VARIATIONS IN THE PRICE OF FOODS AND NUTRIENTS IN THE UK

Ian Crawford

THE INSTITUTE FOR FISCAL STUDIES
WP03/19

Variations in the price of foods and nutrients in the UK

Ian Crawford
Institute for Fiscal Studies

October 21, 2003

Abstract. This paper investigates the patterns and extent of differences in the prices paid for foods, and for the nutrients they contain, amongst households in the UK. The data used are from the National Food Survey and are unit prices, quantities purchased and nutrient conversion factors for each food. The paper first describes the circumstances under which ratios of unit prices are exact measures of cross section price variation. It also discusses a nonparametric method of recovering the underlying prices of nutrients under the maintined assumption that households have rational preferences over the nutritional characteristics of foods. It then uses these data to describe patterns in the variation of the price of food and nutrients with repect to household and regional characteristics.

Acknowledgments. I would like to thank Laura Blow, Martin Browning, Andrew Chesher, Lars Nesheim, Ian Preston and David Ulph for helpful comments and suggestions. Financial support from the Nuffield Foundation (grant ref OPD/00101/G) is gratefully acknowledged. Material from the NFS made available by the MAFF through the ESRC Data Archive has been used by permission of the controller of HMSO. Neither the MAFF nor the ESRC Data Archive bear responsibility for the analysis or the interpretation of the data reported here. The author is responsible for all errors.

Summary

- This paper investigates the patterns and extent of differences in the prices paid for foods, and for the nutrients they contain, amongst households in the UK.
- The data used are from the National Food Survey and are unit prices, quantities purchased and nutrient conversion factors for each food.
- The paper first describes the circumstances under which ratios of unit prices are exact measures of cross section price variation.
- It also discusses a nonparametric method of recovering the underlying prices of nutrients under the maintined assumption that households have rational preferences over the nutritional characteristics of foods.
- It then uses these data to describe patterns in the variation of the price of food and nutrients with repect to household and regional characteristics.

1. Introduction

The aim of this paper is to describe cross-sectional variations in the prices which different households pay for foods and the nutrients they contain. The data used is the 2000 National Food Survey (NFS) which records the quantity and the cost of each separate purchase of 244 different foods. These foods ('minor food categories' as they are labelled in the NFS) are quite finely distinguished; there are, for example, eight types of milk, four types of butter and eight varieties of bread. However, these categories are not *entirely* homogenous, and sometimes not even remotely homogenous. As a result dividing expenditure by quantity for a minor food category does not strictly recover the "price" of an individual good, but rather the unit price — total cost measured in pence divided by total quantity measured in (commensurable) physical units — of groups of similar goods. Consequently, when we are comparing unit prices for minor foods across households, there is a possibility (perhaps minor in some cases, but not in general ignorable) that we are not comparing like with like. This raises two questions: firstly under what circumstances (typically restrictions on consumer's preferences and the nature of the data) *can* we say that cross-section variation in unit prices correctly identifies cross-section variation in true prices; secondly, can we make sure we are comparing like-with-like by going below the level of marketed foods to uncover the households' valuations of the nutrients they contain?

Section 2 of this paper looks at the circumstances under which these types of data can be used to describe cross-section variations in prices. It turns out that there are two important identifying assumptions (one restricting the type of price variation within NFS minor food categories,and one restricting households' preferences) which together allow the ratio of two households' unit prices to be interpreted as true, welfare-constant cross section price indices. These assumptions are then duly made. Section 3 of the paper looks at the implications of a particular structure of households' preferences for foods in which households have preferences over the nutritional characteristics of the foods they buy. It discusses how this structure can be used to identify the households shadow prices/willingness to pay for different nutrients. Section 4 investigates these issues empirically. Section 5 concludes.

2. Identifying variation in food prices

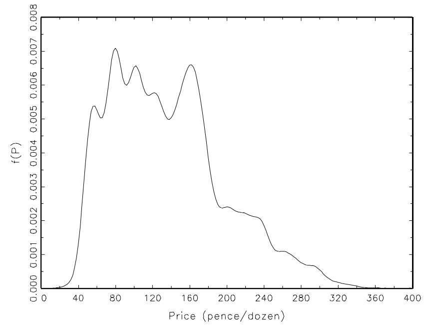
To illustrate these issues and to try to see if anything can be inferred from cross-section variation in per-unit prices I will take a fairly homogenous good: Eggs.

Table 1: Eggs, Descriptive statistics for unit prices.

Egg Prices, p/dozen 2000					
Mean	std.dev	5th	50th	95th	n
137.3	66.4	52.7	127.7	256.5	3166

Figure 1 shows the density of the distribution of the unit price of eggs in 2000 (the prices here and in the rest of this paper are unit prices — denoted p — and have had time-varying effects removed by regression on quarterly dummies. Note

Figure 1: The density of the distribution of egg unit prices



that, within each month the survey rotates around local authorities/regions and only within a quarter can the NFS be said to be nationally representative). The figure shows a perhaps surprisingly large degree of variation and indicates several modes in the distribution, perhaps as many as five at about 55p/dozen , 80p/dozen, 100p/dozen, 125p/dozen and 160p/dozen. Table 1 shows that average unit price of eggs in this sample was 137.3p/dozen, the standard deviation was 66.4p/dozen and the median price paid was 127.7p/dozen with 90% of the distribution lying between about 53p/dozen and 257p/dozen.

Figure 2: Eggs, $f(p,q)$

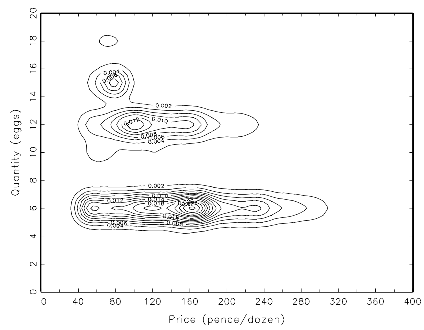


Figure 2 shows the bivariate density of the unit price of eggs and the quantity purchased — denoted q. There are clear mass points in the density at quantities of 1/2 dozen and 1 dozen and 15 eggs. Careful inspection shows that the five mass points shown in figure 1 are present here in the price dimension at these principal quantities. Holding quantity constant there appear to be two or three mass points within each conditional price distribution. However these mass points are slightly off-set with respect to quantity. For example, the three main mass points seem to be at about 80p/dozen for 15 eggs, 100p/dozen for 1 dozen eggs, and 160p/dozen for 1/2 dozen eggs. Does this help to explain the distribution of the unit price of eggs? Possibly: one explanation may be that there are essentially two or three “types” of egg (free-range, barn and battery perhaps, accounting for the modes in the conditional unit price distributions) and that these are bulk-discounted (accounting for the off-set nature of the modes). Each type is discounted by about 1.67p/dozen when bought in quantities of 15 instead of 1 dozen or by 5p/dozen when bought as part of a dozen instead of 1/2 dozen. I suspect that this interpretation is broadly true however it is certainly not the whole picture as the NFS minor food category 12901 which is labelled "eggs" contains many types of eggs (chickens, ducks, geese, quails, ... but not fish eggs which are in minor food category 12001) and these eggs vary in size, in quality (principally freshness I would guess), the location/shop type at which the purchase took place, and the organic/conventional nature of the production process.

To try to understand and the circumstances under which pictures like Figure 1 can tell us about the distribution of egg prices consider a single observation on a purchase and let the observed total quantity of eggs bought be q, at a total cost of x and unit price given by

$$p = \frac{x}{q}. \tag{1}$$

Let the vector $\boldsymbol{\xi}$ denote a list of all of the individual quantities purchased of all of the different types of egg (the quantity of each type is represented in a different element of $\boldsymbol{\xi}$). The total quantity of eggs is the sum over all the different types of eggs

$$q = \mathbf{1}'\boldsymbol{\xi} \tag{2}$$

and spending on eggs is

$$x = \boldsymbol{\rho}'\boldsymbol{\xi} \tag{3}$$

where

$$\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\xi}) \tag{4}$$

is a vector representing the individual prices of different types of eggs. The dependence of $\boldsymbol{\rho}$ on $\boldsymbol{\xi}$ reflects bulk discounting. However let us ignore this by assigning bulk discounted eggs to be yet another "type" and extending $\boldsymbol{\xi}$ to account for this.

The ultimate aim is to compare the prices paid by different households for identical eggs. In other words one would like to see the differences between $\boldsymbol{\rho}_h$ and $\boldsymbol{\rho}_i$ (where h and i subscripts index two households). However, neither $\boldsymbol{\rho}_h$ nor $\boldsymbol{\rho}_i$ are observable. Consider the observed unit price of any given purchase of eggs by household h

$$p_h = \frac{x_h}{q_h} = \frac{\boldsymbol{\rho}_h'\boldsymbol{\xi}_h}{\mathbf{1}'\boldsymbol{\xi}_h}. \tag{5}$$

This is, in effect, a price index which measures the cost of buying the vector $\boldsymbol{\xi}_h$ at prices $\boldsymbol{\rho}_h$ compared to the cost if all prices were one. Further, assuming cost-minimising economic behaviour and also that preferences over eggs are weakly separable,

$$p_h \leq \frac{c\left(\boldsymbol{\rho}_h, u\left(\boldsymbol{\xi}_h\right)\right)}{c\left(\mathbf{1}, u\left(\boldsymbol{\xi}_h\right)\right)} \tag{6}$$

so the observable price index must be less than or equal to the corresponding true cost-of-living/cost-of-eggs index. The denominator of the true index is hypothetical. It tells us the minimum cost of reaching $u\left(\boldsymbol{\xi}_h\right)$ if the prices of all of the different varieties and qualities of eggs were one. The reason for the difference between the right and left hand sides of this inequality is substitution between different egg types. The cost function in the denominator is defined by the problem

$$c(\mathbf{1}, u(\boldsymbol{\xi})) = \min_{\boldsymbol{\xi} \geq 0} \left\{ \mathbf{1}'\boldsymbol{\xi} : u(\boldsymbol{\xi}) \geq u \right\} \tag{7}$$

in which the all of the different eggs are available at the same price (one). This problem is formally identical to one of choosing a clutch of eggs subject only to an overall quantity restriction (since prices are one, total expenditure equals the total quantity). In the pure quality case the household would just choose the highest quality eggs available subject to the constraint that they can only have q eggs. So the demands in this hypothetical clutch of eggs are

$$\boldsymbol{\xi}_h^* = \mathbf{f}(\mathbf{1}, q_h) \tag{8}$$

and

$$\mathbf{1}'\boldsymbol{\xi}_h^* = \mathbf{1}'\mathbf{f}(\mathbf{1}, q_h) = c(\mathbf{1}, u(\boldsymbol{\xi}_h^*)) = q_h \tag{9}$$

and the associated level of (indirect) utility is

$$v(\mathbf{1}, q_h) = u(\boldsymbol{\xi}_h^*). \tag{10}$$

If one compares the utility of the clutch of eggs actually chosen with the hypothetical clutch which would be chosen subject only to the overall quantity restriction then

$$u(\boldsymbol{\xi}_h^*) = v(\mathbf{1}, q_h) \geq v(\boldsymbol{\rho}_h, x_h) = u(\boldsymbol{\xi}_h) \tag{11}$$

and one can measure how close the actual clutch of eggs comes to hypothetical one by determining the proportion of the observed total quantity of eggs which, if allocated ideally, would leave the household as well off as the actual clutch[1]. That is the overall quality of a clutch of eggs (χ_h) is implicitly defined by

$$v(\mathbf{1}, \chi_h q_h) \equiv v(\boldsymbol{\rho}_h, x_h) = u(\boldsymbol{\xi}_h) \tag{12}$$

where, by construction,

$$0 \leq \chi_h \leq 1. \tag{13}$$

It is only strictly possible to call χ_h “quality” if all of the different varieties/qualities of eggs were perfect substitutes. More generally χ_h is the distance function measuring the deviation from the hypothetical clutch and combines pure quality differences with the taste for variety which allows for variation in egg types. Also by definition

$$\chi_h q_h = c(\mathbf{1}, u(\boldsymbol{\xi}_h)) \tag{14}$$

so

$$p_h = \frac{\boldsymbol{\rho}_h'\boldsymbol{\xi}_h}{\mathbf{1}'\boldsymbol{\xi}_h} = \frac{c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h))}{c(\mathbf{1}, u(\boldsymbol{\xi}_h))}\chi_h \tag{15}$$

The unit price is equal to the true price index scaled by the overall quality index χ_h. If the eggs purchased given prices $\boldsymbol{\rho}_h$ were the ideal clutch then $\chi_h = 1$ and the observed unit price would be exactly equal to the true index.

[1] I am very grateful to David Ulph for suggesting this formulation.

Now consider the comparison of prices between two households indexed i and h. In general

$$\frac{p_h}{p_i} = \frac{c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h))}{\mathbf{1}'\boldsymbol{\xi}_h} \frac{\mathbf{1}'\boldsymbol{\xi}_i}{c(\boldsymbol{\rho}_i, u(\boldsymbol{\xi}_i))} \approx \frac{c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h))}{c(\mathbf{1}, u(\boldsymbol{\xi}_h))} \frac{c(\mathbf{1}, u(\boldsymbol{\xi}_i))}{c(\boldsymbol{\rho}_i, u(\boldsymbol{\xi}_i))} \tag{16}$$

and the ratio of unit prices serves as an approximation to a chained true index. Such a chained index lacks much in the way of economic intuition except under homotheticity within the egg clutch in which case

$$\frac{c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h))}{c(\mathbf{1}, u(\boldsymbol{\xi}_h))} \frac{c(\mathbf{1}, u(\boldsymbol{\xi}_i))}{c(\boldsymbol{\rho}_i, u(\boldsymbol{\xi}_i))} = \frac{a(\boldsymbol{\rho}_h)\, u(\boldsymbol{\xi}_h)}{a(\mathbf{1})\, u(\boldsymbol{\xi}_h)} \frac{a(\mathbf{1})\, u(\boldsymbol{\xi}_i)}{a(\boldsymbol{\rho}_i)\, u(\boldsymbol{\xi}_i)} = \frac{a(\boldsymbol{\rho}_h)}{a(\boldsymbol{\rho}_i)} = \frac{c(\boldsymbol{\rho}_h, u^*)}{c(\boldsymbol{\rho}_i, u^*)} \tag{17}$$

when the chained true index is also a unique true cross-section index. Hence under within-egg homotheticity,

$$\frac{p_h}{p^i} \approx \frac{c(\boldsymbol{\rho}_h, u^*)}{c(\boldsymbol{\rho}_i, u^*)} \tag{18}$$

where the quality of the approximation depends on the qualities of the eggs which the comparison households have bought. To see this note that, under homotheticity within the egg clutch,

$$v(\boldsymbol{\rho}_h, x_h) = \frac{x_h}{a(\boldsymbol{\rho}_h)} \quad \text{and} \quad v(\mathbf{1}, \chi_h q_h) = \frac{\chi_h q_h}{a(\mathbf{1})} \tag{19}$$

and

$$c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h)) = a(\boldsymbol{\rho}_h)\, u(\boldsymbol{\xi}_h) = x_h \quad \text{and} \quad c(\mathbf{1}, u(\boldsymbol{\xi}_h)) = a(\mathbf{1})\, u(\boldsymbol{\xi}_h) = \chi_h q_h \tag{20}$$

each of which imply

$$\chi_h = \frac{a(\mathbf{1})}{a(\boldsymbol{\rho}_h)} \frac{x_h}{q_h} = \frac{a(\mathbf{1})}{a(\boldsymbol{\rho}_h)} p_h \tag{21}$$

and quality is independent of the egg budget and p_h measures quality (up to a household-specific constant). This means that one can write

$$\frac{p_h}{p_i} = \frac{a(\boldsymbol{\rho}_h)}{a(\boldsymbol{\rho}_i)} \frac{\chi_h}{\chi_i} = \frac{c(\boldsymbol{\rho}_h, u^*)}{c(\boldsymbol{\rho}_i, u^*)} \frac{\chi_h}{\chi_i}. \tag{22}$$

Hence, under the assumption of homotheticity within the clutch of eggs (but not necessarily between eggs and other goods) the ratio of unit prices is equal to the product of a (unique) true price index and a quality index. In order for the ratio of unit prices to provide an exact cross-section index it is necessary to limit the type of cross-section price variation which is permissible. Suppose that each household's price vector can be written as a household-specific translation of each other's price vectors ($\boldsymbol{\rho}_i = \kappa \boldsymbol{\rho}_h$) then p_h/p_i is also an exact cross-section price index under homotheticity. To see this note that

$$\boldsymbol{\xi}_h = \nabla_{\rho} c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h)) = \nabla_{\rho} a(\boldsymbol{\rho}_h)\, u(\boldsymbol{\xi}_h) \tag{23}$$

which implies

$$p_h = \frac{c(\boldsymbol{\rho}_h, u(\boldsymbol{\xi}_h))}{\mathbf{1}'\boldsymbol{\xi}_h} = \frac{a(\boldsymbol{\rho}_h)}{\mathbf{1}'\nabla_{\rho} a(\boldsymbol{\rho}_h)}. \tag{24}$$

Using the fact that Hicksian demands are homogeneous of degree zero in prices and cost functions are homogeneous of degree one in prices

$$
\begin{aligned}
\frac{p_h}{p_i} &= \frac{a(\boldsymbol{\rho}_h)}{\mathbf{1}'\nabla_\rho a(\boldsymbol{\rho}_h)} \frac{\mathbf{1}'\nabla_\rho a(\boldsymbol{\rho}_i)}{a(\boldsymbol{\rho}_i)} \\
&= \frac{a(\boldsymbol{\rho}_h)}{\mathbf{1}'\nabla_\rho a(\boldsymbol{\rho}_h)} \frac{\mathbf{1}'\nabla_\rho a(\kappa\boldsymbol{\rho}_h)}{a(\kappa\boldsymbol{\rho}_h)} \\
&= \frac{a(\boldsymbol{\rho}_h)}{\kappa a(\boldsymbol{\rho}_h)} \\
\frac{p_h}{p_i} &= \frac{1}{\kappa}
\end{aligned}
\tag{25}
$$

and the ratio of unit prices is a unique exact cross-section index which measures the proportionality between the households' price vectors.[2]

To sum up: the ratio of unit prices is not an exact cross-section price index except under certain strong assumptions regarding preferences for different types of eggs and restrictions on the form of the cross-section price variation. In general, the ratio of unit prices only provides an approximation to an exact cross-section price index under homotheticity of demands within the eggs clutch. The quality of the approximation depends upon the cross-section differences in the quality of different households' clutches of eggs. If one further assumes that the relative price structure for eggs of different quality facing every household is identical but for a scale transformation, then the ratio of unit prices is an exact measure of the cross-section index. There is of course no evidence either way in the data as neither $\boldsymbol{\rho}$ nor $\boldsymbol{\xi}$ are observable. In what follows, therefore, I will take homotheticity within the egg clutch and constant *relative* prices for different egg types across households to be the identifying assumption which will allow us to describe cross-section variation in the price of eggs (and other foods).

3. Identifying variation in the prices of nutrients

The previous section discussed the comparison of unit prices and the problems of making like-with-like comparisons based on these data. This section discusses whether one might get around these issues by trying to uncover the household's valuation (termed their shadow price) of the nutrients contained in these minor foods. Consider an observation on a $K-$vector of food purchases q and the corresponding unit prices p. These food items contain a $J-$vector of nutrients z. In the National Food Survey minor foods are translated into nutrient vectors by means of Nutrient Conversion Factors – further details are given in the following section.. The relationship between the quantity vector and its nutrient content is

$$\mathsf{z} = \mathsf{A}'\mathsf{q} \tag{26}$$

where A is the $(K \times J)$ matrix of nutrient conversion factors. The vector z is a list of the total quantity of each of the J nutrients which consists in the K minor

[2] It is easy to relate this to the previous result by showing that the effect of this assumption is to remove cross-section variation in quality. Since $\chi_h = \frac{a(1)}{a(\rho_h)} \frac{\rho_h^0 \nabla_\rho a(\rho_h)}{1^0 \nabla a(\rho_h)}$ and (using $\boldsymbol{\rho}_i = \kappa\boldsymbol{\rho}_h$) we have $\chi_i = \frac{a(1)}{\kappa a(\rho_h)} \frac{\kappa\rho_h^0 \nabla_\rho a(\kappa\rho_h)}{1^0 \nabla_\rho a(\kappa\rho_h)} = \frac{a(1)}{a(\rho_h)} \frac{\rho_h^0 \nabla_\rho a(\rho_h)}{1^0 \nabla_\rho a(\rho_h)}$ which gives $\chi_i = \chi_h$.

foods. In the National Food Survey J (the number of nutrients) is 44 and K (the number of market goods) is 244 hence $J < K$. I what follows we will suppose that a household has preferences which can be defined in terms of both market goods and the characteristics, in this case nutritional attributes, of those goods. That is their preferences have the following structure:

$$u(\mathsf{q}) = v\left(\mathsf{A}'\mathsf{q}\right) = v(\mathsf{z}) \tag{27}$$

This is known as a linear characteristics structure and the idea is that households' preferences can be expressed in terms of both the foods themselves (market goods) and/or the nutrients (characteristics) they contain, and that characteristics are a linear function of market goods. This represents a restriction on a model in which preferences are defined solely over the market goods as the characteristics representation reduces the dimensions over which preferences are defined. What are the implications of this structure?

If the purchaser is optimising then q solves the problem

$$\min_{\mathsf{q}} \mathsf{p}'\mathsf{q} \text{ subject to } u(\mathsf{q}) = u \tag{28}$$

which has the first order condition

$$\frac{1}{\lambda}\mathsf{p} \geq \nabla u(\mathsf{q}) = \mathsf{A}\nabla v(\mathsf{z}) \tag{29}$$

Following Gorman (1956) define

$$\boldsymbol{\pi} = \lambda \nabla v(\mathsf{z}) \tag{30}$$

then the first order condition can be written as

$$\mathsf{p} \geq \mathsf{A}\boldsymbol{\pi} \tag{31}$$

The vector $\boldsymbol{\pi}$ represents the household's shadow prices/willingness to pay for the nutrients contained in the foods. The relationship between market prices of market goods and the shadow prices of nutrients (31) holds with equality for those goods actually purchased (it is worth remembering that the unit prices of foods which are not purchased are not observed). Suppose that we observe $\tilde{K}$ purchases ($\tilde{K} \leq K$). Then if $\tilde{\mathsf{p}}$ denotes the subset of $\tilde{K}$ unit prices

$$\tilde{\mathsf{p}} = \tilde{\mathsf{A}}\boldsymbol{\pi} \tag{32}$$

where $\tilde{\mathsf{A}}$ is the $\left(\tilde{K} \times J\right)$ submatrix of A corresponding to $\tilde{\mathsf{p}}$. The general solution to this linear equation system is

$$\boldsymbol{\pi} = \tilde{\mathsf{A}}^{+}\tilde{\mathsf{p}} \tag{33}$$

where $\tilde{\mathsf{A}}^{+}$ denotes the Moore-Penrose inverse of $\tilde{\mathsf{A}}$. This result relies on the fact that the matrix of nutrient conversion factors is, by construction, of full column rank. Notice that if the $\tilde{K}$ vector of purchased goods do not contain all of the nutrients then $\tilde{\mathsf{A}}$ will not have full row rank equal to J. If, say, only $\tilde{J}$ nutrients are purchased then only $\tilde{J}$ shadow prices can be identified as long as $\tilde{K} \geq \tilde{J}$. Note that the expression (33) requires the condition[3]

$$\text{rank}(\tilde{\mathsf{A}} : \tilde{\mathsf{p}}) = \text{rank}(\tilde{\mathsf{A}}) \tag{34}$$

in order to provide a solution to (32). In fact this condition (that the price vector is in the column space of the matrix of nutrient conversion factors) is implied by the characteristics structure of preferences and optimising behaviour in the form of the first order condition (31). If this rank condition does not hold then the characteristics structural assumption is invalid[4] and there exists no vector $\boldsymbol{\pi}$ such that (32) holds. There is however, some room for manoeuvre based on relaxing the model in terms of the number of dimensions over which preferences are assumed to be defined. Suppose that the rank condition (34) fails and consider some arbitrary vector $\boldsymbol{\pi}^*$. We know that[5]

$$\mathsf{p} \neq \mathsf{A}\boldsymbol{\pi}^* \tag{35}$$

but that we can always write

$$\mathsf{p} = \mathsf{A}\boldsymbol{\pi}^* + \boldsymbol{\varepsilon}^* \tag{36}$$

where we define

$$\boldsymbol{\varepsilon}^* = \mathsf{p} - \mathsf{A}\boldsymbol{\pi}^* \tag{37}$$

This is equivalent to

$$\mathsf{p} = [\mathsf{A} : \mathsf{I}_K] \begin{bmatrix} \boldsymbol{\pi}^* \\ \boldsymbol{\varepsilon}^* \end{bmatrix} \tag{38}$$

so if we simply augment the A matrix with an I_K matrix it is the case that

$$\text{rank}(\mathsf{A} : \mathsf{I}_K : \mathsf{p}) = \text{rank}(\mathsf{A} : \mathsf{I}_K) \tag{39}$$

and we have restored the rank condition albeit for the augmented matrix $[\mathsf{A} : \mathsf{I}_K]$. One possible way to interpret this is that there are not J but $J + K$ measurable characteristics associated with K foods; J nutritional attributes, and K dimensionless food-specific attributes which are unique to each food. Thus only Eggs possess "Egginess", only Honey possess the attribute "Honeyness", only New Zealand Butter possesses "New-Zealand-Butteriness", and only Frozen Convenience Fish possesses "Frozen-convenience-Fishiness" (to pick some NFS categories at random). The $\boldsymbol{\varepsilon}^*$vector might then be called the shadow prices of these characteristics. This (not very) fancy footwork guarantees the rank condition holds, but since there are still only a maximum of K equations in the system and now there are $K + J$ unknowns, the vector of shadow prices cannot be identified. Perhaps another way to put it is that *any* $\boldsymbol{\pi}-$vector can be rationalised with these characteristics, simply by choosing the right $\boldsymbol{\varepsilon}-$vector. Another interpretation which also give the rank condition, but one which of course ends up in the same place, relies on rewriting (38) as

$$\mathsf{p} = [\mathsf{A} : \boldsymbol{\varepsilon}^*] \begin{bmatrix} \boldsymbol{\pi}^* \\ 1 \end{bmatrix} \tag{40}$$

[3]See Magnus and Neudecker (1988), p.Theorem 11, p. 36.

[4]Note that the converse is not true; the rank condition is a necessary but not sufficient condition for the characteristics structure.

[5]Assuming, without loss of generality but with loss of notation, that all all goods are purchased.

This relies on arguing that $\boldsymbol{\varepsilon}^*$ is a common attribute across all foods, but that it is present in different amounts and carries a shadow price of one. This different interpretation is of no consequence - the informational content is the same both ways and the shadow prices are still unidentified.

A standard empirical response to this sort of problem is to choose $\boldsymbol{\pi}$ (call the choice $\hat{\boldsymbol{\pi}}$) in order to minimise the difference between p and $\mathsf{A}\hat{\boldsymbol{\pi}}$. The key to implementing this approach lies in the decision regarding an appropriate measure of differences; a typical choice is sum of squared residuals, another might be the sum of absolute deviations. The first, for example, would suggest $\hat{\boldsymbol{\pi}} = \mathsf{A}^+\mathsf{p}$ as the estimator[6], which is the same as if the rank condition had not failed, but is fundamentally identical to either of the schemes outlined above since this defines

$$\hat{\boldsymbol{\varepsilon}} = \mathsf{p} - \mathsf{A}\hat{\boldsymbol{\pi}} \tag{41}$$

and $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\varepsilon}}$ can serve as $\boldsymbol{\pi}^*$ and $\boldsymbol{\varepsilon}^*$ in (38) and (40). It should be remembered that when the rank test fails this method of estimating $\boldsymbol{\pi}$ is governed by a criterion which, if we take the model of household preferences seriously, aims to minimise the sum of squares of food-specific shadow prices. It is not all that obvious, to me, why this is a relevant criterion.

Given that precise values of $\boldsymbol{\pi}$ are unidentified when the fundamental rank condition fails, an alternative approach is to try bound them. The system (38) provides no bounds as it stands, but, if we are prepared to make an assumption, then bounds are recoverable. One, not totally offensive, assumption might be that the consumer has non-decreasing marginal valuations of foods and characteristics, i.e. that they are freely disposable. This might be true of market goods (if you were mad enough to buy something you didn't want then you could always not use it/throw it away), but is doubtless not true of nutritional characteristics as they cannot easily be unbundled and disposed of. Nevertheless, this is the assumption we need to make in order to recover a bound on $\boldsymbol{\pi}$ and $\boldsymbol{\varepsilon}$. The reason is that we can then solve (38) as a feasible linear program and solve

$$\max_{\pi_j} \pi_j \text{ subject to } \mathsf{p} = [\mathsf{A} : \mathsf{I}_K] \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\varepsilon} \end{bmatrix} \text{ and } \pi_{i \neq j} \geq 0,\ \boldsymbol{\varepsilon} \geq \mathsf{0}_K \text{ for } j = 1, ..., J$$

which returns the maximum willingness to pay for each nutrient given the assumptions embodied by the constraints. In the empirical work below this is the strategy which will be followed if/when the rank condition (34) fails.

4. An empirical investigation

This section provides some empirical evidence on the matters discussed above. The data are drawn from the 2000 National Food Survey. These data are first described and then are used to look first at the extent of cross-sectional price variations subject to the identifying assumptions required as discussed in section 2, and then at the cross-section variation in the prices of nutrients using the linear characteristics methods describe in section 3.

[6]Which is identical to OLS.

4.1. Data. The National Food Survey is a continuously run stratified sampling survey of households in Great Britain and North Ireland in which households record the amounts and costs of food entering the household during a one week period. The National Food Survey was first established in 1940 to provide information on household food purchases and the nutritional value of the domestic diet for the urban working class. In 1950 the coverage of the survey was widened to the population of Great Britain as a whole and has since been widened to include Northern Ireland. The survey arose from concerns about the quality of diet of the urban working class during food rationing in the 1940's. The NFS has run continuously since 1942 and is believed to be the longest running continuous sampling survey in the world.

In the 2000 NFS the responding sample was 6,700 individual households. The person, male or female, principally responsible for domestic food arrangements provides information about each household. That person is referred to as the main diary keeper. The main diary keeper keeps a record, with guidance from an interviewer, of all food, intended for human consumption, entering the home each day for seven days. The following details are noted for each food item: the description, quantity (in either imperial or metric units) and in respect of purchases the cost. Food items obtained free from a farm or other business owned by the household member or from the hedgerow, a garden or allotment is recorded only at the time it is used. To avoid the double counting of purchases, gifts of food and drink are excluded if a donating household bought them. On a separate questionnaire, details are entered of the characteristics of the family and its members.

The energy value and nutrient content of food obtained for consumption in the home are evaluated using special tables of food composition. The nutrient conversion factors are mainly based on values given in Holland *et al* (1991) and its supplements. The conversion factors are revised each year to reflect changes as a result of any new methods of food production, handling and fortification, and also to take account of changes in the structure of the food categories used in the Survey e.g. changes in the relative importance of the many products grouped under the heading of reduced fat spreads. The nutrient factors used make allowance for inedible materials such as the bones in meat and the outer leaves and skins of vegetables. For certain foods, such as potatoes and carrots, allowance is also made for seasonal variations in the wastage and/or nutrient content. Further allowances are made for the expected cooking losses of thiamin and vitamin C; average thiamin retention factors are applied to appropriate food items within each major food group and the (weighted) average loss over the whole diet is estimated to be about 20 per cent. The losses of vitamin C are set at 75% for green vegetables and 50% for other vegetables. However, no allowance is made for wastage of edible food.

4.2. Variation in prices of foods. The variation in prices within minor food categories is hard to summarise because there is so much of it. I have calculated the coefficient of variation for each minor food and Figure 3 shows the density of its distribution[7]. The mean is 0.6 so, on average, the standard deviation of unit prices is about 60% of the mean, and about 5% of foods have a coefficient of variation which

[7]Only those foods with at least 100 observations were used (leaving 203 foods on which this figure is based).

is larger than the mean price. No foods have a zero coefficient of variation hence all exhibit some degree of variability. In general meats and meat products have the highest degree of variability (an average coefficient of variation of 0.87 amongst these foods). Dairy products and eggs and, perhaps surprisingly, alcoholic drinks exhibit the lowest variation (all around 0.45 on average).

Figure 3: The density of the distribution of the coefficient of variation of food prices

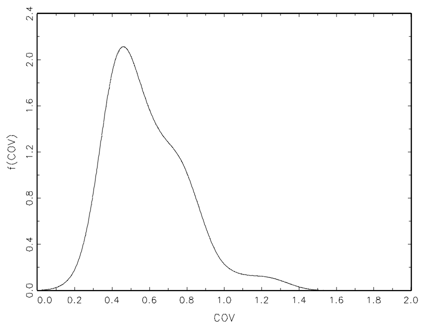


Table 2: Proportion of statistically significant correlations between unit price and household characteristics.

	Positive correlation	Zero	Negative correlation
Family Income	0.6471	0.3480	0.0049
Age of Head of Household	0.2255	0.6324	0.1422
# adults	0.0539	0.8676	0.0784
# children	0.0539	0.5833	0.3627
Pensioners in household	0.0735	0.7598	0.1667
Rural Area	0.1569	0.7402	0.1029
Unemployed	0.0245	0.6373	0.3382
Wales	0.0343	0.8725	0.0931
Scotland	0.1078	0.7990	0.0931
Northern	0.0147	0.7353	0.2500
Yorkshire & Humberside	0.0147	0.6520	0.3333
North Western	0.0539	0.7892	0.1569
East Midlands	0.0245	0.8186	0.1569
West Midlands	0.0343	0.8186	0.1471
South West	0.0490	0.8627	0.0882
East Anglia	0.0490	0.8922	0.0588
Greater London & South East	0.5588	0.4216	0.0196
Northern Ireland	0.3088	0.6176	0.0735

Table 2 reports the proportion of foods for which the unit price is, respectively positively correlated, uncorrelated and negatively correlated (at 90% confidence level) with a range of household and regional variables. For example, prices are positively correlated with family income for 64.71% of foods, negatively correlated with income for 0.49% of foods and insignificantly uncorrelated for the remaining 34.80%. This is the result of calculating the pairwise correlation coefficient between the unit prices and each characteristics individually for each minor food category in the NFS 2000. This gives an overall picture which says that the prices which households pay for foods is positively related to their family income. Note that these correlations are calculated separately for each minor food category. This means (subject to the qualifications outlined in Section 2) that we are comparing like purchases as much as is possible with these data. Subject to the identifying assumptions I have made there is no confounding effect from quality differences. There is a less clear effect with age with 22.55% of foods showing a positive correlation between prices and the age of the head of households and 14.22% showing a negative correlation. The presence of more children in the household seems to be negatively correlated with prices more than it is positively correlated. Prices are also positively correlated with living in a rural area[8] in 15.69% of foods, but negatively correlated in 10.26% of foods, so whilst some foods seem to be more expensive in the countryside, others are cheaper. Of the regional variables the ones which really stand out on the positive side are Greater London & South East (where 55.88% of foods are more expensive) and Northern Ireland. On the negative side the Northern and Yorkshire & Humberside regions have the largest proportions of prices which are negatively correlated with living in that area. Whilst these sorts of general patterns do seem apparent, what is also apparent is the heterogeneity in these correlations across foods – for example, roughly as many foods have prices which are positively correlated with living in Scotland as do those which are negatively correlated.

To describe general between-household differences in food prices, it is necessary to construct household level food price indices. The simplest way to create a household level price index for food is to form

$$\frac{\mathsf{p}_h'\mathsf{q}_h}{\overline{\mathsf{p}}'\mathsf{q}_h} \tag{42}$$

for each household where $\overline{\mathsf{p}}$ is a vector made up of the national average prices of each commodity. This compares the cost of the household's observed food purchases with the cost it would face had it made exactly the same set of purchases whilst facing the set of average prices for each item. The closer the household's purchase prices are to the national average, the closer this measure will be to one. One problem with this object, however, is that variation in this price index across different households will occur because of variations in both prices (p_h) and demand patterns (q_h) so that two households facing the same prices would have different price indices if their patterns of food purchases differed. An alternative index might be

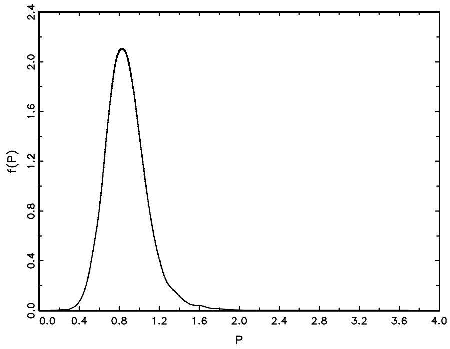
$$\frac{\mathsf{p}_h'\overline{\mathsf{q}}}{\overline{\mathsf{p}}'\overline{\mathsf{q}}} \tag{43}$$

which compares the cost of buying the national average basket of goods ($\overline{\mathsf{q}}$) at average prices, to the cost at the prices faced by the household. This index would give two

[8] Local authority districts with fewer than 0.5 people per acre.

household facing identical prices identical price indices. The problem is that not all households buy all the foods, whereas all the foods are bought by at least some households. This means that the $\bar{\mathsf{q}}$ vector will give weight to some foods which the household does not buy and for which, therefore, there is no corresponding price observation in p_h. This is a real practical problem as the only source of information from which one might impute a value for the missing price is the data generated by other households. Any imputation method base on these data would tend to mask the heterogeneity which we are primarily interested in. This might not be a big problem if it were a question of the odd missing element here or there, but in fact missing elements would be the norm as most households buy far fewer foods than the maximum possible of 244. Price index numbers like (43) are therefore not practicable, without a good deal of aggregation across households, so in what follows (42) is used subject to the caveats outlined. The density of the distribution of household food price indices is shown in Figure (4). In all 10% of the sample have a food price index no more than 0.63 (which means that one in ten households, roughly speaking, face food prices at least 37% lower than average), and 10% of the sample have a food price index of more than 1.13 (which means, roughly speaking, that one in ten households face food prices at least 13% higher than average).

Figure 4: The density of the distribution of household food price indices



The relationship between household food price indices and some key variables are summarised in Table 3. This table shows the results of a linear regression of the food price index on a set of observable household variables; this should be treated as descriptive rather than behavioural – the aim is to provide a simple linear representation of food prices which explains as much of the observed variation around the mean as possible. This regression model explains 95.29% of the variation in food prices seen in the data.

From Table 3 we see that: the food price index generally increases with family income, but does so at a declining rate; the food price index decreases with the age of the head of household, but the rate of decrease slows as age increases; large households those with lots of adults and children have lower than average food prices – the effect of an extra child on food prices seems to be lower than the effect of an extra adult; given the inclusion of an age variable the presence of pensioners in

Table 3: Household food price indices and household characteristics.

	Coefficient	Standard Error	t
Family Income	0.0003143	0.0000184	17.05
Family Income^2	-4.03×10^{-8}	7.04×10^{-9}	-5.73
Age of Head of Household	-0.0055438	0.0009657	-5.74
Age of Head of Household^2	0.0000567	9.45×10^{-6}	6.00
# adults	-0.0545205	0.0040974	-13.31
# children	-0.0355354	0.002637	-13.48
Pensioners in household	-0.0077462	0.0090548	-0.86
Rural Area	0.0050782	0.0076534	0.66
Unemployed	-0.0376757	0.0105506	-3.57
Wales	0.9993008	0.0256803	38.91
Scotland	1.010723	0.0246198	41.05
Northern	0.9535178	0.0250249	38.10
Yorkshire & Humberside	0.9632946	0.0245963	39.16
North Western	0.9785832	0.0243137	40.25
East Midlands	0.9690285	0.0250759	38.64
West Midlands	0.9723389	0.0246047	39.52
South West	0.9940284	0.0245931	40.42
East Anglia	0.9934389	0.0262843	37.80
Greater London & South East	1.045921	0.023354	44.79
Northern Ireland	1.068724	0.0254475	42.00

the household has no significant association with food prices; there is no association between the food price index and living in a rural area; given income, lower food prices are associated with a head of household who is unemployed; higher food price indices are found in Scotland, London and the South East and Northern Ireland. Most of these associations are in line with what was found in the pattern of pairwise correlations in the prices of the individual minor foods. To summarise it seems to be richer-than-average, younger-than-average, and smaller-than-average households in either London and the South East or Northern Ireland who have the highest food price indices.

Table 4 controls for the composition of food purchases and presents regional food price indices for each of the UK standard regions. The figures are based upon

$$\frac{\overline{\mathsf{p}}_r'\mathsf{q}}{\overline{\mathsf{p}}'\mathsf{q}} \tag{44}$$

in which $\overline{\mathsf{p}}_r$ is a regional average price vector, and the other components are as previously defined.

London and the South East and Northern Ireland emerge as the regions with the most expensive food with food there being 10.5% and 4% more expensive than

Table 4: Regional food price indices, 2000.

Region	Food Price Index (National Average=1)	*Std Error*
London & South East	1.105	*0.019*
Northern Ireland	1.040	*0.030*
East Anglia	0.976	*0.054*
Scotland	0.975	*0.034*
South West	0.974	*0.034*
North West	0.964	*0.030*
East Midlands	0.947	*0.037*
West Midlands	0.939	*0.038*
Yorkshire and Humberside	0.933	*0.034*
Wales	0.924	*0.055*
North	0.912	*0.038*

average in those areas. Similarly Wales and North of England appear at the bottom of the list. One surprising feature is the overall spread from top to bottom with food in London and the South East being about 20% more expensive than in the North of England.

4.3. Variation in prices of nutrients. The rank condition (34) on the nutrient conversion factors and purchase unit prices failed for all of the households in the data. None of the households in the 2000 NFS have a linear characteristics structure to their preferences for foods – at least with the set of measured nutritional characteristics in the NFS nutrient conversion tables. By allowing each minor food to have its own shadow value and imposing the assumption that shadow prices are non-negative (as described above) it is possible to recover upper bounds on each individual household's shadow values of nutrients. Table 5 shows descriptive statistics of the upper bounds on the shadow prices for each of the 44 nutrients listed in the tables. The shadow price bounds are measured in pence per unit, and the units are recorded in the left hand column next to the name of the nutrient. For example, the median upper bound on the shadow price of iron is 3.07 pence per milligram, which means that half of the sample have a maximum willingness to pay for iron of just over 3p per milligram, or an expected shadow price of 1.5p per milligram if we assume that the household's shadow prices are realisations of a uniform random variable over the interval. The first thing to note is that the number of shadow price bounds recovered varies across nutrients. The shadow price of alcohol could only be recovered for 1891 households out of a sample of 6585 (there are 6700 households in the 2000 NFS, 115 didn't make any food purchases and so are dropped) i.e. about a quarter of households, whilst the shadow price of carbohydrate, for example, was recovered for 6,519 households (99%). This is simply due to the fact that only the shadow prices of those nutrients which the household is observed to buy can be recovered by this method, and alcohol is a relatively infrequently purchased nutritional characteristic, whilst carbohydrate, calcium and others are very often present in the foods households buy.

The most highly valued nutrients (at least as far as one can tell from the maximum willingness to pay) seem to be: sugars (other than those listed[9]) which are present in yoghurts, infant milks, general fats such as creamed coconut, coconut butter and certina branded food drinks like Horlicks or Ovaltine; maltose which is also partcularly present in branded food drinks, as well as marmarlade and jam; copper, for which livers and nuts are especially good sources; B vitamins (thiamin, riboflavin, and B6) in which meat and yeast extracts, breakfast cereals (especially whole grain types), milk, liver, rice, nuts and eggs are rich; and Vitamin D which is present particularly in milk and sea fish. It should be remembered that these valuations cannot be identified precisely because of the failure of the rank condition and that the Tables indicate the maxima of ranges which all start at zero.

Table 6 is the counterpart of Table 2. It shows the proportions of respectively, significant positive, insignificant, and significant negative correlations between the upper bounds of the nutrient valuations and the same range of variables. The results for, for example, family income show that the (upper bounds of the) valuations of 4.55% nutrients are positively correlated with family income (at 90% confidence), 84.1% have no significant correlation with income either way, and the rest are negative correlated. The middle column which shows the proportions of insignificant correlations tends to have the biggest numbers but there are a number of family characteristics which seem to show similar correlations across nutrients. The number of adults and children in the households seem to be correlated with lower maximum willingness to pay, and living in London and the South East seems to be positively correlated.

The NFS records recommended daily intakes (RDI's) of 18 nutrients[10] for each individual in the survey. These vary principally by age, sex and whether or not the individual is pregnant. Combining these within households gives a household level aggregate RDI figure for this subset of nutrients (denote this 18×1 vector by R_h). These can be combined into an RDI willingness to pay index

$$\frac{\boldsymbol{\pi}_h' \mathsf{R}_h}{\overline{\boldsymbol{\pi}}' \mathsf{R}_h} \tag{45}$$

where $\boldsymbol{\pi}_h$ is the household's shadow price vector, and $\overline{\boldsymbol{\pi}}$ is some reference shadow price vector. The reference vector is taken to be made up of the median shadow price of each nutrient (i.e. the relevant elements of the column of medians in Table 5). The object is the counterpart to equation (42) and suffers from the same drawbacks in that variations in RDI's across households contribute to variations in the index, just as much as variations in the shadow prices of the nutrients. The descriptive statistics for this index are given in table 7 and the density of its distribution is shown in Figure (5).

[9] i.e. not Glucose Fructose, Sucrose, Maltose or Lactose

[10] Vegetable protein, Calcium, Retinol, Carotene, Riboflavin, Nicotinic acid, Tryptophan,Vitamin C, Folate, Starch, Glucose, Zinc, Vitamin B6, Vitamin B12, Phosphorus, Manganese Biotin, Pantothenic acid

Table 5: Upper bounds on the shadow prices of nutrients, descriptive statistics, pence-per-unit.

Nutrient	n	*Mean*	$p\,(10)$	$p\,(25)$	$p\,(50)$	$p\,(75)$	$p\,(90)$
Vegetable protein (g)	6507	1.33	0.40	0.61	0.83	1.12	1.74
Animal protein (g)	6500	1.42	0.67	0.98	1.18	1.48	1.97
Fatty acid saturates (g)	6518	1.90	0.38	0.65	1.19	1.87	3.21
Fatty acid monounsaturates (g)	6516	2.36	0.26	0.68	1.36	2.60	4.33
Fatty acid polyunsaturates (g)	6518	5.11	0.39	0.74	2.40	7.37	11.63
Carbohydrate (g)	6519	0.17	0.04	0.07	0.13	0.19	0.28
Energy $(kcal)$	6519	0.03	0.01	0.01	0.02	0.03	0.05
Calcium (mg)	6519	0.05	0.02	0.03	0.04	0.05	0.06
Iron (mg)	6519	4.37	1.48	2.03	3.07	4.34	6.29
Retinol (μg)	6488	0.16	0.01	0.02	0.04	0.10	0.19
Carotene (μg)	6506	0.05	0.00	0.00	0.01	0.04	0.09
Retinol equivalent (μg)	6510	0.09	0.00	0.01	0.02	0.06	0.13
Thiamin (mg)	6519	32.35	10.78	16.30	24.19	34.19	49.11
Riboflavin (mg)	6519	27.07	12.49	15.50	19.51	26.50	34.77
Nicotinic acid (mg)	6501	9.33	2.25	3.22	5.53	8.66	14.71
Tryptophan (mg)	6500	0.09	0.04	0.06	0.08	0.10	0.14
Niacin equivalent (mg)	6519	2.50	1.00	1.44	2.18	3.02	4.19
Vitamin C (mg)	6512	1.01	0.16	0.24	0.43	0.74	1.43
Vitamin D (μg)	6497	26.01	1.76	2.47	5.98	16.85	58.69
Folate (μg)	6519	0.18	0.06	0.08	0.13	0.20	0.32
Sodium (mg)	6519	0.02	0.00	0.01	0.01	0.02	0.02
Starch (g)	6502	0.53	0.07	0.11	0.16	0.22	0.33
Glucose (g)	6506	3.22	0.55	0.83	1.70	2.62	3.67
Fructose (g)	6443	6.39	0.60	1.00	1.68	2.57	4.42
Sucrose (g)	6505	2.57	0.04	0.34	0.73	1.22	1.94
Maltose (g)	6287	44.56	2.55	4.81	9.88	22.44	40.91
Lactose (g)	6480	7.10	0.71	0.78	0.95	1.32	1.53
Other sugars (g)	6247	61.96	2.21	3.79	9.88	25.38	122.75
Non milk extra sugar (g)	6369	8.72	0.04	0.15	0.41	0.78	1.49
Alcohol (g)	1891	5.01	2.60	3.53	4.54	5.78	7.48
Fibre southgate (g)	6505	1.89	0.57	0.90	1.33	1.92	2.83
Fibre englyst (g)	6505	2.90	0.85	1.29	1.93	2.96	4.94
Potassium (mg)	6519	0.021	0.01	0.01	0.01	0.02	0.03
Magnesium (mg)	6519	0.20	0.071	0.10	0.17	0.26	0.35
Copper (mg)	6512	42.50	13.62	23.46	33.01	44.11	62.01
Zinc (mg)	6518	7.47	3.05	4.14	6.23	9.14	11.92
Vitamin B6 (mg)	6519	26.00	7.78	10.79	17.86	30.12	55.88
Vitamin B12 (μg)	6484	6.37	2.06	2.35	3.20	5.69	11.73
Phosphorus (mg)	6519	0.04	0.02	0.03	0.04	0.052	0.06
Manganese (mg)	6507	15.89	3.22	4.53	7.06	13.98	21.20
Biotin (μg)	6519	1.55	0.47	0.77	1.22	1.87	2.53
Pantothenic acid (mg)	6519	8.67	3.68	5.43	7.16	10.11	13.15
Vitamin E (mg)	6516	7.05	0.46	0.81	3.45	8.59	14.57
Cholestrol (mg)	6502	0.23	0.03	0.05	0.11	0.30	0.50

Table 6: Proportion of statistically significant correlations between maximum willingness to pay and household characteristics .

	Positive correlation	Zero	Negative correlation
Family Income	0.0455	0.8409	0.1136
Age of Head of Household	0	0.6136	0.3864
# adults	0	0.0455	0.9545
# children	0	0.0682	0.9318
Pensioners in household	0	0.5682	0.4318
Rural Area	0.0227	0.7727	0.2045
Unemployed	0.1136	0.7727	0.1136
Wales	0	0.9545	0.0455
Scotland	0.0455	0.9545	0
Northern	0.0227	0.9318	0.0455
Yorkshire & Humberside	0.0682	0.8864	0.0455
North Western	0.0227	0.9545	0.0227
East Midlands	0.0455	0.9318	0.0227
West Midlands	0.0455	0.8636	0.0909
South West	0.0455	0.7273	0.2273
East Anglia	0.0682	0.9318	0
Greater London & South East	0.5227	0.4773	0
Northern Ireland	0.0227	0.7273	0.2500

Figure 5: The density of the distribution of RDI indices.

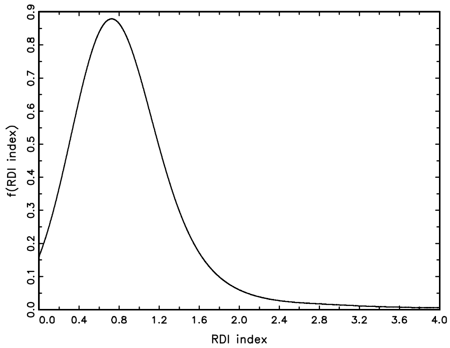


Table 7: The cost of household RDI's, descriptive statistics, pence-per-household.

	n	$Mean$	$Std\ Dev$	Percentiles						
				5	10	25	50	75	90	95
RDI index	6519	1.00	1.73	0.37	0.44	0.58	0.76	1.42	1.93	5.20

Table 8: The cost of household RDI's as a linear function of household characteristics.

	Coefficient	Standard Error	t
Family Income	0.0003337	0.0001566	2.13
Family Income^2	-7.71$\times10^{-8}$	5.97$\times10^{-8}$	-1.29
Age of Head of Household	-0.0263432	0.0082359	-3.20
Age of Head of Household^2	0.0001947	0.0000806	2.42
# adults	-0.2325474	0.0348806	-6.67
# children	-0.1863265	0.022424	-8.31
Pensioners in household	-0.1265312	0.077077	-1.64
Rural Area	0.0032233	0.0651473	0.05
Unemployed	0.1151681	0.0895706	1.29
Wales	2.143054	0.2191738	9.78
Scotland	2.207019	0.2098792	10.52
Northern	2.156001	0.2133906	10.10
Yorkshire & Humberside	2.289666	0.2095713	10.93
North Western	2.141585	0.2074546	10.32
East Midlands	2.286558	0.2137011	10.70
West Midlands	2.178317	0.2098085	10.38
South West	2.117279	0.2099693	10.08
East Anglia	2.257563	0.2243532	10.06
Greater London & South East	2.313766	0.1992628	11.61
Northern Ireland	2.246527	0.2168834	10.36

From Table 8 we see that, just as with the food price index, the RDI index increases with income albeit at a declining rate, and decreases with age with the rate of decrease slowing. Increased numbers of household members, both adults and children, decrease the RDI index. London and the South East remains the region with the highest index followed by East Anglia. The general patterns in willingness to pay for nutrients are similar to those for foods.

Recall that some of the variation described in Table 8 relates to variations in RDI's as well as variations in the shadow prices of nutrients. Table 9 controls for the composition of RDI's and presents regional RDI indices for each of the UK standard regions. The figures are based upon

$$\frac{\overline{\pi}_r' \overline{\mathsf{R}}}{\overline{\pi}' \overline{\mathsf{R}}} \tag{46}$$

in which $\overline{\pi}_r$ is a regional average shadow price vector (calculated from the mid-points of the bounds), and the other components are as previously defined. This is the counterpart of (44).

Table 9: Regional RDI cost indices, 2000.

Region	RDI Price Index (National Average=1)	*Std Error*
Greater London & South East	1.102	0.042
East Anglia	1.054	0.111
Yorkshire & Humberside	1.025	0.076
East Midlands	1.025	0.085
West Midlands	1.004	0.075
Northern	0.990	0.090
Scotland	0.981	0.075
Northern Ireland	0.948	0.066
Wales	0.902	0.097
North Western	0.899	0.071
South West	0.878	0.074

The regions with the highest shadow cost of reaching the national average RDI's for these nutrients are London and the South East, and East Anglia (10% and 5% above average, respectively). Whilst Northern Ireland was relatively high up the list of price indices is about 5% below average for the shadow price index for RDI's.

5. Conclusions

This paper tries to describe the variation in the prices of foods and the prices of nutrients for a representative sample of UK households. The data are unit values – expenditure divided by quantity purchased – for groups of similar foods. This presents an immediate problem in identifying cross section variation in prices because these food groups are not entirely homogenous. There is undoubtedly some variation of the foods which comprise the observed food groups across different households, but two identifying assumptions on household demand behaviour and the nature of within-food-group price variation render this problem irrelevant for the present purpose. These assumptions are duly made and the variation in food prices described. Overall it seems that richer-than-average and younger-than-average households, particularly those which have fewer than average members and which live in Northern Ireland or London and the South East pay the most for food. The paper also develops a simple method for uncovering households' individual valuations of the nutritional characteristics present in the foods they buy. The precise identification of this requires a certain rank condition to hold between the matrix of nutrient contents of foods and their prices. In practice this rank condition fails for every sample household, and precise valuations cannot be recovered from these data. However, by making two assumptions – that households marginal valuations of nutrients and foods are non-negative, and that each food type has its own unique attribute – it is possible to recover bounds on the valuations of nutrients. These assumptions are made and the variation in these bounds with respect to household circumstances is described.

References

[1] DEFRA (2001), *National Food Survey 2000*, http://www.defra.gov.uk/esg/m_publications.htm

[2] Gorman, W.M. (1956), "A possible procedure for analysing quality differentials in the egg market", *London School of Economics*, mimeo, reprinted in *Review of Economic Studies*, 47, 843-856 (1980).

[3] Holland, B., A. A. Welch, I. D. Unwin, D. H. Buss, A. A. Paul and D. A. T. Southgate (1991), *The Composition of Foods* 5th edition, Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food, Royal Society of Chemistry: London

[4] Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus*, Revised Edition (1999), Wiley Series in Probability and Statistics: Chichester.