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

Céline Bonnet†, Pierre Dubois‡, Valérie Orozco§

First Version: January 2007. This Version: January 2009

Abstract

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 caloric intakes according to the gender, age and the 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: H3, I18, D12

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

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

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

§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 of food consumption and its relationship with obesity across individuals and households in France. While nutritional surveys have generally the advantage of providing individual level measures of food consumption, they can usually be done on short periods only and are thus subject to the important daily variations of consumption which makes difficult the documentation of the relationship between the food diet and obesity, or casts doubts on the reliability of the estimation of such relationship in the long run. Thus, with a totally different source of data, 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 their gender, age and body mass index when these characteristics of household members are known. Using data from France, we apply this method 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 1990's. According to the 2003 Decennial Health Survey of INSEE (Paraponaris et al., 2005), the percentage of overweight has increased from 32.9% to 37.5% and from 6.3% to 9.9% for obesity between 1980 and 2003. The health problems related to obesity are consequently increasing. Actually, 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 expenditures 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 to a total of $99.2 billion. Using French data from the Decennial Health Survey, Paraponaris et al. (2005) show that overweight and obesity status reduce 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 the obesity would imply an additional cost in health expenditures in France between 2.1 and 6.2 billion € per year.

Among the causes of obesity epidemic, technological explanations based on the induced relative costs of food products and the surge of calories intake are the most important ones (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 of lipids intake less than 35% of the total food intake) that are not necessarily respected in 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.

As a related study, 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 days 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 don't observe the individual consumption on long periods of time. Actually, the usual nutritional survey provide precise information on food intakes and health outcomes at the individual level but usually on 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 strong approximation errors by respondents. As an example of such nutritional study, the French nutritional data INCA (Individual National Study of Food Consumption) collected by AFSSA (Agence Française de Sécurité Sanitaire des Aliments) consist in a survey of more than 4000 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 on 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 recall are potentially precise but on a 24 hour duration only and subject to mistakes for the consumption of household members who cannot respond themselves like children. The NHANES also collects information through 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 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 on 354 product categories over a period of two years that we matched with nutritional information of all products, we recover individual level estimates of nutrients consumption extending a method introduced by Chesher (1998). In particular, we are able to take into account the differences of Body Mass Index (BMI) into the individual food demands which is an important source of heterogeneity to explain 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 gender and ages and across nutrients such as carbohydrates, lipids or proteins.

It is to be noted that our data present the advantage of providing two years of food demand. Usual nutritional studies typically use a week (or less) of observation of food intakes or dietary history interview of daily intakes that are subject to lack of 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. Using food purchased on a long period seems thus an interesting way of measuring more accurately food demand.

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 an appendix appears in Section 5.

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 on 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. The final data consists of 4 166 households and 11 237 persons present on the period 2001-2003.[1] Concerning households 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 age and gender of all household members, the household income, their employment category, their region of residence and type of residence (location or property), the town size, the diploma and nationality of the person of reference. At the individual level, we have information on weight and height which allowe to compute the body mass index of each individual.

In addition to this home scan data of household food purchases, we collected nutritional information from different sources[2] for all the food products purchased by households and

[1] We removed 2711 households that stopped participating in the survey at the end of 2001 and dropped 3126 households because of missing information on either age, gender, height, or weight for some individuals of the household.

[2] 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, www.tabledescalories.com ....) and food industry

obtained the amount of energy in kilocalories (kcal), in grams of proteins, lipids and carbohydrates per 100g for each of the 2073 products. This nutrient information depends on the product characteristics. For example, we are able to differentiate plain yoghurts according to their fat content. For snacks, we could go until brand level differentiation of nutritional content. Matching this information with households purchases, we get the total amount of nutrients or energy purchased.

Remark 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 code, to fruits and vegetables without bar code and to meat and fish without bar code. For each household, all food purchases are collected except those of one of two categories without bar code (fruits and vegetables or meat and fish), purchases of products with bar code are always collected. 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 5.1. 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 on households with the same set of characteristics. This matching procedure seems quite reliable given that the missing category is supposed to be unrelated to any systematic household consumption behavior and given the rich set of household characteristics that we observe. Moreover, it concerns on average a small percentage of household consumption.

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 (in kilograms) divided by the height (in meter) squared and used by most nutritionists and epidemiologists. Following the World Health Organization and other disease control and prevention institutions, adult individuals are considered as overweight if their BMI is between 25 and 30, and obese if their BMI is over 30. For children, we consider the international definition from the international corpulence curves

companies web sites (Picard, Carrefour, Telemarket, Unifrais, Bridelice, Andros, Florette, Bonduelle, McCain, Nestlé, Avico).

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^2) 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, 2006). Our figures are also consistent with the national percentage of overweight people obtained from Obépi (2006). Indeed, one third of adults suffer from overweight, that is, more than 20 million people in France. Obesity is particularly important for the more than 60 years old since it represents 15% and 14% of the population of men and women respectively. While there is no strong differences on average between males and females for obesity rates, the percentage of overweight is higher for adult men than adult women. On the contrary, for children, there seems to be no strong difference 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	All		4158	17.13 (2.96)	3.01	10.03
less than 16		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

Looking at socioeconomic differences in the population, Table 2 shows that the obesity rate is not the same in the different professional categories for men and women. For instance, female farmers have a weak obesity rate of 4.3% whereas this rate is 10.5% for male farmers.The senior executive profesional class has also the smaller rates of obesity and overweight. The self employer have the highest rate, followed y blue collars. Although, obesity rates differ between males and females, the percentage of obese women by category is not that different from the rate of obese male, while the percentage of overweight women is systematically smaller than the percentage of overweight men.

Sample Means		N	BMI (std.dev.)	% Obese	% Overweight
Male	Farmers	143	25.56 (3.73)	10.49	40.56
	Self employed	207	25.76 (3.21)	13.04	41.55
	Senior executive	556	25.06 (3.30)	7.73	37.77
	Middle manager	1125	25.18 (3.36)	7.47	38.13
	White collar	1001	24.76 (3.62)	8.79	31.67
	Blue collar	1879	25.08 (3.85)	10.32	35.87
	Retired and without professional activity	3490	24.55 (4.08)	10.03	34.01
Female	Farmers	92	24.20 (2.85)	4.35	33.70
	Self employed	54	24.58 (4.81)	12.96	22.22
	Senior executive	243	22.38 (3.32)	3.29	12.76
	Middle manager	1094	23.35 (4.21)	7.40	18.46
	White collar	2957	24.00 (4.60)	11.16	22.02
	Blue collar	347	24.24 (4.46)	12.97	19.88
	Retired and without professional activity	5123	24.16 (4.63)	10.85	26.84

Table 2: Adults BMI and obesity by professional category

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. Thus, we look at the within household proximity of BMI status by defining for each individual its deviation of BMI 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 its 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 one 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 among other household members increases the probability that an individual is obese on average by 7 percent. Interestingly, the presence of obese individuals in the household also increases the probability that an individual be overweight by 7 percentage points. However, the presence of at least one overweight individual has almost no effect on the probability of an individual to be overweight because the marginal effect is around 1% and is barely significant at conventional statistical levels.

3 From Household to Individual Consumption

We first present an econometric model allowing to estimate average individual level food intakes using household food consumption. 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. These conditions rely on conditional moments allowing to identify the average (in the population) consumption of individuals with a given set of characteristics.

Identification

Let's assume that for a person p in a household i at period t, the individual food consumption y_{ipt} (the different measures of food consumption will consist first of calories but we will also consider proteins, lipids and carbohydrates) is

$$y_{ipt} = \beta\left(x_{ipt}\right) + u_{ipt} \tag{1}$$

where x_{ipt} is a vector of individual characteristics of person p and u_{ipt} is a deviation for this person's consumption. Then, the household consumption y_{it} is

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

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

$$E\left(u_{ipt} | x_{i1t}, .., x_{iP(i)t}\right) = 0 \tag{3}$$

implies that

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

which allows to identify β non parametrically.

The assumption (3) implies that in equation (1) $\beta\left(x_{ipt}\right)$ can be interpreted as the average consumption of 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.

Assumption (3) implies that the function $\beta(.)$ is overidentified by the natural additive structure between individual consumptions imposed on total household consumption: $E\left(y_{it}|x_{i1t},..,x_{iP(i)t}\right)=\sum_{p=1}^{P(i)}\beta\left(x_{ipt}\right)$. Denoting $x_{it}=\left(x_{i1t},..,x_{iP(i)t}\right)$, we can test whether $E\left(y_{it}|x_{it}\right)$ is separable across different individuals' characteristics. This can be done after estimating non parametrically $E\left(y_{it}|x_{it}\right)$ and testing that

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

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 such test has low power.

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

$$(A) \qquad E\left(y_{it}|x_{it}\right)=\textstyle\sum_{p=1}^{P(i)}\beta\left(x_{ipt}\right)$$

and

$$(B) \qquad E\left(y_{it}|x_{it},z_{it}\right)=\textstyle\sum_{p=1}^{P(i)}\delta\left(x_{ipt},z_{ipt}\right)$$

are not equivalent and none 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 $E\left(y_{it}|x_{it}\right)=\sum_{p=1}^{P(i)}E\left(\delta\left(x_{ipt},z_{ipt}\right)|x_{it}\right)$ 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) can be true and not (B). For example if (B) is not true because $E(y_{it}|x_{it}, z_{it}) = \sum_{p=1}^{P(i)} \phi(x_{ipt}, z_{i1t})$ where $\frac{\partial}{\partial z_{i1t}} \phi(x_{ipt}, z_{i1t}) \neq 0$. (A) will nevertheless be true if if z_{i1t} is independent of x_{i1t} because then $E(\phi(x_{ipt}, z_{ipt}, z_{i1t})|x_{it}) = \beta(x_{ipt})$.

It is thus important to choose carefully the set of characteristics x_{it} such that the separability assumption is satisfied.

Chesher (1998) introduced this approach 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 other data sets from Italy and the Czech Republic. Allais and Tressou (2007) applied it to decompose the consumption of seafood across individuals to determine their exposition to methylmercury which involves health risks. In these studies, in order to get consistent estimates of average individual consumptions for a person of a given age and gender, authors need to assume that household level deviations ε_{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 BMI of one member is correlated with the age of the other member given the gender and age characteristics of the first member, 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 some 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 individuals will not provide consistent estimates. We will thus prefer to use age (x^1_{ipt}), gender (x^2_{ipt}) and BMI (z_{ipt}). Moreover, conditioning on age and gender only, does not allow to distinguish the consumption of two individuals of the same age and gender but with different anthropometric measures. But, it is likely that a lot of heterogeneity still remains, in particular along the dimension related to the obesity status.

One way to get more precise and relevant individual measures of consumption is thus to condition consumptions on a larger 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:

$$y_{ipt} = \beta\left(x^1_{ipt}, x^2_{ipt}, z_{ipt}\right) + u_{ipt}$$

and

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

where x^1_{ipt}, x^2_{ipt} 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, we choose to specify the function β as follows

$$\beta\left(x^1_{ipt}, x^2_{ipt}, z_{ipt}\right) = \sum_{a=1}^{100}\sum_{g=1}^{2} 1_{\left\{x^1_{ipt}=a, x^2_{ipt}=g\right\}} \beta^g_a \left[\delta^g_0 + \delta^g\left(x^1_{ipt}\right)\left(\frac{z_{ipt} - \overline{z}_{a,g}}{\sigma_{a,g}}\right)\right] \tag{4}$$

where $\delta^g\left(x^1_{ipt}\right) = 1_{\left\{x^1_{ipt}\leq 13\right\}}\delta^g_1 + 1_{\left\{13<x^1_{ipt}<20\right\}}\delta^g_2 + 1_{\left\{x^1_{ipt}\geq 20\right\}}\delta^g_3$, and $\overline{z}_{a,g}$ 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)[3]. With this specification, the continuous part of the function β in z is supposed to be an age and gender specific linear function of the standardized z by gender and age (in years).

Estimation with smoothing

Concerning the estimation of the model parameters, we can obtain consistent estimates using ordinary least squares using the specification (4). However, because of the discrete

[3]Remark that a constant term β_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.

observation of ages, we introduce a smoothing technique (Chesher, 1998) penalizing the non smoothness of the function $\beta\left(x^1_{ipt}, x^2_{ipt}, z_{ipt}\right)$ with respect to the age variable x^1_{ipt}. The method amounts to estimate β (the vector of β^g_a for $a = 1, .., 100$ and $g = 1, 2$) of (4) as

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

with a penalization parameter λ and with $W = I_2 \otimes A$ where I_2 is the identity matrix of size 2 and $A = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & & & \vdots \\ \vdots & 0 & 1 & -2 & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}$ with size 98×100.

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 outside members of the household must be considered. Assuming that all these errors, denoted ς_{it}, are uncorrelated with individual characteristics x^1_{ipt}, x^2_{ipt}, z_{ipt} of household members, the same method of estimation can be applied using the household observed purchases $\widetilde{y}_{it} = y_{it} + \varsigma_{it}$ instead of y_{it} to obtain consistent estimates of β^g_a, δ^g_0, δ^g_1, δ^g_2 and δ^g_3.

Remark that, although for single households, the household consumption $\widetilde{y}_{it}$ is a consistent measure of the individual consumption $\left(E\left(\widetilde{y}_{it}|P(i) = 1\right) = \beta\left(x^1_{ipt}, x^2_{ipt}, z_{ipt}\right)\right)$, it is not necessarily more precise than the estimated $\widehat{\beta}\left(x^1_{ipt}, x^2_{ipt}, z_{ipt}\right)$ because of the measurement error ς_{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 $V\left(\widetilde{y}_{it}|x^1_{ipt}, x^2_{ipt}, z_{ipt}, P(i) = 1\right) = V\left(\varsigma_{it}|x^1_{ipt}, x^2_{ipt}, z_{ipt}, P(i) = 1\right) = V\left(\varsigma_{it}\right)$ and $V(\widehat{\beta}(x^1_{ipt}, x^2_{ipt}, z_{ipt})|P(i) = 1) = \frac{V(\varepsilon_{it}+\varsigma_{it})}{card\left\{i \;|x^1_{ipt}=a, x^2_{ipt}=g, z_{ipt}=c\right\}}$. The second will in general be lower than $V\left(\varsigma_{it}\right)$ if the number of observations such that $x^1_{ipt} = a$, $x^2_{ipt} = g$ and $z_{ipt} = c$ is large enough[4].

[4] 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.

3.2 Empirical Results

We apply the previous method to the total energy purchased by the household during a year that 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 nutrients purchase like proteins, lipids, 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 that by using information in the 2004 representative survey about the average number of meals taken out by household members because this information is not available in 2002 and 2003. In order to take into account the gender and age pattern differences in average meals taken out at the individual level, we apply the same disaggregation method on the number of meals taken out to obtain an average number of meals taken out per individual conditional on his gender and age, and controlling for total household consumption.

Figures 1 and 2 present the graphs of the estimated function $\beta(.)$ (with penalization parameter[5] $\lambda = 300$). These graphs show that the individual energy consumption depends on the body mass index of individuals and is increasing with BMI but with different slopes according to the age of the individual. The slope seems almost zero for young boys but larger and positive for adult men. For women, it seems that the slope of calories intake with respect to BMI is more clearly increasing with BMI at all ages.

[5] Minimizing the difference between the estimated consumption and the observed consumption for household with one individual, we find that the optimal penalization parameter is $\lambda = 700$. However this λ gave too flat estimates on the total sample and thus have chosen to decrease this parameter to 300.

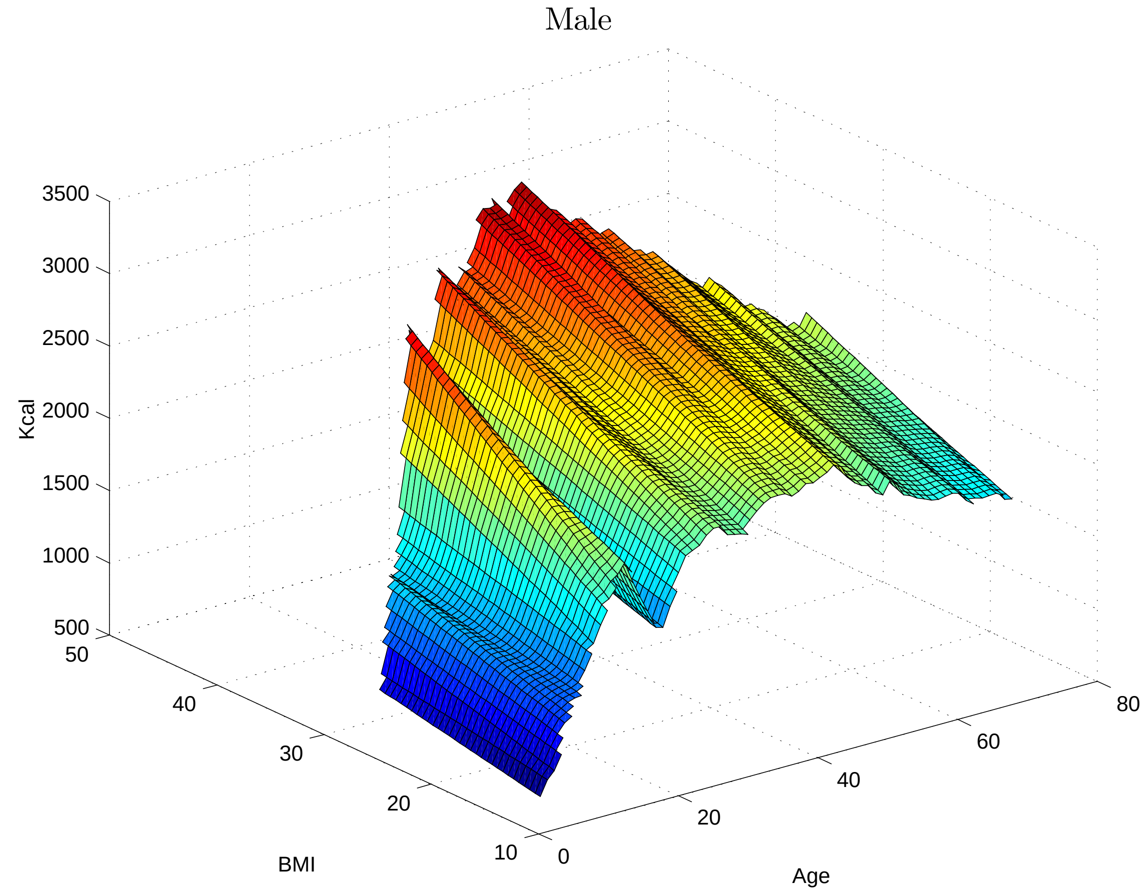


Figure 1: Male Individual Energy Consumption $\beta(age, gender, BMI)$ (kcal/day)

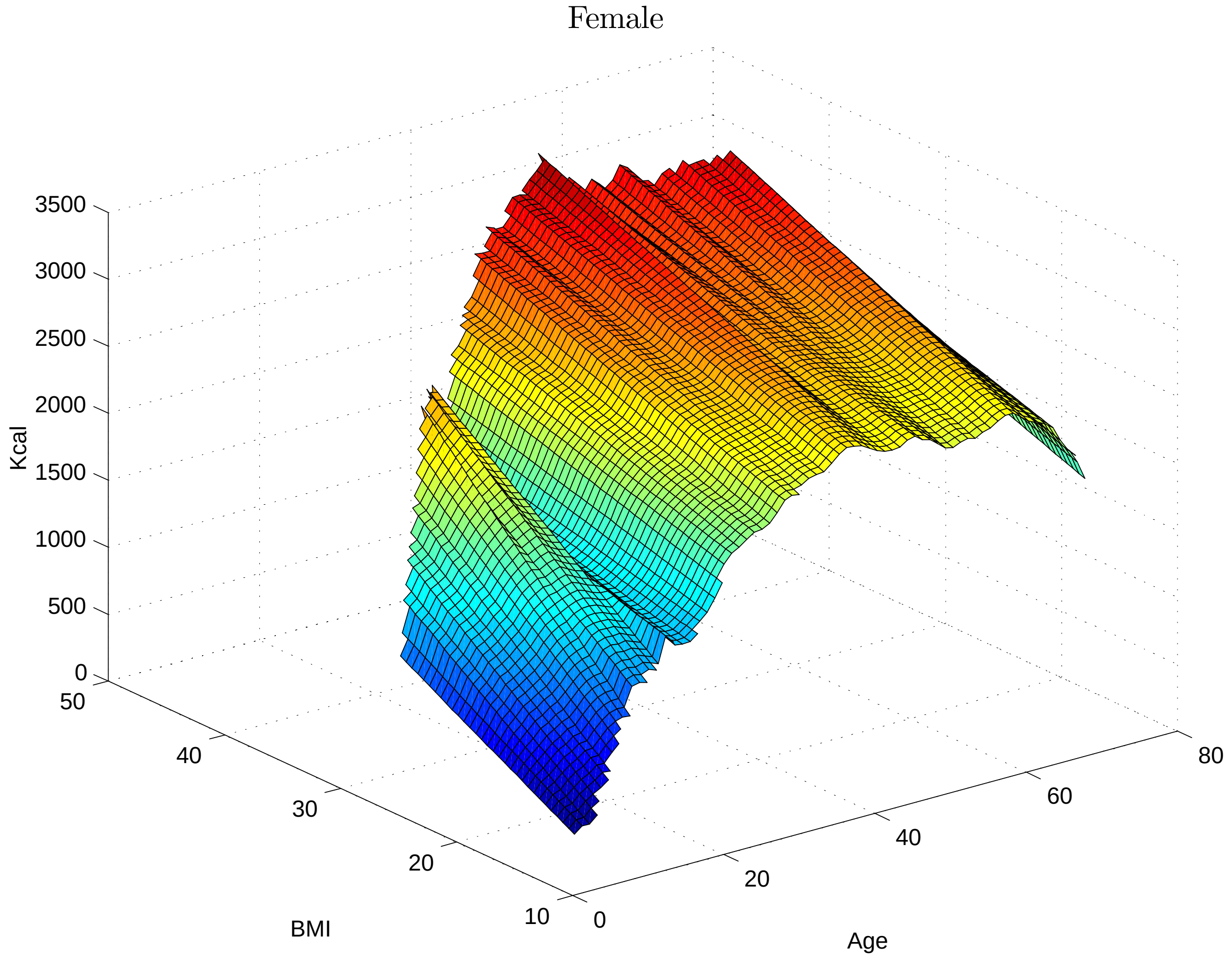


Figure 2: Female Individual Energy Consumption $\beta(age, gender, BMI)$ (kcal/day)

Looking at the projection on the age hyperplane of the function β 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[6]. These graphs show mainly that the energy intake is increasing 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 for women and 60 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 specially during the periods of life of highest consumption. These curves also show that the differences between obese and non obese people are higher for women than for men. Assuming that obesity comes from an excess caloric intake compared to calories expended through physical activity, the difference of obesity between women seems thus to be even more strongly related to differences in calories intakes than for men, perhaps because of less differences in physical activity across women than across 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 expenses) must be the unique source of variation that explains differences in BMI for these individuals.

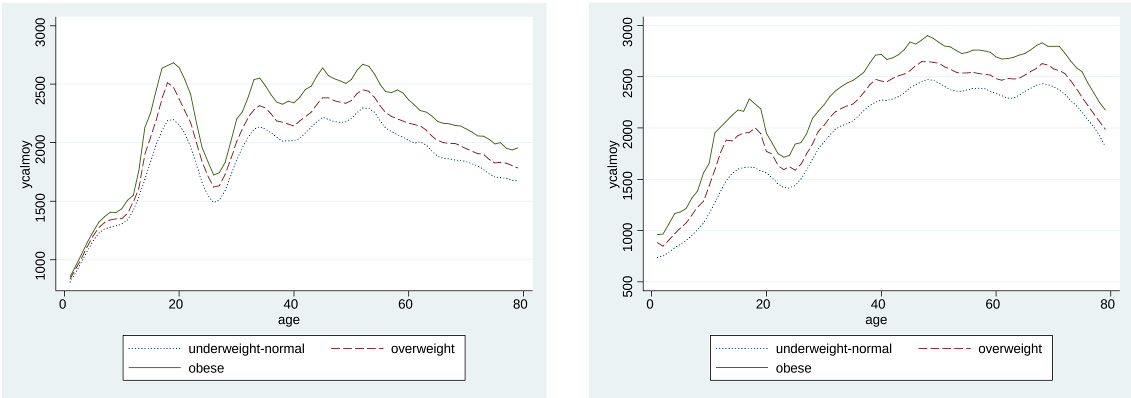
Using the measures of proteins, carbohydrates and lipids, we estimate the corresponding individual quantities of nutrients consumed. Figure 3 also presents the age profiles

[6] In appendix, Figure 6 presents individual calories consumption with 95% confidence intervals.

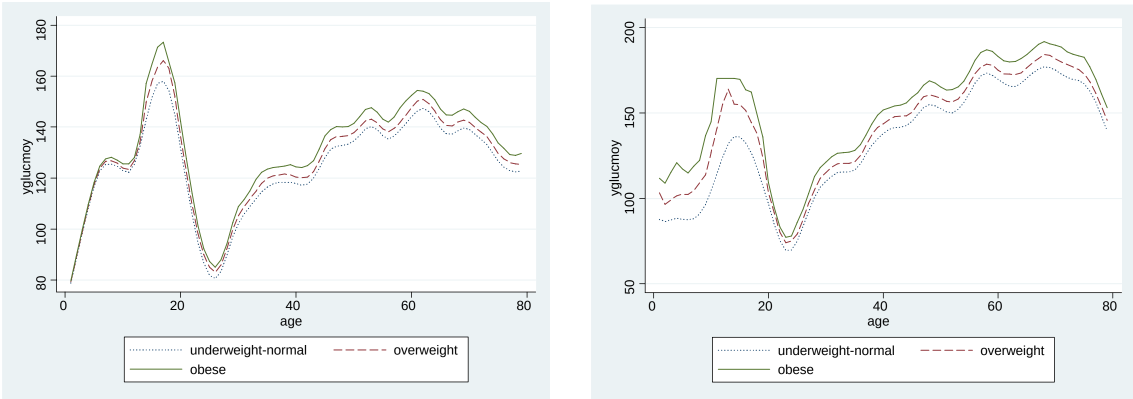
of these estimated individual consumptions for "normal", overweight and obese people. It is interesting to see that the graphs of "heavier" individuals (larger BMI) are always above those of leaner individuals, except for the consumption of carbohydrates by men until 30 where all three graphs are very 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 most important for lipids 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 important. Finally, looking at the shapes of the age profile of carbohydrates, proteins and lipids consumptions, we observe that the increase of consumption until adulthood lasts a few years more for lipids and proteins than for carbohydrates where the pick of consumption is around 18 years old instead of 20 years old for proteins and lipids. Also, the decrease of consumption at old ages appears to begin first with the decrease of carbohydrates, then of lipids and at last of proteins. This is clearly the case of men but much less for women whose decrease is less obvious until 70 years old.

Males ($\lambda = 300$)	Females ($\lambda = 300$)

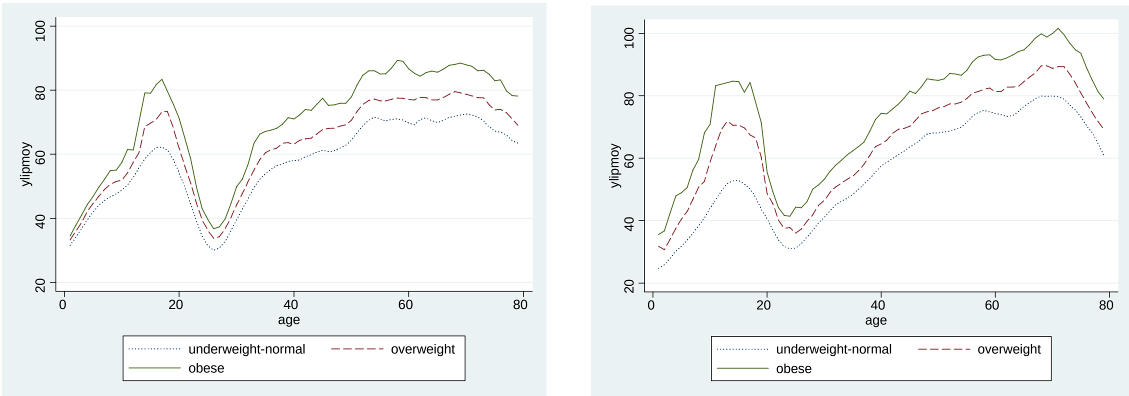
Energy (kcal per day)



Carbohydrates (grams per day)



Lipids (grams per day)



Proteins (grams per day)

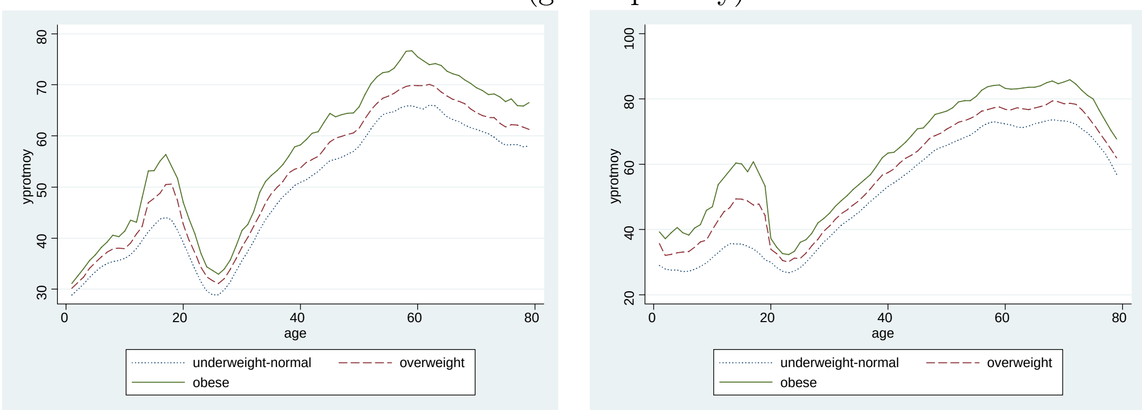


Figure 3: Individual Consumption of Energy, Carbohydrates, Lipids and Proteins

Using the Atwater factors (α, β, γ), known to calculate the energy content of food by nutritionists (Nichols, 1994), the share of energy from proteins, lipids or carbohydrates at the household and individual levels are presented in Tables 3 and 4. Table 3 shows that at the household level, proteins represent 15.9% of energy, lipids 39.7% and carbohydrates 39.6%.

Share of energy from			
Proteins	Lipids	Carbohydrates	Other
$(\alpha = 4)$	$(\beta = 9)$	$(\gamma = 4)$	
0.159 (0.029)	0.397 (0.061)	0.396 (0.068)	0.048 (0.041)

Table 3: Energy per household (N=4377)

Table 4 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 lipids, although the absolute quantities consumed are higher. Ransley et al. (2003) found similarly 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 lipids than men.

		Energy	Share of energy from			
			proteins	lipids	carbohydrates	Other
All	(N=11237 indiv.)	2003.47	0.153	0.402	0.400	0.045
	(N=22469 obs.)	475.03	(0.015)	(0.019)	(0.040)	(0.021)
By gender	Males	1950.34	0.151	0.409	0.391	0.050
	(N=5236)	384.65	(0.014)	(0.016)	(0.046)	(0.023)
	Females	2049.82	0.155	0.397	0.408	0.040
	(N=6001)	537.43	(0.015)	(0.019)	(0.031)	(0.018)
Class of BMI	Gender					
	Males	1832.62	0.147	0.402	0.412	0.040
Normal		376.75	(0.015)	(0.013)	(0.042)	(0.022)
-underweight	Females	1896.95	0.152	0.390	0.418	0.040
		514.48	(0.017)	(0.017)	(0.029)	(0.018)
	Males	2106.91	0.157	0.418	0.361	0.064
Overweight		293.125	(0.009)	(0.011)	(0.029)	(0.016)
	Females	2315.84	0.161	0.408	0.390	0.041
		389.116	(0.009)	(0.014)	(0.022)	(0.020)
	Males	2246.50	0.157	0.430	0.343	0.070
Obese		373.393	(0.008)	(0.014)	(0.032)	(0.017)
	Females	2594.78	0.163	0.419	0.374	0.043
		394.96	(0.006)	(0.014)	(0.022)	(0.016)
Age	Gender					
Adults	Males	2099.65	0.155	0.412	0.376	0.058
		232.61	(0.012)	(0.015)	(0.039)	(0.019)
	Females.	2231.22	0.159	0.398	0.400	0.043
		372.30	(0.012)	(0.015)	(0.026)	(0.016)
Children	Males	1343.49	0.134	0.397	0.450	0.018
		265.51	(0.006)	(0.015)	(0.016)	(0.007)
	Females	1190.09	0.133	0.389	0.448	0.030
		325.08	(0.012)	(0.033)	(0.022)	(0.025)
Note:	$\alpha = 4$, $\beta = 9$, $\gamma = 4$					

Table 4: Estimated energy per person and shares of nutrients

4 Conclusion

We have estimated individual food consumptions, in particular the energy content of food products in kilocalories and the main macronutrients such as lipids, proteins and carbohydrates, from household food consumptions. We have extended 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 find that the individual food consumption is clearly increasing with the BMI for men and women. Overweight and obese people consume more lipids than lean people. We also see that children have a higher consumption in carbohydrates than adults.

These individual food consumptions estimates will allow to adapt public policies aiming at reducing the prevalence of obesity and overweight by identifying the sources of overconsumption and population at risk.

5 Appendix

5.1 Imputation methodology

To overcome the problem of missing data in one of the categories without bar code, 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's 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 ω_{it}^k as the unobserved effects on household consumption of category k that make the household consumption different from the conditional average one: $\omega_{it}^k = y_{it}^k - E\left(y_{it}^k | W_{it}\right)$. We then assume that whether category k consumption is observed or not is independent on the unobserved variable ω_{it}^k given all the observed covariates W_{it}

$$\omega_{it}^k \perp S_{it}^k | W_{it} \tag{5}$$

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

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

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 one. Conditioning on a lot of observed characteristics W_{it} is likely to explain a lot of variation 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 the household heterogeneity by conditioning on as many variables as possible will lead to a non parametric identification of the conditional means $E\left(y_{it}^k|W_{it}, S_{it}^k = 1\right)$ more difficult due to the lack of sufficient observations given the large dimension of W_{it}.

There is a trade-off between matching on a lot of W for a better asymptotic precision and reducing the dimension of W for a better empirical finite sample precision given the sample size. However, reducing the dimensionality of the conditioning set may lead to inconsistencies in the estimates because denoting W_{it}' a sub-set of the vector W_{it}, the assumption $\omega_{it}^k \perp S_{it}^k|W_{it}$ does not imply that $\omega_{it}^k \perp S_{it}^k|W_{it}'$.

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

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

and thus

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

where $P\left(S_{it}^k = 1|W_{it}\right)$ is the propensity score of observing category k. One advantage of such 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 have

$$\begin{aligned} E\left(y_{it}^k|W_{it}', S_{it}^k = 0\right) &= E_{P(W)|W', S_{it}^k=0}\left[E\left(y_{it}^k|W_{it}', P\left(S_{it}^k = 1|W_{it}\right), S_{it}^k = 0\right)\right] \\ &= E_{P(W)|W', S_{it}^k=0}\left[E\left(y_{it}^k|W_{it}', P\left(S_{it}^k = 1|W_{it}\right), S_{it}^k = 1\right)\right] \\ &= \int E\left(y_{it}^k|W_{it}', P\left(S_{it}^k = 1|W_{it}\right), S_{it}^k = 1\right) dF_{P(W)|W', S_{it}^k=0} \end{aligned}$$

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

We also have that

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

We will prefer to use this method which includes the propensity score matching estimation to obtain consistent estimates by conditioning on a lot of observed characteristics.

We thus first estimate the propensity score $P\left(S_{it}^{k}=1|W_{it}\right)$ 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 general,

$$E\left(y_{it}^{k}|P\left(S_{it}^{k}=1|W_{it}\right),S_{it}^{k}=1\right)\neq E\left(y_{it}^{k}|W_{it},S_{it}^{k}=1\right)$$

and we always have that

$$var\left(y_{it}^{k}-E\left(y_{it}^{k}|P\left(S_{it}^{k}=1|W_{it}\right),S_{it}^{k}=1\right)|W_{it}\right)>var\left(y_{it}^{k}-E\left(y_{it}^{k}|W_{it},S_{it}^{k}=1\right)|W_{it}\right).$$

In practice, after some specification tests, the characteristics W_{it}' include the declared household income, the household size, and the household head age class. The characteristics W_{it} consist of the gender and activity status of the individual doing most food purchases in the household, indicators of socioeconomic class divided in 28 categories, indicators of the geographic region, 8 indicators of the level of diploma 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 to estimate the propensity score matching for observation of the categories "fruits and vegetables" or "meat and fish" are not presented for brevity but were done using a probit model.

5.2 Other Tables and Graphs

Males ($\lambda = 300$)	Females ($\lambda = 300$)

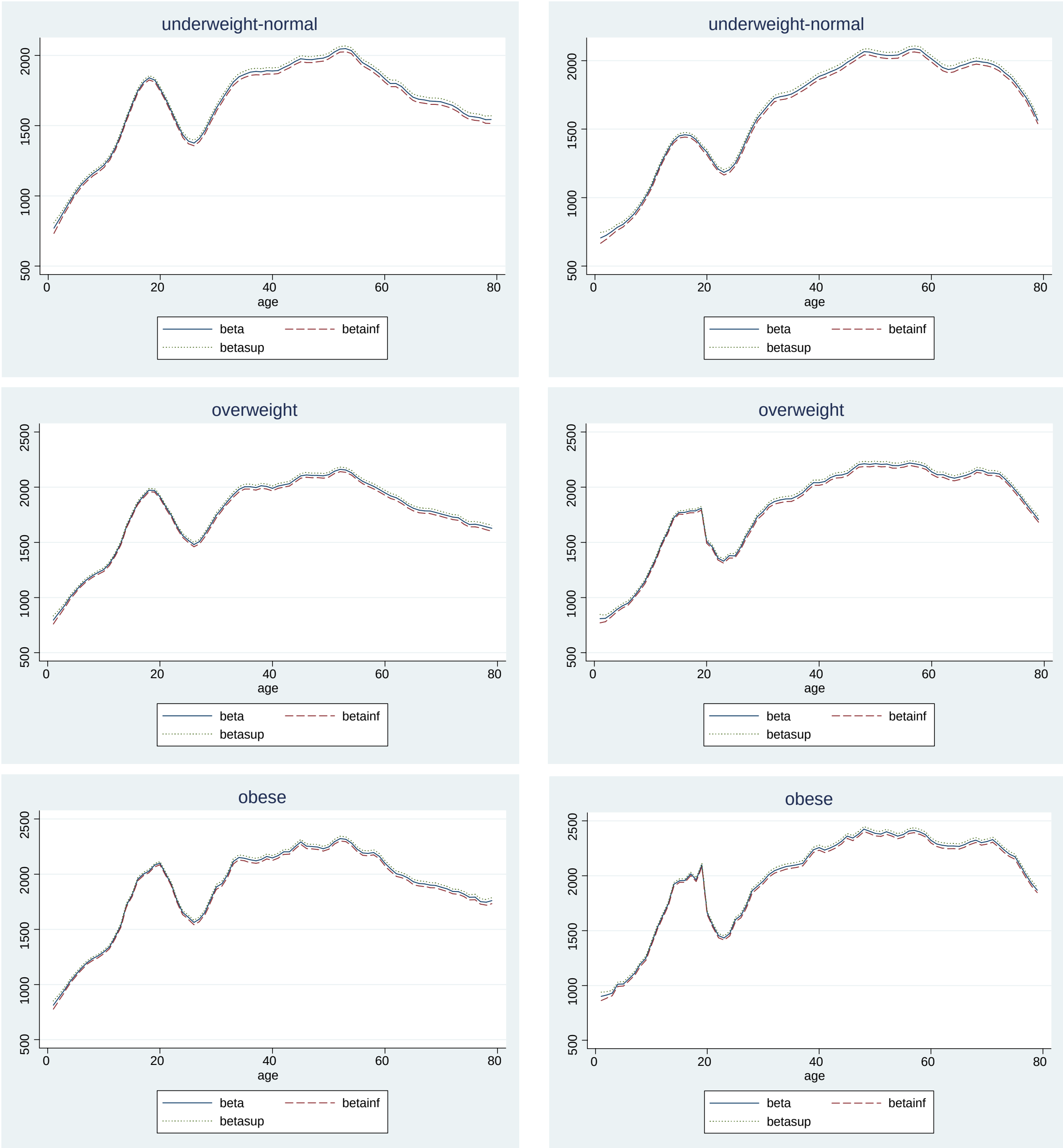


Figure 4: Estimated individual calories consumption by category per day

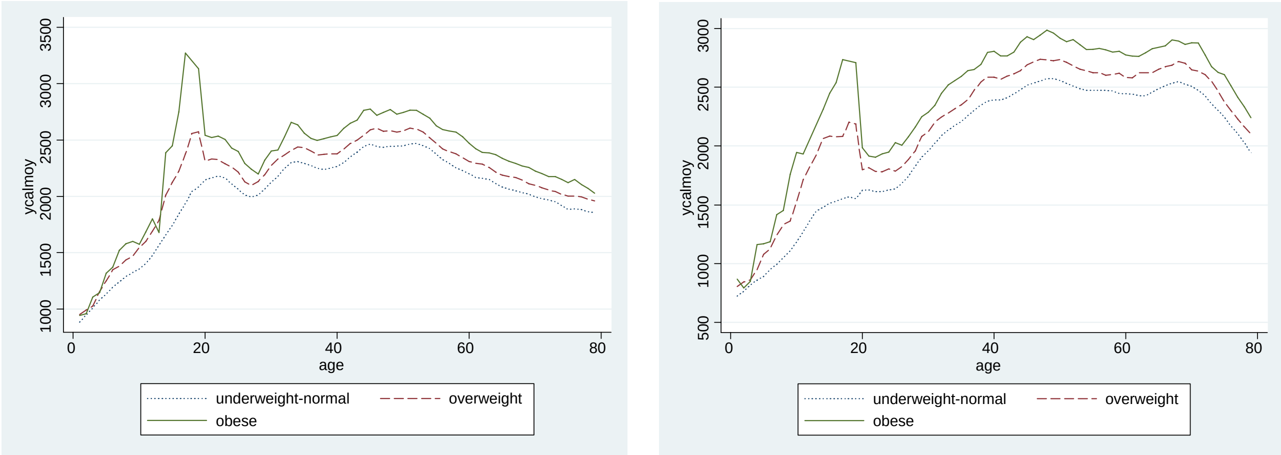


Figure 5: Age profile of estimated calories consumption per day with weight for z_{ipt}

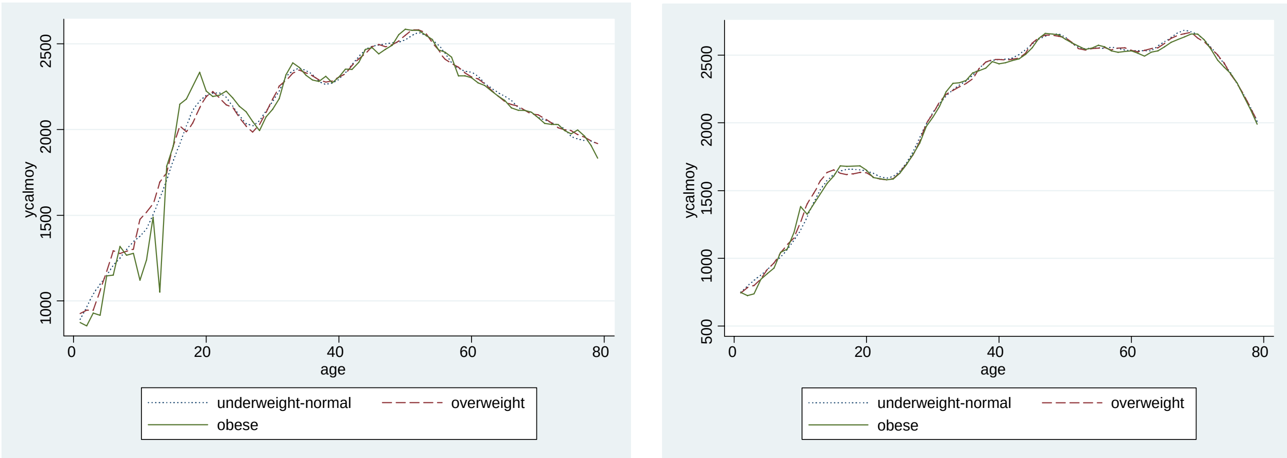


Figure 6: Age profile of estimated calories consumption per day with height for z_{ipt}

References

Allais O. and J. Tressou (2007) "Using decomposed household food acquisitions as inputs of a Kinetic Dietary Exposure Model", forthcoming *Statistical Modelling*

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", ISER 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. (2002) "The growth of obesity and technological change: a theoretical and empirical examination", *NBER* Working Paper 8946

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. (2005) "La consommation d'aliments et de nutriments en France: évolution 1969-2001 et déterminants socio-économiques des comportements", *mimeo*

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 (2006) "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 (1999) "The long-run growth in obesity as a function of technological change", *NBER* Working Paper 7423.

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