

Uni**S**

**University of Surrey**



**Discussion Papers in Economics**

REVEALED PREFERENCE ANALYSIS OF  
CHARACTERISTICS MODELS

By

Laura Blow

(Institute for Fiscal Studies)

Martin Browning

(University of Copenhagen)

&

Ian Crawford

(University of Surrey & IFS)

DP 03/05

Department of Economics  
University of Surrey  
Guildford  
Surrey GU2 7XH, UK  
Telephone +44 (0)1483 689380  
Facsimile +44 (0)1483 689548  
Web [www.econ.surrey.ac.uk](http://www.econ.surrey.ac.uk)

# REVEALED PREFERENCE ANALYSIS OF CHARACTERISTICS MODELS.

Laura Blow

Institute for Fiscal Studies,  
London

Martin Browning

CAM, Institute of Economics,  
University of Copenhagen  
Copenhagen,  
Denmark

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 Laura Anderson, Lars Nesheim, Ian Preston and participants in 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: Martin.Browning@econ.ku.dk

# 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. 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>1</sup> (non-parametric) 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. Where the conditions fail we highlight the role which the introduction of unobserved product attributes 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 gauge how much to produce. In this 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.

## 2 Revealed preference conditions for the characteristics model

### 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

---

<sup>1</sup>See Samuelson (1948), Houthakker (1950), Afriat (1967), Hanoch and Rothschild (1972), Diewert (1973), Diewert and Parkan (1978) and Varian (1982, 1983, 1984).

$\mathbf{p}_t$ . 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  $u(\mathbf{q})$   $q$ -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:

**Afriat's Theorem**<sup>2</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  $\mathbf{z} = \mathbf{F}(\mathbf{q})$ . 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:

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

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 (1954) 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  $(K \times J)$  matrix of constants with  $J \leq K$  and  $\mathbf{A}$  has full column rank. 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.

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:

<sup>2</sup>Afriat (1967), Diewert (1973), Varian (1982).

**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 plugged into the consumer's maximisation problem set out above, then it would perfectly generate the observed data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  for the posited technology  $\mathbf{A}$ .

For a good,  $k$ , 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:

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

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

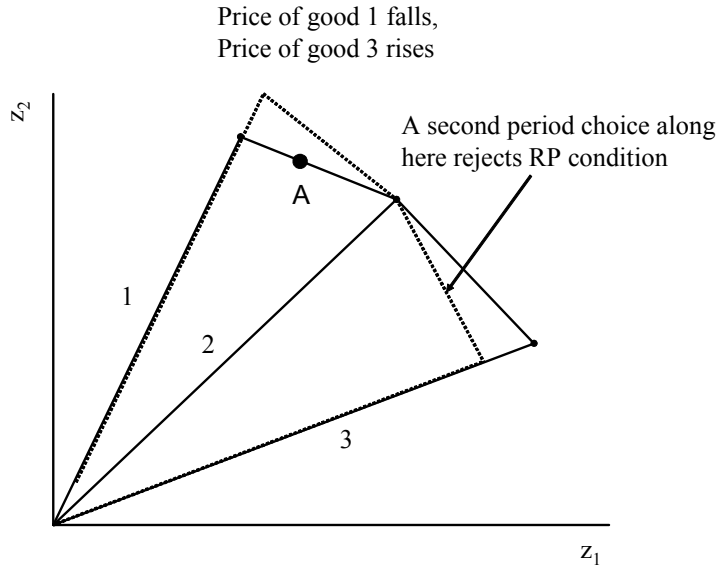
$$\boldsymbol{\pi}_t = (\lambda_t)^{-1} \nabla v(\mathbf{z}_t)$$

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.

FIGURE 1. Revealed preference for characteristics



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<sup>3</sup>.

**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  $z$ -rationalises 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*

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

$$\mathbf{p}_t \geq \mathbf{A} \boldsymbol{\pi}_t, \quad \forall t \quad (A2)$$

$$(\mathbf{p}_t - \mathbf{A} \boldsymbol{\pi}_t) \odot \mathbf{q}_t = \mathbf{0}, \quad \forall t \quad (A3)$$

(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*

$$W_s \leq W_t + \boldsymbol{\sigma}'_t (\mathbf{A}' \mathbf{q}_s - \mathbf{A}' \mathbf{q}_t), \quad \forall s, t \quad (L1)$$

$$\rho_t \mathbf{p}_t \geq \mathbf{A} \boldsymbol{\sigma}_t, \quad \forall t \quad (L2)$$

$$(\rho_t \mathbf{p}_t - \mathbf{A} \boldsymbol{\sigma}_t) \odot \mathbf{q}_t = \mathbf{0}, \quad \forall t \quad (L3)$$

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

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>4</sup>. Condition (A) involves both non-linear functions of unknowns (the  $\lambda_t \boldsymbol{\pi}'_t$  terms) and strict inequality constraints on unknowns ( $\lambda_t > 0$ ). We have therefore included condition (L) since this condition is in the form of the restrictions in a linear programming problem. Consequently we can simply 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

<sup>3</sup>The symbol  $\odot$  denotes the Hardamard product (element-by-element multiplication) of two vectors or matrices of the same dimension; that is  $(a, b) \odot (c, d) = (ac, bd)$ .

<sup>4</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  $\boldsymbol{\pi}_t$  vectors may be recovered uniquely.

always give a result in a finite number of steps, for even moderate sized problems this quickly leads 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 a simple necessary condition for  $z$ -rationalisation; this will also be useful in the next subsections.

### 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>5</sup>. Any set of shadow price vectors that satisfy  $A2$  and  $A3$  must have:

$$\mathbf{p}_t^+ = \mathbf{A}_t^+ \boldsymbol{\pi}_t$$

A necessary condition<sup>6</sup> 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  $\mathbf{p}_t^+$ . If this rank condition holds then a solution for the shadow prices will be given by:

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

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<sup>7</sup> and even if they hold we may not be able to find shadow prices that satisfy:

$$\mathbf{p}_t^0 \geq \mathbf{A}_t^0 \boldsymbol{\pi}_t$$

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. Finally we note that since the characteristics model with  $J < K$  is just a special case of the standard model  $z$ -rationalisation implies  $q$ -rationalisation. Formally:

**Corollary 1** *If there exists a concave utility function  $v(\mathbf{z})$  with  $\mathbf{z} = \mathbf{A}'\mathbf{q}$  which  $z$ -rationalises the data then there exists a concave utility function  $u(\mathbf{q})$  which  $q$ -rationalises the data.*

To summarise, Theorem 1 provides necessary and sufficient conditions for  $z$ -rationalisation which are computationally feasible and indeed straightforward to check. If a given data set and technology matrix satisfies these conditions then there exists at least one utility function, defined over characteristics, which when plugged into the consumer's optimisation problem will generate the observed data without error.

<sup>5</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.

<sup>6</sup>See for example, Magnus and Neudecker, (1991), pp.36-37, Theorem 11.

<sup>7</sup>Another necessary, but not sufficient condition is that the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  satisfies GARP. The conditions in Theorem 1 imply this, but are not implied by it. See Corollary 1 below.

## 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 taking 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*

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

These conditions are identical to, for example, those in conditions (A) and (L) in Theorem 1, except that they do not involve the unobserved prices. 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 in this theorem, but not the other way around. If these conditions hold then we can take any set of implied  $\boldsymbol{\sigma}_t$ 's and  $\rho_t$ 's and simply set:

$$\mathbf{p}_t^0 = \rho_t \mathbf{A}_t^0 \boldsymbol{\sigma}_t$$

(which satisfies (L3) in Theorem 1). In this case the resulting  $\mathbf{p}_t^0$  vectors have the interpretation of being *virtual* market 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  $\boldsymbol{\sigma}_t$ 's and the  $\rho_t$  are not unique, these values will not be uniquely determined either. Note that we also have the corollary (paralleling Corollary 1) that if there exists a data consistent utility function defined over characteristics and virtual prices, then there also exists a data consistent utility function over market goods.



**Corollary 2** *If there exists a concave utility function  $v(\mathbf{z})$  with  $\mathbf{z} = \mathbf{A}'\mathbf{q}$  and virtual prices  $\{\mathbf{p}_t^0\}_{t=1,\dots,T}$  which  $z$ -rationalise the data  $\{\mathbf{p}_t^+, \mathbf{q}_t\}_{t=1,\dots,T}$  then there also exists a concave utility function  $u(\mathbf{q})$  and virtual prices  $\{\mathbf{p}_t^0\}_{t=1,\dots,T}$  which  $q$ -rationalise the data.*

To conclude this section we note that in the presence of missing prices the linear programming condition seems to have a clear advantage over the GARP condition. Firstly they still require the necessary condition  $\text{rank}(\mathbf{A}_t^+ \sim \mathbf{p}_t^+) = \text{rank}(\mathbf{A}_t^+)$ , so that we can reject  $z$ -rationalisation, even with missing prices should we find that  $\text{rank}(\mathbf{A}_t^+ \sim \mathbf{p}_t^+) > \text{rank}(\mathbf{A}_t^+)$ . Furthermore (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 the price of the  $k$ th good is not observed in period  $t$  because it wasn't purchased, but good  $k$  was purchased in period  $s$ . 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  $\mathbf{I}_K$ . This then provides 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.

### 3 Latent characteristics.

#### 3.1 All goods bought.

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 often not 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. Suppose we have some data in which all goods are bought in each period and all prices are observed. 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  $\mathbf{A}$  matrix which has rank less than  $K$  and does  $z$ -rationalise the data.<sup>8</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 unobserved characteristics that do  $z$ -rationalise the data in conjunction with the observed ones. Given a  $K \times H$  matrix  $\mathbf{B}$ , denote this augmented technology matrix by the  $K \times (J+T)$  matrix  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{B}$ . We denote the  $K \times T$  matrix of horizontally concatenated price vectors by  $\mathbf{P}$ . The following gives a useful result:

**Lemma 1** *The following statements are equivalent:*

- 1) *the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1,\dots,T}$  satisfy the conditions for  $q$ -rationalisation and  $\text{rank}(\tilde{\mathbf{A}} \sim \mathbf{P}) = \text{rank}(\tilde{\mathbf{A}})$  for all  $t$*
- 2) *the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1,\dots,T}$  and the augmented matrix  $\tilde{\mathbf{A}}$  satisfy the conditions for  $z$ -rationalisation.*

This is just a special case of Theorem 1 in which (A2) holds with equality (since all goods are bought). Given data which satisfies the  $q$ -rationalisation

<sup>8</sup>The restriction  $\text{rank}(\mathbf{A}) < K$  is because we can always find a 'trivial' characteristics model with  $\text{rank}(\mathbf{A}) = K$  by taking  $\mathbf{A} = \mathbf{I}_K$ .

conditions the issue is then one of finding suitable supplementary characteristics so that the rank condition holds.

**Theorem 3** *If the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  satisfy the conditions for  $q$ -rationalisation, but not the conditions for  $z$ -rationalisation given  $\mathbf{A}$ , then we can always find an augmented matrix  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{B}$  so that the data do  $z$ -rationalise given the augmented matrix  $\tilde{\mathbf{A}}$  where  $\mathbf{B}$  has rank greater than or equal to  $H \equiv \text{rank}(\mathbf{A} \sim \mathbf{P}) - \text{rank}(\mathbf{A})$ .*

The matrix  $\mathbf{A} \sim \mathbf{P}$  has  $J+T$  columns. If we have a large number of time periods relative to the number of market goods ( $T \geq K$ ) then we shall generally have  $\text{rank}(\mathbf{A} \sim \mathbf{P}) = K$  so that the number of latent characteristics,  $H$ , is equal to the number of goods minus the number of observed characteristics ( $K - J$ ). In this case any  $K \times H$  matrix  $\mathbf{B}$  that is orthogonal to  $\mathbf{A}$  will serve to  $z$ -rationalise the data with  $\mathbf{A} \sim \mathbf{B}$ . One way of finding  $\mathbf{B}$  is to regress  $\mathbf{P}$  on  $\mathbf{A}$ :

$$\mathbf{P} = \mathbf{A}\mathbf{\Pi} + \mathbf{E}$$

$(K \times T)$     $(K \times J)(J \times T)$     $(K \times T)$

Then take the estimated residual matrix  $\hat{\mathbf{E}}$  and find a  $K \times H$  matrix  $\mathbf{B}$  that forms a column basis for  $\mathbf{E}$ , which we can do using, for example, principal components.

Two other special cases of an augmented characteristics matrix, where  $\mathbf{B}$  is not restricted to be  $K \times H$ , have been used in the literature. The first option is to add a set of product-specific dummy characteristics to the existing set of characteristics as discussed in Gorman (1956) and Pudney (1981). This approach sets  $\mathbf{B} \equiv \mathbf{I}_K$  and  $\mathbf{\Pi}_B \equiv \mathbf{E}$ . Thus:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \sim \mathbf{I}_K \\ \mathbf{E} \end{bmatrix}_{(K \times [J+K])} \quad \tilde{\mathbf{\Pi}} = \begin{bmatrix} \mathbf{\Pi} \\ \mathbf{E} \end{bmatrix}_{([J+K] \times T)} \quad \text{and} \quad \tilde{\boldsymbol{\pi}}_t = \begin{bmatrix} \boldsymbol{\pi}_t \\ \mathbf{e}_t \end{bmatrix}_{([J+K] \times 1)}$$

where  $\tilde{\boldsymbol{\pi}}_t$  and  $\mathbf{e}_t$  represent column  $t$  of  $\tilde{\mathbf{\Pi}}$  and  $\mathbf{E}$  etc.). Obviously this augmented model achieves the required rank condition by greatly expanding the number of characteristics to  $J + K$ . Thus Gorman suggests that the new characteristics should be made ‘small’. The second representation of essentially the same model is to add a single time-varying unobserved characteristic to the  $\mathbf{A}$  matrix; this is widely used in the empirical IO literature<sup>9</sup>. This approach sets  $\mathbf{B} \equiv \mathbf{E}$  and  $\mathbf{\Pi}_B \equiv \mathbf{I}_T$ . Thus:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \sim \mathbf{E} \\ \mathbf{E} \end{bmatrix}_{(K \times [J+K])} \quad \tilde{\mathbf{\Pi}} = \begin{bmatrix} \mathbf{\Pi} \\ \mathbf{I}_T \end{bmatrix}_{([J+K] \times T)} \quad (3)$$

Alternatively, we can think of this as a time varying technology where

$$\mathbf{p}_t = \tilde{\mathbf{A}}_t \tilde{\boldsymbol{\pi}}_t$$

and

$$\tilde{\mathbf{A}}_t = \begin{bmatrix} \mathbf{A} \sim \mathbf{e}_t \\ \mathbf{E} \end{bmatrix}_{(K \times [J+1])} \quad \text{and} \quad \tilde{\boldsymbol{\pi}}_t = \begin{bmatrix} \boldsymbol{\pi}_t \\ 1 \end{bmatrix}_{([J+1] \times 1)}$$

<sup>9</sup>For example Berry, S., J. Levinsohn, and A. Pakes, (1995).

Obviously the informational content of these three approaches is identical, and only differ in the way  $\mathbf{E}$  is partitioned into additional characteristics and their marginal valuations. Note that  $\mathbf{E}$  itself is not identified by anything other than the initial method of choosing  $\mathbf{\Pi}$ . In the first we want to add as few extra characteristics as possible so we make  $\mathbf{B}$  a column basis for  $\mathbf{E}$ . Obviously  $\mathbf{B}$  is not unique. The second introduces  $K$  market-good-specific dummy characteristics with corresponding marginal valuations given by  $\mathbf{E}$ . The third introduces a new time-varying characteristic which has a marginal valuation of unity in every period.

The last two suggestions discussed above yield essentially ‘trivial’ characteristics models in the sense that they have at least as many characteristics as market goods. Only in the first may we be able to find a non-trivial characteristics model with fewer characteristics than goods would fit the data (i.e. where  $J + H < K$ ). So far we have assumed that the prices of all goods are observed, even the prices of goods which are not purchased. We now relax this assumption.

### 3.2 Zeros in some periods.

The simplest extension is ask whether there is *any* 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’. More generally, we could search for a  $(K \times H)$  matrix  $\mathbf{B}$  such that we can  $z$ -rationalise the data for  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{B}$ . As before, to give a non-trivial model we require  $J + H < K$ . Conceptually this is straightforward: we simply replace  $\mathbf{A}$  with  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{B}$  in 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. To provide the constraint we use the maximum Afriat Efficiency<sup>10</sup> function,  $emax(\cdot)$ , which maps from the observed data to  $[0, 1]$ . This is the value such the revealed preference conditions, allowing for some inefficiency in buying, satisfy GARP; 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

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

This algorithm is straightforward to implement using standard nonlinear optimisation methods. 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  $emax(\mathbf{\Pi}, \mathbf{Z}) < 1$  then  $f(\mathbf{B}) > 0$  and  $\mathbf{B}$  is then altered on the basis of numerically computed gradients with respect to the objective function in order to minimise  $f(\mathbf{B})$ . We have the following result.

<sup>10</sup>See Afriat. S. N. (1973) and the Appendix for a brief discussion.

**Theorem 4** *If Problem 1 has a solution at  $f(\mathbf{B}) = 0$  then there exists at least one suitable  $\mathbf{B}$  matrix such that the data  $\{\mathbf{p}_t, \mathbf{q}_t\}$  and the augmented technology matrix  $\tilde{\mathbf{A}} = \mathbf{A} \sim \mathbf{B}$  satisfy  $z$ -rationalisation.*

Finally we consider the case in which no characteristics are observed at all. This is equivalent to having an unknown transformation matrix  $\mathbf{A}$ . In such a case we may wish to test whether there is *any* linear characteristics model with a given rank  $J < K$  that could be  $z$ -rationalisable with some given data. The approach laid out in Problem 1 and Theorem 4 still applies - suitably adapted with  $\tilde{\mathbf{A}} = \mathbf{B}$  for example. Note, however, that even for medium sized data sets this a ‘large’ nonlinear problem which can take significant time.

## 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; this is done by the use of home scanners which record price, quantity, 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>11</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.

---

<sup>11</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.

TABLE 1. Market shares

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 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 1.20 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.

TABLE 2. Prices

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

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 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:

TABLE 3. Market goods and characteristics

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	
$K = 6$	$J = 3$

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 baseline characteristic common to all milk – we have termed it “miliness” (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 liter 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>12</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 Missing prices.

In these data we can only construct unit prices for the purchases which take place. As already discussed, the two options to deal 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 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.

---

<sup>12</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.

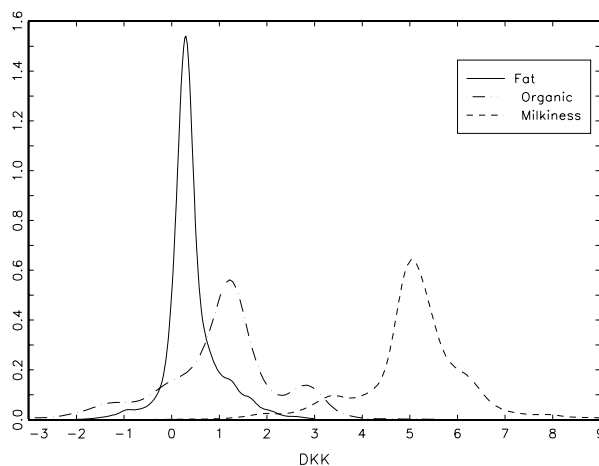
TABLE 4. Test results, with missing prices

	<i>q</i> -rationalisation		<i>z</i> -rationalisation	
	Freq.	Percent	Freq.	Percent
<b>Pass</b>	2341	94%	1891	76%
<b>Fail</b>	159	6%	609	24%
Total	2500	100%	2500	100%

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 passing/failure 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 patterns of passes/failures is largely unexplained by observables.

FIGURE 2. The densities of the distributions of marginal valuations of characteristics



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

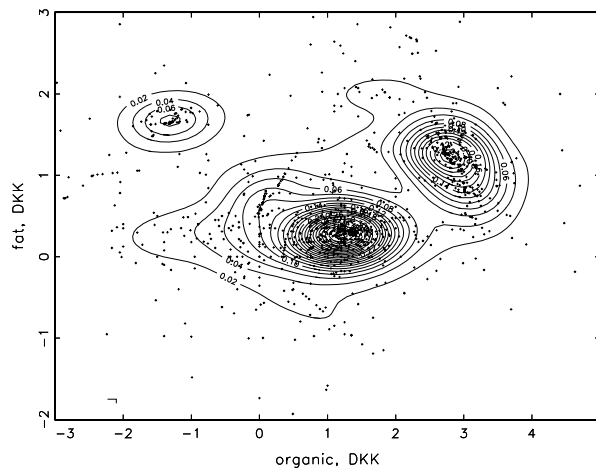
which are rationalisable we are able to recover unique solutions for the  $\pi$ -vectors. For these households (1385 in all) we solve for  $\pi_t$  in each period where a solution can be obtained<sup>13</sup>. Table 6 reports some descriptive statistics for the marginal valuations of each characteristic, and Figure 2 illustrates their sample densities. It is most 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.

TABLE 6. Descriptive statistics of recovered marginal valuations, DKK

	Mean	Std Dev.	Percentile Points					n
			10 <sup>th</sup>	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

On average, milkiness has a valuation of 5.10 DKK per litre with each additional cl of fat 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 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.

FIGURE 3. The joint density of the distribution of the marginal valuations of fat and organic



<sup>13</sup>Note that we may not be able to solve for all of the elements of  $\pi_t$  in every case as this will depend on the standard rank/order requirements for the simultaneous equation systems  $\mathbf{p}_t^+ = \mathbf{A}_t^+ \pi_t$ .



Two dimensions (fat,organic) are illustrated in Figure 3; 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.

#### 4.2.2 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 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.

TABLE 8. Test results, with imputed prices

	$q$ -rationalisation		$z$ -rationalisation	
	Freq.	Percent	Freq.	Percent
<b>Pass</b>	1766	71%	1498	60%
<b>Fail</b>	734	29%	1002	40%
<b>Total</b>	2500	100%	2500	100%

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 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>14</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.

## 5 Conclusions

We have extended revealed preference methods to the analysis of consumer characteristics models. We have derived the necessary and sufficient empirical condi-

<sup>14</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.

tions under which data on the market behaviour of heterogeneous, price-taking, individual consumers are nonparametrically consistent with the consumer characteristics model. We consider the case in which all prices are observed and the case in which only the 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.

## 6 Appendix.

### 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<sup>15</sup>. 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  $W_t = V_t / \lambda^{\min}$ ,  $\rho_t = \lambda_t / \lambda^{\min}$ ,  $\sigma_t^j = \lambda_t \pi_t^j / \lambda^{\min}$ . We then have  $\{W_t, \rho_t \geq 1, \sigma_t\}_{t=1, \dots, T}$  satisfying (L) as required. Conversely, given  $\{W_t, \rho_t \geq 1, \sigma_t\}_{t=1, \dots, T}$  satisfying (L) we can divide (L1) to (L3) by the positive constant  $\rho_t$  and defining  $V_t = W_t / \rho_t$ ,  $\lambda_t = \rho_t^{-1}$  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 Corollary 1

If there exists a function  $v(\mathbf{z})$  that  $z$ -rationalises  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$ , then  $v(\mathbf{A}' \mathbf{q}_t) \geq v(\mathbf{A}' \mathbf{q})$  for all  $\mathbf{q}$  such that  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}$ . Define  $u(\mathbf{q})$  by  $u(\mathbf{q}) = v(\mathbf{A}' \mathbf{q})$ . Then  $v(\mathbf{A}' \mathbf{q}_t) \geq v(\mathbf{A}' \mathbf{q})$  for all  $\mathbf{q}$  such that  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q} \implies u(\mathbf{q}_t) \geq u(\mathbf{q})$  for all  $\mathbf{q}$  such that  $\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}$ . Therefore  $u(\mathbf{q})$   $q$ -rationalises the data. If  $v(\mathbf{A}' \mathbf{q})$  is concave then

$$\begin{aligned} \mu v(\mathbf{A}' \mathbf{q}_t) + (1 - \mu) v(\mathbf{A}' \mathbf{q}_s) &= \mu u(\mathbf{q}_t) + (1 - \mu) u(\mathbf{q}_s) \\ \mu u(\mathbf{q}_t) + (1 - \mu) u(\mathbf{q}_s) &\leq v(\mu \mathbf{A}' \mathbf{q}_t + (1 - \mu) \mathbf{A}' \mathbf{q}_s) \end{aligned}$$

But

$$\begin{aligned} v(\mu \mathbf{A}' \mathbf{q}_t + (1 - \mu) \mathbf{A}' \mathbf{q}_s) &= v(\mathbf{A}' (\mu \mathbf{q}_t + (1 - \mu) \mathbf{q}_s)) \\ &= u(\mu \mathbf{q}_t + (1 - \mu) \mathbf{q}_s) \end{aligned}$$

<sup>15</sup>Gorman (1956) p. 843.

<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.

Hence  $\mu v(\mathbf{A}'\mathbf{q}_t) + (1 - \mu)v(\mathbf{A}'\mathbf{q}_s) \leq u(\mu\mathbf{q}_t + (1 - \mu)\mathbf{q}_s) \implies u(\mathbf{q})$  is concave. ■

**Proof of Theorem 2.**

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

**Proof of Corollary 2.**

Analogous to Corollary 1. ■

**Proof of Lemma 1.**

By Afriat's Theorem GARP for the data  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  is equivalent to the existence of numbers  $\{V_t, \lambda_t > 0\}_{t=1, \dots, T}$  such that

$$V_s \leq V_t + \lambda_t \mathbf{p}'_t (\mathbf{q}_s - \mathbf{q}_t) \quad \text{for all } s, t = 1, \dots, T$$

The rank condition implies the existence of vectors  $\{\boldsymbol{\pi}_t\}_{t=1, \dots, T}$  such that

$$\mathbf{p}'_t = \boldsymbol{\pi}'_t \tilde{\mathbf{A}}'$$

for all  $t$ . Substituting in means that there exists numbers  $\{V_t, \lambda_t > 0\}_{t=1, \dots, T}$  such that

$$\begin{aligned} V_s &\leq V_t + \lambda_t \boldsymbol{\pi}'_t \tilde{\mathbf{A}}' (\mathbf{q}_s - \mathbf{q}_t) \quad \text{for all } s, t = 1, \dots, T \\ V_s &\leq V_t + \lambda_t \boldsymbol{\pi}'_t (\tilde{\mathbf{A}}' \mathbf{q}_s - \mathbf{A}' \mathbf{q}_t) \quad \text{for all } s, t = 1, \dots, T \end{aligned}$$

The rest of the proof is analogous to that for Theorem 1. ■

**Proof of Theorem 3.**

Let  $\boldsymbol{\Pi} = [\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_T]$  be a  $(J \times T)$  matrix. Let

$$\mathbf{E} = \mathbf{P} - \mathbf{A}\boldsymbol{\Pi}$$

$(K \times T) \quad (K \times T) \quad (K \times J)(J \times T)$

Then  $\text{rank}\{\mathbf{E}\} = \text{rank}(\mathbf{A}\boldsymbol{\Pi} \sim \mathbf{P}) - \text{rank}(\mathbf{A}\boldsymbol{\Pi}) \geq \text{rank}(\mathbf{A} \sim \mathbf{P}) - \text{rank}(\mathbf{A}) = H$ . Assume  $\boldsymbol{\Pi}$  is chosen so that  $\text{rank}\{\mathbf{E}\} = H$ . Thus  $\mathbf{E}$  can be written as  $\mathbf{B}\boldsymbol{\Pi}_B$  where  $\mathbf{B}$  is a  $K \times H$  matrix that forms a column basis for  $\mathbf{E}$  and  $\boldsymbol{\Pi}_B = (\mathbf{B}\mathbf{B}')^{-1} \mathbf{B}\mathbf{E}$ . So

$$\mathbf{P} = [\mathbf{A} \sim \mathbf{B}] \begin{bmatrix} \boldsymbol{\Pi} \\ \boldsymbol{\Pi}_B \end{bmatrix}$$

and  $\text{rank}\{\tilde{\mathbf{A}} \sim \mathbf{P}\} = \text{rank}\{\tilde{\mathbf{A}}\}$  where  $\tilde{\mathbf{A}}$  is the augmented  $\mathbf{A}$  matrix. Given this rank condition and the fact that  $\{\mathbf{p}_t, \mathbf{q}_t\}_{t=1, \dots, T}$  is  $q$ -rationalisable Corollary 1 then applies. Since  $\text{rank}\{\mathbf{E}\}$  can never be less than  $H$ ,  $\mathbf{B}$  must always add at least  $H$  characteristics to the model. ■

**Proof of Theorem 4**

If  $f(\mathbf{B}) = 0$  then  $\text{emax}(\boldsymbol{\Pi}, \mathbf{Z}) = 1$  since  $\text{emax}(\boldsymbol{\Pi}, \mathbf{Z}) \in [0, 1]$  and by definition (see the discussion of this function in the Appendix) the data  $\{\boldsymbol{\pi}_t, \mathbf{z}_t\}_{t=1, \dots, T}$  satisfy condition (G) and therefore conditions (A) and (L) in Theorem 1 and their equivalents in Theorem 2. ■

## The Afriat efficiency function.

A GARP 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 GARP then Afriat (1973) suggests 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 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  $\mathbf{q}_t P_e^0 \mathbf{q}_s$ ”  $\equiv e \mathbf{p}'_t \mathbf{q}_t \leq \pi'_t \mathbf{q}_s$  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. The function  $e()$  is a simple numerical search algorithm which finds the largest value of  $e$  such that a given data set satisfies  $GARP(e)$ .

## References

- [1] Afriat, S.N. (1967), “The construction of a utility function from expenditure data”, *International Economic Review*, **8**, 76-77.
- [2] Berry, S., J. Levinsohn, and A . Pakes, (1995), “Automobile prices in market equilibrium,” *Econometrica*, **63**, 841-990.
- [3] Blundell, R., M. Browning and I. Crawford (2003), “Nonparametric Engel curves and revealed preference”, *Econometrica*, 71(1), 205-240.
- [4] Blundell, R. and J-M Robin (2000), “Latent Separability: Grouping Goods without Weak Separability”, *Econometrica*, 68(1), 53-84.
- [5] Crawford I. (2004), “Necessary and sufficient conditions for latent separability”, cemmap working paper CWP02/04.
- [6] Diewert, W.E. (1973), “Afriat and revealed preference theory”, *Review of Economic Studies*, **40**, 419-426.
- [7] Diewert, E. and C. Parkan, (1978), “Tests for consistency of consumer data and nonparametric index numbers”, University of British Columbia, Working Paper 78-27.
- [8] 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.
- [9] Griliches, Z. (1971), *Price Indexes and Quality Change*, Washington: Federal Reserve Board.
- [10] Hanoch, G. and M. Rothschild, (1972), “Testing the assumptions of production theory: a nonparametric approach”, *Journal of Political Economy*, **80**, 256-275.
- [11] Heckman, J., and J. Scheinkman, (1987), “The Importance of Bundling in a Gorman-Lancaster Model of Earning” *Review of Economic Studies*, **54**, 243-55
- [12] Houthakker, H. S. (1950), “Revealed preference and the utility function”, *Economica*, May, 159-174.
- [13] Lancaster, K. (1966), “A new approach to consumer theory”, *Journal of Political Economy*, **74**, 2, 132-157.
- [14] Magnus, J.R. and H. Neudecker, *Matrix Differential Calculus*, Chichester: Wiley.
- [15] Markowitz, H. (1959), *Portfolio Selction: Efficient Diversification of Investments*, New York: Wiley.
- [16] Muellbauer, J. (1974), “Household production theory, quality and the ‘hedonic technique’”, *American Economic Review*, **64**, 977-994.
- [17] Pudney, S. (1981) “Instrumental Variable Estimation of a Characteristics Model of Demand”, *Review of Economic Studies*, **48**, 417-433

- [18] Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, **82**, 34-55.
- [19] Samuelson, P. (1948), "Consumption theory in terms of revealed preference", *Economica*, November, 243-253.
- [20] Stone, R. (1956), *Quantity and Price Indexes in National Accounts*, Paris: Organisation for European Economic Cooperation.
- [21] 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.
- [22] Varian, H. (1982), "The nonparametric approach to demand analysis", *Econometrica*, **50**, 945-973.
- [23] Varian, H. (1983), "Nonparametric tests of consumer behaviour", *Review of Economic Studies*, **50**, 99-110.
- [24] Varian, H. (1984), "The nonparametric approach to production analysis", *Econometrica*, **52**, 579-597.
- [25] Varian, H. (1988), "Revealed preference with a subset of goods". *Journal of Economic Theory*, **46**(1),179-185.