FOOD PREFERENCES SEGMENTATION USING AN AIDS/MIXTURE APPROACH

Matthieu H. Arnoult, Ariane Kehlbacher, C.S. Srinivasan, Rachel McCloy, Richard Tiffin

Contributed Paper prepared for presentation at the 89th Annual Conference of the Agricultural Economics Society, University of Warwick, England
13 - 15 April 2015

FOOD PREFERENCES SEGMENTATION USING AN AIDS/MIXTURE APPROACH

MATTHIEU H. ARNOULT[1,4], ARIANE KEHLBACHER[2],
C.S. SRINIVASAN[2], RACHEL MCCLOY[3], RICHARD TIFFIN[1]

Abstract. Excess weight is a problem affecting over half of the British population, with some categories being more at risk than others, in particular in lower socio-economic groups. In that respect, differentiated dietary behaviours are known to contribute to inequalities in health outcomes. Segmentation is increasingly employed as a means of better targeting policy interventions. While conventional segmentation methods divide the population according to their dietary choices or according to socio-demographic characteristics, a potential flaw in this approach is that people may choose to consume a bad diet for entirely different reasons, or that people from different socio-demographic groups may behave in a similar fashion. We use a novel alternative approach which seeks to segment according to peoples dietary preferences. The method estimates a finite mixture of AIDS. We identify segments which have a degree of homogeneity in their food purchases, while remaining heterogeneous in terms of their socio-demographics. The homogeneity of food purchases within components is less than within components identified using k-means clustering of food choices. We argue that this approach will lead to more effective targeted interventions because they would appeal to the reasons for bad dietary choices rather than the choices themselves.

Keywords: dietary guidelines, household food consumption, segmentation, finite mixtures, demand model

[1] Centre for Food Security, School of Agriculture, Policy and Development, University of Reading, UK
[2] Department of Food Economics and Marketing, School of Agriculture, Policy and Development, University of Reading, UK
[3] Centre for Applied Behavioural Science, School of Psychology & Clinical Language, University of Reading, UK
[4] Correspondence to: Centre for Food Security, University of Reading, Earley Gate, P.O. Box 237, Reading, RG6 6AR, UK, email: m.h.p.arnoult@reading.ac.uk, phone +44 (0)118 378 8154

1. Introduction

As of 2011, excess weight is a problem affecting over 60% of the population in England, with some groups being more at risk than others, in particular in lower socio-economic groups (DH, 2011). In that respect, differentiated dietary behaviours are known to contribute to inequalities in health outcomes (SACN, 2008; DH, 2011). Various microeconomic studies have investigated the link between socio-economic status and food consumption (Johansson et al., 1999; Drewnowski et al., 2007; Darmon and Drewnowski, 2008; Maguire and Monsivais, 2015), and generally conclude that there is a positive relationship between socio-economic position and healthy eating. Personal preferences however remain unaccounted for when segregating according to demographic criteria: Johansson et al. (1999) have found that individuals who claimed to pay attention to their diet had healthier dietary habits, while Drewnowski et al. (2007) argue for behavioural as well as economic interventions to tackle poor eating habits.

Attempts have been made to assess the determinants of food choice in subgroups of the general population, using variables such as income (Park et al., 1996; Han and Wahl, 1998), or age and education level (Cortez and Senauer, 1996). These studies segment their samples prior to analysis, with the former estimating demand systems, and the latter analysing demand changes over time using non-parametric techniques. These analyses have limitations however. Firstly as groupings are exogenously imposed on the population without regard to food preferences within or across groups; and secondly as not all food categories are considered, the focus being often on fruit and vegetables (Han and Wahl, 1998; Bertail and Caillavet, 2004; 2008).

In the present study, we argue that focussing on socio-demographics as a means of segmenting the population is potentially limiting. Individuals who are similar in their socio-demographics may have very different motivations regarding their diet and health. Equally segmenting according to the diet is also potentially inadequate as people choosing a similar diet may do so for entirely different reasons.

When designing interventions to improve diets it is important to address the reasons why people are making bad choices. We therefore use a segmentation approach that differs from previous studies by basing the segmentation on differences in peoples preferences. We do this by assuming that the population can be subdivided into groups of individuals according to food preferences. Within each group preferences are homogeneous and can be represented by the same individual utility function for each member of the group. Differences in preferences between the different groups are reflected in differences in the utility functions across groups. The differences in utility function across groups mean that a different AIDS is defined for each. The classification of individuals into the groups is unknown in advance however. We therefore embed an AIDS within a finite mixtures model to allow us to simultaneously estimate the classification and each of the AIDS for the different groups. Our model extends an earlier study by Bertail and Caillavet (2004; 2008) by including the full set of food items consumed within a household, and secondly by using a Bayesian approach which allows us to account for censoring arising from infrequency of purchase.

2. Method

We combine an AIDS model with a finite mixture model to discriminate components of the population based on their underlying preferences. We assume that the differences in preferences across components will be reflected in differences in the parameters of the AIDS expenditure function, and hence in the share equations which are estimated.

We follow Tiffin and Arnoult (2010) in specifying the AIDS for Bayesian estimation with infrequency of purchase. In this model the usual set of AIDS share equations is augmented with a set of probit equations to model households where no purchase is observed. In the share equations two types of latency arise. First, for the households making no purchase in the period for which they are observed, a latent quantity is used to represent the fact that consumption may occur from previous purchases. Second, where purchases are made, they need to be adjusted to account for the fact

that some stocks may be carried over beyond the period for which the household is observed. The probit equations are also expressed in terms of latent variables which are the continuous counterpart to the observed binary variable. The full demand system incorporating both the share and probit equations is written:

$$\mathbf{z}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

where $\mathbf{z}^* = \left(\mathbf{s}^{*'}, \mathbf{y}^{*'}\right)$. We define $\mathbf{s}^*$ and $\mathbf{y}^*$ as the latent consumption shares and latent continuous probit variables respectively, and $\mathbf{X}$ is block diagonal:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X_1} & 0 \\ 0 & \mathbf{X}_2 \end{pmatrix} \tag{2}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are respectively a matrix of prices and expenditure, and a vector of constants. The share equations are reparameterised to satisfy symmetry, homogeneity and adding-up restrictions (for further details see Tiffin and Arnoult, 2010). The error term $\mathbf{e} = (\mathbf{v}', \mathbf{u}')'$ is assumed to follow a multivariate normal distribution, $\mathbf{e} \sim MVN(0, \Sigma)$.

We assume that there are a fixed number of components $j = 1, \ldots, k$ in the population, each with a different set of food preferences. Each of these components can be represented with a different AIDS as:

$$\mathbf{z}_j^* = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{e}_j \qquad j = 1, \ldots, k \tag{3}$$

where the subscript j indicates the subset of $\mathbf{z}^*$, $\mathbf{X}$ and $\mathbf{e}$ for observations classified as belonging to the jth component, and the coefficient vector $\boldsymbol{\beta}$ varies between components. A finite mixtures model is used to represent this collection of models across the whole population of $i = 1, \ldots, N$ households.

Assuming the same prior for each component, the posterior distribution for a given classification is given by:

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{c}) = \prod_{j=1}^{k} \theta_j^{\sum_{i=1}^{m} c_{ij}} \prod_{i=1}^{N} \left(\boldsymbol{\theta}_j p\left(z_i^*|\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\theta}_j, \boldsymbol{D}, \boldsymbol{c}_i\right) p(\boldsymbol{\beta}) p(\boldsymbol{\Sigma}) p(\boldsymbol{\theta})\right) \tag{4}$$

where c_{ij} is a binary variable which has the value one when household i is classified in component j, and zero otherwise; $0 \leq \boldsymbol{\theta}_j \leq 1$ is the proportion of the sample that belongs to component j; $\boldsymbol{D}$ is the data (both latent and observed); and $p(\boldsymbol{\beta})$, $p(\boldsymbol{\Sigma})$ and $p(\boldsymbol{\theta})$ are the priors. The following conditional posterior distributions, which form the basis of the Gibbs sampler, are obtained from equation 4:

$$
\begin{aligned}
(c_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\theta}, \boldsymbol{D}) &\sim Multin\Big(1, \theta_1 p_1(\cdot), \ldots, \theta_j p_j(\cdot)\Big) \\
&\quad \forall\, i = 1, \ldots, N;\ j = 1, \ldots, k && (5)\\
(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{c}, \boldsymbol{D}) &\sim Dirichlet\left(\sum_{i=1}^{N} c_{i1}, \sum_{i=1}^{N} c_{i2}, \ldots, \sum_{i=1}^{N} c_{ik}\right) && (6)\\
\left(\boldsymbol{\beta}_j|\boldsymbol{\Sigma}_j, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{D}_j\right) &\sim MVN\Big(\left(\boldsymbol{\Sigma}_j^{-1} \otimes \mathbf{X}_j'\mathbf{X}_j\right)^{-1} \\
&\quad \left(\boldsymbol{\Sigma}_j^{-1} \otimes \mathbf{X}_j'\right)\mathbf{z}_j^*, \boldsymbol{\Sigma}_j^{-1} \otimes \mathbf{X}_j'\mathbf{X}_j\Big) && (7)\\
\left(\boldsymbol{\Sigma}_j|\boldsymbol{\beta}_j, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{D}_j\right) &\sim IW\left(\mathbf{e}_j'\mathbf{e}_j, N_j\right) && (8)\\
\left(s_{ij}^*|\boldsymbol{\Sigma}_j, \boldsymbol{\beta}_j, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{D}_{-i}\right) &\sim N\left(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j\right) \qquad \forall\, i = 1, \ldots, N_j && (9)\\
\left(y_{ij}^*|\boldsymbol{\Sigma}_j, \boldsymbol{\beta}_j, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{D}_{-i}\right) &\sim TN^{neg}(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j) \text{ for } y_i = 0;\ j = 1, \ldots, k && (10)\\
\left(y_{ij}^*|\boldsymbol{\Sigma}_j, \boldsymbol{\beta}_j, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{D}_{-i}\right) &\sim TN^{pos}(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j) \text{ for } y_i = 1;\ j = 1, \ldots, k && (11)
\end{aligned}
$$

where $p_j(\cdot)$is the normal density function parameterised for the jth component and $\boldsymbol{D_{-i}}$ is the data omitting the observed variable for which the conditional is defined. Based on these distributions, the following Gibbs sampling algorithm is defined in order to draw a sample from the full posterior distribution:

(0) Randomly allocate households into the components

(1) Within each component , estimate the AIDS coefficients:

  (a) Draw $\boldsymbol{\beta}_j$ using equation 7

  (b) Draw $\boldsymbol{\Sigma}_j$ using equation 8

  (c) Draw latent $\mathbf{z}^*$ using equations 9, 10 and 11

(2) For all components, draw classification c_i using equation 5

(3) Draw weights $\boldsymbol{\theta}$ using equation 6 and compute the probabilities for the multinomial distribution.

(4) Repeat steps 1 to 4 using the newly drawn classification matrix in 2 to allocate households into components, allowing to compute new AIDS coefficients, etc.

3. Data

We use the UK government's Living Costs and Food Survey (LCF) for 2011. Participating households voluntarily fill in a food diary over a 2-week period where all food purchases for home consumption are recorded. The sample is based on 5,692 households in 638 postcodes sectors stratified by Government Office Region (GOR), National Statistics Socio-Economic Classification (NS-SEC), and car ownership (ONS, 2013). For our analysis we compute consumption per adult equivalent, using the OECD equivalisation scale provided in the LCF.

The LCF survey categorises 258 food items as 'food/drink brought home' or 'takeaway brought home', which we have aggregated into 5 major food groups: dairy & eggs; meat & fish; fats, starches, etc. (FSE hereafter); fruit, vegetables & nuts (F&V hereafter); and drinks. Censoring levels are low, ranging from 0.6% for fats, starches, etc., up to 6.9% for drinks (see Table 1 for details on censoring levels and mean expenditure shares).

Table 1. Expenditure shares of the aggregated food groups; censoring levels

	dairy	**meat**	**FSE**	**F&V**	**drinks**
mean share	0.135	0.232	0.246	0.215	0.171
std dev	0.086	0.120	0.109	0.100	0.146
std err	0.001	0.002	0.001	0.001	0.002
censoring	1.4%	3.5%	0.6%	1.6%	6.9%

4. Results

The number of components (k) is predetermined in our model. We found setting $k = 5$ gave the clearest classification: for $k > 5$ some components were so small that estimation was not possible. The components are robust with high recurrence of households within each component at different iterations of the sampler. Table 2 summarises the 5 components obtained: components 1 and 2 are rather large (34% and 42% of the population), while the others are much smaller, accounting for under a quarter of the population. Component 5 in particular stands at just 3%.

Own-price and expenditure elasticities are computed using the estimated coefficients of the component-specific AIDS models, for each household at every iteration of the sampler. Figure1 reports the mean of these for each component.[1]

The top panel of figure 1 shows the own-price elasticities for each of the five components and the food groups within these components. The results show that there is a clear downward gradient from component 1 to 5 for the own-price elasticity of dairy & eggs demand. The reverse is true for fats, starches, etc. The other own-price elasticities are broadly similar across components, except for fruit & vegetables and drinks where demand is more elastic in component 5. These results potentially indicate that preferences for dairy & eggs as well as for fats, starches, etc. are influential in determining the segmentation.

The second panel shows the deviation of expenditures on food per adult equivalent. Notable departures from the mean include component 5 where quantities purchased are above average for dairy & eggs, and below average for meat & fish and fruit & vegetables. Component 4 has below average purchases of all five food groups whilst component 1 is above average for all groups except dairy & eggs.

The bottom panel shows the expenditure elasticities. The widest variation across food groups is exhibited by component 5. Their expenditure elasticity for dairy & eggs is lower than the other components whilst that for meat & fish and fruit & vegetables is higher. This suggests that this group may be constrained by income

[1]Own-price elasticities and expenditure elasticities are presented for all components in appendix tables 3 and 4 (p.18). We also report quantities purchased and associated expenditures in appendix tables 5 and 6 (p.19).

TABLE 2. Component structure

	Cmp1	Cmp2	Cmp3	Cmp4	Cmp5	ALL
n	1919	2355	645	554	182	5655
$n\%$	34%	42%	11%	10%	3%	100%

in their purchase of meat & fish and fruit & vegetables, whilst their purchase of dairy & eggs is very insensitive to both expenditure and price changes. Of the other food groups, fats, starches, etc. has most variation between components with values increasing from component 1 to 5. The higher value for the expenditure elasticity in component 1 is also noteworthy.

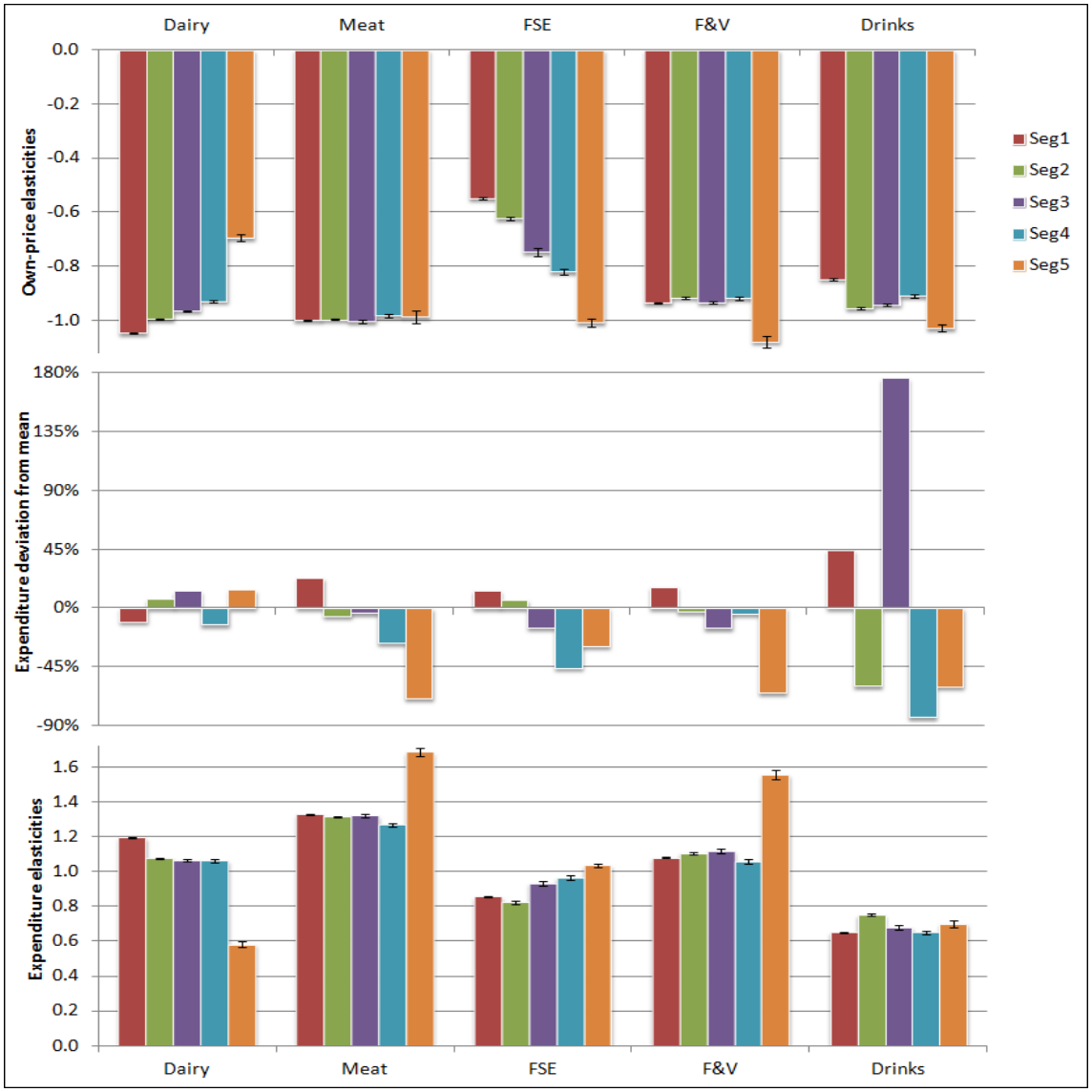

FIGURE 1. Own-price Marshallian elasticities, expenditure figures and expenditure elasticities (from top to bottom; 95% confidence intervals)

Figures 2 and 3 show the nutrient intake implied by each segment's diet. These were calculated using the conversion tables associated with the LCF and are based on consumption at and away from home. The figures also report the recommended maximum and minimum intakes (SACN, 2008; DH, 1991). In the case of alcohol these represent the recommended intakes for men and women respectively (NHS, 2015). The nutrient intakes shown in figure 2 are expressed as relative values of total daily energy intake. For all categories of fats except polyunsaturated fats (PUFAs) and for non-milk extrinsic sugars (NME sugars), the diets of every component are over the recommended limits. There is little variation in nutrient intakes across components. Figure 3 shows the intake of nutrients and alcohol where the recommendations are expressed as absolute daily values. There is more variation across components for all categories considered. For all 'nutrients' except alcohol (and to a lesser extent fruit & vegetables), components 1 to 3 have higher intakes than components 4 and 5: looking at daily energy intake, components 1-3 consume from 3,200 kCal to nearly 3,600 kCal per day whereas components 4 and 5 consume only 2,200 kCal and 2,500 kCal respectively (the sample average is 3,240 kCal). Components 1-3 also exceed the recommended maximum intake of sodium (component 1 has almost twice the recommended daily amount) and cholesterol. Components 1-3 exceed the minimum recommended daily intake of dietary fibre. Components 4 and 5 have lower levels of food consumption and as such they do not exceed the recommended cholesterol intake but they do exceed the recommendation for sodium although component 5's do not do so significantly. Both components 3 and 4 fail to meet the dietary fibre recommendations. No component manages to achieve the daily 400g target ("5-a-day") for fruit and vegetable consumption. Component 1 comes close at 382g, followed by components 2 and 4 at about 350g; component 3 consumes 290g, almost three-quarters of the recommended intake, while component 5 consumes only 122g (1.5 portions). Alcohol consumption varies most across components: components 2, 4 and 5 have very low consumption levels, under the

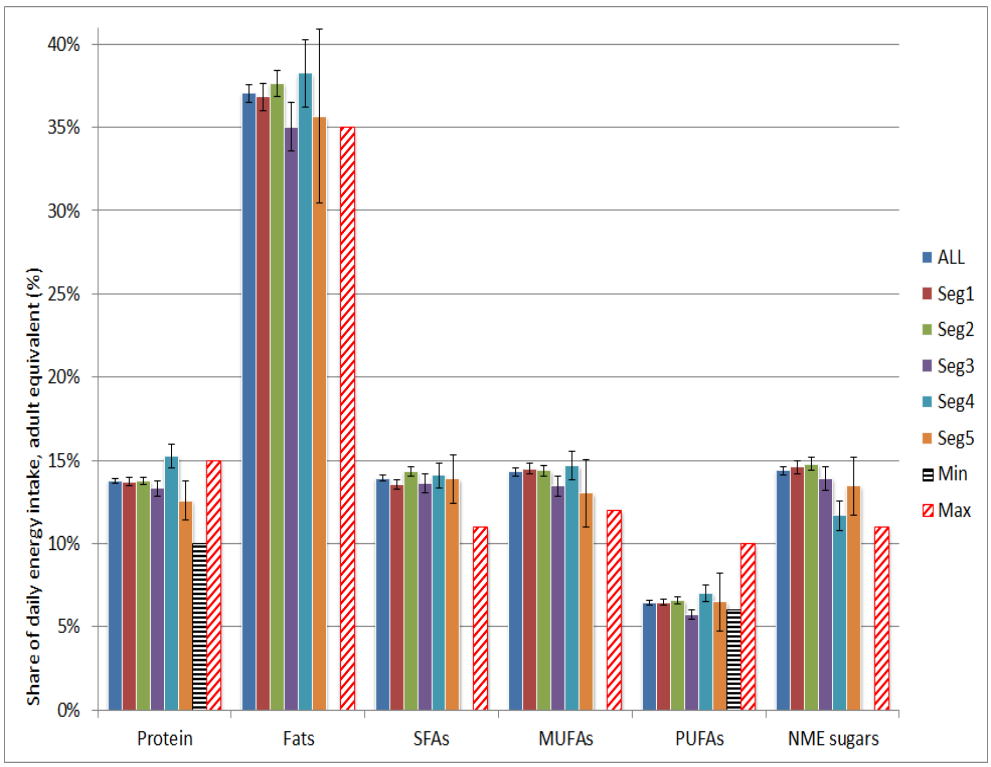

FIGURE 2. Nutrient intake for protein, fats, NME sugars expressed as a percentage of daily energy intake per adult equivalent (bars represent standard errors)

17.5 units per week considered as a safe amount for women,[2] while component 3 consumes 41 units per week, well above the safe amount recommended for men (24.5 units); component 1's consumption is within safe limits at 19 units per week.

We use 2 indicators of diet healthiness developed by the USDA to assess our components' consumption patterns, namely the USDA Healthy Eating Index (HEI; Guenther et al., 2013) and the USDA Score (Volpe and Okrent, 2012). The USDA healthy eating index is scored out of 100 and reflects a household's diet compliance to healthy eating guidelines. For each nutrient a household is awarded a maximum score of 5 to 20 according to which nutrient is considered if it fully complies with recommended intake. Points are removed if consumption departs from the guidelines, potentially reaching 0. Individual nutrient scores are summed and normalised to 100 which is the score achieved by fully compliant households. The USDA score compares a household's expenditure shares to an accepted set of values for broad food categories (dairy, meat, vegetables, etc.); deviation from those accepted values will lower the

[2]The NHS lower risk guidelines recommend no more than 2 to 3 units daily for women, and 3 to 4 units for men; the mid-point intake has been used as a reference here (NHS, 2015).

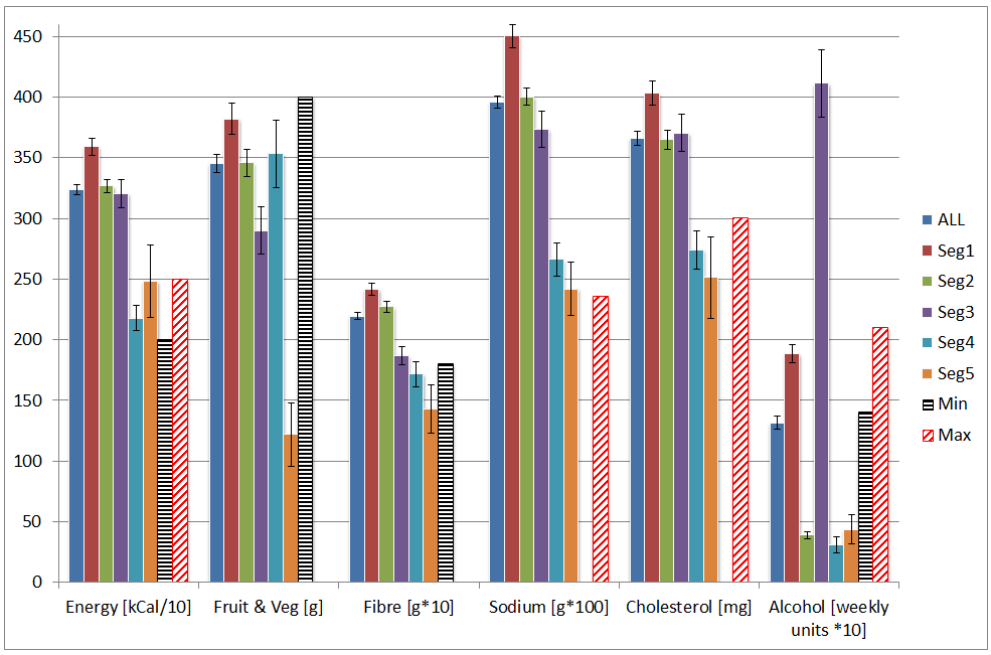

FIGURE 3. Nutrient intake for energy, fruit & vegetable, fibre, sodium, cholesterol, alcohol, per adult equivalent (bars represent standard errors, max and min for alcohol represent recommended intakes for men and women respectively)

overall score for each household. Both methodologies have been adapted to the UK: for the HEI, healthy eating guidelines from SACN (2008) and Department of Health (1991) have been applied, while for the USDA score, suggested expenditure shares in the UK have been extrapolated using the Eatwell plate (PHE, 2014; PHE, 2015). As the Eatwell plate does not include alcohol consumption, this has been excluded from the HEI. In each case, score and index are computed for each household and then averaged at the sample or component level.

Both approaches have been applied to the different baskets of goods purchased by our 5 components, and results are summarised in Figure 4 below and corresponding Appendix Table 8 (p.21), where results have been normalised using the full sample as reference to facilitate comparison.

The USDA healthy eating index does not differ much between components. It shows that component 5 has the worst diet and component 4 has the best. The USDA score varies more between components. Component 2 has the most balanced diet whilst component 5 has the worst balance. This is because component 5 has a low

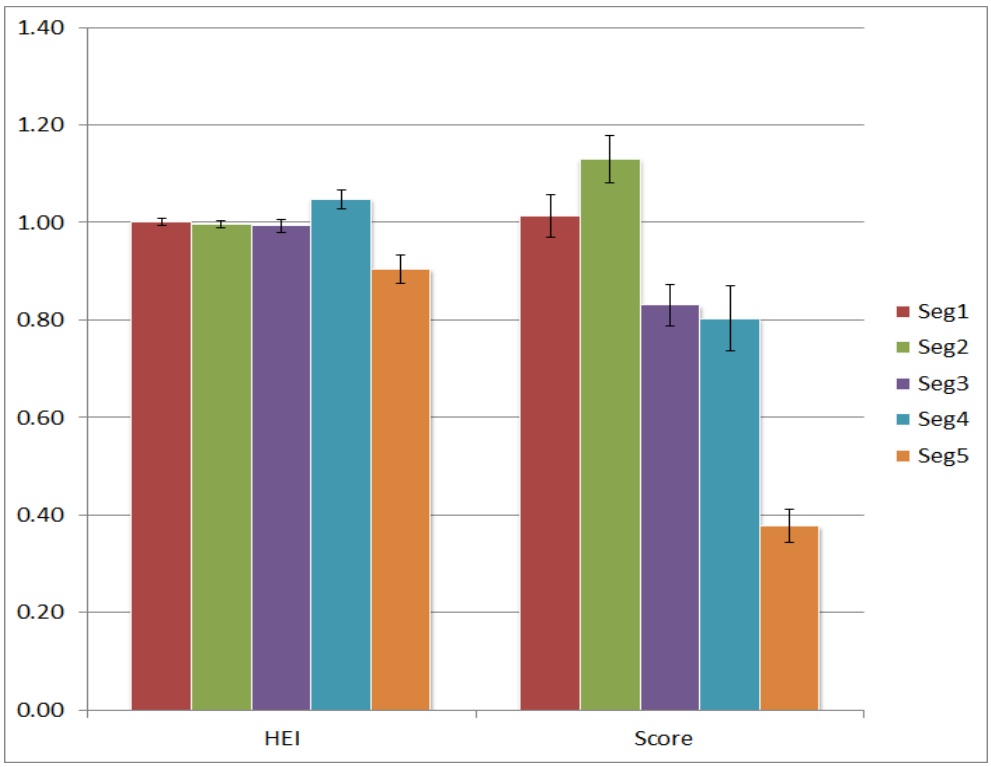

FIGURE 4. USDA measures of diet healthiness, normalised (sample = 1.00)

intake in fruit & vegetables, a relatively high intake in dairy, and a low consumption of fish.

While our segmentation is not reliant upon socio-demographic characteristics, it remains possible to make use of those variables when characterising our components. Component 1 has the highest median income and the highest proportion of households belonging to higher socio-economic classes and higher socio-economic groups. It has the highest proportionate number of workers[3] per household. Component 1 also has the highest proportion of married or cohabiting couples, and we also find a high proportion of children in their teenage years. Component 2 is the largest of all 5 components, and as such has a purchase pattern and socio-demographic features which are close to the sample average. It does however have a higher proportion of retired people and people of lower socio-economic class. Component 3 has the second highest median income and second highest proportion of high SEC/SEG households; it has the second lowest number of children per household, more and older workers per household than average; it is the least ethnically mixed component

[3]Defined as "self-employed, full or part-time employee, or in a government training programme" (ONS, 2013).

(98% classified as white British against a 93% average), with a higher proportion in Yorkshire and the East. Component 4 has the second lowest median income, the highest proportion of pensioners, and is the most ethnically diverse (16% non-white British), with a higher presence in London and the South East, probably indicating an urban population.Component 5 is comprised of the second smallest households, with the lowest number of workers as a proportion of the number of adults. It has the highest number of children per household, with a large proportion of under 5, and the highest proportion of single/divorced or widowed households (over 60%), making this probably a component of single parents. It also has the lowest median income and the highest proportion of lower SEC/SEG. There is a higher presence in the North West and the Midlands.

5. Conclusions

We have used household food consumption data to estimate a demand model which reflects differences in preference between five groups of households in the population. These components have differences in their price and expenditure elasticities. These differences suggest that preferences for dairy & eggs, and fats, starches, etc. food may be influential in differentiating between groups. The differences between groups may however be due to more fundamental cognitive characteristics, such as a framing bias or hyperbolic discounting which might lead consumers to have different utility functions. Components also differ in their consumption patterns, and to a lesser extent their nutrient intakes. Of the 5 components identified, we have isolated 2 components of low consuming households, one of which with very low fruit & vegetables consumption and high dairy & eggs intake (mostly due to liquid milk); a third component is characterised by a very large alcohol intake. It is possible to identify socio-demographics features of the components mainly in terms of income, socio-economic class and group, with components of a higher status being associated with healthier diet. Within components however there is some heterogeneity in socio-demographic composition. We also find that households with similar consumption patterns (be it for food or nutrients) and similar healthiness scores appear in different

components. This accords with our initial observation that households can have similar diets but for different reasons which are reflected in their utility functions. From a policy perspective, the ability to identify which component a particular individual belongs to is clearly desirable. This is one of the attractions of segmenting the population according to socio-demographic characteristics or diet. Because our method relies on unobservable preferences associating individuals with the appropriate component is more difficult. Interventions will need to appeal to the underlying cognitive characteristics which in part determine an individual's preferences so that people effectively self select into the intervention. In further research we will seek to elicit cognitive biases such as resistance to framing, recognizing social norms, etc. (Bruine de Bruine et al., 2007; Kirby et al., 1999).

We re-segmented our sample using a more conventional approach. Applying k-means clustering to household choices regarding the 5 major food yields 5 components which, by and large, do not overlap with our own. Members of components 4 and 5 in the mixture model tend to be found in a combined cluster defined by low levels of consumption in the k-means analysis. The k-means analysis also identifies a component of heavy drinkers though it is much smaller than segment (a fifth in size). The remaining segments identified with k-means clustering do not readily map onto the components of mixture model.

The segmentation method we have introduced identifies groups of consumers based on the reasons for their dietary choices as opposed to the choices themselves. We find that there is not as much variation in diet across segments as one might expect which suggests that people may be choosing similar diets for similar reasons. This suggests that a differentiated approach to the design of interventions which appeals to this reasoning may be successful. In further work we aim to explore the reasoning in more detail and to experiment with possible interventions based on this.

References

Bertail, Patrice, and France Caillavet (2004) "Food consumption and poverty: A segmentation approach." Working paper CORELA/INRA 04-06.

Bertail, Patrice, and France Caillavet (2008) "Fruit and Vegetable Consumption Patterns: A segmentation approach." *American Journal of Agricultural Economics* **90**(3):827–842.

Bruine de Bruine, Wändi, Andrew Parker, and Baruch Fischhoff (2007) "Individual differences in adult decision-making competence." *Journal of Personality and Social Psychology* **92**(5):938–956.

Cortez, Rafael, and Ben Senauer (1996) "Taste changes in the demand for food by demographic groups in the United States: A nonparametric empirical analysis." *American Journal of Agricultural Economics* **78**(2):280–289.

DH (1991) *Dietary reference values for food energy and nutrients in the United Kingdom.* Department of Health, Report on health and social subjects No.41. London: HMSO.

DH (2011) *Healthy lives, healthy people: A call to action on obesity in England.* London: Department of Health.

Darmon, Nicole, and Adam Drewnowski (2008) "Does social class predict diet quality?" *American Journal of Clinical Nutrition* **87**(5):1107–1117.

Drewnowski, Adam, Pablo Monsivais, Matthieu Maillot, and Nicole Darmon (2007) "Low-energy-density diets are associated with higher diet quality and higher diet costs in French adults." *Journal of the American Dietetic Association* **107**(6):1028–1032.

Guenther, Patricia M., Kellie O. Casavale, Jill Reedy, Sharon I. Kirkpatrick, Hazel A.B. Hiza, Kevin J. Kuczynski, Lisa L. Kahle, and Susan M. Krebs-Smith (2013) "Update of the Healthy Eating Index: HEI-2010." *Journal of the Academy of Nutrition and Dietetics* **113**(4):569–580.

Han, Tong, and Thomas I. Wahl (1998) "China's rural household demand for fruit and vegetables." Journal of Agricultural and Applied Economics **30**(1):141–150.

Johansson, Lars, Dag S. Thelle, Kari Solvoll, Gunn-Elin Aa. Bjørneboe, and Christian A. Drevon (1999) "Healthy dietary habits in relation to social determinants and lifestyle factors." *British Journal of Nutrition* **81**(3):211–220.

Kirby, Kris N, Nancy M. Petry, and Warren K. Bickel (1999) "Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls." *Journal of Experimental Psychology: General* **128**(1):78–87.

Johansson, Lars, Dag S. Thelle, Kari Solvoll, Gunn-Elin Aa. , and Christian A. Drevon (1999) "Healthy dietary habits in relation to social determinants and lifestyle factors." *British Journal of Nutrition* **81**(3):211–220.

Maguire, Eva R., and Pablo Monsivais (2015) "Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data." *British Journal of Nutrition* **113**(1):181–189.

NHS (2015) *Alcohol units and guidelines. The lower risk daily guidelines.* Change4Life. National Health Service website.

ONS (2013) *Living Costs and Food Survey.* Technical Report for survey year 2011. London: Office for National Statistics.

PHE (2014) *Your guide to the eatwell plate.* Information leaflet. London: Public Health England.

PHE (2015) *The eatwell plate: How to use it in promotional material.* London: Public Health England & Government Digital Service.

Park, John L., Rodney B. Holcomb, Kellie C. Raper, and Oral Capps, Jr. (1996) "A demand systems analysis of food commodities by U.S. households segmented by income." *American Journal of Agricultural Economics* **78**(2):290–300.

SACN (2008) *The nutritional wellbeing of the British population.* London: Scientific Advisory Committee on Nutrition & The Stationery Office.

Tiffin, Richard (2006) "Bayesian clustering of farm types using the mixtures model." *Journal of Agricultural Economics* **57**(3):547–562.

Tiffin, Richard, and Matthieu Arnoult (2010) "The demand for a healthy diet: Estimating the almost ideal demand system with infrequency of purchase." *European Review of Agricultural Economics* **37**(4):501–521.

Tiffin, Richard, and Ariane Kehlbacher (2015) "Using mixtures to model zero consumption." Working paper. Department of Food Economics and Marketing, University of Reading, UK.

Volpe, R., and A. Okrent (2012) "Assessing the healthfulness of consumers' grocery purchases." Economic Information Bulletin No.102. Washington DC: Economic Research Service, US Department of Agriculture.

Appendix

A.1 Elasticities

Table 3. Own-price Marshallian elasticities; mean values

	Cmp1	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**
dairy & eggs	-1.049	-0.998	-0.965	-0.930	-0.697
meat & fish	-1.001	-0.999	-1.006	-0.985	-0.988
fats, starches, etc.	-0.550	-0.624	-0.748	-0.822	-1.010
fruit, vegetables & nuts	-0.937	-0.918	-0.935	-0.919	-1.081
drinks	-0.848	-0.957	-0.944	-0.912	-1.031

Table 4. Expenditure elasticities; mean values

	Cmp1	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**
dairy & eggs	1.195	1.076	1.067	1.063	0.582
meat & fish	1.329	1.317	1.322	1.269	1.687
fats, starches, etc.	0.857	0.823	0.933	0.966	1.036
fruit, vegetables & nuts	1.079	1.106	1.118	1.060	1.558
drinks	0.651	0.751	0.679	0.652	0.700

A.2 Quantities & Expenditures

TABLE 5. Quantities purchased; deviation from sample mean

	Cmp1	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**
dairy & eggs	-14%	11%	4%	-14%	37%
meat & fish	21%	-4%	-5%	-28%	-70%
fats, starches, etc.	8%	8%	-17%	-42%	-10%
fruit, vegetables & nuts	12%	1%	-14%	-9%	-58%
drinks	39%	-35%	85%	-70%	-44%

TABLE 6. Expenditures; deviation from sample mean

	Cmp1	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**
dairy & eggs	-10%	7%	13%	-13%	14%
meat & fish	23%	-6%	-4%	-27%	-69%
fats, starches, etc.	13%	6%	-15%	-46%	-29%
fruit, vegetables & nuts	16%	-3%	-15%	-4%	-65%
drinks	44%	-60%	176%	-84%	-60%

A.3 Nutrient Intake

TABLE 7. Daily nutrient intake per adult equivalent across segments; for protein, fats and sugar, intake is given as a percentage of total daily energy intake; minima and maxima taken from DH (1991), SACN (2008) and NHS (2015)

	ALL	**Cmp1**	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**	**Min**	**Max**
Protein	13.8%	13.7%	13.8%	13.3%	15.2%	12.6%	10%	15%
Total Fats	37.0%	36.8%	37.6%	35.0%	38.2%	35.7%		35%
SFAs	13.9%	13.5%	14.3%	13.6%	14.1%	13.9%		11%
MUFAs	14.3%	14.5%	14.4%	13.5%	14.7%	13.0%		12%
PUFAs	6.5%	6.5%	6.6%	5.8%	7.0%	6.5%	6%	10%
NME sugars	14.4%	14.6%	14.8%	13.9%	11.7%	13.4%		11%
Energy [kCal]†	3238.4	3591.4	3267.8	3205.1	2177.2	2483.9	2000	2500
Fruit & Veg [g]	345.2	382.0	345.6	290.0	353.4	121.8	400	
Fibre [g]	21.9	24.1	22.7	18.7	17.2	14.3	18	
Sodium [g]	3.96	4.50	4.00	3.73	2.66	2.42		2.36
Cholesterol [mg]	366.0	403.2	365.0	370.2	274.2	251.3		300
Alcohol [weekly units]	13.1	18.8	3.9	41.1	3.1	4.4	17.5	24.5

Values are computed using 'eating-out' data available in the LCF diary
† min and max are recommendations for women and men respectively

A.4 Healthy Index Scores

Table 8. USDA and NHS measures of diet healthiness, normalised (sample = 1.00)

	All	**Cmp1**	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**
USDA HEI	63.3	63.4	63.1	62.9	66.4	57.3
USDA Score	14.9	15.1	16.8	12.4	12.0	5.6