfgang Leininger.

the Law of demand holds if an index exceeds or falls below the critical value? This would make the whole approach more applicable since the results which are derived in the cited contributions and in this paper are only valid for the extreme values of the indices.

4 Aggregation

This section analyzes how the aggregation of subpopulations affects the degree of behavioral differences.

As shown in Kneip (1999) for the coefficient of sensitivity $h(\nu)$ and in Wilke (2000) for the degree of behavioral heterogeneity γ, both measures satisfy some nice properties when considering an aggregation framework. In particular, the degree of behavioral heterogeneity is weakly increasing due to the aggregation of arbitrary subpopulations. This is particularly interesting since in empirical analysis mostly homogeneous subgroups are considered. Taking into account what we have derived in the last section it might be the case that, due to this disaggregation, the structural properties of aggregate demand get lost. In what follows we scrutinize whether the degree of behavioral differences $\phi_k(x)$ possesses similar properties.

The analysis of this section is done for a given x and k. For convenience use the following notations: $\phi_k(x) = \phi$ and $d_k(x) = d$.

Proposition 8 *Suppose there are n households indexed by $i = 1, \ldots, n$ and an intruder household $j = n + 1$. All households have the same income x. The degree of behavioral differences may either increase or decrease when the intruder is embraced. Let $\zeta = \sum_{i=1}^{n} d^{ij}$ for $j = n+1$ and let ϕ^+ denote the degree of behavioral differences of the enlarged population $i = 1, \ldots, n + 1$. Then, we have*

$$\begin{aligned} \phi^+ \geq \phi, \quad & \text{if} \quad \zeta \geq \frac{n}{2}\phi \text{ for } n \text{ even, or} \\ & \text{if} \quad \zeta \geq \frac{n+1}{2}\phi \text{ for } n \text{ odd, and} \\ \phi^+ < \phi \quad & \quad\;\;\; \text{otherwise.} \end{aligned}$$

Proof. Let D^+ denote the upper bound for d^+ of the enlarged population $i = 1, \ldots, n+1$. In order to show that $\phi^+ \geq \phi$ holds, we rewrite this inequality as

$$\frac{d^+}{D^+} \geq \phi.$$

From

$$d^+ - d = \zeta$$

follows

$$\frac{d + \zeta}{D + (D^+ - D)} \geq \phi.$$

Solving for ζ yields

$$\zeta \geq \phi D + (D^+ - D)\phi - d.$$

Since

$$\begin{aligned} D^+ - D &= \frac{n}{2} \text{ for } n \text{ even} \\ D^+ - D &= \frac{n+1}{2} \text{ for } n \text{ odd,} \end{aligned}$$

we obtain the conditions for ζ. ■

Example 4 *If the intruder household $j = n + 1$ is extremely different with respect to the initial population, i.e.*

$$\zeta = n,$$

it follows $\phi^+ \geq \phi$, since $\zeta \geq \phi(n + 1)/2 \geq \phi n/2$.

If the intruder household $j = n + 1$ is relatively different with respect to the initial population and n even, i.e.

$$\zeta \geq n/2$$

it follows $\phi^+ \geq \phi$, since $\zeta = n/2 \geq \phi n/2$. The same reasoning holds for n odd if $\zeta \geq (n + 1)/2$.

Let us now consider the general case, when aggregating some arbitrary disjoint subpopulations $m = 1, \dots, M$ which each consist of n_m households.

Proposition 9 *Given a population $i = 1, \dots, n$. Let $\omega_m = n_m/n$ and $\sum_{m=1}^{M} n_m = 1$. Then aggregation over $m = 1, \dots, M$ disjoint subpopulations weakly increases the degree of behavioral differences:*

$$\phi \geq \sum_{m=1}^{M} \omega_m \phi^m,$$

where γ^m is the degree of behavioral differences of subpopulation m.

Proof. Suppose $M = 2$ without loss of generality. Use the notation $d_m = \sum_j \sum_{i \le j} d^{ij}$ for $i, j = 1, \ldots, n_m$ and $m = 1, 2$. We have to show that

$$\frac{d}{D} \ge \frac{n_1 d_1}{n D_1} + \frac{n_2 d_2}{n D_2} \tag{7}$$

holds for all possible disjoint decompositions of the entire population.

Suppose the first population consists of $n - r$ households indexed by $i = 1, \ldots, n - r$ and the second accordingly of r households indexed by $j = n - r + 1, \ldots, n$. Using this (7) becomes

$$\frac{d}{D} \ge \frac{n - r}{n} \frac{d_1}{D_1} + \frac{r}{n} \frac{d_2}{D_2}.$$

We know that $\sum_{j=n-r+1}^{n} \zeta_j = d - d_1 - d_2$ and $\zeta_j = \sum_i d^{ij}$ for $i = 1, \ldots, n - r$ and $j = n - r + 1, \ldots, n$. The proof is done for n and r even. However, for any even number n and r the upper bound of d_m remains constant when adding an additional household such that n and r are odd numbers, we can infer that the results do not change in this case. We obtain

$$\begin{aligned} D &= \frac{n^2}{4} \\ D_1 &= \frac{(n-r)^2}{4} \\ D_2 &= \frac{r^2}{4}. \end{aligned}$$

and hence,

$$\frac{d_1 + d_2 + \sum_{j=n-r+1}^{n} \zeta_j}{n} \quad \ge \quad \frac{d_1}{n - r} + \frac{d_2}{r}.$$

Therefore,

$$\sum_{j=n-r+1}^{n} \zeta_j \ge \frac{r}{n - r} d_1 + \frac{n - r}{r} d_2. \tag{8}$$

The proposition is proven by showing that inequality (8) holds. Since it is complicated to understand the reasoning of the general proof, it is first shown for some specific cases:

1. $d_1 = d_2 = 0$: trivial. Both subpopulations are extremely homogeneous.
2. $d_1 = D_1$, $d_2 = 0$: extreme behavioral differences of subpopulation 1 implies $\zeta_j = (n-r)/2$ for $j = n - r + 1, \ldots, n$ (Figure 6a). Hence,

$$\sum_{j=n-r+1}^{n} \zeta_j = \frac{r(n-r)}{2} \ge \frac{r}{n-r} \frac{(n-r)^2}{4}$$

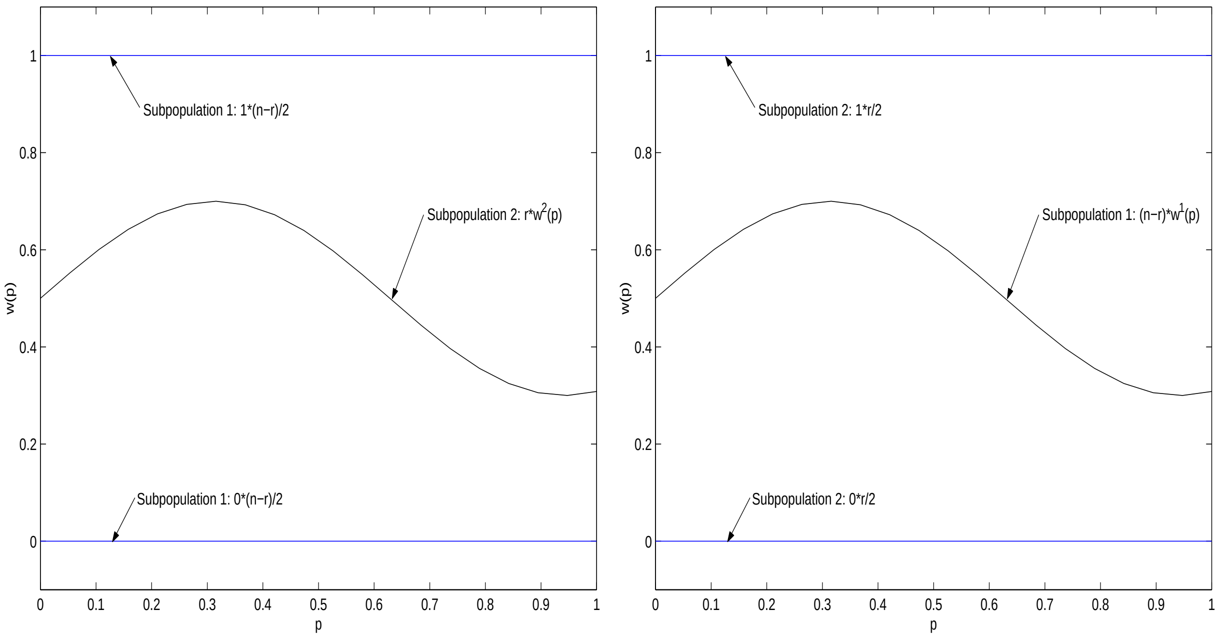

Figure 6: Illustration for $\sum \zeta_j = r(n-r)/2$, if one of the subpopulations consists of extremely differently behaving households; for simplicity: $K = 1$.

holds. Moreover, it holds for $d_2 \in [0, D_2]$, whereby it is an equality for $d_2 = D_2$.
3. $d_1 = 0, d_2 = D_2$: extreme behavioral differences of subpopulation 2 implies $\sum \zeta_j = r(n-r)/2$ (Figure 6b). Hence,

$$\sum_{j=n-r+1}^{n} \zeta_j = \frac{r(n-r)}{2} \geq \frac{n-r}{r}\frac{r^2}{4}$$

holds. Moreover, it holds for $d_1 \in [0, D_1]$, whereby it is an equality for $d_1 = D_1$.
General Case: $d_1 = D_1 - \delta_1$, $d_2 = D_2 - \delta_2$, where $\delta_1 \in [0, D_1]$ and $\delta_2 \in [0, D_2]$. We obtain:

$$\begin{aligned}\sum_j \zeta_j &\geq \frac{r}{n-r}(D_1 - \delta_1) + \frac{n-r}{r}(D_2 - \delta_2) \\ &= \frac{r}{n-r}\left[\frac{(n-r)^2}{4} - \delta_1\right] + \frac{n-r}{r}\left[\frac{r^2}{4} - \delta_2\right] \\ &= \frac{r(n-r)}{2} - \frac{r}{n-r}\delta_1 - \frac{n-r}{r}\delta_2\end{aligned}$$

and after rewriting $\delta_1 = \epsilon_1(n-r)/2$ and $\delta_2 = \epsilon_2(r)/2$, we obtain

$$\sum_j \zeta_j \geq \frac{r(n-r)}{2} - \frac{r}{2}\epsilon_1 - \frac{n-r}{2}\epsilon_2$$

Indeed, we have

$$\begin{aligned}\sup\{\sum_j \zeta_j\} &= \frac{r(n-r)}{2} + \frac{r}{2}\epsilon_1 + \frac{n-r}{2}\epsilon_2 \\ \inf\{\sum_j \zeta_j\} &= \frac{r(n-r)}{2} - \frac{r}{2}\epsilon_1 - \frac{n-r}{2}\epsilon_2,\end{aligned}$$

where the sup and inf are taken over the set of all possible compositions of the subpopulations. ■

Propositions 8 and 9 are in accordance with the findings of Wilke (2000) and Kneip (1999), since they find the same properties for γ and $h(\nu)$. However, for the case of behavioral differences we can infer another property:

Proposition 10 *Any population which possesses extreme behavioral differences, i.e.* $\phi = 1$, *can be decomposed into two extreme homogenous subgroups, i.e.* $\phi_1 = \phi_2 = 0$, *such that aggregate demand of both subgroups separately satisfy the Law of demand.*

The two subgroups clearly consist of the different types of Cobb-Douglas households. Note, however, that this might serve as a justification that in some specific cases a decomposition into homogenous households does not destroy structural properties.

References

[1] Becker, G.S. (1962). Irrational Behavior and Economic Theory. *Journal of Political Economy*, 70: 1–13.

[2] Butler, R.J. and McDonald, J.B.(1989). Using Incomplete Moments to Measure Inequality. *Journal of Econometrics*, 42: 109–119.

[3] Giraud, G. and Maret, I. (2001). Behavioral Heterogneity in Large Economies. Working Paper, Universite Louis Pasteur, Strasbourg.

[4] Grandmont, J.M. (1992). Transformation of the Commodity Space, Behavioral Heterogeneity, and the Aggregation Problem. *Journal of Economic Theory*, 57(1): 1–35.

[5] Hildenbrand, W. and Kneip, A. (1999). On Behavioral Heterogeneity. SFB 303 Discussion Paper No. A-589, Universität Bonn.

[6] Kirman, A.P. (1992). Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives*, 6(2): 117-136.

[7] Kneip, A. (1999). Behavioral Heterogeneity and Structural Properties of Aggregate Demand. *Journal of Mathematical Economics*, 31: 49-74.

[8] Maret, I. (2001). Modelling Behavioral Heterogeneity in Demand Theory. BETA Working Paper No. 2001-04, Universite Louis Pasteur, Strasbourg.

[9] Villemeur, E. de (1998). Essays on Consistency in Economic Modelling. PhD Thesis, European University Institute.

[10] Villemeur, E. de (1999). Aggregation of Demand and Distribution of Characteristics: A Difficulty in Modelling Behavioral Heterogeneity. THEMA Discusssion Paper No. 9938, Universite de Cergy-Pontoise.

[11] Villemeur, E. de (2000a). On the Index of Heterogeneity proposed by W.Hildenbrand and A.Kneip. Manuscript, Universite de Toulouse.

[12] Villemeur, E. de (2000b). Behavioral Complementarity (Not Heterogeneity) Causes the Law of Demand. Discussion Paper, Universite de Toulouse.

[13] Wilke, R. (2000). A Note on Behavioral Heterogeneity and Aggregation. Discussion Paper in Economics No.00-05, Universität Dortmund.

Essay 4

Semiparametric Estimation of Consumer Demand

April 22, 2002

Abstract

This essay suggests two approaches to the estimation of mean expenditure shares. The estimation of demand systems is usually done either with parametric or nonparametric estimators. This paper focuses on semiparametric estimation procedures that only parameterizes the heterogeneity in the behavior of individuals. The particular individual behavior is supposed to be largely unknown. The resulting flexibility due to this specification leads to a smaller risk of misspecification in comparison to parametric estimation. Moreover, due to a low dimensionality of the nonparametric part there is a higher rate of convergence in comparison to nonparametric estimation. Simulations investigate the finite sample performance.

1 Introduction

Since Hildenbrand and Kneip (1999), it is well known that a sufficient degree of behavioral heterogeneity in the population implies certain structural properties of mean demand, such as the law of demand. Wilke (2000a) has shown that focusing on homogeneous subgroups may make those structural properties vanish. Villemeur (2000) argues that only a specific kind of behavioral heterogeneity, i.e. complementary behavior, and not generic behavioral heterogeneity causes those structural properties.

The aim of this paper is to present two semiparametric models for the estimation of the mean expenditure share of a population of heterogeneously behaving households. Since we just have limited information about how exactly households behave, it cannot be our main purpose to parameterize their exact behavior. Therefore we use nonparametric estimators. In order to ensure a low dimensionality of the underlying problem, we embed a parametric component into the model. This parametric part does not explicitly model the consumption behavior of the households but takes into account some systematic behavioral heterogeneity in the population. Three main aspects contribute to an interesting overall performance of the presented approaches:

- Considering a part of the model to be unknown allows for more flexibility and hence implies more robustness against misspecification in comparison to pure parametric estimators.
- The combination of a nonparametric and a parametric part results in a higher convergence rate in comparison to nonparametric estimators. In particular, estimates are more reliable in areas with low density in the data. As becomes apparent from Figure 1, the tail behavior of a nonparametric estimate can be unreliable. This is often due to low density of the data and weak finite sample performance of the estimators at the boundaries.
- Modelling of differences in the behavior of the households, but not the behavior itself, lets behavioral heterogeneity play a role as a determinant of the shape of the mean expenditure share.

The first estimation approach is based on the shape preserving transformation technique as presented in Härdle and Marron (1990). The authors provide a theoretical framework for the pooling of unknown subgroup expenditure shares of similar shape, yielding a more

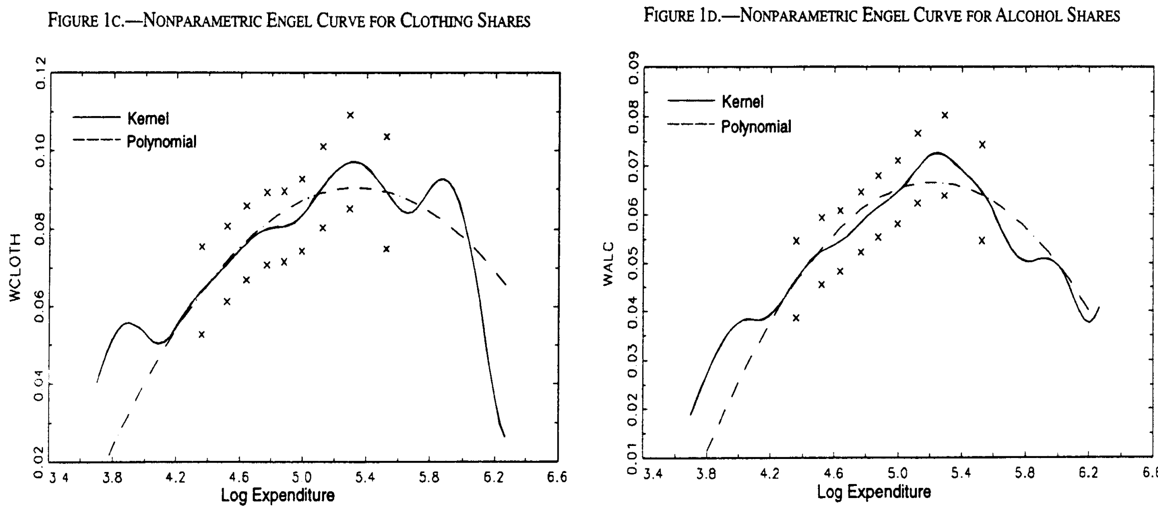

Figure 1: Source: Banks, Blundell and Lewbel (1997)

accurate estimate of the mean expenditure share. The second approach uses similar transformations, but makes use of a particular functional form of subgroup expenditure shares, which has been provided by economic theory. The second approach uses multivariate parametric approximations for the pooling of the unknown functions.

The two semiparametric estimators suggested differ somewhat to what we can mostly find in the econometric literature. The reason is that the mean expenditure share is considered to be a weighted average of unknown functions, which are related by parametric transformations either in the spaces of the covariables or in the space of the response variable. Therefore, we propose the mean function is estimated nonparametrically, while finding the estimates of the transformation parameters.

The paper is organized as follows: Section 1.1 presents the economic model and main specifications for applied demand theory. Section 1.2 provides a brief survey of the estimators: the local polynomial smoother is used as a nonparametric regression estimator; the Parzen-Rosenblatt estimator is used as a nonparametric density estimator; the variance estimator is the most commonly used estimator for homoscedastic models. Then, in order to determine the optimal constant and optimal variable bandwidth for the nonparametric regression estimation, we briefly introduce some bandwidth selection rules. Section 2 develops the semiparametric estimation methods for the mean expenditure share and provides some examples. Section 3 analyzes the finite sample performance of these methods through simulations. Section 4 concludes.

1.1 The Economic Model

The population consists of $i = 1, \ldots, n$ households.

Definition 1 *The expenditure share of household $i = 1, \ldots, n$ for good $k = 1, \ldots, K$ is defined as the function $M_k^i(x) := M_k(x, w_{ik})$, where $w_{ik} \in \mathbb{R}^P$ is a vector of parameters that captures the household specific behavior for good k and $x \in \mathbb{R}_+$ denotes the disposable household income.*

Thus, the household expenditure shares only differ in some parameters, while the functional form is the same for all i. In contrast to many contributions that are concerned with theoretical and empirical demand analysis, we require very mild assumptions on the individual behavior, which are more or less of technical nature. Therefore, the functional form of the expenditure shares is to a large extent unknown. In order to take into account some common facts about data, let us introduce the concept of homogenous subgroups of the population. This is particularly useful when working with cross section data, since in this case we do not observe a household more than once.

Definition 2 *Two households i and j are homogeneous in terms of behavior if they are contained in the same homogenous subgroup, i.e. their demand behavior is sufficiently similar:*

$$\|M_k^i(x) - M_k^j(x)\|_2 \leq \epsilon,$$

where $\epsilon > 0$ is an arbitrarily chosen constant and $\| * \|_2$ denotes the Euclidean distance.

In order to obtain reliable estimates, we need the following assumption:

Assumption 1 *For a small $\epsilon > 0$, there exists a disjoint decomposition of the population into a finite number of sufficiently large homogeneous subpopulations $z \in \mathbb{I}$, where $\mathbb{I}$ denotes an index set.*

If two households are contained in the same homogenous subpopulation z, one can define a group specific expenditure share $M_k(x, w_{zk})$, where $w_{zk} \in \mathbb{R}^P$ is a vector of parameters that fully describes the specific behavior of subpopulation z for good k. Figure 2 shows that expenditure shares of homogeneous subpopulations could be in fact linked in a parametric way.

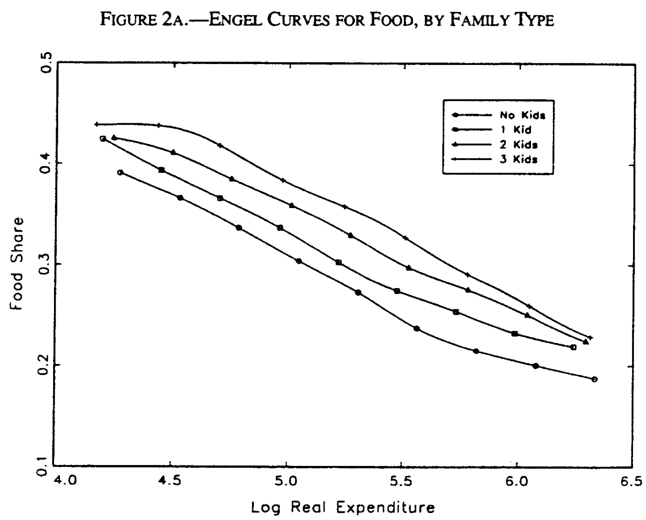

Figure 2: Source: Banks, Blundell and Lewbel (1997)

Definition 3 *The mean expenditure share for good k of the population at a given level of income, x, is given by*

$$M_k(x) := \frac{1}{n}\sum_{i=1}^{n} M_k^i(x).$$

Furthermore it is useful to point out the following property:

Lemma 1 *The mean expenditure share can be written as*

$$M_k(x) = M_k(x, w_k),$$

where $w_k = \frac{1}{n}\sum_i w_{ik}$, if the Functions M^i are linear in w_{ik} for all i.

The economic and econometric literature derives and considers many functional forms of the mean expenditure share, most of which are linear in the parameters. Three frequently used specifications are

$$M_k(x) = \alpha_k + \beta_k \ln x \textbf{ (PIGLOG)} \quad (1)$$

$$M_k(x) = \alpha_k + \beta_k \ln x + \gamma_k (\ln x)^2 \textbf{ (QUAIDS)} \quad (2)$$

$$M_k(x) = \alpha_k + \beta_k \ln x + g_k(x) \textbf{ (PLM)}, \quad (3)$$

where $\alpha_k = \sum_{i=1}^{n} \alpha_{ik}/n$, $\beta_k = \sum_{i=1}^{n} \beta_{ik}/n$ and $\gamma_k = \sum_{i=1}^{n} \gamma_{ik}/n$ are parameters expressing the mean consumption behavior of the population for good $k = 1, \ldots, K$. According to economic theory, the parameters may depend on the prices of goods. For simplicity, we do not consider this general form, and omit this price dependency. Engel curves of the form

$$M_k(x) = \alpha_k x + \beta_k x \ln x \textbf{ (Working-Leser)} \quad (4)$$

have been introduced by Working (1963) and Leser (1943). For some goods, Working and Leser's specification is well supported by empirical analysis, as is the PIGLOG specification. However, expenditure shares are not generally log-linear. This is in accordance with Gorman's (1981) theoretical findings, which showed that utility based exactly aggregable demand systems have a maximum rank equal to three, where the maximum rank of a demand system is defined as the maximum rank of the $(K \times P)$ matrix of the demand system coefficients. Here, P denotes the number of parameters in a demand system, i.e. $P = 2$ in (4) and (1), and $P = 3$ in (2). Banks et al. (1997) deal with rank three demand systems. They show that the Quadratic Almost Ideal Demand System (QUAIDS) is the only rank three demand system that is consistent with utility maximization and has expenditure shares which are linear in a constant, in log income and in some other differentiable function of income. Note that in (1), (2) and (4) M_k is linear in w_{ik}.

Since the purpose of this paper is to suggest two new semiparametric methods, we do not consider mean expenditure shares of the common semiparametric specification (3). These models are known as partially linear models (PLM) and represent a common extension of the PIGLOG specification. The function $g_k(x)$ is an unknown smooth function that captures all nonlinearities in x. It is estimated with a nonparametric estimator. Robinson (1988) has shown $\sqrt{n}$-consistency of the parameter estimates, while using the Nadaraya-Watson estimator as the estimator for g_k. Wilke (2000b) compares the Nadaraya-Watson estimator to the local linear smoother and derives differences in asymptotic properties for both cases. His simulations indicate that the local linear smoother possesses a superior finite sample performance in this class of models. For further analysis let us omit the index k, i.e. focus on one good.

1.2 The Estimators

Local Polynomial Regression Let (X, Y) be observable random variables with $(X_i, Y_i)_{i=1,\dots,n}$ independent realizations. Suppose that

$$E(Y|X) = M(X)$$

holds, where

$$Y|X \sim N(E\left(Y|X\right), \sigma^2)$$

and M is an unknown smooth function of $\mathcal{C}^\nu$, the space of ν times differentiable functions, for which the following series around x_0 exists

$$M(X) = M(x_0) + M'(x_0)(X - x_0) + \frac{M''(x_0)}{2!}(X - x_0)^2 + \ldots + \frac{M^{(\nu)}(x_0)}{\nu!}(X - x_0)^\nu.$$

Let $K : \mathbb{R} \mapsto \mathbb{R}$ be the Kernel function which satisfies:

- $\int K(u)du = 1$
- K has compact support and is bounded
- All odd order moments of K vanish, i.e.

  $$\int u^l K(u)du = 0$$

  for all odd integers $l > 0$. This is for example satisfied for symmetric functions.

Under the foregoing regularity conditions on the model and the Kernel function, the local polynomial regression estimator at the evaluation point x is the solution to

$$\min \sum_i \left(Y_j - \sum_{j=0}^{\nu} a_j (X_i - x)^j \right)^2 K_h(X_i - x)$$

with respect to a_j where $K_h(X_i - x) = K\left((X_i - x)/h(x)\right)/h(x)$ and $h(x)$ denotes either a variable bandwidth or a constant bandwidth h. The natural estimator for $\hat{M}(x_0)$ is $\hat{a_0}$. In fact, it is the solution to the following weighted least squares problem:

$$\hat{a} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}Y, \tag{5}$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \ldots & (X_1 - x_0)^\nu \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x_0) & \ldots & (X_n - x_0)^\nu \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, a = \begin{pmatrix} a_0 \\ \vdots \\ a_\nu \end{pmatrix}$$

and $\mathbf{W} = \text{diag}\{K_h(X_i - x_0)\}$. The performance of $\hat{a}_0$ is better when choosing ν odd. Fan and Gijbels (1996) give a detailed description of local polynomial modelling and derive adequate rules in order to choose ν.

Density estimation Under the usual conditions on the marginal density $f(X)$, it is estimated by

$$\hat{f}(x) = \frac{1}{n}\sum_i K_h(X_i - x),$$

which is the so called Parzen-Rosenblatt estimator.

Variance estimation The unknown variance σ^2 is estimated by

$$\hat{\sigma}^2 = \sum_i \left(Y_i - \hat{M}(X_i)\right)^2 \hat{f}(X_i). \tag{6}$$

Our model can be extended to a heteroscedastic one with an unknown variance function. In this case the local polynomial variance function estimator of Ruppert, Wand, Holst and Hössjer (1997) can be used.

Optimal Choice of the Bandwidth The choice of the bandwidth is crucial in nonparametric regression and density estimation. The correct choice of bandwidth is necessary for obtaining pointwise or uniform consistency of the nonparametric estimators.

Optimality of a bandwidth requires that it minimizes the asymptotic mean integrated squared error (AMISE) of the estimator. The asymptotically optimal constant bandwidth for regression function estimation when choosing $\nu = 1$ and using the Epanechnikov-Kernel is given by

$$h_{opt} = 1.719 \left[\frac{\int \sigma^2 w(x)/f(x)dx}{\int M''(x)^2 w(x)dx}\right]^{1/5} n^{-1/5}, \tag{7}$$

where $w : \mathbb{R} \mapsto \mathbb{R}$ is an arbitrary weight function with $\int_{\mathbb{R}} w(x)dx = 1$. The function f denotes the marginal density of X, whilst x are the design points. To obtain the value of the constant, refer to Fan and Gijbels (1996). Note that $2\hat{a}_2$ may serve as an estimate of M'' and $\hat{\sigma}^2$ is an estimate of σ^2 as defined in the foregoing paragraph. The optimal variable bandwidth is proportional to

$$h_{opt}(x) \propto \left[\frac{\sigma^2}{M''(x)^2 f(x)}\right]^{1/5} n^{-1/5}, \tag{8}$$

where the denominator has to be bounded away from 0 for all x. If the denominator tends to 0, the optimal constant bandwidth can be chosen. One can show (see Fan and Gijbels (1992)) that the AMISE when using a constant bandwidth selection is at least the AMISE when using a variable bandwidth selection. Figure 3c shows the resulting h and $h(x)$, based on the shape of $f(X)$ (Figure 3a) and $M''(X)$ (Figure 3b).

Plug-In Method As a matter of fact, the nonparametric estimators depend on the bandwidth, and the optimal bandwidth depends on unknown quantities that have to be estimated. In turn, these estimates depend upon the initial selection of the bandwidth. Therefore, it is straightforward to estimate the unknown quantities M, M'', h and σ^2 in an iterative way:

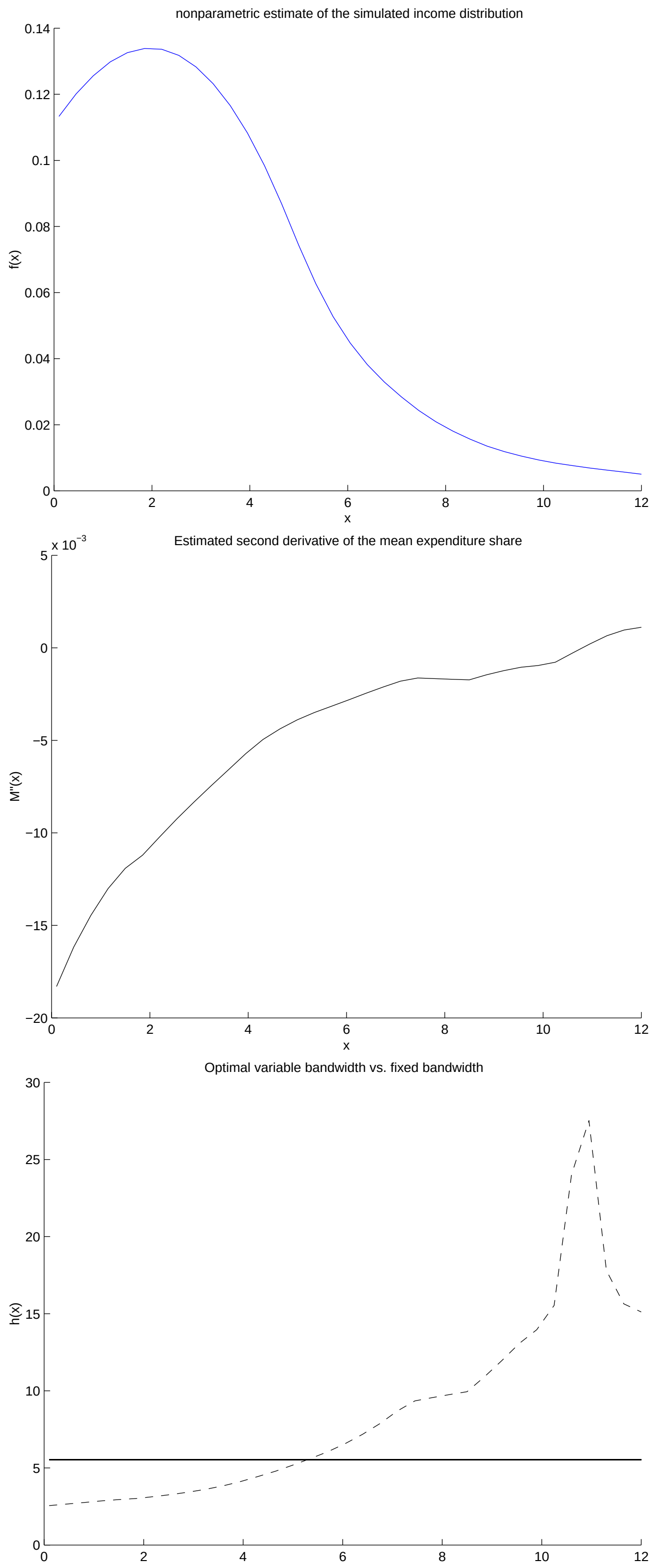

Figure 3: 3a) density estimate of $f(X) = LogN(0.5, 1)$, 3b) estimated 2nd derivative of $M(X) = 0.3 + 0.0167\text{lnX}$, 3c) constant vs. variable bandwidth

1. Crude determination of a bandwidth (h_0) by using a rule of thumb.

2. Use this bandwidth in order to estimate the unknown quantities.

3. Compute new bandwidths while using $\hat{M}$, $\hat{M}''$, $\hat{\sigma^2}$ and $\hat{f}$.

4. Repeat steps 2 and 3 until convergence.

This method has been suggested by for example Fan and Gijbels (1996).

2 Semiparametric Estimation

2.1 Transformation of Covariables

This section presents an estimation procedure for the estimation of a unknown mean regression function. It captures an idea of Härdle and Marron (1990), who consider similar shaped regression functions which are related in a parametric way. Shape preserving transformations of these curves are used to obtain a more accurate final estimate of the regression function, that makes use of all available data. Accordingly, we describe a method that allows for pooling subsamples in order to obtain a more precise nonparametric estimate of the mean regression function.

Suppose (X, Y, Z) are observable random variables with $(X_i, Y_i, Z_i)_{i=1,\dots,n} \in \mathbb{R} \times \mathbb{R} \times \mathbb{I}$ independent realizations. There are two independent covariates: let X denote the household income and Z denote an index devoted to a homogeneous subgroup z of the households i. Hence card$\{\mathbb{I}\}$ is the number of homogeneous subgroups. Furthermore let the following equation hold:

$$E(Y|X,Z) = A_Z\left(m(T_Z(X))\right) \ , \tag{9}$$

where $A_Z : \mathbb{R} \mapsto \mathbb{R}$ and $T_Z : \mathbb{R} \mapsto \mathbb{R}$ are known bijective transformation functions with unknown parameter vectors α_z and β_z respectively. Note that the unknown function $m \in \mathcal{C}^\nu : \mathbb{R} \mapsto \mathbb{R}$ does not depend on Z, where ν denotes a positive integer.

In general, the underlying model (9) can be used to combine univariate data sets that are linked via shape invariant or by shape preserving transformations. The subsamples may represent either the results of experiments or disaggregated subsamples of a larger data set, i.e. homogeneous subgroups of households (Figure 2).

Before going into the details of the estimation, it is necessary to shed some light on the transformation functions and on the discrete covariable Z. The parameter of the transformation functions differs in z. Therefore one has card$\{\mathbb{I}\}$ vectors of unknown parameters. For reasons of identification it is indispensable either to have prior information on the parameter values for one $z \in \mathbb{I}$ or to assume $\|\alpha\| = \|\beta\| = 1$. Let us consider the first case and denote z_0 as the corresponding z. It will serve as a basis for the transformations. It is reasonable, if possible, to take the basis such that the transformation functions A_Z and T_Z are the identity functions. Then, the relationship (9) simplifies to

$$E(Y|X, Z = z_0) = m(X).$$

We require one additional assumption on the structure of the model:

Assumption 2 $A_Z^{-1}(Y|X,Z) \sim N\left(E(A_Z^{-1}(Y|X,Z)), \sigma_Z^2\right)$

This assumption ensures the applicability of usual (weighted) least squares estimation.

Remark 1 *The conditional density functions of $X|Z$ are independent of Z, i.e. $f(X|Z) = f(X)$, $\forall Z$. This is due to the independency of X and Z.*

Assumption 3 *If $T_z(x)$ is not the identity function for all $z \in \mathbb{I}$, the function m has to be nonlinear and noncycling on its observed support. Moreover the parameter space of β has to be restricted to some extent in order to ensure that the nonparametric estimates are indeed comparable on the same support.*

A discussion of Assumption 3 can be found in Wilke (2001) for a similar class of models. Suppose that Assumptions 1, 2, 3, the regularity conditions on the model and on the Kernel function hold. A five step estimation procedure in order to estimate the mean regression function $M(X)$ is presented in what follows:

1. For all $z \in \mathbb{I}$ and all admissible α_z and β_z, determine $T_z(X)$ and $A_z^{-1}(Y)$.

2. Estimate $m(T_z(X))$ by minimizing

$$\sum_{i|Z_i=z} \left(A_z^{-1}(Y_i) - \sum_{j=0}^{\nu} a_j \left(T_z(X_i) - s\right)^j \right)^2 K_h\left(s - T_z(X_i)\right)$$

   with respect to a_j for all z, α_z and β_z, where s are the design points. Consequently, $\hat{a}_0 = \hat{m}_{\alpha_z, \beta_z}$.

The estimation of the transformation parameters α_z and β_z is done by weighted least squares. In order to measure the goodness of the transformations we use the integrated squared error (ISE). Minimizing the ISE corresponds to finding the most accurate fit of the transformations. Accordingly, let L be a loss function given by

$$L(\alpha_z, \beta_z) = \int_{\mathbb{R}} \left(\hat{m}(T_{z_0}(x)) - \hat{m}_{\alpha_z,\beta_z}(T_z(x))\right)^2 w(x)dx, \tag{10}$$

where $w(x)$ is a nonnegative weight function. It is reasonable to choose $w(x) = \hat{f}(x)$, because in this case the reliability of the nonparametric subgroup regression estimates is implemented into the problem.

3. In order to obtain $\hat{\alpha}_z$ and $\hat{\alpha}_z$ minimize $L(\alpha_z, \beta_z)$ with respect to α_z and β_z for all $z \in \mathbb{I}$. Denote the resulting parameter estimates as $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. The size of these matrices depends on the number of parameters (columns) and on card$\{\mathbb{I}\}$ (rows).

4. The vectors of the weighted mean transformation parameters $\hat{\alpha}$ and $\hat{\beta}$ are obtained by

$$\hat{\alpha} = \sum_{z \in \boldsymbol{I}} \frac{\text{card}\{Z_i = z\}}{n} \hat{\boldsymbol{\alpha}}_z \quad \text{and} \quad \hat{\beta} = \sum_{z \in \boldsymbol{I}} \frac{\text{card}\{Z_i = z\}}{n} \hat{\boldsymbol{\beta}}_z,$$

   where $\hat{\boldsymbol{\alpha}}_z$ denotes the z'th row of $\hat{\boldsymbol{\alpha}}$, using the same notation for $\hat{\boldsymbol{\beta}}$.

The unknown regression function m can now be estimated by using all observations.

5. Estimate m with data of the form $\left(\hat{A}_{Z_i}^{-1}(Y_i), \hat{T}_{Z_i}(X_i)\right)$. Accordingly,

$$\sum_{i=1}^{n} \left(\hat{A}_{Z_i}^{-1}(Y_i) - \sum_{j=0}^{\nu} a_j(\hat{T}_{Z_i}(X_i) - s)^j\right)^2 K_h\left(s - \hat{T}_{Z_i}(X_i)\right)$$

   is minimized with respect to a_j, where s are the design points. The solution $\hat{a}_0$ corresponds to $\hat{m}$. Note that this estimate of m takes into account *all* observations. Therefore, it determines the shape of m in the most accurate way.

6. Let A and T denote the transformation functions with average parameter values α and β. The mean regression function M is defined as

$$M(X) = A\left(m(T(X))\right),$$

   if A_Z and T_Z are linear in α_Z and β_Z. In this case, the natural estimator of M is given by

$$\hat{M}(X) = \hat{A}\left(\hat{m}(\hat{T}(X))\right).$$

Using the full sample size for the final nonparametric estimate is the main advantage of this approach. Asymptotic properties are not treated in detail here but consistency can be shown using the same framework as in Wilke (2001): under appropriate regularity conditions on the model and the bandwidth selection rule, the nonparametric estimators are uniformly consistent. The consistency of the parametric estimator can then be shown by using the nonlinear least squares estimation framework.

The above defined estimation procedure can also perform well in the case of similarly shaped regression functions m_z, i.e. limited dependency on z:

$$\|m_{z_0}(x) - m_z(x)\|_2 \leq \epsilon$$

for all $z \in I\!I$, where $\epsilon \geq 0$ is an arbitrarily chosen constant. The trade-off is as follows: Using all data improves the identification of m, but combining different functions may result in a biased estimate of the parameters. For further details see for example Pinkse and Robinson (1995).

2.2 Examples

Härdle and Marron (1990) and Pinkse and Robinson (1995) Härdle and Marron (1990) consider the following example. It treats transformations of the vertical axis and the horizontal axis. In particular, they consider

$$T_Z(X) = X + \beta_Z \quad \text{and} \quad A_Z\left(m(T_Z(X))\right) = m(T_Z(X)) + \alpha_Z.$$

Consequently, the matrices of unknown parameters reduce to vectors and therefore α and β, the vectors of average transformation parameters, consist of only one element.

Härdle and Marron have shown consistency and asymptotic normality of the resulting parameter estimates when using the Nadaraya-Watson Kernel estimator with nonstochastic regressors. Moreover, they derived asymptotic results for a wider class of transformation functions that are also treated in the next example.

Pinkse and Robinson consider shape preserving transformation functions in a stochastic regressors framework. Let A be an affine transformation function that is taken to be

$$A_Z(m) = \alpha_{1,Z} + \alpha_{2,Z}m(T_Z(x))$$

and let $T_Z(X)$ be an invertible transformation function with unknown parameters β. The loss function for $Z = z$ is therefore given by

$$L(\alpha_{1,z}, \alpha_{2,z}, \beta_z) = \int_{\mathbb{R}} \left(\hat{m}(T_{z_0}(x)) - \alpha_{1,z} - \alpha_{2,z}(\hat{m}_{\alpha_z,\beta_z}(T_z(x)))\right)^2 w(x)dx$$

and the final estimate of m is obtained by minimizing

$$\sum_{i=1}^{n} \left(\frac{Y_i - \hat{\alpha}_{1,Z_i}}{\hat{\alpha}_{2,Z_i}} - \sum_{j=0}^{\nu} a_j (\hat{T}_{Z_i}(X_i) - s)^j \right)^2 K_h \left(s - \hat{T}_{Z_i}(X_i) \right)$$

with respect to a_j.

Pinkse and Robinson have shown $\sqrt{n}$-consistency and asymptotic normality of the parameter estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\beta}$. However, in their paper they have chosen a specification of the loss function that is based on the definition of the Nadaraya-Watson estimator and different to what we have presented here. As shown in Wilke (2001) by simulations, the finite sample performance of an estimator using their particular specification is worse in many applications, because it can involve larger expected bias of the nonparametric estimators.

PIGLOG expenditure shares Suppose that although economic theory proposes that Engel curves are of the PIGLOG form as given in (1), there is little faith in this specification. In particular, the specification $m(x) = \ln x$ seems too specific. Therefore, we intend to identify the shape of $m(x)$ with the help of a nonparametric estimator. Accordingly, we assume

$$M(X, Z) = A_Z\left(m(X)\right) = \alpha_{1,Z} + \alpha_{2,Z} m(X),$$

where the function T is now the identity function and A is an affine transformation function. Therefore we have a special case of the Pinkse and Robinson model and equivalent asymptotic properties of the parameter estimates can be inferred immediately. In order to estimate m, we have to minimize

$$\sum_{i=1}^{n} \left(\frac{Y_i - \hat{\alpha}_{1,Z_i}}{\hat{\alpha}_{2,Z_i}} - \sum_{j=0}^{\nu} a_j (X_i - x)^j \right)^2 K_h \left(x - X_i \right)$$

with respect to a_j. Note that for this and the following example we require Assumption 3.

PIGLOG expenditure shares with price dependency Suppose that the population is homogenous, i.e. there is only one subgroup, and we observe this group at various dates. By definition, panel data fulfills this condition, but cross section may satisfy it, too. Assume that the expenditure share of this group depends on time due to price variation, for example. The model can then be written as

$$M(X,Z) = A_Z\left(m(X)\right) = \alpha_{1,Z} + \alpha_{2,Z} m(X),$$

where $Z = Z_i$ now represents an observation date $Z_i \in I\!\!I$. One may also combine this specification with the modelling of heterogeneity. Then, we have expenditure shares depending on time and subgroup specific behavior.

2.3 Semiparametric fitting of basis functions

In this section, the regression function is assumed to satisfy

$$E(Y|X,Z) = m\left(X,Z\right), \tag{11}$$

where $X \in \mathbb{R}$ and $Z \in I\!\!I$, the covariables, are again independent observable random variables with $i = 1,\ldots,n$ independent realizations. The function $m \in \mathcal{C}^\nu : \mathbb{R} \times I\!\!I \mapsto \mathbb{R}$ is an unknown function which does not necessarily have the bijective property. Let X_i denote the income of a household and let Z_i denote the index devoted to its homogeneous subgroup. In order to keep things simple, let us again assume

$$Y|X,Z \sim N\left(E(Y|X,Z), \sigma_z^2\right)$$

and remark that $f(X|Z) = f(X)$ as in the previous approach.

The idea of this section is to approximate the function m by a linear combination of a known basis function $T : \mathbb{R} \mapsto \mathbb{R}^{\mathrm{P}}$, which has an image of full rank, and unknown transformation parameters $t_z \in \mathbb{R}^{\mathrm{P}}$:

$$m(X|Z) \approx t_Z' T(X)$$

where P, the dimension of the approximation, determines the degree of smoothing and should in fact be the rank of the underlying demand system. In many cases the basis functions can be chosen by reference to a guess that has been provided by theoretical analysis. Ramsay and Silverman (1997) call this approach "fitting of basis functions". Like in Section (2.1), the first aim is to estimate the values of the transformation parameters t_z. Using matrix notation, the least squares estimator for t_z is given by

$$\hat{t}_z = (\mathbf{T}_z'\mathbf{T}_z)^{-1}\mathbf{T}_z' Y_z \tag{12}$$

where $\mathbf{T}_z$ is the card$\{I\!I\} \times P$ matrix of basis function values with $T(X_i)'$ as the i'th row such that $Z_i = z$. Y_z denotes the card$\{Z_i = z\} \times 1$ vector of $Y_i|Z_i = z$. Note that $\mathbf{T}_z$ has full rank. There is obviously some relationship to the loss function criterion from the previous section, since we are minimizing the norm $\|Y_z - \mathbf{T}_z t_z\|_2$. Moreover, this estimator can easily be adjusted by a weighting scheme. The corresponding loss function is then given by

$$L(t_z) = \int_{\mathbb{R}} \left(\hat{m}_z(x) - t_z' T(x)\right)^2 w(x) dx,$$

where $\hat{m}_z(x)$ is a nonparametric estimate of $m(x, z)$. As already mentioned in the previous section, we propose a nonparametric estimate of the marginal density of x as a weight function. Minimizing the loss function is equivalent to minimizing the weighted norm. The weighted least squares estimator for t_z can therefore be written as

$$\hat{t}_z = (\mathbf{T}_z' \mathbf{W} \mathbf{T}_z)^{-1} \mathbf{T}_z' \mathbf{W} \hat{m}_z$$

where $\mathbf{T}_z$ is the same as in (12) and $\mathbf{W}$ denotes the matrix of weights. Note that $\mathbf{W}$ does not depend on z. Solving the equation for all $z \in I\!I$ yields $\hat{\boldsymbol{t}}$, the card$\{I\!I\} \times P$ matrix of parameter estimates.

In order to estimate the mean regression function, we have to compute the averaged estimated transformation parameters $\hat{t}$ as in the foregoing section. Then, we are able to transform the data to the mean level before estimating the mean regression function, $M(X)$, with the local polynomial smoother by using *all* observations.

Example: QUAIDS Suppose we have a model of the form

$$E(Y|X, Z) := t_{1Z} + t_{2Z}\ln X + t_{3Z}(\ln X)^2$$

where $Y \sim N(E(Y|X, Z), \sigma_Z^2)$. Accordingly, $T(X)' = (1 \; \ln X \; (\ln X)^2)$, $t_z' = (t_{1z} \; t_{2z} \; t_{3z})$ and $P = 3$. The mean regression function is given by

$$M(X) = t_1 + t_2 \ln X + t_3 (\ln X)^2$$

where

$$t_p = \sum_{z \in I} \frac{\text{card}\{Z = z\}}{n} t_{zp} \quad \text{for } p = 1, 2, 3.$$

The loss function now becomes

$$L(t_z) = \int_{\mathbb{R}} \left(\hat{m}_z(x) - t_{1z} - t_{2z}\ln x - t_{3z}(\ln x)^2\right)^2 w(x) dx.$$

After computing the estimated mean transformation parameters $\hat{t}$, one has to transform the original data to the mean level $(\bar{Y}_i, X_i)$, where

$$\bar{Y}_z = Y_z + (\hat{t} - \hat{t}_z)' T(X_z)$$

for all $z \in \mathbb{I}$. Then the mean regression function M is estimated by minimizing

$$\sum_{i=0}^{n} \left(\bar{Y}_i - \sum_{j=0}^{\nu} a_j (X_i - x)^j \right)^2 K_h (x - X_i)$$

with respect to a_j, where $\hat{a}_0 = \hat{M}$.

Example: Working-Leser The estimation of Working-Leser Engel curves is similar and differs only in the rank, i.e. $P = 2$, and the basis function specification: $T(X)' = (X \; X \ln X)$.

3 Simulations

This section investigates the performance of the two proposed semiparametric estimation procedures by simulation studies. In order to measure the performance, we run a series of 50 simulations with five different estimators:

- **A** nonparametric estimator using a fixed bandwidth without taking into account Z, i.e. nonparametric estimate of $E(Y|X)$.
- **B** semiparametric fit using a fixed bandwidth for the mean regression function but a variable bandwidth for the subgroup regression functions
- **C** nonparametric estimator using a variable bandwidth without taking into account Z
- **D** semiparametric fit using a variable bandwidth
- **E** parametric fit; OLS is used in order to estimate the subgroup regression functions and the mean regression function.

In all simulations we have $z \in \{1, 2, 3\} = \mathbb{I}$. Therefore $\text{card}\{\mathbb{I}\} = 3$. The simulated data is such that $\text{card}\{Z_i = 1\} = \text{card}\{Z_i = 2\} = \text{card}\{Z_i = 3\} = 200$. Hence, in terms of X it is a random effects model and in terms of Z it is a fixed effects model (in contrast to the theoretical part). The samples are drawn from the following distributions:

$$(Y|X, Z) \sim N(E(Y|X, Z), 0.05)$$

and

$$f(X) = LogN(0.5, 1).$$

Matlab 5.3 serves as the software environment. We use the random number generators of the toolbox 'Statistics'. All non- and semiparametric estimators are self- programmed.

In order to investigate the performance of the Methods A-E, we apply them in six different models:

True Model Specification

1. PIGLOG (Example of Section 2.2) ⇒ Table 1, Figure 4
2. QUAIDS, (Example of Section 2.3) ⇒ Table 2, Figure 5

Misspecification

3. Specification: QUAIDS
   True Model: PIGLOG ⇒ Table 3, Figure 6
4. Specification: PIGLOG
   True Model: QUAIDS ⇒ Table 4, Figure 7
5. Specification: QUAIDS
   True Model: PIGLOG with an additional sine-function ⇒ Table 5, Figure 8
6. Specification: PIGLOG
   True Model: PIGLOG with an additional sine-function ⇒ Table 6, Figure 9

The true values of the transformation parameters differ across the six models as is apparent from the tables.

Resulting Performance In order to measure the performance of the estimators A-E, the mean averaged squared error (MASE) is used:

$$MASE = \frac{1}{S}\sum_{s=1}^{S} MSE(s) = \frac{1}{S}\sum_{s=1}^{S}\left[\left(E\hat{M}(s) - M(s)\right)^2 + \text{var}(\hat{M}(s))\right],$$

where $s \in (0, 12]$ denote the evaluation points. The nonparametric regression function estimates are obtained with the local polynomial smoother, where ν is always chosen to be one (local linear smoother), except for the estimation of the second derivative of the regression function, which is done with $\nu = 3$.

Tables 1-6 and Figures 4-9 present the results of the finite sample performance of the methods. The following list summarizes the main findings:

- The nonparametric estimators A and C have serious problems in areas with low density in the data, e.g. relatively rich people. This is in accordance with the empirical findings of Banks et al. (1997) (see Figure 1). In addition, estimator A in Case 4 indicates a linear underlying model structure, but in fact the true model is far linear. However, estimators A and C would become more precise by increasing the sample size.

- The parametric estimator E is biased in the case of misspecification. In this case, it is unable to detect the true underlying model structure. Under the true model specification it outperforms the other estimators since the rate of convergence of parametric estimators is higher.

- The choice of a variable bandwidth in five cases results in a lower MASE than choosing a fixed bandwidth. Therefore estimator D should be preferred over B, and C over A. This is in accordance with the theoretical findings of Fan and Gijbels (1992).

- Considering all model specifications, estimator D appears to be a good choice for the estimation of mean expenditure shares.

4 Conclusion

Two semiparametric approaches to the estimation of mean expenditure shares have been presented and have been compared to usual parametric and nonparametric methods. It is well known that parametric estimators perform best if the underlying model specification is correct, otherwise they are not consistent. Nonparametric estimators are consistent for a broader class of models but require more observation in order to obtain precise results. They have a poor performance in areas with low density in the data. The simulation study shows that the two suggested semiparametric estimators reduce these main disadvantages. They are more flexible than parametric estimators and have a higher convergence rate than nonparametric estimators. The latter is of particular importance when working with small samples. The simulations attest this aspect of finite sample performance of the estimators. In addition, the simulations indicate that choosing a variable bandwidth is superior to choosing a fixed bandwidth.

It remains to conclude that the proposed methods for the modelling of behavioral heterogeneity perform well in finite samples. Hence, the econometrician is not obliged to suppose household behavior as a result of utility maximization. Behavioral heterogeneity need no longer be considered a burden, but rather an important determinant of the mean expenditure share.

Performance		
Method	MASE	MASE/MASE(E)
A	$8.9443e^{-004}$	6241%
B	$6.7665e^{-005}$	472%
C	$9.2802e^{-004}$	6475%
D	$5.7859e^{-005}$	403%
E	$1.4332e^{-005}$	100%

Parametric Part				
Method	$\alpha = 0.3$	$\beta = 0.0167$	var$\hat{\alpha}$	var$\hat{\beta}$
D	0.2999	0.0167	$0.4757e^{-004}$	$0.5059e^{-004}$
E	0.3001	0.0170	$0.1525e^{-004}$	$0.1582e^{-004}$

Table 1: PIGLOG, Specification=True Model,$\alpha + \beta lnx$

Performance		
Method	MASE	MASE/MASE(E)
A	$9.8756e^{-004}$	582%
B	$1.0351e^{-004}$	61%
C	$9.5053e^{-004}$	560%
D	$6.8715e^{-005}$	41%
E	$1.6966e^{-004}$	100%

Parametric Part						
Method	$\alpha = 0.533$	$\beta = 0.05$	$\lambda = 0$	var$\hat{\alpha}$	var$\hat{\beta}$	var$\hat{\lambda}$
D	0.5250	0.0658	-0.0066	$0.2879e^{-004}$	$0.1874e^{-004}$	$0.0839e^{-004}$
E	0.5338	0.0499	-0.0029	$0.2525e^{-004}$	$0.1616e^{-004}$	$0.06e^{-004}$

Table 2: *QUAIDS, Specification=True Model,* $\alpha + \beta lnx + \lambda(lnx)^2$, $\lambda_z \neq 0$ for all z

Performance		
Method	MASE	MASE/MASE(E)
A	$1.5e^{-003}$	1340%
B	$6.6994e^{-005}$	60%
C	$9.7729e^{-004}$	873%
D	$4.7617e^{-005}$	43%
E	$1.1194e^{-004}$	100%

Parametric Part						
Method	$\alpha = 0.5333$	$\beta = 0.033$	$\lambda = 0$	var$\hat{\alpha}$	var$\hat{\beta}$	var$\hat{\lambda}$
D	0.5308	0.0366	-0.0012	$0.2428e^{-004}$	$0.2164e^{-004}$	$0.1002e^{-004}$
E	0.5332	0.0334	-0.0021	$0.1919e^{-004}$	$0.1781e^{-004}$	$0.0528e^{-004}$

Table 3: *Specification: QUAIDS, True Model:* $\alpha + \beta lnx + \lambda(lnx)^2$, $\lambda_z = 0$ for all z

Performance		
Method	MASE	MASE/MASE(E)
A	$6.1e^{-003}$	16%
B	$5.1e^{-003}$	13%
C	$8.1126e^{-004}$	2%
D	$3.1116e^{-004}$	1%
E	$3.92e^{-002}$	100%

Parametric Part				
Method	$\alpha = 0.3$	$\beta = 0.00167$	var$\hat{\alpha}$	var$\hat{\beta}$
D	0.4191	0.0564	$0.3800e^{-004}$	$0.2950e^{-004}$
E	0.3834	0.1013	$0.1582e^{-004}$	$0.1591e^{-004}$

Table 4: *Specification: PIGLOG, True Model:* $\alpha + \beta lnx + \lambda(lnx)^2$, $\lambda_z = 0.1$ for all z

Performance		
Method	MASE	MASE/MASE(E)
A	$2.37e^{-002}$	527%
B	$8.0753e^{-004}$	18%
C	$2.14e^{-002}$	476%
D	$3.0088e^{-004}$	7%
E	$4.5e^{-003}$	100%

Parametric Part						
Method	$\alpha = 0.5333$	$\beta = 0.0333$	$\lambda = 0$	var$\hat{\alpha}$	var$\hat{\beta}$	var$\hat{\lambda}$
D	0.6050	0.0238	-0.0240	$0.3447e^{-004}$	$0.7777e^{-004}$	$0.2316e^{-004}$
E	0.5981	0.0240	-0.0119	$0.1677e^{-004}$	$0.1762e^{-004}$	$0.0539e^{-004}$

Table 5: *Specification: QUAIDS, True Model:* $\alpha + \beta lnx + \lambda(lnx)^2$*+0.1sinx*, $\lambda_z = 0$ for all z

Performance		
Method	MASE	MASE/MASE(E)
A	$47.7e^{-003}$	96%
B	$1.7e^{-003}$	35%
C	$4.5e^{-003}$	92%
D	$9.4963e^{-004}$	19%
E	$4.9e^{-003}$	100%

Parametric Part				
Method	$\alpha = 0.5$	$\beta = 0.0167$	var$\hat{\alpha}$	var$\hat{\beta}$
D	0.5402	0.0018	$0.6281e^{-004}$	$0.6281e^{-004}$
E	0.5564	-0.0077	$0.1839e^{-004}$	$0.1434e^{-004}$

Table 6: *Specification: PIGLOG, True Model:* $\alpha + \beta lnx + 0.1 sinx$

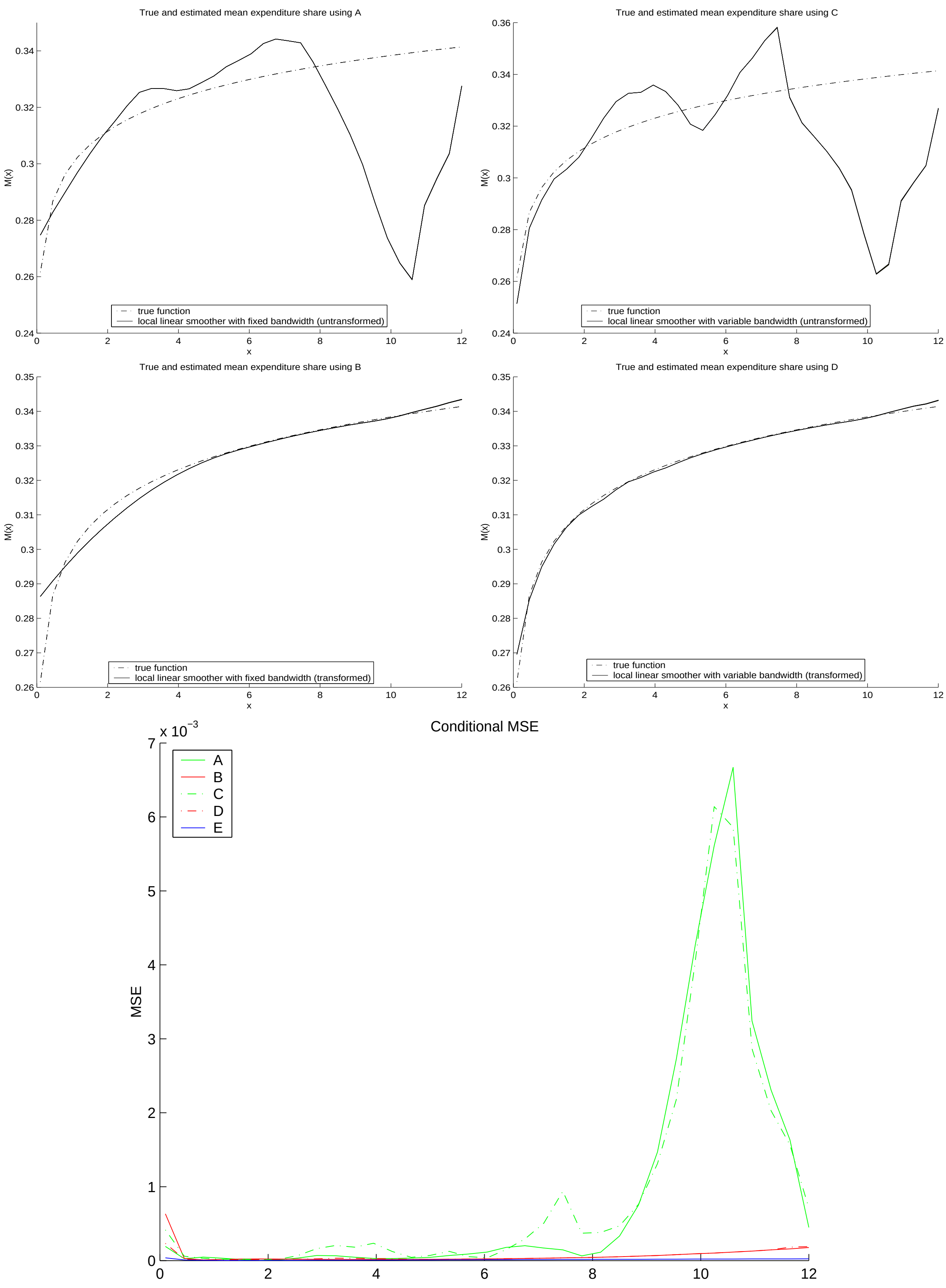

Figure 4: Case 1; PIGLOG

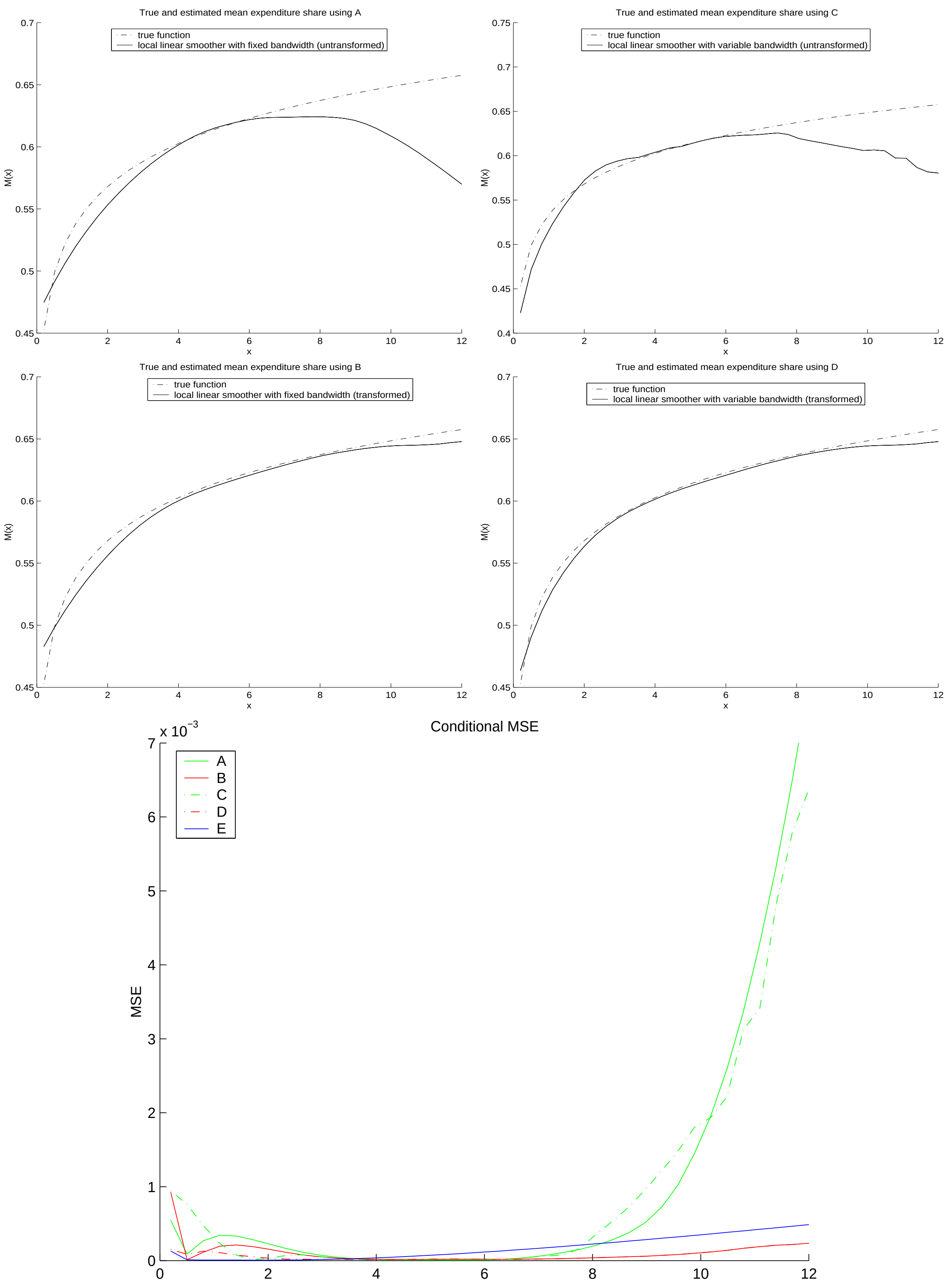

Figure 5: Case 2; QUAIDS

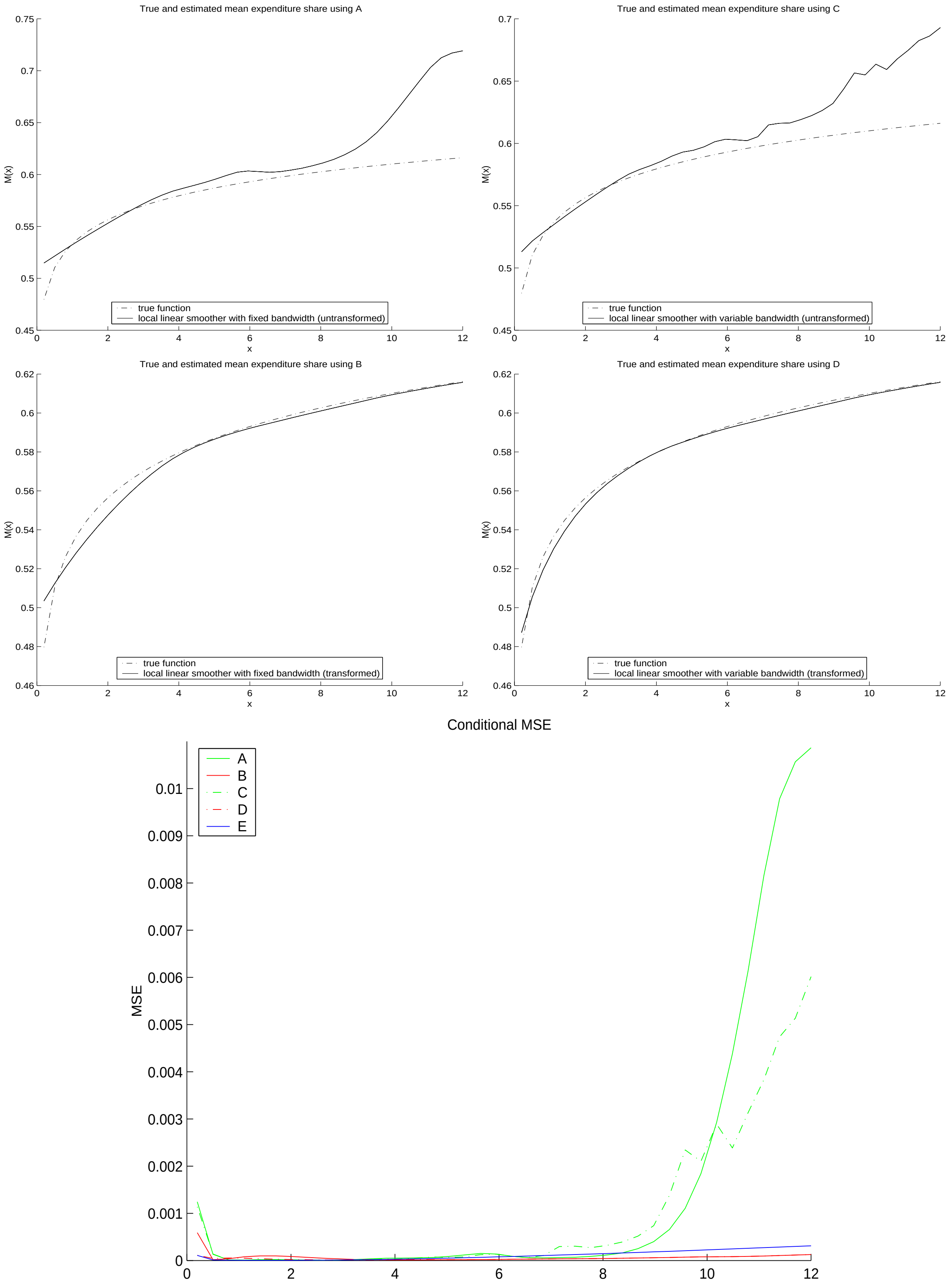

Figure 6: Case 3; Specification: QUAIDS, True Model: PIGLOG

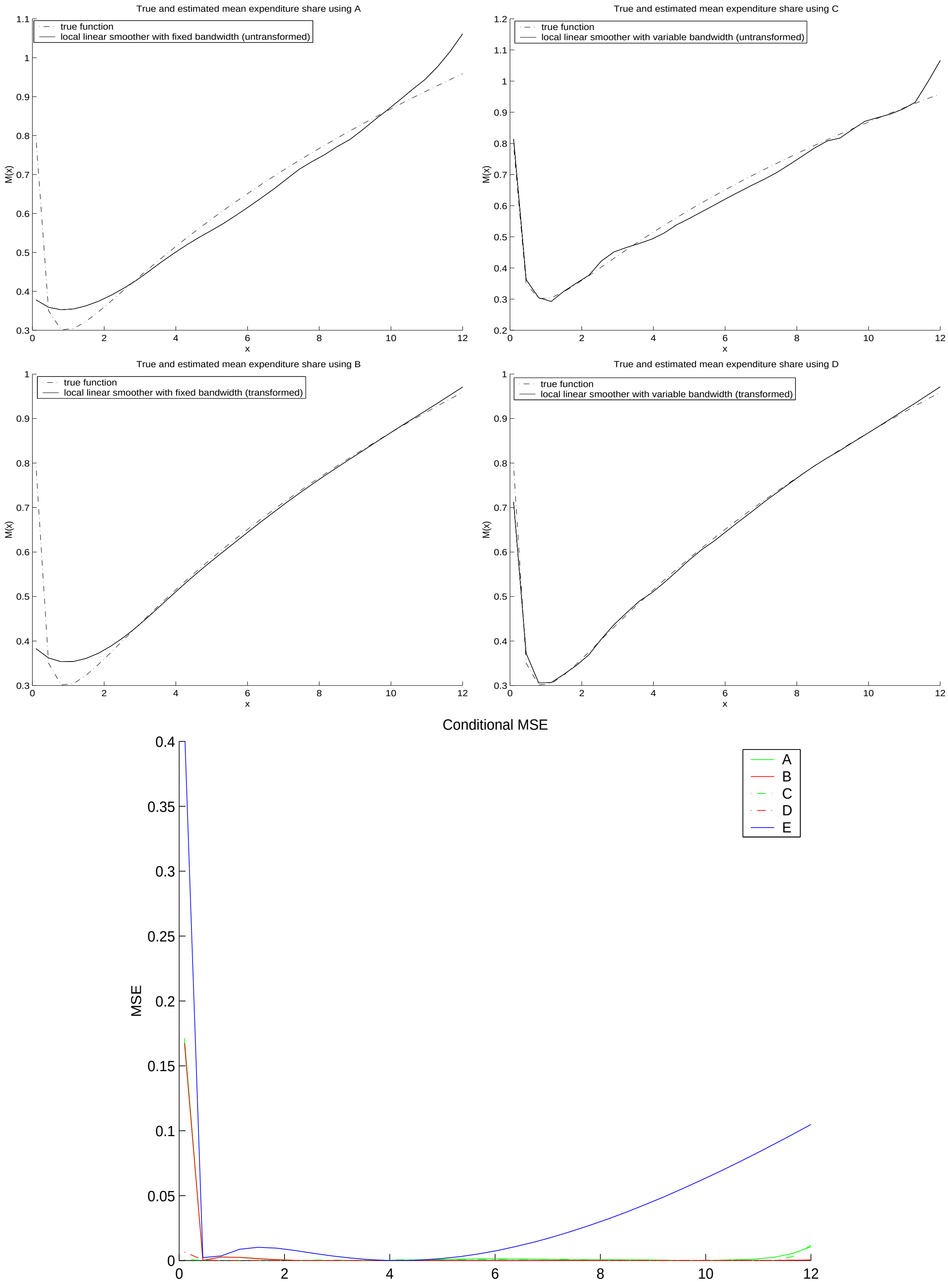

Figure 7: Case 4; Specification: PIGLOG, True Model: QUAIDS

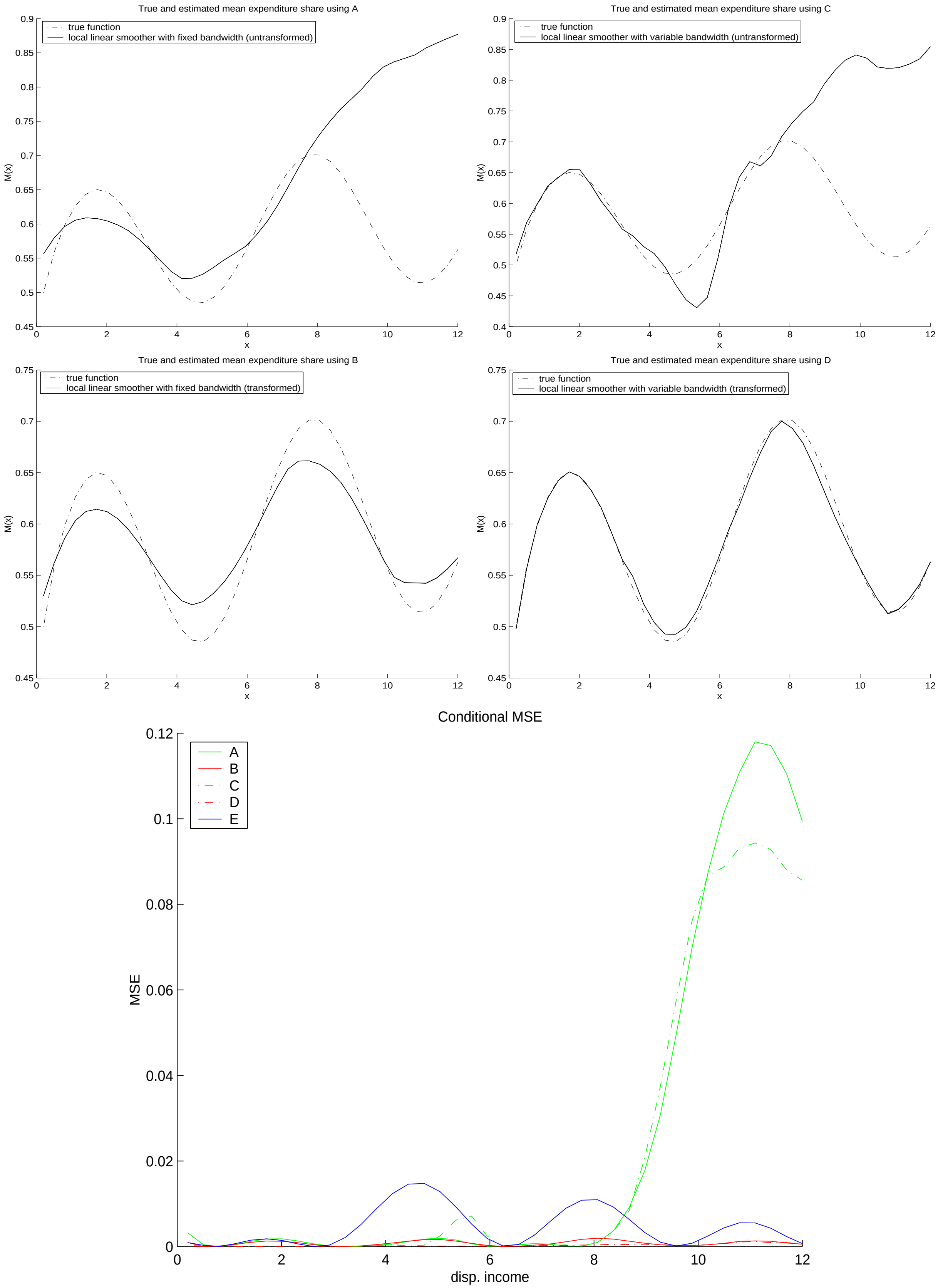

Figure 8: Case 5; Specification: QUAIDS, True Model: PIGLOG + Sine-function

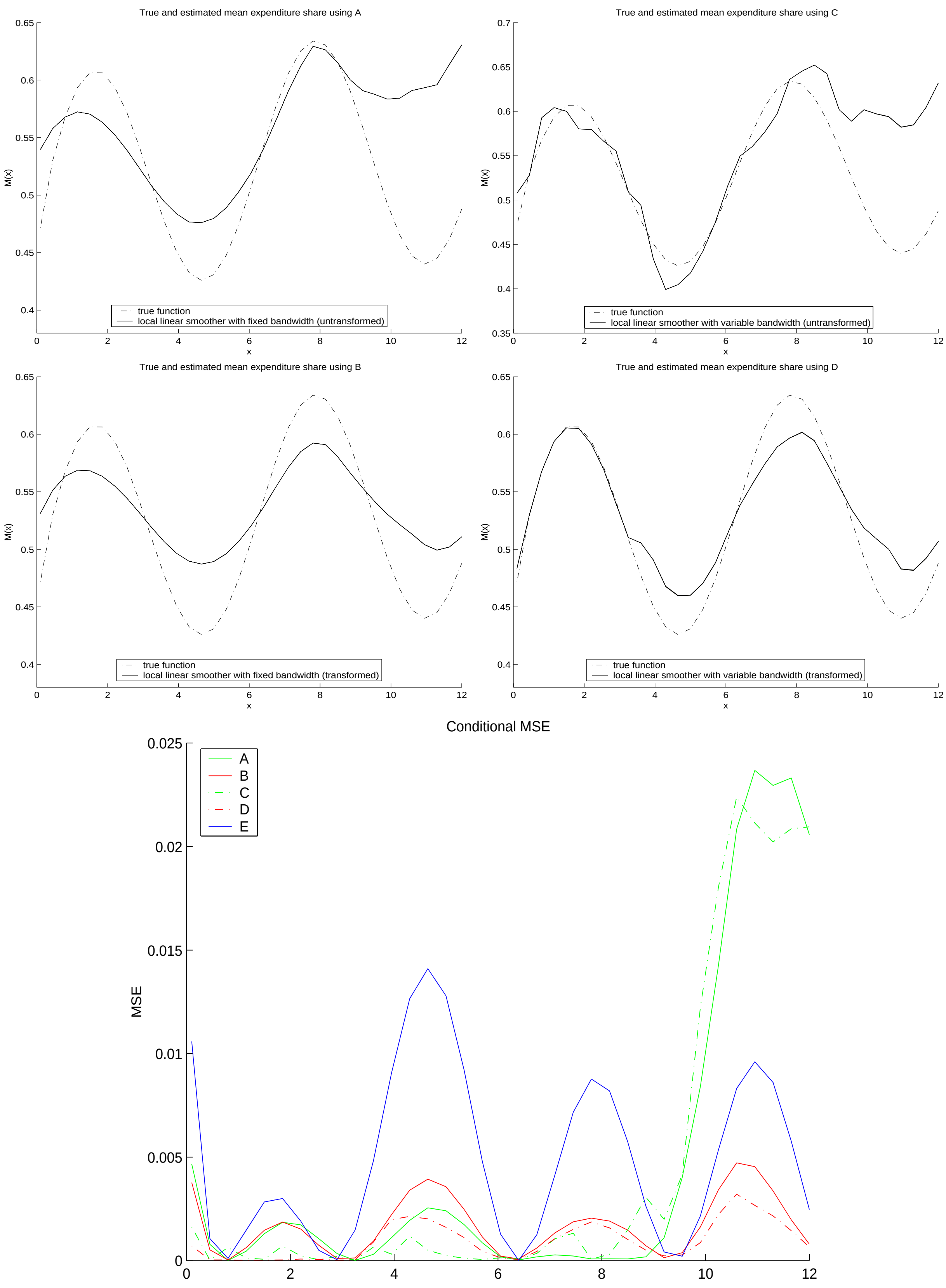

Figure 9: Case 6; Specification: PIGLOG, True Model: PIGLOG + Sine-function

References

[1] Banks, J. and Blundell, R. and Lewbel, A., Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics*, Vol.LXXIX (No 4)(1997), 527–539.

[2] Fan, J. and Gijbels, I., Variable Bandwidth and Local Linear Regression Smoothers. *Annals of Statisitcs*, Vol.20 (No 4) (1992), 2008–2036.

[3] Fan, J. and Gijbels, I., *Local Polynomial Modelling and its Applications.* (Chapman and Hall, 1st ed., London, 1996).

[4] Fan, J. and Hall, P. and Martin, A., On Local Smoothing of Nonparametric Curve Estimators. *Journal of the American Statistical Association*, Vol.91 (No 433) (1996), 258–266.

[5] Gorman, W., Some Engel Curves. In: *The Theory and Measurement of Consumer Behavior*, (Angus Deaton (Ed.), Cambridge University Press, Cambridge, 1981)

[6] Härdle, W., Mammen, E. and Müller, M., Testing Parametric versus Semiparametric Modelling in Generalized Linear Models. Discussion Paper (1996) Universität Heidelberg

[7] Härdle, W. and Marron, J.S., Semiparametric Comparision of Regression Curves. *Annals of Statistics*, Vol.18 (No 1) (1990), 63–89.

[8] Hildenbrand, W. and Kneip, A., On Behavioral Heterogeneity. SFB 303 Discussion Paper No. A-589 (1999) ,Universität Bonn.

[9] Leser C.E.V., Forms of Engel-Functions. *Econometrica*, Vol.31 (1963), 694–703.

[10] Pinkse, C. and Robinson, P., Pooling Nonparametric Estimates of Regression Functions with a similar shape. In: *Advances in Econometrics and Quantitative Economics* (Editor, Verlag, Ort, 1995), 172–195.

[11] Ramsay,J. and Silverman B., *Functional Data Analysis* (Springer, New York, 1997)

[12] Robinson,P.M., Root-N-Consistent Semiparametric Regression. *Econometrica*, Vol.56 (No 4) (1988), 931–954

[13] Ruppert,D. and Wand,M.P. Holst,U. and Hössjer,O., Local Polynomial Variance-Function Estimation. *Technometrics*, Vol.39 (No 3) (1997), 262–273

[14] Villemeur, E.B. , Behavioral Complementarity (not heterogeneity) Causes the Law of Demand.. Manuscript (2000), Universite de Sciences Sociales Toulouse.

[15] Wilke, R. , A Note on Behavioral Heterogeneity and Aggregation. Discussion Paper in Economics No. 00-05 (2000a), Universität Dortmund.

[16] Wilke, R., Multivariate Quasi Likelihood Estimation. Discussion Paper in Economics No. 00-08 (2000b), Universität Dortmund.

[17] Wilke, R., Semiparametric Estimation of Regression Functions under Shape Invariance Restrictions. *mimeo*, University College London.

[18] Working, H., Statistical Laws of Family Expenditure. *Journal of the American Statistical Association*, Vol.38 (1943), 43–56

Essay 5

Semiparametric Estimation of Regression Functions Under Shape Invariance Restrictions

April 22, 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.

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 sufficient conditions in order to obtain consistent estimates. Moreover, a 4 step estimation procedure is defined which is computationally feasible for large samples and convenient to implement. 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 horizontal and vertical shifts with unknown parameters. The econometrician wants to identify the unknown functions as accurately as possible and 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).

The two main theoretical articles concerned with this class of semiparametric models are 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

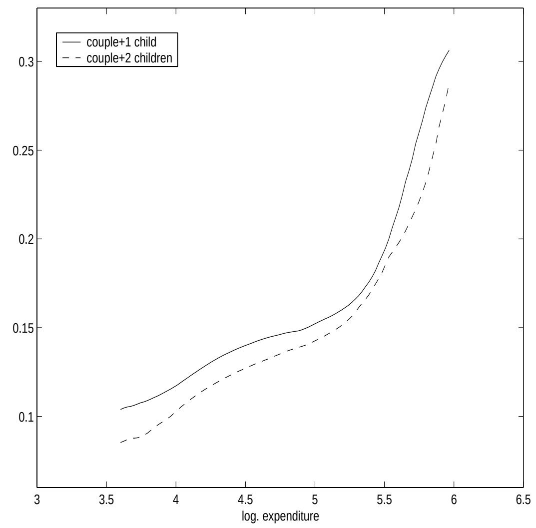

Figure 1: Nonparametric estimates of transport expenditure shares.

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 to carefully 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 sample 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 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:

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

where T is an invertible parametric transformation with $T^{-1}_{\mu_0}(W_i) \in \mathcal{X}_2$ 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) now becomes:

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

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:

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

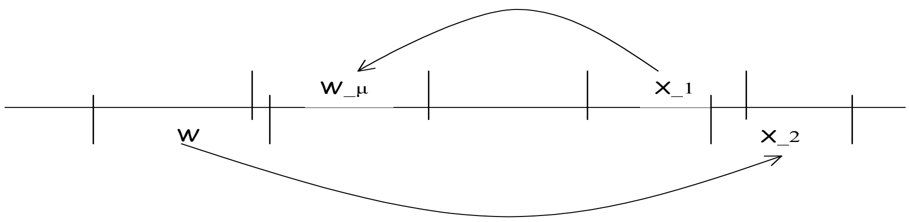

Figure 2: Relationship between $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{W}$ and $\mathcal{W}_\mu$.

where

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

and

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

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 previously 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 to 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 to 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.

Therefore, this formulation of the estimation procedure induces 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 might not 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

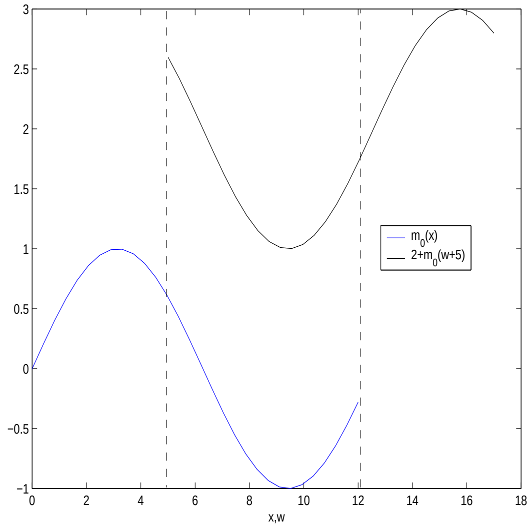

Figure 3: Intersection of observed support: $y = sin(0.5x)$, $z = 2 + sin\big(0.5(w + 5)\big)$, $\mathcal{X}_1 = [0, 12]$, $\mathcal{W} = [0, 12]$ and $\mathcal{X}_2 = [5, 17]$.

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 (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.

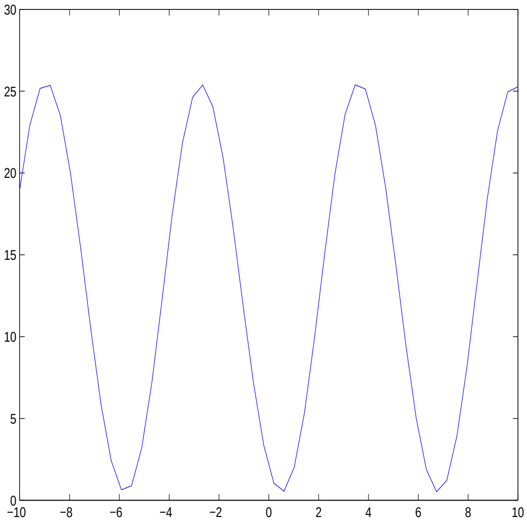

Figure 4: Multiple minima of the loss function: $y = sin(0.5x)$, $z = 5 + 0.5sin(0.5(x - c))$, $C = [-10, 10]$, $c_0 = 0.5$.

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 4 step estimator defined above 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 results for the two 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 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 due to 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 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.

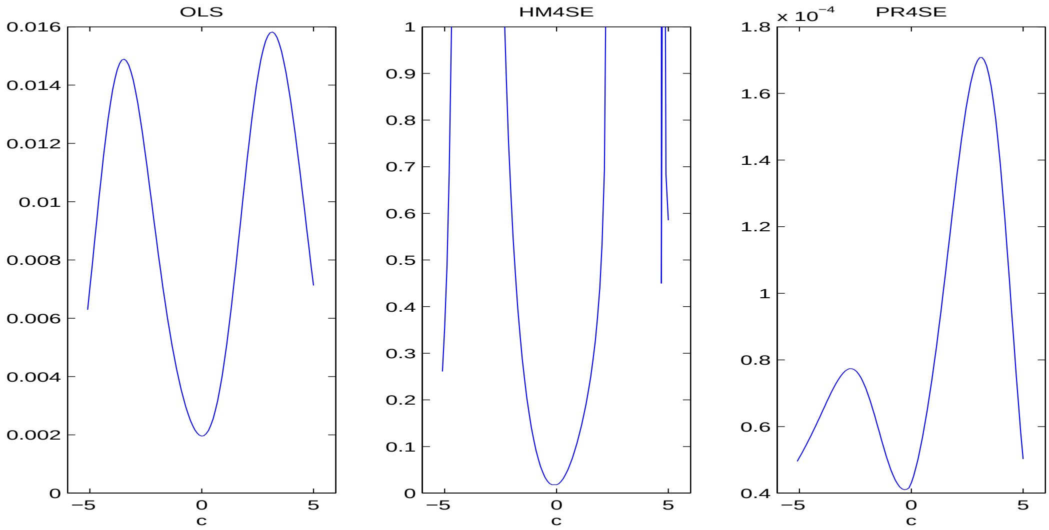

Figure 5: Mean conditional Loss functions $L(c|a, b)$ of the first Monte Carlo Series ($c_0 = 0$): a) parametric b)HM 4 step estimator c) PR 4 step estimator.

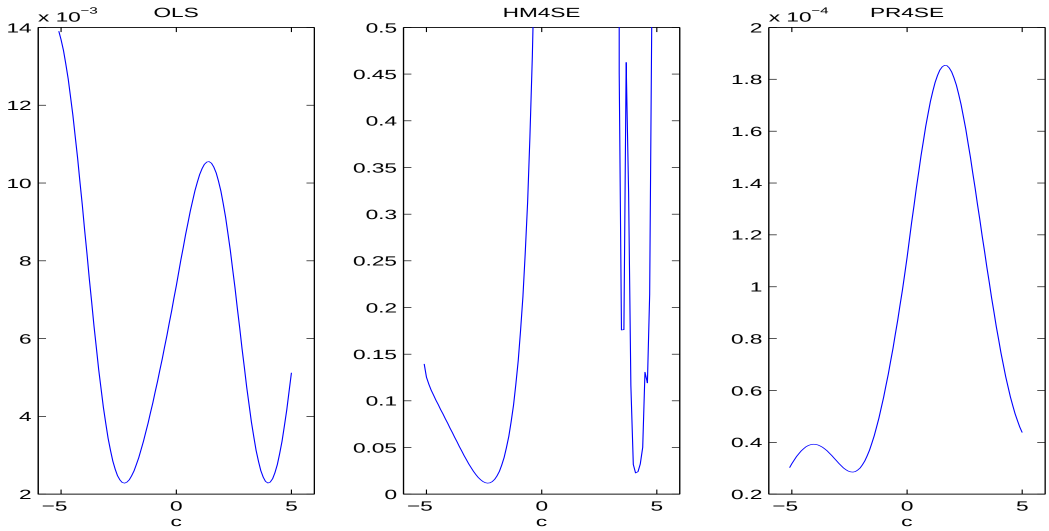

Figure 6: Mean conditional Loss functions $L(c|a, b)$ of the second Monte Carlo Series ($c_0 = 4$): a) parametric b)HM 4 step estimator c) PR 4 step estimator.

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 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.

The next paragraph discusses why the two estimators behave differently.

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 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

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

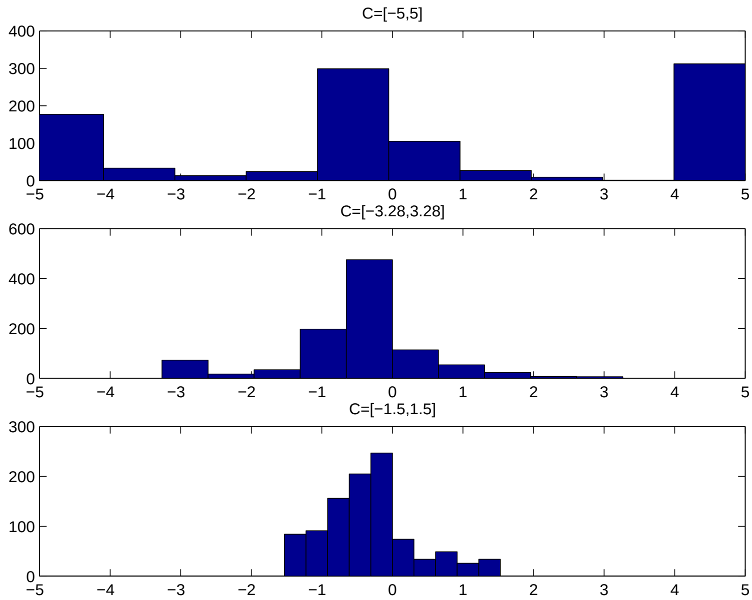

Figure 7: Three histograms for the distribution of $\hat{c}$ obtained with the Pinkse-Robinson 4 step estimator using different supports of c. First Monte Carlo series ($c_0 = 0$).

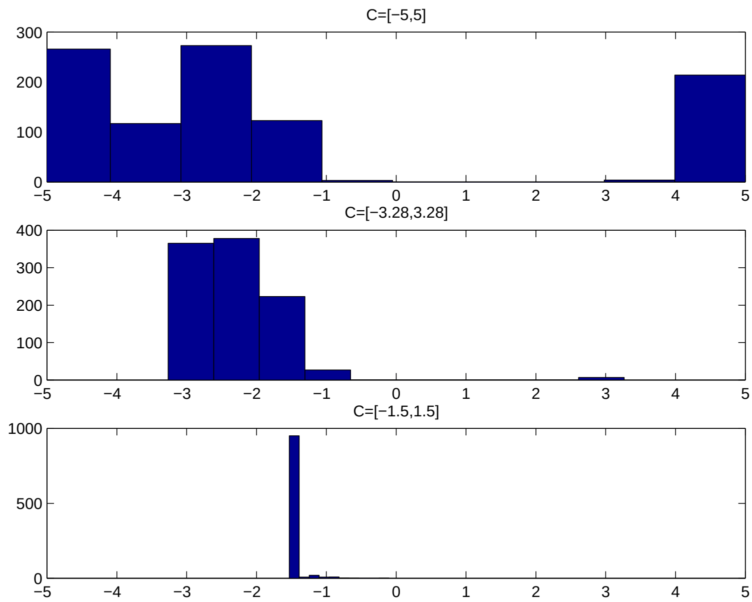

Figure 8: Three histograms for the distribution of $\hat{c}$ obtained with the PR 4 step estimator using different supports of c. Second Monte Carlo series ($c_0 = 4$).

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 particularly 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 between, and which estimator is preferable depends on the specific situation. In small samples the second point should clearly dominate the first, 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 $\mathcal{W} \cap \mathcal{W}_{c_0}$. 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).

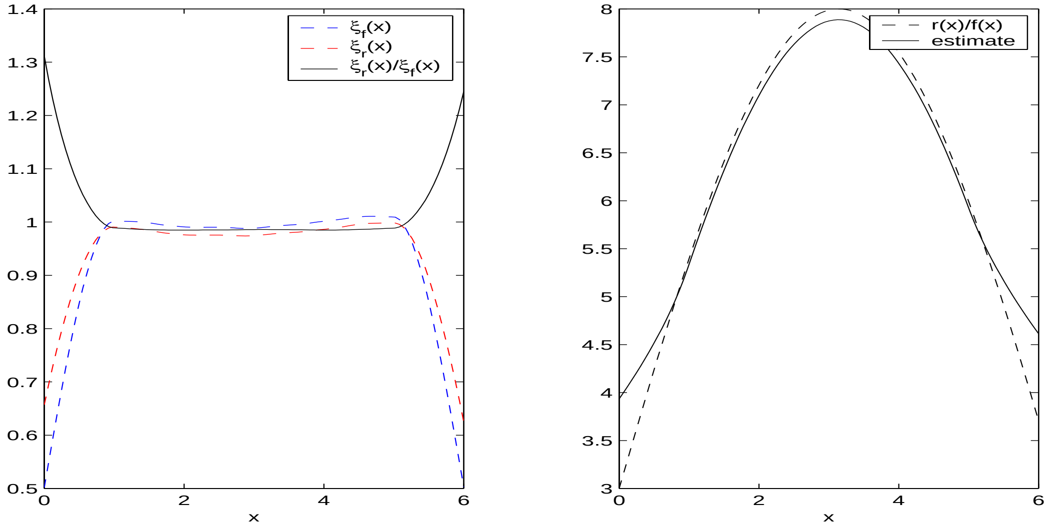

Figure 9: Proportionality of the bias: $N = 2000$, $y = 3 + 5sin(0.5x)$, $x \sim U[0, 6]$, $h = 0.5$, mean of 1000 samples

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 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 indeed has 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:

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

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

$$\hat{m}_1(x_t-c_0) \;=\; a_0+b_0m_0(x_t-c_0)+\epsilon_1(w_t^c,N) \tag{7}$$

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

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.*

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

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.*

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

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

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

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

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

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.*

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

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 will 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:

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

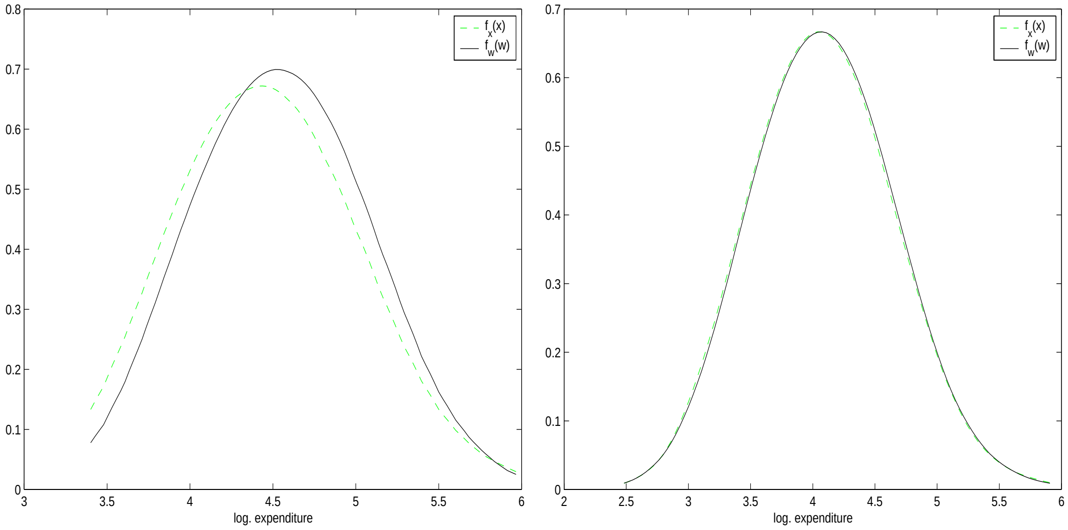

Figure 10: FES I: Kernel estimates of the marginal distribution of log expenditure a) FES I b) FES II

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 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 indicates 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 case 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). In this case the loss function (4) becomes:

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

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: $\hat{c}$ 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 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 with 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 here the local linear smoother instead of the Nadaraya-Watson estimator.

FES II is also 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 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.

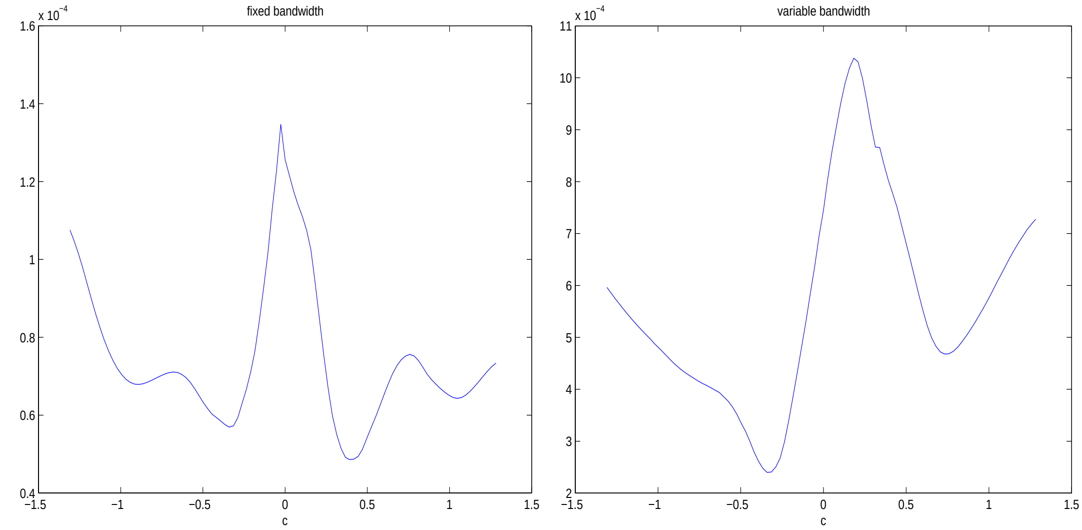

Figure 11: FES I: Loss function of EPLM

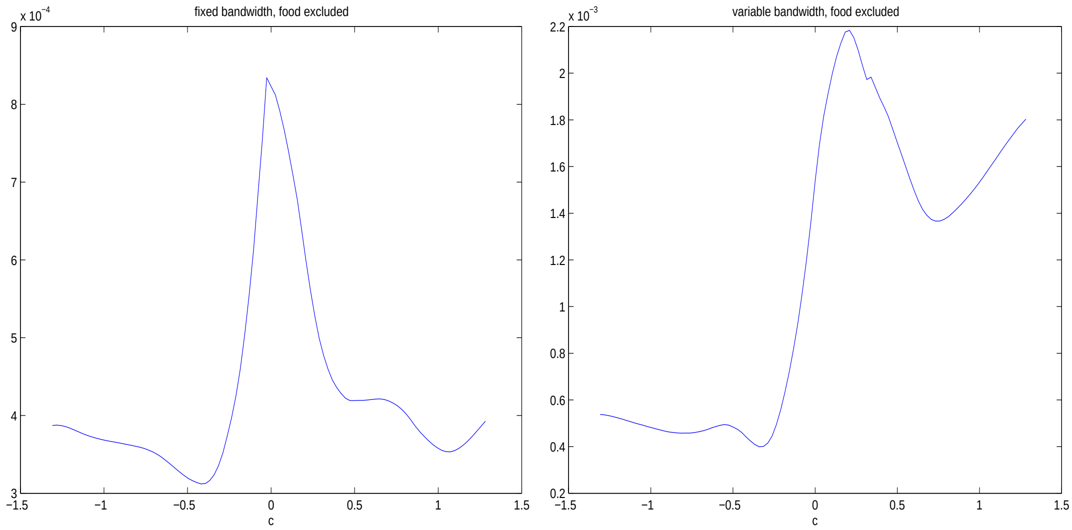

Figure 12: FES I: Loss function of EPLM, J=5 (food excluded).

	fixed bandwidth	variable bandwidth
expenditure shares	$\hat{a}^{(j)}$	
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

	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

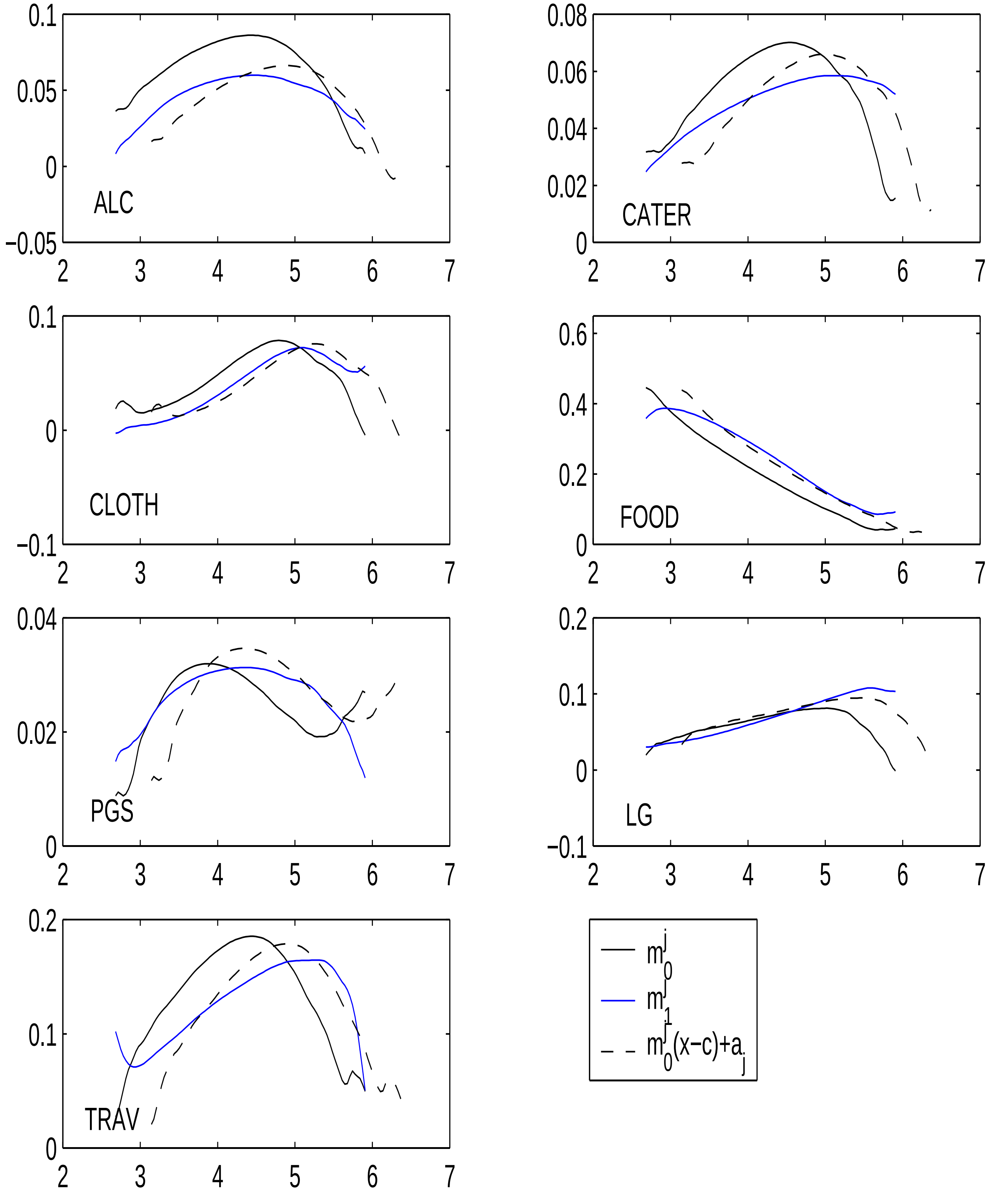

Figure 13: FES II, EPLM, fixed bandwidth

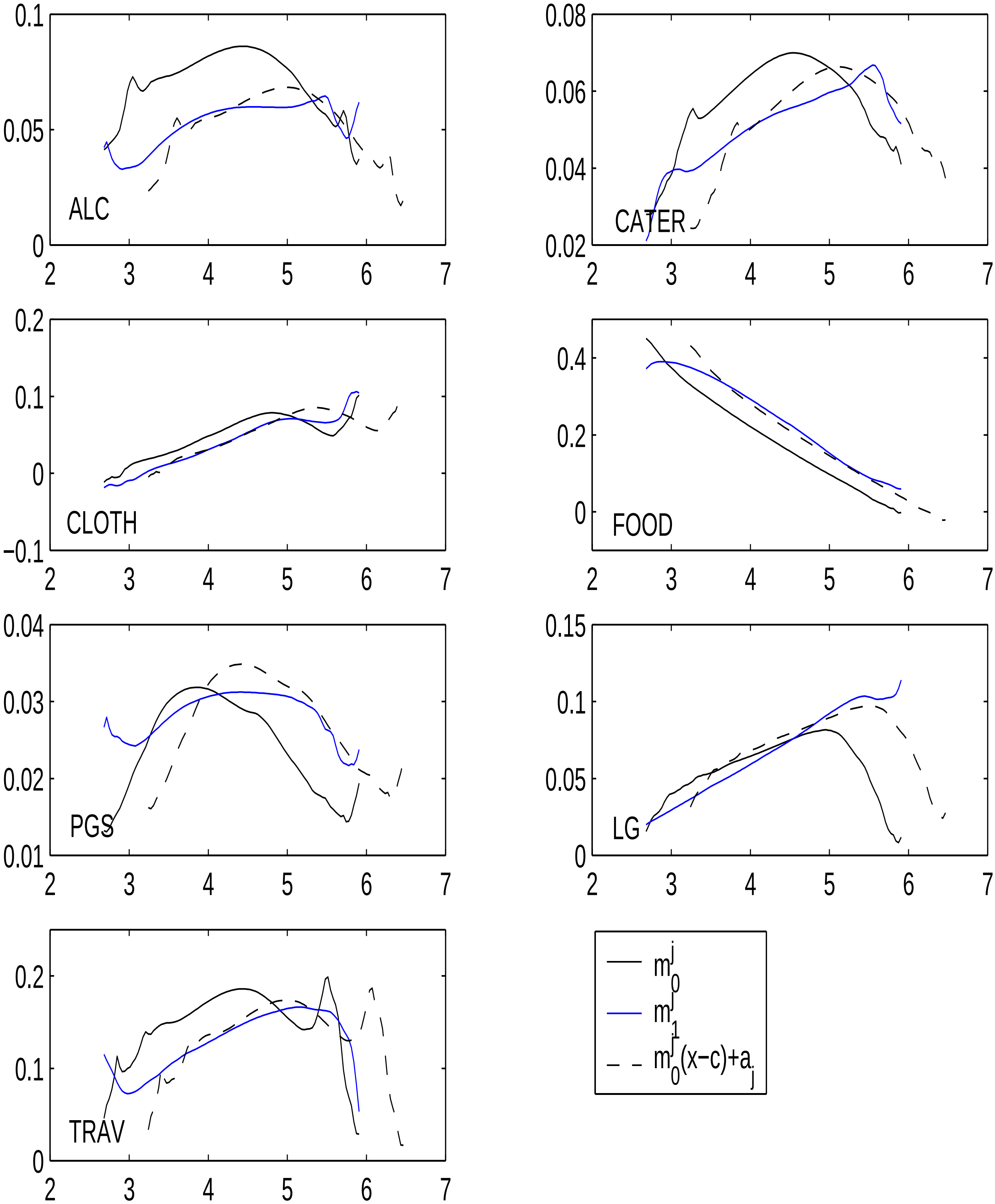

Figure 14: FES II, EPLM, variable bandwidth

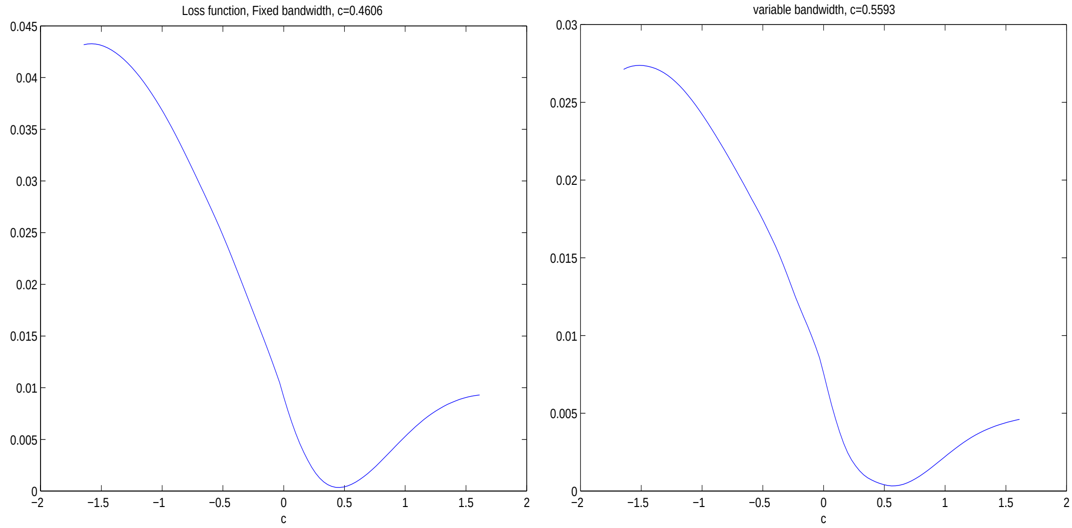

Figure 15: FES II, EPLM, Loss functions

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. 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

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

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 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:

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

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 probability limit of the loss function, i.e. $\text{plim}L_N(a,b,c)$ =$\text{plim}A_1$, has a unique minimum at a_0, b_0, c_0, i.e.

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

We have to check the necessary and the sufficient conditions.

The first order conditions are:

$$\partial_a \text{plim}A_1(a,b,c) = -\frac{2E\big[a_0 - a + b_0 m_0(x-c_0) - b m_0(x-c)|x-c \in \mathcal{W}\big]}{\big[F(\overline{x}(c)) - F(\underline{x}(c))\big]} = 0 \quad (9)$$

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

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

From (9) and (10) we obtain

$$\hat{a} = a_0 + E\big[b_0 m_0(x-c_0) - b m_0(x-c)|x-c \in \mathcal{W}\big] \quad (12)$$

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

Substituting for a yields:

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

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

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

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

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

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. ■

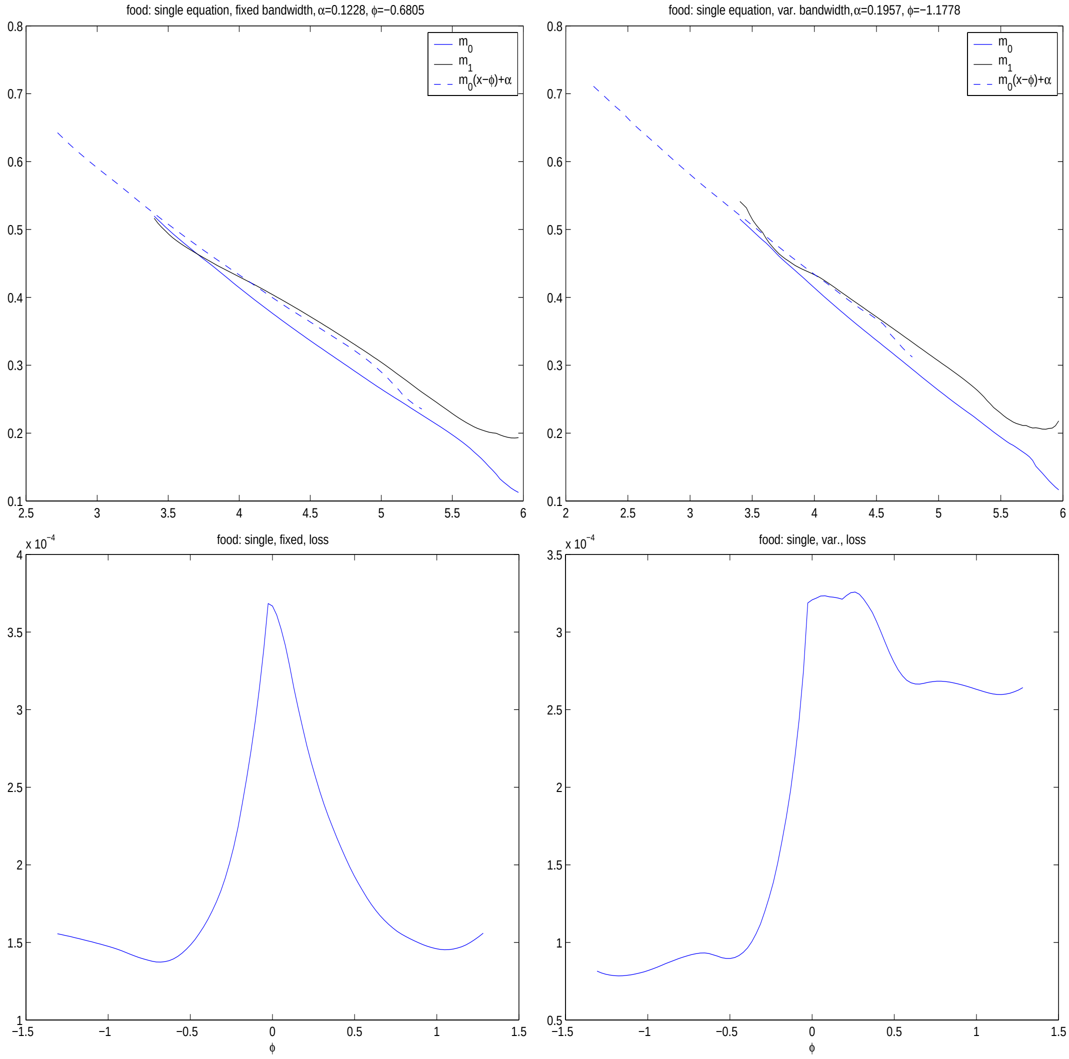

Figure 16: FES I: Food expenditure share.

Appendix 2: Estimation results

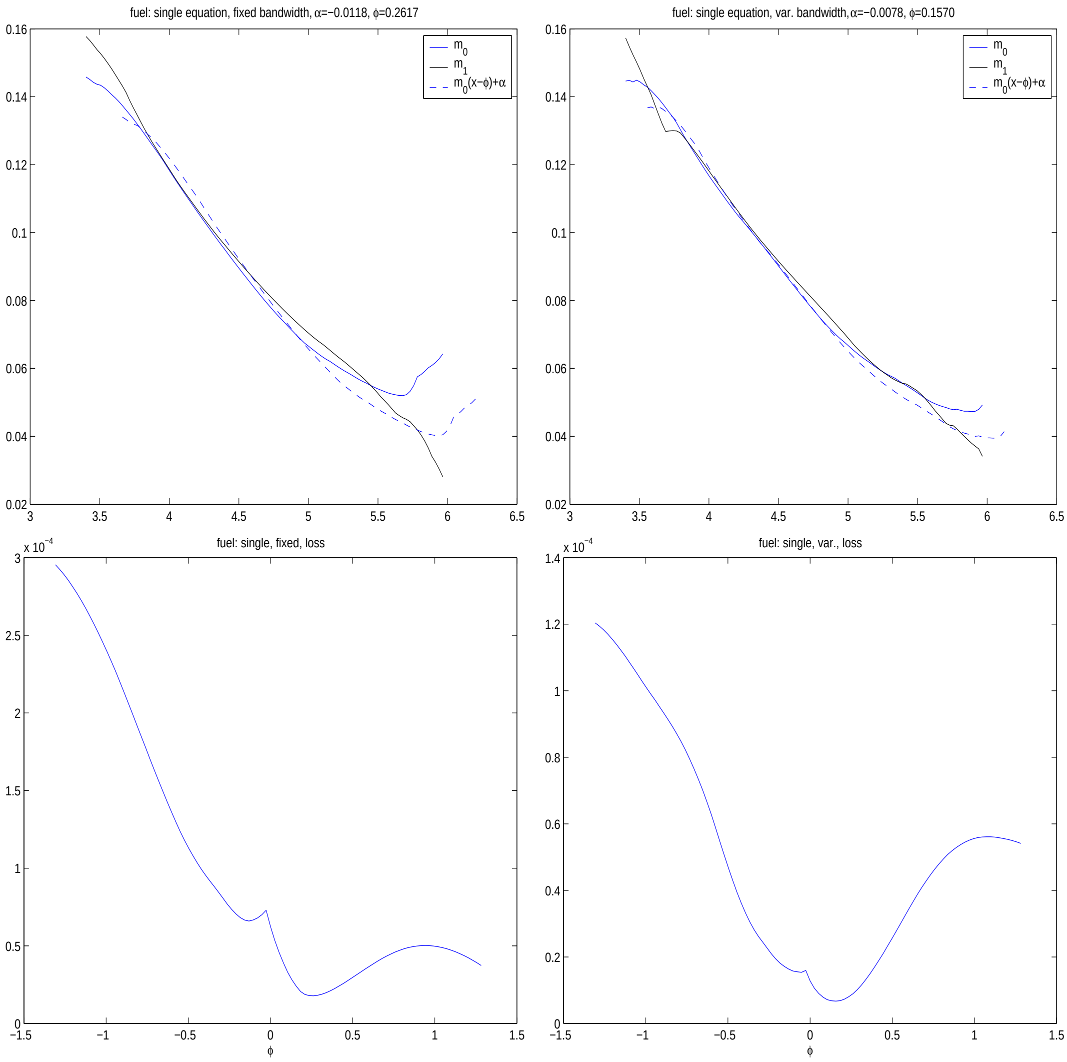

Figure 17: FES I: Fuel expenditure share.

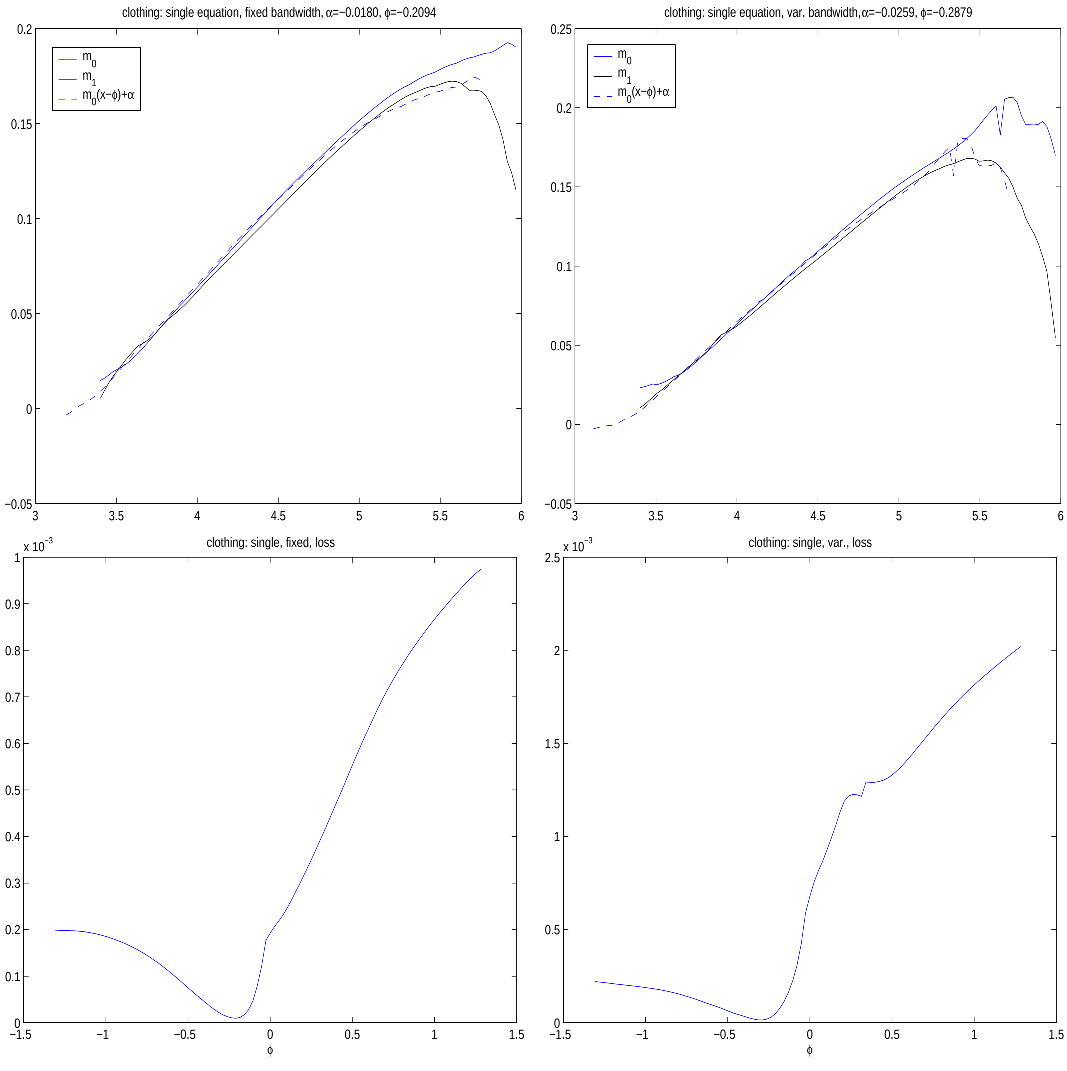

Figure 18: FES I: Clothing expenditure share.

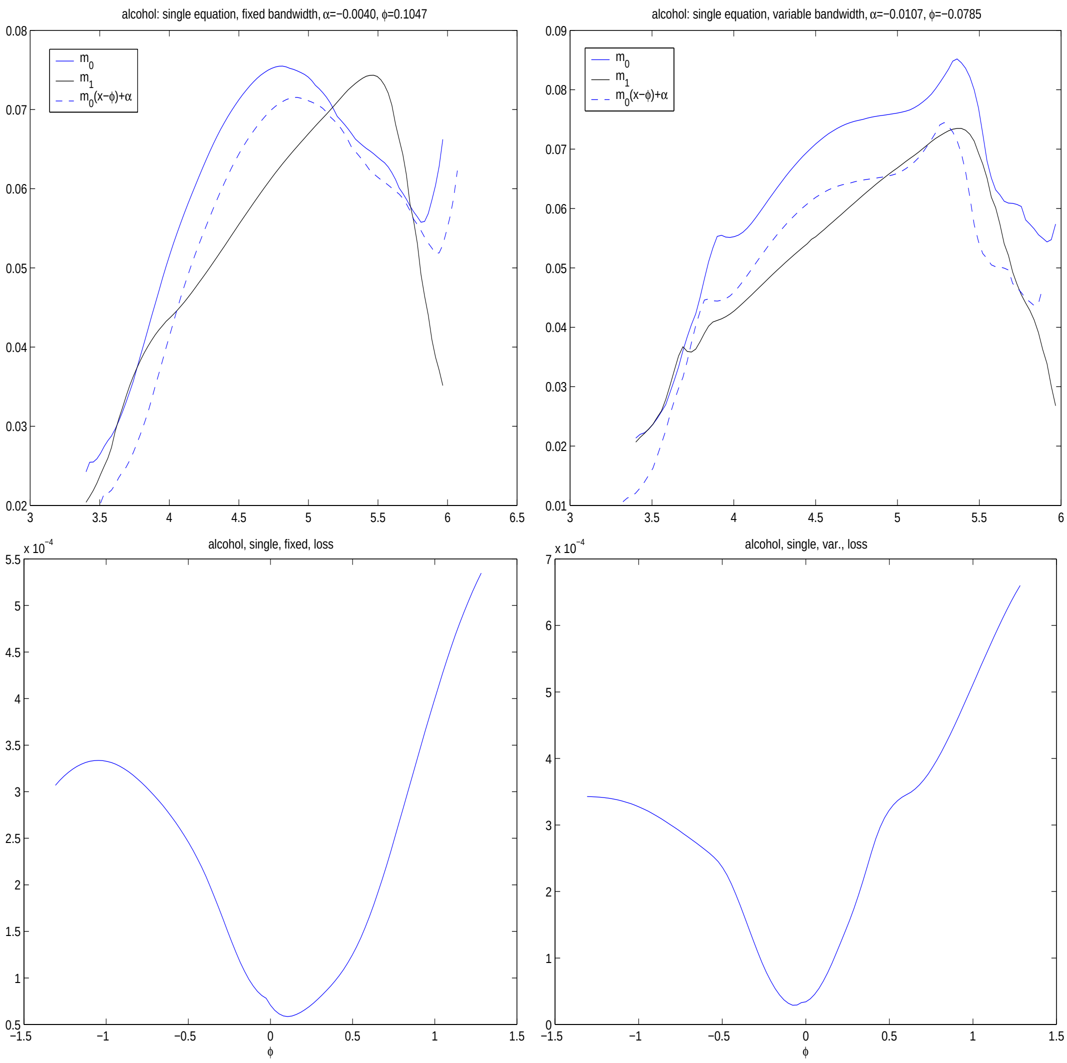

Figure 19: FES I: Alcohol expenditure share.

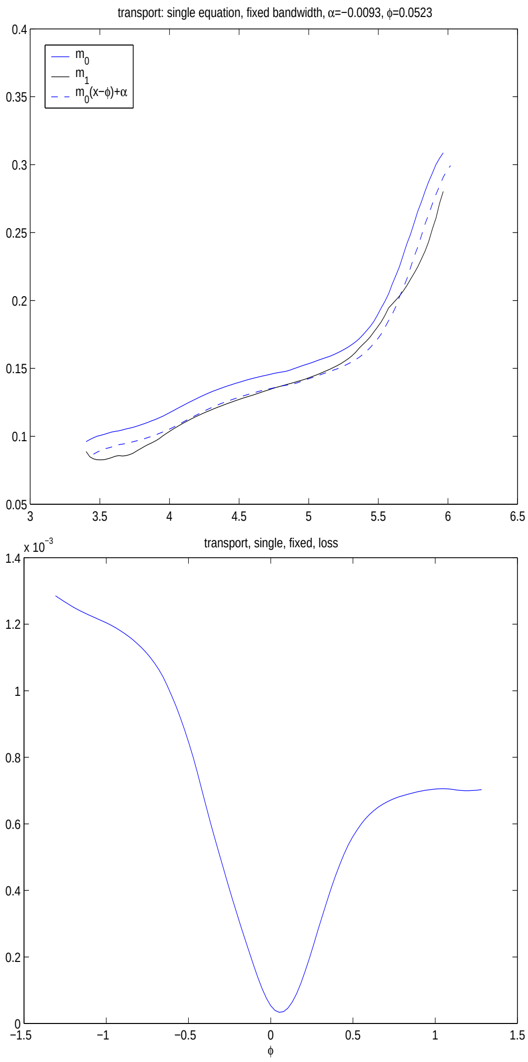

Figure 20: FES I: Transport expenditure share.

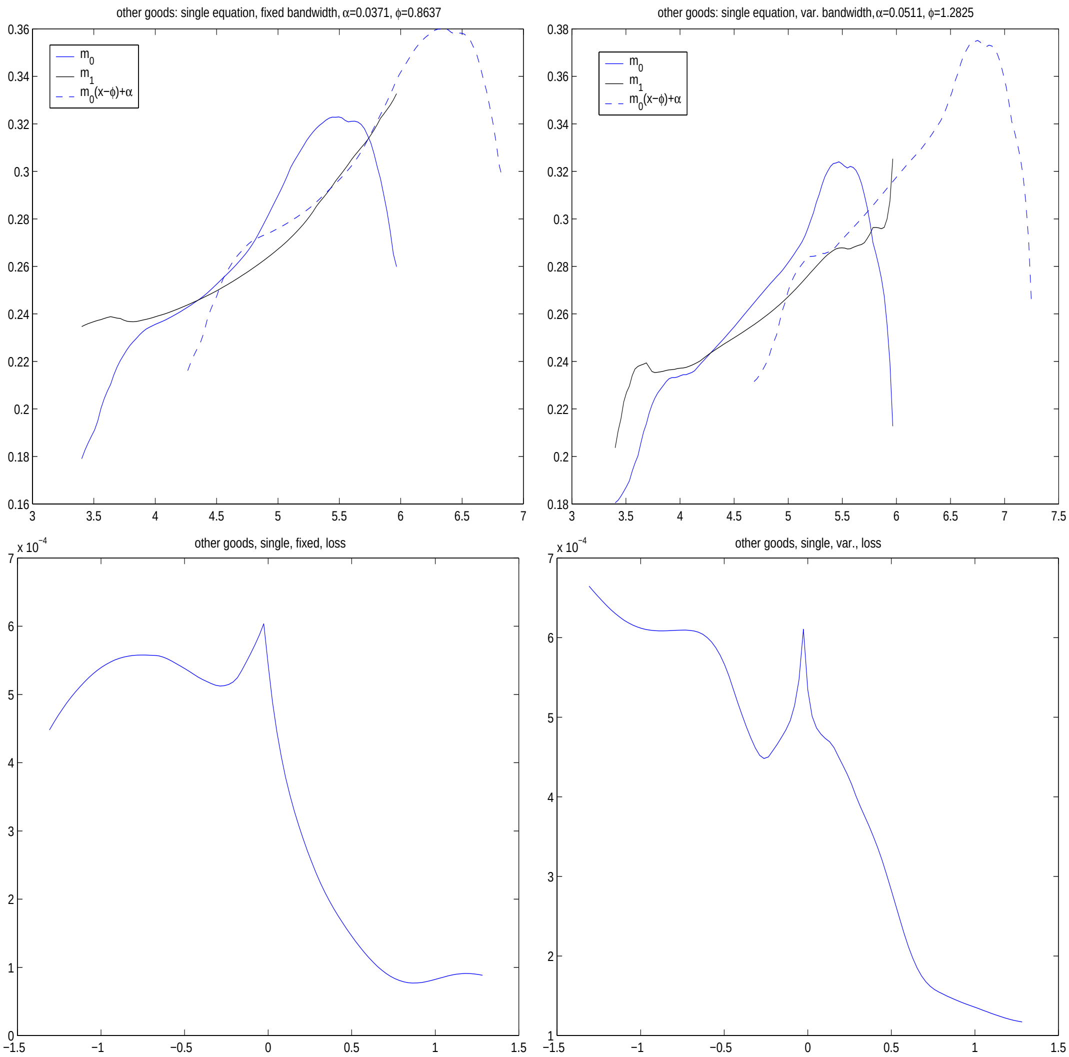

Figure 21: FES I: Other goods expenditure share.

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.

“Ich versichere, daß die vorstehende Dissertationsschrift von mir selbständig ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde, und daß ich alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, als solche gekennzeichnet habe.“