AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Food preference segmentation using an AIDS mixture: An application to the UK

Ariane Kehlbacher[1,4], Matthieu Arnoult[2], Chittur Srinivasan[1], Rachel McCloy[3], and Richard Tiffin[2]

[1] Department of Food Economics and Marketing, School of Agriculture, Policy and Development, University of Reading, UK

[2] Centre for Food Security, School of Agriculture, Policy and Development, University of Reading, UK

[3] Centre for Applied Behavioural Science, School of Psychology and Clinical Language, University of Reading, UK

[4] Corresponding author: a.kehlbacher@reading.ac.uk

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Copyright 2015 by Kehlbacher et al. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract. Levels of obesity and overweight in the UK are high with certain groups within the population particularly affected. The customary approach is to identify at risk groups based on their socio-demographics or their observed unhealthy food choices. This approach fails to acknowledge that households with similar socio-economic background may behave very differently and that households make unhealthy food choices for very different reasons. In this study we segment households according to their underlying food preference using an Almost Ideal Demand System (AIDS) mixture model. We identify five household segments that are similar in their food preferences and therefore in how they would respond to policy interventions. The food purchasing patterns of households in each of the five segments tend to be similar but households differ in terms of socio-demographics. This information needs to be taken into account when designing a targeting mechanism for policy interventions to improve diets.

Keywords: household food consumption, segmentation, finite mixture model, Almost Ideal Demand System

1. Introduction

Over 60% of the population in England are overweight with some individuals being more at risk than others (DH, 2011). Differences in dietary behaviours are known to contribute to inequalities in health outcomes (SACN, 2008; DH, 2011). Various microeconomic studies investigate 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 find a positive relationship between socio-economic status and healthy diets. Attempts have been made to assess the determinants of food choice by certain groups using variables such as income (Park et al., 1996; Han and Wahl, 1998), or age and education (Cortez and Senauer, 1996). These studies exogenously impose the grouping of the population and they do not consider all food categories but focus on fruit and vegetable intake instead (Han and Wahl, 1998; Bertail and Caillavet, 2004; 2008). Preferences that underlie food choices remain unaccounted for in these studies even though they are important drivers of food choice behaviour. Moreover, individuals with similar socio-demographics may have very different motivations regarding their diet and health, and households may choose similar diets for very different reasons. We argue that it may be better to identify households that are similar in their food preferences and therefore in how they would respond to policy interventions. Different to previous studies, we segment households according to their food preferences. We assume that the population can be split into several groups according to food preferences. Food preferences are homogeneous within groups but differ across groups as reflected by differences in the utility functions across groups. For each group a different AIDS can be estimated and the demand coefficients can be mapped to the group specific indirect utility function using Roy's identity. We use a finite mixture of AIDS to estimate each household's classification to a given group and the coefficients of each group's AIDS. Our model extends a study by Bertail and Caillavet (2004; 2008) in that it considers all food items not only fruit and vegetables, and in that it uses a Bayesian approach in the estimation which allows us to account for parameter uncertainty and censoring arising from infrequency of purchase.

2. Method

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

We follow Tiffin and Arnoult (2010) in estimating an AIDS using Bayesian inference and accounting for infrequency of purchase. The set of share equations is augmented with a set of probit equations modelling households with no observed purchase. 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 purchase 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 $\mathbf{z}^* = \left(\mathbf{s}^{*\prime}, \mathbf{y}^{*\prime}\right)$. 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 constant vector, and are the variables in the probit and demand equations respectively. The share equations are reparameterised to satisfy symmetry, homogeneity and adding-up restrictions. 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; and $\boldsymbol{D}$ is the data (both latent and observed). The following conditional posterior distributions, which form the basis of the Gibbs sampler, are obtained from equation 4:

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

where $p_k(\cdot)$is the probability of a specific household of belonging to component j, and which is used to adjust the proportion θ_j, and 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:

(1) Within each component $j = 1, \ldots, k$, 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
(4) Compute the probabilities $p_j(\cdot)$ as $p_j(\cdot) = \frac{\text{p.d.f.}(\hat{e}_j)\cdot\theta_j}{\text{p.d.f.}(\hat{\mathbf{e}})\cdot\boldsymbol{\theta}}$, where $\hat{e}_j = \mathbf{z}^* - \mathbf{X}\hat{\boldsymbol{\beta}}_j$ are the fitted residuals for component j, and p.d.f. stands for the multivariate normal probability density function.
(5) 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 Living Costs and Food Survey (LCF) for 2011. Participating households record food purchases for consumption at home over a 2-week period. The sample of 5,692 households is stratified by Government Office Region (GOR), National Statistics Socio-Economic Classification, 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 k of components is predetermined in our model because for $k < 5$ some components were highly unstable whilst for $k > 5$ some components were so small that estimation was not possible. We obtained robust components with recurrence of households within each component, confirming that allocation was not random. 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 consists of 3% of the population.

Elasticities are computed based on the estimated coefficients of the component-specific AIDS models, for each household at every iteration of the sampler, and mean values are then taken within each component. Figure 1 presents for each component own-price and expenditure elasticities in relation with expenditure patterns.[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.

[1]Own-price elasticities and expenditure elasticities are presented for all components in appendix tables 3 and 4 (p.19). We also report quantities purchased and associated expenditures in appendix tables 5 and 6 (p.20).

Table 2. Component structure

	Cmp1	**Cmp2**	**Cmp3**	**Cmp4**	**Cmp5**	**All**
$\boldsymbol{n}$	1919	2355	645	554	182	5655
$\boldsymbol{n\%}$	34%	42%	11%	10%	3%	100%

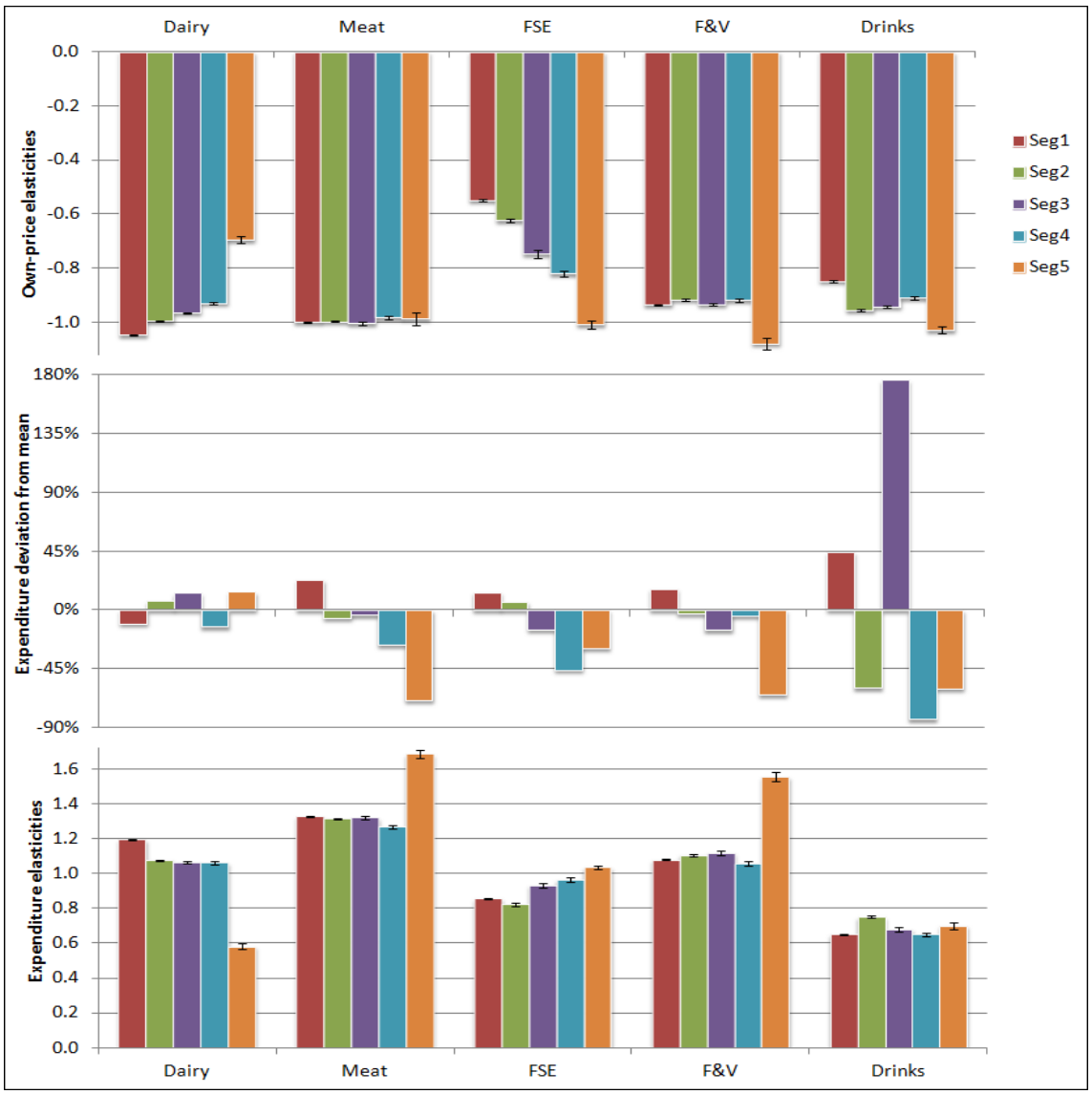

FIGURE 1. Own-price Marshallian elasticities, expenditure figures and expenditure elasticities (from top to bottom; 95% credible intervals)

The second panel of Figure 1 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 of Figure 1 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 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.

Nutrient intake was derived using nutrient conversion tables provided by the LCF complementary data. For nutrients expressed as relative values of total daily energy intake (Figure 2 and top half of appendix table 7, p.21), we observe that for fats (all categories with the exception of polyunsaturated fats, PUFAs) and for non-milk extrinsic sugars (NME sugars), all components are over the recommended limits. Looking at total fats, the maximum recommended is 35% of the daily energy intake and the sample average is 37%, ranging from 35% for component 3 to just over 38% for component 4. There is little variation in saturated fats (SFAs) intake with an average of 14% for a maximum recommended of 11%. There is more variation regarding monounsaturated fats (MUFAs) ranging from 13% for component 5 up to close to 15% for component 4 (while the maximum is set at 12%). NME sugars show also some variation, from under 12% for component 4 to almost 15% for component 2 (maximum set at 11%). For PUFAs, all components are just above the minimum recommended of 6%, and well under the upper limit of 10% with an average of 6.5%. Regarding protein, all components are within the recommended range (10 to 15%) with an average of just under 14%.

Regarding nutrients expressed as absolute daily values (Figure 3 and bottom half of Table 7, Appendix p.21), there is more variation across components for all categories considered. We observe that for all categories but alcohol (and to a lesser extent fruit & vegetables), components 1 to 3 are much higher up on the scale than components 4 and 5: looking at daily energy intake, components 1-3 are indeed consuming from 3,200 kCal to nearly 3,600 kCal per day whereas components 4 and 5 are consuming around 2,200 kCal and 2,500 kCal respectively (sample average at 3,240 kCal). From this high food consumption (well above the recommended daily 2,000 kCal and 2,500 kCal for women and men respectively) stems an excess intake of sodium (almost twice the recommended daily amount for component 1) and cholesterol for these components. As an upside maybe, this allows those 3 components to achieve the daily recommended minimum intake of dietary fibre, if only just for component 3. In contrast, as components 4 and 5 have very modest food consumptions in comparison to other components (though within the energy guidelines for men and women), they do not exceed their cholesterol intake, and while both components exceed the recommended daily amount for sodium, component 5's intake

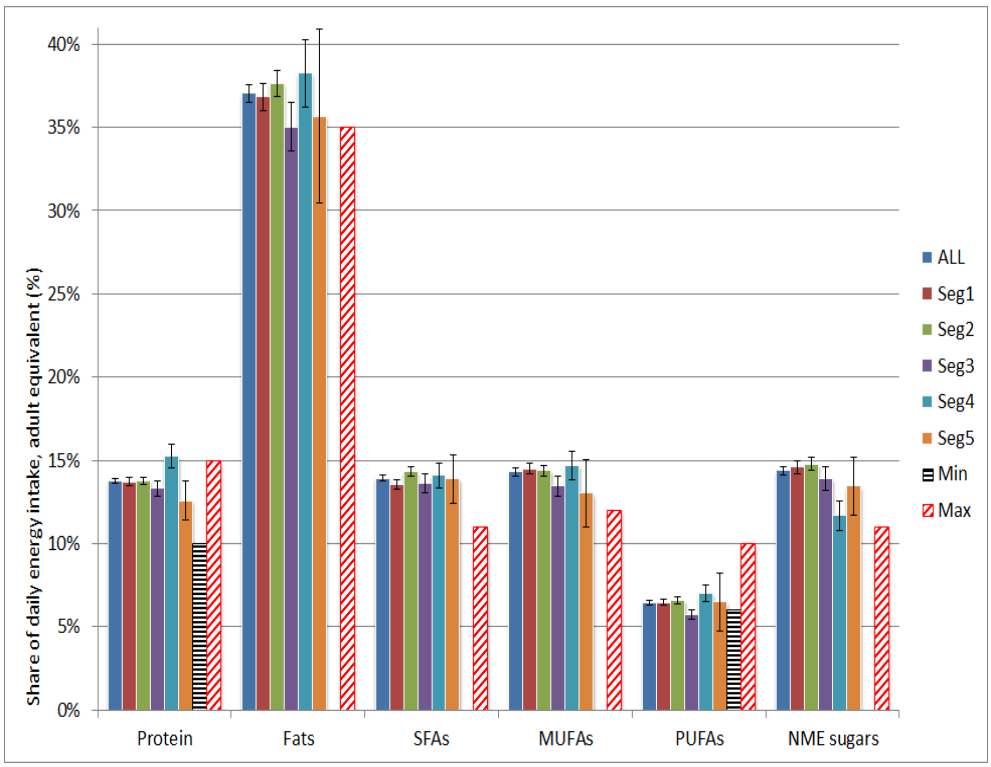

FIGURE 2. Nutrient intake for protein, fats, NME sugars expressed as a percentage of daily energy intake per adult equivalent (standard errors)

is not statistically significant. As a consequence of their low food consumption though, both components struggle to meet their dietary fibre recommendations. Regarding fruit & vegetables intake, not a single component manages to achieve the daily 400g target ("5-a-day"), though component 1 comes relatively close at 382g, followed by components 2 and 4 at about 350g; component 3 is consuming 290g, almost three-quarters of the recommended intake, while component 5 is trailing very far behind at 122g (1.5 portions out of the recommended 5-a-day). Alcohol consumption across components finally does not quite follow the pattern observed for the other nutrients: components 2, 4 and 5 have very low consumption levels, under the 17.5 units per week considered as a safe amount for women,[2] while component 3 consumes a staggering 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 two 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 HEI is scored out of 100 and reflects a household's diet compliance to healthy eating guidelines. For any

[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).

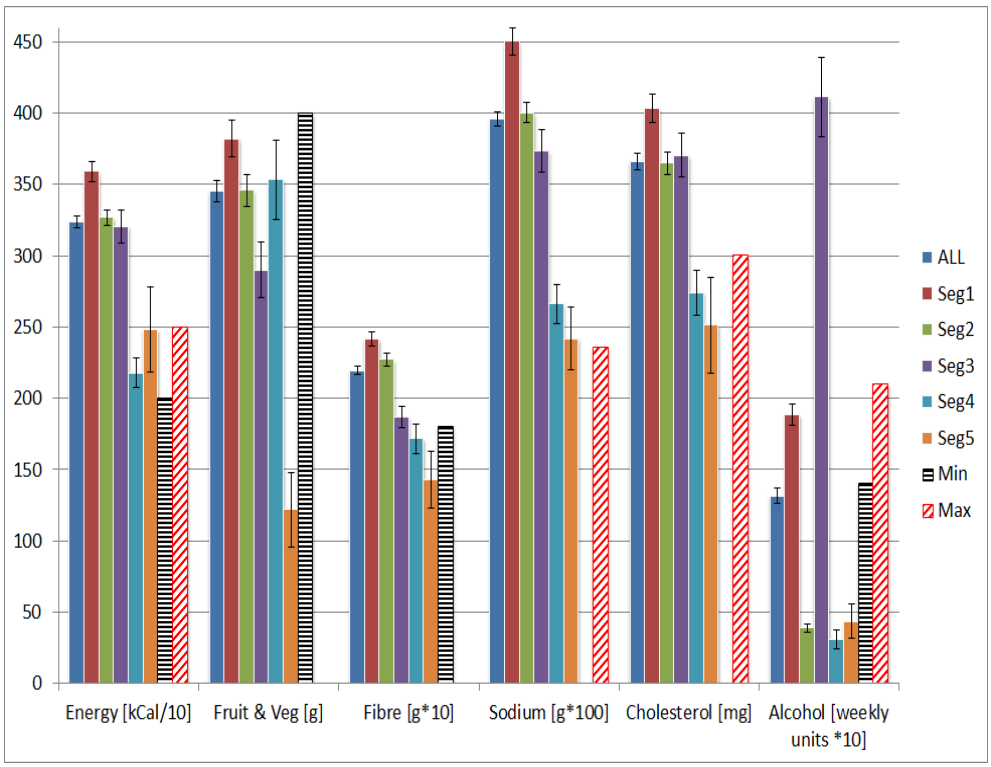

FIGURE 3. Nutrient intake for energy, fruit & vegetable, fibre, sodium, cholesterol, alcohol, per adult equivalent (standard errors)

given household, for each nutrient a set number of points are given for full compliance with recommended intake, and points are removed if this household's consumption departs from the guidelines, potentially reaching 0. Individual nutrient scores are summed up and normalised to 100 (maximum score achieved by fully compliant households). The USDA score takes a different approach in measuring the healthfulness of a household's grocery basket by comparing its expenditure shares to an accepted set of values for broad food categories (dairy, meat, vegetables, etc.); deviation from those accepted values will lower the overall score for each household. The USDA score is computed twice, taking into account food categories for which zero consumption has been recorded (Score 1), or discarding those values (Score 2): the rationale behind Score 2 is not to penalise those households that have not recorded any purchase of a particular food category over the survey period (thus lowering their score) while actually consuming from stock at the time. Both methodologies have been adapted to the UK: for the HEI, healthy eating guidelines from SACN (2008) and Department of Health (2011) 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).

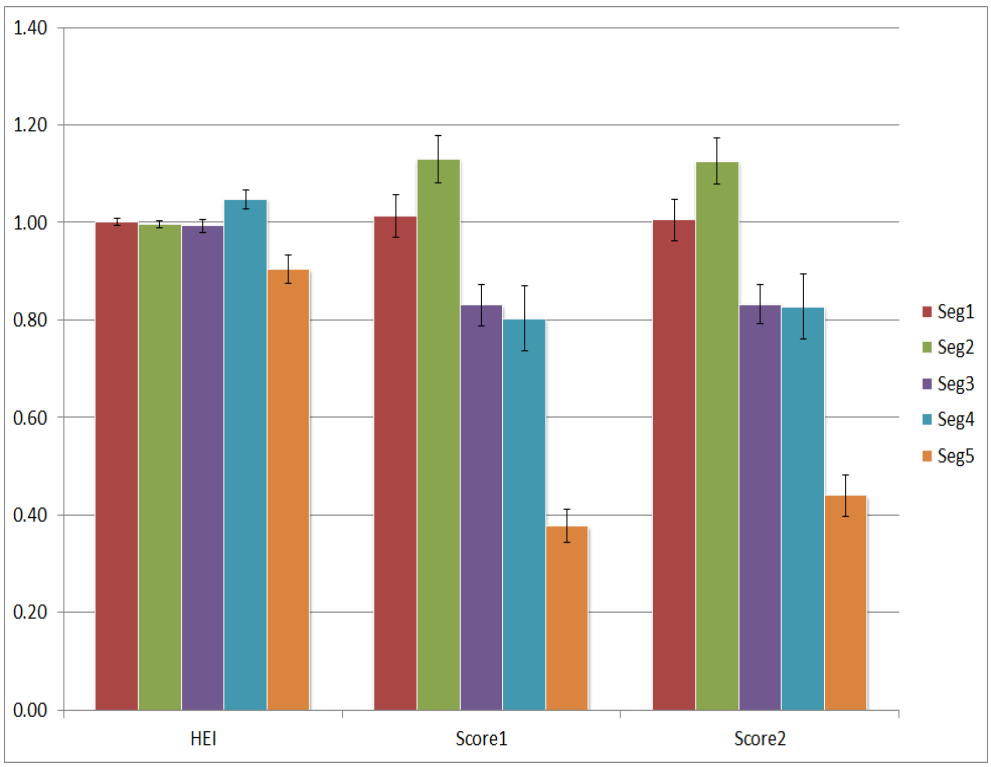

FIGURE 4. USDA and NHS measures of diet healthiness, normalised (sample = 1.00)

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.22), where results have been normalised using the full sample as reference to facilitate comparison. Finally, as the Eatwell plate does not take alcohol consumption into account, this has been excluded from the HEI.

We observe that the USDA HEI does not discriminate much between components, indicating only a less healthy diet for component 5, and a slightly better one for component 4. The USDA scores offer more variation however. Accounting for censoring does not affect the results much, and this can be explained by the the relatively low level of censoring observed in the data (see Table 1); component 2 fares better than average while components 3 and 4 fare worse and component 5 much worse. As far as component 5 is concerned, this can be explained by its low intake in fruit & vegetables, its relatively high intake in dairy, and its low intake consumption of fish relative to other components. The apparent contradicting results obtained from both indices can be explained by their difference in approach: while the HEI focuses on dietary guidelines compliance, the USDA score focuses on a balanced diet across major food groups; in that respect, a component can be well noted under one index and poorly under the other. Both measures however are

in agreement to find component 5's diet of a poorer quality, both in terms of nutrient requirements and general balance.

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's consumption is relatively close to, but above sample average though it consumes less dairy, mostly liquid milk, and visibly more drinks (both soft and alcoholic drinks). It is the only component to consume more than average of the fruit & vegetables and meat & fish categories, and in particular it is the only component to exhibit a fish consumption higher than average. It is the component with the highest median income and the highest proportion of households belonging to higher socio-economic classes (SEC) and higher socio-economic groups (SEG). As a proportion of the total number of people in the household, we also find the highest number of workers[3] as a proportion of the number of adults per household. Regarding marital status, this component exhibits the highest proportion of married or cohabiting couples, and we also find a high proportion of children in their teenage years. Concerning elasticities, component 1 is the most price and income elastic for dairy, and is the most price inelastic and second most income inelastic for fats, starches, etc. It also the most price and income inelastic for drinks.

Component 2 is the largest of all 5 components, and as such exhibits consumptions patterns and socio-demographic features which are close to the sample average. We can notice however that the consumption level for drinks is well under the sample mean (-35% for all drinks, -81% for alcohol); component 2 is also the most income elastic for drinks and the least income elastic for fats, starches, etc. Regarding household composition, we observe a higher proportion of retired people and of lower SEC and SEG.

Component 3 has a consumption generally close to the sample average except for drinks for which we observe a massive over consumption (+85%), driven by alcohol intake (+254%). Consumption of fruit & vegetables is under average (-14%) but even more so for fruit, fresh (-23%) or otherwise (-19%). The component is income elastic for meat

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

& fish, fruit & vegetables, and dairy & eggs, while income inelastic for drinks. This is the second highest median income and second highest proportion of SEC/SEG households; we also observe the second lowest number of children per household, more and older workers per household than average; it is the least ethnically mixed component (98% classified as white British against a 93% average), with a higher proportion in Yorkshire and the East.

Component 4 under-consumes from all categories, as would be expected from its low energy intake. Consumption of fats, starches, etc. is much lower than average (-42%), with all subcategories equally concerned (fats, breads and baked goods, cereals, jams, etc.). Consumption of meat & fish is lower than average by 28%, but fish is not as affected, with consumption only 15% lower. Fruit & vegetables is the least affected of the main food categories, and consumption of fresh fruit is even above the sample average (+9%), though all other subcategories are under-consumed. Finally, drinks consumption is much lower (-70%). This component is income elastic for meat & fish, dairy & eggs, and fruit & vegetables, while income inelastic for drinks. In terms of socio-demographics, the component comprises of the smallest households, with the lowest number of children per household. It also 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 indicating an urban population.

Component 5 like component 4 above has a low consumption which falls under the sample average for all major food categories, with the exception of the dairy & eggs group, for which we observe an above average of 37%. Consumption of meat & fish is the lowest observed (-70%), as is the consumption of fruit & vegetables (-58%, fresh subcategories in particular). Consumption of dairy & eggs is driven by liquid milk (+60%), as consumption of other subcategories is below average. Component 5 is the most price and income inelastic for dairy & eggs, and the most price and income elastic for fats, starches, etc., and fruit & vegetables as well as being the most price elastic for drinks and the most income elastic for meat & fish. In terms of socio-demographics, it comprises households of the second smallest average size, with the lowest number of workers as a proportion of the number of adults. We find the highest average number of children per household, with a large proportion of under 5 year olds, and the highest proportion of single/divorced

or widowed households (over 60%), making this a component consisting largely of single parents. The component also has the lowest median income and the highest proportion of lower SEC/SEG. There is a higher presence of households located in the North West and the Midlands.

Throughout components, a few similarities regarding elasticities and consumption patterns are noticeable. Some components are price elastic for food categories which they consumes in lesser quantities (e.g., component 1 and dairy & eggs, component 5 and fruit & vegetables, fats, starches, etc., drinks, to a lesser extent meat & fish), while these same components are price inelastic for foods which they over consume (component 1 and drinks, component 5 and dairy & eggs). Component 5 is highly income elastic for meat & fish, and fruit & vegetables, 2 food categories for which this component is well under the sample mean, thus indicating an income constraint; regarding dairy & eggs however, component 5 is income inelastic, indicating that the consumption level is not income-dependent. As noted earlier, all components are relatively income inelastic for drinks while having different consumption levels, thus implying different preferences for these components (satiety or dislike) rather than an income constraint.

For comparison, we estimate a mixture using expenditures on the five major food categories as the dependent variable. A k-means cluster of 5 components only vaguely resembles our segmentation result. Our two components of low consuming households are mostly allocated to a component of low consuming households, while a component of heavy drinkers has been isolated again, though it is much smaller than the one we identified. Also, where elasticities discriminated on dairy and fats, starches, etc., no such pattern is apparent with k-means clustering, thus showing that segmenting on consumption alone (quantities or expenditures) cannot account for preferences (sensitivity to price through elasticities) in the same way that our approach does, and yields a substantially different segmentation. The implications in terms of policy interventions are certainly essential since different segmentation would lead to different sub-populations being targeted with different outcomes.

Considering consumption patterns as observed across components, the question arises whether they are the product of a household's preferences, or rather that of the household's economic circumstances? We observe that demand by households in component 5 for products which they do not purchase often such as meat & fish or fruit & vegetables is more expenditure elastic possibly indicating a willingness-to-buy limited by an income constraint. The same component has income inelastic demand for dairy & eggs which they buy in large quantities (liquid milk in particular), a buying pattern possibly dictated by the presence of children. Recalling that income elasticities for drinks are very similar across components in spite of rather different consumption levels, we can infer that components are not constrained, but rather that they are consuming according to their preferences; the high intake of component 3 in particular cannot be explained by its socio-demographic features alone.

5. Conclusions

In this paper we use household food consumption data to estimate demand models the coefficients of which serve as discriminants in a finite mixture setting. Combining AIDS and mixture model allows to segment households into five components with each group having homogeneous underlying food preferences. The components have different own-price as well as income elasticities, with segmentation occurring chiefly based on dairy & eggs, and fats, starches, etc. food groups. Components are differ in terms of their consumption patterns and to a lesser extent in terms of their nutrient intakes. Of the five components identified, we have isolated two components containing low consuming households, one of which has very low fruit & vegetables intake and high dairy & eggs intake (mostly due to liquid milk); a third component is characterised by a high alcohol intake. There are some tendencies in terms of socio-demographic characteristics of components, mainly in terms of income, SEC/SEG, with components of a higher status being linked to a more abundant and healthier diets according to our healthiness indices adapted from the literature. Overall components remain however heterogeneous in their socio-demographic make-up, and results indicate that households with similar food or nutrient intake and similar healthiness scores are allocated to different components: this confirms our starting assumption that households can have similar diets as a result of different sets of preferences and therefore elasticities.

From a policy perspective, ease of identification of the components is essential in designing interventions specifically aimed at populations in need and socio-demographics provide a convenient way of selecting target populations. While our approach does not rely on socio-demographics, segmenting on the basis of preferences should not be a barrier to designing policy interventions. We assume that differences in consumption are the result of differences in utility functions and ultimately of differences in susceptibility to cognitive biases. People therefore may self-select into responding to a given intervention according to their susceptibility to given cognitive biases, in the same way they would respond to advertising signals or cues in general. These cognitive biases could be elicited using a dedicated questionnaire based on previously validated measures, such as resistance to framing, recognizing social norms, etc. (Bruine de Bruine et al., 2007; Kirby et al., 1999).

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 1	14.9	15.1	16.8	12.4	12.0	5.6
USDA Score 2	15.1	15.2	17.0	12.6	12.5	6.6