

# A nonparametric characteristics model of the demand for milk.

Laura M. Andersen\*,    Laura Blow\*\*,    Martin Browning\*\*\*  
   Ian Crawford\*\*

\* AKF, Copenhagen;

\*\* Institute for Fiscal Studies, London.

\*\*\* CAM, University of Copenhagen;

July 31, 2003

## Abstract

Characteristics models in demand analysis capture the idea that people value goods not for the commodity itself but for the characteristics (or attributes) or embodied in the good. For example, agents may care about the fat content and the taste of different sorts of milk but not the actual type of milk. When we have fewer characteristics than types of good the theory imposes restrictions on observables.

We present a revealed preference characteristics model analysis of the demand for milk in Denmark.

## 1 Introduction.

We present an analysis of the demand for milk using a characteristics model in which agents value different types of milk for their characteristics (attributes) such as fat content, whether the milk is produced under environmentally friendly conditions ('oko' to use the Danish abbreviation) and taste. In this paper we consider only linear characteristics models. We develop Afriat-Varian style revealed preference tests for linear characteristics models that obviate the need for assuming a particular functional form for preferences. The principal questions we address are:

- are the data on milk prices and demand nonparametrically consistent with a linear characteristics model with a known technology?
- as well as the characteristics mentioned above, are other (latent) attributes also needed to rationalise the data?

- if the data are not consistent with a known technology, are they consistent with any non-trivial linear characteristics model?<sup>1</sup>
- if the data can be rationalised with a linear characteristics model, what are the implied bounds on the willingness to pay for a characteristic such as ‘oko’? Here we can make a distinction between the market valuation and the valuation for particular households.
- can a revealed preference analysis yield bounds on the value of a new market good that bundles attributes in some way that has not previously been offered?

Milk is a particularly simple commodity to model since it is non-durable, non-storable (so that there is no inventory problem to deal with), absolutely homogeneous within milk types and we can observe some (or even all) of the characteristics in our data. Thus milk is one of the simplest commodities we could think of modelling (another is eggs, as in the classic paper in the field, Gorman (1980)). Despite this simplicity we shall see that important modelling and substantive issues arise. It is our hope that highlighting these for such a simple commodity will be useful for modelling goods such as soap powder, cars and personal computers.

The empirical analysis uses an unusual data set in which we have the records of purchases of milk (and other goods) for a large number of Danish households over several years (a maximum of 208 weeks for some households). We conduct separate analyses for each household so that we do not have to model between household heterogeneity explicitly.<sup>2</sup> Since so much of the subsequent analysis is motivated by the nature of our data, in the next section we present an detailed description of the data.

In the third section we consider the analysis of demand data using revealed preference techniques. A necessary condition for being able to rationalise demand behaviour with a characteristics model is that the data satisfy the Generalised Axiom of Revealed Preference (GARP); see Varian (1982). If the data do pass GARP we can then ask whether the data are also consistent with the existence of a non-trivial linear characteristics model. The analysis is conducted for two cases; the first in which we know the linear technology (and all characteristics are observed) and the second in which we do not know the technology. Additionally we discuss one way to allow for measurement errors in prices for RP type tests.

---

<sup>1</sup>As we discuss below, a non-trivial characteristics model is one in which there are fewer characteristics than market goods.

<sup>2</sup>Or, rather, we deal with heterogeneity by allowing that preferences might be completely different from household to household.

Type	Fat content	Oko	Taste	Share
1	0.1%	No	Regular	14.9
2	1.5%	No	Regular	47.2
3	3.5%	No	Regular	13.7
4	0.1%	Yes	Regular	7.7
5	1.5%	Yes	Regular	7.9
6	3.5%	Yes	Regular	2.8
7	0.5%	No	Buttermilk	4.5
8	0.5%	Yes	Buttermilk	1.25

Table 1: Types of milk

## 2 A first look at the data.

### 2.1 The market for milk in Denmark.

We present a description of the milk market in Denmark during our sample period. This will include a discussion of the types of milk and the marketing structure. Discuss our sample period with mention of the introduction of 'mini-milk' in the final year. Detailed description of how prices are set.

### 2.2 A description of the price data.

Describe how the data are collected.

For each household, daily records of expenditures (in Danish Kroner) and the quantity bought (in liters) of various types of milk are recorded. The types are differentiated by fat content, whether the milk is produced under environmentally congenial conditions ('oko') and taste. Table 1 records the characteristics of the eight types that we work with. We also record the market shares for the different types (these are derived for our total sample).<sup>3</sup>

As can be seen 75% of milk purchases are for regular non-oko milk. For oko milks, low fat is relatively popular (as compared to non-oko milk). We construct weekly unit values for each type of milk and each of nine regions<sup>4</sup> by dividing the weekly expenditure on that type of milk by all households in our survey in that region by the weekly quantity in the region. We shall refer to these week-region specific unit values as prices; note that these are absolute prices and are *not* deflated by any general price index. We have a number of missing values in our price data because no one in our sample bought that particular type of milk in that region and week. Table A1 in the

<sup>3</sup>We exclude a small number of milk purchases that are not categorised and we also exclude flavoured milk.

<sup>4</sup>The regions are Copenhagen, North Zealand, South Zealand, Funen, Bornholm, East Jutland, North Jutland, West Jutland and South Jutland.

Type	ADF	Mean	SD	ARMA(1, 1)	AR	MA	AR(1)	MA(1)
1	-2.6	5.1	0.11	5.1	0.99	-0.91	32.9	36.0
2	-4.4	5.3	0.16	4.2	-0.01	-0.20	4.9	4.2
3	-3.0	6.1	0.15	3.4	0.91	-0.74	22.3	24.7
4	-2.9	6.4	0.12	5.7	0.98	-0.75	37.9	65.1
5	-5.0	6.6	0.20	3.4	0.96	-0.92	5.2	5.2
6	-4.7	7.5	0.18	2.4	-0.47	0.61	5.9	5.3
7	-4.8	5.2	0.21	8.8	0.98	-0.95	15.1	15.1
8	-4.3	6.5	0.16	0.1	0.92	-0.84	4.5	4.9
ADF is the ADF unit root test statistic. The 5% value for a unit root is -2.9.								
ARMA(1, 1) is the $\chi^2(2)$ statistic for ARMA(2, 2) $\rightarrow$ ARMA(1, 1).								
AR and MA are the estimated parameter values from the ARMA(1, 1).								
AR(1) is the $\chi^2(3)$ statistic for restricting an ARMA(2, 2) to an AR(1).								
MA(1) is the $\chi^2(3)$ statistic for restricting an ARMA(2, 2) to an MA(1).								

Table 2: Price statistics

Appendix gives the number of weeks in which we do not observe a particular price for a particular region. We have very few observations in regions 3 (South Zealand) and 5 (Bornholm) so we drop these from our analysis. We choose to impute missing prices for the remaining missing prices using a simple interpolation scheme; details are given in the Appendix.

Since the prices and their properties play an important role in the analysis below we present here a time series analysis of the prices for one region (Copenhagen). For this region we do not have any missing values for any prices. Figure 1 presents the plots for 208 periods (1997.1 to 2000.52) for milk types 2, 3 and 5. The y-axis gives the price per liter in Danish Kroner (about 7 kroner to one Euro). The first feature of this figure is that there is a clear (upwards) break in prices for types 1 and 2 in weeks 15/16 and 194/195. This break is also seen in all of the other non-oko price series but not in any of the oko milk price series. In some of our analysis below it is convenient to work with stationary series, so we restrict attention to weeks 16 to 194 (a total of 179 time periods). In our demand analysis we use the whole sample since the discrete, persistent and exogenous change in the relative price of oko milk obviously helps the identification of price effects. On the restricted period there is no visual evidence of a time trend in absolute prices over the 179 weeks of the sample; below we shall present formal tests for stationarity. The other important feature of the graphs is that higher fat milk is more expensive (compare types 1 and 2) and oko milk is more expensive (compare types 2 and 5). To supplement this visual analysis, we present some statistics for the eight price series in Table 2. milk).

The first column presents the augmented Dickey-Fuller (ADF) unit root test statistic (with 5 lags of differences included in the regression) for the sample

from weeks 16 to 194.<sup>5</sup> Although types 1 and 4 are marginal we shall proceed as though none of series has a unit root. Given this, we present means and standard deviations in columns 2 and 3. The means show that higher fat content milks cost more (*ceteris paribus*) as does oko milk. The differences between the mean price for non-oko and the comparable oko variety are all about 1.3 Kroner per liter. The other columns of Table 2 examine the univariate process that each price series follows. We start with an ARMA(2, 2) with Gaussian errors. We then test for restricting this to an ARMA(1, 1). The results for this test are given in column 4; mostly we do not reject the restrictions. The final two columns give the test statistics for restricting the processes to being even simpler than the ARMA(1, 1). As can be seen, we reject these simpler models in most cases. Accepting that a common model of an ARMA(1, 1) is appropriate, we present the parameter estimates in the columns headed AR and MA. One major surprise from this analysis is how different the processes are. For example, the price for regular 1.5% milk (type 2) displays no autoregressive component and only a small and negative MA component. On the other hand, regular 0.1% milk (type 3) seems to have an autoregressive parameter that is almost unity and a very large and negative MA component.

In revealed preference analysis only relative prices matter. In figure 2 we present the path of the relative price of type 5 (medium fat, oko) milk to the price of type 2 milk (medium fat, non-oko) milk. This displays considerable variability with a premium for oko of between 10% and 35% (in the middle periods). This high variability is of considerable interest for the identification of price effects in a model that assumes intertemporal additivity but we must, of course, allow that some of it may be due to demand shocks.

### 2.3 A description of the quantity data.

The prices given above are assumed common to all households in the particular week and region. The purchases by households are, of course, specific to the household. Here we present an informal analysis of the demand patterns for a small number of households who are present in our sample for a large number of contiguous weeks. To minimise 'taste changes' we also select on the household not having any change in household structure, except for the ageing of the members. Describe sample.

Graphs of 'aggregate' data. For a given region (Copenhagen) graph aggregate budget shares for non-oko and oko regular milks. Also graphs of three fat types for non-oko and oko. Finally buttermilk, non-oko and oko.

For illustrative purposes we consider seven households who are in our sample for almost 208 years and who consist of a married couple with no one else present during the whole data period. Table ?? presents the mean

---

<sup>5</sup>Including a deterministic trend gives very similar results.

	Household						
	A	B	C	D	E	F	G
Quantity	2.7	3.2	2.2	3.7	5.4	3.4	2.0
Type 1	0	0	0	0	0.1	0	15.7
Type 2	0	7.4	33.4	45.4	95.0	9.2	21.8
Type 3	99.5	82.5	48.5	32.5	0.5	2.6	3.9
Type 4	0	0	0	0	3.7	0.9	12.8
Type 5	0	0.2	0	0	0.7	70.0	43.6
Type 6	0.5	0.2	0.2	0	0	16.5	2.2
Type 7	0	9.5	1.3	21.6	0	0.8	0.1
Type 8	0	0	16.6	0.5	0	0.1	0
Quantities in litres per week.							
Budget shares (multiplied by 100)							

Table 3: Individual budget shares

budget shares (averaged over time) for the seven households for each of the eight goods.

Most obvious features:

- some households buy mostly one type of milk (A, B and E).
- lots of zeros over whole four years
- lots of heterogeneity

Figure ?? presents the time path of budget shares for types 2, 3 and 7 for household D. As can be seen, there seems to be a switch from type 3 to type 2 milk at around week 80. The research question is whether we can account for these changes assuming unchanging preferences.

The sample we work with is a sample of households who are present for at least 156 weeks of the 208 we consider and who live in the same region for the whole period. The numbers of such families are (133, 112, 62, 136, 60, 59, 66) for regions (1, 2, 4, 6, 7, 8, 9) respectively for a total of 628 households in all.

### 3 The linear characteristics model.

#### 3.1 The basic framework.

Suppose we observe quantity bought by an individual household and price data on  $K$  market goods  $\mathbf{q} = [q^1, \dots, q^K]'$  and  $\mathbf{p} = [p^1, \dots, p^K]'$  for in periods  $t = 1, \dots, T$ . Afriat and Varian have established empirical conditions under which the data  $\{\mathbf{q}_t, \mathbf{p}_t\}_{t=1 \dots T}$  can be rationalised as the maximisation of a utility function  $u(\mathbf{q})$  with standard properties. In this and the next section

we extend these conditions to the case where the data are generated by a utility maximising consumer whose preferences have the following structure:

$$u(\mathbf{q}) = v(A\mathbf{q}) \tag{1}$$

where  $A$  is a  $J \times K$  non-negative matrix with full row rank  $J < K$ . The  $J$ -vector  $\mathbf{z} = \mathbf{A}\mathbf{q}$  is a vector of *characteristics* or *intermediate goods* or *attributes*. To illustrate the use of characteristics consider our milk data. We have eight types of milk. One set of characteristics we could take are: ('milkiness', 1.5% fat, 3.5% fat, oko, regular taste) so that  $J = 5$ .<sup>6</sup> This gives the following  $A$  matrix:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

This matrix has rank five. The technology is not unique, in the sense that  $BA$  is also a linear technology for any non-singular  $J \times J$  matrix; this corresponds to a linear redefinition of the intermediate goods. For example, we could replace the first row with a vector so that the first characteristic is '0.1% fat'. If we assume that preferences are linear over fat content then we have only three characteristics, (fat,oko,taste) with following rank 3 transformation matrix:

$$A = \begin{bmatrix} 0.1 & 1.5 & 3.5 & 0.1 & 1.5 & 3.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

It is important to test for this structure since if it is accepted we can put much tighter bounds on the market valuation of mini-milk (the new type of regular milk that was introduced with an 0.5% fat content). Below we shall present tests for whether we can replace a five factor model by a three factor model.

## 4 The hedonic pricing model.

### 4.1 Known technology.

A particularly important special case of the linear characteristics model is when prices are such that the available characteristics for a fixed outlay on market goods all lie on a hyperplane; this is the hedonic pricing model. Figure ?? illustrates for a three good, two characteristics model. In the

---

<sup>6</sup>Note that buttermilk is different in both taste and fat content but we can only allow for one of these characteristics.

left hand panel we have that the available budget set is not linear and we do not have a hedonic model. In the right panel the price of good 2 has been increased until the boundary of the budget set is linear. The defining characteristic of a hedonic pricing model is that market prices are simply the weighted sums of the prices of the underlying shadow prices,  $\pi_t^j$ :

$$\mathbf{p}_t = A' \tilde{\pi}_t \left( \frac{\mu_t}{\lambda_t} \right) = A' \boldsymbol{\pi}_t \quad (2)$$

where the  $\tilde{\pi}_t^j$ 's are *shadow prices* for the characteristics.

We begin by considering the case in which the linear technology  $A$  is known. In this case the full set of characteristics  $\mathbf{z}_t = A\mathbf{q}_t$  are also observed if market purchases are observed. From equation (2) we see that we need to find shadow prices  $\boldsymbol{\pi}_t$  such that  $\mathbf{p}_t = A'\boldsymbol{\pi}_t$  for all  $t$ . For each  $t$  this is a system of  $K$  equations in  $J < K$  unknowns so that a solution does not generally exist. The following gives the well-known condition for existence, where  $(A'|\mathbf{p}_t)$  is the  $K \times (J+1)$  matrix of  $A'$  and  $\mathbf{p}_t$  stacked horizontally.

**Lemma 1** *There exists a  $\boldsymbol{\pi}_t$  satisfying  $\mathbf{p}_t = A'\boldsymbol{\pi}_t$  if and only if  $\text{rank}(A'|\mathbf{p}_t) = J$ . If this condition holds then  $\boldsymbol{\pi}_t$  is given uniquely by:*

$$\boldsymbol{\pi}_t = (AA')^{-1} A\mathbf{p}_t \quad (3)$$

Some remarks:

- It is important to recognise that given any  $A$  and  $\mathbf{p}_t$  we can always define  $\boldsymbol{\pi}_t$  uniquely in this way, but this  $\boldsymbol{\pi}_t$  will only satisfy  $\mathbf{p}_t = A'\boldsymbol{\pi}_t$  exactly if the rank condition holds.
- Even if the rank condition holds,  $A$  is nonnegative and market prices are positive we may have  $\pi_t^j < 0$  for some  $j, t$  so that shadow prices may not always be positive.
- A result that is useful below is that:

$$\text{rank}(A'|\mathbf{p}_t) = J \Rightarrow A'(AA')^{-1} A\mathbf{p}_t = \mathbf{p}_t \quad (4)$$

The implications of this for the revealed preference conditions are given by:

**Proposition 2** *If  $\text{rank}(A'|\mathbf{p}_t) = J$  for all  $t$  then  $\{\mathbf{q}_t, \mathbf{p}_t\}$  satisfies GARP if and only if  $\{\mathbf{z}_t, \boldsymbol{\pi}_t\}$  satisfies GARP.*

**Proof.** *Suppose the rank condition holds and  $\{\mathbf{q}_t, \mathbf{p}_t\}$  satisfies GARP. GARP for  $\{\mathbf{q}_t, \mathbf{p}_t\}$  implies:*

$$\mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}_s \text{ and } \mathbf{p}'_s \mathbf{q}_s \geq \mathbf{p}'_s \mathbf{q}_w \Rightarrow \mathbf{p}'_w \mathbf{q}_w \leq \mathbf{p}'_w \mathbf{q}_t \quad (5)$$



(where we have, without loss of generality, taken a chain of length 3,  $\{t, s, w\}$ ). Given the rank condition we can find unique shadow prices for period  $t$  that satisfies  $\mathbf{p}_t = A\boldsymbol{\pi}_t$  (and similarly for  $s$  and  $w$ ). Substituting for market prices this gives:

$$\boldsymbol{\pi}'_t A \mathbf{q}_t \geq \boldsymbol{\pi}'_t A \mathbf{q}_s \text{ and } \boldsymbol{\pi}'_s A \mathbf{q}_s \geq \boldsymbol{\pi}'_s A \mathbf{q}_w \Rightarrow \boldsymbol{\pi}'_w A \mathbf{q}_w \leq \boldsymbol{\pi}'_w A \mathbf{q}_t \quad (6)$$

so that:

$$\boldsymbol{\pi}'_t \mathbf{z}_t \geq \boldsymbol{\pi}'_t \mathbf{z}_s \text{ and } \boldsymbol{\pi}'_s \mathbf{z}_s \geq \boldsymbol{\pi}'_s \mathbf{z}_w \Rightarrow \boldsymbol{\pi}'_w \mathbf{z}_w \leq \boldsymbol{\pi}'_w \mathbf{z}_t \quad (7)$$

which is GARP on  $\{\mathbf{z}_t, \boldsymbol{\pi}_t\}$ .

Conversely, for GARP on  $\{\mathbf{z}_t, \boldsymbol{\pi}_t\}$  we examine inequalities such as:

$$\boldsymbol{\pi}'_t \mathbf{z}_t \geq \boldsymbol{\pi}'_t \mathbf{z}_s \quad (8)$$

Substituting for  $\mathbf{z}_t$  and  $\boldsymbol{\pi}_t$  and using (4) this implies:

$$\mathbf{p}'_t A' (AA')^{-1} A \mathbf{q}_t \geq \mathbf{p}'_t A' (AA')^{-1} A \mathbf{q}_s \Rightarrow \mathbf{p}'_t \mathbf{q}_t \geq \mathbf{p}'_t \mathbf{q}_s \quad (9)$$

so that GARP holds for  $\{\mathbf{q}_t, \mathbf{p}_t\}$  if it holds for  $\{\mathbf{z}_t, \boldsymbol{\pi}_t\}$ . ■

Given these results we have a clear path for revealed preference testing for a characteristics model. First we test for GARP for the observed data  $\{\mathbf{q}_t, \mathbf{p}_t\}$ . Clearly if this fails there is no utility model that rationalises the data. Strictly GARP tests for the milk group being a separable group. If the test fails we can attribute this to several reasons: the group is separable but the demand for milk products is not ‘rational’; the group is not separable (for example, we do not include substitutes such as other types of drink or other sources of calcium and fat); the group is separable but preferences for milk products change over time; the group is separable but preferences for milk products are not time separable (there are habits); the prices or quantities are measured with error. We shall return to some of these issues below. If the market data satisfy GARP, then we test for the rank condition. If this passes, then a characteristics model with this linear technology rationalises the data. If the demand data satisfy GARP but the prices fail the rank condition then we conclude that there is no characteristics model that rationalises the data *with the given linear technology*. If this happens, two issues arise. First, we can ask: is there any linear technology that is consistent with the data? This is dealt with in the next sub-section. The second issue allows that prices might be measured with error.

## 4.2 Noisy prices.

Suppose that we assume that true prices do satisfy the rank condition in each period but that prices are measured with an additive error:

$$\mathbf{p}_t = A' \hat{\boldsymbol{\pi}}_t + \boldsymbol{\eta}_t \quad (10)$$

where  $\boldsymbol{\eta}_t$  is a serially uncorrelated vector process with zero mean and constant covariance matrix  $\Sigma_\eta$ . Let  $P$  be the  $K \times T$  matrix of observed prices stacked horizontally. If we have more time periods than goods ( $T > K$ ) then if  $\boldsymbol{\eta}_t$  has a non-degenerate distribution, we will almost surely have  $\text{rank}(P) = K$  so that the exact rank condition of the previous section will never hold. One procedure in this case is to regress prices in period  $t$  on the transposed characteristics matrix  $A'$  to give estimates of the shadow prices:

$$\hat{\boldsymbol{\pi}}_t = (AA')^{-1} A\mathbf{p}_t \quad (11)$$

and then to define ‘predicted prices’:

$$\hat{\mathbf{p}}_t = A'\hat{\boldsymbol{\pi}}_t = A'(AA')^{-1} A\mathbf{p}_t \quad (12)$$

If (and only if) the price vector in period  $t$  satisfies the rank condition of the previous subsection then  $\hat{\mathbf{p}}_t = \mathbf{p}_t$  (see equation (4)). Given predicted prices, the following gives a test for a characteristics model.

**Proposition 3** *The following are equivalent:*

- (i)  $\{\mathbf{q}_t, \hat{\mathbf{p}}_t\}$  satisfies GARP
- (iii)  $\{A\mathbf{q}_t, \hat{\boldsymbol{\pi}}_t\}$  satisfies GARP
- (ii) there is a characteristics model with linear technology  $A$  that rationalises the data.

It is important to emphasise that GARP tests can pass for  $\{\mathbf{q}_t, \hat{\mathbf{p}}_t\}$  but fail for  $\{\mathbf{q}_t, \mathbf{p}_t\}$  (and vice versa) unless the rank condition holds for each period.

This framework can also be used to give the means of the shadow prices,  $\boldsymbol{\mu}_\pi$ , and hence valuations for the different characteristics. The obvious estimator to take for the mean is:

$$\bar{\boldsymbol{\pi}} = \frac{1}{T} \sum \hat{\boldsymbol{\pi}}_t = (AA')^{-1} A \frac{1}{T} \sum \mathbf{p}_t = (AA')^{-1} A\bar{\mathbf{p}} \quad (13)$$

Under the stationarity assumption, this is a consistent estimator of  $\boldsymbol{\mu}_\pi$ .

Although the solution given in (11) is standard in the static factor analysis literature it ignores the time series structure of shadow and market prices. To take account of this, assume that shadow prices follow a stationary first order vector autoregression:

$$\boldsymbol{\pi}_t = (I - B)\boldsymbol{\mu}_\pi + B\boldsymbol{\pi}_{t-1} + \mathbf{u}_t \quad (14)$$

where  $B$  is a non-singular  $J \times J$  matrix with the usual stationarity condition. The  $J$ -vector  $\mathbf{u}_t$  is assumed to be serially uncorrelated with mean zero and a constant contemporaneous covariance matrix  $\Sigma_u$ . Together equations (14) and (10) constitute a state space representation with the former as the transition equations and the latter as the measurement equations (see, for example, Harvey (1990) or Hamilton (199?)). More specifically, this is a

dynamic factor model; in this state space form it was introduced by Engle and Watson (1981). In our structure we have a fixed and small number of variables ( $K$  prices) and large  $T$ . Thus our structure is somewhat simpler than for dynamic factor models with large  $K$  and  $T$  (see, for example, Bai and Ng (2002)) or for models with  $K \gg T$ , as in Stock and Watson (1998). Taking account of the time series structure will provide a lower variance estimator for the mean of the shadow prices. It will also facilitate testing for additional factors or an alternative characteristics structure without the need to use the tests for rank developed in, for example, Cragg and Donaldson (1997). The obvious estimation procedure is to use Kalman filter based ML. An alternative is based on the reduced form of the state space representation. From the (lagged) measurement equations we have:

$$\boldsymbol{\pi}_{t-1} = (AA')^{-1} A\mathbf{p}_{t-1} - (AA')^{-1} A\boldsymbol{\eta}_{t-1}$$

If we substitute the transition equations into the measurement equations and use this then we have the reduced form:

$$\begin{aligned} \mathbf{p}_t = & A'(I - B)\boldsymbol{\mu}_\pi + A'B(AA')^{-1}A\mathbf{p}_{t-1} \\ & - A'B(AA')^{-1}A\boldsymbol{\eta}_{t-1} + \boldsymbol{\eta}_t + A'\mathbf{u}_t \end{aligned} \quad (15)$$

so that prices follow a vector ARMA(1, 1) (see Pena and Box (1987)):

$$\mathbf{p}_t = \mathbf{d} + C\mathbf{p}_{t-1} + \mathbf{v}_t + \Theta\mathbf{v}_{t-1}$$

This is consistent with the time series properties for prices found in section 2. From this framework we can derive predictions of shadow prices that are alternative to those derived from (11).

### 4.3 Additional latent attributes.

We can ask: given the technology described by  $A$ , do we need an additional (latent) attribute to adequately describe the evolution of prices?

### 4.4 Unknown linear technology.

Suppose that we have market data  $\{\mathbf{q}_t, \mathbf{p}_t\}_{t=1, \dots, T}$  that satisfy GARP and that prices and quantities are measured without error. We suspect that this may be rationalisable by a linear technology characteristics model but the form of the technology (and the number of characteristics) is unknown.<sup>7</sup> Clearly if we allow for as many characteristics as goods ( $J = K$ ) then we can always do this by setting  $A = I_K$ . Thus we shall seek a *non-trivial linear characteristics model* with  $J < K$ .

---

<sup>7</sup>This clearly requires that characteristics are not observed by the analyst (otherwise we could estimate the technology). We always assume that agents observe all characteristics.

**Proposition 4** *Suppose the market data  $\{\mathbf{q}_t, \mathbf{p}_t\}$  satisfy GARP. We can find a non-trivial linear characteristics model that rationalises the data if and only if  $\text{rank}(P) < K$ .*

**Proof.** *Consider the  $K \times T$  matrix of prices  $P$  and let  $\text{rank}(P) = J < K$ . Let  $\Pi$  be a  $J \times T$  row basis matrix for  $P$  (that is, every row of  $P$  can be expressed as a linear combination of rows of  $\Pi$ ). By definition  $\Pi$  has rank  $J$ . Now consider the set of equations  $P = A'\Pi$  where  $A$  is a  $J \times K$  matrix. By the definition of  $\Pi$  a solution  $A$  exists and is given by:*

$$A = (\Pi\Pi')^{-1}\Pi P' \quad (16)$$

*Now take the characteristics model with  $\mathbf{z}_t = A\mathbf{q}_t$  and  $\boldsymbol{\pi}_t$  as the  $t$ 'th column of  $\Pi$  (so that  $\mathbf{p}_t = A'\boldsymbol{\pi}_t$ ). This satisfies the conditions of Proposition 1 and hence we have found a linear characteristics model that rationalises the data.*

*Conversely, suppose that there is a non-trivial linear characteristics model that rationalises the data. Suppose that this has  $J$  characteristics. Then proposition 1 gives that  $\text{rank}(A'|P) = J$  which implies that  $\text{rank}(P) \leq J < K$ . ■*

A number of remarks:

- Given GARP for a particular household, the condition for the existence and construction of a linear technology depends only on market prices that household faces and not on the market purchases.
- If we have fewer time periods than goods ( $T < K$ ) then  $\text{rank}(P) \leq T < K$  so that we can always find a linear technology that rationalises the data if it satisfies GARP.
- There is nothing to ensure that all of the elements of  $A$  are non-negative even if all market prices are positive. We conjecture that there is some way to choose the row basis matrix so that  $A$  is always non-negative. Consider, for example, the following technology with  $T = 7$  and  $K = 6$ :

$$P = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 4 & 5 \\ 2 & 2 & 2 & 3 & 5 & 5 & 5 \\ 1 & 2 & 3 & 4 & 5 & 4 & 5 \\ 1 & 2 & 2 & 3 & 4 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 & 4 & 5 \\ 3 & 4 & 4 & 6 & 9 & 9 & 10 \end{bmatrix} \quad (17)$$

This has rank 3. If we take a random row basis matrix then we always end up with at least one negative element in each row of  $A$ . On the other hand, if we take the row basis given by rows 1, 2 and 4:

$$\Pi = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 4 & 5 \\ 2 & 2 & 2 & 3 & 5 & 5 & 5 \\ 1 & 2 & 2 & 3 & 4 & 4 & 5 \end{bmatrix} \quad (18)$$

we have:

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (19)$$

which does satisfy our non-negativity restriction.

#### 4.5 Noisy prices.

In general we will have many more periods than goods ( $T \gg K$ ) and any noise in observing prices will almost surely give  $\text{rank}(P) = K$ . On the other hand we might posit that ‘true’ prices  $P^*$  do have rank less than  $K$ . One informal check is to examine the eigenvalues of the  $K \times K$  matrix  $PP'$ . More acceptable would be to conduct a test on the rank of  $P$ .

Use dynamic factor model to determine the number of factors and the form of the transformation matrix  $A$ .

### 5 Testing for a linear characteristics model.

#### 5.1 Known technology.

The previous section dealt with testing for a hedonic pricing model and was concerned solely with market prices. In this section we move to a revealed preference analysis of individual demands. We provides tests for whether a time series of emands for a given agent can be rationalised by utility function that has a linear characetristics form given in equation (1) and reproduced eher for convenience:

$$u(\mathbf{q}) = v(A\mathbf{q}) \quad (20)$$

where  $A$  is a  $J \times K$  matrix with full row rank.

### 6 Conclusions.

#### References

- [1] Bai, Jushan and Serena Ng (2002), “Determining the number of factors in approximate factor models”, *Econometrica*, 70(1), 191-222.
- [2] Cragg, John and Stephen Donald (1997), “Inferring the rank of a matrix”, *Journal of Econometrics*, 76, 223-250.
- [3] Gorman, W.M. (1980), “...”, *Review of Economic Studies*,
- [4] Gregory, Alan, A. Head and J. Raynauld (1997), “Measuring world business cycles”, *International Economic Review*, 38, 677-701.
- [5] Hamilton, James (199?), *Time series analysis*,

- [6] Harvey, Andrew (1990), *Forecasting, structural time series models and the Kalman filter*, Cambridge University press, Cambridge.
- [7] Pena, Daniel and George Box (1987), "Identifying a simplifying structure in time series", *Journal of the American Statistical Association*, 82, 836-843.
- [8] Stock, James and Mark Watson (1998), "Diffusion indexes", NBER Working Paper 6702.
- [9] Varian, H.(1982), "The Nonparametric Approach to Demand Analysis", *Econometrica*, 50, 945-974.

## A Imputation of prices.

The following Table gives the mean number of households in each region per week and the number of missing values for prices. Note that prices are based on the full set of observations and not just those that are included in our final elected sample of households. We drop regions 3 and 5 since we have so few observations. For the other regions we have relatively few missing price values for types 1 to 7 (11 out of a total of 1456 prices) so we simply use the simple interpolation scheme:

$$p_{r,t}^i = 0.5 * (p_{r,t-1}^i + p_{r,t+1}^i)$$

where  $r$  denotes region,  $i$  denotes the milk type and  $t$  is time. For the milk type 8 which has 44 missing values, we use ??.

Region	# obs	Number of missing prices for good:							
		1	2	3	4	5	6	7	8
1	389	0	0	0	0	0	0	0	0
2	308	0	0	0	0	0	0	0	0
3	40	1	0	0	36	15	86	0	97
4	164	0	0	0	0	0	0	0	27
5	19	1	0	0	46	151	156	29	139
6	336	0	0	0	0	0	0	0	0
7	173	0	0	0	0	0	3	0	4
8	163	0	0	0	2	1	5	0	2
9	182	0	0	0	0	0	0	0	11

Table 4: Table Caption

## B Figures.

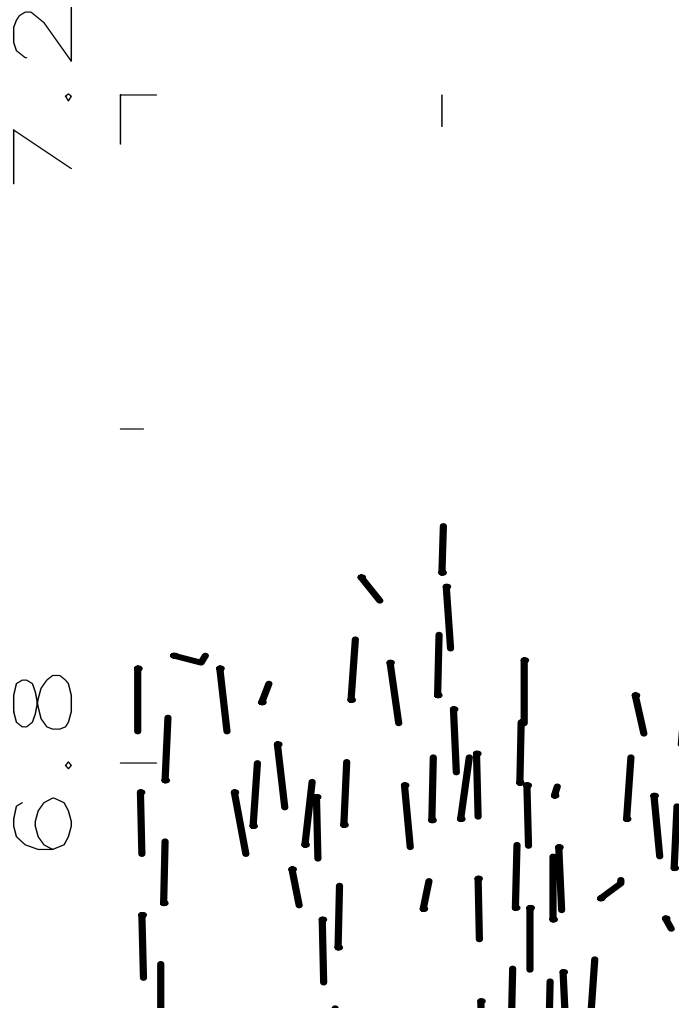


Figure 1: Prices for three types of milk.



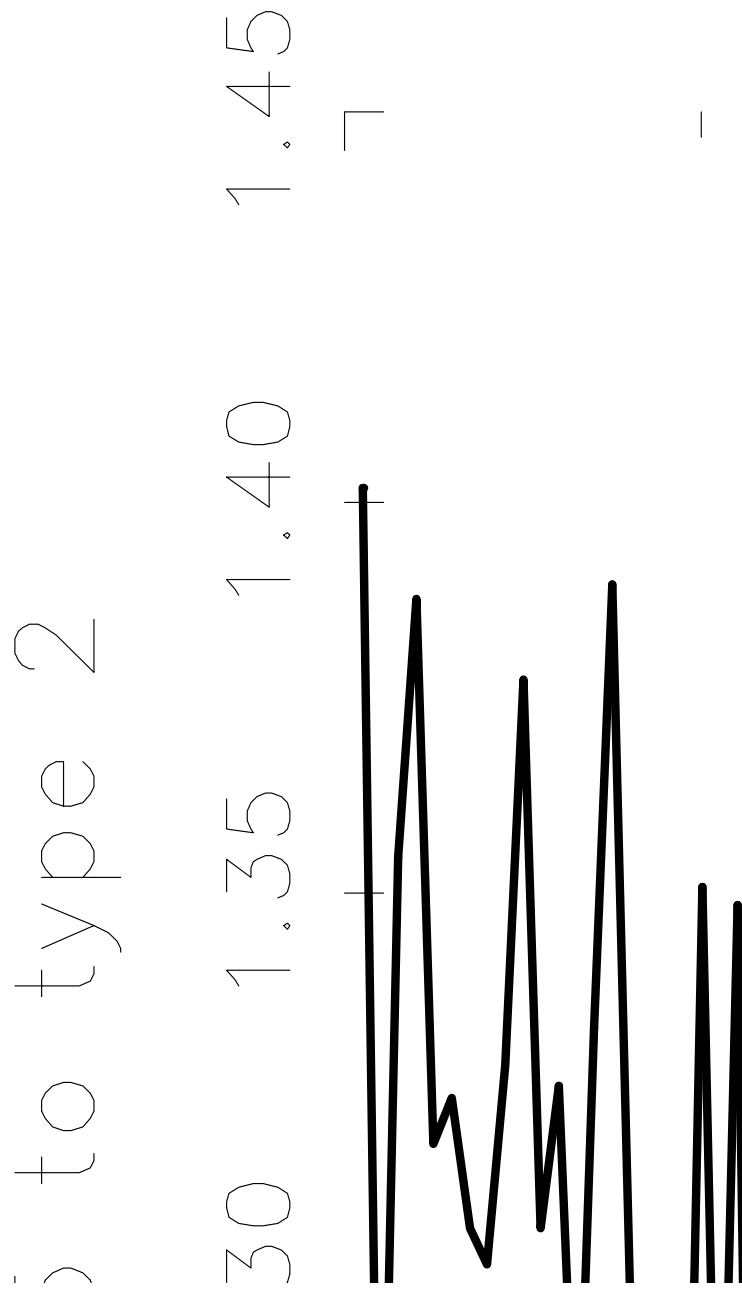


Figure 2: Relative price for oko to non-oko milk.

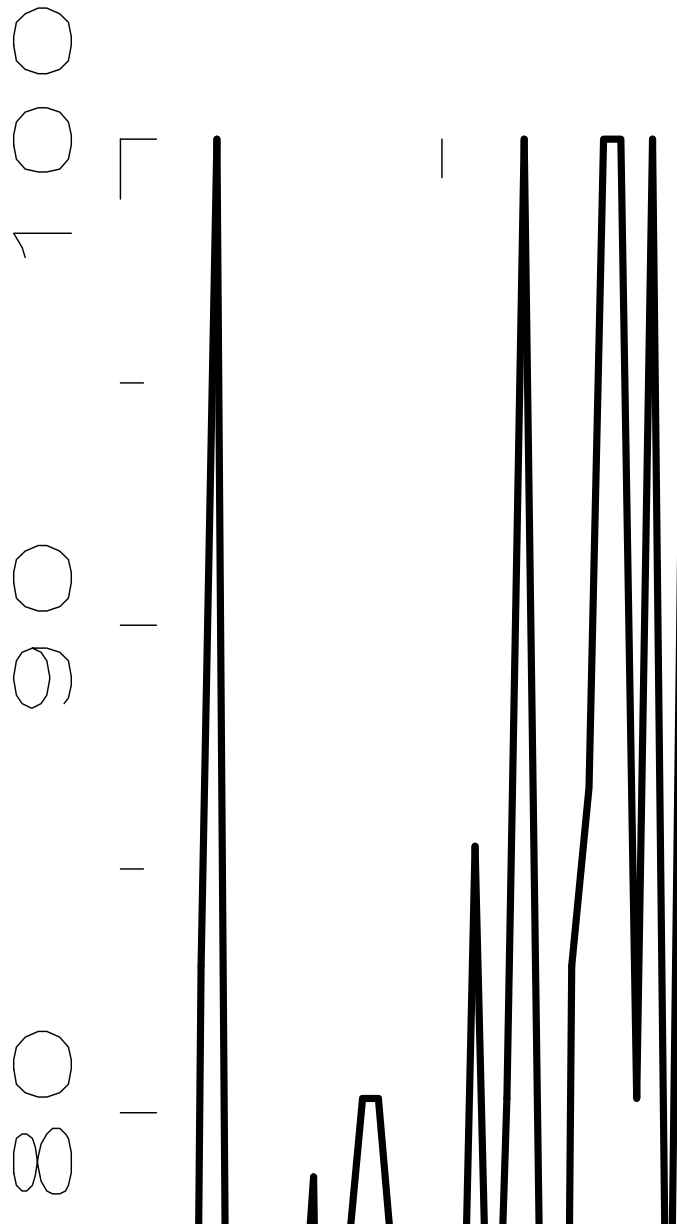


Figure 3: Budget shares for one household.