Nonparametric Demand Systems and a Heterogeneous Population

Stefan Hoderlein
Mannheim University

September 2003
- Very Preliminary -

Abstract

This paper is concerned with the econometric modelling of the demand behavior of a population with heterogeneous preferences under minimal assumptions. More speci…cally, we characterize the implications of the assumption that the Slutsky matrix is negative semide…nite and symmetric across a heterogeneous population without assuming anything on the functional form of individual preferences, and very little about their distribution. In the same spirit, implications of a linear budget set are being considered.
Solutions for several sources of endogeneity, like measurement error and endogeneous preference are considered. The consequences of functional form restrictions are also explored. First empirical results using new nonparametric regression techniques establish that the Weak Axiom holds across the population, while Utility maximization is somewhat less well accepted.

1 Introduction

Economic theory yields strong implications for the actual behavior of individuals. This is particularly true for demand theory, where a couple of well-known restrictions like Slutsky symmetry arise. All restrictions imposed by rationality on demand behavior are qualitative in nature, which means that they do not predict a speci…c functional relationship among a set of variables. To test the implications of rational behavior, by and large two strands of literature have emerged. The …rst uses revealed preference theory, is nonparametric in nature and concentrates on violations of the Strong Axiom in observable data. Key contributions are Afriat (1967) and Varian (1982). More recently, a similar approach has been suggested by Blundell, Browning and Crawford (2002).
The second strand of literature tests a couple of restrictions on demand behavior,

using fully speci…ed parametric demand systems. This literature dates back to at least the …fties (Stone (1954)), but has really peaked with the advent of fully ‡exible functional form demand systems. More recent examples are the Translog, Jorgenson et al. (1982), the AIDS, Deaton and Muellbauer (1980), Blundell et al (1993), or the "exact QUAIDS", Banks et al.(1997), see also Lewbel (1999) for a comprehensive survey. Obviously, both approaches have its limitations: The …rst usually leads to tests of low power, as price movements are dwarfed by movements in income, and concentrates on one speci…c property only. The second su¤ers from the limitations that demands take a certain functional form and that the introduction of preference heterogeneity has not been solved very successfully (see, e.g. Brown and Walker, 1989).

Our aim in this paper is to lay the foundations for nonparametric demand systems, ideally combining the advantages of both approaches: Being nonparametric in nature, i.e. not specifying any functional form, and still able to judge the restrictions imposed by rationality robustly as well as comprehensively. Additionally, we want to allow for unobserved heterogeneity in preferences. Furthermore, we will include the formation of preferences, an issue that has been rightfully emphasized recently, e.g., by McFadden (2001), or - particularly forcefully - by Manski (2000).

The structure of this paper will be as follows: in the next section we introduce the main concepts, and derive the …rst major theoretical result that speci…es under what conditions key elements of demand theory can be recovered from applied models, provided we have a heterogeneous population with completely general heterogeneity of unknown type. In particular, our interest centers on the key elements of individual rationality. For instance, we concentrate on the negative semide…niteness and symmetry of the Slutsky matrix in a heterogeneous population, and we give a new characterization of both in terms of observables. In the third section we consider modi…cations of the benchmark scenario of the second section: Restricting mildly the way in which parts of the unobserved heterogeneity enter, we show that we may then recover in particular Slutsky symmetry in a new fashion. Other important extensions of the basic model concern the use of additional information like exclusion restrictions or other sources of data. While the latter allows us to determine the in‡uence of parts of the unobserved heterogeneity, the former may be used to weaken some remaining restrictive assumptions. Finally, we give an overview of preliminary results, and close this paper with a brief summary.

2 The Demand Behavior of a Heterogeneous Population

As already mentioned in the introduction, our main aim in this paper is to model a population heterogeneous in preferences without assuming anything on the functional form of individual demands and still retain testable implications of Economic theory.

To this end, we start by introducing a framework for modelling a heterogeneous population.
Demand theory assumes that the demand of all individuals is the result of a well behaved utility maximization problem, yielding a demand function

$$w_i = \phi(p; y_i; u_i) \ , \tag{2.1}$$

where $w_i; p$ and y_i are budget shares, log prices and log total expenditure, vectors of length $L; L$ and $1;$ respectively. Furthermore, $u_i = u_i(\cdot)$ denotes the individual's utility function. Throughout, we restrict ourselves to continuously differentiable demand functions, which restricts preferences to be itself continuous, strictly convex and locally nonsatiated, with utility function everywhere twice differentiable. Also, the use of total expenditure instead of income is justified by the assumption of additive separability of the preferences over time, a strong assumptions which nevertheless underlies all of the applied demand literature(with rare exceptions, e.g. Browning (1991), Hoderlein (2002a)). This assumption allows to abstract from all issues pertaining to an uncertain future,and will be denoted by (Add).
The existence of the $\phi(\cdot)$ functional (from now on called theoretical microrelation) can be derived from the argmax operator, i.e. a rule that relates these variables. The theoretical properties of this functional are as follows: For fixed u_i, say u_0; $\phi(\cdot; \cdot; u_0)$ behaves like a standard rational demand function, which obeys the usual conditions of rational behavior, e.g. the compensated price derivatives form the negative semidefinite and symmetric Slutsky matrix.
In order to avoid technical difficulties arising with the differentiation on function spaces, we shall assume henceforth that u_i may be completely described by a finite fixed vector $v_i = (v_{1i}; \ldots; v_{Mi})$ of parameters[1]. Therefore we consider ϕ as a $[0; 1]^L$ valued function defined on $\mathbb{R}^L_+ \times \mathbb{R}_+ \times \mathbb{R}^M$; continuously differentiable in p and y: Also, for simplicity of exposition, we consider p to be a positive nonrandom vector. This is immaterial for our argumentation as the same arguments go through if prices depend on time series randomness alone, while other variables exhibit cross-section variation, see Hoderlein (2002a).
If we interpret each individual as a realization from an underlying population, we can give the equivalent formulation to (2.1) in terms of random variables. We assume that $(W_i; Y_i; V_i)$ and all other random variables to appear below, denoted as random vector by G_i; are iid with $(W_i; Y_i; V_i; G_i) \sim (W; Y; V; G)$, where the latter denote the population variables.

Assumption 2.1 Let all variables and functions be as defined above. Demand is then given by

$$W = \phi(p; Y; V) \tag{2.2}$$

[1] This does not mean that the concepts can not be defined more generally, see Hoderlein (2002a), who uses Frechet-derivatives (see Luenberger (1997)). Little is, however, gained in terms of Economic understanding.

As our aim is to establish the link between the theoretical microrelation and its empirical counterpart, we consider the conditional average. The conditioning here is on observables, where the set of observables obviously depends on the information at hand. In demand analysis this is cross-section data, in which case the conditioning σ-field must consist of all the current observables.
To capture the endogeneity in preferences and clarify the importance of observables and unobservables, we assume that every preference is endogenous in the following sense: it depends on the individuals' current observable and unobservable attributes, denoted as random vectors by Z and A respectively. Here, Z denotes all observable household attributes (like age, household size, etc.). The variable A in turn is meant to capture individual specific unobservables. These could in principle be time-varying as well as infinite dimensional, however, for simplicity of exposition we desist from this greater generality and consider only the case of a finite $(S \times 1)$ and time invariant vector[2]. This leads to the following

Assumption 2.2 Let all variables be as defined below. Then

$$V = \vartheta(Z; A); \tag{2.3}$$

where ϑ is a fixed Borel-measurable $\mathbb{R}^M$-valued function defined on the set $\mathcal{Z} \times \mathcal{A}$ of possible values of $(Z; A)$:

So far we have defined all main components of our framework. To state the next assumption, which ensures that interchanging differentiation and integration is well defined, as well as for statement of the proposition, we need the following notation: Let μ_G be the distribution of a random variable G, and denote by $\mu_{G|H}$ the conditional distribution of G given H:
Let $m(p; y; z) = E[W|Y = y; Z = z] = E[\varphi(p; Y; V)|Y = y; Z = z]$ denote the empirical regression function, and finally let $D_x f$ denote the derivative of a function f with respect to x; whose dimension will be obvious from the context. Moreover, whenever convenient we suppress the arguments of the respective functions.

Assumption 2.3: (Bounded Convergence) There exists a function g; such that

$$\left\| \begin{pmatrix} D_y\varphi(p; y; \vartheta(z; a)) \\ D_p\varphi(p; y; \vartheta(z; a)) \end{pmatrix} \right\| \leq g(a) \text{ ;with } \int g(a)\mu_A(da) < \infty ;$$

uniformly in $(p; y; z)$:[3]

Finally, we specify all dependence assumptions

[2] Both complications can be handled by the methods below.

[3] Among the primitive economic conditions that ensure that this assumption holds are: strict convexitiy, local nonsatiation and continuity of the preferences generated by ϑ, a linear budget constraint and $p >> 0$:

Assumption 2.4: $\mu_{A|Y,Z} = \mu_{A|Z}$:

Basically, this assumption states that - conditional on Z - income and unobserved heterogeneity are distributed independently. This is obviously a strong assumption, needed in this strength due to the generality of the other assumptions. Under what conditions this may be relaxed is one of topic of section 3.
Given these assumptions and notations, we are in the position to state the following propositions on the relation of theoretical and empirical quantities, where we focus on the following questions:

1. How are the empirically obtained derivatives $(D_y m; D_p m)$ with respect to prices and income related to the theoretical ones $(D_y \varphi; D_p \varphi)$?

2. How and under what kind of assumptions do elements of observable behavior allow inference on key elements of Economic theory. Especially, what does observable behavior tell us about homogeneity, adding up as well as negative semidefiniteness and symmetry of the Slutsky-matrix

$$S(p;y;v) = D_p \varphi(p;y;v) + D_y \varphi(p;y;v)\varphi(p;y;v)' + \varphi(p;y;v)\varphi(p;y;v)' + \mathrm{diag}\,\{\varphi(p;y;v)\}$$

These concepts are commonly known as "rationality" in this scenario[4], and shall be subject of Proposition 2.2. We start with Proposition 2.1 which establishes the relationship between the derivatives:

Proposition 2.1

Let all the variables and functions be as defined above. Let (Add) and (A2:1) - (A2:3) be true. Then follows that (i)

$$D_p m(p;Y;Z) = E[D_p \varphi(p;Y;V)|Y;Z]$$

If in addition (A2.4) holds, we have (ii)

$$D_y m(p;Y;Z) = E[D_y \varphi(p;Y;V)|Y;Z]$$

Moreover, if V is Z-measurable; then (iii)

$$D_y m(p;Y;Z) = D_y \varphi(p;Y;V) \text{and } D_p m(p;Y;Z) = D_p \varphi(p;Y;V).$$

Proof: Appendix.

Parts (i) and (ii) of this proposition state that each individual's empirically obtained marginal effect is the best approximation (in the sense of minimizing distance with respect to L_2-norm) to the individual's theoretical marginal effect. For price derivatives, this holds under virtually no conditions at all, for income derivatives we have

[4] We adopt this language. For other definitions of rationality, see Chiappori and Rochet (1987).

to invoke the additional assumption A2:4, because the individually varying income effects are not to be confounded with the individually varying preference heterogeneity. In this general scenario, this is as close as current observables allow us to get to the true marginal effects[5].

Usually, the empirical coefficients will still be an average across individuals with the same realization of Z, and the preference-induced heterogeneity will still be bigger than the observed heterogeneity. However, the second part of the proposition gives a condition on the information needed for both to coincide: all individual randomness that affects demand must be fully captured by current observables.

Regarding the average across a population or a subgroup, the following corollary holds:

Corollary 2.2

Let all the variables and functions be as defined above. Let (Add) and (A2:1) - (A2:4) be true. Then follows $E[D_y m(p;Y;Z)|F] = E[D_y \phi(p;Y;V)|F]$ $\forall j$; and for any $F \subseteq \sigma\{Y;Z\}$: In particular $E[D_y m(p;Y;Z)] = E[D_y \phi(p;Y;V)]$ $\forall j$: A similar condition holds for D_p under (Add), (A2:1) - (A2:3):

Proof: Appendix.

Thus, the average of the empirical marginal effects over the whole population or over a subgroup coincides almost surely with the true average marginal effect across population or subpopulation.

Another trivial corollary concerns the standard practise of inferring something about elasticities from the observed regression function. Again, we need some notations:

Let $V[G;H|O]$ denote the conditional covariance (matrix) between G and H conditional on O and $V[H|O]$ be the conditional (co-)variance (matrix) of H: In both cases the dimensionality should become clear from the context. Moreover, let π_{ij} denote the i-th price elasticity of good j, let η_j denote the income elasticity of good j, let $m_{log}(p;y;z) = E[\log(W)|Y = y;Z = z]$ and δ_{ij} be Kronecker's delta

Corollary 2.3

For the price and income elasticities, the following holds: $E[\pi_{ij}|Y;Z] = D_{p_i} m_{log;j} - \delta_{ij}$ and $E\left[\eta_j|Y;Z\right] = D_y m_{log;j} + 1$; where $m_{log;j}$ is the j-th element of m_{log}. In particular, unless the condition

$$V\left[D_y \phi_j; \frac{1}{\phi_j}|Y;Z\right] - \frac{E\left[D_y \phi_j|Y;Z\right]}{E\left[\phi_j|Y;Z\right]} E\left[\frac{E\left[\phi_j|Y;Z\right]}{\phi_j} - 1|Y;Z\right] = 0$$

holds, $E\left[\eta_j|Y;Z\right] \neq \frac{D_y m_{(j)}}{m_{(j)}} + 1$; where $m_{(j)}$ is the j-th element of m:

[5] Note that A2:4 could be relaxed to a local independence condition $D_y \mu_{A|Y;Z}(a;y;z) = 0$; $(y;z) \in [y_0;y_1] \times [z_0;z_1]$ for fixed y_0, y_1; z_0, z_1; if we were just interested in the marginal effects of a subgroup of the population.

Proof: Appendix.

It is instructive to note that the elasticities have to be calculated from the log budget share regressions (which is only possible provided $W > 0$). In particular, $E\left[\varepsilon_j | T, Z\right]$ equals $\frac{D_y m_{(j)}}{m_{(j)}} + 1$ only if the aforementioned condition is fulfilled[6], for which there is no a priori reason.

We turn now to the question which economic properties carry through to the observable spaces. This problem bears some similarities with the literature on aggregation over agents in demand theory, because taking conditional expectations can be seen as an aggregation step, as long as the measurability condition of $P2.1$ (iii) is not met. With the new notation, $m_2(p, y, z) = E[WW' | Y = y, Z = z]$ and $\operatorname{diag}\{m\}$ denoting the matrix having the m_j, $j = 1, \dots, L$ on the diagonal and zero off the diagonal, we are in the position to state the following

Proposition 2.4

Let all the variables and functions be as defined above, and (Add), (A2.1) - (A2.3) be true.

(i) If ψ fulfills $\iota'\psi = 1$ (a.s.) $\Rightarrow$ $\iota' m = 1$ (a.s.).

Let additionally (A2.4) hold as well. Then follows that

(ii) If ψ fulfills $\psi(p + \theta, Y + \theta, V) = \psi(p, Y, V)$ (a.s.) $\Rightarrow$ $D_p m\iota + D_y m = 0$ and $m(p + \theta, Y + \theta, Z) = m(p, Y, Z)$ (a.s.).

(iii) If $\overline{S}$ is negative semidefinite (nsd) (a.s.) $\Rightarrow$ $\overline{D_p m} + D_y m_2 + 2(m_2 - \operatorname{diag}\{m\})$ is nsd (a.s.), where $\overline{D_p m} = D_p m + D_p m'$.

(iv) If S and $V[D_y\psi, \psi' | Y, Z]$ are symmetric (a.s.) $\Rightarrow$ $D_p m + D_y m m'$ is symmetric (a.s.).

(v) Let V be Z measurable $\Leftrightarrow$ $\{S$ is symmetric and nsd iff $D_p m + D_y m m' + m_2 - \operatorname{diag}(m)$ is symmetric and nsd$\}$.

Moreover, if V is Z measurable, the converse holds in (i) and (ii) as well.

Proof: Appendix.

The importance of this proposition lies in the fact that it allows testing the key elements of rationality without having to specify the functional form of the individual

[6] In this condition, the second term measures to a certain extent the degree of nonlinearity present in ψ_j. If this were zero then (*) would reduce to $V\left[D_y\psi_j, \frac{1}{\psi_j} | Y, Z\right] = 0$.

demand function. Suppose we see any of these conditions rejected in the observable (generally nonparametric) regression at a position y, z, p. Recalling the interpretation of the conditional expectation as average (over a “neighborbood”) this proposition tells us that there exists a set of positive measure of the population (“some individuals in this neighborhood”) which does not conform with the postulates of rationality. This is the case regardless of how rich our information about heterogeneity is: If our information set is poorly, and we are nevertheless able to identify a local average for which one of the conditions is violated, then it must be a fortiori violated if our information set increases.

If we believe the information to be complete - see case (v) - then we may directly identify these individuals, for then they are completely characterized by their observables. Moreover, the reverse implication is perhaps even more significant. Statements linking the observed model $D_p m + D_y m m'$ to individual behavior[7], namely the S, are only true if V is Z measurable, i.e. if all individual heterogeneity has been captured by observables. This is a fortiori true for the parametric literature. Appending “an additive error capturing unobserved heterogeneity” and proceeding as usual is not a solution either. Note that we may always append an additive error, since $m = \phi + (m - \phi) = \phi + \varepsilon$. The crux is now that the error is generally a function of y and p, as was already noted by Brown and Walker (1989). For instance, the nonsymmetric part of the Slutsky matrix becomes

$$S = D_p m + D_y m m' + D_p \varepsilon + (D_y m)\,\varepsilon' + (D_y \varepsilon)\, m' + (D_y \varepsilon)\,\varepsilon' ,$$

and the last four terms will not vanish under general specification of ϕ. But even if we restrict the way unobserved heterogeneity enters, as is done in the third section, there will be an averaging interpretation. More importantly, as shown below, new correction terms and expressions arise. Thus, the standard practise must be understood as assuming that there be no unobserved preference heterogeneity.

Returning to Proposition 2.2., one should note a key difference between negative semidefiniteness and symmetry. For the former we may provide an “if” characterization without any assumptions other than the basic (see (iii)). To obtain something equivalent for symmetry, we have to invoke the additional assumption about the conditional covariance matrix. This matrix is unobservable - at least without any further assumptions. Note that this assumption is (implicitly) implied in all of the literature, since only then we can unambiguously check for symmetry using $D_p m + D_y m m'$, which is the standard practice.

Note further some parallels with the aggregation literature in economic theory: Only adding up and homogeneity carry immediately through to the conditional average. This result is similar in spirit to the Mantel-Sonnenschein theorem, where only these two properties are inherited by aggregate demand. Furthermore, it is also well known in this literature that the aggregation of negative semidefiniteness (usually shown for

[7] For instance: “All individuals display a negative semidefinite Slutsky matrix, as is evident from the empirical results”.

the Weak Axiom) is more straightforward than that of symmetry. Finally, a matrix similar to $V[D_y Á; Á^0]$ has been used in this literature (as "increasing dispersion", see Jerison (1984)).

Lewbel (1990, but especially 2001, Theorem 1) and Brown and Walker (1989) give results in a similar spirit. While Brown and Walker concentrate on the consequences for the error structure, this approach is more closely related to Lewbel's. There are, however, some key di¤erences: The result linking negative semide…niteness to observables, i.e. P 2:2 (iii) is new. Additionally, Lewbel characterizes symmetry through $V[D_y Á; Á^0 jY; Z] = 0$, which is of course more restrictive as our result (iv). Finally, the approximation and conditional averaging interpretation of the non-measurable case is new.

As a last consequence we obtain a characterization of the functional forms of the regression. In particular, Blundell et al. (2002), establish that regressions additive in income and preference parameter, di¤erentiable in both variables, must have income entering (log-)linearly. The same results is likely to carry through to the observable regression. Too see this, suppose that there is "mixed" term of the form $´(Y; V)$. But for the observable regression to be additive, we must have that $E[´(Y; V)jY; Z] = 0$; a strong assumption. Thus, as long as unobserved preference heterogeneity is conditionally independent of income, which was a necessary assumption for identi…cation, additive observable regressions must have been caused by additive models in the unobservable world. This restricts the use of additive models severely. An alternative model that retains theory consistency and is econometrically tractable, is the extended additive model of Hoderlein (2002a), and Christopeit and Hoderlein (2002) This model is used in the application below.

Thus far we have established that the most commonly used assumptions may be weakened dramatically, without loosing the ability to test the key elements of rationality. However, we still had to invoke some assumptions, out of which the assumed conditional independence of preference heterogeneity and "income", as well as the covariance assumption in (iv) are arguably the most troubling. Given the generality of our model (2.1) the strength of these requirements comes as no surprise. We now turn to the question in which way we may weaken them.

3 Endogeneity

In this section we show how the framework introduced may be extended to tackle some of the most common sources of endogeneity. It is a reoccurring theme in this paper that emphasis is given to structural modeling, i.e. explicitly taking into account the various sources of endogeneity. First we shall focus on the implication of measurement error in the income variable, and we establish that large parts of our statements may be preserved, even in the presence of measurement error.

As was already emphasized in the second section, violations of the conditional inde-

pendence of unobservables assumption may constitute an important source of endogeneity. Endogeneity related to the formation of preference is a good example. We show how this issue can be modeled and that additional information may be used to tackle the endogeneity coming from this particular source of unobserved heterogeneity, yielding yet another correction term.
But we also give a general treatment of endogeneity in this framework, related to nonparametric IV. Finally, we show how the implications of functional form restrictions may be used. We want to emphasize that the order in which these issues are being treated does not imply anything about their importance.

3.1 Measurement Error

Measurement errors are often cited as a cause for unsatisfactory empirical results. The advantage of the projection-based approach is that the measurement error may be treated as another element of the projection, implying that some of the properties may hold even in the polluted data, or may at least be found after correction. Recall our baseline model

$$W = Á(p; Y; V).$$

It is often assumed that instead of the random scalar Y we only observe a mismeasured random scalar X, where $X = Y + Q$: Here Q is another scalar random variable, assumed to be independent of Y with mean zero and ...nite variance. As above, instead of $D_y Á$ we may observe it's closest approximation $E[D_y Áj X; Z]$; now of course with the mismeasured variable in the conditioning set. To obtain this quantity is now the goal, as is it the closest approximation (in the sense of minimizing a L_2 distance) given our information, and some or all of the economic properties may have testable implications.
More speci...cally, we focus on the relationship between $E[D_y ÁjX; Z]$ and $D_x E[ÁjX; Z]$, which is of course the derivative of the nonparametric regression of W on $X; p$ and Z. We discuss this is in the baseline scenario. To this end, we introduce the following assumption

Assumption 3.1.5: assume that, conditional on $Z; Y$ has a absolutely continuous distribution with density $f_{Y jZ}(y; z)$, such that
(i) $\lim_{y \,! \,\S\, 1} f_{Y jZ}(y; z) = 0 \; 8z$
(ii) $Y jZ = z \;\mathbf{v}\; N(¹(z); ¾^2(z))$

We shall also make use of the following notation: $M(p; x; z) = E[Á(p; Y; V)jX = x; Z = z]$:

The following proposition is a consequence:

Proposition 3.1.1 Let all the variables and functions be as de...ned above, and let (Add) and (A2:1)-(A2:4) and (A3:1:5) (i) be true

(i) Then follows that

$$D_x E[\varphi(p;Y;V)|X=x;Z=z] = E[D_y\varphi(p;Y;V)|X=x;Z=z] + \kappa;$$

where

$$\kappa = E\left[W\left[D_x \log f_{X|Z}(X;Z) - D_y \log f_{Y|Z}(Y;Z)\right]|X=x;Z=z\right]$$

(ii) If in addition (A3:1:5) (ii) holds, then

$$D_x E[\varphi(p;Y;V)|X=x;Z=z] = (1-\psi(x;z))E[D_y\varphi(p;Y;V)|X=x;Z=z] + \eta;$$

where

$$\psi(x;z) = \frac{V[Y|X=x;Z=z]}{V[Y|Z=z]};$$

and η contains higher order terms in y. Suﬃcient for $\eta = 0$ is

$$\int V\left[\frac{D_y^2\varphi(p;y_r;V)}{2}(Y-y_0)^2;Y|X=x;Z=z\right]\mu_{Y;V|X;Z}(dy_0;dv;x;z) = 0;$$

with $y_r = \lambda y_0 + (1-\lambda)Y$:

Proof: Appendix.

Remark 3.1: Although it appears to be suﬃcient for $\kappa = 0$ that $D_x \log f_{X|Z}(x;z) = D_y \log f_{Y|Z}(y;z)$ $\forall y;x;z$; this condition is completely implausible. To see this, take any fixed z (so that we may skip the dependence on z) and note that $D_x \log f_X(x) = D_y \log f_Y(y)$ can only be fulfilled if neither side depends on their respective argument, i.e. $D_y \log f_Y(y) = D_x \log f_X(x) = h$: Thus, $f_X(x) = \exp[hx] + c$; where h and c are constants. Since x is log income, which ranges from $+\infty$ to $-\infty$, this is a violation of assumption A3:1:5 (i); save for the case when $h = 0$; i.e. X is uniformly distributed, which is empirically rejected. Since the correction expression is hard to simplify further without any additional assumption, we invoke the much more plausible A3:1:5 (ii): It states that "true income" has a lognormal distribution. Since it is known that the unconditional distribution of X is approximately lognormal, this may have been caused by a proportionate measurement error on a true underlying lognormal Y. Additionally, note that

$$\int V\left[\frac{D_y^2\varphi(p;y_r;V)}{2}(Y-y_0)^2;Y|X=x;Z=z\right]\mu_{Y;V|X;Z}(dy_0;dv;x;z) = 0;$$

is fulfilled if $D_y^2\varphi = 0$: This would be the case with most of the commonly used functional forms, in particular the almost ideal type.

Remark 3.2: Note that

$$\begin{aligned} V[Y|Z=z] &= E[V[Y|X=x;Z=z]|Z=z] + V[E[Y|X=x;Z=z]|Z=z] \\ &\geq E[V[Y|X=x;Z=z]|Z=z]; \end{aligned}$$

so that $0 < E[\psi(X;Z)|Z=z] \le 1$; and we have **attenuation** (on average over the income range).

The implications for the economic properties, in particular for Slutsky negative semidefiniteness, are summarized in the following

Proposition 3.1.2 Let all the variables and functions be as defined above, and let (Add) and (A2:1)-(A2:4) be true

(i) Then follows that adding up of φ is inherited by M.

If in addition (A3:1:5) holds, then

(ii) Homogeneity of φ implies that $D_pM\iota + D_xM < 0$ if $D_xM > 0$; that $D_pM\iota + D_xM > 0$ if $D_xM < 0$ and that $D_pM\iota = 0$ if $D_xM = 0$:

(iii) Sufficient for
S is nsd (a.s.) $\Rightarrow \overline{D_pM} + D_xM_2 + 2(M_2 - \mathrm{diag}(M))$ is nsd,
is that $E[D_y[\varphi\varphi']|X;Z]$ is positive semidefinite

(iv) If S is symmetric (a.s.), $D_xM = 0$ and $V[D_y\varphi\varphi'|X=x;Z=z]$ is symmetric $\Rightarrow D_pM$ is symmetric (a.s.).

Proof: Appendix.

Remark 3.3: Note how unevenly the measurement error diminishes the strength of the testable implications: Some implications remain largely unaltered: Besides the trivial adding up restriction it is in particular negative semidefiniteness that proves robust. In particular, $E[D_y[\varphi\varphi']|X;Z]$ pds has to be assumed. This is of course implied by $D_y[\varphi\varphi']$ psd, a property of, e.g., homothetic preferences, but the aggregation literature has given some other examples that lead to an average (in our case: conditional average) income effect matrix that is psd, for instance increasing dispersion (Jerison (1984)). In remarkable contrast, the already weak implications of symmetry are now confined to the extreme case of a purely homothetic average ($D_xM = 0$). Also, homogeneity is weakened to a sign property.

3.2 Preference Formation

It is a common to assume stable preferences during the process of decision making. However, it is also widely acknowledged that these "stable" preferences - besides incorporating truly idiosyncratic elements - have also been formed by the social environment. A prime candidate for such an environment would of course be the upbringing in a family, but an individual's preferences might also be influenced by becoming parent, by the colleagues at work, etc. To capture the endogeneity in preferences and clarify the importance of observables and unobservables, we assume that every preference is endogenous in the following sense: it depends on the individuals' past observable and unobservable attributes, denoted as random vectors by $Z^{¡}$ and A respectively. $Z^{¡}$ re‡ects the dependence of preferences on the past "social environment" where preferences have been shaped[8]. It is debatable whether this set might also contain past Economic choice variables, for then forward-looking individuals could in‡uence future preferences by current decisions. This is a question of myopia. To give Economic theory some predictive power we shall exclude this possibility, so that $Z^{¡}$ contains only past attributes and no choice variables, but we shall pick up this point when discussing exclusion restrictions and instrumental variables.

Assumption 3.2.2 Let all variables be as de…ned below. Then

$$V = \#(Z^{¡}; A);$$

where # is a …xed Borel-measurable R^M-valued function de…ned on the set $Z^{¡} £ A$ of possible values of $(Z^{¡}; A)$:

As a matter of fact, technically we may allow for Z being an argument of # as well, so that $V = \#(Z; A)$ would be a nested case. However, our focus is really on the preference formation in the past, so we retain the notation (2.3). The next assumption is about the nature of the stochastic process generating Z

Assumption 3.2.3

$$Z = h(Z^{¡} + U)$$

where h is one to one and onto, and $\bar{h} = h \pm g$; $g(x; y) = x + y$; is a …xed, Borel-measurable R^G-valued function de…ned on the set $Z^{¡} £ U$ of possible values of $(Z^{¡}; U)$:

This assumption clari…es how past and present are linked. The leading proponent would be a (Markovian) VAR, i.e. $Z_t = B(Z_{t¡1} + U_t) = BZ_{t¡1} + C_t$; where t denotes time, and B is a nonrandom matrix, but A3:1:3 allows for more general structures as well[9]. This assumption can be relaxed if we have additional information on the stochastic process generating Z; see section 3.3. below.

Assumption 3.2.4: (Dominated Convergence) There exists a function g; such that

$$\left| \left(\frac{D_y Á(p; y; \#(h^{¡1}(z) ¡ u; a))}{D_y Á(p; y; \#(h^{¡1}(z) ¡ u; a))} \right) \right| · g(a; u) \text{ ; with } \int g(a; u) ¹_{AU}(da; du) < 1;$$

[8] "Past" in this sense may well include the immediate past. Moreover, allowing current attributes to in‡uence demand actually simpli…es the analysis.

[9] For instance, the more distant past (t ¡ ¿, say) may a¤ect current Z: Furthermore, it could be generalized to $Z = h(g(Z^{¡}) + g(U))$. The role of the additivity assumption inside the h function is to insure that the Jacobian determinant in a change-of-variable formula is unity:

uniformly in $(p; y; z)$: Here, h^{i1} is the inverse function of h. A similar condition holds for D_p:[10]

Thus far we have just modi…ed the above assumption to …t our new scenario. To prevent this from being an exercise in pure modelling aesthetics, we have to introduce some additional elements that may allow us to make use of the richer model structure. It will come from the following observation:
In many countries, large panel-data sets have become available recently. These contain a lot of information about the time-series evolution of the distribution of many variables of interest to our discussion, in particular on the joint distribution of $Y; Z; Z^{i}$ and U, but usually lack information on the demand. This is, as in our benchmark model of the second section, still contained in the cross section only. The question becomes how to nevertheless pro…t from this additional source of information?
To see that this observation allows some loosening of assumption, consider the modi…ed dependence assumptions

Assumption 3.2.5: The de…ned distributions obey the following restrictions:
(i) $¹_{AjY;Z;U} = ¹_{AjZ;U}$,
(ii) assume that, conditional on Y and $Z; U$ has a absolutely continuous distribution with density $f_{UjY;Z}$: Moreover U has bounded support, with $f_{UjY;Z}$ bounded away from 0 on the entire support, and that $f_{UjY;Z}$ is di¤erentiable with respect to y for any $y; z; u$. $¹_{UjY;Z} = ¹_{UjZ}$:

Remark 3.4: Note that in A3:2:5 there are now two sources of potentially unobserved heterogeneity. As in the benchmark scenario, the component A cannot be recovered. Therefore we have to invoke an assumption of similar type as above. However,we will be able to capture the in‡uence of the other source of unobserved heterogeneity from the panel data. Thus, there is no need for a (potentially to strong) independence assumption.

As in the general scenario of the second section, but now after a change of variables discussed in the appendix, we have

$$m(p; y; z) = \int_U \int_A Á(p; y; \#(h^{i1}(z) ¡ u; a)) ¹_{AjY;Z;U}(da; u; y; z) ¹_{UjY;Z}(du; y; z)$$

Taking the derivative with respect to y; using A3:3:5 ,and rearranging yields

$$\begin{aligned} n(p; y; z) &= \int_U \int_A D_y Á(p; y; \#(h^{i1}(z) ¡ u; a)) ¹_{AjZ;U}(da; u; z) ¹_{UjY;Z}(du; y; z) + \\ &\quad \int_U \int_A Á(p; y; \#(h^{i1}(z) ¡ u; a)) D_y f_{UjY;Z}(u; y; z) ¹_{AjZ;U}(da; u; z) du \end{aligned}$$

[10] Among the primitive economic conditions that ensure that this assumption holds are: strict convexitiy, local nonsatiation and continuity of the preferences generated by #, a linear budget constraint and $p >> 0$:

As above the first rhs term is $E[D_y\phi(p,X,V)|Y,Z]$, while the second becomes

$$\int_U\int_A \phi(p,y,\vartheta(h^{-1}(z)-u,a))\frac{D_y f_{U|Y,Z}(u,y,z)}{f_{U|Y,Z}(u,y,z)}\mu_{A|Z,U}(da,u,z)\mu_{U|Y,Z}(du,y,z),$$

which is $E\left[W D_y \log\left(f_{U|Y,Z}(U,Y,Z)\right)|Y=y,Z=z\right]$ and shall be denoted as $Cor(p,y,z)$. Since y and z were chosen arbitrarily, we can summarize this argument in the following lemma, which extends P2.1 to this scenario:

Lemma 3.2.1 Let all the variables and functions be as defined above, and let (Add) and (A2.1),(A3.2.2)-(A3.2.5) be true. Then follows that $E[D_y\phi(p,X,V)|Y,Z] = D_y m - Cor$ and $E[D_p\phi(p,X,V)|Y,Z] = D_p m$. Moreover, if $\vartheta(h^{-1}(Z)-U,A)$ is Z,U-measurable, then follows that $D_y\phi = D_y m - Cor$ (a.s.) and $D_p\phi = D_p m$ (a.s.).

Proof: Given in text.

Remark 3.5: This shows that we may use distributional information obtained from panel data to circumvent the conditional independence assumption and obtain the derivatives. Of course, in panels it is also possible to estimate an individual specific fixed effect, say α. This α may reflect the influence of one or more elements of A. The same argument as made above for U can be applied to this α. As we saw in the benchmark, we may allow for a higher dimensional A, and it is very likely that the fixed effect will not capture all individual-specific effects contained in A.

The economic properties, in particular Slutsky negative semidefiniteness, are summarized in the following

Proposition 3.2.2 Let all the variables and functions be as defined above, and let (Add) and (A2.1),(A3.2.2)-(A3.2.4) be true

(i) Then follows that adding up of ϕ is inherited by m.

If in addition (A3.2.5) holds, then

(ii) Homogeneity of ϕ implies that $D_p m\iota + D_y m - Cor = 0$.

(iii) If S is nsd (a.s.), $\overline{D_p m} + D_y m_2 - Cor2 + 2(m_2 - diag(m))$ is nsd.

(iv) If S is symmetric (a.s.) and additionally $V[E[D_y\phi|Y,Z,U], E[\phi'|Y,Z,U]|Y,Z]$ is symmetric $\Rightarrow$ $D_p m + D_y m m' - Cor m'$ is symmetric (a.s.).

(v) Finally, if $\vartheta(h^{-1}(Z)-U,A)$ is Z,U-measurable,
$\Rightarrow$ $S = D_p m + D_y m m' - Cor m' + m_2 - diag(m)$ (a.s.).

Proof: Appendix.

Remark 3.6: These results illustrate that knowing parts of the distribution of U may allow for significant progress and how this progress takes place. Not only can we relax or abolish the conditional independence assumptions. The conditions on $V[D_y\phi, \phi'|Y, Z]$ for obtaining symmetry can also be relaxed significantly. From totally ignoring it in the most general model, to getting parts of it through specifying the way A enters in a general fashion, and to obtaining it completely when also restricting U to exert only first order effects.

3.3 Exclusion Restrictions and Instrumental Variables

A general way of treating endogeneity is of course given by the instrumental variables (IV) paradigm. In our framework we propose a solution using the key assumption of the control function (CFIV) approach. CFIV can be seen as one possible generalization of linear IV to our more general setting, and has been formalized by Newey, Powell and Vella (1999). The setting is straightforward: The problem comes from a possible violation of A2:5, i.e. $\mu_{A|Y,Z} \neq \mu_{A|Y}$. Assume there exists a random variable G (the "instruments") with the following properties: Let $X = Y - E[Y|G]$; that is X are the residuals from a projection of Y on the G-space. Then, $\mu_{A|X,Y,Z} = \mu_{A|X,Z}$ holds (from now on called Assumption 3.3.5). It is a consequence that all statements of the second section remain true, with the augmented sigma algebra, $\sigma\{X, Y, Z\}$ in place of $\sigma\{Y, Z\}$.[11]

3.4 Specifying the Functional Form

Specifying functional forms for ϕ is hazardous as we may exclude a lot of possible preference specifications. Perhaps the most sensible way is to use the following obvious consequence of the argumentation in the first section

$$\begin{aligned} W &= E[W|Y, Z] + (W - E[W|Y, Z]) \\ &= m(p, Y, Z) + \varepsilon(p, Y, Z, A), \end{aligned}$$

with $E[\varepsilon(p, Y, Z, A)|Y, Z] = 0$. Note further that the error will not be homoscedastic under general assumptions, as was shown by Brown and Walker (1989), and the covariance matrix is singular, due to the budget identity. We adopt the common practise of deleting one equation. The remaining equations (again L, for simplicity) have a regular covariance matrix.
A sufficient, but not necessary, condition for $E[\varepsilon(p, Y, Z, A)|Y, Z] = 0$ to hold is:

Assumption 3.4.1: Let $\phi = m + \varepsilon$ as defined above and assume that there exist a

[11] It is interesting to note that actually what is only needed is $E\left[W D_y \log f_{A|X,Y,Z}|Y, X, Z\right] = 0$, i.e. the conditional orthogonality of two functions of A. However, without distributional assumptions, little gain can be made from this observation.

$R^L \times R^K$ valued function Σ and a R^K valued function λ defined on $Y \times Z \times U$ and A respectively, s.th.

$$\varepsilon(p;Y;Z;A) = \Sigma(p;Y;Z)\lambda(A);$$

if it is combined with $E[\lambda(A)|Y;Z] = 0$; which taken together, yields a complete specification of the model. However, it turns out useful to consider a more general model where $E[\lambda(A)|Y;Z] = 0$ is relaxed. Consider the following assumption:

Assumption 3.4.5: A2.5 (ii) holds. Instead of A2.5 (i), assume that for λ defined above we have

$$\text{(i) } E[\lambda(A)|Y;Z] = \zeta(Z) \text{ and (ii) } V[\lambda(A)|Y;Z] = \Lambda(Z);$$

where ζ and Λ are a K vector and a $K \times K$ matrix valued function respectively.

Having allowed for the conditional first moments of λ to depend on all variables but Y, we may think of $A3{:}4{:}1$ in a different way. Of course, it encompasses the conditionally mean independent case. But it can also be seen as a K-th order Taylor expansion in an one-dimensional single index, say b, or as a linear expansion in a K-dimensional vector λ: Note the relaxation in the dependence compared to above, as now all functions of A and Y may be correlated, save for those given in $A3{:}4{:}5$. Moreover, as illustrated in Proposition 3.4.1 below, only $A3{:}4{:}5$ (i) is needed to obtain the best projection of the marginal effects. In contrast, for nsd and symmetry of the Slutsky matrix we will need $A3{:}4{:}5$ (ii) as well. Analogously to the previous sections, symmetry is the property most difficult to obtain, and while we may remove the assumption that $V[D_y\phi;\phi'|Y;Z]$ is symmetric in $P2{:}2$; we shall need some identification assumptions. They take the following form

Assumption 3.4.6 There exists a $K \times K$ matrix valued function P such that

$$\text{(i) } V[\lambda(A)|Y = y; Z = z] = P(z)P(z)';$$
$$\text{(ii) } \bar{\Sigma}(y;z) = \Sigma(y;z)P(z) \text{ is symmetric for all } (y;z):$$

The first part is merely a restatement of the second part of $A3{:}4{:}5$ as such a decomposition of the covariance matrix exists naturally. In contrast, (ii) is a strong assumption, but necessary because we have to solve a system of quadratic equations, which has between 0 and 2^K solutions in general. It may be Consider now the residuals of the regression, namely $\varphi = W - E[W|Y;Z]$ and $\Gamma = E[\varphi\varphi'|Y;Z]$. With this notation, part (ii) may be relaxed to: there exists a unique decomposition of Γ such that $\Gamma = \bar{\Sigma}\bar{\Sigma}'$. Since in applications it is necessary to choose a certain decomposition (see below), we choose the stronger version. Note that in this case $K < L$ is not possible since then the covariance matrix would be singular. Thus, it is necessary to have $K \geq L$:

Lemma 3.4.1 Let (Add),$(A3{:}4{:}1)$; $(A2{:}2)$-$(A2{:}4)$,$(A3{:}4{:}5)$ be true.

Let $\Gamma(y;z) = E[\varphi\varphi'|Y = y; Z = z]$; where $\varphi = W - E[W|Y = y; Z = z]$. Then follows from A3.4.6 (i) that $\Gamma(y;z) = \Sigma^{\frac{1}{2}}(y;z)\Sigma^{\frac{1}{2}}(y;z)'$; and from (i) and (ii) that $\Sigma^{\frac{1}{2}}(y;z) = \Gamma(y;z)^{\frac{1}{2}}$. Moreover, $D_y\Sigma^{\frac{1}{2}}(y;z) = D_y\left(\Gamma(y;z)^{\frac{1}{2}}\right)$.

Proof:
Take $(y;z)$ fixed, but arbitrary.
Since $\Gamma(y;z) = E[\varphi\varphi'|Y = y; Z = z] = V[\varphi|Y = y; Z = z]$, it follows that $\Gamma(y;z)$ is pds and therefore there exists a unique decomposition $\Gamma(y;z) = R(y;z)^2$; where R is a square, symmetric matrix (actually, a matrix valued function at a fixed position).
Moreover, $\Gamma(y;z) = \Sigma(y;z)V[\lambda(A)|Y = y; Z = z]\Sigma(y;z)' = \Sigma^{\frac{1}{2}}(y;z)\Sigma^{\frac{1}{2}}(y;z)'$;
due to A3:4:6. By the uniqueness of R follows that $\Sigma^{\frac{1}{2}}(y;z) = R(y;z) = \Gamma(y;z)^{\frac{1}{2}}$; and taking derivatives completes the proof.

This lemma shows the strength of the requirements needed to obtain $\Sigma^{\frac{1}{2}}(y;z)$ from the conditional variance of the error term. Under A3:4:6 (i) we can only identify $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}\prime}$; and thus $D_y\left(\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}\prime}\right) = \left(D_y\Sigma^{\frac{1}{2}}\right)\Sigma^{\frac{1}{2}\prime} + \Sigma^{\frac{1}{2}}\left(D_y\Sigma^{\frac{1}{2}}\right)'$. However, in order to say something about the symmetry of the Slutsky-matrix we must be able to say something about $(D_y\Sigma)*\Sigma'$; which is impossible without invoking A3:6 (ii),(iii).
For the following proposition, let $\phi = m + \Sigma\lambda$; $n = E[W|Y; Z]$ and $n_2 = E[WW'|Y; Z]$.

Proposition 3.4.2

Let all the variables and functions be as defined above, and let (Add) and (A3:4:1),(A2:2)-(A2:4); (A3:4:5)(i) be true.

(i) The results of Proposition 2.1. continue to hold.

(ii) If ϕ fulfills additionally $\iota'\phi = Y$ (a:s:) $\Rightarrow$ $\iota'n = Y$ (a:s:):

(iii) If ϕ fulfills $\phi(\lambda p; \lambda Y; V) = \phi(p; Y; V)$ (a:s:) $\Rightarrow$ $n(\lambda p; \lambda Y; Z) = n(p; Y; Z)$ (a:s:):

(iv) If S is nsd (a.s.) and (A3:4:5) (ii) holds additionally
$\Rightarrow$ $\overline{D_p n} + D_y n_2$ is nsd (a.s.), where $\overline{D_p n} = D_p n + D_p n'$:

(v) Let S be symmetric (a.s.).
Additionally, assume that (A3:4:5) (ii) and (A3:4:6) hold,
$\Rightarrow$ $D_p n + D_y nn' + \left(D_y\Gamma^{\frac{1}{2}}\right)\Gamma^{\frac{1}{2}}$ is symmetric (a.s.).

Proof: Appendix.

Remark 3.7: 1. In this scenario, the difference between symmetry and semidefiniteness, i.e. between utility maximization and the weak axiom, is obvious. In particular, we have to restrict the covariances in the symmetry case, while for nsd we do not have to invoke something similar as we can still use the (symmetric) second moment regression. For the same reason, we do not have to invoke the "identification of $\overline{\Sigma}$" assumption A3.6.

2. In comparison with P2.2 (iv), note that we can relax $V[D_y\phi, \phi'|Y, Z]$ symmetric now, reducing possible sources of bias in this model.

3. The remaining bit may be recovered from the covariance matrix.

By similar reasoning and with similar results, one may further extend the model defined by A3.1 to include second order effects. Suppose that $K = L$ and that the model were given by $\phi = m + \Sigma\lambda + \Psi\lambda_2$, where λ_2 is the $\frac{L(L+1)}{2}$ vector of squared elements of λ, i.e. $\lambda_1^2, \lambda_1\lambda_2, \ldots$ and Ψ is a $L \times \frac{L(L+1)}{2}$ matrix containing all second order derivatives (in a Taylor-expansion). Additionally, restrict all conditional moments up to fourth order not to be functions of y. Thus, we may introduce more generality in the functional form of the theoretical microrelation at the expense of restricting the conditional distribution of unobservables further, arriving again at full generality and full independence in the limit. We do not elaborate on this point further. Instead, we look for alternative ways to increase the overall information available in the system.

4 Preliminary Empirical Results

In this section we state preliminary results for the general framework of section 2 only.

4.1 The Econometric Model

As is obvious from the discussion above, there is no such thing as a single model for conducting the whole analysis. However, nonparametric regressions of various quantities, in particular of $E[W_i|Y_i, Z_i]$ and $E[W_i W_i'|Y_i, Z_i]$ play a key role. Of course, one can assume that all variables are jointly normal, so as to arrive at a linear model. But why should we restrict ourselves from the outset? Therefore it seems natural to apply regression methods more general than linear OLS. The leading proponent is the well-known nonparametric regression model

$$W_i = m(G_i) + \varepsilon_i, \quad i = 1, 2, \ldots, \tag{4.1}$$

which models the dependence of the budget shares W_i on a $d+1$-dimensional random vector $G_i = (Y_i, Z_i', p')'$. The error term ε_i is assumed to be independent of G_i, with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2$, and m is the mean regression function. For our purposes, however, this model is infeasible due to the curse of dimensionality, i.e. the fact that the precision of any estimator decreases exponentially with d. However, the most popular alternative, namely additive models are at odds with economic theory, as we saw above.

As our interest centers on a particular set of variables, others, often household observables like age of household head, are of less importance. The econometric model we propose extends the additive model, but is consistent with theory and allows exactly to model the impact of this set of particular variables in more detail. The model is given by

$$W_{il} = k_l(Y_i; p) + l_l(Z_i) + g_l(Y_i; p)' \lambda(Z_i) + \varepsilon_i; \quad i = 1, 2, \ldots; \quad l = 1, \ldots, L \qquad (4.2)$$

where $\lambda : \mathbb{R}^d \to \mathbb{R}^S$ is a known vector valued function; k_l; l_l; and $g_l' = (g_{1;l}, \ldots, g_{S;l})$ with $g_{s;l}(\cdot)$; $s = 1, \ldots, S$; are smooth, but otherwise unrestricted unknown functions. Furthermore, subscript l denotes the demand for the l-th good. Details of an estimator for this model based on local quasi-differencing can be found in Hoderlein (2002a) and Christopeit and Hoderlein (2002). Here it suffices to say that the estimator is optimal by any criterion and easy to implement. In particular, the choice of bandwidth, a parameter that governs the complexity of the model, can be done as suggested in the second reference, largely analogous to local polynomial modelling.

4.2 The Data

We start by giving a brief overview of the data, of the methods of data clearance and of the definitions of variables involved, and discuss the already mentioned issues of the estimation process.

4.2.1 The Data: FES

Every year, the FES reports the income, expenditures, demographic composition and other characteristics of about 7,000 households. The sample surveyed represents about 0.05% of all households in the United Kingdom. The information is collected partly by interview and partly by records. Records are kept by each household member, and include an itemized list of expenditures during 14 consecutive days. The periods of data collection are evenly spread out over the year. The information is then compiled and provides a repeated series of yearly cross-sections.

4.2.2 Grouping of Goods, Income Definition and Data Clearance

All the goods are grouped into five categories, namely food, housing, travel and leisure, personal expenses, alcohol and tobacco. The category food consists of the subcategories food bought and eating outside of home, which are self explanatory. In contrast to this, housing is a more heterogeneous category; it consists of rent or mortgage payments as well as household items like furniture, but also DIY and water charges are subsumed here. Personal expenses consist mainly of clothing and personal goods (such as chemistry, jewelry etc.) and of personal services. Travel and leisure is again a rather mixed category, with travel including expenditures on car and public

transport, while leisure covers audio-visual articles, toys and holidays.
Since alcohol and tobacco are known to su¤er from serious underreporting, they are omitted. Additionally, personal expenses su¤er from infrequent purchases (recall that the recording period is 14 days) and are thus underreported. We excluded those persons with zero expenditure on personal expenses, and also those with the 0.5% highest expenditure levels for each composite good, reducing the total population by roughly 5%.
Income is constructed as in the de…nition of "household below average income study" (HBAI). It is roughly de…ned as net income after taxes, but including state transfers. This is done in both data sets to de…ne nominal income. Real income is then obtained by dividing through the retail price indices.

4.3 Issues in Estimation

4.3.1 Stone-Lewbel Cross Section Prices

The problem with the estimation of price e¤ects is closely tied to the fact that price nonstationary. As such a cointegration based analysis should be performed. However, there is a possibility we may circumvent the di¢culties associated with this issue. It comes from the fact that we are grouping goods to form composite goods, and that we can control this grouping since we have expenditure data on each single good. The standard practice of using a single price index amounts - as noted already by Stone (1954) and more recently by Lewbel (1989, 1999) - to assuming that all individuals consume all goods within a certain compostitum of goods in the same proportion, meaning that they have identical "within group" Cobb-Douglas (CD) preferences. This is an extremely unrealistic assumption that not only can, but actually should be relaxed. Moreover, dispensing with this assumption can be done at no extra costs, but with the extra bene…t of obtaining CS prices.
The alternative approach - and here we follow Lewbel (1989) - can be sketched as follows: For each compostitum the price for an individual is obtained by weighting the prices of good j by the individuals share of the expenditures of good j from total expenditure for all goods in this compositum. For details we refer to Lewbel (1989), where it is shown that this construction amounts to assuming that individuals have di¤erent CD preferences for all goods within a group, while individuals are allowed to have completely arbitrary preferences between various groups of goods.

4.3.2 The Issue of Dimensionality of the Vector of Characteristics

Recall that in our model we have neither restricted the Z_i vector - that is the vector of characteristics - nor the functional form of the $l_t(¢)$ function, or all the other conditional expectations involving Z_i. Moreover, our econometric model is geared for continuous data. Although we show in Hoderlein (2002a) that the curse of dimensionality does not a¤ect this model, we use principal components to reduce Z_{it}

to some three orthogonal components. This leads then to an implicit specification of $I_t(Z_{it}) = I_t(\phi_{1t}^T Z_{it}, \phi_{2t}^T Z_{it}, \phi_{3t}^T Z_{it})$. We define $Z_{it,new}^T = (\phi_{1t}^T Z_{it}, \phi_{2t}^T Z_{it}, \phi_{3t}^T Z_{it})$, This has a couple of advantages: 1 It yields continuous covariates. 2. The small sample performance is likely to be good. 3. Due to the orthogonality of the new regressors, we may use a diagonal bandwidth matrix. Since we normalize the components, we can further apply the same amount of smoothing in every direction. 4. collinearity is excluded. The econometric model (4.2) has the additional advantage of including the λ-term, which may include the original Z_{it} in full dimensionality, thus giving a semi-parametric control for the process of dimensionality reduction. Indeed, in the application, the remainder parts of the Z_{it} yield only insignificant t and F-statistics, with associated p-values close to one.

4.3.3 Choice of Bandwidth

First experiments with the bandwidth, a parameter that governs local model complexity, suggest that theoretically optimal bandwidths, in the sense defined in Hoderlein (2002a), results in somewhat undersmoothed estimates. We believe this however to be somewhat problematic, since local rises and dips in income elasticities, for instance, are hard to interpret and are most probably not a “feature” of reality. Thus, the choice of bandwidth is guided largely by economic intuition on the images displayed.

4.3.4 Tests

Here we describe briefly two tests: The first is a test for negative semidefiniteness (nsd) of the Slutsky-matrix, the second a test for symmetry. Both tests are performed at 300 “representative” positions in the population.

1. Testing for nsd uses the fact that a matrix is nsd iff all eigenvalues are smaller than zero. Moreover, in our case all eigenvalues are real as the matrix appearing in P2.2 (iii) is symmetric. Having estimated $m, m_2, \overline{D_p m}$ as well as $D_y m_2$, we simply bootstrap all eigenvalues by naive bootstrapping. Hence, if the empirical distribution of the largest eigenvalue over 1000 bootstrap replications does not cover within its 2.5 and 97.5 percentile the 0 we conclude that the biggest eigenvalue is significantly negative. There are two potential pitfalls: The first is multiplicity of the eigenvalues. Since the empirical distributions of all eigenvalues appear to have disjunct range this issue seems not to be problematic. The other issue is that of a parameter on the boundary of the parameter space (Andrews (1993)). We offer no solution to this problem. However, we note that ”most of the time” the whole support of the empirical distribution of the largest eigenvalue appears to be negative.

2. Testing for symmetry is a bit involved, as it involves cross section restrictions that seem to be hardly compatible with the nonparametric approach taken. However, it remains possible under mild assumptions. There are two distinct viewpoints one

can adopt. The first is a "pointwise" one: Since every fixed position represents an average over a population, tests at fixed positions are warranted. Here we may use the key observation that the derivative estimators form a "local SURE" system on the transformed data. In our econometric model, setting $\hat{s}^{jk}(\zeta) = \hat{s}^{kj}(\zeta)$, at a fixed position $\zeta = p, y, z$, is

$$\left(\hat\beta^j_k(\zeta) + \hat\beta^j_x(\zeta)\hat\alpha^k(\zeta)\right) - \left(\hat\beta^k_j(\zeta) + \hat\beta^k_x(\zeta)\hat\alpha^j(\zeta)\right) = 0;$$

where $\hat\beta^j_k(\zeta) = D_{p_k}\hat m^j(\zeta)$; $\hat\beta^j_x(\zeta) = D_y\hat m^j(\zeta)$ and $\hat\alpha^j(\zeta) = \hat m^j(\zeta)$; for all $j = 1, \ldots, L-1$, $k = j+1, \ldots, L$, yielding $L(L-1)/2$ symmetry restrictions. While this restriction looks nonlinear, after taking the differences in speed of convergence into account its asymptotics are as if it were a linear restriction. To see this, consider the following t-statistic. For ease of notation, in the denominator we have already concentrated on the variance parts belonging only to the two equations involved.

$$t^*_{sy}(\zeta; j) = \frac{\hat\beta^j_k(\zeta) + \hat\beta^j_x(\zeta)\hat\alpha^k(\zeta) - \hat\beta^k_j(\zeta) - \hat\beta^k_x(\zeta)\hat\alpha^j(\zeta)}{\left\{\nabla_\theta \hat g' \hat\Sigma(\zeta) \nabla_\theta \hat g\right\}^{\frac{1}{2}}};$$

where $\nabla_\theta \hat g = \left(\hat\beta^j_x(\zeta) \;\; 0\ldots0 \;\; 1/h \;\; 0\ldots0 \;\; \hat\alpha^k(\zeta)/h \;\; -\hat\beta^k_x(\zeta) \;\; 0\ldots0 \;\; -1/h \;\; 0\ldots0 \;\; -\hat\alpha^j(\zeta)/h\right)'$ and $\hat\Sigma(\zeta)$ is a consistent estimator of the covariance matrix of the scaled coefficients $\theta = (\alpha; h\beta)$:

Here the $1/h$ is due to the fact that the variances are defined on the h-scaled β, and taking the differences in speed of convergence into account. To understand the asymptotic behaviour of this statistic, consider first the numerator

$$\begin{aligned} &\sqrt{nh^d}\,h\left[\hat\beta^j_k(\zeta) + \hat\beta^j_x(\zeta)\hat\alpha^k(\zeta) - \hat\beta^k_j(\zeta) - \hat\beta^k_x(\zeta)\hat\alpha^j(\zeta)\right] \\ = &\sqrt{nh^{d+2}}\left[\hat\beta^j_k(\zeta) + \hat\beta^j_x(\zeta)\alpha^k(\zeta) - \hat\beta^k_j(\zeta) - \hat\beta^k_x(\zeta)\alpha^j(\zeta)\right] \\ &+ h\left[\hat\beta^j_x(\zeta)\right]\sqrt{nh^d}\left(\hat\alpha^k(\zeta) - \alpha^k(\zeta)\right) - h\left[\hat\beta^k_x(\zeta)\right]\sqrt{nh^d}\left[\hat\alpha^j(\zeta) - \alpha^j(\zeta)\right]; \end{aligned}$$

so that the first expression on the rhs converges by a trivial extension of P1 in distribution, while for the second and, for instance, for the third we have that $\sqrt{nh^d}\left[\hat\alpha^j(\zeta) - \alpha^j(\zeta)\right] \xrightarrow{d} N$; $\hat\beta^k_x(\zeta) \xrightarrow{p} \beta^k_x(\zeta)$; and thus $h\left[\hat\beta^k_x(\zeta)\right]\sqrt{nh^d}\left[\hat\alpha^j(\zeta) - \alpha^j(\zeta)\right] \xrightarrow{p} 0$: Hence, only the variance of the derivative estimators has to be taken into account, so that, instead of $\nabla_\theta \hat g$; we may use the following restriction in the denominator $\hat\Psi_\theta' \hat\Sigma_\Gamma(\zeta)\hat\Psi_\theta$,
where $\hat\Psi_\theta$ is defined as $\left[0 \;\; 0\ldots0 \;\; 1 \;\; 0\ldots0 \;\; \hat\alpha^k(z) \;\; 0 \;\; 0\ldots0 \;\; -1 \;\; 0\ldots0 \;\; -\hat\alpha^j(z)\right]$ and $\hat\Sigma_\Gamma(\zeta)$ is just the covariance matrix of the derivatives.

The second point of view is an "overall" one. When it comes to the population, or certain subpopulations, we may - instead of looking at a grid of positions - consider a single statistic. In particular, sample counterparts to

$$\sum_{j;k>j} \int \left\{ D_{p_k}^j m(\zeta) + D_y m^j(\zeta) m^k(\zeta) - \left[D_{p_j} m^k(\zeta) + D_y m^k(\zeta) m^j(\zeta) \right] \right\}^2 \mu_{YZ}(d\zeta) = 0$$

may be considered. The most natural choice is

$$\frac{1}{n} \sum_{j;k>j} \sum_i \left\{ \left(\hat{b}_k^j(\zeta_i) + \hat{b}_x^j(\zeta_i) \hat{b}^k(\zeta_i) \right) - \left(\hat{b}_j^k(\zeta_i) + \hat{b}_x^k(\zeta_i) \hat{b}^j(\zeta_i) \right) \right\}^2 .$$

The distribution theory for this statistic is transferred to a companion paper, Haag and Hoderlein (2003).

4.4 Preliminary Results

Using the model (4.2) in combination with the previously sketched tests, we obtain the following result in $t = 1985$: for eleven cells we are not able to reject the null of indefiniteness. We conclude, that the Weak Axiom appears to hold for roughly 97% of the population. This result does not change significantly for other time periods. Symmetry in turn seems harder to obtain: Roughly speaking, it seems to hold only for 60% of a population at a time. Thus we conclude that the Weak Axiom is almost uniformly accepted across the population, while Utility maximization is less well accepted.

5 Summary

In this paper we introduce a new framework which allows to model the demand behavior of a population with heterogeneous preferences of unknown type. Additionally, we allow for these preferences to be formed in the past by social interactions. We focus on the question what can be learned about this population from data, and how this can be done. Specifically, we focus on the four properties usually considered in demand system analysis: adding up, homogeneity of degree zero and negative semidefiniteness as well as symmetry of the Slutsky matrix. We establish that even in the most general scenario, all these quantities can be identified from nonparametric regression analysis under a conditional independence assumption. Furthermore, we give new characterizations for most of these objects in terms of observables.
We establish that the standard practise is a subcase with very restrictive assumptions, e.g. all preference heterogeneity is covered by observables. Moreover, we show how the main restrictive assumption, namely the conditional independence assumption, may be relaxed, if one has, for instance, additional information.

Preliminary results are indications for the strength of both economic theory and this approach. The Weak Axiom, arguably the core property of rationality, appears to hold uniformly across the population. Symmetry, of the Slutsky matrix is less well accepted, but then it is one of the results of this paper that the identification of symmetry rests on stronger assumptions.
Finally, this approach may be extended to any applied economic field, where heterogeneity of agents is to be modelled empirically.

6 Appendix

Proof of Proposition 2.1:

Ad $(i),(ii)$ First recall that, by definition, $0 < W < \infty$. Thus, the expectation exists and $E[|W|] \leq k < \infty$ (the same holds for the second moment). From this follows that all conditional expectations exist as well, and are even bounded. Let now p,y,z be fixed, but arbitrary. Then, inserting $A2.1$

$$D_y m(p,y,z) = D_y E[W|Y=y,Z=z] = D_y \int_A \phi(p,y,\vartheta(z,a))\,\mu_{A|Y,Z}(da,y,z)$$

Under $A2.4$, the rhs equals $D_y \int_A \phi(p,y,\vartheta(z,a))\,\mu_{A|Z}(da,z)$, and using the dominated convergence assumption $A2.3$, we obtain

$$\int_A D_y \phi(p,y,\vartheta(z,a))\,\mu_{A|Z}(da,z) \tag{A.1}$$

But due to $A2.4$ this is a version of $E[D_y\phi|Y=y,Z=z]$. Upon inserting random variables for the fixed z,y the statement follows. The proof is identical for D_p, save for the fact that we do not need $A2.4$.

For the part (iii) of the proposition, simply note that if A is Z-measurable

$$E[D_y\phi(p,Y,\vartheta(Z,A))|Y=y,Z] = D_y\phi(p,y,\theta(Z))$$

for any y and some function θ. ■

Proof of Corollary 2.2:

By iterated expectations and $P2.1$,

$$\begin{aligned} E[D_y\phi(p,Y,V)|F] &= E[E[D_y\phi(p,Y,V)|Y,Z|F] \\ &= E[D_y E[\phi(p,Y,V)|Y,Z|F] \\ &= E[D_y m(p,Y,Z)|F] \end{aligned}$$

for any $F \subseteq \sigma\{Y,Z\}$. The same holds of course for the trivial sigma algebra $\{\emptyset,\Omega\}$.

Proof of Corollary 2.3:

Consider $E[\pi_{jj}|T,Z]$ first. Note that

$$D_{p_j} m_{\log,j} = D_{p_j} E[\log(W_j)|Y,Z] = E\left[D_{p_j}\log(W_j)|Y,Z\right];$$

Since $D_{p_j}\log(W_j) = D_{p_j}\log(Q_j/Y) + 1$, the statement follows. $E[\pi_{ij}|Y,Z]$ and $E\left[\eta_j|Y,Z\right]$ by similar reasoning. Note further that

$$\frac{D_y m_{(j)}}{m_{(j)}} = \frac{D_y E\left[\phi_j|Y,Z\right]}{E\left[\phi_j|Y,Z\right]} = \frac{E\left[D_y\phi_j|Y,Z\right]}{E\left[\phi_j|Y,Z\right]},$$

but the rhs equals

$$\frac{E\left[D_y\phi_j|Y,Z\right]}{E\left[\phi_j|Y,Z\right]} = E\left(\frac{D_y\phi_j}{\phi_j}|Y,Z\right) + \frac{E\left[D_y\phi_j|Y,Z\right]}{E\left[\phi_j|Y,Z\right]} E\left[\frac{E\left[\phi_j|Y,Z\right]}{\phi_j} - 1|Y,Z\right] - V\left(D_y\phi_j, \frac{1}{\phi_j}|Y,Z\right).$$

Thus, only if the last two terms cancel, $\frac{D_y m_{(j)}}{m_{(j)}} - 1 = E\left[\eta_j|T,Z\right]$.

Proof of Proposition 2.4:

Ad (i) Assume adding up $\iota'\phi = 1$ (a.s.). Taking conditional expectations produces $\iota' m = E[\iota'\phi|Y,Z] = 1$ (a.s.), by which $\iota' m = 1$ (a.s.) is obvious.

Ad (ii) Assume homogeneity holds across the population, i.e. $\phi(p+\lambda, y+\lambda, V) = \phi(p,y,V)$ (a.s.) for all p,y. Thus

$$\begin{aligned} m(p,y,z) &= \int_A \phi(p,y,\vartheta(a,z))\,\mu_{A|Y,Z}(da;y,z) \\ &= \int_A \phi(p+\lambda,y+\lambda,\vartheta(a,z))\,\mu_{A|Y,Z}(da;y,z) \end{aligned}$$

But since $\mu_{A|Y,Z} = \mu_{A|Z}$, we have that $\mu_{A|Y,Z}(da;y,z) = \mu_{A|Z}(da;z) = \mu_{A|Y,Z}(da;y+\lambda,z)$. Thus,

$$\int_A \phi(p+\lambda,y+\lambda,\vartheta(a,z))\,\mu_{A|Y,Z}(da;y,z) = \int_A \phi(p+\lambda,y+\lambda,v)\,\mu_{A|Y,Z}(da;y+\lambda,z) = m(p+\lambda,y+\lambda,z)$$

Ad (iii); Note that for any random matrix $A(\omega)$ we have if $p'A(\omega)p \leq 0$ for all ω, it follows that upon taking expectations w.r.t. an arbitrary probability measure[12] μ

$$\int p'A(\omega)p\,\mu(d\omega) \leq 0 \iff p'\int A(\omega)\mu(d\omega)p \leq 0, \text{ for all } p \in \mathbb{R}^L.$$

[12] For conditional probability measures this works similarly in the spaces under consideration.

From this S nsd (a.s.) $\Rightarrow E[S|Y,Z]$ nsd (a.s.) is immediate. Let $E[S|Y,Z] = B$, and note that since the definition of negative semidefiniteness of a square matrix B of dim L involves the quadratic form, $p'Bp \le 0$, we see that if we put $\tilde{B} = B + B'$, we have

$$p'\tilde{B}p = p'Bp \text{ for all } p \in R^L,$$

and $\tilde{B}$ symmetric, implying that B is negative semidefinite if and only if $\tilde{B}$ is negative semidefinite. From

$$\begin{aligned} B &= E[S|Y,Z] \\ &= E[D_p\psi|Y,Z] + E[D_y\psi\psi'|Y,Z] + E[\psi\psi'|Y,Z] + E[\mathrm{diag}(\psi)|Y,Z] \\ &= B_1 + B_2 + B_3 + B_4 \end{aligned}$$

follows that $\tilde{B} = B + B' = B_1 + B_2 + B_3 + B_4 + B_1' + B_2' + B_3' + B_4' = \tilde{B}_1 + \tilde{B}_2 + 2(B_3 + B_4)$, since B_3 and B_4 are symmetric. Thus we have that

$$S \text{ nsd (a.s.)} \Rightarrow \tilde{B}_1 + \tilde{B}_2 + 2(B_3 + B_4) \text{ nsd (a.s.)}$$

From P2.1 it is apparent that $\tilde{B}_1 = D_pm + D_pm'$. To see that $\tilde{B}_2 = D_ym_2(p,y,z)$, first note that due to the boundedness of W the second moments and conditional moments exist, so that

$$D_ym_2(p,y,z) = D_yE[WW'|Y = y, Z = z] = D_y\int_A \psi(p,y,\vartheta(z,a))\psi'(p,y,\vartheta(z,a))\mu_{A|Y,Z}(da,y,z)$$

Finally, by a modification of A2.3, we have

$$D_y\int_A \psi(p,y,\vartheta(z,a))\psi'(p,y,\vartheta(z,a))\mu_{A|Y,Z}(da,y,z) = E[D_y(\psi\psi')|Y = y, Z = z],$$

but the rhs equals $E[D_y\psi\psi' + \psi D_y\psi'|Y = y, Z = z]$ which is $\tilde{B}_2$. B_3 and B_4 are trivial. Upon inserting random variables, the statement follows.

Ad (iv) First note that S symmetric implies that $K = D_p\psi + D_y\psi\psi'$ is symmetric, which implies that $E[K|Y,Z]$ is symmetric since

$$A_{ij} = E[K_{ij}|Y,Z] = E[K_{ji}|Y,Z] = A_{ji}.$$

This implies in turn that

$$\begin{aligned} E[K|Y,Z] &= E[D_p\psi|Y,Z] + E[D_y\psi\psi'|Y,Z] \\ &= E[D_p\psi|Y,Z] + E[D_y\psi|Y,Z]\,E[\psi'|Y,Z] + V[D_y\psi,\psi'|Y,Z] \end{aligned}$$

is symmetric, from which $E[D_p\phi|Y,Z] + E[D_y\phi|Y,Z]E[\phi'|Y,Z]$ is symmetric since $V[D_y\phi,\phi'|Y,Z]$ is assumed to be symmetric.
By Proposition 2.1. this equals $D_p m + D_y m m'$.

Ad (v) Consider first the case the implication of V is Z measurable:
The 'if' part follows from (iv) and the observation that under measurable V, $V[D_y\phi,\phi'|Y,Z] = 0$, and thus symmetric, by which $K = D_p m + D_y m m'$.
For the 'only if' we argue by contradiction: Assume $K_{ij} \neq K_{ji}$. We have to show now that $A_{ij} \neq A_{ji}$. But $A = K$ under measurable V, so that the result is obvious. This shows also why the converse does not hold under $V[D_y\phi,\phi'|Y,Z] = 0$ alone, because then $K_{ij} \neq K_{ji}$ does not necessarily imply $E[A_{ij}|Y,Z] \neq E[A_{ji}|Y,Z]$.

Consider now the reverse case, i.e. that
{S is symmetric and nsd iff $D_p m + D_y m m'$ is symmetric and nsd} implies V is Z measurable.
This is equivalent to:
If V is not Z measurable $\Rightarrow$

either {S is symmetric, nsd does not imply $D_p m + D_y m m'$ is symmetric, nsd}

or {$D_p m + D_y m m'$ is symmetric and nsd does not imply S is symmetric and nsd}.

The first statement can be true which is implied by $P2.4$ (iv) for $z = V[D_y\phi,\phi'|Y,Z]$ not symmetric. Also the second may be true. To give an example where under non-measurability of V $D_p m + D_y m m'$ is symmetric but S is not, consider a two goods example, with two possible realizations, where the superscript $l = 1,2$ denote these two realizations. Assume that $\pi^1 = \pi^2 = \frac{1}{2}$. Assume further that $D_y\phi = 0$ and that

$$S^1 = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}, \quad S^2 = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$$

Note that $E[S|Y,Z] = \begin{pmatrix} -1 & 1.5 \\ 1.5 & -1 \end{pmatrix}$ which is symmetric although the "individual" Slutsky matrices have not been so. $\square$

Proof of Proposition 3.1.1 Start by noting that

$$E[D_y\psi(p,Y,V)|X=x,Z=z] = -\int_{A\times Y}\psi(p,y,\theta(a,z))D_y f_{A,Y|X,Z}(a,y,x,z)\,da\,dy;$$

by integration by parts, and that

$$D_x E[\psi(p,Y,V)|X=x,Z=z] = \int_{A\times Y}\psi(p,y,\theta(a,z))D_x f_{A,Y|X,Z}(a,y,x,z)\,da\,dy,$$

by dominated convergence. Rewriting this into one expression,

$$\begin{aligned}
& E[D_y\psi(p,Y,V)|X=x,Z=z] \\
= \; & D_x E[\psi(p,Y,V)|X=x,Z=z] \\
& + (E[D_y\psi(p,Y,V)|X=x,Z=z] - D_x E[\psi(p,Y,V)|X=x,Z=z]) \\
= \; & D_x E[\psi(p,Y,V)|X=x,Z=z] \\
& - \int_{A\times Y}\psi(p,y,\theta(a,z))D_y f_{A,Y|X,Z}(a,y,x,z)\,da\,dy \\
& - \int_{A\times Y}\psi(p,y,\theta(a,z))D_x f_{A,Y|X,Z}(a,y,x,z)\,da\,dy \\
= \; & D_x E[\psi(p,Y,V)|X=x,Z=z] \\
& - \int_{A\times Y}\psi(p,y,\theta(a,z))\left[D_y f_{A,Y|X,Z}(a,y,x,z) + D_x f_{A,Y|X,Z}(a,y,x,z)\right]da\,dy
\end{aligned}$$

Rewrite the correction term using

$$\begin{aligned}
f_{Y,X|Z}(y,x,z) &= f_{Y,Q|Z}(y,x-y,z) = f_{Y|Z}(y,z)f_Q(x-y)\,, \\
f_{X|Z}(x,z) &= \int f_{Y|Z}(s,z)f_Q(x-s)ds\,, \\
D_y f_{Y,X|Z}(y,x,z) &= D_y f_{Y|Z}(y,z)f_Q(x-y) - f_{Y|Z}(y,z)D_q f_Q(x-y)\,, \\
D_x f_{Y,X|Z}(y,x,z) &= f_{Y|Z}(y,z)D_q f_Q(x-y)\,, \\
D_x f_{X|Z}(x,z) &= \int f_{Y|Z}(s,z)D_q f_Q(x-s)ds,
\end{aligned}$$

we obtain

$$
\begin{aligned}
& -\int_{A\times Y} \phi(p,y,\theta(a,z))\left[D_y f_{A,Y|X,Z}(a,y,x,z) + D_x f_{A,Y|X,Z}(a,y,x,z)\right] da\,dy \\
= & -\int_{A\times Y} \phi(p,y,\theta(a,z))\left[D_y f_{Y|X,Z}(y,x,z) + D_x f_{Y|X,Z}(y,x,z)\right] f_{A|Z}\, da\,dy \\
= & -\int_{A\times Y} \phi(p,y,\theta(a,z))\frac{1}{f_{X|Z}}\left[D_y f_{Y,X|Z} + D_x f_{Y,X|Z} - D_x f_{X|Z} f_{Y|X,Z}\right] f_{A|Z}\, da\,dy \\
= & -\int_{A\times Y} \phi(p,y,\theta(a,z))\frac{f_Q(x-y)}{\left(f_{X|Z}(x,z)\right)^2}\left[D_y f_{Y|Z} f_{X|Z} - D_x f_{X|Z} f_{Y|Z}\right] f_{A|Z}\, da\,dy \\
= & \int_{A\times Y} \phi(p,y,\theta(a,z))\frac{f_Q(x-y) f_{Y|Z}}{f_{X|Z}(x,z)}\left[D_x \log f_{X|Z} - D_y \log f_{Y|Z}\right] f_{A|Z}\, da\,dy;
\end{aligned}
$$

where we suppressed the arguments whenever obvious. But since

$$\frac{f_Q(x-y) f_{Y|Z}(y,z)}{f_{X|Z}(x,z)} f_{A|Z}(a,z) = f_{A,Y|X,Z}(a,y,x,z);$$

the last rhs equals

$$
\begin{aligned}
& E\left[W\left[D_x \log f_{X|Z}(X,Z) - D_y \log f_{Y|Z}(Y,Z)\right] | X = x, Z = z\right] \\
= & \; D_x \log f_{X|Z}(x,z)\, E[W | X = x, Z = z] \\
& - E\left[W D_y \log f_{Y|Z}(Y,Z) | X = x, Z = z\right].
\end{aligned}
$$

Using the fact that

$$
\begin{aligned}
& E\left[W D_y \log f_{Y|Z}(Y,Z) | X = x, Z = z\right] \\
= & \; E\left[D_y \log f_{Y|Z}(Y,Z) | X = x, Z = z\right] E[W | X = x, Z = z] \\
& + V\left[W, D_y \log f_{Y|Z}(Y,Z) | X = x, Z = z\right];
\end{aligned}
$$

an that Integration by parts yields

$$
\begin{aligned}
& E\left[D_y \log f_{Y|Z}(Y,Z) | X = x, Z = z\right] \\
= & \; \frac{1}{f_{X|Z}(x,z)} \int_{A\times Y} D_y f_{Y|Z}(y,z) f_Q(x-y) f_{A|Z}(a,z)\, da\,dy \\
= & \; \frac{1}{f_{X|Z}(x,z)} \int_{A\times Y} f_{Y|Z}(y,z) D_q f_Q(x-y) f_{A|Z}(a,z)\, da\,dy \\
= & \; \frac{1}{f_{X|Z}(x,z)} \int_{A\times Y} D_x f_{Y,X|Z}(y,x,z) f_{A|Z}(a,z)\, da\,dy \\
= & \; D_x \log f_{X|Z}(x,z);
\end{aligned}
$$

we obtain

$$D_x E[\phi(p;Y;V)|X=x;Z=z] = E[D_y\phi(p;Y;V)|X=x;Z=z]+\xi;$$

where

$$\xi = V\left[\phi(p;Y;V);D_y \log f_{Y|Z}(Y;Z)|X=x;Z=z\right]$$

But since $Y|Z \sim N(\mu(z);\sigma^2(z))$;

$$D_y \log f_{Y|Z}(y;z) = -\frac{y-E[Y|Z=z]}{V[Y|Z=z]}$$

we have

$$V\left[\phi(p;Y;V);D_y \log f_{Y|Z}(Y;Z)|X=x;Z=z\right] = \frac{-V[\phi(p;Y;V);Y|X=x;Z=z]}{V[Y|Z=z]}:$$

By a partial second order Taylor-expansion

$$\begin{aligned} V[\phi(p;Y;V);Y|X=x;Z=z] &= D_y\phi(p;y_0;V)V[Y-y_0;Y|X=x;Z=z] \\ &\quad + V\left(\frac{D_y\phi(p;y_r;V)}{2}(Y-y_0)^2;Y|X=x;Z=z\right); \end{aligned}$$

and using

$$\int V\left(\frac{D_y^2\phi(p;y_r;V)}{2}(Y-y_0)^2;Y|X=x;Z=z\right)\mu_{Y;V|X;Z}(dy_0;dv;x;z)=0$$

we obtain

$$\begin{aligned} & V\left[\phi(p;Y;V);D_y \log f_{Y|Z}(Y;Z)|X=x;Z=z\right] \\ = & \int D_y\phi(p;y_0;v)V\left[Y;D_y \log f_{Y|Z}(Y;Z)|X=x;Z=z\right]\mu_{Y;V|X;Z}(dy_0;dv;x;z)+0 \\ = & -E[D_y\phi(p;Y;V)|X=x;Z=z]\,V[Y|X=x;Z=z]: \end{aligned}$$

Hence,

$$D_x E[\phi(p;Y;V)|X=x;Z=z] = (1-\psi)E[D_y\phi(p;Y;V)|X=x;Z=z];$$

where

$$\psi = \frac{V[Y|X=x;Z=z]}{V[Y|Z=z]}:$$

Proof of Proposition 3.1.2

Ad (i) Taking conditional expectations as above.

Ad (ii) Note from P3.1.1 that $D_xM = (1 - \psi)E[D_y\phi|X, Z] < E[D_y\phi|X, Z]$, with $\psi > 0$. Since homogeneity implies $E[D_p\phi|X, Z]\iota + E[D_y\phi|X, Z] = 0$, we have three cases: If $D_y\phi < 0$, which implies that $E[D_y\phi|X, Z] < 0$, we have $0 = D_pM\iota + E[D_y\phi|X, Z] > D_pM\iota + D_xM$. Second, if $D_y\phi > 0$ the argument may be reversed. Third, if $D_y\phi = 0$, $E[D_y\phi|X, Z] = 0$ and $D_xM = 0$.

Ad (iii) As in P2.4, S nsd implies $E\left[S|X, Z\right]$ nsd or

$$E\left[\overline{D_p\phi}|X, Z\right] + E[D_y[\phi\phi']|X, Z] + 2(E[\phi\phi'|X, Z] - \mathrm{diag}(E[\phi|X, Z])) \leq 0$$

in a matrix sense. As above $E\left[\overline{D_p\phi}|X, Z\right] = \overline{D_pM}$, $E[\phi\phi'|X, Z] = M_2$ and $\mathrm{diag}(E[\phi|X, Z]) = \mathrm{diag}(M)$, so that it is only the second term that needs closer inspection. By the same argument as in P3.1.1, $D_xM_2 = (1 - \psi)E[D_y[\phi\phi']|X, Z]$, since only the function changes from ϕ to $\phi\phi'$. Thus

$$\overline{D_pM} + D_xM_2 + \psi E[D_y[\phi\phi']|X, Z] + 2(M_2 - \mathrm{diag}(M)) \leq 0.$$

Moreover, ψ is a scalar with $\psi(X, Z) = \dfrac{V[Y|X, Z]}{V[Y|Z]} > 0$, so that $\psi E[D_y[\phi\phi']|X, Z] \geq 0$ in a matrix sense. Thus

$$\begin{aligned} & \overline{D_pM} + D_xM_2 + 2(M_2 - \mathrm{diag}(M)) \\ \leq\ & \overline{D_pM} + D_xM_2 + \psi E[D_y[\phi\phi']|X, Z] + 2(M_2 - \mathrm{diag}(M)) \leq 0, \end{aligned}$$

which shows the statement.

Ad (iv) If $D_xM = 0$ and $V[D_y\phi, \phi'|X = x, Z = z]$ the only nonsymmetric term in

$$E[D_p\phi|X, Z] + E[D_y[\phi\phi']|X, Z] + 2(E[\phi\phi'|X, Z] - \mathrm{diag}(E[\phi|X, Z]))$$

is $E[D_p\phi|X, Z]$, which implies D_pM symmetric.

Proof of Proposition 3.2.1:

Let p, y, z be fixed, but arbitrary. Then, as above

$$D_y m(p,y,z) = D_y E[W|Y = y, Z = z] =$$

$$D_y \int_{Z^- \times A} \varphi(p, y, \vartheta(z^-, a))\, \mu_{Z^- A|Y,Z}(dz^-, da; y, z)$$

Note that by the change of variable lemma (with $h^{-1}(z)$ the inverse function of h evaluated at z) and using the fact that the Jacobian determinant of the transformation of (A, Z^-) to $(A, h^{-1}(z) - U)$ equals unity - this term becomes

$$D_y \int_{U \times A} \varphi(p, y, \vartheta(h^{-1}(z) - u, a))\, \mu_{A,U|Y,Z}(da, du; y, z) =$$

$$D_y \int_U \int_A \varphi(p, y, \vartheta(h^{-1}(z) - u, a))\, \mu_{A|U,Z}(da; z, u)\, \mu_{U|Z}(du; z)\ .$$

The rest is in the text.

Proof of Proposition 3.2.2:

Ad (i) As in P2.4.

Ad (ii) Assume homogeneity holds across the population, i.e. $\varphi(\lambda p, \lambda y, V) = \varphi(p, y, V)$ (a.s.) for all p, y. Thus

$$\begin{aligned} m(p,y,z) &= \int_V \varphi(p, y, v)\, \mu_{V|Y,Z}(dv; y, z) \\ &= \int_V \varphi(\lambda p, \lambda y, v)\, \mu_{V|Y,Z}(dv; y, z) \end{aligned}$$

But since $\mu_{V|Y,Z} = \mu_{V|Z}$, we have that $\mu_{V|Y,Z}(dv; y, z) = \mu_{V|Z}(dv; z) = \mu_{V|Y,Z}(dv; \lambda y, z)$. Thus,

$$\int_V \varphi(\lambda p, \lambda y, v)\, \mu_{V|Y,Z}(dv; y, z) = \int_V \varphi(\lambda p, \lambda y, v)\, \mu_{V|Y,Z}(dv; \lambda y, z) = m(\lambda p, \lambda y, z)$$

Ad (iii) As in P2.4 (iii), with the exception that $\bar{B}_2 = D_y m_2(p,y,z) - Cor2$. To see this, note that

$$
\begin{aligned}
& E[D_y\phi\phi' + \phi D_y\phi' | Y=y, Z=z] \\
= \; & E[D_y(\phi\phi')|Y=y, Z=z] \\
= \; & D_y m_2(p,y,z) \\
& - \int_U\int_A \phi(p,y,\vartheta(h^{-1}(z)-u,a))\phi(p,y,\vartheta(h^{-1}(z)-u,a))' D_y f_{U|Y,Z}(u,z,y)du\,\mu_{A|Z}(du,z) \\
= \; & D_y m_2(p,y,z) - E[WW' D_y f_{U|Y,Z}(u,z,y)|Y=y, Z=z],
\end{aligned}
$$

Denoting $E[WW' D_y f_{U|Y,Z}(u,z,y)|Y=y, Z=z] = Cor2(p,y,z)$, yields the statement.

Ad (iv) As previously.

Proof of Lemma 3.4.

Start by noting that

$$
\begin{aligned}
& V[E[D_y Á|Y;Z;U];E[Á^0|Y;Z;U]|Y;Z] \\
& = V[D_y E[Á|Y;Z;U];E[Á^0|Y;Z;U]|Y;Z] \\
& = V[D_y(E[Á|Y;Z;U] ¡ E[Á|Y;Z]);E[Á^0|Y;Z;U] ¡ E[Á^0|Y;Z]|Y;Z] \\
& = V[D_y ¥U;U^0¥^0|Y;Z] \\
& = D_y ¥ ¢ ¥^0
\end{aligned}
$$

Proof of Proposition 3.4

The model is given by
$Á = m(p;y;z;u) + §(p;y;z;u)¸(a)$; s.th. $E[¸|Y;Z;U] = ³(Z;U)$ and $V[¸|Y;Z;U] = ¤(Z;U)$:

Ad (i) Assume $p^0Á = Y$ (a:s:): Then follows $p^0E[Á|Y;Z] = Y$ and $p^0E[E[Á|Y;Z;U]|Y;Z] = Y$: But this is $p^0E[m|Y;Z] + E[§|Y;Z]³ = Y$ and $p^0n = Y$ is immediate.

Ad (ii) Similar argument as in P2:2: (ii):

Ad (iii) By the same argument as in P2:2 (iii) follows that if S is nsd than so is

$$\overline{E[S|Y;Z]} = \frac{\overline{E[D_p Á|Y;Z]} + E[D_y\{ÁÁ^0\}|Y;Z]}{2};$$

where $\overline{W}$ denotes the symmetrized version of the matrix W: Since

$$
\begin{aligned}
\frac{1}{2}E[D_y\{ÁÁ^0\}|Y;Z] &= \frac{1}{2}E[D_y ÁÁ^0 + ÁD_y Á^0|Y;Z] \\
&= \frac{1}{2}\overline{D_y nn^0} ¡ \overline{D_y f}nn^0 + \frac{1}{2}\overline{V[D_y Á;Á^0|Y;Z]}
\end{aligned}
\tag{A.3}
$$

and

$$
\begin{aligned}
\frac{1}{2}D_y n_2 &= \frac{1}{2}D_y E[ÁÁ^0|Y;Z] \\
&= \frac{1}{2}\overline{D_y nn^0} + \frac{1}{2}D_y V[Á|Y;Z]
\end{aligned}
$$

follows that

$$D_y n_2 ¡ E[D_y ÁÁ^0|Y;Z] = D_y V[Á|Y;Z] + 2\overline{D_y f}nn^0 ¡ \overline{V[D_y Á;Á^0|Y;Z]}: \tag{A.4}$$

The …rst term on the rhs equals

$$D_y E\,[E\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]\ ¡\ D_y\,(E\,[Á jY;Z]\,E\,[Á^0 jY;Z]) + D_y E\,[V\,[Á jY;Z;U]\,jY;Z]:$$

By the measurability assumption $E\,[V\,[Á jY;Z;U]\,jY;Z] = §L§^0$. Moreover, $D_y\,(E\,[Á jY;Z]\,E\,[Á^0 jY;Z]) = \overline{D_y nn^0}$:
Turning to the last term in (A:4), note that this equals

$$\overline{E\,[E\,[D_y Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]}\ ¡\ \overline{E\,[D_y Á jY;Z]\,E\,[Á^0 jY;Z]} + \overline{E\,[V\,[D_y Á;Á^0 jY;Z;U]\,jY;Z]}$$

The third term is $\overline{D_y §L§^0}$ and the second $\overline{D_y nn^0}$: Due to L not a function of y;

$$D_y\,(§¤§^0)\ ¡\ \overline{D_y §¤§^0} = 0;$$

and since

$$\overline{E\,[D_y Á jY;Z]\,E\,[Á^0 jY;Z]} = \overline{D_y nn^0}\ ¡\ 2\overline{D_y f}\,nn^0$$
$$\overline{E\,[D_y Á jY;Z]\,E\,[Á^0 jY;Z]}\ ¡\ D_y\,(E\,[Á jY;Z]\,E\,[Á^0 jY;Z]) =\ ¡\,2\overline{D_y f}\,nn^0$$

so that the di¤erence in (A:4) reduces to

$$D_y E\,[E\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]\ ¡\ \overline{E\,[E\,[D_y Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]}: \quad (A.5)$$

Consider the second term. By A3:4:4, A3:4:5

$$\overline{E\,[E\,[D_y Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]} = E\,[D_y\ fE\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]g\,jY;Z]:$$

and

$$\begin{aligned} E\,[D_y\ fE\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]g\,jY;Z] &= D_y E\,[E\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]\ ¡ \\ &\quad \overline{D_y f}\,E\,[E\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]: \end{aligned}$$

Consequently, (A:5) reduces to
$\overline{D_y f}\,E\,[E\,[Á jY;Z;U]\,E\,[Á^0 jY;Z;U]\,jY;Z]$: But this equals,

$$\begin{aligned} &\overline{D_y f}\,(E\,[E\,[ÁÁ^0 jY;Z;U]\,jY;Z] + E\,[V\,[Á;Á^0 jY;Z;U]\,jY;Z]) \\ =\ &\overline{D_y f}\,(n_2 + §L§^0)\,; \end{aligned}$$

and $\frac{1}{2}\overline{E\,[SjY;Z]} = \dfrac{D_p n + D_y n_2 + \overline{D_y f}\,(n_2 +\ ¡)}{2}$ follows.

Ad (iv) From

$$\begin{aligned} E\,[SjY;Z] &= E\,[D_p Á jY;Z] + E\,[D_y ÁÁ^0 jY;Z] \\ &= D_p n + E\,[D_y Á jY;Z]\,E\,[Á^0 jY;Z] + V\,[D_y Á;Á^0 jY;Z] \qquad (A.6) \\ &= D_p n + E\,[D_y Á jY;Z]\,E\,[Á^0 jY;Z] + E\,[V\,[D_y Á;Á^0 jY;Z;U]\,jY;Z] + \\ &\quad V\,[E\,[D_y Á jY;Z;U]\,;E\,[Á^0 jY;Z;U]\,jY;Z] \end{aligned}$$

The second to last term is symmetric by assumption. The second term is obvious by Lemma 3:2; and only the third term needs to be simpli…ed. But this equals

$$E[D_y §V[¸jY;Z;U]§^0jY;Z] = D_y §L(Z)§^0 = D_y \hat{§}\hat{§}^0$$

Since, by Lemma 3.1, ¡ $= E['\,'^0jY;Z] = R^2$ and thus ¡$^{\frac{1}{2}} = \hat{§}$, the result follows.

Ad (v) Under A3:9; by Lemma 3.3 $V[E[D_y ÁjY;Z;U];E[Á^0jY;Z;U]jY;Z] = D_y ¥¢¥^0$: Inserting into (A.6) produces the result.

Ad (vi) As in P2:2 (v):

Literature:

AFRIAT, S. N. (1967); The Construction of Utility Functions from Expenditure Data, International Economic Review, 8, 67-77.

BANKS, J., BLUNDELL, R. and A. LEWBEL, (1997); Quadratic Engel Curves and Consumer Demand, Review of Economics and Statistics, 527-539.

BROWN, R. and M.B. WALKER, (1989); The Random Utility Hypothesis and Inference in Demand Systems, Econometrica, 57, 815-829.

BROWNING, M., (1991); A Simple Nonadditive Preference Structure for Models of Household Behavior over Time, Journal of Political Economy, 99, 607-37.

BLUNDELL, R., BROWNING, M. and I. CRAWFORD, Nonparametric Engel Curves and Revealed Preferences, IFS working paper, 2002.

BLUNDELL, R., PASHARDES, P. and G. WEBER, (1993); What Do We Learn About Consumer Demand Patterns from Micro Data?, American Economic Review, 83, 570-597.

CHIAPPORI, P.A. and J.C. ROCHET, (1987), Revealed Preferences and Di¤erentiable Demand: Notes and Comments, Econometrica, 55, 687-691.

CHRISTOPEIT, N. and S.G.N. HODERLEIN (2002); Estimation of Models with Additive Structure via Local Quasi-di¤erencing, Working Paper, Mannheim University Working Paper.

DEATON, A., and J. MUELLBAUER, (1980); An almost ideal Demand System, American Economic Review, (70), 312-26.

HODERLEIN, S., (2002), Econometric Modelling of Heterogeneous Consumer Behavior - Theory, Empirical Evidence and Aggregate Implications, PhD Thesis, LSE.

HODERLEIN,S., (2002); Nonparametric Demand of a Heterogeneous Population, unpublished manuscript, Mannheim University.

JERISON, M., (1984); Aggregation and Pairwise Aggregation of Demand when the Distribution of Income is …xed, Journal of Economic Theory, 33, 1-31.

JORGENSEN, D., LAU, L. and T. STOKER, (1982); The Transcedental Logarithmic Model of Individual Behavior. In: BASMAN, R. and G. RHODES (Eds.), Advances in Econometrics, Vol 1. JAI Press.

LEWBEL, A., (1989); Identi…cation and Estimation of Equivalence Scales under Weak Separability, Review of Economic Studies, 56, 311-16.

LEWBEL, A., (1990); Income Distribution Movements and Aggregate Money Illusion, Journal of Econometrics, 43, 35-42.

LEWBEL, A., (1999); Consumer Demand Systems and Household Expenditure, in PESARAN, H. and M. WICKENS (Eds.), Handbook of Applied Econometrics, Blackwell Handbooks in Economics.

LEWBEL, A., (2001); Demand Systems With and Without Errors, American Economic Review, 611-18.

LUENBERGER, D.G., (1997); Optimization by Vector Space Methods, Wiley&Son, New York.

MAMMEN, E. , (1993); Bootstrap in Linear Models, Annals of Statistics, 21, 255-285.

MANSKI, C., (2000); Economic Analysis of Social Interaction, Journal of Economic Perspectives, 14, 115-136

McFADDEN, D., (2001); Economic Choices, American Economic Review, 351-79.

NEWEY, W.K., J.L. POWELL and F. VELLA (1999); Nonparametric Estimation of Triangular Simultaneous Equations Models, Econometrica, 67, 565-603.

STONE, J.R.N., (1954); Linear Expenditure Systems and Demand Analysis, Economic Journal, 64, 511-24.

VARIAN, H., (1982); The Nonparametric Approach to Demand Analysis; Econometrica, 50, 945-974.