Session Number:Monday Evening Session
Session Title:Econometric Issues in the Multidimensional Measurement and Comparison of Economic Well-Being
Paper Number: 4
Session Organizer: Jean-Yves Duclos
Discussant: Michael Grimm

Paper Prepared for the 28th General Conference of
The International Association for Research in Income and Wealth
Cork, Ireland, August 22 – 28, 2004

HEALTHY DIET AND POVERTY IN FRANCE : A SEGMENTATION APPROACH

PATRICE BERTAIL and FRANCE CAILLAVET

For additional information please contact:

France Caillavet
CORELA / INRA – 65, Bd de Brandebourg – 94205 Ivry-sur-Seine
France.Caillavet@ivry.inra.fr
Fax: 33.1.49.59.69.90 – Tel: 33.1.49.59.69.83

HEALTHY DIET AND POVERTY IN FRANCE: A SEGMENTATION APPROACH.

(Running Title: **Healthy diet and poverty**)

by

Patrice BERTAIL * and **France CAILLAVET****

*CREST, LSA, Timbre J340, 3, ave Pierre Larousse, 92245 Malakoff.

**CORELA / INRA - 65, Bd de Brandebourg, F-94205 Ivry-sur-Seine

Tel: 33-1-49.59.69.83 – Fax: 33-1-49.59.69.90

Email : France.Caillavet@ivry.inra.fr

Abstract

This paper questions the relevance of income for measuring the link between poverty and healthy diet. We use a segmentation approach to cluster households, based on a finite mixture of QAIDS/AIDS models of fruit and vegetables consumption. The endogeneity of total expenditure within each class is taken into account and tested. An E-M estimation procedure which reduces to compute iteratively weighted instrumental estimators is developed and applied to a French panel data. We also propose a subsampling technique to determine the optimal number of classes. We obtain 6 homogeneous classes, reflecting specific income and price elasticities. We find that, for the deprived class, fruit and vegetables demand accurately characterizes economic and social inequality.

Key Words and Phrases: *JEL classification:* C51;C52;D12;I12

1 Introduction

Food deprivation is one basic dimension of poverty[1]. Our main interest in this paper is to define a concept of poverty revealed by food consumption patterns, in the perspective of inequalities in standards of living including health. Our hypothesis underlying this approach is that economic deprivation and social exclusion do imply specific food consumption patterns which are not entirely explained by monetary poverty and may have important health consequences. Consumption analysis shows that low-income or low-status populations have significantly different behaviors in several areas: quantities consumed, quality of food, variety of products, distribution of consumption between at-home and away-from-home segments,... Nutritional studies acknowledge the role played by food behavior and lifestyle in health inequalities. Several studies point out that higher social class groups tend to have healthier diets, one of the main differences lying in fruit and vegetable consumption (Roos et al. 2000, Johansson et al. 1999). Since this product category plays a major role in nutrition, it has developed as a special tool for socio-economic policy. Increases in fruit and vegetable consumption is one of the most prominent goals of Health Programs in several developed countries: Food Pyramid recommendations in the USA (USDA 2000), the Programme National Nutrition-Santé (Ministère de l'Emploi et de la Solidarité 2000) in France, the Five-a-day Programme in UK (Department of Health 2000). As such we may use the differentiation power of this food category.

We will focus on the food dimension of deprivation, with two aims: the first is to clarify the relationship between monetary poverty and healthy food consumption (or deprivation); the second is to establish income classes from a relevant healthy food deprivation scale, so as to escape the usual frame of quantile thresholds. Deprivation is widely distributed in the income scale, since social workers report that an income shock (for example a health problem) sends 3rd quintile households into poverty.

We assess the existence of different types of food demand in the population. We want to estimate a demand system for different homogeneous categories of population, the concept of homogeneity (in food consumption behavior) being itself defined by

[1] Sen (for example 1984) made popular the multidimensionality of poverty.

the demand system. In this perspective we will focus on a segmentation approach, that is a method for clustering individuals based on the homogeneity of their food consumption patterns, without constituting *ex ante* categories based on income. For this we perform a clusterwise regression analysis (see DeSarbo and Cron, 1988) of food demand based on a finite mixture of components, without making any prior assumptions as to the number of classes, and taking into account the specificities of demand systems. In our case we will be essentially interested in estimating a finite mixture of AIDS and QAIDS models, taking into account the endogeneity of total expenditure in the model.

The outline of the paper is as follows. In part 2 we begin with a short description of the QAIDS model which serves as a basis for the segmentation procedure. We describe the mixture and show how it is possible to combine E-M type algorithms and standard iterated weighted least square estimators to obtain convergent estimators of the parameters of the model. We also propose a test based on subsampling the log-likelihood ratio to establish the number of components to retain in the model. In part 3, we give the French data to which the model is applied. Our results are presented in part 4. It is shown that income is not the main structuring variable for differentiation and inequalities in food consumption, although the most deprived class corresponds to the lowest income households. Our more striking result is that food consumption has no correlation with income in the poorest class, and this constitutes a rather surprising result in demand analysis.

2 A finite Mixture QAIDS model for food demand

2.1 Theoretical framework

We describe the household h consumption behavior[2] during the week t by the Quadratic Almost Ideal Demand System model (QAIDS) which was developed by Banks, Blun-

[2] Food consumption is studied in the framework of the household production model (Becker, 1965).

dell and Lewbel (1997). The budget shares, w_{ih}, have the following form[3] :

$$w_{ih} = \mu_{ih} + \sum_{j=1}^{N} \gamma_{ij} \ln p_{jh} + \beta_i \left[\ln x_h - \ln a(p_h)\right] + \lambda_i \frac{\left[\ln x_h - \ln a(p_h)\right]^2}{b(p_h)} + u_{ih} \ , \qquad (1)$$

for $i = 1, .., N$ food categories and $h = 1, .., H$ households with

$$\ln(a(p_h)) = \alpha_0 + \sum_{i=1}^{N} \mu_{ih} \ln p_{ih} + \frac{1}{2} \sum_{i,j=1}^{N} \gamma_{ij} \ln p_{ih} \ln p_{j,h} \ \text{ and } b(p_h) = \prod_{i=1}^{N} p_{ih}^{\beta_i}$$

($a(p)$ may be interpreted as a price index), where p_{ih} is the price of a food category i for the household h, x_h is the total food expenditure of the household h and α_i, γ_i, β_i and λ_i are the parameters to estimate. We also define $p_h = (p_{1h}, ..., p_{Nh})$. Because we cannot identify coefficient α_0 in this model, we will impose $\alpha_0 = 0$. In order to take into account heterogeneity in behavior and identify the model, the parameters $\mu_{i,h}$ are modelled as a linear form $\mu_{ih} = \alpha_{i0} + Z'_{kh}\alpha_{ik}$, where Z_{kh} is a vector of household characteristics, and $k = 1, ...K$. The first column of Z_h is assumed to be a vector of 1. The standard AIDS (Almost Ideal Demand System) of Deaton and Muellbauer (1980) also corresponds to the equation (1) without the quadratic expenditure term but including the price index $a(p_h)$.

It may be proved that this system derives from some cost minimization, if it satisfies the restrictions imposed by the theory i.e. additivity and homogeneity of degree zero of the budget shares in prices and total expenditure and the symmetry of Slutsky's matrix. This implies the following constraints :

Adding up is satisfied when $\sum_{i=1}^{N} \alpha_{i0} = 1, \sum_{i=1}^{n} \alpha_{ik} = 0, \ \forall k \ , \sum_{i=1}^{N} \gamma_{ij} = 0,$ $\forall j, \sum_{i=1}^{N} \beta_i = 0$, and $\sum_{i=1}^{N} \lambda_i = 0$.

Homogeneity and symmetry require $\sum_{j=1}^{N} \gamma_{ij} = 0, \forall i$ and $\gamma_{ij} = \gamma_{ji}$.

In the following, let $\alpha = (\alpha_1, ...\alpha_N)$ be the vector of constants, $\beta = (\beta_1, ..., \beta_N)$ the vector of expenditure parameters, $\gamma = [\gamma_{i,j}]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}$ the matrix of cross price parameters and $\lambda = (\lambda_1, ..., \lambda_N)$ the impact vector of the quadratic term. Define also $\theta = (\alpha, \beta, \gamma, \lambda)$. We denote by Ξ the set of all parameters θ satisfying the constraints. Since the shares sum to 1, one equation may be dropped to ensure that the residuals

[3] In the following, we drop the index t for a clearer presentation.

have a full rank covariance matrix. Thus we may write the model in the following normalized functional form

$$w_{ih} = f_i(p_h, Z_h, x_h, \theta) + u_{ih},\ i = 1, ..., N-1\ ,\quad w_{N,h} = 1 - \sum_{i=1}^{N-1} w_{i,h}$$

and globally

$$w_h = (w_{1h}, w_{2h,...,}w_{N-1,h})' = f(p_h, Z_h, x_h, \theta) + u_h\ , \tag{2}$$

with the notations $u_h = (u_1, u_2, ..., u_{N-1})'$ and $f = (f_1, ..., f_{N-1})'$.

One of the main problem when estimating demand systems is the endogeneity of the expenditure x_h. To solve this problem, the income y_h is generally chosen as an instrument and a model is assumed in the form $\ln x_h = I_h\delta + \eta_h$, with $I_h = (1, \ln y_h, p_h, Z_h)$ and $E(\eta_h|I_h) = 0,\ V(\eta_h|I_h) = \sigma^2$.

To test and correct for the endogeneity, it is now common to use an augmented regression framework. This procedure has been adapted by Blundell and Robin (1997) to the case of a non linear model with applications to a QAIDS demand model. As shown in their paper, to perform and prove the validity of this method we need the following additional assumptions on the structure of the residual in the demand system mainly $u_h = \rho^0\eta_h + \varepsilon_h$, with $\rho_0 \in \mathbb{R}^{N-1}$ (which may be interpreted as a measure of the endogeneity of x_h, conditionally to the information I_h) and $E(\varepsilon_h|I_h, \eta_h) = 0\ ,\ V(\varepsilon_h|I_h, \eta_h) = \Sigma$.

We assume in addition that (ε_h, η_h) are independent for $h = 1, ...H$. Under these hypotheses, we may estimate the whole demand system with a two steps procedure. The first step involves regressing the expenditure on the income and the exogenous variables to get an estimator $\widehat{\delta}$ and plug the predictive value of the residuals $\widehat{\eta}_h = x_h - I_h\widehat{\delta}$ into the original demand system to correct for endogeneity. However as revealed by a previous study (Caillavet 2002) estimating a QAIDS on the whole population does not allow to adequately describe the economic and social inequalities underlying the demand system. To address this problem of heterogeneity, we propose the use of finite mixture regression or clusterwise regression models in the following section.

2.2 Segmentation: from discrimination to model based cluster analysis

There is a huge literature about classification and mixture in the statistical field : relevant ideas may be found in standard non-probabilistic data analysis (for instance K-means) but also in mixture modelling (see McLachlan and Peel 2000) and statistical learning (Freedmann, Hastie and Tibshirani 2001). We would simply like to recall that there are essentially two aspects of classification, which have clearly different goals :

- discrimination also known as supervised learning in the statistical learning field. In that case the individuals are supposed to belong to some pre-specified class. In our framework, this approach is essentially useful when one wants to predict the behavior or the consumption of an individual belonging to a specific class. Methods of segmentation have seldom been used in the economic food demand literature: most of the approaches considered may be categorized as discriminating. Indeed, papers considering various segments of demand generally impose the class ex ante. The most widely used criterion for this typology is monetary income categories (Park, Holcomb, Raper and Capps Jr 1996), with particular focus on poor and non poor populations across poverty lines. Nonetheless a growing body of applied literature mentions the lack of significance of income on food consumption.

- cluster analysis or unsupervised learning in statistical learning. In this approach the classes are not known a priori but are rather determined by the data itself. The main idea is to consider that observations arise from unobserved homogeneous classes, in some unknown proportions according to some finite mixture of distributions describing the behavior of each class. The goal is to recover the underlying classes and estimate the parameter of each component of the model. Finite mixture models are described at length for instance in Titterington, Smith, and Makov (1985) and McLachlan and Peel (2000). In food demand analysis, an attempt in that direction is developed in Jensen and Manrique (1998), in which they use a Goldfeldt-Quandt test to determine two (or a fixed number of) homogeneous segments of the population. However we have some doubt about the statistical significance of such a test, which relies on the existence of a switching model along some ordered variable, for instance the income, and the existence of only two models for two distinct categories (the poor and non-poor)[4].

In general clustering algorithm operates on feature vectors of fixed dimension, but some earlier studies have also been focusing on clustering regression models (see DeSarbo and Cron 1988, Wedel and DeSarbo 1995)). Following these ideas, we assume that there exists K classes (K unknown but bounded) characterized by the relation (2) with some parameters $\theta_{(s)},\ \Sigma_{(s)},\ s = 1, ..., K$ with $\theta_{(s)} \neq \theta_{(s^*)}$ for $s \neq s^*$. For $s = 1, .., K$, we put $\theta_{(s)} = (\alpha_{(s)}, \beta_{(s)}, \gamma_{(s)}, \lambda_{(s)})$ where the parameters are indexed by (s). Moreover there is no reason for the total food expenditure elasticities to be the same throughout all the classes, so we assume that the instrumental rgeressions hold with some parameter $(\delta_{(s)}, \sigma^2_{(s)}, \rho^0_{(s)})$, which may be different on each class. Actually, there is no reason to assume that the model is the same for each class, for instance it may be that in our case the quadratic part of the QAIDS is not significant for some particular sub-classes (Caillavet 2002). For simplicity, we assume that for each class the residuals $(\varepsilon_{h,(s)}, \eta_{h(s)})$ are i.i.d gaussian, h=1,...,H. For mixture models, the gaussian assumption is not restrictive in this way (see McLachlan 2000). Finally, because the number of parameters to estimate is large in relation to the number of individuals that we have in our data, we will assume that this matrix is diagonal $\Sigma_k = diag(\zeta_k)_{k=1,..H}$. This (strong) hypothesis will facilitate the estimation procedure but may be relaxed in situations where a great deal of data is available.

Because of the exogeneity of the demand, this will however ensure that we can efficiently correct the endogeneity problem by a simple iterated two step procedure.Under these conditions and the conditions outlined above, the likelihood of each household may be represented by the "leek" function

$$\Upsilon(w_h, x_h|I_h, \Theta) = \sum_{s=1}^{K} \pi_s f_s(w_h, x_h|I_h, \Theta_{(s)}) \ ,$$

where $f_s(w_h, x_h|I_h, \Theta_{(s)})$ is the likelihood of (w_h, x_h) conditionally to I_h for each in-

[4] Another interesting approach which may be seen as intermediate between the two approaches mentioned above has been explored by Jarque (1987): it is actually a special example of what is known in the statistical literature as a mixed model (see for instance Searle, Cassela, McCulloch, 1992). We believe that some statistical work is still needed before validating this approach.

dividual in class s.In the following we use the simplified notation $\phi(x) = \exp(-\frac{1}{2}x'x)$ for $x \in R^p$, for all p. If $f_s^w(.)$ denotes the likelihood of w_h conditionally to both x_h and I_h and and $f_s^x(.)$ respectively the likelihood of w_h conditionally to I_h, by conditioning arguments, we have under the preceding hypotheses,

$$f_s(w_h, x_h|I_h, \Theta_{(s)}) = f_s^w(w_h|x_h, I_h, \Theta_{(s)}) f_s^x(x_h|I_h, \Theta_{(s)}) = \frac{1}{(2\pi)^{N/2} \det(\Sigma_{(s)})^{1/2} \sigma_{(s)}}$$

$$\phi(\Sigma_{(s)}^{-1/2} \left\{w_h - f(p_h, Z_h, x_h, \theta_{(s)}) - \rho_{(s)}^0 (x_h - I_h \delta_{(s)})\right\} \phi(\left\{x_h - I_h \delta_{(s)}\right\} / \sigma_{(s)}) \ ,$$

where $\Theta_{(s)} = (\theta_{(s)}, \Sigma_{(s)}, \delta_{(s)}, \sigma_{(s)}^2, \rho_{(s)}^0)$, $\theta_{(s)} \in \Xi$ and $\Theta = (\Theta_{(1)}, ..., \Theta_{(K)})$ and $\pi = (\pi_1, ..., \pi_K)$. The probabilities $\pi_s's$ may be interpreted as membership probabilities and thus characterize the size of the class.

If we have some convergent estimators of the parameters π and Θ say $\widehat{\pi} =(\widehat{\pi}_1, ...\widehat{\pi}_K)$ and $\widehat{\Theta}$, the analysis of the classes may be performed by using the classic posterior Bayes formula which gives the estimated posterior probability of belonging to the class. This essentially means that the membership to the class is determined by the range of the price and income elasticities and the coefficients of the socio-demographic variables on an homogeneous segment. However it may be difficult to identify the main determinants of this "homogeneity" and to explain the composition of the classes. A possible solution with regard to this problem is to assume that the probabilities $\pi_i's$ are themselves functions of the sociodemographic characteristics. In that case, the model may be seen as a combination of cluster and discrimination analysis. Because we do not want in this paper to make any prior assumptions on the variables explaining the composition of the class, we will not pursue this approach. We will thus focus on a pure cluster modelling analysis and explain a posteriori the compositions of the classes by a multivariate probit (or logit) model with some socioeconomic explanatory determinants.

Assume that the number of classes is fixed equal to K. By the independence assumption on the households, the likelihood of the whole data is thus

$$\Upsilon_H(\Theta) = \prod_{h=1}^{H} \Upsilon(w_h, x_h|p_h, Z_h, y_h, \Theta) \ .$$

It is known that the maximum likelihood is not correctly defined in mixture models. The most popular algorithm in that kind of situation is the E-M algorithm developed by Dempster, Laird and Rubin (1977). Since the algorithm may also converge to a local maximum, the use of different starting values and several runs of the algorithm are generally necessary to validate the final results. The use and details on E-M algorithm in mixture regression models may be found in De Sarbo and Cron (1988).

An iterated E-M algorithm

The main idea of E-M. algorithm in mixture models is to consider that the underlying classes $(C_1, ..., C_H)$ play the role of hidden data. The E-step (Expectation step) calculates the expected value of the full log-likelihood with respect to the distribution of membership variables (conditionally to all the others). Starting from some previously established classes $C_1, ..., C_K$ and values of the parameters $\widehat{\Theta}$ and $\widehat{\pi}$, the E-step consists in calculating the posterior probability of membership of individual h to class s^* that is

$$\widehat{\widehat{\pi}}_{(s^*),h} = \frac{\widehat{\pi}_{(s^*)} f_{s^*}(w_h, x_h | p_h, Z_h, y_h, \widehat{\Theta}_{(s)})}{\sum_{s=1}^{K} \widehat{\pi}_{(s)} f_s(w_h, x_h | p_h, Z_h, y_h, \widehat{\Theta}_{(s)})} \ if \ h \in C_{s^*} \ . \tag{3}$$

The expected full log-likelihood (evaluated at $\widehat{\Theta}$) is given by

$$\sum_{h=1}^{H} \sum_{s=1}^{K} \widehat{\widehat{\pi}}_{(s),h} \ln(\widehat{\pi}_{(s)} f_s(w_h, x_h | p_h, Z_h, y_h, \Theta_{(s)})) \ .$$

The M-step (maximization step) now consists in finding the MLE for each $\Theta_{(s)}$. Since we can factorize the likelihood in each $\Theta_{(s)}$, and getting rid of the term independent of Θ, this amounts to finding the m.l.e. for each log-likelihood, $s = 1, ..., K$, given by (C is a generic constant independent of the data and the parameters)

$$l_s(\Theta_{(s)}) = \sum_{h=1}^{H} \widehat{\widehat{\pi}}_{(s),h} \ln(f_s(w_h, x_h | p_h, Z_h, y_h, \Theta_{(s)}))$$

Following Blundell and Robin (1997), it is easy to see that the factorization into two distinct parts (the augmented demand system and the instrumental regression model) implies that we can first estimate the instrumental equation separately and then obtain the m.l.e. for the parameter in the demand system by plugging the estimated value of $\delta_{(s)}$ say $\widehat{\widehat{\delta}}_{(s)}$ into it. But under the hypothesis of normality, this in turn amounts in estimating the parameter of the demand system as if the variance of each individual was weighted by the inverse of its posterior membership probability $\widehat{\widehat{\pi}}_{(s),h}$. In the case of the AIDS model, the m.l.e. estimators are exactly the weighted least-square estimators on the instrumental equation and all the shares equations, with weights given by (3) (see detail in the full technical report Bertail and Caillavet, 2003). Thus at each step of the E-M algorithm, we can use the usual two step estimation procedures of the demand system by just multiplying the terms in the models by $\sqrt{\widehat{\widehat{\pi}}_{(s),h}}$. Notice that in the case of the QAIDS, the whole quadratic part should be multiplied by this quantity but not the variables inside. We just have to take care of the estimators of the variance, which in our case are given respectively by $\widehat{\widehat{\sigma}}_s^2 = \left(\sum_{h=1}^H \widehat{\widehat{\pi}}_{(s),h}\right)^{-1} \sum_{h=1}^H \widehat{\widehat{\pi}}_{(s),h}(x_h - I_h\widehat{\widehat{\delta}}_{(s)})^2$ and $\widehat{\Sigma}_{(s)} = \left(\sum_{h=1}^H \widehat{\widehat{\pi}}_{(s),h}\right)^{-1} \sum_{h=1}^H \widehat{\widehat{\pi}}_{(s),h}\widehat{\varepsilon}_h\widehat{\varepsilon}_h{}^t$,where

$$\widehat{\varepsilon}_h = w_h - f(p_h, Z_h, x_h, \widehat{\widehat{\theta}}_{(s)}) - \widehat{\rho}^0_{(s)}(x_h - I_h\widehat{\widehat{\delta}}_{(s)})$$

are the estimated residuals. The algorithm now consists in iterating the E-M procedure by assigning, at each step. In the final step, when convergence occurs the variance of the estimators of the parameters and their transformations (for instance the compensated or uncompensated elasticities) are obtained by evaluating the Hessian and using Slutsky's theorem.

Choice of the number of classes

Despite of a lot of recent work on the subject, the choice of the number of relevant classes (when it is not fixed by some *a priori* arguments) still remains an important and difficult issue. The original proposal of DeSarbo and Cron (1988) is to base the choice on the comparison of an Akaike criterion for an increasing number of classes. Another natural and almost equivalent proposition is to look at the value of the likelihood and use a likelihood ratio test. However it is known that since this amounts to consider that the π_s may take value on the boundary, the limiting distribution of the likelihood ratio test is non-standard so that it may be difficult to implement especially in this two step procedures. The heuristic Akaike criterion $AIC(K)$ proposed in DeSarbo and Cron (1988) is transposed to our case. We here define it as $AIC(K) = 2\ ML(K) - p(K)$, where $p(K)$ is a penalization of the form $p(K) = 2\{(j_1 + j_2)K + K - 1\}/H$,where j_1 and j_2 are respectively the number of parameters for the demand system and for the instrumental regression). $(j_1 + j_2)K + K - 1$ is the effective number of parameters estimated in the K clusterwise regression model. The AIC criterion tries to balance between the adequacy of the model and its parsimony. However, Wedel and DeSarbo (1995) argued that the AIC criterion tends to choose a too small number of components and that the so called Bayesian Information Criterion may lead to a better choice. It is here defined as $BIC(K) = 2ML - \log(H) * \{(j_1 + j_2)K + K - 1\}/H$. The purpose is to find the optimal value of K which maximizes these information criteria.

Using bootstrap methods for determining the number of components in a mixture model has also been proposed by several authors (see for instance McLachlan and Peel, 2000). However it is known that this is typically the kind of situation in which the bootstrap may fail because of the lack of uniformity ofd, subsampling techniques are known to work in the most general framework including statistics with an unknown rate of convergence (see Bertail, Politis and Romano 1999). The idea is simply to draw (B times) without replacement some subsets of fixed size b_H and to estimate the log-likelihood ratio of K against $K + 1$ for each of these subsets. The quantile of order 1-α of the subsampling distribution may serve as the basis for constructing the rejection region. Typically the null hypothesis is rejected if the observed likelihood ratio on the whole sample is greater than the value of this estimated quantile. A detailed description of the procedure as well as a proof of its asymptotic validity in our framework are given in Bertail and Caillavet (2003).

3 Data and methods

3.1 Data

We use a French panel data from 1997 conducted by SECODIP containing 5348 households. Households are asked to register their food purchases for at-home consumption on the basis of daily purchases through the use of a scanner. The total quantity purchased of each good and the total amount of money spent on each good are recorded. For this study, we work at the aggregate level of large food categories on an annual expenditure basis. Consequently, we have no zero values at that level in our consumption data, and we can estimate directly a demand system without controlling for the participation in the market, which is one of the main issues in the food demand literature. Households are assigned to 2 groups, and are requested the model on its boundary. Indeed to register their purchases of a restricted set of fresh food products. Hence the list of products covered by each group differ, although there is some overlap. If the two lists put together represent practically all possible food products with a wide variety, unfortunately it means that we have no complete information on food purchases for a given household. Though a clear limitation of our data, SECODIP survey remains in France the only current data source allowing the computation of unit values for food, by registering quantities and expenditures for detailed products of food-at-home. [5]

Healthy food is not a strictly defined category[6] but some food categories invariably stand out when we are considering positive influences on health: fruit and vegetables are widely considered to be an indicator of dietary habits conducive of good health. This is explicit in nutritional studies, where it is used as a proxy to characterize good dietary habits (Fuhrer et al. 2002; Morland and Diez Roux 2002), as well as in socio-economic studies on the role played by food in inequality (Feng and Chern 2000). In view of its prevalence in positive nutrition and taking into account the social differentiation of its consumption, suggested by several studies, we focus here on fruit and vegetable consumption. Moreover, we can avoid the drawbacks of dealing with such a heterogeneous category as "healthy food", where different dietary

[5] Since 1991, information on food consumption is only available in French Family Expenditure Surveys (FES). In the 1995 FES, data are limited to expenditure recording. In 2000-2001 FES, some quantities have been introduced. Unfortunately, these data are not yet available.

[6] A healthy diet is considered to mean abundance in grains, vegetables, fruit, and restrictions on total fat, saturated fat, and cholesterol, combined with moderate and regular physical activity, at least from the perspective of preventing food-related diseases such as coronary heart disease, cancer, stroke, and diabetes.

patterns are mixed. In SECODIP survey, fruit and vegetable purchases are registered by 2005 households with complete information. As noted above, we do not have information for all food categories, so that we will consider that fruit and vegetables are separable from other food and non food budget, for poor as well as for non poor populations.

Our survey does not register home production. According to 1991 data, it would represent 26% of fruit and vegetable expenditure. This underestimation of the real consumption does not prevent us from calculating price and income elasticities which are meaningful for socio-economic policies. We here adopt the same classification as in a previous study: products are separated in N=9 groups, following their type (fruit vs vegetables) and the degree of processing[7]. Thus we distinguish : fresh fruit; canned and frozen fruit; dry fruit; fruit juices; fresh potatoes; fresh vegetables; canned and dry vegetables; frozen vegetables; and vegetable convenience foods.

The **dependent variables** are constituted by the nine budget shares of fruit or vegetable categories in total fruit and vegetable expenditure.

Since prices are not available in our survey, unit values are used as proxies. First, for each household, we obtain unit values by dividing expenditure by the quantities for each food category defined above. Second, we calculate the means of these unit values by cluster. Here we build the clusters by concatenating two variables: the region, and the size of the residential area. Finally, the unit values obtained are imputed to each household according to its own cluster. The unit value of a food category is calculated by dividing expenditure by the quantity recorded for this food category.

Socio-demographic influences are controlled through the introduction of the age of the household head, the maximum educational level in the household (between

[7]Caillavet (2002) estimated an AIDS on (a priori) subclasses and found major differences in the range of elasticities. The 1st quartile (compared to the remaining population) showed lower income elasticities for fresh or frozen products (then healthier products would be less elastic) and higher income elasticities for more processed products. Regarding direct price elasticities, higher values were found for the 1st quartile with respect to fresh fruit, dry fruit, and convenience vegetable products.

spouses), and household size. The age of the reference person indicates the stage of the life cycle at which we find the household. Household composition is represented by the number of members. Education increases household and market efficiency, and consequently income. Education may also be considered as a proxy for efficiency and may condition the household production function accordingly. However, its effect is not clear a priori. For example, a higher level of education may increase the opportunity cost of time, but at the same time the utility of leisure activities varies according to education. Location and rural/urban effects are captured respectively through 8 and 4 dummies. A variable indicating the availability of a kitchen garden and/or an orchard will control for the absence of home production in the data.

Table 1 presents the summary statistics of the variables used for estimation[8].

3.2 Methodology

One of the purposes of our classification of households is the **determination of the number of segments** into which they are to be classified. The initial classes are set up with quantiles of household income (quartile for $K = 4$, quintile for $K = 5$ and so on). For this, we use a measure taking into account demographic units to obtain an equivalent adult household income, following the Oxford/OECD scale.

The procedure has been run for 4, 5, 6 and 7 classes. The corresponding values were obtained for the mean log-likelihood $=ML(K)$. We also give the corresponding AIC and BIC criteria.We also performed the subsampling log-likelihood ratio test described above and detailed in appendix 2. For this, we choose $b_H = 400$ and repeated the subsampling procedure 500 times. The 95% quantile of the subsampling distribution of the log-likelihood ratio test (which is equal with our notation to $R(K) = 2H(ML(K+1) - ML(K))$) for testing K components against K+1 is also given

[8] For the further econometric analysis, reference categories are indicated in italics.

below

K	$ML(K)$	$AIC(K)$	$BIC(K)$	$R(K)$	$SUB(K,K+1)$
4	16.42	31.760	32.091	$360,5$	$175,4$
5	16.51	31.669	32.084	1198.4	182.8
6	16.81	31.999	32.496	155.6	187.3
7	16.85	31.809	32.389	.	.

We can see that the highest increase in the log-likelihood was obtained between 5 and 6 components. This jump in the likelihood is strong evidence that there are at least 6 significant components. Minimization of both the AIC and BIC criteria yields the choice $K = 6$.

The subsampling test also gives us the choice of $K = 6$ at 5%. It may be noted that if we had only considered the AIC and BIC criteria and looked at $K = 4$ and $K = 5$, we would in fact have concluded that the right number of classes is $K = 4$. We have actually tried to estimate the model for more than 7 classes to test the robustness of our results, but we encountered serious convergence problems for $K = 8$. We either got estimators of the MLE yielding a likelihood smaller than what we obtained with $K = 6$ or $K = 7$, or non-convergence of the algorithm when changing the starting values. This strongly indicates some identifiability problem for $K = 8$. Consequently, based on our previous tests, we have retained the number of $K = 6$ classes. For $K = 6$, more than 92% of the households belong to a specific class with a posterior probability close to 1. This means that the discriminatory power of the model is very high. The remaining 8% have a large probability (superior to 0.7 except for a few "misclassified" individuals) to belong to at least one class and smaller probability to belong to the others. To have a better view of the classes, using a standard bayesian rule, we have decided to allocate the households to the class for which they have the largest final posterior probability. The characteristics of the groups corresponding to this classification are given in table 2. Other methods based on a simultaneous estimation of the model and the composition of the classes will be studied elsewhere.

The demand system is estimated simultaneously for the 6 classes revealed through the segmentation process. Let us recall that those classes identify homogeneous segments of fruit and vegetable demand. The estimations obtained with the QAIDS model seldom show significativeness of the quadratic term in expenditure for some classes and may yield convergence problems for large K ($\geq$6). Consequently we only present here the AIDS specification results. We hope to validate these results on the QAIDS on a much larger sample in the future. Concerning the exogeneity of the total food expenditure, the results indicate that it is rejected except in the lowest income class.

4 Empirical Results

4.1 Relevance of income as a criterion for segmentation

A preliminary look at class characteristics (table 2) shows that the main structuring variable here is *not* income. The gross measure of household income ranks the lowest strata in K1, then in K4. The remaining four classes have levels closely grouped together above these two. Instead, a relevant ranking appears when we use adult-equivalent income. Family composition is then an important variable. Note the small size of this first segment (corresponding both to the lowest income and adult-equivalent income class): it represents only 7.8% of the population, well below the first quartile used in preceding studies, and not even very close to the decile, which is the most common income interval considered in poverty studies.

This is only partially coherent with Jensen and Manrique's (1998) classifying approach. They used total expenditure as a measure of income for classifying households into income groups and for estimating a complete demand system. On our sample, fruit and vegetable expenditure gives also the same ranking as income, but does not provide the best explanation for heterogeneity of food behavior. Jarque's clustering (1987) obtained a segmentation of households explained by occupational category more than by income classes.

4.2 Determination of homogeneous classes in fruit and vegetable demand

These results are based on a multinomial logit performed to explain the membership to a class (table 6), on the estimation of the determinants of fruit and vegetable expenditure by segment (table 3[9]), and on the computation of expenditure and price elasticities at the mean point for each class (tables 4 and 5). We will focus particularly on the first segment which we identify as the deprived class (table 2, where maximum values are in bold and minimum in italics).

K1: a multi-deprived class, where FV expenditure is low and insensitive to economic variables

This segment simultaneously holds numerous aspects of deprivation: the lowest consumption, the lowest household income, and the highest proportion of households where the reference member has no education. It is constituted mainly by retired people (25% of households), workers: 25%, employees: 23% and mostly has a rural profile. It should be noticed that a great proportion of retired people in this class come from an original class with a much higher household income. It suggests that K1 may be a grouping of heterogeneous categories, one being a muti-deprived population, another being an elder population whose consumption patterns are directly linked to health conditions and the position in the life cycle. Specific patterns of consumption are visible in the budget structure. We find the highest budget share in fresh potatoes and in most processed categories: canned and frozen fruit, dry fruit, canned and dry vegetables, frozen vegetables, vegetable based convenience food. This expenditure is compensated by a very inferior proportion of fruit juices, and of fresh vegetables other than potatoes. The high share of the budget does not, nonetheless, indicate a significant consumption in absolute terms (table 2). Moreover, this segment has the lowest rank for fresh fruit (less than twice the amount purchased by K5), and fresh vegetables consumption (1 to 3).

The real specificity of K1 is that fruit and vegetable expenditure is insensitive to economic variables : there is no correlation with income, in opposition to the 5 other segments. In food demand studies it is a rather surprising result to find that

[9] Significance tests are indicated at 10% by *, 5% by ** and 1% by ***.

income has no influence. Let us recall that, at the global level of food consumption, a higher income elasticity in low-income households was expected. Nor do we register either the effect of prices on fruit and vegetable expenditure. This may reflect a very strong income restriction leaving no real choice in food consumption, in response to incompressible categories in the budget. The absence of economic determinants does not mean a poor explanatory power of our model, since the R2 value obtained is one of the highest among segments (>58%). The main determinants of expenditure are to be found in socio-demographic characteristics with older, larger households, and landlords producing a positive effect, and low levels of education and self-consumption possibilities having a negative effect. For other segments, socio-demographic variables share explanatory power with income and prices. Despite this general lack of sensitivity to economic variables, all fresh products categories (fruit, vegetables, and potatoes) do have significant price and expenditure elasticities. Note the high values of own price elasticities for fresh fruit and potatoes, since an increase in price would induce a considerable decrease in purchases (as in K4). With regard to expenditure elasticities, fresh fruit and fresh vegetables (potatoes excluded) have high values (>1) compared with the other segments. We observe few substitution effects, only between the three fresh products categories, in particular between fresh fruit and fresh vegetables.

K2: deprivation in monetary resources and high sensitivity to income and prices

This is the segment of families with children: K2 registers the highest household size, the younger reference member in low to median income households. They live in rural or very big cities (except Paris) and frequently have the possibility of home production. Their consumption is marked by the importance of fruit juices (highest budget share) and of convenience food, which consists of processed products suitable for active households with kids, where time as well as money is a scarce resource. Expenditure appears very sensitive to variations in price, with considerably higher elasticities for canned and frozen fruit than for fresh fruit, and also for frozen and fresh vegetables as opposed to potatoes or canned vegetables. Note the effect of the educational level (negative for no education, positive for both categories over the primary level, which constitutes the reference category). Expenditure elasticities are high (over 1) for most products, and are the highest for convenience food: any income or price variation will have a visible effect on consumption. Moreover, this segment registers a great number of substitution and complementarity effects. Fruit and vegetable budgeting seems to be very flexible in its composition and very responsive to prices.

K3 : deprivation in fruit and vegetable consumption

K3 households show deprivation in expenditure, mainly led by canned vegetable consumption, though a median income. These households have a low fruit and vegetable expenditure despite not being unduly constrained by monetary resources. They consist of fewer farmers, reduced possibilities of home production, while their consumption appears orientated towards canned vegetables (with very high price elasticities).

K4: deprivation in education but taste for fresh products

K4 records deprivation in education and shows a low fruit and vegetable expenditure, except for fresh products. This is a segment dominated by retired household characteristics: more than 1/3 of the class are retired. They have the lowest household size and median monetary resources. They maintain a low expenditure level, but with a very high share of fresh products, especially fresh fruit: it makes up 40% of the budget (compensated through low expenditure of canned fruit and convenience food). This segment shows a real taste for fresh products, and obviously has time to prepare food and use traditional recipes... Let us recall that retired households are also found in K1, but their consumption combines a low expenditure with a low share of fresh products.

K5: high resources level and highest fruit and vegetable expenditure

These households have a high adult equivalent income (the highest gross income) and the highest level of education among segments. Urban areas are particularly represented, with fewer possibilities for home production. They have the highest level of expenditure, also the first rank in fresh vegetable (except potatoes) budget share and the lowest in canned and frozen vegetables. All fresh categories are price and expenditure sensitive, with elasticities close to 1, therefore simply reflecting the variations of environment without extra effect.

K6: high household equivalent income and taste for fresh fruit

K6 has the highest income and a low household size, and has a considerable proportion of retired households with a good educational level. With a good level of expenditure, it has the highest budget share for fresh fruit purchases (41%, very close from K4, and similarly a very low share of frozen vegetable consumption). Expenditure elasticities are quite high for fresh vegetables (potatoes and other), but surprisingly very low for fresh fruit. We find higher price elasticities for potatoes and canned products than for fresh ones.

4.3 Deprivation and healthy food demand

Three segments reveal households deprived in some dimension of economic status: K1 on several aspects, K2 on monetary grounds, K4 on educational terms. Low consumption of fruit and vegetables is evidenced in K1 and K3. When we consider healthy food proxied by total fruit and vegetable expenditure, it corresponds to gross income ranking. But taking into account the degree of processing of products purchased leads to different conclusions. A preceding work (Caillavet 2002) on the same data based on a *a priori* classification between 1st quartile of income and the remaining population suggested that healthier products (fresh and frozen) would be less income and price elastic. This is not so clear-cut here for our 6-segment results. The segmentation found here corresponds to a ranking following adult equivalent household income, but it happens to coincide with increasing quantities per capita of fresh products: fruit, vegetables and potatoes. Moreover, in terms of expenditure, it corresponds to increasing budget shares of fresh products (except potatoes), which, as for vitamins concern, has the closest relationship with a healthy dietary pattern.

Fresh fruit purchases do have higher price elasticities in low income and expenditure classes. Note that fresh fruit consumption and fresh vegetable consumption segmentate households in a different way whether this is in terms of budget shares or through their response to economic signals. Fresh fruit concerned consumers are K4 and K6 with inverse characteristics of elasticities (over or under 1, giving a more or less significant response to a variation in income and price), both including an important proportion of retired households. Fresh vegetable concerned consumers are K5 and to a lesser extent, K2.

5 Conclusions

In this study we have tried to establish a deprivation ranking based on fruit and vegetable consumption, as a principal indicator associating healthy eating habits and poverty. Traditional indicators are based on monetary measures or on a wider set of indicators such as living conditions. More recent definitions of the level of welfare such as the UNDP one (1998), include health inequalities as a major criterion. Among these, nutrition plays a fundamental two-fold role: a short term one as an element of consumption and welfare, and a more long-term one as in preventing diseases and contributing to longevity. In this context, we have developed a methodology to segment the population on the criterion of homogeneity of food behavior, using a finite-mixture model. For this, we proceed to a clusterwise regression analysis of fruit and vegetable demand based on a gaussian mixture of components. We obtain six homogeneous segments of the population. Important structuring variables are shares of fresh products in fruit and vegetable expenditure and household composition but not gross income classes. Estimating a demand system on each of these segments, we find very different price and expenditure elasticities, reflecting the diversity of demand for fruit and vegetables in the population. The finding of an income-independent segment of food demand (fruit and vegetable expenditure) has to be stressed: it seems to correspond to a multi-deprived population in terms of income, education, ...(K1). The affectation of retired households with higher income in this class shows K1 as a grouping of heterogeneous categories which have in common a low fruit and vegetable expenditure and share of fresh products. This sign of unhealthy diet would result from strong economic restrictions or particular patterns linked to health and life cycle conditions. In any case, we see clearly that monetary deprivation does not always imply unhealthy eating habit, since the level of education plays a role (K2). At the same time, an average level of income does not guarantee a satisfactory consumption (K3). Traditional eating patterns linked to resources in time certainly explain the importance of fresh products in K4 and K6 households, at different levels of monetary constraints. It is also confirmed by the fact that the conjunction of large resources and high education (high opportunity cost) does not give rise to the healthiest consumption : the highest level of fruit and vegetable expenditure does not in fact imply the highest share of fresh products (K5).

Evidencing different segments of demand suggests targeting specific nutrition recommendations and policies following the population. Concerning poverty, does monetary poverty correspond to deprivation in healthy eating? On the basis of our data, this appears to be true in France to only a limited extent: for a segment of less than 8% of the population. With regard to the remaining segments, educational level or traditional eating habits partially appear to compensate for monetary deprivation. On the other hand, deprivation in healthy eating clearly affects a broader population, which may enjoy a median income level. Thus food consumption patterns may constitute a useful indicator of health deprivation and inequalities rather than of monetary poverty.

References

Banks, J., R. Blundell, and A. Lewbel. "Quadratic Engel Curves and Consumer Demand". *Rev. Econ. Stat.* 79 (1997): 527-539.

Bertail, P. and F. Caillavet. "Food Consumption and Poverty: a segmentation approach". Working Paper, CORELA,INRA, 2003. www.ivry.inra.fr/corela/wp/CaillavetWP0406.pdf.

Bertail, P., D.N. Politis, and J.P. Romano. "On subsampling estimators with unknown rate of convergence". *J. Amer. Stat. Assoc.,* 94 (1999): 569-579.

Blundell, R., and J.M. Robin. "Estimation in Large and Disaggregated Demand Systems: An Estimator for Conditionally Linear Systems". *J. Applied Econometrics*, 14 (1997): 209-232.

Caillavet, F. . "Low-income households and healthy food consumption in France". Working Paper, CORELA, INRA, 2002.

Deaton, A., and J. Muellbauer." An Almost Ideal Demand System". *Am.Econ. Rev.*, 70 (1980): 312-326.

Dempster, A.P., N.M. Laird, and D.B. Rubin. "Maximum likelihood for incomplete data via the E-M algorithm", *J.R.S.S.(B)* 39(1977): 1-38.

Department of Health. "Increasing fruit and vegetable consumption: a national priority".2000. *http://www.doh.gov.uk/fiveaday/index.htm*

DeSarbo, W.S., and W.L. Cron. "A maximum likelihood methodology for clusterwise linear regression". *J. Classification* 5 (1988): 249-283.

Feng, X., and W.S. Chern. "Demand for Healthy Food in the United States", Paper presented at AAEA annual meeting, Tampa, Florida, 30 July - 2 August 2000.

Freedman, D., T. Hastie, and R. Tibshirani. *Element of statistical learning: data mining, inference and prediction*, Wiley, N.Y., 2001.

Fuhrer R., M.J. Shipley, J.F. Chastang, A. Schmaus, I. Niedhammer, S.A. Stansfeld, M. Goldberg, and M.G. Marmot. "Socioeconomic Position, Health, and Possible Explanations : A Tale of Two Cohorts".*Amer. J. Pub. Health* 92 (2002): 1290-1294.

Jarque, M.C. "An application of limited dependent variable models to household expenditure analysis in Mexico". *J.Econometrics* 36 (1987): 31-53.

Jensen, H.H., and J. Manrique. "Demand for food commodities by income groups in Indonesia", *Appl. Econ.* 30 (1998): 491-501.

Johansson, L., D.S. Thelle, K. Solvoll, G.A. Bjorneboe, and C. Drevon. "Healthy dietary habits in relation to social determinants and lifestyle factors". *Brit. J. Nutr.* 81 (1999): 211-220.

McLachlan, G.J., and D. Peel. *Finite Mixture Models*. New York: John Wiley, 2000.

Morland, K., S. Wing, and A. Diez Roux. "The Contextual Effect of the Local Food Environment on Resident's Diets: The Atherosclerosis Risk in Communities Study". *Amer. J.Pub. Health* 92 (2002): 1761-1767.

Park J.L., R.B.Holcomb, K.C.Raper, and O. Capps Jr . "A Demand Systems Analysis Segmented by Income". *Amer. J. Agr. Econ.* 78 (1996): 290-300.

Politis, D.N., and J.P.Romano. "Large sample confidence regions based on subsamples under minimal assumptions". *Ann. Statist.* 22 (1994): 2031-2050.

Roos G., L. Johansson, A. Kasmel, , J. Klumbiené, and R. Prättälä. "Disparities in vegetable and fruit consumption: European cases from the north to the south". *Pub. Health Nutr.* 4 (2000): 35-43.

Searle, S.R., G. Cassela, and C.E. McCulloch. *Variance Components.* New York: John Wiley & Sons, Inc., (pp. 275-286) 1992.

Sen, A. "*Resources, Values and development"*. Cambridge: Harvard University Press, 1984.

Titterington, D. M., A. F. M.Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley, 1985.

United Nations Development Programme. *Human Development Report* 1998. New York, 1998.

Wedel, M., W.S. DeSarbo. "A mixture likelihood approach for generalized linear models". *J. Classification* 12 (1995): 21-56.

Acknowledgements: We thank Martin Minchom and Michael Visser (Inra-Corela) for careful reading of the manuscript, as well as Stefan Lollivier (Ensae) for his remarks on a previous version. Special thanks to Véronique Nichele (Inra-Corela) for providing a preliminary version of a Stata QAIDS estimation program and her critical reading. This research has benefited from a grant from INRA-NHSA program "A.I.P. Comportement Alimentaire" directed by Ph. Verger.

Table 1: **Description of the sample**, SECODIP 1997 - n=2005

Variable			**Mean**	**S.E.**
Household income (F/month)		income	12958.23	6776.54
Adult-equiv Household income		AEincome	6233.48	3636.92
Fruits and vegetables expenditure (F/year)		FVexp	3086.94	1841.14
Household size		hsize	3.04	1.45
Max level of education in the household	primary school	educ1	0.18	0.39
	<bac	*educ2*	0.40	0.49
	bac	educ3	0.18	0.39
	>bac	educ4	0.23	0.42
Availability of a kitchen garden	no availability	*gard0*	0.45	0.50
	orchard only	gard1	0.09	0.28
	kitchen garden only	gard2	0.15	0.36
	both	gard3	0.31	0.46
Inactive woman in the household (including retired)		inactw	0.43	0.50
Administrative region	Paris region	zeat1	0.15	0.35
	Parisian bassin	zeat2	0.18	0.39
	North	zeat3	0.07	0.26
	East	zeat4	0.11	0.32
	West	zeat5	0.15	0.36
	South-west	zeat7	0.10	0.30
	Center-east	zeat8	0.13	0.33
	Mediterranean region	*zeat9*	0.10	0.30
householder			0.62	0.49
Size of residential area	rural	resid1	0.25	0.43
	$R < 50\ 000$ inhabitants	*resid2*	0.25	0.43
	$50\ 000 < R < 200\ 000$	resid3	0.16	0.37
	$R > 200\ 000$	resid4	0.34	0.47
Age of the household head	21-40 years	age1	0.29	0.45
	41-50	age2	0.28	0.45
	51-60	*age3*	0.16	0.37
	61-96	age4	0.27	0.44

Table 2: **Characteristics of homogeneous classes in fruit (F) and vegetable (V) demand**

Mean values		**K1**	**K2**	**K3**	**K4**	**K5**	**K6**
income		*10915*	13340.43	13147.53	12483.08	**13374.71**	13150.16
Aeincome		*5628.81*	5920.57	6161.38	6268.9	6454.44	**6644.85**
FVexp(F/month)		*1932.59*	3030.26	2718.57	2974.95	**3865.67**	3226.6
Budget shares (%)	Fresh F	*21.51*	26.52	25.7	40	35.66	**41**
	Canned and frozen F	**6.26**	3.07	4.73	*1.26*	1.45	3.6
	Dry F	**9.96**	5.23	3.71	5.46	*2.35*	2.4
	F juices	*5.07*	**18.68**	11.46	7.75	5.52	5.51
	Potatoes	**6.26**	*2.67*	4.11	4.21	4.47	3.83
	Fresh V	19.8	*19.31*	25.92	27.26	**33.42**	28.25
	Canned and dry V	17.02	12.41	15.04	9.14	8.5	10.44
	Frozen V	**5.34**	3.97	1.66	*0.5*	3.94	0.29
	V convenience food	**8.8**	8.15	7.68	*4.41*	5.69	4.69
Quantities/cap (kg)	Fresh F	*21.07*	29.84	25.56	52.89	**55.59**	55.27
	Canned and frozen F	**3.82**	2.41	3.31	*1.19*	1.63	2.76
	Dry F	**2.38**	1.76	1.13	2.1	1.29	*1.12*
	F juices	*5.86*	**27.89**	16.9	14.36	12.76	9.68
	Potatoes	13.8	*8.35*	11.66	16.28	**17.36**	15.1
	Fresh V	**17.01**	20.38	24.19	34.32	*49.04*	39.53
	Canned and dry V	7.05	7.81	**8.49**	*6.57*	7.68	7.61
	Frozen V	3.35	3.18	1.43	0.52	**4.16**	*0.34*
	V convenience food	3.63	4.8	4.19	*2.83*	**4.89**	3.06
Socio-demo	age	50.22	*45.92*	47.47	**54.02**	52.02	52.15
	educ1	**31.33**	*14.1*	15.56	21.54	19.95	16.04
	educ2	42	40.96	38.52	38.15	36.43	44.97
	educ3	12	18.88	20.25	19.38	17.4	19.18
	educ4	*14.67*	26.06	25.68	20.92	**26.22**	19.81
	hsize	2.88	**3.33**	3.2	*2.86*	2.98	2.85
	inactw (%)	38	*32.45*	41.98	**51.69**	48	44.03
	householders (%)	57.33	*56.91*	58.52	64.61	63.11	**71.38**
	farmers (%)	1.33	*3.99*	**1.23**	3.38	1.39	1.26
	retired (%)	25.33	*16.49*	20.49	**35.38**	28.31	32.39
	no self-consumption (%)	47.33	43.88	46.91	45.54	**47.8**	*37.42*
n		150	376	405	325	431	318

Table 3: **Estimation of fruit and vegetable expenditure (log) by class**

	K1			**K2**			**K3**			**K4**			**K5**			**K6**		
	Coef.		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*
age1	-0.342	*	0.181	-0.222	***	0.084	-0.382	***	0.087	-0.618	***	0.055	-0.198	***	0.078	-0.414	***	0.074
age2	-0.113		0.185	0.186	**	0.086	-0.335	***	0.087	-0.362	***	0.051	-0.100		0.072	-0.182	***	0.070
age4	0.398	**	0.184	0.524	***	0.108	-0.106		0.098	-0.030		0.052	0.073		0.079	0.133	*	0.075
hsize	0.288	***	0.051	0.145	***	0.023	0.113	***	0.024	0.092	***	0.014	0.101	***	0.020	0.198	***	0.021
educ1	-0.512	***	0.142	-0.169	**	0.085	-0.347	***	0.083	0.528	***	0.047	-0.355	***	0.071	0.038		0.064
educ3	-0.050		0.183	0.246	***	0.075	0.076		0.075	0.320	***	0.047	-0.038		0.067	-0.156	***	0.057
educ4	0.192		0.179	0.189	***	0.074	0.023		0.073	0.189	***	0.050	-0.060		0.064	0.105	*	0.061
zeat1	-0.296		0.261	-0.088		0.133	0.102		0.135	-0.182	**	0.080	-0.289	***	0.113	0.143		0.110
zeat2	-0.905	***	0.230	0.048		0.109	-0.020		0.110	-0.475	***	0.068	0.012		0.090	-0.078		0.093
zeat3	-0.875	***	0.325	-0.150		0.136	-0.089		0.148	-0.434	***	0.084	-0.060		0.137	0.178		0.122
zeat4	-0.842	***	0.249	-0.007		0.122	-0.039		0.126	-0.219	***	0.077	-0.019		0.094	0.339	***	0.096
zeat5	-0.862	***	0.259	-0.101		0.122	-0.251	*	0.135	-0.133	*	0.073	-0.121		0.107	0.064		0.101
zeat7	-1.163	***	0.240	0.122		0.122	0.161		0.117	-0.344	***	0.071	0.069		0.106	-0.269	**	0.109
zeat8	-0.830	***	0.273	-0.021		0.115	0.068		0.121	0.272	***	0.070	-0.278	***	0.096	0.002		0.096
resid1	-0.484	***	0.164	0.074		0.082	-0.141	*	0.077	-0.230	***	0.047	-0.012		0.069	-0.242	***	0.059
resid3	0.367	*	0.192	0.153		0.098	-0.220	*	0.098	-0.391	***	0.059	0.152	*	0.079	-0.113	*	0.067
resid4	-0.143		0.170	0.147	*	0.087	0.008		0.083	-0.065		0.050	0.020		0.072	-0.064		0.064
gard1	0.199		0.201	0.116		0.103	-0.094		0.108	-0.241	***	0.064	0.173	*	0.096	-0.401	***	0.080
gard2	-0.006		0.189	-0.194	***	0.081	0.047		0.084	0.143	***	0.052	-0.090		0.071	0.167	***	0.064
gard3	-0.502	***	0.152	-0.001		0.070	-0.256	***	0.073	-0.503	***	0.047	-0.168	***	0.061	0.021		0.056
inact woman	0.049		0.127	0.076		0.066	0.137	**	0.063	-0.196	****	0.039	0.205	***	0.057	0.108	**	0.052
householder	0.292	***	0.122	-0.142	**	0.064	0.135	**	0.064	0.092	***	0.038	0.134	***	0.058	0.102	**	0.051
lp1	0.533		0.970	0.399		0.548	-1.003	**	0.506	1.163	***	0.333	-0.934	**	0.461	1.752	***	0.379
lp2	-0.030		0.412	0.285		0.227	-0.157		0.216	-0.172		0.121	0.091		0.160	-0.203		0.150
lp3	-0.710		0.442	-0.476	**	0.240	-0.105		0.262	-0.315	**	0.152	0.026		0.199	0.066		0.181
lp4	-0.351		0.563	-0.112		0.258	0.237		0.234	0.672	***	0.144	-0.072		0.205	0.139		0.187
lp5	0.009		0.368	-0.186		0.205	-0.282		0.218	-1.226	***	0.118	0.318	*	0.180	0.183		0.145
lp6	-0.733		0.686	-0.068		0.434	-0.489		0.428	0.284		0.218	-0.150		0.402	-0.439		0.319
lp7	1.381	*	0.746	0.122		0.212	0.052		0.274	-0.643	***	0.187	0.487	**	0.239	-0.121		0.189
lp8	0.129		0.221	0.333	***	0.121	-0.108		0.108	-0.087	*	0.071	0.172	*	0.105	-0.107		0.081
lp9	0.418		0.342	-0.401	**	0.191	0.330	**	0.159	-0.031		0.103	-0.072		0.140	-0.044		0.119
income(log)	0.065		0.116	0.382	***	0.075	0.468	***	0.067	0.146	***	0.039	0.279	***	0.056	0.272	***	0.054
cons	2.805		2.852	2.651	*	1.529	5.385	***	1.528	5.421	***	0.873	3.819	***	1.234	0.589		1.181
R2	0.581			0.4386			0.4094			0.7346			0.3728			0.6229		

Table 4: **Price and expenditure elasticities by class**

Price Elasticities		**K1**			**K2**			**K3**			**K4**			**K5**			**K6**	
	Coef.		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*
Fresh F	-2.237	***	0.506	-1.097	***	0.232	-1.089	***	-0.054	-1.910	***	0.298	-0.225		0.207	-0.581	***	0.268
Canned and frozen F	1.396		2.137	-1.939	***	0.364	-0.439		*0.479*	-1.010	***	0.158	-0.881	***	0.118	-1.868	***	0.47
Dry F	1.374		2.420	-0.005		0.480	-1.251	***	-0.014	-0.427		0.519	-0.785	***	0.165	-1.078	***	0.189
F juices	-1.295	***	0.470	-0.262		0.499	-1.929	***	0.394	-1.759	***	0.313	-1.241	***	0.166	-0.866	***	0.211
Potatoes	-3.423	***	*1.550*	-0.701	***	0.185	-1.733	***	-0.264	-4.283	***	0.780	-0.814	***	0.311	-1.44	***	0.295
Fresh V	-1.114	***	0.486	-1.494	***	0.191	-0.001		-0.226	-0.351		0.225	-1.1	***	0.214	-0.727	***	0.251
Canned and dry V	0.108		*2.296*	-0.559	*******	0.230	-2.440	***	-0.119	-1.670	***	0.258	-0.712	***	0.175	-1.037	***	0.212
Frozen V	0.295		0.987	-1.762	***	0.409	-0.835	***	-0.070	-1.019	***	0.061	-0.246		0.345	-0.998	***	0.034
V convenience food	-0.779		0.889	-1.214	***	0.364	-0.195		*0.305*	-0.956	***	0.204	-0.901	***	0.227	-1.19	***	0.221

Expenditure elasticities		**K1**			**K2**			**K3**			**K4**			**K5**			**K6**	
	Coef.		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*
Fresh F	1.584	***	0.294	1.290	***	0.094	0.999	***	0.053	1.867	***	0.117	1.171	***	0.099	0.397	***	0.126
Canned and frozen F	5.579	***	2.071	1.516	***	0.285	*1.010*	***	0.326	0.431	*	0.234	0.937	***	0.141	-0.569		0.537
Dry F	1.454		2.496	1.294	***	0.342	0.703	***	0.151	0.449		0.371	1.056	***	0.159	1.093	***	0.187
F juices	0.836		0.616	-0.543	*	0.309	1.198	***	0.168	1.189	***	0.219	0.851	***	0.147	0.628	***	0.204
Potatoes	-2.152		1.600	1.314	***	0.165	*1.474*	***	0.188	-1.022	*	0.600	0.179		0.305	2.252	***	0.339
Fresh V	1.348	***	0.484	1.198	***	0.084	*1.041*	***	0.067	0.790	*******	0.099	1.116	***	0.11	2.016	***	0.148
Canned and dry V	-0.675		1.423	0.653	***	0.179	0.827	***	0.147	0.109		0.173	0.921	***	0.129	0.57	***	0.194
Frozen V	-0.541		1.835	0.457		0.509	*1.268*	***	0.163	0.803	***	0.162	-0.459		0.552	1.089	***	0.077
V convenience food	-0.016		1.115	1.208	***	0.302	0.644	***	0.241	0.835	***	0.200	1.094	***	0.283	1.041	***	0.308

Table 5: **Compensated price elasticities by class**

K1	Fresh F		C/fro F		Dry F		F juic.		Potat.		Fresh V		C/dry V		Froz. V		V conv.	
Fresh F	-1.727	***	-0.208		0.194		0.321	*	0.396	*	0.91	***	0.424		-0.168		-0.143	
Canned and frozen F	-2.167		1.569		1.146		-0.207		-0.516		0.588		1.32		0.204		-1.938	
Dry F	1.471		0.832		1.436		-0.705		-1.253		1.217		-3.21		-0.33		0.542	
F juices	1.088	*	-0.067		-0.315		-1.215	***	-0.19		0.268		0.449		-0.096		0.078	
Potatoes	3.152	*	-0.393		-1.315		-0.446		-3.51	**	-0.268		2.031		-0.356		1.104	
Fresh V	1.109	***	0.069		0.196		0.097		-0.041		-0.758		-0.793		0.189		-0.068	
Canned and dry V	1.176		0.351		-1.176		0.368		0.709		-1.806		0.029		-0.129		0.477	
Frozen V	-2.201		0.256		-0.571		-0.372		-0.587		2.039		-0.609		0.282		1.762	*
V convenience food	-0.717		-0.932		0.358		0.116		0.697		-0.279		0.862		0.674	*	-0.78	
K2	Fresh F		C/fro F		Dry F		F juic.		Potat.		Fresh V		C/dry V		Froz. V		V conv.	
Fresh F	-0.768	***	-0.116	*	0.084		0.19	**	-0.018		-0.085		0.636	***	0.066	**	0.01	
Canned and frozen F	-1.208	**	-0.408		0.196		0.365		1.565		-1.884	***	0.862		-0.033		0.543	
Dry F	0.635		0.143		-1.221	***	-0.696	***	-0.557		1.481	***	0.025		0.159		0.031	
F juices	0.644	**	0.119		-0.311	***	-1.815	***	0.343		0.402		0.295		-0.148	**	0.471	***
Potatoes	-0.145		1.193	***	-0.585	***	0.805	***	-1.673	***	-0.29		0.874	***	-0.01		-0.169	
Fresh V	-0.104		-0.22	***	0.238	***	0.145		-0.044		0.274		-0.109		-0.02		-0.159	**
Canned and dry V	1.767	***	0.229		0.009		0.242		0.305	***	-0.248		-2.344	***	0.106	*	-0.066	
Frozen V	0.865	**	-0.041		0.276		-0.572	**	-0.017		-0.22		0.501	*	-0.803	***	0.011	
V convenience food	0.051		0.261		0.02		0.698	***	-0.107		-0.656	**	-0.118		0.004		-0.154	
K3	Fresh F		C/fro F		Dry F		F juic.		Potat.		Fresh V		C/dry V		Froz. V		V conv.	
Fresh F	-0.681	***	0.114	**	0.009		-0.148		0.181	***	0.483	***	0.023		0.092		-0.073	
Canned and frozen F	1.189	**	-1.892	***	0.468		-0.403		-0.422	**	1.697	***	-0.504		-0.367	*	0.233	
Dry F	0.07		0.34		0.05		0.227		-0.263		-1.017	*	0.074		-0.758	***	1.277	***
F juices	-0.502		-0.131		0.101		-0.314		-0.055		0.511		-0.512	*	0.632	***	0.268	
Potatoes	1.442	***	-0.321	**	-0.275		-0.129		-0.648	***	-0.227		0.095		0.009		0.054	
Fresh V	0.589	***	0.198	***	-0.163	*	0.184		-0.035		-1.177	***	0.329	***	0.013		0.062	
Canned and dry V	0.063		-0.134		0.027		-0.419	*	0.033		0.749	***	-0.483	**	0.091		0.073	
Frozen V	1.208		-0.462	*	-1.312	***	2.451	***	0.015		0.14		0.431		-1.751	***	-0.72	
V convenience food	-0.365		0.112		0.845	***	0.397		0.034		0.257		0.132		-0.275		-1.136	***

Table 5 continued :**Compensated price elasticities by class**

K4	Fresh F	C/fro F	Dry F	F juic.	Potat.	Fresh V	C/dry V	Froz. V	V conv.
Fresh F	-1.309 ***	0.142 ***	0.067	-0.066	0.668 ***	-0.247	0.652 ***	0.082 ***	0.01
Canned and frozen F	1.482 ***	-0.996 ***	-0.232	0.851 ***	-0.8 ***	0.473 *	-0.92 ***	0.012	0.13
Dry F	0.512	-0.169	-0.408	1.046 ***	-0.301	-0.919	-0.001	-0.165 **	0.407
F juices	-0.222	0.276 ***	0.467 ***	-1.646 ***	0.036	0.468	0.643 ***	0.126 ***	-0.148
Potatoes	5.312 ***	-0.61 ***	-0.316	0.084	-4.324 ***	0.745	-1.166 **	-0.241 *	0.517 *
Fresh V	-0.301	0.055 *	-0.148	0.169	0.114	-0.142	0.058	0.037 **	0.158 **
Canned and dry V	1.811 ***	-0.245 ***	**0**	0.527 ***	-0.407 **	0.131	-1.657 ***	-0.048	-0.113
Frozen V	1.082 ***	0.015	-0.286 **	0.489 ***	-0.398 *	0.403 **	-0.226	-0.999 ***	-0.079
V convenience food	0.048	0.063	0.269	-0.219	0.326 *	0.65 **	-0.204	-0.03	-0.903 ***

K5	Fresh F	C/fro F	Dry F	F juic.	Potat.	Fresh V	C/dry V	Froz. V	V conv.
Fresh F	-0.453 *	0.097	-0.025	-0.024	0.157 ***	-0.077	0.271 ***	0.007	0.047
Canned and frozen F	1.008	-1.886 ***	-0.269	-0.953 ***	-0.077	-0.305	1.469 ***	0.078	0.934 ***
Dry F	-0.191	-0.195	-1.032 ***	0.779 ***	0.232 *	0.471	-0.376 **	0.014	0.299 ***
F juices	-0.081	-0.309 ***	0.348 ***	-0.806 ***	0.02	0.371	0.012	-0.038 *	0.484 ***
Potatoes	1.247 ***	-0.059	0.243 *	0.046	-1.349 ***	0.065	-0.437	0.059	0.187
Fresh V	-0.094	-0.036	0.076	0.134	0.01	-0.194	0.11	0.049 ***	-0.055
Canned and dry V	0.753 ***	0.391 ***	-0.138 **	0.01	-0.153	0.251	-0.971 ***	0.06 ***	-0.204 *
Frozen V	0.098	0.098	0.024	-0.149 *	0.097	0.524 ***	0.282 ***	-0.971 ***	-0.002
V convenience food	0.237	0.449 ***	0.198 ***	0.717 ***	0.118	-0.226	-0.368 *	-0.001	-1.123 ***

K6	Fresh F	C/fro F	Dry F	F juic.	Potat.	Fresh V	C/dry V	Froz. V	V conv.
Fresh F	0.152	-0.027	-0.055	0.051	-0.049	-0.107	0.01	-0.094 *	0.119 *
Canned and frozen F	-0.28	-0.852 ***	-0.102	-0.097	0.09	1.036 ***	-0.061	0.22 ***	0.047
Dry F	-0.415	-0.074	-0.74 ***	0.088	0.075	0.511 *	0.212	0.13	0.213 *
F juices	0.172	-0.032	0.039	-1.16 ***	0.101	0.83 ***	0.417 ***	-0.167 **	-0.201 **
Potatoes	-0.391	0.069	0.079	0.238	-0.807 ***	-0.072	0.366	0.242	0.276
Fresh V	-0.131	0.121 ***	0.082 *	0.299 ***	-0.011	-0.805 ***	0.178 **	0.165 ***	0.102
Canned and dry V	0.028	-0.016	0.078	0.342 ***	0.128	0.405 **	-0.605 ***	-0.173 ***	-0.186 **
Frozen V	-1.238 *	0.276 ***	0.225	-0.648 **	0.399	1.777 ***	-0.817 ***	-0.257	0.284
V convenience food	0.599 *	0.023	0.141 *	-0.298 **	0.174	0.419	-0.335 **	0.109	-0.831 ***

Table 6 : **Multinomial logit explanations for class membership**

	K1			K2			K3			K4			K6		
	Coef.		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*	*Coef.*		*S. E.*
inc1	0.753	*	0.391	-0.277		0.327	-0.144		0.332	-0.33		0.275	-0.037		0.308
inc2	1.019	***	0.378	0.307		0.297	0.413		0.298	0.042		0.267	0.164		0.299
inc3	0.648	*	0.352	-0.319		0.281	-0.186		0.277	-0.012		0.234	0.001		0.266
inc5	0.053		0.359	-0.389		0.263	-0.212		0.25	-0.553	***	0.231	-0.046		0.249
inc6	0.212		0.405	-0.535	*	0.312	-0.357		0.295	-0.505	*	0.269	-0.144		0.292
inc7	0.004		0.418	-0.255		0.294	0.039		0.271	-0.417		0.254	0.133		0.274
inc8	0.231		0.549	-0.264		0.413	-0.894	**	0.443	-0.417		0.355	-0.051		0.4
age1	0.304		0.31	0.75	***	0.255	0.791	***	0.263	-0.108		0.226	0.531	**	0.248
age2	0.096		0.294	0.288		0.252	0.48	*	0.256	-0.03		0.212	0.416	*	0.237
age4	-0.032		0.336	-0.164		0.287	0.838	***	0.285	0.353		0.223	0.194		0.255
hsize	0.242	***	0.081	0.025		0.07	0.288	***	0.067	0.082		0.062	-0.09		0.068
educ1	0.745	***	0.265	0.312		0.243	0.198		0.244	0.147		0.202	-0.196		0.233
educ3	-0.086		0.291	0.094		0.226	0.471	***	0.208	0.015		0.2	-0.08		0.211
educ4	-0.364		0.304	-0.27		0.226	0.136		0.215	-0.025		0.195	-0.298		0.211
zeat1	0.058		0.392	0.213		0.309	0.123		0.323	0.259		0.289	-0.108		0.314
zeat2	0.03		0.359	0.071		0.304	0.157		0.3	0.397		0.265	0.247		0.284
zeat3	-0.215		0.465	-0.2		0.399	0.341		0.367	0.599	*	0.325	0.089		0.367
zeat4	-0.169		0.407	-0.289		0.348	0.046		0.327	0.22		0.291	0.076		0.311
zeat5	0.065		0.376	-0.099		0.332	0.509	*	0.306	0.388		0.279	0.051		0.306
zeat7	-0.214		0.422	-0.01		0.348	0.266		0.33	0.289		0.296	0.055		0.323
zeat8	-0.663		0.411	-0.208		0.314	-0.194		0.316	-0.195		0.285	0.017		0.292
resid1	0.673	***	0.276	0.039		0.237	-0.128		0.216	0.077		0.196	-0.192		0.211
resid3	0.141		0.332	-0.115		0.269	-0.041		0.239	0.121		0.21	-0.245		0.23
resid4	0.363		0.302	0.424	*	0.231	0.075		0.221	-0.045		0.201	-0.142		0.212
gard2	0.082		0.365	0.391		0.321	0.416		0.301	0.195		0.267	0.462		0.289
gard3	-0.045		0.309	0.051		0.259	0.605	***	0.224	0.079		0.21	0.282		0.228
gard4	0.102		0.263	0.673	***	0.212	0.665	***	0.205	0.31	*	0.182	0.545	***	0.199
inactw	-0.243		0.225	-0.164		0.192	-0.292		0.182	0.127		0.161	0.011		0.179
householder	0.261		0.228	-0.044		0.184	0.042		0.179	0.115		0.16	0.12		0.175
cons	-2.706	***	0.568	-0.968	**	0.446	-2.337	***	0.45	-0.606		0.386	-0.535		0.42