Developments in Nonparametric Demand Analysis: Heterogeneity and Nonparametrics

Stefan Hoderlein

Brown University

December 30, 2008

Abstract

This paper discusses new development in nonparametric econometric approaches related to empirical modeling of demand decisons. It shows how diverse recent approaches are, and what new modeling options arise in practise. We review work on nonparametric identification using nonseparable functions, semi-and nonparametric estimation approaches involving inverse problems, and nonparametric testing approaches. We focus on classical consumer demand systems with continuous quantities, and do not consider approaches that involve discrete consumption decisions as are common in empirical industrial organization. Our intention is to give a subjective account on the usefulness of these various methods for applications in the field.

Keywords: Nonparametric, Integrability, Testing Rationality, Nonseparable Models, Demand, Nonparametric IV.

1 Introduction

Economic theory yields strong implications for the actual behavior of individuals. In the standard utility maximization model for instance, economic theory places significant restrictions on individual responses to changes in prices and wealth, the so-called integrability constraints. However, economic theory says very little about the functional form of the preference ordering of any given individual and the homogeneity of any given population. Indeed, in applications using microdata, heterogeneity seems to be a prevalent feature of the data: any least squares regression explains only around 10% of the variation in the data. While measurement error is certainly partially responsible for the large amount of unobserved variation, it is very likely that a large fraction of the variation in the data is due to unobserved preference heterogeneity on behalf of the individuals surveyed. Put reversely, it is a common empirical finding that individuals with very similar values of the regressors show widely different values of the dependent variable, a fact that is only hard to reconcile solely with measurement errors, in particular since the same finding is obtained through several different methods of data recording, including scanner data and diaries that are widely believed to be less measurement error ridden. For both reasons - unrestricted heterogeneity and lack of natural functional form assumptions - nonparametric methods seem well suited tools for data analysis.

This chapter gives a - necessarily subjective - account of recent research related to nonparametric modeling of consumer decisions, with a particular focus on unobserved heterogeneity. This section focuses on demand systems in the spirit of Deaton and Muellbauer (1980), or Lau, Jorgenson and Stoker (1982). Neither do we discuss approaches related to nonparametric revealed preference type of approaches (see Blundell, Browning and Crawford (2008) and Cherchye, Crawford, de Rock and Vermeulen (this issue) for an overview), nor do we review related approaches on discrete choice in empirical IO (see Berry, Levinsohn, Pakes (1995) and Ackerberg, Benkard, Berry, and Pakes (2008) for an overview). Our focus is on traditional demand analysis involving continuous quantities. Parametric versions of approaches put forward in this paper are, e.g., the Translog demand system, Jorgenson et al. (1982), and the Almost Ideal demand system, Deaton and Muellbauer (1980)). Our analysis is also closely related to textbook modeling of demand in economic theory (see for instance Mas-Colell et al. (2005), Chapters 2- 4).

2 Models and Literature

Setup: On individual level, we define a demand function to be a causal relationship between budget shares, a $[0, 1]$ valued random L-vector denoted by W, and regressors of economic importance, namely log prices P and log total expenditure Y, real valued random vectors of length L and 1, respectively. Let $X = (P', Y)' \in \mathbb{R}^{L+1}$. To capture the notion that preferences vary across the population, we assume that there is a random variable $V \in \mathcal{V}$, where $\mathcal{V}$ is a Borel space[1], which denotes preferences (or more generally, decision rules). We assume that heterogeneity in preferences is partially explained by observable differences in individuals' attributes (e.g., age), which we denote by the real valued random G-vector Q. Hence, we let $V = \tau(Q, A)$, where τ is a fixed $\mathcal{V}$-valued mapping defined on the sets $\mathcal{Q} \times \mathcal{A}$ of possible values of (Q, A), and where the random variable A (taking again values in a Borel space $\mathcal{A}$) covers residual unobserved heterogeneity in a general fashion. With this notation, we can summarize a structural model of a heterogenous population through single equations, as shall become obvious in the following.

Models: To begin our discussion, suppose that on individual level the model is given by a linear relationship between all random variables. Specifically, we take the standard applied shortcut of using the approximate almost ideal model of Deaton and Muellbauer (1980), where we circumvent the nonlinearity in the income effect by using log real income as regressor. On individual level, the model is then given by

$$w = \beta^0 + \beta^{1\prime} p + \beta^2 y,$$

where all variables are as defined above, but the small letters indicate that we view an individual as a realization from a underlying heterogenous population (we could add a subscript i to all variables, but desist for brevity of notation. Small letters are from now on individuals or fixed positions, capital letters denote population quantities). The coefficients β vary across the heterogeneous population, and are a functions of the observable and unobservable determinants of the preferences, q and a. We will call the fact that all individuals share the same preference ordering, a "type" of individuals - in our example, everybody is of almost ideal type, and only the specific value of the parameters varies across the population. If we consider the entire population, our "heterogeneous approximately almost ideal" population may be formalized as:

$$W = \beta^0(Q, A) + \beta^1(Q, A)'P + \beta^2(Q, A)'Y. \tag{2.1}$$

.What has been done traditionally is to assume that $\beta^j(Q, A)$. $j = 0, 1, 2$, has an additive structure, e.g., $\beta^j(Q, A) = \beta^j_1 + \beta^j_2 Q + \beta^j_3 A$. This correponds to the standard notions of random

[1] Technically: $\mathcal{V}$ is a set that is homeomorphic to the Borel subset of the unit interval endowed with the Borel σ-algebra. This includes the case when V is an element of a polish space, e.g., the space of random piecewise continuous utility functions.

coefficient models, and leads to estimation of β_1^j and β_2^j by heteroscedasticity robust least squares methods.

There are three drawbacks of such an approach. First, even if one accepts linearity in p and y, the additive specification in Q and A is arguably ad hoc. Second, we may only determine (at best), mean parameters, and do not obtain the distribution of coefficients across the population. Finally, the random coefficient structue is at odds with correcting endogeneity using instruments that are mean independent of residuals.

The first point is arguable not a major concern, though one could do better. The second point is, however, restrictive if the interest centers, say on the worst and least afected parts of the population. The last point is commonly overlooked in applications, but should be a major concern. We will take up these points in the second section, and, in particular, we will see that in this case in general the distribution of random coefficients ("preference heterogeneity") is still identified.

A more general alternative is when we do not specify the entire population to be linear (or nonlinear with a finite dimensional, but heterogeneousparameter)- Instead we assume that the every individual in the population is characterized by a smooth (in the sense of differentiable) function ϕ only, i.e., the model is on individual level given by:

$$w = \phi(p, y).$$

This nests of course the case of a population consisting of a nonlinear parametric type. It is a little less obvious how a heterogeneous population looks like. To see this, assume that there is a (potentially infinte) dimensional parameter that varies across the population, (denoted U, for utility). Then, the population would be $W = \phi(P, Y, V)$. Assuming that the prefrence parameter V is determined by Q and A, we obtain that our model for the heterogeneous population is given by

$$W = \phi(P, Y, Q, A). \tag{2.2}$$

In contrast to the heterogeneous type AI population given in (2.1), this population is only identified, if we obsere any indivdiual in this population infinitely often. Certainly, a single cross section is not suffficient to identify features of interest in the population. The obvious question that arises is: what can we learn from data about this model? And how does endogeneity of preferences afftect what we can learn? The answer is given in the third section, where we consider this model in more detail.

Related Literature - the Nonparametric Additive Errors Approach: A key feature of both (2.1) and (2.2) is the nonparametric character of the distribution of preferences. Indeed, this separates this line of research from other work on that introduces nonparametric elements in a traidtional model with additive errors. Indeed, while one may criticize the additive error approach as being restrictive on heterogeneity (see Hoderlein (2002, 2008)), there has been significant progress on nonparametric modeling related to consumer demand. Suppressing Q, the standard nonparametric model is defined by

$$Y = m(X) + \varepsilon, \tag{2.3}$$

with errors ε that are mean independent of regressors. This model is of course not very useful in high dimensional demand applications because of the curse of dimensionality, and there have been several suggestions in the demand literature to impose structure. As was pointed out in Blundell, Browning and Crawford (2003), there is an issue with imposing the general forms of structure suggested in the econometrics literature: The imposed structure may be incompatible with economic theory in all but trivial cases. For instance, an additive model where $m(X) = \sum m_j(X_j)$, and $j = 1, .., d$ denotes the j-the elemnt of X, is only in line with economic theory in conumer demand if it is linear.

This has lead several researchers to suggest partially linear models with scaling factors, see Pendakur (1999) and Blundell, Duncan and Pendakur (1998). The general aim of this research is to tackle the curse of dimensionality by being parametric in the price dimension (where there is little variation in the data), and nonparametric in the income dimension, where there is a lot. To this date, like many nonparametric methods, these models have not been employed frequently, although their performance is quite satisfactory, and reveals very interesting findings, like the fact that many consumers seem to share the same shape of the demand function, just shifted by different values of the intercept. In a similar spirit (i.e., being parametric in the price dimension) is also Pendakur's EASI approach (Lewbel and Pendakur (2008), Pendakur and Sperlich (2008), the latter one is semiparametric). All of these models aim at greater flexibility of in particular the income effect on consumption.

Introducing endogeneity in these models proves to be a major challenge, as the standard mean independence restriction of instruments Z from errors (e.g., ε in (2.3)) leads in the case of semi- or nonparametric models generally to so called ill-posed inverse problems, meaning that the estimator can be obtained by inverting an integral equation. For instance, in model (2.3), the equation that has to be solved for m is

$$\mathbb{E}\left[Y|Z\right] = \mathbb{E}\left[m(X)|Z\right]. \tag{2.4}$$

In the language of this literature, the conditional expectation is treated as an operator T acting on a function m (think of this operator as a matrix). Hence, what is required is inversion of T. However, this inversion is non trivial and not well behaved, and to achieve a stable solution regularization has to be employed[2]. Such an approach has been put forward in Blundell, Chen and Kristensen (2007). From an applied perspective, this adds another layer of difficulties to an already involved analysis, while from a theoretical point of view many problems (including such important issues as having additional exogenous control variables) remain unresolved. From our perspective, the main drawback is the lack of compatibility with a heterogeneous population, see Hoderlein and Holzmann (2008) for a discussion. The core argument can easily be illustrated with the linear random coefficients model. Suppose the population is of the approximate AI type, with parameters that vary across the population, i.e.,

$$W = \beta^0(A) + \beta^1(A)'P + \beta^2(A)'Y, \tag{2.5}$$

where for simplicity we suppressed the dependence on Q. Assume in addition that we have correlation between preferences A and Y. A common heuristic in demand is that total expenditure is also determined by a part of the preference ordering. Assume, however, that we have instruments Z (including the exogenous prices P), such that

$$Y = \alpha + \gamma' Z + U,$$

where α, γ are fixed coefficients, and U are residuals. Suppose in addition that instruments are independent of both A and U jointly[3]. Then, (2.5) can be rewritten in a form involving an additive error,

$$W = \beta^0 + \beta^{1\prime}P + \beta^{2\prime}Y + \varepsilon, \tag{2.6}$$

where $(\beta^0, \beta^1, \beta^2)$ denotes the mean of the vector of random coefficients, and $\varepsilon = \beta^0(A) - \beta^0 + \left(\beta^1(A) - \beta^1\right)' P + \left(\beta^2(A) - \beta^2\right)' Y$ contains the preference heterogeneity. Then it is straightforward to show that $\mathbb{E}\left[\varepsilon|Z\right] \neq 0$ unless there is no randomness in $\beta^2(A)$. In Hoderlein and

[2]Roughly speaking, Ridge regression, i.e., using $\left[\mathbf{X}'\mathbf{X} + \lambda\mathbf{I}\right]^{-1}$, instead of $\left[\mathbf{X}'\mathbf{X}\right]^{-1}$ in OLS, is the closest analog in parametric regression. In this analogy, adding $\lambda\mathbf{I}$ would be the regularization, and λ is the regularization parameter. This regularization is performed in the case when the inversion of the matrix $X'X$ would produce an unstable result

[3]This is stronger than needed, but economically plausible.

Holzmann (2008) we show a more general case, and argue that this issue is only aggravated in the nonparametric case. The upshot of this dicussion is that the "additive error mean independent of instruments restriction" requires the population to be identical but for an additive shift parameter. This is clearly not a very attractice heterogeneous population, and the effort going into more refined versions of nonparametric IV estimators seems hard to be justified. This seems to raise questions about the usefulness of instruments as a means to correct for endogeneity. However, as we will show below, instruments may still be used in a heterogeneous population, albeit in a different fashion.

Other Developments: Demand has more facets than just obtaining unrestricted estimates of consumers reaction to prices and income. Indeed, economic theory places restrictions - in particular the so called integrability constraints - on demand systems which can be either tested or imposed. Imposing restrictions of consumer demand on nonparametric models has been done in several papers. Tripathi (2003) imposes homogeneity on nonparametric regression models of the form discussed in (2.2). Lewbel and Linton (2004) discuss how to impose homotheticity on a (demand) function. Haag, Hoderlein and Pendakur (2008) show how to impose negative symmetry of the Slutsky matrix. Hall and Yatchew (2008) estimate a nonparametric regression if data on the derivatives are also present.

Less common in applied work than nonparametric estimation are nonparametric testing approaches. Testing integrability constraints dates back at least to the early work of Stone (1954), and has spurned the extensive research on (parametric) flexible functional form demand systems. Nonparametric analysis of some derivative constraints was performed by Stoker (1989) and Härdle, Hildenbrand and Jerison (1991), but none of these has its focus on modelling unobserved heterogeneity. We will discuss recent work on testing below, but first focus on estimation approaches.

3 Linear Models in a Heterogeneous Population - Nonparametric Estimation of the Density of Random Coefficients

We start now by sketching the estimator for the density of random coefficients in the heterogeneous approximate AI model, introduced in Hoderlein, Klemelä and Mammen (2008, henceforth HKM). For simplicity, we focus on the case of a scalar dependent variable, and neglect the dependence on Q_i. Throughout this chapter, we will always assume to have i.i.d. sequence of random vectors (W_i, X_i, β_i), $i = 1, \ldots, n$, with $W_i \in \mathbb{R}$ and $X_i, \beta_i \in \mathbb{R}^d$, where $d = L + 1$, with the following structural relationship between the variables:

$$W_{ii} = \beta_i' X_i, \qquad i = 1, \ldots, n. \tag{3.1}$$

Our goal is to estimate the density of the vector β_i, which we denote by $f_\beta : \mathbb{R}^d \to \mathbb{R}$. The key identification assumption is that X_i and β_i are independent. Note that we require at this point full independence, which may seem a strong assumption. It may be relaxed, however, at the expense of introducing instruments.

To derive the estimator, the data is transformed by dividing it through the norm $\|X_i\|$, so that $S_i = \|X_i\|^{-1} X_i$ and $T_i = \|X_i\|^{-1} Y_i$. The key observation is now that the conditional density of U given S is given by the Radon transform of the density f_β:

$$f_{T|S}(t|s) = Rf_\beta(s, t), \tag{3.2}$$

see again HKM (2008) for details. We construct now an estimator by inverting the operator R much like a matrix. Then,

$$f_\beta(t, s) = R^{-1} f_{T|S}(t|s). \tag{3.3}$$

This, however, is an ill-posed inverse problem. The inverse operator is not smooth, i.e. small changes in the argument may result in big changes in the value. To solve this problem, one has to use a regularized inverse A_h of the Radon transform. We show in HKM (2008) that a regularized inverse is given by the operator A_h, defined by

$$(A_h f_{T|S})(\xi) = \int_{\mathbf{S}_{d-1}} \int_{-\infty}^{\infty} K_h(s^T \xi - t) \frac{f_{TS}(t, s)}{f_S(s)} \, dt d\mu(s), \qquad \xi \in \mathbb{R}^d, \tag{3.4}$$

where μ is the Lebesgue measure on the unit sphere $\mathbf{S}_{d-1}$. The definition of K_h is involved, see HKM (2008). However, its properties will turn out to make it similar to a smoothing kernel. A natural sample counterparts estimator is given by,

$$\hat{f}_\beta(b) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{f}_S(S_i)} K_h \left(S_i^T b - U_i \right), \qquad b \in \mathbf{R}^d, \tag{3.5}$$

and hence this estimator has a structure similar to nonparametric kernel density estimators.

Despite the appealing structure, this estimator suffers from two shortcomings that make this apporach unattractive for many applications: 1. It requires strong assumptions on the distribution of regressors. If a random intercept is included, the assumptions effectively rule out joint normality of regressors. While it may be the case that violations of this assumption do not seriously affect the results in applications, this issue remains not clear, and obviously a concern in applications. 2. The estimator exhibits slow rates of convergence, and the curse of dimensionality is twice as bad as it is standard nonparametric density estimation (in terms of rates, $n^{-1/2d+3}$ instead of $n^{-1/d+4}$). This is not a shortcoming of the specific estimator we propose. Indeed, in HKM (2008) we show that the rate is the optimal one for this nonparametric problem. Translated into applications, this means that it is virtually impossible to estimate any high dimensional regression without further assumptions. In HKM (2008), we give examples of such assumptions, which are similar in spirit to partially linear modeling. However, in economic terms these specifications mean that, say, the income effect varies across the population, but the price effect stays invariant across the population. This is clearly restrictive, and further research is required on how to combine these method with semiparmetric structure so as to make it applicable with finite amounts of data.

Testing economic restrictions is also not always straightforward. Suppose that the restriction we want to test is whether $R(\beta_i) < 0$, for all $i = 1, .., n$. Essentially, it requires to check whether the density $f_{R(\beta)}$ is positive on a subset of $\mathbb{R}_0^+$ with positive probability. One could employ, e.g. a L_2 distance test, but no formal model has as of yet been proposed.

A final issue that remains to be discussed is the treatment of endogeneity. In HKM (2008), we specified the relationship between endogenous regressors and instruments to be linear without variation in parameters (i.e., random coefficients). This is clearly restrictive, and more general forms of dependence in a heterogeneous population are being called for. In the following section we discuss less restrictive specifications of the functional form of all relationships in the system. But the drawback of such approaches is that we loose identification in general, and are only able to work of the consequences.

4 Nonlinear Models in a Heterogeneous Population - the Implications for Testing Rationality Restrictions

Moving away from linearity makes the model both less restrictive, and harder to identify and estimate. If we do not want to assume that the heterogeneous population is linear from the outset, and we want to allow for several types of individuals, we could consider a heterogeneous population consisiting of two types of individuals, one linear in coefficients like model (3.1), one with a nonlinear, but parametric, both with (finite) parameters that vary across the population, may be formalized as

$$W = \mathbf{1}\{A_0 > 0\}\, X'A_1 + \mathbf{1}\{A_0 \leq 0\}\, g(X, A_2), \tag{4.1}$$

for a known function g and random vector $A = (A_0, A_1', A_2')'$ of parameters that vary across the population (here we have set $A_1 = \beta$). In this model, f_A is not identified. More generally, there may be infinitely many types, and the parameters may be infinitely dimensional, and hence we formalize the heterogeneous population as $W = \phi(X, A)$, for a general mapping ϕ. Still, for any fixed value of A, say a_0, we obtain a demand function having standard properties. The hope is now that when averaging over the unobserved heterogeneity A, rationality properties of individual demand may still be preserved by some structure.

Discussion of this type of models has a long tradition in comsumer demand, or rather, in the econometrics literature that is bordering consumer demand. Models of the form $W = \phi(X, A)$ havebeen introduced by Röhrig (1988). However, he considered these models under the added restriction of monotonicity in a scalar unobservable. This line of research was popularized and greatly expended by the work of Matzkin (see Brown and Matzkin (1996), Matzkin (2003) and Altonji and Matzkin (2005)), but also other authors Chesher (2003) and Imbens and Newey (2008). The predominat theme in this literature is the monotonicity assumption already mentioned. In a nutshell, what is required is to make assumptions that allow to link heterogeneity in responses as given for instance in the various quantiles to unobservables. In a system of equations, which is of course the situation typically encountered in demand, monotonicity is usually enriched by the assumption of triangularity, which means that the first equation is monotonic in exactly one scalar unobservable, the second equation has two scalar unobservable, where the first is the same as in the first equation, and the function is monotonic in the second equation, the third equation has three unobservables, etc. While this model might be sensible in sequential decision making, it is arguably ad hoc in consumer demand, where all equations are driven by the same infinite dimensional unobservable, namely the individual's preference ordering.

Hence, in Hoderlein (2002, 2008), we have argued in favor of thinking of A in $W = \phi(X, A)$ as infinite dimensional random element that varies across the population. As already mentioned above, this model is not identified, not even by exploiting all distributional information in the data, which is given, e.g., through regression quantiles. Indeed, in Hoderlein and Mammen (2007, 2009), we show that in this setup the partial derivative of the conditional α-quantile of Y given $X = x$, denoted $k_\alpha(x)$, with respect to the first component ∂_{x_1}, is related to the underlying heterogeneous population ϕ through

$$\partial_{x_1} k_\alpha(x) = \mathbb{E}[\partial_{x_1}\phi(X, A)|X = x, Y = k_\alpha(x)],$$

provided $A \perp X$. In words, even the derivative of the conditonal quantile is a local average over a heterogeneous population. The same is actually true for the conditional mean

$$\partial_{x_1} m(x) = \mathbb{E}[\partial_{x_1}\phi(X, A)|X = x],$$

under essentially the same assumptions. While in particular the latter result is straightforward, this approach has several advantages and features.

1. It provides a direct link between the unobservable world of the nonlinear heterogeneous population and features of the joint distribution.

2. Since averaging is a transformation that preserves many features of the unnderlying heterogeneous population ("the ϕ world") remain preserved, even though ϕ is not identified.

3. Endogeneneity can be handled in a control function fashion. Regressions remain useful toold for analyzing data.

4. Both conditional means and quantiles convey information about the underlying heterogeneous population. It is not the case that one is the structural model, while the other is useless. In fact, none of them is the structural model, but by averaging certain features of the observable distribution of the data inherit properties of the structural model.

5. Sometimes the objects that inherit features and the strutural objects in the underlying heterogeneous population differ.

Let us elaborate on some of these points: While the first two points are obvious, point 3. is the first that need some elaboration. Indeed, what we show in Hoderlein (2005, 2008) and in Hoderlein and Mammen (2007, 2009) is that a control function IV structure generalizes all the way to the nonseparable nonmonotonic setup. Indeed the system of equations is defined through

$$W = \phi(X, \tau(Q, A)) \tag{4.2}$$

$$X = \vartheta(Z, Q, U) \tag{4.3}$$

where ϕ is a fixed R^L-valued Borel mapping defined on the sets $\mathcal{X} \times \mathcal{V}$ of possible values of (X, V). Analogously, μ is a fixed R^{L+1}-valued Borel mapping defined on the sets $\mathcal{S} \times \mathcal{Q} \times \mathcal{U}$ of possible values of (S, Q, U). Moreover, ϑ has to be invertible in U. If we let $S = (Q', U')'$, then we require that

$$F_{A|S,Z} = F_{A|S} \tag{4.4}$$

This assumption is the key identification assumption and discussed in detail in Hoderlein (2005, 2008). Assume for a moment all regressors were exogenous, i.e. $Z \equiv X$ and $U \equiv 0$. Then this assumption states that X, in our case total expenditure and prices, and unobserved heterogeneity are independently distributed, conditional on individual attributes, and control function residuals. If there is endogeneity, we would usually specify ϑ to be additive in U, e.g., $\vartheta(Z, Q, U) = \psi(Z, Q) + U$, where $\mathbb{E}\left[U|Z, Q\right] = 0$. In this case, we can think of ϑ as separating X into an exogenous part $\psi(Z, Q)$ and an endogenous U. Once we condition on U, we X is independent of A. The upshot of this discussion is that even in the case of endogeneity a modified version of a mean or quantile regression including control function residuals helps to identify the object of interestest, $\mathbb{E}[\partial_{x_1}\phi(X, A)|X, Q, U]$, which is an best unbiased approximation of the effect of interest $\partial_{x_1}\phi$.

To see point 4,.we note that already

$$\partial_{x_1} k_\alpha(x) = \mathbb{E}[\partial_{x_1}\phi(X, A)|X = x, Y = k_\alpha(x)],$$

and

$$\partial_{x_1} m(x) = \mathbb{E}[\partial_{x_1}\phi(X, A)|X = x],$$

defined two different objects of interest. Both of these might be of interest, as are more elaborate projections in the case when we have instruments and control function residuals, see Hoderlein (2005, 2008) and Hoderlein and Mammen (2007, 2009).

To see point 5, consider the Slutsky matrix, i.e. the matrix of utiliy compensated price deriavtives. In the form usually considered in the demand literature, this is in the underling heterogeneous population (defined by ϕ, x, and v), takes the form

$$\mathfrak{S}(x,v) = D_p\phi(x,v) + \partial_y\phi(x,v)\phi(x,v)' + \phi(x,v)\phi(x,v)' - diag\left\{\phi(x,v)\right\}.$$

Here, $diag\left\{\phi\right\}$ denotes the matrix having the ϕ_j, $j = 1,..,L$ on the diagonal and zero off the diagonal. To discuss this object, we need again some notation. Let $\mathbb{V}\left[G,H|\mathcal{F}\right]$ denote the conditional covariance matrix between two random vectors G and H, conditional on some σ-algebra $\mathcal{F}$, and $\mathbb{V}\left[H|\mathcal{F}\right]$ be the conditional covariance matrix of a random vector H. We will also make use of the second moment regressions, i.e. $m_2(\xi,q) = \mathbb{E}[WW'|X=\xi, Q=q]$. We also abbreviate negative semidefiniteness by *nsd*. Finally, for any square matrix B, let $\overline{B} = B + B'$. The following statement replicates results in Lewbel (2001) and Hoderlein (2002, 2008).

Theorem 1: *Let all the variables and functions be defined as above. Let $A \perp X|Q$. Then, the following implications hold almost surely:*

$\mathfrak{S}$ *nsd* $\Rightarrow \overline{D_p m}(X,Q) + \partial_y m_2(X,Q) + 2\left(m_2(X,Q) - diag\left\{m(X,Q)\right\}\right)$ *nsd*,

However, *if and only if* $\mathbb{V}\left[\partial_y\phi,\phi|X,Q\right]$ *is symmetric we have*

$\mathfrak{S}$ *symmetric* $\Rightarrow\ D_p m(X,Q) + \partial_y m(X,Q)m(X,Q)'$ *symmetric*,

almost surely. For a proof, see Hoderlein (2008).

The importance of this proposition lies in the fact that it allows for testing the key elements of rationality without having to specify the functional form of the individual demand functions or their distribution in a heterogeneous population. Indeed, the only element that has to be specified correctly is the conditional independence assumption, in this case $A \perp X|Q$, which is implied by $A \perp (X,Q)$ and corresponds to a completely exogenous setting. But, as already mentioned above there is a variety of independence conditions that in general lead to different structures, see again Hoderlein (2008). Testing can be pointwise, at a fixed position (x_0, q_0), see Hoderlein (2002, 2008), or integrated across the population, see Haag, Hoderlein and Pendakur (2008), depending on whether one aims at a statement for the population in total, or whether one focusses on subpopulations, see the discussion in the next section.

The fifth point is also well illustrated by theorem 1. The structure that inherits negative semidefniteness, i.e. $\overline{D_p m}(X,Q) + \partial_y m_2(X,Q) + 2\left(m_2(X,Q) - diag\left\{m(X,Q)\right\}\right)$ differs from the one in the underlying heterogeneous population $\mathfrak{S}(X,V)$. Lewbel (2001) was the first to point out that if we, in order to check for negative semidefiniteness, replicate $\mathfrak{S}(X,V)$ by considering $\mathfrak{S}(X,Q)$, we require $\mathbb{V}\left[\partial_y \phi, \phi | X, Q\right]$ to be negative semidefinite as well. However, this assumption is not testable and simply has to be assumed to hold.

5 Nonparametric Tests of Rationality Restrictions

A final issue that deserves mentioning from the perspective of a heterogeneous population is how to test economic restrictions. Take the issue of symmetry: For instance, under assumptions stated above, the null hypothesis is that the matrix $S(x,q) = D_p m(x,q) + \partial_y m(x,q) m(x,q)'$ is symmetric at every (x_0, q_0). This can be tested in several ways. We would like to mention two: The first is pointwise, i.e. one picks a “representative grid” of positions at which to evaluate the truth of the hypthesis, and use standard pointwise asymptotics to derive the behavior of the test.

There are several problems with such an approach: 1, How does one select the grid? An obvious answer is to draw from the empirical cdf of the data, by repeatedly drawing with replacement. 2. There is a problem of testing multiple hypothesis. Indeed, in five percent of the cases, we would be bound to reject even a valid null hypothesis. There is a fix for this, see Hoderlein (2008), but the more fundamental problem of the axiomatic foundation of such an apporach remains unresolved. 3. The individual tests have low power, due to the pointwise and nonparametric character. 4. One always tests implications only, but this is a process of information loss. It may well be the case that, at every point (x_0, q_0), all individuals violate symmetry, but in the conditional average these violations cancel. 5. The elements of the asymptotic distribution are difficult to estimate, and hence a bootstrap procedure is being called for. In Hoderlein (2008) we argue that the bootstrap is consistent.

An alternative are more elaborate nonparametric tests, in particular so called L_2- distance tests. To illustrate how these tests are constructed, ignore again for simplicity the dependence of S on q, and let $s^{jk}(x)$ denote the (j,k)-th element of $S(x)$. The null hypothesis is then that

$$H_0 \colon \mathbb{P}(s^{jk}(X) = s^{kj}(X), \forall\, j \neq k) = 1$$

and the alternative is that there is at least one pair (j,k) with $s^{jk}(x) \neq s^{kj}(x)$ over a significant range. We may express the alternative as $H_1\colon \mathbb{P}(s^{jk}(X) = s^{kj}(X), \forall\ j \neq k) < 1$. The null hypothesis is equivalent to the condition that the L_2-distance of these functions is zero. Using a nonnegative and bounded weighting function $a(x)$ this can be written as

$$\Gamma_S = \mathbb{E}\Big(\sum_{j<k}(s^{jk}(X) - s^{kj}(X))^2 a(X)\Big) = 0.$$

Of course this is only equivalent to the null hypothesis if $a(X)$ is nonzero over the whole support of X. However the weighting function allows to restrict the test to certain regions of the explanatory variables. Secondly, assuming that the support of $a(X)$ is strictly contained in the support of the density, boundary problems can be avoided.

A nonparametric test statistic may be constructed by the analogy principle. Using estimators $\partial_k \hat{m}_h^j(x)$ for the partial derivatives of the regression functions $\partial_k m_j(x)$, are used as estimators for the derivatives. A possible test statistic is then

$$\hat{\Gamma}_S = \frac{1}{n}\sum_{j=1}^{M-2}\sum_{k=j+1}^{M-1}\sum_{i=1}^{n}(\partial_k \hat{m}_h^j(X_i) + \hat{m}_h^k(X_i)\partial_x \hat{m}_h^j(X_i) - \partial_j \hat{m}_h^k(X_i) - \hat{m}_h^j(X_i)\partial_x \hat{m}_h^k(X_i))^2 A_i. \quad (5.1)$$

The large sample behavior of this test statistic is derived in Haag, Hoderlein and Pendakur (2008), but is even more cumbersome to estimate. As such, the 5th problem above is not alleviated by such an approach. However, the bootstrap works indeed well, and is recommended to applied reserachers. This is particularly the case, because it is well known that this is an instance where asymptotic normality is a poor approximation to the true finite sample behavior of the test statistic.

How about the other four problems of the pointwise test. Obviously, points 1 and 2 disappear. Also, the third point is mitigated, as the power is dramatically improved by considering the average. However, there is an added difficulty now. Because we are aggregating over the (observable!) X dimension we may actually mix up areas where people are rational with those where they are not. As such, we may reject rationality if we found that the two functions $s^{jk}(x)$ and $s^{kj}(x)$ differ only for a subset of the heterogeneous population. This problem, together with the fourth issue mentioned above may result in potentially misleading answers: We may reject rationality even if it is only relatively few individuals who are actually not rational.

6 Conclusion

In this chapter we have discussed the implications of the recently emphasized issue of preference heterogeneity for state of the art econometric modeling. We have shown that some well establsihed approaches may have difficulties, and that there are potentially new approaches that may reveal new and interesting features of the data. Many of these approaches, like nonparametric random coefficinet models are still in their early stages, and a number of problems are to be resolved before they become meianstream tools for applied econometrics.

References

[1] ACKERBERG, D., BENKARD, L., BERRY, S., and A.PAKES, (2009), Econometric Tools for Analyzing Market Outcomes, Chapter 63 in *Handbook of Econometrics*, vol. 6A, James J. Heckman and Ed Leamer, eds. North-Holland Press.

[2] ALTONJI, J., and R.MATZKIN, (2005); Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors, *Econometrica*, 73, 1053 - 1103.

[3] BLUNDELL, R., BROWNING, M., and I. CRAWFORD (2003); Nonparametric Engel Curves and Revealed Preference, *Econometrica*, 71, 205-240.

[4] BLUNDELL, R., CHEN, X., and D. KRISTENSEN (2007); Semi-Nonparametric Estimation of Shape Invariant Engel Curves under Endogeneity, *Econometrica*, forthcoming.

[5] BLUNDELL, R., A. DUNCAN and K. PENDAKUR (1998); "Semiparametric Estimation of Consumer Demand", Journal of Applied Econometrics 13, pp 435-461.

[6] BROWN, D. and R.MATZKIN (1996), "Estimation of Nonparametric Functions in Simultaneous Equation Models, with an Application to Consumer Demand", *Working Paper, Northwestern University.*

[7] CHESHER, A., (2003), "Identification in Nonseparable Models", *Econometrica,* 71, 1405-1443.

[8] CHERCHYE, L., CRAWFORD, I., De ROCK, B. and F. VERMEULEN (2009), " The Revealed Preference Approach to Demand".

[9] DAROLLES, S., FLORENS, J.-P., and E. RENAULT (2003), "Nonparametric Instrumental Regression", *Working Paper, IDEI.*

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

[11] HAAG, B. and HODERLEIN, S. (2006); Nonparametric Testing Zero Homogeneity, Working Paper, Brown University.

[12] HÄRDLE, W., HILDENBRAND, W., and M. JERISON (1991); Empirical Evidence on the Law of Demand, *Econometrica,* 59, 1525-1549.

[13] HILDENBRAND, W. (1993), "Market Demand: Theory and Empirical Evidence", Princeton, Princeton University Press.

[14] HODERLEIN, S., (2002): "Econometric Modelling of Heterogeneous Consumer Behaviour - Theory, Empirical Evidence and Aggregate Implications," *PhD Thesis*, *LSE.*

[15] HODERLEIN, S. (2008): "How many Consumers are Rational? " *Working Paper, Mannheim.*

[16] HODERLEIN, S., and E. MAMMEN (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 1513 - 1518.

[17] IMBENS, G. and W. NEWEY (2003), "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity", *Working Paper, MIT.*

[18] 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.

[19] KIHLSTROM, R.; A. MAS-COLELL and H. SONNENSCHEIN (1976); The Demand Theory of the Weak Axiom of Revealed Preference, *Econometrica*, 44, 971-978.

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

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

[22] MAS-COLELL, A., M. WHINSTON and J. GREEN (1996), "*Microeconomic Theory* ", Oxford University Press.

[23] MATZKIN, R. (2003), "Nonparametric Estimation of Nonadditive Random Functions", *Econometrica*, 71, 1339-1376

[24] PENDAKUR, K. (1999), "Estimates and Tests of Base-Independent Equivalence Scales", Journal of Econometrics 88(1), pp 1-40.

[25] QUAH, J. (2000), Weak Axiomatic Demand Theory. Working Paper, Nuffield College, Oxford.

[26] ROEHRIG, C., "Conditions for Identification in Nonparametric and Parametric Models", *Econometrica,* 56, 433-447

[27] STOKER, T. (1989); Tests of Additive Derivative Constraints, *Review of Economic Studies,* 56, 535-552.