



**DEPARTMENT OF ECONOMICS**

**DISCUSSION PAPER SERIES**

**REVEALED PREFERENCE ANALYSIS OF  
CHARACTERISTICS MODELS**

**Laura Blow, Martin Browning and Ian Crawford**

Number 282

October 2006

Manor Road Building, Oxford OX1 3UQ

# REVEALED PREFERENCE ANALYSIS OF CHARACTERISTICS MODELS.

Laura Blow  
Institute for Fiscal Studies,  
London

Martin Browning  
Department of Economics,  
Oxford University

Ian Crawford  
University of Surrey,  
cemmap, Institute for Fiscal Studies

## Abstract

Characteristics models have been found to be useful in many areas of economics. However, their empirical implementation tends to rely heavily on functional form assumptions. In this paper we develop a revealed preference approach to characteristics models. We derive the necessary and sufficient empirical conditions under which data on the market behaviour of heterogeneous, price-taking consumers are nonparametrically consistent with the consumer characteristics model. Where these conditions hold, we show how information may be recovered on individual consumer's marginal valuations of product attributes. In some cases marginal valuations are point identified and in other cases we can only recover bounds. Where the conditions fail we highlight the role which the introduction of unobserved product attributes can play in rationalising the data. We implement these ideas using consumer panel data on the Danish milk market.

**Key Words:** Product characteristics, revealed preference.

**JEL Classification:** C43, D11.

**Acknowledgements:** We are very grateful to the Editor, two anonymous referees, Laura Anderson, Lars Nesheim, Ian Preston and participants in various seminars and in the W. M. (Terence) Gorman Memorial Workshop held in London in June 2004 for their help and comments. Funding from the ESRC, grant number R000239865, and from the Danish National Research Foundation through its grant to CAM is gratefully acknowledged. Correspondence to: Laura\_Blow@ifs.org.uk, Martin.Browning@econ.ku.dk, i.crawford@surrey.ac.uk.

# 1 Introduction.

The idea that consumers have preferences over the characteristics of market goods, as developed by Gorman (1956), Lancaster (1966), Muellbauer (1974) and Rosen (1974), has turned out to be an extremely fruitful one.<sup>1</sup> For example, it is central to much applied microeconomic work on price indices (for example, Stone (1956) as an early example), quality change (Griliches (1971)), location decisions (Tinbergen (1959)), labour market allocation (Heckman and Scheinkman (1987)), finance (Markowitz (1959)) and the analysis of markets for differentiated products (Berry, Levinsohn and Pakes (1995)). However, the empirical implementation of characteristics models tends to rely heavily on functional form assumptions. In this paper we develop a revealed preference<sup>2</sup> (nonparametric) approach to the empirical analysis of characteristics models. We derive the necessary and sufficient conditions under which data on the market behaviour of heterogeneous and price-taking consumers are nonparametrically consistent with the consumer characteristics model. One important feature of our analysis is that we show how to allow for the fact that prices might not be observed for goods that are not bought. This development is of considerable use in itself (since the non-observation of prices is common) but it also facilitates the analysis of finding virtual (reservation) prices for goods that are not bought and also the pricing of new goods. Where our conditions fail, we highlight the role which the introduction of product attributes which are observed by consumers but not the researcher ('latent characteristics') can play in rationalising the data. Where these conditions hold, we show how information may be recovered on individual consumer's marginal valuations of product attributes. In some cases marginal valuations are point identified and in other cases we can only recover bounds.

We apply the revealed preference techniques we develop to consumer panel data on the Danish milk market. The two major characteristics that vary across different types of milk are fat content and whether it is produced under 'organic' conditions. Both aspects are important for substantive issues. Given the concern over increasing obesity in all high income countries there is interest in identifying the marginal valuation of fat and how this is distributed across the population. Our techniques allow us to identify the distribution of these valuations. As regards the second characteristic, 'organic', interest here centres on how much extra consumers are willing to pay for milk that is arguably more healthy (for example, because cows that produce organic milk are never treated with antibiotics) and increases animal welfare. Given that organic milk is more expensive to produce it is important to establish the demand so that farmers or milk marketing boards can assess the demand for organic milk. In our empirical analysis we use data which allows us to follow individual households over very long periods so that we can deal with heterogeneity in a fully nonparametric way by treating each household as an individual time series. We find that the demand paths of a majority of households can be rationalised by a linear characteristics model defined over three attributes. We show that the distribution of preferences over characteristics is multi-modal, suggesting the existence of distinct types of consumers.

## 2 Revealed preference conditions.

### 2.1 The Afriat-Varian conditions.

We first present the usual revealed preference (RP) conditions for the standard model in which agents have 'direct' preferences over market goods. Let  $\mathbf{q}_t$  be a  $(K \times 1)$  vector of quantities of market goods bought in period  $t$  at market prices  $\mathbf{p}_t$ . Given a data set  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$ , we ask when these data be rationalised by a utility function, in the following sense:

**Definition 1** A utility function  $u(\mathbf{q})$   $\mathbf{q}$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  if  $u(\mathbf{q}_t) \geq u(\mathbf{q})$  for all  $\mathbf{q}$  such that  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}$ .

The standard results (see Varian (1982) for the definition of GARP) are the classic Afriat-Varian conditions:

---

<sup>1</sup>We choose to term this class of models 'characteristics' models but other terms are also used. In particular, the term 'hedonic' models (first used in this context by Court in 1939, see Goodman (1998)) is widely used.

<sup>2</sup>See Samuelson (1948), Houthakker (1950), Afriat (1967), Hanoch and Rothschild (1972), Diewert (1973), Diewert and Parkan (1978) and Varian (1982, 1983, 1984).

**Afriat's Theorem**<sup>3</sup>. The following statements are equivalent:

1. there exists a utility function  $u(\mathbf{q})$  which is continuous, non-satiated and concave which  $q$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$ .
2. there exist numbers  $\{U_t, \lambda_t > 0\}_{t=1, \dots, T}$  such that

$$U_s \leq U_t + \lambda_t \mathbf{p}'_t (\mathbf{q}_s - \mathbf{q}_t) \quad \forall s, t = 1, \dots, T$$

3. the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  satisfy the Generalised Axiom of Revealed Preference (GARP).

These conditions have been applied to aggregate time series data (see, for example, Varian (1982)) and (less frequently) to household data (see Blundell, Browning and Crawford (2003)).

## 2.2 RP conditions for the characteristics model.

The characteristics model posits that, rather than having preferences over market goods directly, agents have preferences over the characteristics or attributes that these goods embody. The transformation from a  $K$ -vector of market goods,  $\mathbf{q}$ , to a  $J$ -vector of characteristics,  $\mathbf{z}$ , is given by the mapping  $\mathbf{z} = \mathbf{F}(\mathbf{q})$ . In all that follows we assume that  $\mathbf{F}(\cdot)$  is increasing and concave. Preferences are then defined over characteristics  $v(\mathbf{z})$ . This gives a derived utility function over market goods  $u(\mathbf{q}) = v(\mathbf{F}(\mathbf{q}))$ . The consumer choice model for given prices  $\mathbf{p}$  and outlay  $x$  is:

$$\max_{\mathbf{q}} v(\mathbf{z}) \quad \text{subject to } \mathbf{z} = \mathbf{F}(\mathbf{q}) \text{ and } \mathbf{p}'\mathbf{q} \leq x, \mathbf{q} \geq \mathbf{0}$$

This model takes prices as given for individual agents and we follow this treatment in this paper. Thus we adopt the Muellbauer (1974) and Gorman (1956) tradition of focusing on the demand side, and abstract from any supply side simultaneity issues (Rosen (1974)).

The most widely analysed and applied version of the consumer characteristics model is one in which characteristics are a linear function of the demands for market goods. That is  $\mathbf{z} = \mathbf{A}'\mathbf{q}$  where  $\mathbf{A}$  is a  $(K \times J)$  matrix of constants with  $J < K$  and  $\mathbf{A}$  has full column rank<sup>4</sup>. The technology matrix  $\mathbf{A}$  records the amounts of each of the characteristics present in one unit of each of the market goods. In this paper we concentrate on this linear version of model but note that most of what follows also applies to non-linear characteristics with concave technologies. The analogue for non-linear technologies of the first theorem below is given in appendix 1. In this section we assume that  $\mathbf{A}$  is known to the researcher; in section 3 we discuss the case in which we have incomplete knowledge of the transformation from goods to characteristics.

Our initial focus is on the circumstances under which data can be nonparametrically rationalised by this model. In this context the term 'rationalise' is defined as follows:

**Definition 2** A utility function  $v(\mathbf{z})$   $z$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  for the technology  $\mathbf{A}$  if  $v(\mathbf{z}_t) = v(\mathbf{A}'\mathbf{q}_t) \geq v(\mathbf{z})$  for all  $\mathbf{z}$  such that  $\mathbf{z} = \mathbf{A}'\mathbf{q}$  and  $\mathbf{p}'_t\mathbf{q}_t \geq \mathbf{p}'_t\mathbf{q}$ .

This states that a utility function rationalises observed choices if it assigns an equal or higher value to those bundles of characteristics which the consumer chooses, than it does to those alternative bundles of characteristics which could have feasibly been produced from affordable bundles of market goods. If a utility function  $z$ -rationalises the data, this means that were it used in the consumer's maximisation problem set out above, then it would generate exactly the observed data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  for the posited technology  $\mathbf{A}$ . Clearly  $z$ -rationalisation for *any*  $\mathbf{A}$  matrix implies  $q$ -rationalisation.

For a good which is purchased, the first order condition from the linear characteristics model gives the following characterisation of its price as a weighted sum of the shadow prices of its characteristics:

$$p_t^k = \mathbf{a}_k \boldsymbol{\pi}_t = \sum_{j=1}^J a_{kj} \pi_{jt} \quad (1)$$

<sup>3</sup> Afriat (1967), Diewert (1973), Varian (1982).

<sup>4</sup> Note that we are only interested in the case in which  $J < K$  since, as we later show, if the number of characteristics is greater than or equal to the number of goods, then the characteristics model is nonparametrically indistinguishable from the standard preferences-over-products model.

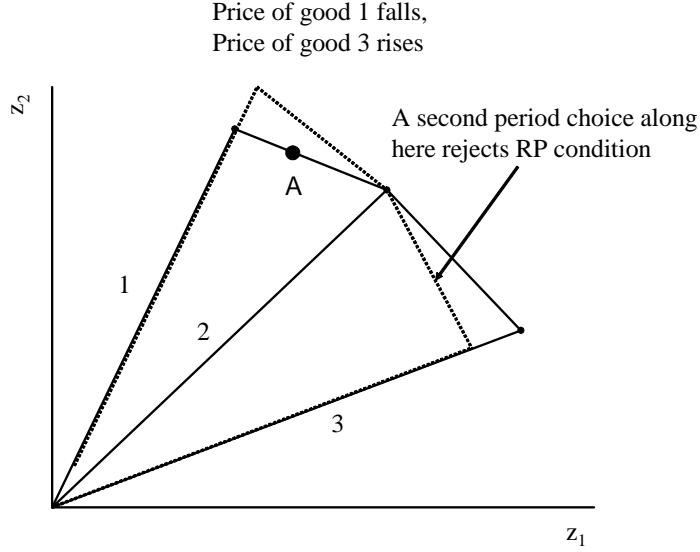


Figure 1: Revealed preference for characteristics models.

where  $\mathbf{a}_k$  denotes the  $k$ th row of  $\mathbf{A}$  and:

$$\pi_{jt} = (\lambda_t)^{-1} v_j(\mathbf{z}_t)$$

Thus the shadow price of a characteristic is defined as its marginal utility normalised by the marginal utility of total expenditure ( $\lambda_t$ ) (see Gorman (1956), equation (5)). That the market price of a good that is bought can be viewed as a linear combination of the underlying shadow prices is the most important feature of characteristics models. If good  $k$  is not bought then we have the inequality:

$$p_t^k \geq \mathbf{a}_k \boldsymbol{\pi}_t \quad (2)$$

so that the market price is too high relative to the subjective valuation of the embodied attributes.

The key to deriving necessary and sufficient conditions for the characteristics model can be shown in a simple figure. In Figure 1 we present a graphical three good, two attribute illustration of the revealed preference conditions. Initially the agent faces the constraints given by the kite shape defined by solid lines and chooses point  $A$  at which she is consuming goods 1 and 2. Suppose now that the price of good 1 falls and the price of good 3 rises, so that she now faces the dotted constraint set. If she now switches to buying some of good 3 and none of good 1 then we cannot rationalise her choices with convex preferences. If, on the other hand, she continues to buy both of goods 1 and 2, or even only one of them, we can find indifference curves that rationalise the choices. Our task is now to extend this insight into feasible testable conditions for several goods and characteristics. The next theorem gives the necessary and sufficient conditions for the characteristics model (the proof is given in appendix 2).

**Theorem 1** *The following statements are equivalent.*

(P) *there exists a utility function  $v(\mathbf{z})$  which is non-satiated, continuous and concave in characteristics which  $\mathbf{z}$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  for given  $\mathbf{A}$ .*

(A) *there exist  $2T$  numbers  $\{V_t, \lambda_t > 0\}_{t=1, \dots, T}$  and  $T$   $J$ -vectors  $\{\boldsymbol{\pi}_t\}_{t=1, \dots, T}$  such that*

$$V_s \leq V_t + \lambda_t \boldsymbol{\pi}_t' (\mathbf{A}' \mathbf{q}_s - \mathbf{A}' \mathbf{q}_t), \quad \forall s, t \quad (A1)$$

$$p_t^k \geq \mathbf{a}_k \boldsymbol{\pi}_t, \quad \forall k, t \quad (A2)$$

$$p_t^k = \mathbf{a}_k \boldsymbol{\pi}_t \text{ if } q_t^k > 0, \quad \forall k, t \quad (A3)$$

(L) there exist numbers  $\{U_t, \rho_t \geq 1, \}_{t=1, \dots, T}$  and vectors  $\{\sigma_t\}_{t=1, \dots, T}$  such that

$$U_s \leq U_t + \sigma_t' (\mathbf{A}' \mathbf{q}_s - \mathbf{A}' \mathbf{q}_t), \quad \forall s, t \quad (L1)$$

$$\rho_t p_t^k \geq \mathbf{a}_k \sigma_t, \quad \forall k, t \quad (L2)$$

$$\rho_t p_t^k = \mathbf{a}_k \sigma_t \text{ if } q_t^k > 0, \quad \forall k, t \quad (L3)$$

(G) the data  $\{\pi_t, \mathbf{A}' \mathbf{q}_t\}_{t=1, \dots, T}$  pass GARP for some choice of  $\pi_t$  such that (A2) and (A3) are satisfied.

One important feature of these conditions is that they do not impose that the shadow prices (the  $\pi_t^j$ 's in (A)) are non-negative; that is, agents may have a negative valuation for some characteristics. Of course, some of the shadow prices must be positive otherwise condition (L2) could not hold. Conditions (A2) and (A3) impose the linear pricing condition (1). Conditions (A) and (G) are the characteristics model analogues of the conditions in Afriat's Theorem. However, both present practical difficulties for testing. Condition (G) requires that we first find the shadow prices in order to implement a GARP test and there is no known general algorithm to do this in a finite number of steps<sup>5</sup>. Condition (A) involves both non-linear functions of unknowns (the  $\lambda_t \pi_t'$  terms) and strict inequality constraints on unknowns ( $\lambda_t > 0$ ). To overcome these problems in implementing the test for rationalisation we have derived condition (L). This condition is in the form of the restrictions in the first step of a linear programming problem. Consequently we can employ standard linear programming techniques to find, in a finite number of steps, whether there exists a feasible set of unknowns which satisfy these constraints (the first step of all linear programming algorithms). Nevertheless, whilst this approach will always give a result in a finite number of steps, for large problems this may lead to a large dimension problem since we have  $T(J+2)$  unknowns,  $T(T-1) + T(K+J)$  inequality constraints and  $TK$  equality constraints. Given this, it is convenient to derive an easily tested necessary condition for  $z$ -rationalisation.<sup>6</sup>

### 2.3 A necessary condition for $z$ -rationalisation.

Denote the sub-vector of period  $t$  prices for which demands are positive as  $\mathbf{p}_t^+$ , and let  $\mathbf{A}_t^+$  be corresponding sub-matrix of  $\mathbf{A}$  (with  $\mathbf{p}_t^0$  and  $\mathbf{A}_t^0$  denoting the complementary sub-vectors and sub-matrices for goods for which demands are zero)<sup>7</sup>. Any set of shadow price vectors that satisfy (A2) and (A3) must have:

$$\mathbf{p}_t^+ = \mathbf{A}_t^+ \pi_t$$

A necessary condition for this system of equations to have a solution is that:

$$\text{rank}(\mathbf{A}_t^+ \sim \mathbf{p}_t^+) = \text{rank}(\mathbf{A}_t^+)$$

where the  $\sim$  symbol denotes horizontal concatenation of the matrix  $\mathbf{A}_t^+$  with the extra column  $\mathbf{p}_t^+$ . Note that this simple necessary condition means that, in contrast to standard revealed preference tests which require at least two observations, the characteristics model is potentially rejectable with a single observation on a consumer; for example, if the agent buys all three goods in the first period of the example in the last subsection. If this rank condition holds then a solution for the shadow prices will be given by:

$$\pi_t = [\mathbf{A}_t^+]^{-1} \mathbf{p}_t^+$$

where the operator  $[\mathbf{X}]^{-1}$  denotes the generalised (Moore-Penrose) inverse of the not-necessarily-square matrix  $\mathbf{X}$ . These rank conditions are very easy to check and if they fail for some  $t$  then we know that the data can never satisfy (A2) and (A3). Note, however, that these conditions are only necessary and even if they hold we may not be able to find shadow prices that satisfy:

$$\mathbf{p}_t^0 \geq \mathbf{A}_t^0 \pi_t$$

<sup>5</sup>The computational problem is akin to that encountered in revealed preference tests of weak functional separability (see Varian (1983)). See below for a discussion of the particular circumstances under which the  $\pi_t$  vectors may be recovered uniquely before we test for GARP.

<sup>6</sup>The empirical application we present later is a small problem by industry standards and does not need a preliminary test of these necessary conditions.

<sup>7</sup>The  $t$  subscript on  $\mathbf{A}_t^+$  and  $\mathbf{A}_t^0$  reflects the fact the pattern of goods purchased can vary from period to period.

The rank conditions are particularly useful if  $\text{rank}(\mathbf{A}_t^+) = J$  since then we can solve *uniquely* for the shadow price vectors. If this holds in every period, then we can use GARP and condition (G) for testing which is computationally very rapid.

If the rank condition holds but the agent buys a subset of goods containing fewer than the full set of characteristics so that  $\text{rank}(\mathbf{A}_t^+) < J$ , then it will not be possible to obtain unique estimates of all the shadow prices. The general solution will be given by:

$$\boldsymbol{\pi}_t = [\mathbf{A}_t^+]^{-1} \mathbf{p}_t^+ - (\mathbf{I}_J - [\mathbf{A}_t^+]^{-1} \mathbf{A}) \boldsymbol{\mu}$$

where  $\boldsymbol{\mu}$  is an arbitrary  $J$ -vector and  $\mathbf{I}_J$  is the identity matrix. The exact nature of the solution depends on the form that  $(\mathbf{I}_J - [\mathbf{A}_t^+]^{-1} \mathbf{A})$  takes. It may be possible to find unique solutions for some shadow prices (in which case the corresponding row of  $(\mathbf{I}_J - [\mathbf{A}_t^+]^{-1} \mathbf{A})$  is a row of zeros). Others can only be expressed as a set of simultaneous equations. In this case, whilst there are an infinite number of solutions for the shadow prices, they are not independent of each other. Finally, some shadow prices will not be identified at all; that is, they could take any value, independent of the other shadow price solutions. In these last two cases we could bound the values by imposing other conditions, for example that shadow prices cannot be negative. In general, then, we can only identify (polyhedral) sets of shadow prices.

## 2.4 Missing prices.

Thus far we have assumed that the researcher always observes the price of all goods, even of those goods that the consumer does not buy in a particular period. For some data structures this is not the case and we only observe prices the agent faced (or can construct unit values) for goods that are bought. One possible procedure is to impute the missing prices, perhaps by using the prices paid by other consumers in the same region and period. Another is to use suitable published price indices. The problem with any such imputation is that we can never know how much the outcome of the test depends on the imputation. An alternative procedure is to include the missing prices as unknowns and search for values so that the constructed data satisfy the conditions. Since we can always implicitly set the prices of the goods which are not bought very high, this obviously makes it easier to satisfy the conditions and so the resulting test will be weaker in this sense.

We adopt the same notation as the last subsection with  $\mathbf{p}_t^+$  for the prices of the goods bought in period  $t$ ;  $\mathbf{p}_t^0$  for the prices not observed in period  $t$  (since these goods were not bought) and  $\mathbf{p}_t = \{\mathbf{p}_t^+, \mathbf{p}_t^0\}$  arranged in the correct order. Formally we have the following Afriat conditions for the characteristics model with missing prices:

**Theorem 2** *The following statements are equivalent:*

- (P) *there exists a utility function  $v(\mathbf{z})$  which is non-satiated, continuous and concave in characteristics and there exist prices  $\{\mathbf{p}_t^0\}_{t=1,\dots,T}$  which z-rationalise the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1,\dots,T}$  for given  $\mathbf{A}$ .*
- (A) *there exist numbers  $\{V_t, \lambda_t > 0\}_{t=1,\dots,T}$  and vectors  $\{\boldsymbol{\pi}_t\}_{t=1,\dots,T}$  such that*

$$\begin{aligned} V_s &\leq V_t + \lambda_t \boldsymbol{\pi}_t' (\mathbf{A}' \mathbf{q}_s - \mathbf{A}' \mathbf{q}_t), \quad \forall s, t \\ \mathbf{p}_t^+ &= \mathbf{A}_t^+ \boldsymbol{\pi}_t, \quad \forall t \end{aligned}$$

- (L) *there exist numbers  $\{W_t, \rho_t \geq 1, \}_{t=1,\dots,T}$  and vectors  $\{\boldsymbol{\sigma}_t\}_{t=1,\dots,T}$  such that*

$$\begin{aligned} W_s &\leq W_t + \boldsymbol{\sigma}_t' (\mathbf{A}' \mathbf{q}_s - \mathbf{A}' \mathbf{q}_t), \quad \forall s, t \\ \rho_t \mathbf{p}_t^+ &= \mathbf{A}_t^+ \boldsymbol{\sigma}_t, \quad \forall t \end{aligned}$$

Conditions (A) and (L) are identical to those in Theorem 1, except that they do not involve the inequalities on prices of goods that are not bought. Once again we give the standard version of the condition and the linear programming form which is relatively easy to check. The test given in this theorem is weaker than the tests given in the Theorem 1 in the sense that passing the test on data in which all prices are observed (even if some quantities are zero) implies the conditions

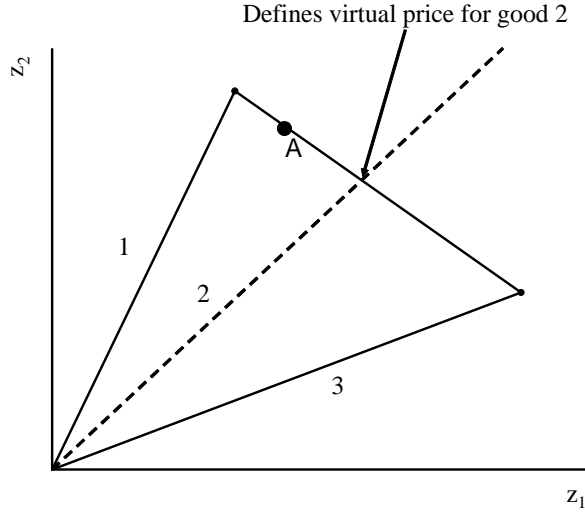


Figure 2: Finding the virtual price for a market good.

in this theorem, but not the other way around. If condition (L) holds then we can take any set of implied  $\sigma_t$ 's and  $\rho_t$ 's and simply set:

$$\mathbf{p}_t^0 = \rho_t^{-1} \mathbf{A}_t^0 \sigma_t \quad (3)$$

(which satisfies (L3) in Theorem 1). In this case the resulting  $\mathbf{p}_t^0$  vectors have the interpretation of being *virtual* prices; that is, at these prices consumers are just on the verge of buying market goods that they did not buy in period  $t$ . If the  $\sigma_t$ 's and the  $\rho_t$  are not unique, these values will not be uniquely determined either and we can only identify sets of virtual prices. In the case of set identification, we can sometimes find a lower bound on a particular virtual price by minimising the relevant element of  $\mathbf{p}_t^0$  in (3) subject to the constraints in (L); this is a standard linear programming problem. A graphical illustration is given in Figure 2 which shows an example with three market goods of which the agent only buys two (goods 1 and 3) and ends up at point A. From this figure we see that the price of good 2 must have been high enough to ensure that the end of the possibility line for that good was inside the triangle defined by the goods that were bought. The lowest possible price that good 2 could have had is given by the intersection with the budget line, as shown; if the price had been lower than this, then the agent would have bought good 2. This simple procedure for finding virtual prices compares favourably with the much more involved procedures for parametric models that require either special functional forms or use of the Legendre transform (see Neary and Roberts (1980) and Browning (1983)). Note, however, that this illustration in which the consumer buys as many goods as there are characteristics may be unduly optimistic. If in this example the consumer only buys good 1, then we cannot rule out that she positively dislikes attribute 1 and the virtual price is unbounded below. To achieve finite lower bounds we would need to put monotonicity constraints on the marginal valuations, and this may not be acceptable in all contexts.

To conclude this section we note that the conditions we have established are also useful for testing  $q$ -rationalisation. In the presence of missing prices the linear programming condition for  $q$ -rationalisation has a clear advantage over the GARP condition. As pointed out by Varian (1988), GARP-type tests are generally ruled out by missing price data because inner products such as  $\mathbf{p}_t' \mathbf{q}_s$  can involve missing prices; for example, if  $p_t^k$  is missing because  $q_t^k = 0$ , but  $q_s^k > 0$ . However, the framework described in Theorem 2 can be easily adapted to testing for  $q$ -rationalisation when there are missing prices by simply replacing  $\mathbf{A}$  with the identity matrix  $\mathbf{I}_K$ .<sup>8</sup> This then provides

<sup>8</sup>Varian (1988) establishes that if one price is not observed then there are no RP restrictions. Our context is different since we assume that a price is not observed in a particular period if and only if the good is not bought in that period; in this case there are testable RP restrictions (as shown).

a weaker test of GARP ( $q$ -rationalisation) in the presence of missing price data in the same sense that Theorem 2 provides a weaker test of the characteristics model than is given in Theorem 1.

## 2.5 Pricing new goods.

One of the primary motivations for using characteristics models is to price new goods. In our context, a new good is simply an extra row of the  $\mathbf{A}$  matrix that is not co-linear with any other row. The key insight into allowing for new goods in the revealed preference approach is that a new good is exactly like a good that has always been available, but was not bought before the introduction date for the new good. Hicks (1940) explicitly states this formal equivalence between pricing new goods and rationing (at zero). Given this, we can use the apparatus for dealing with missing prices established in the previous subsection to find the price at which a new good might be bought when it becomes available. In particular, we can find the maximum price for the new good that will ensure that it might be bought by finding the lowest virtual price as discussed after equation (3). In this light, the situation shown in Figure 2 can be viewed as pricing a new good, good 2. Of course, the prices of the other goods might also change consequent on the introduction of the new good (and might even be driven out of the market) and this simple procedure does not take this into account; rather we can state the maximum price a new good would have had to have to have any chance of being bought in an observed period. In the empirical analysis below we shall present an example of finding the maximum price for a new good.

## 3 Latent characteristics.

In all of the above we have assumed that we observe all of the characteristics of any good that is bought. In practice this is rarely the case. The leading example is when goods differ in ‘quality’ and this is not recorded, even though it might be reflected in the market price. Other examples are ‘freshness’ and ‘redness’ for apples, ‘elegance’ and ‘ease of use’ for domestic appliances and ‘political slant’ for newspapers. This has long been recognised as a problem in the literature and there are basically two approaches to dealing with latent characteristics. The first is due to Gorman (1956) (see Pudney (1981) for an empirical implementation). This is to allow that the name of the good is also a characteristic. This gives  $J + K$  characteristics with an augmented technology  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{I}_K$ . Then the pricing equation is:

$$p_t^k = \sum_{j=1}^J a_{kj} \pi_{jt} + e_t^k \quad (4)$$

where  $e_t^k$  is the shadow price of the ‘name’ of market good  $k$  in period  $t$ . Formally, of course, this is vacuous since we can always choose  $e_t^k = p_t^k$  and  $\pi_t^j = 0$  for all  $k, j$  and  $t$  so Gorman suggests choosing the values of the  $e_t^k$ ’s so as to minimise their impact of these auxiliary attributes in the pricing equation. A hedonic regression of prices on observable attributes (the  $a_{kj}$ ’s) is an obvious route. The second approach, which is formally equivalent and which is widely used in the IO literature which generally considers the case of discrete choice, is to assume that there is only one latent characteristic but it is time varying (and has a marginal price of unity) (see Berry, Levinsohn, and Pakes, (1995) and Bajari and Benkard (2004)).

In this section we discuss how to account for latent characteristics in the context of revealed preference tests. Suppose that the data satisfies the  $q$ -rationalisation conditions but not the  $z$ -rationalisation conditions for a given  $\mathbf{A}$ . One possible reaction is to seek an alternative transformation matrix which has rank less than  $K$  and does  $z$ -rationalise the data.<sup>9</sup> If, however, we really do believe that the characteristics given by  $\mathbf{A}$  are objects of preferences (but not the full set) then a more natural procedure is to seek a supplementary set of no more than  $K - J - 1$  unobserved characteristics that do  $z$ -rationalise the data in conjunction with the observed ones

The simplest extension is ask whether there is a single vector  $\mathbf{b}$  that we can concatenate to  $\mathbf{A}$  such that we can  $z$ -rationalise the data for  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{b}$ . This is equivalent to searching for one unobserved characteristic. A natural interpretation of this additional latent attribute is that it represents ‘quality’. Conceptually this is straightforward: we simply replace  $\mathbf{A}$  with  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{b}$  in

<sup>9</sup>Without the rank restriction we can always find a ‘trivial’ characteristics model with the transformation matrix being the identity matrix.

Market share (%)	Conventional milk			Organic milk		
	3.5%	1.5%	0.1%	3.5%	1.5%	0.1%
By value	14.68	45.79	14.16	4.00	10.72	10.64
By volume	13.65	49.26	15.54	3.03	9.16	9.36

Table 1: Market shares

Theorem 2 and add  $\mathbf{b}$  to the list of unknowns. Unfortunately the resulting problem no longer has a set of constraints that is linear in the unknowns so we cannot use standard linear programming techniques to ensure the existence or non-existence of such a set of parameters. Instead we have to use a nonlinear optimisation approach; details are given in appendix 3. If such a vector exists it will not be unique and it will not have any necessary interpretation; for example, we can always choose  $\mathbf{b}$  to be orthogonal to the columns of  $\mathbf{A}$ . If we cannot find one latent characteristic that gives  $z$ -rationalisation, the search can be extended to two latent characteristics (and so on up to  $K - J - 1$  latent characteristics) in the obvious way. In the empirical analysis below we present an example of such a search.

## 4 Application.

### 4.1 The data.

In this section we apply some of the ideas outlined above to data on purchases of (cow’s) milk in a Danish consumer panel. The data cover 2,500 households during 1999 and 2000; these households comprise all types ranging from young singles to couples with children to elderly couples. The sample is representative of the Danish population (the round number of 2,500 households is a coincidence). Over the two years, each household keeps a strict record of everything they buy in each shopping trip and record the price, quantity, and the characteristics of the good, the store used etc. 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 testing is appropriate. Thus for each month we have the quantity and expenditure for each type of milk; from this we construct a unit value for any milk bought as the monthly expenditure divided by the monthly quantity. Since we cannot construct a unit value for a type that is not bought we are in the missing price context discussed in subsection 2.4.

We concentrate on the 6 main types of milk which differ according to fat content (0.1% (skimmed), 1.5% (semi-skimmed) and 3.5% (full-fat)) and conventional/organic production methods<sup>10</sup>. Organic milk is produced under strictly supervised conditions that are meant to be better for the cows and to produce more healthy milk (for example, antibiotics cannot be used for ‘organic’ herds). Table 1 below shows the market shares by value and volume of these six types in our data. Overall conventional milk commanded the majority of the market with a share of about 75% on both measures, and semi-skimmed milk is the most popular type by fat content with nearly 60% of the market by both measures. This is mainly due to the pattern of conventional milk sales; within the organic segment of the market, skimmed and semi-skimmed have equal shares.

Table 2 gives some descriptive statistics on the prices of the different types of milk in Danish kroner (DKK) per litre (very approximately, 8 DKK equals one Euro or \$1.30 US). Organic milk is more expensive than conventionally produced milk by roughly 1.25 DKK/litre, and within the organic/conventional split there is a clear gradient with respect to fat content: the higher the fat content the higher the price with a price difference of roughly 1 DKK/litre between full-fat and skimmed milk. We treat preferences over milk characteristics as being a separable group from all other characteristics. This is a questionable assumption since the fat that is contained in milk is also to be found in other products such as butter and cheese. If we drop separability then we are left with an impossibly wide problem (hundreds of goods and dozens of characteristics); in this respect

<sup>10</sup>We drop speciality milks such as buttermilk and chocolate milks. Note too that container size may be a relevant priced characteristic. In what follows we do not consider container size for simplicity, but we note that container size and bulk discounting could be incorporated into our approach by defining milks sold in containers of different volumes as different market goods and by including the container volume as a characteristic.

Prices (DKK/litre)	Conventional milk			Organic milk		
	3.5%	1.5%	0.1%	3.5%	1.5%	0.1%
Mean	6.11	5.34	5.09	7.34	6.51	6.30
Median	6.00	5.30	5.00	7.31	6.50	6.27
Std. Dev.	0.62	0.56	0.31	0.38	0.33	0.23
Coeff of variation	0.10	0.10	0.06	0.05	0.05	0.04

Table 2: Prices

Market goods	Characteristics
Conventional milk, 3.5% fat	“Milkiness”
Conventional milk, 1.5% fat	Fat content
Conventional milk, 0.1% fat	“Organic”
Organic milk, 3.5% fat	
Organic milk, 1.5% fat	
Organic milk, 0.1% fat	
$K = 6$	$J = 3$

Table 3: Market goods and characteristics

it would be highly desirable to extend revealed preference tests to latent separability (see Crawford (2004) and Blundell and Robin (2000)). The corresponding characteristics model has  $K = 6$  market goods, and  $J = 3$  characteristics as given in Table 3. The transformation matrix is given by:

$$\mathbf{A} = \begin{bmatrix} 1 & 3.5 & 0 \\ 1 & 1.5 & 0 \\ 1 & 0.1 & 0 \\ 1 & 3.5 & 1 \\ 1 & 1.5 & 1 \\ 1 & 0.1 & 1 \end{bmatrix}$$

The first characteristic is the basic characteristic common to all milk – we have termed it “milkiness” (the combination of water, calcium, particular vitamins etc. that are common to all milks) – and it is measured in litres. If a consumer were to buy one litre of completely fat free, conventionally produced milk (were such a product available in this market) this is the characteristic they would consume. Fat content is measured in centilitres per litre and is linear with respect to market goods: a litre of skimmed and a litre of semi-skimmed together produce 1.6 cl’s of fat by volume<sup>11</sup>. The interpretation of the organic/conventional characteristic is less straightforward. It is measured by an indicator taking the value 1 if the milk is produced under organic conditions, and 0 otherwise. Whatever the organic characteristic represents (perhaps a warm glow, perhaps the absence of antibiotic residues in the milk) this specification of the technology says that twice as much of it is produced by buying 2 litres of milk, than is produced by 1 litre; which is not to say, however, that the consumer then values it twice as much since the utility function is concave in attributes.

## 4.2 Results.

### 4.2.1 Rationalisability.

**Missing prices.** In these data we can only construct unit prices for the purchases which take place. As already discussed, the two ways of dealing with this are either to impute prices and apply the test described in Theorem 1 or to apply the weaker test in Theorem 2 which only requires

<sup>11</sup>A less restrictive characteristics model would allow for differences in the taste of milk according to the fat content. In that case full fat is not simply a linear combination of skimmed and semi-skimmed and we would have four characteristics (skimmed, semi-skimmed, full-fat and organic) rather than the three we have taken. Other more exotic mappings from market goods to characteristics are clearly conceivable but we are interested in seeing how well this very simple approach does in explaining the data.

	$q$ -rationalisation	$z$ -rationalisation
<b>Pass</b>	2341 (94%)	1891 (76%)
<b>Fail</b>	159 (6%)	609 (24%)

Table 4: Test results, with missing prices

the observed prices. We present the results for both tests beginning with the weaker test (since if a household fails this then they obviously cannot satisfy Theorem 1 however we were to impute the missing prices). For households satisfying the conditions we present the marginal valuations of characteristics we are able to recover using our methods. We then impute the missing prices using region/month cell medians from observed unit prices and re-test applying the conditions in Theorem 1. Note that in carrying out these test we consider each household separately so that we allow for complete heterogeneity amongst households. This is with respect to whether or not an individual household’s behaviour is  $z$ -rationalisable at all; the form of their preferences if they indeed are rationalisable and their marginal valuations of the characteristics.

To begin with we take the time series of price and quantity data for an individual household and the 3-factor technology matrix and apply phase one of the linear programme defined by condition (L) in Theorem 2. We record whether a feasible solution exists. We then take the data for the next household and repeat the exercise. We also apply the test of  $q$ -rationalisation in which we replace  $\mathbf{A}$  with an  $\mathbf{I}_K$  matrix. The results are given in Table 4. We find that for the great majority of the sample households there exist suitable virtual prices and utility functions which  $q$ -rationalise their observed behaviour. We also find that for three-quarters of the sample (and 81% of those who are  $q$ -rationalisable) behaviour is also  $z$ -rationalisable.

We also investigate whether we can explain the pattern of pass/fail using some of the observable characteristics of the survey households. The controls were: dummy variables for single person households, households composed of couples, the presence of children, the presence of pre-school age children, the main shopper being male, the presence of retired people and living in Copenhagen and continuous variables for the total volume of milk purchased by the household and the mean age of the adults in the household. In a linear regression of a  $q$ -rationalisability indicator on these variables and a constant, only the male shopper dummy was significant (positively). For the same regression for  $z$ -rationalisability the dummies for single person households, retired household members and the purchase volume were all positively significant. The pre-school children dummy was negatively significant. However, the overall ability of these control variables to fit the patterns of passes and failure in the data was very weak. The  $R^2$  of linear regressions of  $q$ -rationalisability and  $z$ -rationalisability indicators on these controls were 0.0075 and 0.032 respectively. We conclude that the pattern of pass/fail is largely unexplained by observables.

The linear programme test only indicates the existence of Afriat numbers and vectors of marginal valuations of characteristics which satisfy the conditions for the model; it does not recover them uniquely. However, for many households which are rationalisable we are able to recover unique solutions for the  $\boldsymbol{\pi}$ -vectors. For these households (1385 in all) we solve for  $\boldsymbol{\pi}_t$  in each period where a solution can be obtained<sup>12</sup>. Table 5 reports some descriptive statistics for the marginal valuations of each characteristic, and Figure 3 illustrates their sample densities. It is important to note that these are no longer representative of the population since they are highly selected. For example, a household that only ever buys one type of milk will not have a unique shadow prices and consequently is not represented in these samples. To fully characterise the joint distributions of tastes would require partial identification methods and would produce bounds rather than point estimates. On average, milkiness has a valuation of 5.10 DKK per litre with each additional cl of fat content increasing the valuation by 0.43 DKK and organic attracting a premium of about 1 DKK. The valuations of the milkiness and the organic characteristic are evidently highly heterogeneous, whilst the valuations of fat are less so. Milkiness always has a positive marginal valuation but there are some negative valuations for the other characteristics: just under 10% of valuations of fat

<sup>12</sup>Note that we may not be able to solve for all of the elements of  $\boldsymbol{\pi}_t$  in every case as this will depend on the standard rank/order requirements for the simultaneous equation systems  $\mathbf{p}_t^+ = \mathbf{A}_t^+ \boldsymbol{\pi}_t$ .

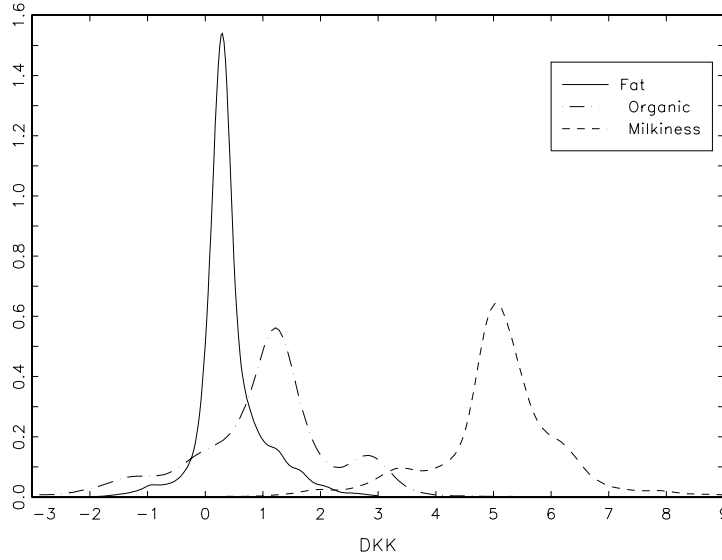


Figure 3: The distributions of marginal valuations of characteristics.

	Mean	Std Dev.	10 <sup>th</sup>	Percentile Points				n
				25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	
Milkiness	5.10	1.24	3.53	4.80	5.07	5.56	6.27	6829
Fat	0.43	0.59	0.00	0.19	0.32	0.56	1.16	4965
Organic	0.99	1.26	-0.63	0.46	1.20	1.53	2.72	3121

Table 5: Descriptive statistics of recovered marginal valuations, DKK

recovered are negative and nearly 17% of organic are negative. These negative marginal valuations do not conflict with the theory and are not inconsistent with the idea that the consumption of market goods has non-negative marginal utility overall.

For some households which buy a sufficiently wide range of goods we are able simultaneously to recover their valuations of all of the characteristics. Two dimensions (fat, organic) are illustrated in Figure 4; note that once again a household may appear several times in this distribution. There appear to be three distinct modes: one small group (top left) shows a high fat valuation and a low valuation for organic; a larger group (top right) has a high valuation of organic and a moderate one for fat; a large middle group which values organic but not fat. It seems that households are fairly stable with respect to which of these groups they principally belong with 77% of households appearing in only one of these groups, 20% appearing in two groups (most often the big low fat-moderate organic and the top right moderate fat-moderate organic groups) and only 3% of households having valuations which appear in all three.

**Imputed prices.** We now impute the missing prices using the medians of region/time period cells and test for  $q$ -rationalisation and  $z$ -rationalisation using Theorem 1. With the missing data filled in this manner, we first test (using Afriat’s Theorem) for  $q$ -rationalisation for each individual household to see whether there exists an admissible utility function  $u(\mathbf{q})$  defined over products which rationalises their behaviour. For those that fail we know that  $z$ -rationalisation is out of the question. For those where a suitable  $u(\mathbf{q})$  exists we apply the conditions in Theorem 1. From Table 6 we see that using the imputed prices, 71% of the sample are  $q$ -rationalisable: the rest are not and since (taking the imputed prices to be correct)  $q$ -rationalisation is a necessary condition for  $z$ -rationalisation, these households cannot satisfy the conditions in Theorem 1. Setting them aside we then test the remaining 1766 households. We find that, using the imputed prices, 60% of

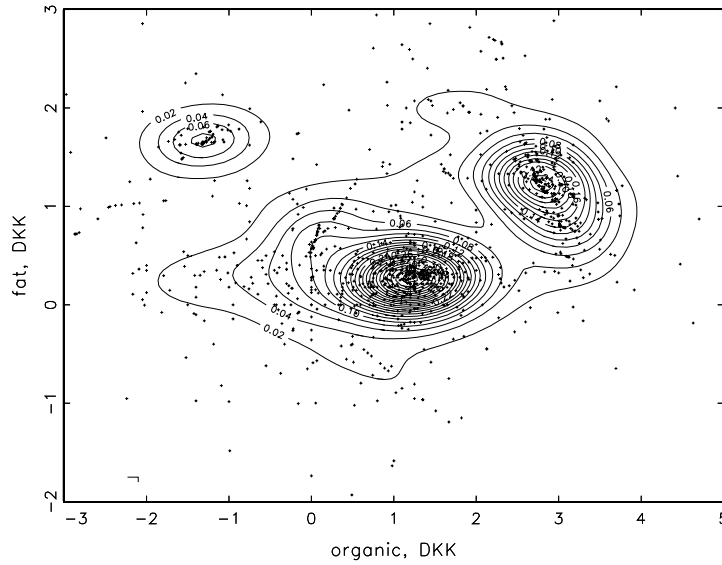


Figure 4: The joint distribution of the marginal valuations of fat and organic.

	<i>q</i> -rationalisation	<i>z</i> -rationalisation
<b>Pass</b>	1766 (71%)	1498 (60%)
<b>Fail</b>	734 (29%)	1002 (40%)

Table 6: Test results with imputed prices

the sample are *z*-rationalisable (this represents 84% of those who satisfy the prerequisite of being *q*-rationalisable). As before, we examined the ability of observable household characteristics to predict the pattern of passes/failures in the data. Once more, though individual controls were statistically significant<sup>13</sup> the overall  $R^2$  for *q*-rationalisability and *z*-rationalisability were extremely low: 0.0402 and 0.0325 respectively. Once more we conclude that the pattern of passes/failures is essentially random with respect to household observables.

#### 4.2.2 Latent characteristics.

The original test results (see Table 4) indicated that the behaviour of 18% of the sample was *q*-rationalisable, but not *z*-rationalisable. The question arises: how many additional unobserved characteristics would be necessary to *z*-rationalise their behaviour? The first point to note is that since  $K = 6$  and  $J = 3$  and these households are all *q*-rationalisable then we know that 3 additional characteristics must be sufficient. The substantive question is therefore: what is the *minimum* number of additional unobserved characteristics required? To establish this we ran the procedure described in section 3 and appendix 3, individually, for each of the 450 households reported in Table 4 which satisfied *q*-rationalisation, but failed *z*-rationalisation<sup>14</sup>. We ran the search initially for a single missing characteristic. For those households where a single unobserved characteristic was insufficient we re-ran the algorithm searching for a two missing characteristics. The results are reported in Table 7. The first two columns in Table 7 are identical to those in Table 4. These show that (testing in the presence of missing prices) 94% of the sample were *q*-rationalisable but only 76% were *z*-rationalisable with the measured characteristics. The next three columns show how the

<sup>13</sup>In the *q*-rationalisability regression the dummy variables for retired household members, male shoppers and the purchase volume measure were all positively significant, whilst age was negatively related. In the *z*-rationalisability regression the results were the same with the addition of the living in Copenhagen dummy variable which was negatively significant.

<sup>14</sup>We used the `fminsearch` routine in MatLab 7.04 (R14). Details of the search problem are given in Appendix 3.

	<i>q</i> -rationalisation	<i>z</i> -rationalisation.			
		# unobserved characteristics			
		0	1	2	3
<b>Pass</b>	2341 (94%)	1891 (76%)	2204 (88%)	2236 (89%)	2341 (94%)
<b>Fail</b>	159 (6%)	609 (24%)	296 (12%)	264 (11%)	159 (6%)

Table 7: Unobserved characteristics, with missing prices

Prices (DKK/litre)	Conventional milk			Organic milk			
	3.5%	1.5%	0.1%	3.5%	1.5%	NEW 0.5%	0.1%
Mean	6.11	5.34	5.09	7.34	6.51	<b>5.92</b>	6.30
Median	6.00	5.30	5.00	7.31	6.50	<b>6.32</b>	6.27
Std. Dev.	0.62	0.56	0.31	0.38	0.33	<b>1.29</b>	0.23
Coeff of variation	0.10	0.10	0.06	0.05	0.05	<b>0.22</b>	0.04

Table 8: Descriptive statistics reservation prices for a new milk

pass rates change as we allow for unobserved characteristics: first one, then two, etc.. This shows that the pass rate increases to 88% if we allow for one unobserved characteristic, and to 89% if we allow for two. As noted above if we allow for three missing characteristics so that the number of characteristics is equal to the number of market goods, then all *q*-rationalisable households can be *z*-rationalised and the pass rate reaches the *q*-rationalisable level of 94%.

#### 4.2.3 A hypothetical new good.

Finally we present an illustration of the way in which a new good might be treated in this framework<sup>15</sup>. To investigate the possibility of using the recovered shadow prices of the characteristics to value a new good we have invented a hypothetical new product: an organic milk with 0.5% fat content. The linear technology for a litre of this new product is

$$\mathbf{A}^* = [1 \ 0.5 \ 1]$$

Using the shadow prices recovered above we are able to predict the reservation price of this hypothetical good from the linear pricing equation  $\mathbf{p}^* = \mathbf{A}^* \boldsymbol{\pi}$  using the  $\boldsymbol{\pi}$ -vectors recovered in section 4.2. The results are given in Table 8 which, for the sake of easy comparison, combines the results for the new good (second column from right, in bold type) with the observed prices of the existing goods drawn from Table 2. The distribution of reservation prices has a somewhat heavy lower tail which pulls the mean below the median price. However, we note that the median price of 6.32 DKK/litre is between the median prices of organic semi-skimmed milk (6.50 DKK/litre) and skimmed milk (6.27 DKK/litre) and is more expensive (by about 1.2 DKK to 1.3 DKK) than a similar price-point located between the prices of semi-skimmed and skimmed conventional milks. Given the evidence in Table 8 (and in Table 2) of an increasing fat-content/price gradient and an organic/conventional mark-up at the median, the predicted price for this new hypothetical product seems intuitively plausible.

## 5 Conclusions.

We have extended revealed preference methods to the analysis of consumer characteristics models. We have derived the necessary and sufficient empirical conditions under which data on the market behaviour of a price-taking consumer is nonparametrically consistent with the consumer characteristics model. We consider the case in which all prices are observed and the case in which only the

<sup>15</sup>We are grateful to a referee and the Editor for suggesting this experiment.

price of the purchased goods are observed. Where these conditions hold, we show how information may be recovered on individual consumer's marginal valuations of product attributes. Where the conditions fail we highlight the role which the introduction of unobserved product attributes can play in rationalising the data. We have implemented these ideas using a consumer panel data on the Danish milk market. Specifically we consider the purchase of six milk types using a long household panel. We find that most households who satisfy the conventional Afriat-Varian consistency conditions also satisfy those for a simple three characteristics model. We supply estimates of the distribution of tastes for the three characteristics.

# Appendices.

## Appendix 1: The Nonlinear Characteristics Model.

In all that follows we define  $\mathbf{z} \equiv \mathbf{F}(\mathbf{q})$  where  $\mathbf{F}(\mathbf{q})$  is increasing and concave. Given this we have the following proposition.

**Theorem** *The following statements are equivalent.*

(P) *there exists a non-satiated, continuous and concave utility function  $u(\mathbf{z})$  which  $z$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  for the technology  $\mathbf{z} = \mathbf{F}(\mathbf{q})$ .*

(A) *there exist  $2T$  scalars  $\{V_t, \lambda_t > 0\}_{t=1, \dots, T}$  and  $T$   $J$ -vectors  $\{\boldsymbol{\pi}_t\}_{t=1, \dots, T}$  such that*

$$V_s \leq V_t + \lambda_t \boldsymbol{\pi}'_t (\mathbf{z}_s - \mathbf{z}_t) \quad (\text{A1})$$

$$\boldsymbol{\pi}'_t (\mathbf{z}_s - \mathbf{z}_t) \leq \mathbf{p}'_t (\mathbf{q}_s - \mathbf{q}_t) \quad (\text{A2})$$

for all  $s, t = 1, \dots, T$ .

(G) *There exists  $T$   $J$ -vectors  $\{\boldsymbol{\pi}_t\}_{t=1, \dots, T}$  such that A2 holds and the data  $\{\boldsymbol{\pi}_t, \mathbf{z}_t\}_{t=1, \dots, T}$  pass GARP.*

### Proof

(P)  $\Rightarrow$  (A) : For convenience we shall take  $u(\cdot)$  to be differentiable; if we do not do this then we have everywhere to replace gradients with tangents. By concavity of  $u(\mathbf{z})$  we have  $u(\mathbf{z}_s) \leq u(\mathbf{z}_t) + \nabla u(\mathbf{z}_t)' (\mathbf{z}_s - \mathbf{z}_t)$  and optimising behaviour implies  $\mathbf{p}_t \geq \nabla \mathbf{F}(\mathbf{q}_t) \boldsymbol{\pi}_t$  with equality when  $q_t^k > 0$  and where  $\lambda_t \boldsymbol{\pi}_t = \nabla u(\mathbf{z}_t)$  by definition (Gorman (1956)).

. Now consider the concavity condition for the technology and pre-multiply by  $\boldsymbol{\pi}_t \geq 0$

$$\begin{aligned} \mathbf{z}_s &\leq \mathbf{z}_t + \nabla \mathbf{F}(\mathbf{q}_t)' (\mathbf{q}_s - \mathbf{q}_t) \\ \Rightarrow \boldsymbol{\pi}'_t \mathbf{z}_s &\leq \boldsymbol{\pi}'_t \mathbf{z}_t + \boldsymbol{\pi}'_t \nabla \mathbf{F}(\mathbf{q}_t)' (\mathbf{q}_s - \mathbf{q}_t) \end{aligned}$$

for all  $s, t = 1, \dots, T$ . Substituting in  $\mathbf{p}'_t \geq \boldsymbol{\pi}'_t \nabla \mathbf{F}(\mathbf{q}_t)'$  preserves the inequality giving

$$\boldsymbol{\pi}'_t \mathbf{z}_s - \mathbf{p}'_t \mathbf{q}_s \leq \boldsymbol{\pi}'_t \mathbf{z}_t - \mathbf{p}'_t \mathbf{q}_t$$

which is the weak axiom of profit maximisation, WAPM (Varian (1984)). Re-arranging gives (A2).

(A)  $\Rightarrow$  (P) : From condition (A2) we have  $\boldsymbol{\pi}'_t \mathbf{z}_t - \mathbf{p}'_t \mathbf{q}_t \geq \boldsymbol{\pi}'_t \mathbf{z}_s - \mathbf{p}'_t \mathbf{q}_s$ . Hence  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}_s \Rightarrow \boldsymbol{\pi}'_t \mathbf{z}_t \geq \boldsymbol{\pi}'_t \mathbf{z}_s$ . Condition (A1) is, by Afriat's Theorem, equivalent to the existence of a concave, continuous utility function  $u(\mathbf{z})$  such that for any  $\mathbf{z}$  with  $\boldsymbol{\pi}'_t \mathbf{z}_t \geq \boldsymbol{\pi}'_t \mathbf{z}$  it is the case that  $u(\mathbf{z}_t) \geq u(\mathbf{z})$  (that is to say, condition (A1) means that there exists some  $u(\mathbf{z})$  which  $q$ -rationalises the data  $\{\boldsymbol{\pi}_t, \mathbf{z}_t\}_{t=1, \dots, T}$ ). Combining these results we have that for any  $\mathbf{z} = \mathbf{F}(\mathbf{q})$  such that  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}$  then  $\boldsymbol{\pi}'_t \mathbf{z}_t \geq \boldsymbol{\pi}'_t \mathbf{z}$  and there exists a suitable utility function  $u(\mathbf{z})$  such that  $u(\mathbf{z}_t) \geq u(\mathbf{z})$ . Hence there exists a concave, continuous utility function  $u(\mathbf{z})$  with  $\mathbf{z} = \mathbf{F}(\mathbf{q})$  which  $z$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t, \mathbf{z}_t\}_{t=1, \dots, T}$ .

(G)  $\iff$  (A) : Given some  $\boldsymbol{\pi}_t$  satisfying condition (A2), the equivalence between conditions (A1) and GARP on the data  $\{\boldsymbol{\pi}_t, \mathbf{z}_t\}_{t=1, \dots, T}$  follows from, for example, Theorem 1 in Varian (1983) by interpreting the  $\boldsymbol{\pi}_t$  vector as the price vector for the characteristics. ■

## 5.1 Appendix 2: Proofs of Theorems 1 and 2.

### Proof of Theorem 1

$(P) \Rightarrow (A)$  : By concavity of  $v(\mathbf{z})$  we have  $v(\mathbf{z}_s) \leq v(\mathbf{z}_t) + \nabla v(\mathbf{z}_t)'(\mathbf{z}_s - \mathbf{z}_t)$  and optimising behaviour implies  $\mathbf{p}'_t \geq \pi'_t \mathbf{A}'$  with equality when  $q_t^k > 0$  where  $\lambda_t \pi_t = \nabla v(\mathbf{z}_t)$  by definition. Therefore we have numbers  $\{V_t, \lambda_t > 0, \pi_t\}_{t=1, \dots, T}$  (where  $\lambda_t > 0$  follows from non-satiation with respect to market goods) which satisfy  $V_s \leq V_t + \lambda_t \pi'_t (\mathbf{z}_s - \mathbf{z}_t)$  for all  $s, t$ .

$(A) \Rightarrow (P)$  : From condition (L2) we have  $\mathbf{p}'_t \mathbf{q}_t = \pi'_t \mathbf{A}' \mathbf{q}_t = \pi'_t \mathbf{z}_t$ , since the locations in  $\mathbf{p}_t$  where the prices are greater than the corresponding locations in  $\pi'_t \mathbf{A}'$  occur only where the corresponding locations in  $\mathbf{q}_t$  are zero<sup>16</sup>. The linear structure, however, ensures that  $\mathbf{p}'_t \mathbf{q} \geq \pi'_t \mathbf{A}' \mathbf{q} = \pi'_t \mathbf{z}$ . Hence (A2) gives  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q} \Rightarrow \pi'_t \mathbf{z}_t \geq \pi'_t \mathbf{z}$ . Condition (A1) is, by Afriat's Theorem, equivalent to the existence of a concave, continuous utility function  $v(\mathbf{z})$  such that for any  $\mathbf{z}$  with  $\pi'_t \mathbf{z}_t \geq \pi'_t \mathbf{z}$  it is the case that  $v(\mathbf{z}_t) \geq v(\mathbf{z})$  (that is to say, condition (A1) means that there exists some  $v(\mathbf{z})$  which  $q$ -rationalises the data  $\{\pi_t, \mathbf{z}_t\}_{t=1, \dots, T}$ ). Combining these results we have that for any  $\mathbf{z} = \mathbf{A}' \mathbf{q}$  such that  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}$  then  $\pi'_t \mathbf{z}_t \geq \pi'_t \mathbf{z}$  and there exists a suitable utility function  $v(\mathbf{z})$  such that  $v(\mathbf{z}_t) \geq v(\mathbf{z})$ . Hence there exists a concave, continuous utility function  $v(\mathbf{z})$  with  $\mathbf{z} = \mathbf{A}' \mathbf{q}$  which  $z$ -rationalises the data  $\{\mathbf{p}_t, \mathbf{q}_t, \mathbf{z}_t\}_{t=1, \dots, T}$ .

$(L) \iff (A)$ . Given  $\{V_t, \lambda_t > 0, \pi_t\}_{t=1, \dots, T}$  which satisfy (A) we can then simply normalise on  $\lambda^{\min} = \min_t \{\lambda_t\}_{t=1, \dots, T}$  and define  $U_t = V_t / \lambda^{\min}$ ,  $\rho_t = \lambda_t / \lambda^{\min}$ ,  $\sigma_t^j = \lambda_t \pi_t^j / \lambda^{\min}$ . We then have  $\{U_t, \rho_t \geq 1, \sigma_t\}_{t=1, \dots, T}$  satisfying (L) as required. Conversely, given  $\{U_t, \rho_t \geq 1, \sigma_t\}_{t=1, \dots, T}$  satisfying (L) we can divide (L1) to (L3) by the positive constant  $\rho_t$  and defining  $V_t = U_t / \rho_t$ ,  $\lambda_t = \rho_t^{-1}$  and  $\pi_t^j = \sigma_t^j / \rho_t$  we have condition (A).

$(A) \iff (G)$  : Given some  $\pi_t$  satisfying condition (A), the equivalence between condition (A) and GARP on the data  $\{\pi_t, \mathbf{A}' \mathbf{q}_t\}_{t=1, \dots, T}$  follows from, for example, Theorem 1 in Varian (1983) by interpreting the  $\pi_t$  vector as the price vector for the characteristics. ■

### Proof of Theorem 2.

Obvious from comparison with Theorem 1 whilst noting that given  $\pi_t$  such that  $\mathbf{p}_t^+ = \mathbf{A}_t^+ \pi_t$  for all  $t = 1, \dots, T$ , then we can simply use them to construct the unobserved prices from  $\mathbf{p}_t^0 = \mathbf{A}_t^0 \pi_t$ . ■

<sup>16</sup>Linearity is over-sufficient in this sense as  $\mathbf{p}'_t \mathbf{q}_t = \pi'_t \mathbf{z}_t$  would also result from  $\mathbf{z}(\mathbf{q})$  being a nonlinear homogenous of degree one function.

### Appendix 3: Finding a latent characteristic.

To provide the constraint we use the idea of Afriat Efficiency (see Afriat (1973)). A GARP-type test can be interpreted as a test of two sub-hypotheses: (1) the consumer has rational preferences, and (2) the consumer is an efficient programmer. If the data violates revealed preference tests then Afriat (1973) suggested modifying (2) whilst maintaining (1). He suggests a form of partial efficiency, and introduces the efficiency parameter  $e$  where  $0 \leq e \leq 1$ . The consumer is now allowed to waste a fraction  $(1 - e)$  of their budget through optimisation error. This is done by modifying the revealed preference relation  $R^0$  to:

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

This efficiency concept can be used to define a weaker consistency test:

$$GARP(e) : \mathbf{q}_s R_e \mathbf{q}_t \Rightarrow \text{Not } \mathbf{q}_t P_e^0 \mathbf{q}_s$$

where “Not  $\mathbf{q}_t P_e^0 \mathbf{q}_s$ ”  $\equiv e \mathbf{p}'_t \mathbf{q}_t \leq \mathbf{p}'_t \mathbf{q}_s$  and where  $R_e$  denotes the transitive closure of  $R_e^0$ . If  $e = 1$  then  $GARP(e)$  is equivalent to  $GARP$ . If  $e = 0$  then there is no restriction on behaviour. To provide the constraint in this problem we use a simple numerical search algorithm which finds the largest value of  $e$  such that a given data set satisfies  $GARP(e)$ . We call this maximum Afriat Efficiency function,  $emax(\cdot)$ , which maps from the observed data to  $[0, 1]$ : a value of 1 means that the data passes GARP and all data passes GARP for a value of zero. The following problem will have a solution at zero if there exists a matrix  $\mathbf{B}$  which satisfies  $z$ -rationalisation

**Problem**  $\min_{\mathbf{B}} f(\mathbf{B}) = (1 - emax(\mathbf{\Pi}, \mathbf{Z}))$  where the  $t$ th column of  $\mathbf{\Pi}$  is  $\boldsymbol{\pi}_t = [\mathbf{A}_t^+ \sim \mathbf{B}_t^+]^{-1} \mathbf{p}_t^+$  and  $\mathbf{Z} = [\mathbf{A} \sim \mathbf{B}]' \mathbf{Q}$  where  $\mathbf{Q}$  is the  $(K \times T)$  matrix of horizontally concatenated quantity vectors.

This algorithm is straightforward to implement using standard nonlinear optimisation methods (we used the `fminsearch` routine in MatLab 7.04 (R14)). It starts with an initial guess at  $\mathbf{B}$ , computes  $(\mathbf{\Pi}, \mathbf{Z})$  as described and then computes the Afriat efficiency parameter at which the data  $(\mathbf{\Pi}, \mathbf{Z})$  pass GARP. If the data  $(\mathbf{\Pi}, \mathbf{Z})$  fail GARP then  $emax(\mathbf{\Pi}, \mathbf{Z}) < 1$  and the objective function  $f(\mathbf{B}) > 0$ . The parameter matrix  $\mathbf{B}$  is then altered on the basis of numerically computed gradients  $\nabla f(\mathbf{B})$  in order to attempt to further minimise  $f(\mathbf{B})$ . If the problem has a solution at  $f(\mathbf{B}) = 0$ , then at these values of  $\mathbf{B}$  we know that  $emax(\mathbf{\Pi}, \mathbf{Z}) = 1$  which means that the data  $(\mathbf{\Pi}, \mathbf{Z})$  pass GARP (condition (G) in Theorem 1) and hence that  $\{\mathbf{p}_t^+, \mathbf{q}_t\}_{t=1, \dots, T}$  and  $[\mathbf{A} \sim \mathbf{B}]$  are  $z$ -rationalisable.

## References

- [1] Afriat, S.N. (1967), "The construction of a utility function from expenditure data", *International Economic Review*, **8**, 76-77.
- [2] Bajari, Patrick and Lanier Benkard (2004), "Demand estimation with heterogeneous consumers and unobserved product characteristics: a hedonic approach", NBER Working Paper 10278.
- [3] Berry, S., J. Levinsohn, and A. Pakes, (1995), "Automobile prices in market equilibrium," *Econometrica*, **63**, 841-990.
- [4] Blundell, R., M. Browning and I. Crawford (2003), "Nonparametric Engel curves and revealed preference", *Econometrica*, 71(1), 205-240.
- [5] Blundell, R. and J-M Robin (2000), "Latent Separability: Grouping Goods without Weak Separability", *Econometrica*, 68(1), 53-84.
- [6] Browning, Martin (1983), "Necessary and sufficient conditions for conditional cost functions", *Econometrica*, 51(3), 851-856.
- [7] Crawford I. (2004), "Necessary and sufficient conditions for latent separability", cemap working paper CWP02/04.
- [8] Diewert, W.E. (1973), "Afriat and revealed preference theory", *Review of Economic Studies*, **40**, 419-426.
- [9] Diewert, E. and C. Parkan, (1978), "Tests for consistency of consumer data and nonparametric index numbers", University of British Columbia, Working Paper 78-27.
- [10] Gorman, W. M. (1956), "A possible procedure for analysing quality differentials in the eggs market", in *Review of Economic Studies*, (Oct 1980) **47**, 843-856.
- [11] Goodman, Allen (1998), "Andrew Court and the invention of hedonic price analysis", *Journal of Urban Economics*, 44, 291-298.
- [12] Griliches, Z. (1971), *Price Indexes and Quality Change*, Washington: Federal Reserve Board.
- [13] Hanoch, G. and M. Rothschild, (1972), "Testing the assumptions of production theory: a nonparametric approach", *Journal of Political Economy*, **80**, 256-275.
- [14] Heckman, J., and J. Scheinkman, (1987), "The Importance of Bundling in a Gorman-Lancaster Model of Earning" *Review of Economic Studies*, **54**, 243-55
- [15] Hicks, John (1940), "The valuation of the social income", *Economica*, 7, 105-124.
- [16] Houthakker, H. S. (1950), "Revealed preference and the utility function", *Economica*, May, 159-174.
- [17] Lancaster, K. (1966), "A new approach to consumer theory", *Journal of Political Economy*, **74**, 2, 132-157.
- [18] Markowitz, H. (1959), *Portfolio Selection: Efficient Diversification of Investments*, New York: Wiley.
- [19] Muellbauer, J. (1974), "Household production theory, quality and the 'hedonic technique'", *American Economic Review*, **64**, 977-994.
- [20] Neary, Peter and Kevin Roberts (1980), "The theory of household behaviour under rationing", *European Economic Review*, 13, 25-42.
- [21] Pudney, S. (1981) "Instrumental Variable Estimation of a Characteristics Model of Demand", *Review of Economic Studies*, **48**, 417-433

- [22] Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, **82**, 34-55.
- [23] Samuelson, P. (1948), "Consumption theory in terms of revealed preference", *Economica*, November, 243-253.
- [24] Stone, R. (1956), *Quantity and Price Indexes in National Accounts*, Paris: Organisation for European Economic Cooperation.
- [25] Tinbergen, J. (1959), "On the Theory of Income Distribution," in *Selected Papers of Jan Tinbergen*, edited by L. H. Klaassen, L. M. Koyck, and H. J. Witteveen, Amsterdam: North-Holland.
- [26] Varian, H. (1982), "The nonparametric approach to demand analysis", *Econometrica*, **50**, 945-973.
- [27] Varian, H. (1983), "Nonparametric tests of consumer behaviour", *Review of Economic Studies*, **50**, 99-110.
- [28] Varian, H. (1984), "The nonparametric approach to production analysis", *Econometrica*, **52**, 579-597.
- [29] Varian, H. (1988), "Revealed preference with a subset of goods". *Journal of Economic Theory*, **46**(1), 179-185.