

# Household Food Consumption, Individual Caloric Intake and Obesity in France\*

Céline Bonnet,<sup>†</sup> Pierre Dubois,<sup>‡</sup> Valérie Orozco<sup>§</sup>

First Version: January 2007. This Version: May 2011

## Abstract

We show how to use a long period of observation of all food purchases at the household level in order to infer the profile of average individual caloric intakes according to the gender, age and body mass index of household members. Using data from France, we apply this method to analyze the relationship between obesity and individual food consumption. The results show that obese or overweight individuals do absorb more calories at all ages but with differences that vary across gender and ages and across food nutrients such as carbohydrates, lipids or proteins.

**Key words:** Body Mass Index, obesity, nutrients, energy, individual food consumption

**JEL codes:** I18, D12

---

\*We thank Namanjeet Ahluwalia, Andrew Chesher, Nicole Darmon, Catherine Esnouf, Rachel Griffith, Vincent Réquillart, François-Charles Wolff for useful comments, all remaining errors are ours.

<sup>†</sup>Toulouse School of Economics (GREMAQ, INRA), 21 Allée de Brienne, F-31000 Toulouse

<sup>‡</sup>Toulouse School of Economics (GREMAQ, INRA, IDEI) and CEPR, 21 Allée de Brienne, F-31000 Toulouse

<sup>§</sup>Toulouse School of Economics (GREMAQ, INRA), 21 Allée de Brienne, F-31000 Toulouse

# 1 Introduction

The objective of this paper is to study the variation in food consumption and its relationship with obesity across individuals in France from household food purchases. Standard nutritional surveys directly make individual-level measures of food consumption but they can usually be done over short periods only and are thus subject to the large daily variations in consumption which make it difficult to document the relationship between food diet and obesity, or cast doubts over the reliability of the estimation of such relationships in the long run. In this paper, we show how to use a long period of observation of all food purchases at the household level to infer the profile of average individual food intakes according to gender, age and body mass index. Using data from France, we apply this method in order to analyze the relationship between individual food consumption and obesity, which is a growing health problem in France.

Obesity and overweight have actually been increasing in France since the 1990s. According to the 2003 Decennial Health Survey by INSEE (Paraponaris et al., 2005), the percentage of overweight increased from 32.9% to 37.5% and obesity from 6.3% to 9.9% between 1980 and 2003. The health problems related to obesity are consequently increasing. Obesity has been linked to various medical conditions such as hypertension, high cholesterol, coronary heart diseases, type 2 diabetes, psychological disorders such as depression, and various types of cancer. In the US, obesity costs more in medical care expenditure than cigarette smoking — around \$75 billion in 2003 — because of the long and costly treatments for its complications (Grossman and Rashad, 2004). Including the indirect costs such as lost days of work and reduced productivity in addition to direct costs such as personal health care, hospital care, physician services and medications, Wolf and Colditz (2006) estimate the total cost of obesity in the US in 1995 at a total of \$99.2 billion. Using French data from the Decennial Health Survey, Paraponaris et al. (2005) show that the overweight and obesity status reduces the employability of workers. Moreover, using a prevalence-based approach identifying the costs incurred during 1992 by obese people, Levy et al. (1995) find a conservative estimate of direct and indirect

costs of obesity for France of more than €1.8 billion for direct costs and €0.1 billion for indirect costs. Controlling for age, gender, professional categories and alcohol and cigarettes consumptions, Emery et al. (2007) find that obesity would imply an additional cost in health expenditure in France of between €2.1 and €6.2 billion per year.

Among the causes of the obesity epidemic, technological explanations based on the induced relative costs of food products and the surge of calorie intake are the most important (Lakdawalla and Philipson 2002; Cutler, Glaeser and Shapiro 2003; Cutler and Glaeser 2005; Chou, Grossman and Saffer 2004). Health policy has recommended dietary intakes of nutrients (for example a reduction in lipids to less than 35% of the total food intake) that are not necessarily respected by the population and such observation is not necessarily easy. The simple question of the relationship between body mass indices and food consumption is thus quite important for policy recommendations.

Others have studied such question differently. Nichèle et al. (2005) study the long term evolution of nutrition in France and its link to obesity but assume an equal division of food among household members. Ransley et al. (2003) use supermarket receipts to estimate the energy and fat content of food purchased by lean and overweight families in the UK, considering the entire household and not its members. On a 28-day basis, they find that overweight households purchase significantly more energy and fat per adult equivalent than lean households.

Our objective is to analyze the relationship between food consumption and obesity in France at the individual level, avoiding the household aggregation bias because households often consist of parents and children that obviously have very different energy requirements. The problem in the study of this relationship is that in general we do not observe individual consumption over long periods of time. Actually, the standard nutritional surveys provide precise information on food intakes and health outcomes at the individual level but usually over a relatively short period of time. Another type of survey consists in food frequency questionnaires that do not provide precise information on quantities and that are subject to major approximation errors by respondents. As an example of

such a nutritional study, the French nutritional INCA (Individual National Study of Food Consumption) by AFSSA (Agence Française de Sécurité Sanitaire des Aliments) consists in a survey of more than 4,000 individuals who have to fill in declaration books of all their food consumption during one week. In the US, the National Health and Nutrition Examination Survey (NHANES) collects information on dietary behavior through personal interviews over a 24-hour dietary recall (survey participants of 12 years and older complete the dietary interview on their own, while proxy respondents report for children and other persons who cannot self-report). These dietary recalls are potentially precise but over a 24-hour period only and are subject to mistakes regarding the consumption of household members who cannot respond themselves, such as children. The NHANES also collects information by means of a food frequency questionnaire mailed to respondents. Information on the frequency of consumption of foods and food groups during the previous 12 months is collected, but is obviously less precise and subject to larger recall errors.

Using a large French data set recording individual characteristics (age, gender, weight and height) and household food purchases covering 354 product categories over a two-year period which we matched with nutritional information on all products, we recover individual-level estimates of nutrient consumption by extending a method introduced by Chesher (1998). In particular, we are able to take into account the differences in Body Mass Index (BMI) in the individual food demands; this is an important source of heterogeneity explaining the link between food consumption and obesity. The results show that obese or overweight individuals do absorb more calories at all ages but with differences that vary across genders and ages and across nutrients such as carbohydrates, fats or proteins.

It is to be noted that our data present the advantage of providing two years of food demand. Standard nutritional studies typically use a week (or less) of observation of food intakes or dietary history interviews on daily intakes that are subject to poor memory, measurement errors and subjective perception mistakes. As the relationship between food

intakes and the body mass index may vary strongly on a daily basis or from one week to another, it seems more relevant from a policy point of view to be able to estimate the relationship between energy intakes and obesity on a long-run basis.

Section 2 presents the data and some descriptive statistics. Section 3 shows how to use household level data to obtain estimates of individual level consumption. Section 4 concludes and Section 5 contains an appendix.

## 2 Data and Descriptive Statistics

### 2.1 Data sources

Our dataset relies on different sources. First, we used home scan data from the TNS-WordPanel company, providing information on household purchases covering 354 product categories over two years (2001-2002) in France for more than 8,000 French households. As we are interested in the impact of food consumption on individual health indicators such as obesity, we also use household and individual characteristics (including anthropometric measures), available yearly between 2001 and 2003<sup>1</sup>. Concerning household purchases, we observe the quantity purchased, the price, as well as a large set of characteristics of purchased goods (identified by their bar code). Our data also provide a detailed set of demographic characteristics of the household for each year, such as the number of persons and the number of children, the household income, their employment category, their region of residence and type of residence (rental or owned), the town size, the qualifications and nationality of the person of reference. At the individual level, we have information on age and gender and also weight and height which allow us to compute the body mass index of each individual.

In addition to these home scan data of household food purchases, we collected nutritional information from different sources<sup>2</sup> for all the food products purchased by house-

---

<sup>1</sup>We removed 2,711 households that stopped participating in the survey at the end of 2001 and dropped 3,126 households because of missing information on age, gender, height, or weight for some individuals in the household. We finally used information of 4166 households on the two-year period that represent 11,237 individuals.

<sup>2</sup>The different sources that allow us to build the dataset are: the Regal Micro Table, Cohen and Sérog (2004), nutritional web sites ([www.i-dietetique.com](http://www.i-dietetique.com), [www.tabledescalories.com](http://www.tabledescalories.com) ...) and food industry companies web sites (Picard, Carrefour, Telemarket, Unifrais, Bridelice, Andros, Florette, Bonduelle,

holds and obtained the amount of energy in kilocalories (kcal) and in grams of proteins, fats and carbohydrates per 100g for each of the 2,073 products. This nutrient information depends on the product characteristics. For example, we are able to differentiate plain yoghurt according to their fat content. For snacks, we could go as far as brand level differentiation of nutritional content. Matching this information with household purchases, we obtain the total amount of nutrients or energy purchased.

Note that the available data on food consumption at the household level concern all food categories. These food items are classified into three categories corresponding to products with bar codes, to fruits and vegetables without bar codes and to meat and fish without bar codes. For each household, all food purchases are collected except those in one of two categories without bar codes (fruits and vegetables or meat and fish) and purchases of products with bar codes are always collected<sup>3</sup>. To overcome this problem of missing data for some food categories for some households, we implement a procedure of imputation at the household level which is detailed in Appendix 1. The method consists in using the full set of observed household characteristics to impute the unobserved value (quantity and expenditure of the unobserved food category) with the average value observed in households with the same set of characteristics. This matching procedure is reliable given that the missing category is unrelated to any systematic household consumption behavior and given the rich set of household characteristics that we observe. Moreover, on average it concerns a small percentage of household consumption. Indeed, only 9.7% of the total food consumption is estimated from the imputation method.

## 2.2 Descriptive Statistics on Obesity

We use the Body Mass Index (BMI) to define obesity. The BMI is a measure of corpulence status defined as the weight divided by the height squared (expressed in kilograms per square meter) and it is used by most nutritionists and epidemiologists. In accordance with the World Health Organization and other disease control and prevention institutions, adult

---

McCain, Nestlé, Avico).

<sup>3</sup>We observe food purchases for products with bar codes and meat and fish without bar codes for 60% of the sample, for products with bar codes and fruits and vegetables without bar codes for 38% of the sample and only for products with bar codes for 2% of the sample.

individuals are considered as overweight if their BMI is between 25 and 30, and obese if their BMI is over 30. For children, we use the definition from the international corpulence curves for boys and girls under 18 years old (Cole et al. 2000), which defines thresholds according to gender and age.

In Table 1, we can see that the average BMI is 23 (kg/m<sup>2</sup>) and that 9% of individuals in our survey can be considered obese. This percentage of obese people is consistent with the national figures in France obtained from other studies (Obépi-Roche, 2009). Our figures are also consistent with the national percentage of overweight people. Indeed, one third of adults are overweight, that is, more than 20 million people in France. Obesity is particularly prevalent among people over 60 years old since it represents 15% and 14% of the population of men and women respectively. While there is no great difference on average between males and females for obesity rates, the percentage of overweight is higher for adult men than adult women. For children, there seems to be no large differences between girls and boys.

Sample Means		N	BMI (std.dev.)	% Obese	% Overweight	
All		22469	23.04 (4.93)	8.71	25.59	
Adults	All	18311	24.39 (4.25)	10	29.12	
	Male	All	8401	24.86 (3.82)	9.53	35.25
		16-60 years old	6340	24.37 (3.81)	7.71	31.32
		more than 60	2061	26.35 (3.48)	15.14	47.31
	Female	All	9910	23.99 (4.54)	10.40	23.93
		16-60 years old	7217	23.40 (4.53)	9.15	19.26
more than 60		2693	25.55 (4.19)	13.78	36.43	
Children less than 16	All	4158	17.13 (2.96)	3.01	10.03	
	Male	2067	17.14 (2.99)	3.43	10.45	
	Female	2091	17.12 (2.93)	2.58	9.61	

**Table 1:** BMI, obesity and overweight

Thanks to the exhaustive data on weight and height of all household members in the survey, it is also possible to look at the within-household correlation of body mass indices. Actually, food consumption is largely a household activity and it is often argued that individual consumption is greatly influenced by the household. We therefore look at the within-household proximity of BMI status by defining for each individual his or her BMI deviation from the average BMI of individuals of the same age and gender. Looking at the

within-household standard deviation of these excess BMI (positive if above the mean and negative if below), we find that they are significantly positively correlated to the average BMI of the household, even after controlling for household size and demographics. There is apparently more heterogeneity of individual body mass indices within more "obese" households. To deal with obesity and food consumption issues, it would be better to look at food demands at the individual level rather than at the household level in order to take into account this household heterogeneity in terms of corpulence.

Moreover, looking at the probability that an individual is obese, it is highly positively correlated with the fact that one other individual in the household is obese, the marginal effect being of 7 percentage points, which means that the presence of at least one obese person among the other household members is associated with a 7 percent increase in probability that an individual is obese. Interestingly, the presence of obese individuals in the household is also associated with a 7 percentage points increase in the probability that an individual is overweight. On the contrary, the presence of at least one overweight individual is not correlated with the probability of an individual being overweight. While these correlations cannot be given any causality interpretation, they are significant of the sources heterogeneity of individuals and households BMIs.

### **3 From Household to Individual Consumption**

We first present an econometric model allowing an estimate of average individual level food intakes using household food purchases. We then apply this method to our data and present the empirical results in France.

#### **3.1 Method of Identification and Estimation**

Using the household measure of food consumption, we first present conditions under which "average" individual consumptions can be identified and estimated, where the "average" must be understood as the mean conditional on a set of individual characteristics. These conditions rely on conditional moments allowing us to identify the average (in the population) consumption of individuals with a given set of characteristics.



### Identification

Let us assume that for a person  $p$  in a household  $i$  at period  $t$ , the individual food consumption  $y_{ipt}$  is

$$y_{ipt} = \beta(x_{ipt}) + u_{ipt} \quad (1)$$

where  $x_{ipt}$  is a vector of individual characteristics of person  $p$  and  $u_{ipt}$  is a deviation for this person's consumption. The different measures of food consumption will consist first of calories and then proteins, fats and carbohydrates. Then, household consumption  $y_{it}$  is

$$y_{it} = \sum_{p=1}^{P(i)} y_{ipt} = \sum_{p=1}^{P(i)} \beta(x_{ipt}) + \varepsilon_{it} \quad (2)$$

where  $\varepsilon_{it} = \sum_{p=1}^{P(i)} u_{ipt}$  and  $P(i)$  is the number of individuals in the household  $i$ .

Assuming that  $\forall p, i, t$

$$E(u_{ipt} | x_{i1t}, \dots, x_{iP(i)t}) = 0 \quad (3)$$

implies that

$$E(\varepsilon_{it} | x_{i1t}, \dots, x_{iP(i)t}) = 0$$

which allows us to identify  $\beta$  non parametrically.

The assumption (3) implies that in equation (1)  $\beta(x_{ipt})$  can be interpreted as the average consumption by individuals whose characteristics are equal to  $x_{ipt}$  and  $u_{ipt}$  is interpreted as the deviation from the mean of individual caloric intake of this person. This assumption (3) also implies that the function  $\beta(\cdot)$  is overidentified by the natural additive structure between individual consumptions imposed on total household consumption:  $E(y_{it} | x_{i1t}, \dots, x_{iP(i)t}) = \sum_{p=1}^{P(i)} \beta(x_{ipt})$ . Denoting  $\mathbf{x}_{it} = (x_{i1t}, \dots, x_{iP(i)t})$ , we can test whether  $E(y_{it} | \mathbf{x}_{it})$  is separable across different individuals' characteristics. This can be done after estimating non parametrically  $E(y_{it} | \mathbf{x}_{it})$  and testing that

$$\frac{\partial^2 E(y_{it} | x_{i1t}, \dots, x_{iP(i)t})}{\partial x_{irt} \partial x_{ist}} = 0 \text{ for all } r \neq s \text{ from } \{1, \dots, P(i)\}.$$

However, estimating second derivatives of a non parametric conditional mean regression leads to very imprecise estimates and we are thus never able to reject the null hypothesis of separability only because of the large standard errors of the estimates of these derivatives. But this is clearly because this test has low power.

The separability assumption of the conditional mean of household consumption depends on the crucial choice of covariates  $\mathbf{x}$ . If one defines different covariates  $\mathbf{x}$  and  $\mathbf{z}$  for each individual, denoting  $\mathbf{x}_{it} = (x_{i1t}, \dots, x_{iP(i)t})$  and  $\mathbf{z}_{it} = (z_{i1t}, \dots, z_{iP(i)t})$ , the assumptions

$$(A) \quad E(y_{it}|\mathbf{x}_{it}) = \sum_{p=1}^{P(i)} \beta(x_{ipt})$$

and

$$(B) \quad E(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}) = \sum_{p=1}^{P(i)} \delta(x_{ipt}, z_{ipt})$$

are not equivalent and neither is more general than the other.

Actually, it could be that (B) is true but not (A) or the contrary. Firstly, by the law of iterated expectations, assumption (B) implies that  $E(y_{it}|\mathbf{x}_{it}) = \sum_{p=1}^{P(i)} E(\delta(x_{ipt}, \mathbf{z}_{ipt})|\mathbf{x}_{it})$  which is not necessarily separable between any  $x_{ipt}$  and  $x_{ip't}$ , for example if some  $z_{ipt}$  is correlated with  $x_{ip't}$  given  $x_{ipt}$ , which shows that (A) is not true in this case. Secondly, (A) may be true and not (B). For example if (B) is not true because  $E(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}) = \sum_{p=1}^{P(i)} \phi(x_{ipt}, z_{i1t})$  where  $\frac{\partial}{\partial z_{i1t}} \phi(x_{ipt}, z_{i1t}) \neq 0$ . (A) will nevertheless be true if  $z_{i1t}$  is independent of  $x_{i1t}$  because then  $E(\phi(x_{ipt}, z_{ipt}, z_{i1t})|\mathbf{x}_{it}) = \beta(x_{ipt})$ . It is thus important to choose carefully the set of characteristics  $\mathbf{x}_{it}$  such that the separability assumption is satisfied.

Chesher (1998) introduced this approach with UK data using gender and age for the individual characteristics  $x_{ipt}$  that are typically observed in the demographic composition of households. De Agostini (2005) and Miquel and Laisney (2001) applied this methodology to data sets from Italy and the Czech Republic. Allais and Tressou (2009) applied it to break down the consumption of seafood across individuals in France to determine their exposure to methylmercury, which involves health risks. In these studies, in order to obtain consistent estimates of average individual consumptions for a person of a given age and gender, the authors need to assume that household-level deviations  $\varepsilon_{it}$  (which are the sum of individual-level deviations  $u_{ipt}$ ) are not correlated with the demographic composition of the household in terms of age and gender. In particular, if the obesity status of individuals is correlated with age and gender and also with individual food consumption (which is intuitively likely to be the case), then biased estimates of individual

consumption by age and gender will be obtained. For example if in couples, the man's BMI is correlated with the woman's age given the man's age, then it is likely that the separability assumption of the household conditional mean consumption will not be true when using only gender and age as covariates. Looking directly at these correlations can thus provide an indication on whether conditioning on gender and age will give consistent estimates of average individual consumptions. Taking the example of couples, regressing the BMI of the man on the age of the woman controlling for the man's age with year dummies for his own age, we find significant correlations between the man's BMI and the woman's age, indicating that conditioning on gender and age of both partners will not provide consistent estimates. We will thus prefer to use age ( $x_{ipt}^1$ ), gender ( $x_{ipt}^2$ ) and BMI ( $z_{ipt}$ ). Moreover, conditioning on age and gender only does not allow us to distinguish the consumption of two individuals of the same age and gender but with different anthropometric measures. It is also likely that a lot of heterogeneity in individual consumption is related to the obesity status.

One way to get more precise and relevant individual measures of consumption is thus to condition consumptions on a wider set of observed characteristics and in particular on an anthropometric measure  $z_{ipt}$  that will be the body mass index, such that individual and household consumptions are written as:

$$y_{ipt} = \beta (x_{ipt}^1, x_{ipt}^2, z_{ipt}) + u_{ipt}$$

and

$$y_{it} = \sum_{p=1}^{P(i)} \beta (x_{ipt}^1, x_{ipt}^2, z_{ipt}) + \varepsilon_{it}$$

where  $x_{ipt}^1$ ,  $x_{ipt}^2$  are respectively gender and age of individual  $p$  in household  $i$  at time  $t$  and  $z_{ipt}$  corresponds to the body mass index at the beginning of the period.

### *Specification*

As age and gender are discrete variables and BMI is continuous, after some specification search, we choose to specify the function  $\beta$  as follows:

$$\beta (x_{ipt}^1, x_{ipt}^2, z_{ipt}) = \sum_{a=1}^{100} \sum_{g=1}^2 1_{\{x_{ipt}^1=a, x_{ipt}^2=g\}} \beta_a^g \left[ \delta_0^g + \delta^g (x_{ipt}^1) \left( \frac{z_{ipt} - \bar{z}_{a,g}}{\sigma_{a,g}} \right) \right] \quad (4)$$

where  $\delta^g(x_{ipt}^1) = 1_{\{x_{ipt}^1 \leq 13\}}\delta_1^g + 1_{\{13 < x_{ipt}^1 < 20\}}\delta_2^g + 1_{\{x_{ipt}^1 \geq 20\}}\delta_3^g$ , and  $\bar{z}_{a,g}$  and  $\sigma_{a,g}$  are respectively the mean and the standard deviation of the body mass index for individuals of age  $a$  and gender  $g$  (100 years is the maximum age in the population)<sup>4</sup>. With this specification, the continuous part of the function  $\beta$  in BMI  $z$  is supposed to be an age and gender specific linear function of the standardized BMI by gender and age (in years).

### *Estimation with smoothing*

Concerning the estimation of the model parameters, we can obtain consistent estimates by ordinary least squares using the specification (4). However, because of the discrete observation of ages, we introduce a smoothing technique (Chesher, 1998) penalizing the non smoothness of the function  $\beta(x_{ipt}^1, x_{ipt}^2, z_{ipt})$  with respect to the age variable  $x_{ipt}^1$ . The method amounts to estimating  $\beta$  (the vector of  $\beta_a^g$  for  $a = 1, \dots, 100$  and  $g = 1, 2$ ) of (4) as follows

$$\hat{\beta} = (x'x + \lambda^2 W'W)^{-1} x'y$$

where  $\lambda$  is a penalization parameter,  $W = I_2 \otimes A$ ,  $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and the matrix  $A$  of size

$$98 \times 100 \text{ is } A = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}.$$

### *Measurement errors*

Although these home scan data provide precise and reliable information on all food purchases, measurement errors due to the lack of observation of wasted food or of food consumed by people outside the household must be considered. Assuming that all these errors, denoted  $\varsigma_{it}$ , are uncorrelated with individual characteristics  $x_{ipt}^1, x_{ipt}^2, z_{ipt}$  of household members, the same method of estimation can be applied using the observed household purchases  $\tilde{y}_{it} = y_{it} + \varsigma_{it}$  instead of  $y_{it}$  to obtain consistent estimates of  $\beta_a^g, \delta_0^g, \delta_1^g, \delta_2^g$  and  $\delta_3^g$ .

Note that, although for single households, the household consumption  $\tilde{y}_{it}$  is a consistent

---

<sup>4</sup>Note that a constant term  $\beta_0$  can be added to the previous specification and interpreted as a waste. Empirically we will prefer the previous specification except for the disaggregation of lipids where it matters for fats used for cooking. The constant is then associated with the presence of a woman in a household.

measure of the individual consumption ( $E(\tilde{y}_{it}|P(i) = 1) = \beta(x_{ipt}^1, x_{ipt}^2, z_{ipt})$ ), it is not necessarily more precise than the estimated  $\hat{\beta}(x_{ipt}^1, x_{ipt}^2, z_{ipt})$  because of the measurement error  $\varsigma_{it}$ . Actually, assuming that measurement errors are independent of household size (for simplicity but the same result is also obtainable in general), the variances of each estimator are  $V(\tilde{y}_{it}|x_{ipt}^1, x_{ipt}^2, z_{ipt}, P(i) = 1) = V(\varsigma_{it}|x_{ipt}^1, x_{ipt}^2, z_{ipt}, P(i) = 1) = V(\varsigma_{it})$  and  $V(\hat{\beta}(x_{ipt}^1, x_{ipt}^2, z_{ipt})|P(i) = 1) = \frac{V(\varepsilon_{it} + \varsigma_{it})}{\text{card}\{i | x_{ipt}^1 = a, x_{ipt}^2 = g, z_{ipt} = c\}}$ . The second will in general be lower than  $V(\varsigma_{it})$  if the number of observations such that  $x_{ipt}^1 = a$ ,  $x_{ipt}^2 = g$  and  $z_{ipt} = c$  is large enough<sup>5</sup>.

### 3.2 Empirical Results

We apply the previous method to the total energy purchased by the household over a year which we have been able to construct thanks to the data on all food purchases matched with the collected nutritional information. We then apply it to measures of nutrient purchase such as proteins, fats, and carbohydrates.

As the results of the estimation correspond to average yearly food consumption at home, we re-scale the estimated individual food consumption to obtain an average daily food intake. We do this by using information from the 2004 representative survey on the average number of meals away from home by household members because this information is not available in 2002 and 2003. In order to take into account the gender and age pattern of meals away from home, we apply the same disaggregation method on the number of meals out of home to obtain an average per individual conditional on his gender and age, and total household consumption.

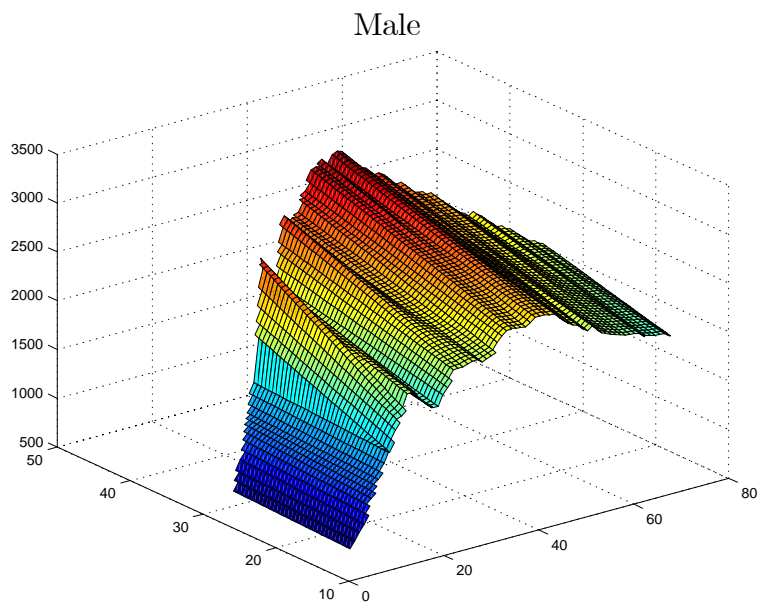
Figures 1 and 2 present the graphs of the estimated function  $\beta(\cdot)$  (with penalization parameter<sup>6</sup>  $\lambda = 500$ ). These graphs show that the individual energy consumption depends on the body mass index of individuals and increases with BMI but with different slopes according to the age of the individual. The slope seems almost zero for young boys but

---

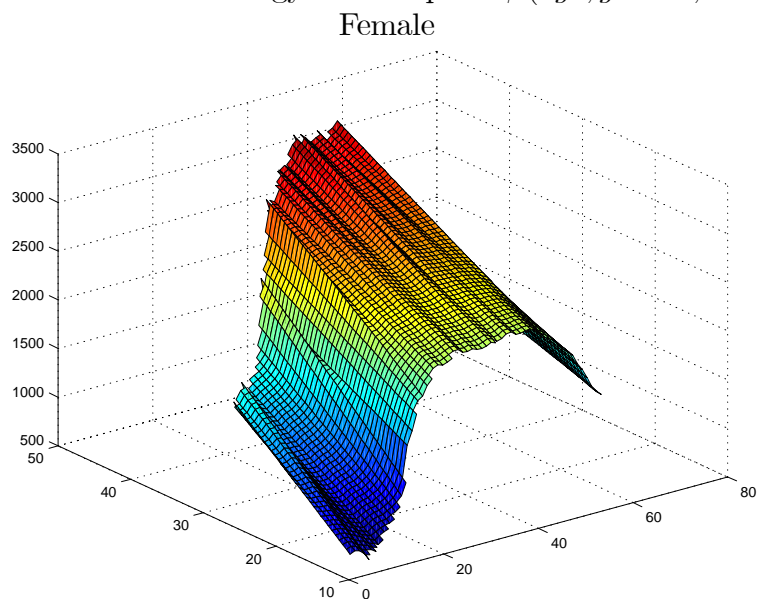
<sup>5</sup>Note that in our data the average size of these classes of individuals with the same age, gender and obesity status is 10. Moreover, only 3.4% of the individuals belong to a class with no other individuals.

<sup>6</sup>Minimizing the sum of squared differences between the estimated consumption and the observed consumption for household with one individual, we find that the optimal penalization parameter is  $\lambda = 500$ .

larger and positive for adult men. For women, it seems that the caloric intake increases more clearly with BMI at all ages.



**Figure 1:** Male Individual Energy Consumption  $\beta(\text{age}, \text{gender}, \text{BMI})$  (kcal/day)



**Figure 2:** Female Individual Energy Consumption  $\beta(\text{age}, \text{gender}, \text{BMI})$  (kcal/day)

Looking at the projection on the age hyperplane of the function  $\beta$  is also interesting to examine the age profile of individual energy consumption. Figure 3 presents such projections for three chosen categories of individuals defined as obese, overweight, and underweight or normal called "normal"<sup>7</sup>. These graphs show mainly that the energy intake increases until 18 years old for both girls and boys (with a stagnation for boys between 8 and 11). Then, the energy consumption decreases until 25 and increases again until 70

<sup>7</sup>In appendix, Figure 6 presents individual calories consumption with 95% confidence intervals.

for women and 55 for men. Chesher (1998) obtains the same shape of individual consumptions with respect to age with data from the UK. We also estimated these individual consumptions using the weight or the height instead of the BMI in the disaggregation method and we find the same shape of consumption with respect to age (see figures 7 and 8 in appendix). Finally, we can see that even if the age profiles of energy consumption have similar shapes across the three categories of individuals defined as "normal", overweight and obese, the levels of energy consumption are clearly higher for obese than for overweight and for overweight than for "normal". Moreover, it seems that overweight and obese people do consume more calories, especially during the periods of highest consumption. These curves also show that the differences between obese and non obese people are greater for women than for men. Assuming that obesity comes from an excess caloric intake compared to calories expended through physical activity, the difference in obesity among women thus seems to be even more strongly related to differences in caloric intakes than for men, perhaps because of fewer differences in physical activity among women than among men. Similarly, the striking feature that caloric intakes of obese, overweight and "normal" young boys does not seem to differ can be interpreted by the fact that physical activity (caloric expense) might be the sole source of variation that explains differences in BMI for these individuals.

Using the measures of proteins, carbohydrates and fats, we estimate the corresponding individual quantities of nutrients consumed. Figure 3 also presents the age profiles of these estimated individual consumptions for "normal", overweight and obese people. It is interesting to see that the graphs of "heavier" individuals (larger BMI) are always above those of leaner individuals, except for the consumption of carbohydrates by men where all three graphs are very close. These graphs also show that boys under 10 have very similar consumption patterns for energy and all nutrient consumptions whether obese or not, which is not true for girls. More obese girls do consume more than less obese girls and this is true for all nutrient measures. The differences between obese, overweight and "normal" people in terms of food intake is relatively the greatest for fats where, for

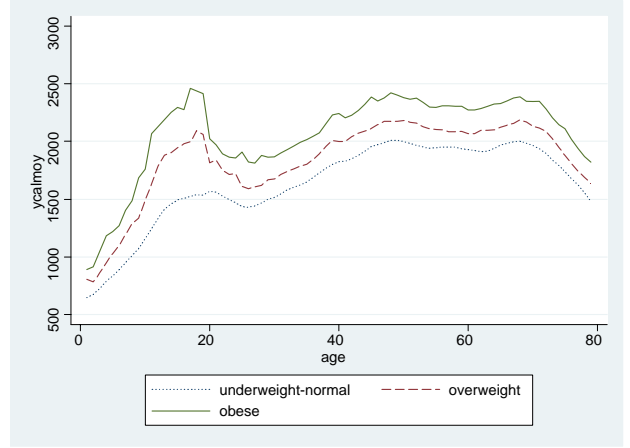
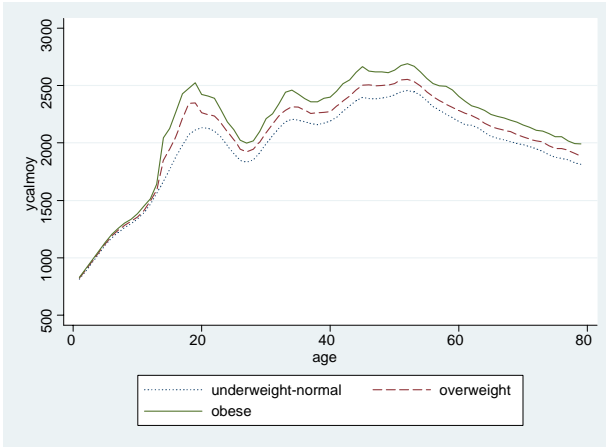
example, after 35 years old, the obese eat on average more than 20% more fat than "normal weight" individuals. They also eat more proteins and carbohydrates but the difference is relatively less great. Finally, looking at the shapes of the age profile of carbohydrate, protein and fat consumptions, we observe that the increase in consumption lasts until 18 years old for boys and until 14 years old for girls whatever the nutrient. Also, the decrease in consumption at old age appears to be quite the same whatever the nutrient considered. This is clearly after 60 years old for men and after 70 years old for women.



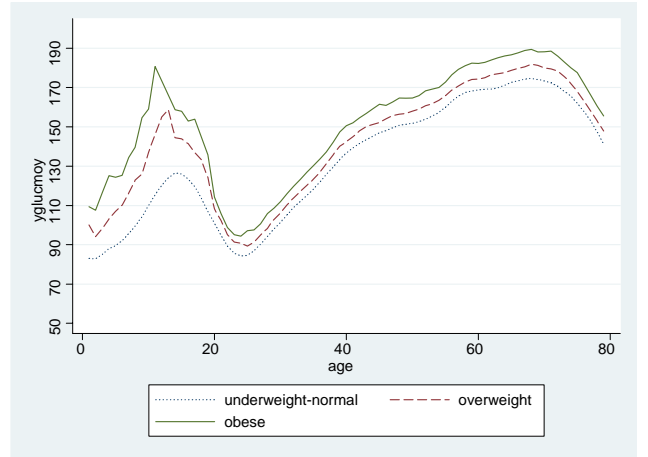
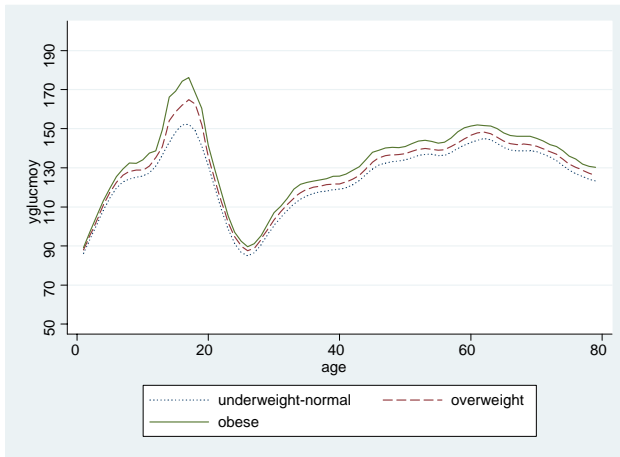
Males ( $\lambda = 500$ )

Females ( $\lambda = 500$ )

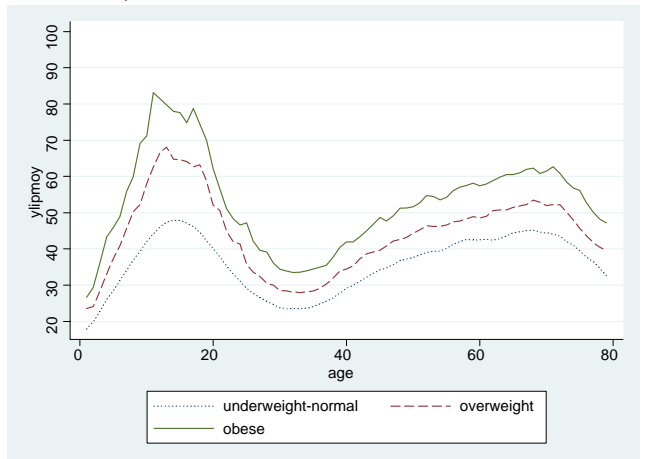
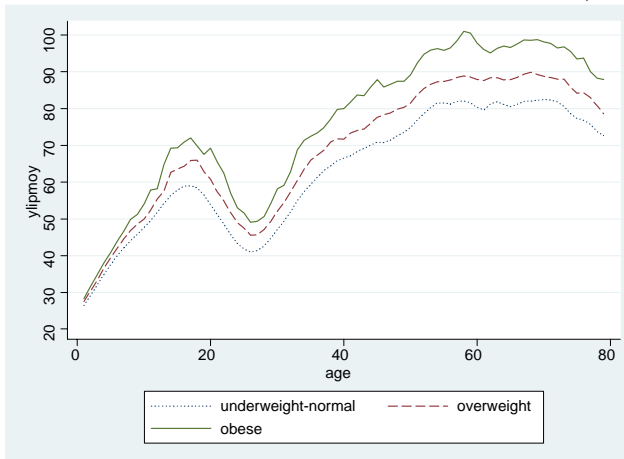
Energy (kcal per day)



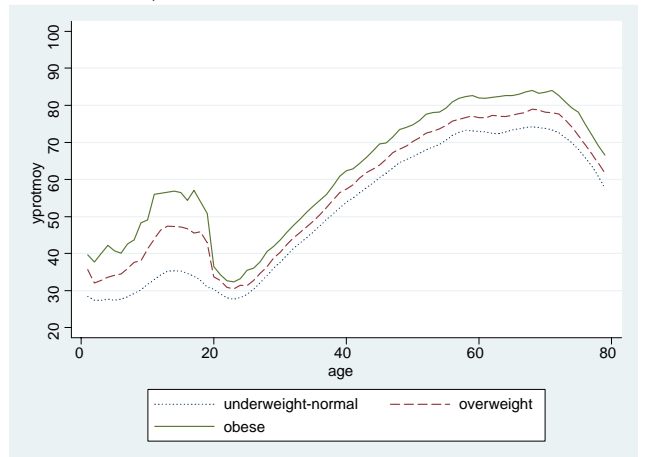
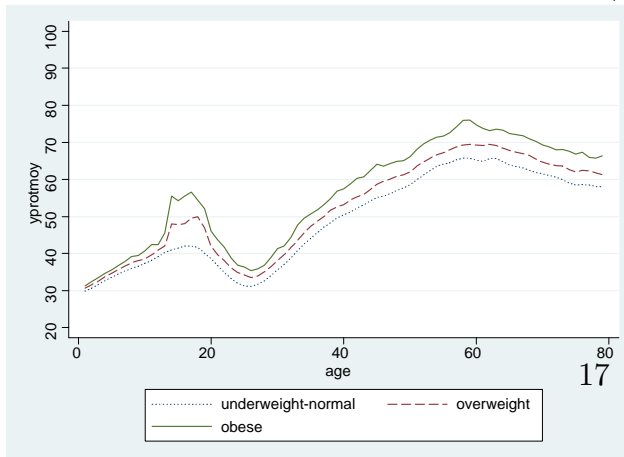
Carbohydrates (grams per day)



Fats (grams per day)



Proteins (grams per day)



**Figure 3:** Individual Consumption of Energy, Carbohydrates, Lipids and Proteins

Using the Atwater factors  $(\alpha, \beta, \gamma)$ , used by nutritionists to calculate the energy content of food (Nichols, 1994), the share of energy from proteins, fats or carbohydrates at the household and individual levels are presented in Tables 2 and 3. Table 2 shows that at the household level, proteins represent 15% of energy, fats 43% and carbohydrates 38%.

Share of energy from			
Proteins	Fats	Carbohydrates	Other
$(\alpha = 4)$	$(\beta = 9)$	$(\gamma = 4)$	
0.147	0.430	0.379	0.042
(0.024)	(0.050)	(0.060)	(0.034)

**Table 2:** Energy per household (N=4166)

Table 3 shows that, at the individual level, the share of energy from carbohydrates is a little lower for more obese people who consume more of their energy in fats, although the absolute quantities consumed are higher. Ransley et al. (2003) similarly found in the UK that overweight households purchase a larger share of their energy in fat than lean households. Concerning children, they consume more of their energy from carbohydrates and less from proteins. In terms of gender comparison, women consume more of their energy in carbohydrates and less in fats than men.

		Energy	Share of energy from			
			proteins	fats	carbohydrates	Other
All	(N=11237 indiv.)	1881	0.164	0.361	0.428	0.047
	(N=22469 obs)	(425)	(0.029)	(0.077)	(0.064)	(0.023)
By gender	Males	2035	0.145	0.426	0.378	0.051
	(N=5236)	(416)	(0.010)	(0.034)	(0.057)	(0.026)
	Females	1747	0.181	0.304	0.472	0.042
	(N=6001)	(386)	(0.029)	(0.057)	(0.028)	(0.019)
Class of BMI	Gender					
	Males	1917	0.141	0.410	0.398	0.051
Normal		(420)	(0.009)	(0.028)	(0.056)	(0.029)
-underweight	Females	1624	0.178	0.305	0.481	0.037
		(355)	(0.032)	(0.062)	(0.026)	(0.017)
	Males	2204	0.150	0.447	0.348	0.055
Overweight		(314)	(0.006)	(0.021)	(0.040)	(0.019)
	Females	1955	0.189	0.299	0.459	0.052
		(270)	(0.021)	(0.045)	(0.022)	(0.014)
	Males	2283	0.154	0.468	0.341	0.037
Obese		(390)	(0.005)	(0.029)	(0.045)	(0.020)
	Females	2194	0.189	0.310	0.440	0.061
		(296)	(0.013)	(0.032)	(0.020)	(0.014)
Age	Gender					
Adults	Males	2209	0.146	0.438	0.357	0.058
		(211)	(0.009)	(0.023)	(0.043)	(0.024)
	Females	1873	0.190	0.291	0.474	0.045
		(259)	(0.023)	(0.050)	(0.028)	(0.018)
Children	Males	1325	0.137	0.376	0.462	0.025
		(262)	(0.008)	(0.024)	(0.015)	(0.017)
	Females	1149	0.139	0.367	0.465	0.029
		(326)	(0.015)	(0.040)	(0.026)	(0.013)

Note:  $\alpha = 4, \beta = 9, \gamma = 4$

**Table 3:** Estimated energy per person and shares of nutrients

## 4 Conclusion

In this paper, we estimate average individual food consumptions, in particular the energy content of food products in kilocalories and the main macronutrients such as fats, proteins and carbohydrates, from household food consumptions. Doing so, we extend the disaggregation method of Chesher (1998) taking into account more heterogeneity in individual food consumptions than in usual studies, adding anthropometric measures like the Body Mass Index to age and gender.

We finally find that individual food consumption in terms of caloric intake clearly

increases with the BMI for men and women at all ages, suggesting that potential differences in physical activities do not reverse the relationship. An exception concerns the obesity and overweight of boys under 10 years old which seems not correlated with their total caloric intake, suggesting that differences in caloric expenditures may explain variations in BMI. We also find that overweight and obese people consume more fats than lean people in terms of share of their food caloric intake, obese people eating 20% more fat than normal weight individuals. We find that children have a higher consumption in carbohydrates than adults. However, their consumption of carbohydrates is not different across obesity status.

The age profile of carbohydrate, protein and fat consumptions shows that the consumption increase until adulthood lasts a few years more for fats and proteins than for carbohydrates. Also, the decrease in consumption at old age appears to begin first with the decrease in carbohydrates, then fats and last proteins.

These individual food consumptions estimates can provide some background information to adapt public policies aiming at reducing the prevalence of obesity and overweight by identifying the sources of overconsumption and the population at risk. These results suggest some possible targeting of prevention policies according to the gender and age of the population at risk, with policies that should insist on the reduction of total caloric intake in different proportions according to age and that could also target different categories of nutrients (carbohydrates, fats, proteins).

## 5 Appendix

### 5.1 Imputation methodology

To overcome the problem of missing data in one of the categories without bar codes, we implement a procedure of imputation at the household level.

Let  $y_{it}^k$  be the household consumption for category  $k = 1, 2, 3$  and let us define  $S_{it}^k \in \{0, 1\}$  equal to 1 only if  $y_{it}^k$  is observed. We also observe in the data a large set of variables  $W_{it}$  for household  $i$  at time  $t$  such that we define  $\omega_{it}^k$  as the unobserved effects

on household consumption of category  $k$  that make the household consumption different from the conditional average:  $\omega_{it}^k = y_{it}^k - E(y_{it}^k|W_{it})$ . We then assume that whether category  $k$  consumption is observed or not is independent of the unobserved variable  $\omega_{it}^k$  given all the observed covariates  $W_{it}$

$$\omega_{it}^k \perp S_{it}^k | W_{it} \quad (5)$$

This independence implies the mean independence of  $y_{it}^k$  given  $W_{it}$  with the observation of  $y_{it}^k$  :

$$E(y_{it}^k | W_{it}, S_{it}^k = 1) = E(y_{it}^k | W_{it}, S_{it}^k = 0)$$

This implies that the conditional mean of household food consumption of category  $k$  is the same on the sample for which it is observed and the one where it is not observed. Households with characteristics  $W_{it}$  will thus have the same average consumption of category  $k$  on the sample for observed consumption and the sample of unobserved consumption. Conditioning on a lot of observed characteristics  $W_{it}$  is likely to explain a lot of variations across households and thus provides a way to impute the consumption of unobserved food categories of some households with the observed consumption of "similar" households. However, taking into account household heterogeneity by conditioning on as many variables as possible will lead to a more difficult non parametric identification of the conditional means  $E(y_{it}^k | W_{it}, S_{it}^k = 1)$  due to the lack of sufficient observations given the large dimension of  $W_{it}$ .

But, one can also use the fact that (5) implies (Rosenbaum and Rubin, 1983)

$$\omega_{it}^k \perp S_{it}^k | P(S_{it}^k = 1 | W_{it})$$

and thus

$$E(y_{it}^k | P(S_{it}^k = 1 | W_{it}), S_{it}^k = 1) = E(y_{it}^k | P(S_{it}^k = 1 | W_{it}), S_{it}^k = 0)$$

where  $P(S_{it}^k = 1 | W_{it})$  is the propensity score of observing category  $k$ . One advantage of such an implication is that we can then condition on a unidimensional variable, the propensity score, and thus solve the dimensionality problem of conditioning on a large set of variables.

Moreover, we can apply this to any conditional mean for intervals of the propensity score, conditional on some  $W'_{it}$ , because we have

$$\begin{aligned} E(y_{it}^k | W'_{it}, S_{it}^k = 0, p \leq P(S_{it}^k = 1 | W_{it}) \leq p') &= \int_p^{p'} E(y_{it}^k | W'_{it}, S_{it}^k = 0, P(S_{it}^k = 1 | W_{it})) dF_{P(W_{it}) | W', S_{it}^k} \\ &= \int_p^{p'} E(y_{it}^k | W'_{it}, S_{it}^k = 1, P(S_{it}^k = 1 | W_{it})) dF_{P(W_{it}) | W', S_{it}^k} \end{aligned}$$

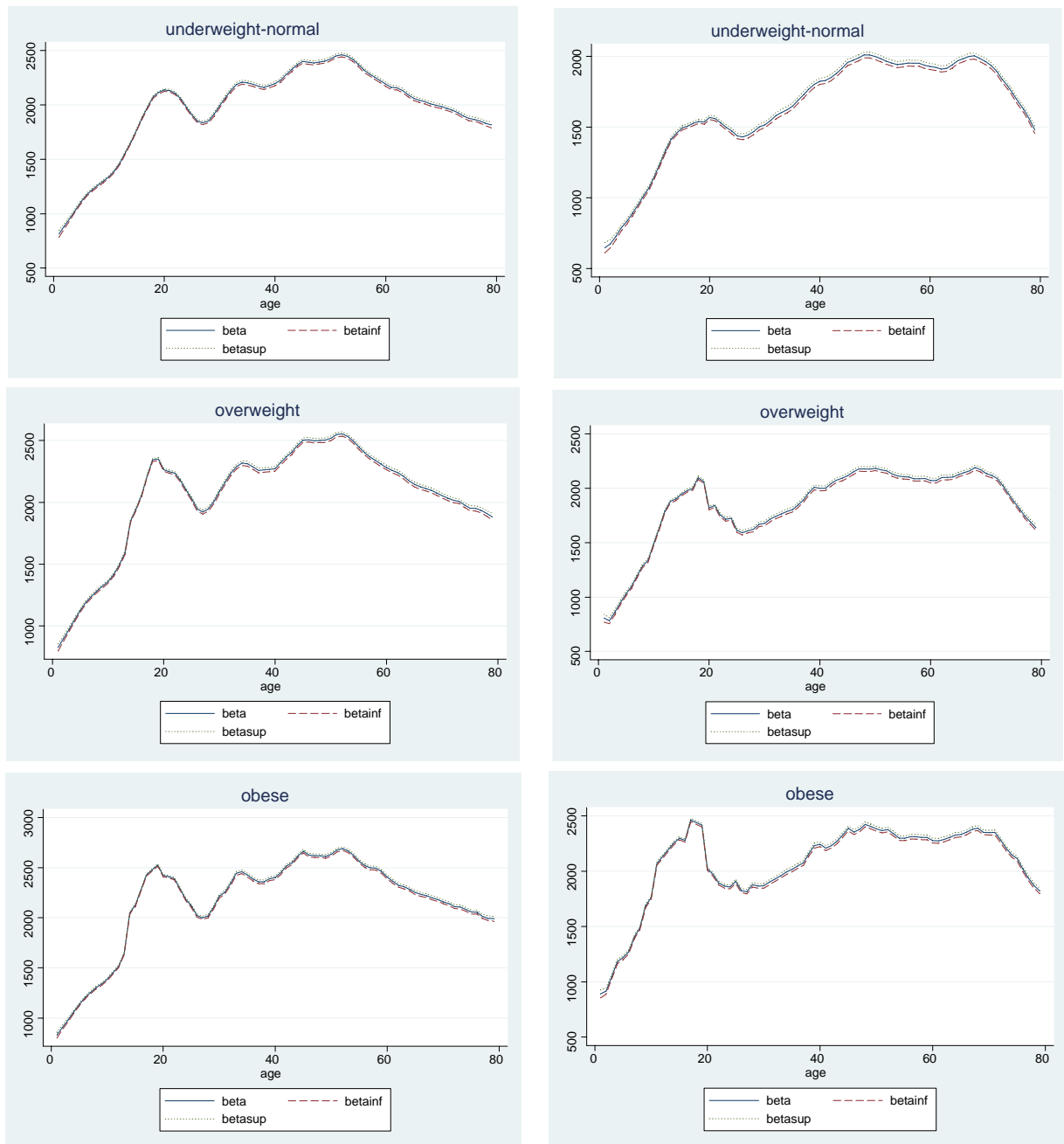
where  $F_{P(W_{it}) | W', S_{it}^k=0}$  denotes the conditional cumulative distribution function of the propensity score given  $W'$  and  $S_{it}^k = 0$ .

We thus first estimate the propensity score  $P(S_{it}^k = 1 | W_{it})$  and then impute the unobserved category  $k$  households consumption with propensity score  $p$  with the average observed household consumption of category  $k$  food products with the same propensity score.

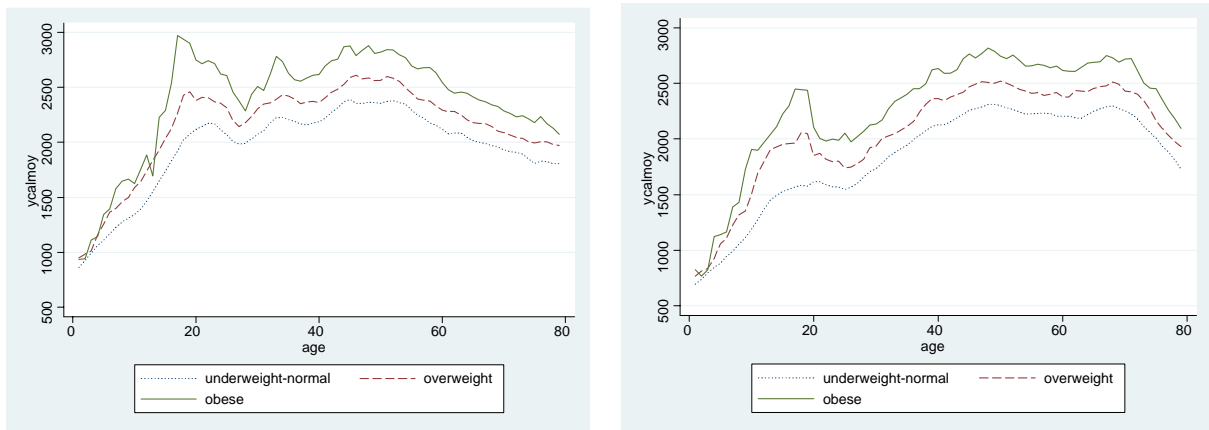
In practice, after some specification tests, the characteristics  $W_{it}$  include the declared household income, the household size, and the age class of the household head. The characteristics  $W_{it}$  consist of the gender and activity status of the individual making most food purchases in the household, indicators of socioeconomic class divided into 28 categories, indicators of the geographic region, 8 indicators of the level of qualification of the reference person, indicators of the citizenship of the reference person, the number of children under 16, the number of children under 6, 7 dummy variables for the type of housing, 8 dummy variables for urban, rural and municipality population size.

The regressions allowing us to estimate the propensity score matching for observation of the "fruits and vegetables" or "meat and fish" categories are not presented for brevity, but were done using a probit model.

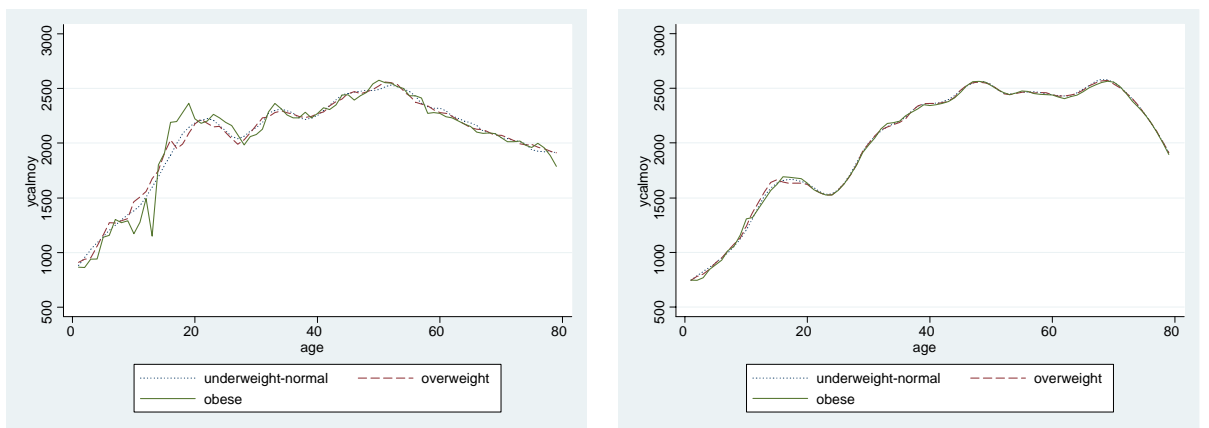
## 5.2 Other Tables and Graphs



**Figure 4:** Estimated individual calorie consumption by category per day



**Figure 5:** Age profile of estimated calorie consumption per day with weight for  $z_{ipt}$



**Figure 6:** Age profile of estimated calorie consumption per day with height for  $z_{ipt}$



## References

- Allais O. and J. Tressou (2009) "Using decomposed household food acquisitions as inputs of a Kinetic Dietary Exposure Model", *Statistical Modelling*, 9, 1, 27-50
- Chesher A. (1997) "Diet Revealed?: Semiparametric Estimation of Nutrient Intake - Age Relationships", *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 160, 3., 389-428
- Chesher A. (1998) "Individual demands from household aggregates: Time and Age variation in the composition of diet", *Journal of Applied Econometrics*, 13, 505-524
- Cohen J. M. and Sérog P. (2004) "Savoir manger : Le guide des aliments", Editions Flammarion
- Cole J.T., M.C. Bellizzi, K.M. Flegal and W.H. Dietz (2000) "Establishing a standard definition for child overweight and obesity worldwide: international survey", *British Medical Journal*, 320, 1-6
- Cutler D. M., Glaeser E., Shapiro J. M. (2003) "Why Have Americans Become More Obese?", *Journal of Economic Perspectives*, 17, 3, 93-118
- De Agostini P. (2005) "The relationship between food consumption and socioeconomic status: Evidence among British youths", ISEER working paper
- Emery C. , J. Dinet, A. Lafuma, C. Sermet, B. Khoshnood, F. Fagnani (2007) "Cost of obesity in France" *La Presse Médicale* 36, 6, 832-40.
- Grossman M. and I. Rashad (2004) "The Economics of Obesity", *The Public Interest*, 156, 104-112
- Lakdawalla D., Philipson T. (2009) "The growth of obesity and technological change: a theoretical and empirical examination", *Economics and Human Biology*, 7,3, 283-293
- Lakdawalla D., Philipson T., Bhattacharya J. (2005) "Welfare enhancing technological change and the growth of obesity", *American Economic Review*, 253-257.
- Levy E., Levy P., Le Pen C., Basdevant A. (1995) "The economic cost of obesity: the French situation" *International Journal of Obesity*, 19, 11, 788-792
- Miquel R. and F. Laisney (2001) "Consumption and Nutrition; Age-Intake Profile for

Czechoslovakia 1989-92", *Economics of Transition*, Vol. 9 (1) 115-151

Nichèle, V., Andrieu, E., Boizot, C., Caillavet, F. and Darmon, N. (2008) "L'évolution des achats alimentaires :30 ans d'enquête auprès des ménages en France, *Cahier de Nutrition et de Diététique*, 43, 3, 123-130

Nichols B. L. (1994) "Atwater and USDA Nutrition Research and Service: A Prologue of the Past Century" *Journal of Nutrition*, 124, 9, 1718-1727

Obépi-Roche (2009) "Enquête épidémiologique nationale sur le surpoids et l'obésité", INSERM / TNS HEALTHCARE SOFRES / ROCHE

Paraponaris A., B. Saliba and B. Ventelou (2005) "Obesity, weight status and employability: Empirical evidence from a French national survey", *Economics and Human Biology*, 3, 241-258

Philipson T. J. and R. A. Posner (2003) "The long-run growth in obesity as a function of technological change", *Perspectives in Biology and Medicine*, 46, 3, 87-108

Ransley, J.K. Donnelly, H. Botham, T.N. Khara, D.C. Greenwood, J.E. Cade (2003) "Use of supermarket receipts to estimate energy and fat content of food purchased by lean and overweight families" *Appetite* 41, 141-148

Wolf, A. and G. Colditz (2006) "Current Estimates of the Economic Cost of Obesity in the United States", *Obesity Research*, 6, 2, 97-106