How Many Types Are There?

Ian Crawford
University of Oxford and IFS

Krishna Pendakur
Simon Fraser University

Abstract

We suggest an elementary nonparametric (revealed preference) method which will partition cross-section consumption data into the *minimal* number of preference groups such that all of the data are rationalisable by standard utility theory. This provides a simple, theory-driven way of investigating unobserved preference heterogeneity in consumption microdata, and easily extends to any choice model which has a nonparametric (revealed preference) characterisation. In a 500 observation cross-sectional dataset on milk demand, we find that not much unobserved heterogeneity is needed: at least 4 and not more than 5 utility functions are required to fully explain all observed variation in behaviour.

Key Words: Unobserved heterogeneity, revealed preference.

JEL Classification: C43, D11.

Acknowledgements: Pendakur acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada. We are very grateful Richard Blundell, Martin Browning, Andrew Chesher, Simon Sokbae Lee, Arthur Lewbel and J. Peter Neary for helpful comments. Remaining problems are translation errors between British and Canadian English.

1 Unobserved Heterogeneity in Microdata

One of the most striking features of consumer microdata is the great heterogeneity in choice behaviour which is evident, even amongst economic agents which are otherwise similar in all observable respects. This presents researchers with a difficult problem - how to model behaviour in a way which accommodates this heterogeneity and yet preserves theoretical consistency and tractability.

One rather robust response is to demand that everything should be explainable by the theory in terms of observables alone. This view is typified by Becker and Stigler (1977):

> "Tastes neither change capriciously nor differ importantly between people." G. Becker and G. Stigler *De Gustibus Non Est Disputandum,* AER, 1977

The research agenda which follows from this view is one which tries to explain differences in observed behaviour without recourse to unobserved heterogeneity in tastes, but instead purely in terms of the theory and observable differences in constraints, characteristics of market goods and characteristics of agents. From this point of view, resorting to unobserved preference heterogeneity in order to rationalise behaviour is a cop-out; it is an admission of failure on the part of the theory.

From this perspective it is therefore a matter for some regret that measures of fit in applied work on microdata are typically very low - that is, the theory performs poorly. As a result, the belief that unobserved heterogeneity is an inescapable and essential part of the modeling problem has become the dominant view in the profession. This approach was summarised by the joint 2000 Nobel laureates as follows.

> "In the 1960's, rapidly increasing availability of survey data on individual behavior ... focussed attention on the variations in demand across individuals. It became important to explain these variations as part of consumer theory, rather than as *ad hoc* disturbances". D. McFadden, *Nobel Lecture*

> "Research in microeconometrics demonstrated that it was necessary to be careful in accounting for the sources of manifest differences among apparently similar individuals. ... This heterogeneity has profound consequences for economic theory and for econometric practice." J. Heckman, *Nobel Lecture*

In applied microeconometrics, the standard approach has been to pool data and to model the behaviour of individual economic agents as a combination of a common component and an idiosyncratic component which reflects unobserved heterogeneity. In its least sophisticated form, this amounts to interpreting additive error terms as unobserved preference heterogeneity parameters. Recently, it has become clear that such an approach typically requires a combination of assumptions on the functional form of the statistical model and the distribution of unobservable heterogeneity. Contributions here include McElroy (1987), Brown and Walker (1989), Lewbel (2001) and Lewbel and Pendakur (2008). Broadly, the current consensus is on unobserved heterogeneity is that: it is a fundamental feature of microdata; if neglected, it makes econometric estimation and identification difficult; and it is rather hard to deal with convincingly, especially in models where heterogeneity is not additively separable.

Whilst the dominant empirical methods, by and large, proceed by *pooling* agents, the approach which we develop here is based on *partitioning*. We work from the basis of revealed preference (RP) restrictions (developed first in Afriat 1967, Diewert 1973 and Varian 1982). At heart, revealed preference restrictions are sets of inequality restrictions on observables (prices and demands), which provide necessary and sufficient conditions for the existence of an unobservable (a well behaved utility function representing the consumer's preferences) which rationalises the data. RP restrictions are usually applied to longitudinal data on individual consumers and are used to check for the existence and stability of well-behaved preferences. In this paper we apply this kind of test to pooled data on many different consumers (though, as we describe below, our idea applies to many contexts with optimising agents). In this context, RP restrictions are interpretable as a check for the *commonality* of well-behaved preferences.[1]

Of course, this is a rather simplistic idea. The very notion that such a check might pass and that the choices of all of the consumers in a large microeconomic dataset could be explained perfectly by a single common utility function is, as Lewbel (2001) points out, "implausibly restrictive". The real problem is what to do if (or more likely when) the data do not satisfy the RP restrictions. Dean and Martin (2010) provide one type of solution: they show how to find the largest subset of the data that *do* satisfy (some of) the RP restrictions. However, their approach leaves some of the data as unexplained by the optimising model.

The contribution of this paper is to provide a different (and complementary) set of strategies

[1] Our methods can easily be extended to cover panel data contexts. We discuss this in section 7.3 of the Appendix.

for the case where the pooled data violate the RP restrictions. Here, some amount of preference heterogeneity is necessary in order to model those data—we need more than just one utility function. The question is *how many do we need*? Is it very many (perhaps as many as there are observations), or just a few? This paper shows how to find out the minimum number of types (utility functions) neccessary to fully explain all observed choices in a data set.

In seeking the *minimum* number of utility functions necessary to rationalise behaviour, we keep with Friedman's (1957) assertion that we don't want the true model, which may be unfathomably complex; rather, we want the simplest model that is not rejected by the data. Occam's Razor applies here: we know that we can fully explain behaviour with a model in which every agent is arbitrarily different from every other, but that model is not useful for modeling or predicting behaviour. Instead, we want to pool agents to the maximum possible degree that is consistent with our theoretical model. If the minimum number of types (utility functions) is very large relative to the number of observations, then modeling strategies with a continuum of types, or with one type for each agent (such as fixed effects models), might be appropriate. In contrast, if the minimum number of types is small relative to the number of observations, then modeling strategies with a small number of discrete types, such as those found in macro-labour, education choice, and empirical marketing models, might be appropriate.

We argue that our approach offers three benefits which may complement the standard approaches to unobserved heterogeneity in empirical work. Firstly, it provides a framework for dealing with heterogeneity which is driven by an economic model of interest. Secondly, it is elementary: our approach does not require statements about the empirical distributions of objects we can't observe or functional structures about which economic theory is silent. This contrasts with the standard approach of specifying *a priori* both the distribution of unobserved preference heterogeneity parameters and its functional relationship with observables. Thirdly, it provides a practical method of partitioning data so that the observations in each group are fully theory-consistent. This contrasts with approaches wherein only part of the model (the part which excludes the unobserved heterogeneity) satisfies the theory.

We implement our strategy with a 500 observation cross-sectional dataset on Danish milk demand. These data have individual-level price, quantity and product characteristics information, and so are ideal for the application of RP methods. We find that at least 4, and not more than 5, types (utility functions) are needed to completely explain all the observed variation in consumption behaviour.

The paper is organised as follows. We begin with a description of the cross-sectional data on household expenditures and demographics which we use in this study. We then investigate whether these data might be rationalised by partitioning on observables which form the standard controls in microeconometric models of spending patterns. We then set out a simple method for partitioning on unobservables, and consider whether the results from these partitioning exercises can be a useful input to econometric modelling of the data. The final section draws some conclusions.

2 The Data

In this paper we focus on the issue of rationalising cross-sectional household-level data on spending patterns with the standard static utility maximisation model of rational consumer choice. This approach can readily be extended to other more exotic economic models which have a nonparametric/revealed preference characterisation (examples are given in the discussion and in section 7.4 of the Appendix). The data we use are on 500 Danish households and their purchases of six different types of milk. These households comprise all types ranging from young singles to couples with children to elderly couples. The sample is from a survey of households which is representative of the Danish population. Each household keeps a strict record of the price paid and the quantity purchased as well as the characteristics of the product. We aggregate the milk records to a monthly level, partly to minimise the computational burden and partly to allow us to treat milk as a non-durable, non-storable good, so that the intertemporally separable model which we are invoking is appropriate. An attractive feature of these data is that there is variation in prices in the cross section which is not due to unobserved differences in product qualities.

Descriptive statistics are given in Table 1. In what follows let $I = \{i : i = 1, ..., 500\}$ denote the index set for these observations and let $\{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$ denote the price-quantity data. We will also make use of a list of observable characteristics of each household and these are represented by the vectors $\{\mathbf{z}_i\}_{i \in I}$.

Table 1: Descriptive Statistics				
	Mean	Min	Max	Std. Dev
$\{\mathbf{w}_i\}_{i=1,..,500}$	**Budget Shares**			
Conventional Full Fat	0.1688	0	1.0000	0.3158
Conventional Semi-skimmed	0.4255	0	1.0000	0.4102
Conventional Skimmed	0.1521	0	1.0000	0.2932
Organic Full Fat	0.0374	0	1.0000	0.1394
Organic Semi-skimmed	0.0977	0	1.0000	0.2237
Organic Skimmed	0.1185	0	0.9951	0.2669
	Total Expenditure (DK)			
Total Expenditure	66.1986	4.8222	345.1279	58.5765
$\{\mathbf{p}_i\}_{i=1,..,500}$	**Prices** $(DK\ litre)$			
Conventional Full Fat	6.1507	3.3068	11.3289	0.4652
Conventional Semi-skimmed	5.4104	4.0919	7.9567	0.4304
Conventional Skimmed	5.1524	4.1619	6.2075	0.1814
Organic Full Fat	7.3335	6.1188	8.6597	0.1860
Organic Semi-skimmed	6.4968	5.0565	8.5374	0.2187
Organic Skimmed	6.2679	5.5312	7.9684	0.1501
$\{\mathbf{z}_i\}_{i=1,..,500}$	**Demographics**			
Singles $\{0,1\}$	0.3260	0	1.0000	0.4692
Singles Parents $\{0,1\}$	0.0420	0	1.0000	0.2008
Couples $\{0,1\}$	0.3500	0	1.0000	0.4774
Couples with children $\{0,1\}$	0.2300	0	1.0000	0.4213
Multi-adult $\{0,1\}$	0.0520	0	1.0000	0.2222
Age $(Years)$)	47.8600	18.0000	87.0000	15.5240
Male HoH $\{0,1\}$	0.92	0	1.0000	0.27156

Given these data the question of interest is whether it is possible to rationalise them with the canonical utility maximisation model. The classic result on this issue is provided by *Afriat's Theorem* (see especially Afriat (1967), Diewert (1973) and Varian (1982, 1983)). Afriat's Theorem shows that the generalised axiom of revealed preference (GARP) is a necessary and sufficient condition for the existence of a well-behaved utility function $u(\mathbf{q})$ which exactly rationalises the data in the sense that $u(\mathbf{q}_i) \geq u(\mathbf{q})$ for all $\mathbf{q}$ such that $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}$. That is, if $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j \Leftrightarrow \mathbf{q}_i R^0 \mathbf{q}_j$ denotes a direct revealed preference relation and R is the transitive closure of R^0 then GARP is defined by the restriction that $\mathbf{q}_i R \mathbf{q}_m \Rightarrow \mathbf{p}_i'\mathbf{q}_i \leq \mathbf{p}_i'\mathbf{q}_m$. If observed demands $\{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$ satisfy these GARP inequalities, then there is a single utility function (preference map) that can rationalise all observed demands. If not, then there is not.[2]

We checked the data for consistency with GARP and it failed. No single utility function exists which can explain the choices of all of these households—Lewbel's (2001) warning seems to be justified. So we now turn to the question: *how many well-behaved utility functions are required to rationalise these price-quantity microdata?* Obviously 500 utility functions, each one rationalising each observation,will be over-sufficient. The next two section explore the idea of conditioning on observables and unobservables in order to find a *minimal* necessary partition of these data.

3 Partitioning on Observables

The aim of this section is to use nonparametric (revealed preference) methods to partition the data into a mutually exclusive, exhaustive set of types such that the preferences of all of the households of a given type can be represented by a single utility function. This involves simply stratifying the data according to observables and running RP tests within groups, then refining the stratification on observables until the RP test for commonality of preferences within types is satisfied. To this end we first stratified the data into groups according to household structure and created 5 groups: single person household, single parents, couples, couples with children and multi-adult households. The partition was insufficient to be able to rationalise the data within these groups.

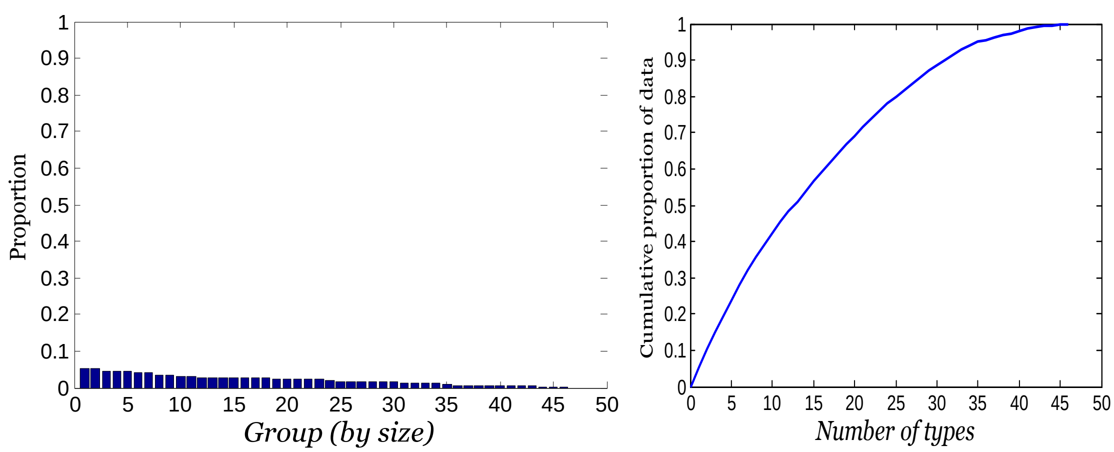
We next ordered the household of each structure by the age of the head of household and, beginning with the youngest, we sequentially tested the RP condition in order to see whether we could rationalise behaviour by a further partition on age into contiguous bands. This too proved impossible because there were instances of households with the same structure whose heads of household were the same age whose behaviour was not mutually rationalisable. Having first split by household structure, and then split by age and not yet found a rationalisation

[2]This discussion assumes that consumers are perfect optimisers. Our methods can easily be adapted to the situation where they are imperfect optimisers, see Section 7.2 in the Appendix.

for the data we further split by region (there were nine regional indicators in the data). This too failed to rationalise the data as there were instances of households with identical structure and age living in the same region who were irreconcilable with a common utility function. We then looked at the gender of the household head. This, finally, produced a rationalisation of the data with 46 types defined by household structure/age/region/gender.

The left hand panel of Figure 1 shows the distribution of group sizes with the groups ordered largest to smallest. This shows that the largest groups consist of approximately 5% of the data (there are two such groups) whilst the smallest (the 44th, 45th and 46th on the left of the histogram) consist of singletons. The right hand panel of Figure 1 shows the cumulative proportion of the data explained by the rising numbers of types. The first ordinate shows that approximately 5% of the data are rationalisable by one type (the most numerous) and approximation 10% by two most numerous types. Ten types are needed to rationalise half the data.

FIGURE 1. Partitioning on observables



We draw the following lessons from this exercise. Firstly the resulting partition is dependent upon the order in which one takes the observables - the method is path dependent. Secondly the partition is not parsimonious - each type only accounts for around 2% of the data on average.

4 Partitioning on Unobservables

In this section we consider partitioning on unobservables. As before we are interested in trying to split the data into as few (and as large) groups as we can such that all of the households within each group can be modelled as having a common well-behaved utility function. However this time we will not use observables like those used above to guide/constrain us. The simplest "brute force" approach would be to check the RP restrictions within all of the possible subsets of the data and retain those which form the minimal exclusive exhaustive partitions of the data. This is computationally infeasible as there are 2^{500} such subsets. Instead we have designed two simple algorithms which will provide two-sided bounds on the minimal number of types in the data.The details of the algorithms need not detain us here (they are described in section 7.1 of the Appendix).

We ran the algorithm on our data and found that the minimal number of types was between 4 and 5. That is one needs at least 4, and not more than 5, utility functions to completely rationalise all the observed variation in choice behaviour observed in these data in terms of income and substitution effects. For our upper bound of 5 types our algorithm also delivers a partition of the data into the groups, such that within-groups a single utility function is sufficient to rationalise all the observed behaviour. Table 2 gives the average budget shares for each group delivered by our upper bound algorithm and Figure 2 shows the distribution of types and gives same information as Figure 1 on the same scale for ease of comparison and in order to show how parsimonious this partition is in comparison.

FIGURE 2. Partitioning on unobservables

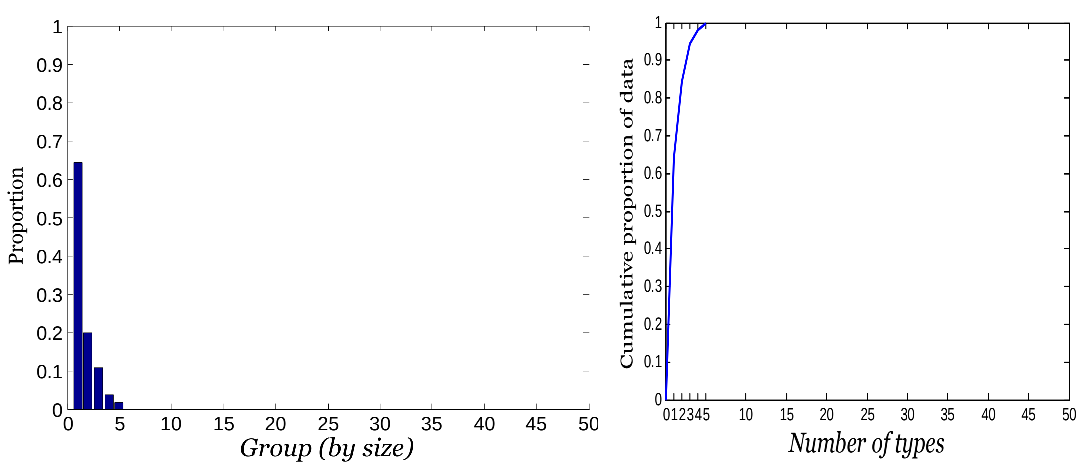


Table 2: Average Budget Shares Across Types		Conventional Milk			Organic Milk		
Sample Means	Group N	Full-fat	Semi	Skim	Full-fat	Semi	Skim
pooled	500	0.168	0.425	0.152	0.037	0.097	0.118
Type 1	321	0.160	0.496	0.143	0.024	0.075	0.100
Type 2	100	0.155	0.285	0.205	0.070	0.121	0.162
Type 3	53	0.239	0.351	0.074	0.044	0.144	0.147
Type 4	18	0.134	0.256	0.148	0.032	0.258	0.170
Type 5	8	0.292	0.195	0.357	0.128	0.017	0.009

Our expectation was that the observable characteristics of households would be the crucial determinants of type-membership. However, a multinomial logit model of group membership conditional on demographic characteristics (age and sex of household head, number of members, number of children, geographic location, etc) has a (McFadden unadjusted) pseudo-R^2 of only 5.4%. That is, observable characteristics of households are essentially uninformative regarding which of the five types to which a household is assigned. The implication here is that, in a framework where we want to use the minimum number of types, unobserved preference heterogeneity is vastly more important than observed demographic heterogeneity.

5 Estimation of Preferences

The incorporation of unobserved preference heterogeneity into demand estimation is a theoretically and econometrically tricky affair. Matzkin (2003, 2007) proposes a variety of models and estimators for this application, all of which involve nonlinearly restricted quantile estimators, and most of which allow for unobserved heterogeneity which has arbitrary (but monotonic) effects on demand. These models are difficult to implement, and, as yet, only Matzkin (2003, 2007) has implemented them. Lewbel and Pendakur (2009) offer an empirical framework that incorporates unobserved preference heterogeneity into demand estimation that is easy to implement, but which requires that unobserved preference parameters act like fixed effects, pushing the entire compensated budget share function up or down by a fixed factor.

Given the difficulty of incorporating unobserved preference heterogeneity beyond a fixed effect, it is instructive to evaluate how our 5 utility functions differ from each other. Since group 5 has only 8 observations assigned to it, we leave it out of this part of the analysis. For the remaining groups we estimate group-specific demand systems. Since we know that, within each type, there exists a single preference map which rationalises all of the data we need not worry about unobserved heterogeneity in our estimation. We know that there is a single integrable demand system which exactly fits the data for each group. The problem we face is that we do not know the specification of that demand system so our main econometric problem is finding the right specification. We take the simplest possible route at this point and estimate a demand system with a flexible functional form - the Quadratic Almost Ideal (QAI) demand system (Banks, Blundell and Lewbel 1997)). The idea is that such a model should be flexible enough to fit the mean well and that the interpretation of the errors is solely specification error[3]. By estimating budget-share equations for each of our four largest groups we characterise what their Engel curves look like and test whether or not including group dummies in budget share equations (as in Lewbel and Pendakur (2009)) is sufficient to absorb the differences across these utility functions.

TABLE 3: QAI Estimates of levels at median constraint

		Conventional Milk			Organic Milk		
Group	Group N	Full-fat	Semi	Skim	Full-fat	Semi	Skim
pooled	500	0.154	0.400	0.163	0.041	0.100	0.0142
		0.018	*0.024*	*0.017*	*0.008*	*0.013*	*0.015*
group 1	321	0.155	0.434	0.173	0.020	0.085	0.133
		0.022	*0.030*	*0.021*	*0.007*	*0.014*	*0.017*
group 2	100	0.153	0.287	0.194	0.089	0.092	0.184
		0.032	*0.044*	*0.041*	*0.021*	*0.028*	*0.032*
group 3	53	0.195	0.330	0.091	0.070	0.130	0.184
		0.055	*0.064*	*0.033*	*0.027*	*0.036*	*0.041*
group 4	18	0.084	0.171	0.295	0.052	0.209	0.190
		0.061	*0.075*	*0.079*	*0.028*	*0.061*	*0.057*

Table 3 gives predicted budget-shares for each group, evaluated at a common constraint defined by the vector of median prices and the median milk expenditure level. (These are

[3]Measurement error is much more cumbersome to consider in a revealed-preference context, so we do not consider it here.

the level coefficients in the QAI regressions for each group, where prices and expenditures are normalised to 1 at the median constraint.) The point estimates differ quite substantially across groups, and a glance at the estimated standard errors shown in *italics* shows that the hypothesis that these point estimates are the same value is heartily rejected.

The estimated linear and quadratic terms are also quite different across group, but these are estimated with rather large standard errors. We do not present these coefficients, but rather focus on tests of whether or not the coefficients are the same across groups. The joint hypothesis that all these groups have the same linear and quadratic terms is rather weakly rejected—the sample value of the test statistic is 45.4, and under the Null it is distributed as a χ^2_{30} with a p-value of 3.5%. Individually, only groups 2 and 4 show evidence that they differ from group 1 in terms of the total expenditure responses of budget shares (they test out with p-values of 8% and 1%, respectively).

In contrast, the estimated price responses of budget shares differ greatly across groups. The test that they share the same price responses has a sample value of 382, is distributed under the Null as a χ^2_{45} with a p-value of less than 0.1%. Further, any pairwise test of the hypothesis that two groups share the same price responses rejects at conventional levels of significance.

One can also test the hypothesis the heterogeneity across the types can be aborbed into level effects. Not surprisingly, given that we reject both the hypotheses that total expenditure effects are identical and that price effects are identical, this test is massively rejected. The test statistic has a sample value of 405, and is distributed under the Null as a χ^2_{75} with a p-value of less than 0.1%.

One problem with using the QAI demand system to evaluate the differences across groups is that there is no reason to think that the functional structure imposed by the QAI demand system is true. An alternative approach is to use nonparametric methods. These methods have the advantage of not imposing a particular functional form on the shape of demand. They have the disadvantage of suffering from a severe curse of dimensionality, because in essence one needs to estimate the level of the function at every point in the support of possible budget constraints. The dimensionality problem is that this support grows fast with the number of goods in the demand system. A nonparametric approach that does not suffer from the curse of dimensionality is to try to estimate averages across the support of budget constraints.

In the top panel of Table 4, we present the average over all observed budget constraints of the nonparametric estimate of budget shares for each group. For the nonparametric analysis, we study only the 3 largest groups, totaling 474 observations. For each group, we nonparametrically estimate the budget share function evaluated at each of the 474 budget constraints, and report its average over the 474 values, as well as a simulated standard error in *italics*.[4]

Table 4: Nonparametric Estimates, averaged over all constraints

		Conventional Milk			Organic Milk		
Group	Group N	Full-fat	Semi-fat	Skimmed	Full-fat	Semi-fat	Skimmed
Average Levels							
group 1	321	0.157	0.465	0.158	0.027	0.078	0.115
		0.015	*0.020*	*0.014*	*0.006*	*0.014*	*0.013*
group 2	100	0.168	0.292	0.191	0.074	0.092	0.184
		0.022	*0.035*	*0.029*	*0.013*	*0.015*	*0.028*
group 3	53	0.201	0.362	0.073	0.034	0.128	0.201
		0.056	*0.056*	*0.019*	*0.017*	*0.032*	*0.056*
Average Semi-Elasticity wrt Expenditure							
group 1	321	-0.020	0.013	0.003	-0.001	-0.016	0.022
		0.022	*0.028*	*0.016*	*0.007*	*0.009*	*0.014*
group 2	100	-0.048	0.068	-0.084	-0.010	0.000	0.074
		0.028	*0.061*	*0.040*	*0.013*	*0.025*	*0.050*
group 3	53	0.028	-0.074	-0.029	-0.002	-0.054	0.132
		0.082	*0.094*	*0.027*	*0.017*	*0.035*	*0.099*

The top panel of Table 4 shows average levels that are broadly similar to the sample averages reported in Table 2. However, those reported in Table 4 differ in one important respect:

[4]Nonparametric estimates on each budget constraint (before averaging) are computed following Haag, Hoderlein and Pendakur (2009). It is well-known that average derivative estimators suffer from boundary bias. Although the estimates in Table 4 do not trim near the boundaries, estimates which do trim near the boundaries yield the same conclusions.

Standard errors are simulated via the wild bootstrap using Radamacher bootstrap errors. Nonparametric estimators only suffer from specification error in the small sample. Such error disappears as the sample size gets large. Further, unobserved heterogeneneity need not cause a deviation from the regression line, because such heterogeneity is not necessary after our grouping exercise. Thus, the wild bootstrap, which bases simulations on resamples from an error distribution, is actually an odd fit to the application at hand. An alternative is to resample from budget constraints (rather than from budget shares) to simulate standard errors. These simulated standard errors are much smaller, and make the groups look sharply different from each other in terms of both average levels and average slopes.

whereas those shown in Table 2 are averages across the budget constraints in each group, those reported in Table 4 are averages across the budget constraints of all groups. That is, whereas the sample averages in Table 2 mix the effects of preferences and constraints, the nonparametric estimates in Table 4 hold the budget constraints constant. These numbers suggest that there is a quite a lot of preference heterogeneity. For example, Group 1 and Group 2 have statistically significantly different average budget shares for most types of milk.

Given that unobserved heteregeneity which can be absorbed through level effects can fit into recently proposed models of demand, it is more important to figure out whether or not the slopes of demand functions differ across groups. The bottom panel of Table 4 presents average derivatives with respect to the log of expenditure (that is, the expenditure semi-elasticities of budget share functions), again averaged over the 474 observed budget constraints, with simulated standard errors shown in *italics*.

Clearly, the estimated average derivatives are much more hazily estimated than the average levels. But, one can still distinguish groups 1 and 2: the skimmed conventional milk budget share function of group 2 has a statistically significantly lower (and negative) expenditure response than that of group 1. No other pairwise comparison is statistically significant. However, the restriction that the average derivatives are the same across groups combines 10 t-tests like this, two restrictions for each of the 5 independent equations. One can construct a nonparametric analogue to the the joint Wald test of whether or not the three groups share the same expenditure responses in each of the 6 equations. This test statistic has a sample value of 24.3 and is distributed as a χ^2_{10} with a p-value of 0.7%.[5]

The picture we have of the heterogeneity in the milk data is as follows. First, we can *completely* explain all the variation of observed behaviour with variation in budget constraints and 4 or 5 preference maps (ie. ordinal utility functions). Second, the groupings are not strongly related to observed characteristics of households. That is, the primary heterogeneity here is *un*observed. Third, the groups found by our upper bound algorithm are very different from each other, mainly in terms of how budget shares respond to prices, but also in expenditure responses. That the budget-share equations of the groups differ by more than just level effects suggests that unobserved preference heterogeneity may not act like 'error terms' in regres-

[5] If we use the alternative resampling strategy which provides tighter standard errors (outlined in the previous footnote), then the test that the average derivatives are the same for all 3 groups is rejected in each of the 5 independent equations, and, not surprisingly, rejected for all 5 together.

sion equations, and thus may not fit into models recently proposed to accomodate preference heterogeneity in consumer demand modeling.

6 Discussion

We consider an elementary method of partitioning data so that it can be explained perfectly, and in a way which admits the minimal necessary heterogeneity. We argue that our approach offers three benefits which may complement the more established microeconometrics treatment of unobserved heterogeneity. Firstly it provides a framework in which to study heterogeneity which is driven by the economic model of interest. Secondly it does not require statements about the distributions of objects we can't observe or functional structures about which economic theory is silent. Thirdly it provides a practical method of partitioning data so that the observations in each group are precisely theory-consistent rather than just approximately so. This allows researchers to estimate group-specific demand models without fear of the complications which arise in the presence of unobserved heterogeneity.

Throughout this paper we have focused on consumer data and on the canonical utility maximisation model. This is mainly for expositional reasons and it is important to point out what we are proposing can easily be applied to the analysis of heterogeneity in any microeconomic model of optimising behaviour which admits a RP-type characterisation. This is an increasingly wide class which includes profit maximisation and cost-minimisation models of competitive and monopolistic firms, models of intertemporal choice, habits, choice under uncertainty, collective household behaviour, characteristics models, firm investment as well as special cases of all of these models which embody useful structural restrictions on preferences or technology (e.g. weak separability, homotheticity and latent separability)[6]. To adapt our approach to any of those models, one simply replaces the GARP check in all the algorithms (detailed in section 7.1 of the Appendix) with the appropriate RP check. The point is that our strategy for assessing heterogeneity in the consumer demand framework is in principle applicable to any environment where agents are assumed to be optimising something.

In the empirical illustration we characterise the amount of heterogeneity necessary to com-

[6] Afriat (1967), Diewert (1973), Varian (1982, 1983a, 1983b, 1984), Hanoch and Rothschild (1972), Browning (1989), Crawford (2003, 2007), Bar-Shiva (1992), Cherchye, De Rock and Vermuelen (2007),Browning, Blow and Crawford (2008),

pletely rationalise the observed variation in Danish milk demands. We find that very few types are sufficient to rationalise observed behaviour completely. Our results suggest that Becker and Stigler had it wrong in *De Gustibus ..*: preferences do indeed differ both capriciously and importantly between people. The capriciousness is that although in the three decades since Becker and Stigler's assessment, we have learned much about how to deal with preference heterogeneity that is correlated with observables, it seems that the more important kind of heterogeneity is driven by unobservables. Our results also suggest that models which use a small number of heterogeneous types—such as those found in macro-labour models, education choice models, and a vast number of empirical marketing models—may in fact be dealing with unobserved heterogeneity in a sufficient fashion. In contrast, models like Lewbel and Pendakur (2009), in which unobserved preference heterogeneity is captured by a multidimensional continuum of unobserved parameters could well be overkill.

7 Appendix

7.1 Partitioning Algorithms

Notation: For an arbitrary set A, $\mathcal{P}(A)$ denotes the power set (set of all subsets) of A The number of elements of A is denoted by $|A|$. For arbitrary sets A and B, $A \backslash B$ denotes A minus B: all elements from A that are not in B.

Brute Force Algorithm
Inputs: $I, \{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$. Outputs: N and G.

1. If $\{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$ satisfies RP set $N = 1, G = I$ and goto (6).
2. Set $H = \left\{h : h \in \mathcal{P}(I),\ \{\mathbf{p}_i, \mathbf{q}_i\}_{i \in h} \text{ satisfies } RP\right\}$
3. Set $J = \{j : j \in \mathcal{P}(H),\ j = I\}$
4. $N = \min\{|j| : j \in J\}$.
5. Set $G = \{j : j \in J,\ |j| = N\}$.
6. End.

The outputs of the "Brute Force" algorithm are $N =$ the number of types and $G =$ a set containing all of the exclusive and exhaustive partitions of the data into N subsets such that the data in each type satisfy RP. This algorithm works by simply enumerating all of the subsets of the data, chekcing RP conditions within those subsets and then finding the minimal partition based on those subsets which satisfy RP.

Upper Bound Algorithm.

Inputs: $I, \{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$. Outputs: $\overline{N}$ and $\overline{G}$

1. If $\{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$ satisfies RP set $\overline{N} = 1, \overline{G} = I$ and goto (8).
2. Select $i \in I$ with uniform probability, set $I = I \backslash i$.
3. Set $\overline{G}_1 = \{i\}$, set $\overline{G} = \left\{\overline{G}_1\right\}$
4. If $I \neq \varnothing$, select $i \in I$ with uniform probability, set $I = I \backslash i$, set $E = \overline{G}$. Else if $I = \varnothing$ goto (8)
5. If $E \neq \varnothing$, select $g = \arg\max\left\{|g| : g \in \overline{G}\right\}$, set $E = \overline{G} \backslash g$. Else goto (8)
6. If $\{\mathbf{p}_j, \mathbf{q}_j\}_{j \in \{g,i\}}$ satisfies RP set $\overline{G} = \overline{G} \backslash g$, set $g = g \cup i$, set $\overline{G} = \overline{G} \cup g$ and goto (4); else goto (7).
7. Set $\overline{G}_{|\overline{G}|+1} = \{i\}$, set $\overline{G} = \overline{G} \cup \overline{G}_{|\overline{G}|+1}$, goto (4).
8. $\overline{N} = |\overline{G}|$
9. Stop.

The outputs of the Upper Bound algorithm are $\overline{N}$ = the upper bound on the number of types and $\overline{G}$ = a set containing an exclusive and exhaustive partition of the data into $\overline{N}$ subsets such that the data in each type satisfy the RP conditions. The algorithm works on the principle of randomly ordering the data and trying to construct groups which satisfy RP conditional on that ordering. As new observations are drawn it tries to add them to the existing partition and starts by placing them in the largest group available. If an observation cannot be added to an existing group it is used to initialise a new group. The upper bound algorithm begins by picking a single observation at random without replacement. This forms the basis for the first group. It then chooses the next observation at random also without replacement and tests whether the two satisfy RP. If they do they are placed together in the first group. If they don't the new observation is used to begin a new group. The next observation is then drawn and, starting with the largest existing group an RP test is used to determine whether it can by placed in that group. It is placed into the first group where it satisfies RP. If no such group can be found amongst the exists groups the observation is used to start a new group. The algorithm continues in this way until the dataset is empty and all observations have been assigned to groups. Since this algorithm relies on a random ordering of the data we run it a number of times and retain the minimum partition over these independent runs.

Lower Bound Algorithm.

Inputs: $I, \{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$.. Output: $\underline{N}$

1. If $\{\mathbf{p}_i, \mathbf{q}_i\}_{i \in I}$ satisfies RP set $\underline{N}= 1, g = 1$ and goto (6).
2. Select $i \in I$ randomly with uniform probability, set $I = I \backslash i$ and $g = i$
3. If $I \neq \varnothing$, select $j \in I$ with uniform probability, set $I = I \backslash j$, else if $I = \varnothing$ goto (6)
4. If the dataset $\{\mathbf{p}_j, \mathbf{q}_j, \mathbf{p}_i, \mathbf{q}_i\}_{\forall j,i \in \{j,g\}}$ violates RP set $g = g \cup j$. Goto (3).
5. $\underline{N}= |g|$
6. Stop.

The output of the Lower Bound algortihm is $\underline{N}$ = the lower bound on the number of types. This algorithm works on the principle that if we can find $\underline{N}$ observations which violate RP in all pairwise tests conducted between themselves, then there muct be at least $\underline{N}$ groups (since none of these observations could ever be placed in the same group). It begins by selecting an initial observation at random without replacement. It then picks another observation without replacement and tests RP. If the pair satisfy RP the second observation is dropped and a new observation selected. However if the pair violate RP the new observation is retained. We now have two observations which violate RP. A third observation is now selected from the data without replacement. This is tested against each of the observations currently held. If it violates in pairwise tests against all of them then it is retained. Otherwise it is dropped. The algorithm continues in this way until the dataset is exhausted. At the end of the process the algorithm has collected together a set of observations which all violate pairwise RP test conducted between each of them. The number of these observations gives the lower bounds $\underline{N}$. As before the algorithm is reliant on a random ordering of the data. We therefore run the process a number of times and retain the maximum value of $\underline{N}$ we find.

7.2 Allowing for Optimisation Errors

So far we have treated the data as error free. We now consider optimising error by consumers. Our treatment of this issue is not original to this paper[7]. The point of this section is merely to show, briefly, that this firmly-established treatment can be applied in our context.

Afriat (1967) interpreted RP checks as a conflation of two sub-hypotheses: theoretical consistency *and* the idea that economic agents are efficient programmers. If the data violate the conditions then it may be that some consumers have made optimisation errors. His suggestion was that, instead of requiring exact efficiency, a form of partial efficiency is allowed. This is achieved by introducing a parameter $e \in [0, 1]$ (the Afriat efficiency parameter) such that

$$e\mathbf{p}_t'\mathbf{q}_t \geq \mathbf{p}_t'\mathbf{q}_s \ \Leftrightarrow \mathbf{q}_t R_e^0 \mathbf{q}_s$$

The weaker form of GARP is then

$$\mathbf{q}_j R_e \mathbf{q}_i \Rightarrow e\mathbf{p}_i'\mathbf{q}_i \leq \mathbf{p}_i'\mathbf{q}_j$$

where R_e is the transitive closure of R_e^0. The interpretation of e is as the proportion of the consumer's budget which they are allowed to waste through optimisation errors. This parameter is used to modify the restrictions of interest to allow for a weaker form of consistency (see Afriat (1967)). To admit optimisation error into the partitioning approach all we need to do is to specify a level for e in advance and insert the modified RP restriction into the algorithms at the appropriate steps (step (2) in the exact algorithm, step (6) in the upper bound and step (4) in the lower bound algorithm). It is then straightforward to examine how the results vary with the required efficiency level.

The effects of allowing for optimisation errors is to reduce the amount of heterogeneity which is needed to rationalise the data. Admit enough error and it is possible to rationalise almost anything. As a result, running the algorithms without these adaptations will give a "worst-case" assessment, delivering the "maximum minimum" number of groups.

We implemented this methodology with our sample of 500 observations of household milk demands. Clearly if there is enough optimisation error, then one can explain any behaviour with just one utility function. This is what we observe for $e \leq 0.781$. For e greater than

[7]The treatment of optimisation errors in RP tests is due to Afriat (1967) and that of measurement error is due to Varian (1985).

this, more than one utility function may be required to explain the variation in behaviour that we observe. For $e \geq 0.90$, more than one utility function is definitely required to explain the variation in observed behaviour.

TABLE A.1: Bounds on the Minimum Number of types for various e

e	Number of Types	
	Lower bound	Upper bound
0.78	1	1
0.80	1	2
0.85	1	2
0.90	2	3
0.95	3	4
1.00	4	5

7.3 Panel Methods

So far we have considered cross-section data. Clearly in cross section data, where each consumer is observed pnly once, some degree of commonality in preferences is necessary in order to make progress in applied work. However panel data generally holds out the hope of identifying more about individuals than is possible with cross section data. Indeed panel data has two important features in terms of identifying types in our framework. Firstly, repeated observation on individuals allows them to distinguish their type more clearly through their behaviour. Secondly, repeated observations mean that stability of preferences becomes an important factor.

Recalling our main question: how many sets of preferences are needed to rationalise the data? a natural way to proceed is to first check GARP for each individual consumer and then to seek to allocate consumers into type groups. Given a set of individually GARP-consistent consumers, the algorithms described above can be applied almost without modification.

Of course some consumers will individually fail GARP and the question arises what to do with them. One answer is to simply set them aside as their behaviour is not rationalisable with the model of interest. However, this would not be in the spirit of taking rationality as a maintained assumption. A second possibility is to allow for enough optimisation errors in the way we describe above. However, this strategy would also affect the grouping of people (because admitting optimisation error would tend to decrease the amount of RP violations). Thirdly, we could consider alternative models for their intertemporal behaviour, such as a habits model (see Crawford 2008).

People change. Although economists like to invoke immutable preferences, we all know that our preferences can change, sometimes in a dramatic fashion. So, a final alternative is to "break them up into little pieces". By that we mean that we can take the data on an individual and search within it for contiguous sub-periods during which their behaviour is rationalisable. We can then treat each of these sub-periods as a separate individual (which they are in the sense that each one potentially requires a different utility function to model it) and run the partitioning algorithms as before. Because this approach sits wholly inside our basic framework, without the need for discarding data, including optimisation errors or writing down a dynamic structure for utility functions, it is in some sense the simplest option, and therefore our preferred one.

7.3.1 Panel Results

We use the same 500 households as in our cross-sectional analysis, but use a sequence of up to 24 months of milk consumption data for each household. These panel data on Danish milk prices and expenditures are detailed in Crawford (2010).

Using the third strategy mentioned above ("break people into pieces"), which preserves all of the data from our 500 households, we implement our model. The number of groups needed to completely rationalise these data is at least 12 and not more than 31. This is quite surprising. After all the standard approach to panel data in applied econometric analysis is to use "fixed effects", which in this context would imply 434 groups. Our results show that this is at least 15 times as many groups as are really necessary, and therefore is radically overspecified.

Fixed effects are a bad match to these data for 2 more important reasons. First, Blundell, Duncan and Pendakur (1998) show that fixed effects in budget shares are consistent with rationality restrictions only if budget shares are linear in the natural logarithm of total expenditure. Second, in our exploration of cross-sectional data above, we showed that *both* levels *and* derivatives of budget-share equations vary across groups. Thus, fixed level effects do not adequately capture the differences across groups.

7.4 Other Contexts

The RP methods outlined in this paper can be easily adapted to other optimising models. The table below gives RP restrictions corresponding to utility maximisation (Afriat 1967, Diewert 1973, Varian 1982, 1983a, 1983b, 1984), profit maximisation (Hanoch and Rothschild 1972) and cost minimisation (Hanoch and Rothschild 1972). Corresponding restrictions are available for models of intertemporal choice (Crawford 2010), habits (Crawford 2010), choice under uncertainty (Browning 1989), collective household behaviour (Cherchye, De Rock and Vermuelen 2007), characteristics models (Browning, Blow and Crawford 2008), as well as special cases of all of these models which embody useful structural restrictions on preferences or technology (e.g. weak separability, homotheticity and latent separability), Table: Example Restrictions

Model	$\mathbf{X}_i$	T
Utility Max.	$\{\mathbf{p}_i, \mathbf{q}_i\}$	$\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j \; \mathbf{p}_j'\mathbf{q}_j \geq \mathbf{p}_j'\mathbf{q}_k \ldots \geq \mathbf{p}_r'\mathbf{q}_s \Rightarrow \mathbf{p}_s'\mathbf{q}_s \leq \mathbf{p}_s'\mathbf{q}_i \; \forall i, j \in I$
Profit Max.	$\{\mathbf{p}_i, \mathbf{w}_i, \mathbf{q}_i, \mathbf{x}_i\}$	$[\mathbf{p}_i : \mathbf{w}_i]' [\mathbf{q}_i : -\mathbf{x}_i] \geq [\mathbf{p}_i : \mathbf{w}_i]' [\mathbf{q}_j : -\mathbf{x}_j] \, \forall i, j \in I$
Cost Min.	$\{\mathbf{w}_i, \mathbf{q}_i, \mathbf{x}_i\}$	$\mathbf{w}_i'\mathbf{x}_i \leq \mathbf{w}_i'\mathbf{x}_j$ and $\mathbf{q}_i \leq \mathbf{q}_j \forall i, j \in I$

Note: $(\mathbf{p}_i, \mathbf{w}_i, \mathbf{q}_i, \mathbf{x}_i)$ denotes consumer/output prices, input prices, consumer demands/sales and factor demands respectively.

In order to apply our methods to another optimising model, one simply replaces the GARP restrictions used in the body of this paper with the corresponding restrictions on optimising behaviour driven by the model of interest. Below, we briefly demonstrate how this works by applying our methods to a model of firm cost minimisation. Instead of seeking the minimum number of distinct preference maps necessary to rationalise observed consumption choices, we seek the minimum number of distinct technologies necessary to rationalise the observed input demand choices.

7.4.1 Firm Data

Here we provide an illustration of the application of partitioning to firm data. The data relate to 281 Danish Farms observed in 1990. These are Danish Farm Association Service data gathered through a voluntary consultancy scheme and for each farm the data includes detailed annual accounts of variable costs and earnings for each production line with corresponding accounts measures of most inputs and outputs. We measure five outputs {milk, two types of beef, and two types of crops} and we observed 46 inputs - like fodder, cattle, fertiliser, pesticides, and the services from labour, land, building and machine capital. The farms recorded the transactions prices for each of their inputs and outputs. In this application we are interested in unobserved technological heterogeneity and the economic model of interest is the canonical cost minimisation model:

$$\min_{\mathbf{x}} \ \mathbf{w}_i'\mathbf{x} \text{ subject to } \mathbf{x} \text{ is in } V(\mathbf{q}_i)$$

where $\mathbf{q}_i$ denotes a vector recording the quantities of the outputs of firm i and $\mathbf{x}_i$ is a vector recording the quantities of the firm's inputs. Technology is denoted by the input requirement set $V(\mathbf{q}_i)$ which is a closed, non-empty, monotonic, nested and convex set which consists of all input vectors $\mathbf{x}$ that can produce at least the output vector $\mathbf{q}_i$.The observable consequences of this model are summarised in the following theorem.

Theorem: (Hanoch and Rothschild (1972), Diewert and Parkan (1983), Varian (1984)). *The following conditions are equivalent:*
(1) there exists a family of nontrivial, closed, convex, positive monotonic input requirement sets $\{V(\mathbf{q})\}$ *such that the data* $\{\mathbf{q}_t, \mathbf{w}_t, \mathbf{x}_t\}$ *solves the problem* $\min_{\mathbf{x}} \ \mathbf{w}_i'\mathbf{x}$ subject to $\mathbf{x}$ is in $V(\mathbf{q}_i)$ for each $i = 0, 1, ..., N$.
(2) if $\mathbf{q}_j \geq \mathbf{q}_i$ *then* $\mathbf{w}_i'\mathbf{x}_j \geq \mathbf{w}_i'\mathbf{x}_i$ *for all* i *and* j.

The condition in (2) is the Weak Axiom of Cost Minimisation (Varian, 1983) and it provides the partitioning criteria: if the data for two firms are such that (2) does not hold then the two firms concerned cannot have the same technology. We simply replace GARP with WACM in the algorithms described in the paper and find the following bounds on the number of technological types:

$$3 \leq N_G \leq 4$$

It appears that very few production technologies are required to model (precisely) these very disaggregated data in which firms are able to choose from many inputs and produce several outputs. Furthermore as with the consumer cross section data it turns out that a single production technology will fit the majority of the data:

Table A.2: The cumulative distribution of technological types

Number of Types	1	2	3	4
Percent of Sample Explained	80%	94%	98%	100%

References

[1] Afriat, S.N., (1967),"The construction of a utility function from expenditure data", *International Economic Review,* **8**, 76-77.

[2] Bar-Shira, Z., (1992) "Nonparametric Test of the Expected Utility Hypothesis" *American Journal of Agricultural Economics* **74**(3), 523-533

[3] Blundell, R. M. Browning and I. Crawford (2003), "Nonparametric Engel Curves and Revealed Preference", *Econometrica*, **71**(1), 205-240.

[4] Blundell, R. M. Browning and I. Crawford (2007), "Improving Revealed Preference Bounds on Demand Responses" *International Economic Review*, 48(**4**), 1227-1244.

[5] Brown, B. and M .B. Walker, (1989), "The random utiity hypothesis and Inference in Demand Systems", *Econometrica*, **57**(4), 815-829.

[6] Browning, M. (1989), "A nonparametric test of the life-cycle rational expections hypothesis" *International Economic Review*, 30(4), 979-992

[7] Browning, M. Blow, L. and I. Crawford, (2008), "Revealed Preference Analysis of Characteristics Models" *Review of Economic Studies*, **75**(4),.1-19

[8] Cherchye, L. B. De Rock and F. Vermeulen, (2007) "The collective model of household consumption: a nonparametric characterization" *Econometrica* ,**75**, 553-574.

[9] Crawford, I. (2010), "Habits Revealed", forthcoming *Review of Economic Studies*.

[10] Dean, M. and D. Martin, (2010), "How Rational Are Your Choice Data?", unpublished working paper,

[11] Diewert, W.E., (1973), "Afriat and revealed preference theory", *Review of Economic Studies*, **40**, 419-426.

[12] Gorman, W. M. (1956), "A possible procedure for analysing quality differentials in the eggs market", *Review of Economic Studies*, (October 1980) **47**, 843-856.

[13] Haag, B., Hoderlein, S. and K. Pendakur (2009), "Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems", *Journal of Econometrics* **153** (2009) pp. 33-50.

[14] Hanoch, G. and M. Rothschild (1972). "Testing the Assumptions of Production Theory: A Nonparametric Approach," *Journal of Political Economy*, 80(2), 256-75

[15] Hill, M.O., (1973) "Diversity and evenness: A unifying notation and. its consequences". *Ecology* **54**, 427-432

[16] Lancaster, K. (1966), "A new approach to consumer theory", *Journal of Political Economy*, **74**, 2, 132-157.

[17] Lewbel, A., (2001), "Demand systems with and without errors", *American Economic Review*, **91**(3), 611-618.

[18] Lewbel, A. and K. Pendakur, (2008), "Tricks with Hicks: The EASI Demand System", *American Economic Review*, forthcoming.

[19] Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus*, Chichester: Wiley.

[20] McElroy, M., (1987), "Additive general error models for production, cost and derived demand or share systems", *Journal of Political Economy*, 95(4) 737-757.

[21] Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, **69**(2), 307-342

[22] Shannon, C. E., (1948), "A Mathematical Theory of Communication" *The Bell System Technical Journal.* **27**, 379-423

[23] Theil, H., (1965) "The Information Approach to Demand Analysis" *Econometrica* **33**, 67-87

[24] Varian, H., (1982), "The nonparametric approach to demand analysis", *Econometrica*, **50**, 945-973.

[25] Varian, H., (1983a), "Nonparametric tests of consumer behaviour", *Review of Economic Studies*, **50**, 99-110.

[26] Varian, H., (1983b), "Nonparametric Tests of Models of Investor Behavior" *Journal of Financial and Quantitative Analysis*, 18(3), 269-278.

[27] Varian, H. R (1984), "The Nonparametric Approach to Production Analysis," *Econometrica* **52**, 579-97.

[28] Varian, H.R. (1985) "Nonparametric Analysis of Optimizing Behavior with measurement Error"; *Journal of Econometrics*, **30**, 445-458.

[29] Varian,. H., (1988), "Revealed preference with a subset of goods". *Journal of Economic Theory*, **46**(1),179-185.