Food Consumption and Obesity in France: Identification of Causal Effects and Price Elasticities

Céline Bonnet*, Pierre Dubois†, Valérie Orozco‡

This Version: December 2009§

Abstract

The objective of this paper is to identify the causal effects between food consumption, calories intakes and body mass index (BMI) at the individual level. We then estimate the effects of food prices on consumption and analyze the relevance of policies of taxation for the reduction of obesity in the population. Decomposing the price elasticity of BMI into the causal relationship between obesity and calories consumption and the price elasticity of consumption, we can determine the impact of food consumption on BMI. Using French data recording household and individual characteristics, households food purchases (quantities and expenditures) over a period of two years and nutrition information of all products, we recover individual level estimates of calories consumption following Chescher (1998). This individual consumption allows to analyze food demands and its price elasticity as well as their relationships with obesity at the individual level, avoiding the household aggregation bias.

Key words: Body Mass Index, obesity, taxation, nutrients, individual food consumption
JEL codes: H3, I18, D12

*Toulouse School of Economics (GREMAQ, INRA), 21 Allée de Brienne, F-31000 Toulouse.
†Toulouse School of Economics (GREMAQ, INRA, IDEI), 21 Allée de Brienne, F-31000 Toulouse, corresponding author.
‡Toulouse School of Economics (GREMAQ, INRA), 21 Allée de Brienne, F-31000 Toulouse.
§We thank Namanjeet Ahluwalia, Philippe Bontems, Richard Blundell, Carol Propper, Nicole Darmon, Catherine Esnouf, Rachel Griffith, Atsushi Inoue, Thierry Magnac, Lars Nesheim, Vincent Réquillart, Tom Vukina, Wally Thurman for useful discussions and comments on a previous version, as well as seminar participants at Toulouse School of Economics, North Carolina State University, Public Health Seminar in Toulouse, Royal Economic Society Conference in Warwick, all remaining errors are ours.

1 Introduction

The objective of this paper is to study the relationship between obesity and food consumption across individuals in France, in order to provide recommendations for public policies aiming at reducing the prevalence of obesity.

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 or obese people in France has increased since 1980 from 32.9% to 37.5% for overweight and from 6.3% to 9.9% for obesity with less overweight for women but more obesity in 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 disease, diabetes, psychological disorders such as depression, and various types of cancer. In the US, the obesity costs more in annual 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 estimates 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.

The different explanations of obesity

The causes of the obesity epidemic have been studied recently, mostly in the US and several factors are shown to explain why obesity increased in developed economies. Technological explanations based on the induced relative costs of food products and food consumption are among the most important ones. Actually, the genetic component may play an important role in explaining why a given individual is obese. But genetic characteristics in the population change very slowly, and so they clearly cannot explain why obesity has increased so rapidly in recent decades. Economists have thus been proposed explanations of obesity looking at technological changes, changes in taste and consumer habits, and at changes in the social environment. According to Lakdawalla and Philipson (2002), declines in the real prices of grocery food items caused a surge in calories intake that can account for as much as 40 percent of the increase in the body mass index of adults since 1980. Technological advances in agriculture caused grocery prices to fall and these declines caused consumers to demand more groceries. Technological changes in the home kitchen seem to have also fostered more calories intake because of new tools responsible for reduction in the time spent preparing meals at home (Cutler, Glaeser, and Shapiro, 2003, Cutler and Glaeser 2005). Microwaveable meals and other foods that are easy to cook are desirable because they are quicker to prepare but they also have generally higher contents in calories and fat. Other factors have contributed to the growth in obesity like the decline in physical activity since 1980. Chou, Grossman, and Saffer (2004) find that the per capita number of fast-food and full-service restaurants, the prices of a meal in each type of restaurant, food consumed at home, cigarettes, and alcohol, and clean indoor air laws, explain a substantial amount of the trend of obesity in the US since the 70s. Classen and Hokayem (2005) show that demographic factors and mothers' obesity status and education also affect youth obesity (in the US). For example, obese mothers are at least 23% more likely to have an overweight youth than their peers with a Body Mass Index (BMI) in the acceptable range. This result can come from either genetic transmission of obesity or from a technological explanation due to the food production process within the household. Anderson, Butcher, and Levine (2003b) find that the rise in average hours worked by mothers can account for as much as one-third of the growth in obesity among children in certain families. In part, the rise in obesity seems to have been an unintended consequence of encouraging women to become more active in the workforce.

Policy responses

Overall, technological advances in agriculture caused grocery prices to fall and these declines seem to have caused consumers to demand more groceries. Government policy only heightened the effect by encouraging overproduction. Now governments think about taxation tools in order to go against these fall in prices for unhealthy food.

Taxing food in order to influence consumer behavior to improve public health and nutrition has been envisaged. Usually, such taxation schemes are first designed to collect taxes to increase government budgets eventually dedicated to health related prevention and are rarely designed and expected to change food behavior. According to Caraher and Cowburn (2005), the evidence about the impact of food taxes on consumer behavior is unclear. Moreover, many food taxes have been withdrawn after short periods of time due to industry lobbying. The common propositions are usually to combine taxes on unhealthy foods with subsidies of healthy foods. Moreover, this is often done in specific contexts and closed settings such as schools for example. Mytton et al. (2007) examine the effects on nutrition and health of extending value added tax (VAT) to a range of foods in the UK. They use consumption and elasticities estimated elsewhere from aggregate data to show that a carefully targeted fat tax can reduce cardiovascular diseases. Epstein et al. (2007) assess the influence of price changes of low-energy-density and high-energy-density foods on mother's food purchases in a laboratory food-purchasing experience. They show that increasing the price of high energy density foods would lead people to change their buying habits and switch to low-energy-density foods. Gandal and Shabelansky (2009) find that, after controlling for income levels and other factors, high "price-sensitivity" for food products is associated with high obesity rates. Abay (2006) finds that Egypt's food subsidy program, which reduces price of the dense caloric food, may be in part responsible for increased obesity for women with children.

Objective and contributions

Our objective is to study the price elasticity of Body Mass Index (BMI), determine some policy implications on taxation. For that, we break down the price elasticity of BMI into the causal relationship between obesity and calories consumption and the price elasticity of consumption. In this way, we can determine the impact of food consumption on BMI. It is important for the welfare of individuals to know whether an achievable reduction of BMI is due to more or less strong reduction of consumption and from which type of food items.

The problem in the study of the relationship between BMI and calories consumption is that in general we don't observe the individual consumption on long periods of time. Using French data recording household food purchases (quantities and expenditures) over a period of two years that we matched with nutrition information of all products, we recover individual level estimates of calories consumption conditional on some individual characteristics as in Chesher (1998). This individual consumption allows to analyze food demands and its price elasticity as well as their relationship with obesity at the individual level, taking into account unobserved heterogeneity of individuals who obviously can have very different energy requirements. Also, we are able to take into account the differences of BMI into the individual food demand which is an important source of heterogeneity to explain the link between food consumption and obesity. Once the food consumption of each member of households are recovered over a period of two years, we estimate the relationship between individual food consumption and body mass index. We find strong significant relationships between food calories and BMI, confirming the usefulness of policies aiming at reducing food calories intakes. We also take into account heterogeneity in terms of observable and unobservable characteristics. Actually, unobservable heterogeneity can be related to the metabolic requirements of a given individual and is likely to bias the estimates of the relationship between BMI and consumption. We then estimate price-elasticities of household food demand as well as individual demands that are necessary for any analysis of public policies aiming at reducing the prevalence of obesity or other nutritional problems among the population. We categorize food products in 8 categories and estimate group level price elasticities of calories. We find that the "junk food" category has a large price elasticity. This group of products of high density in energy represents a non negligible part of total energy intakes by individuals and thus price changes in this category can lead to significant changes in diets. Taxing such category can thus be an efficient recommendation without changing the availability of alternative cheap calories from lower density energy food items (such as pasta, rice).

Using the same source of data for household food purchases, Nichèle et al. (2005) study the evolution of nutrition in France between 1969 and 2001 but assume an equal division of food among household members. Boizot and Etilé (2005) study the relationship between BMI and the food prices of household purchases on French adults. They assume that adults are at their steady-state body mass index and look at the prices of the household purchases from 7 food groups. They study the relationship between local price indices of these food groups and Body Mass Indices of individuals in a cross-section. They find positive or negative correlations depending on the food groups showing that prices matter. However, without accounting for the energy content of food purchased, variations in prices of a category could lead to intra-group substitution of products of different energy content and thus bias the estimates of the price elasticity of BMI. Moreover, the heterogeneity of individuals in terms of metabolism can also generate spurious correlations between these prices and BMI. 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. On a 28 days basis, they find that overweight households purchase significantly more energy and fat per adult equivalent than lean households.

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 causal relationship of food intakes on the body mass index may differ strongly on a week compared to a longer period because of complex metabolic mechanisms, and because weekly or daily energy demand may not be time separable and involve much more unobservable state dependence, it seems much more relevant to be able to estimate the long run effect of energy intakes on obesity. 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 presents the causality analysis between food consumption and weight. Section 5 presents the estimates of consumption elasticities. Section 6 performs some policy simulations of taxation on the prevalence of obesity. Section 7 concludes and some appendix are in Section 8.

2 Data and Descriptive Statistics

2.1 Data sources

We use different sources of data. First, we used home scan household data from TNS World Panel, providing information on household purchases on 354 product categories over two years (2001-2002) for more than 8 000 French households. As we are interested in the impact of food consumption on health, we also use household and individual characteristics (including anthropometric measures) over the years 2001-2002-2003. The final data amount to 4 166 households and 11 187 persons present on the period 2001-2003.[1] Concerning purchases, we observe the quantity purchased, the price, as well as a large set of characteristics of purchases like the characteristics of all goods (identified by their bar code) and the identity of the retailer at which it has been purchased. Our data also provide a detailed set of demographic characteristics of the household such as the income, the number of persons and the number of children, their employment category, their region of residence and type of residence (location or property), the town size, the diploma of the person of reference, the nationality of the person of reference. At the individual level in the household, we have information on age, sex but also weight and height that allows to compute the body mass index of individuals every year.

We collected nutritional information from different sources[2] for all the food products purchased by households and matched the nutrient information depending on the product characteristics. The final nutrition data contain information about the amount of calories for 2073 products. Our detailed matching

[1] From the initial 10 003 households in the data, we dropped 2711 households who were not present in both 2002 and 2003, 457 households because of missing age of each person of the household, 69 households for problem of numbering of the individuals in the household, 35 households for a problem of height measurement (person whose height falls), 16 households for a problem of incorrect gender or age, 1880 households for which we didn't have the height or weight of at least one member.

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

procedure allowed us to differentiate products very precisely. For example, for plain yoghurt, we are able to differentiate them precisely according to their fat content. For snacks, we went until brand level differentiation. Thus, we know the number of calories, proteins, lipids and carbohydrates per 100g of product for each of the 2073 products thanks to the matching of quantities purchased with the amounts of nutrients.

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 the two categories without bar code. Thus purchases of products with bar code are always collected but either fruits and vegetables or meat and fish without bar code are not collected. To overcome this problem of missing data, we implement a procedure of imputation at the household level which is detailed in Appendix 8.1. The method consists in using the full set of observed household characteristics to impute the unobserved value (quantity and expenditures 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

The Body Mass Index (BMI) is a measure of nutritional status defined as the weight (in kilograms) divided by the height (in meter) squared that is used by most nutritionists and epidemiologists to define obesity. Following the World Health Organization and other disease control and prevention institutions, adult individuals with a BMI over 25 are considered as overweight and obese if their BMI is over 30. For children, thresholds depending on age and gender must be used to define overweight and obesity for boys and girls under 18 years old (Cole et al., 2000).

In Table 1, we can see that the average BMI is 23.12 (kg/m^2) and that almost 9% of individuals in our survey are obese. This percentage of obese people is consistent with the national figures in France obtained from other studies (Obépi, 2006) as well as figures on the percentage of overweight people. Indeed, one third of adults suffer from overweight, that is, more than 20 million people in France. Obesity is particularly important for people above 60 years old, with 15% of obese men and 14% of obese women. 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. This shows that the distribution of weight between men and women is more skewed for women than for men. On the contrary, there seems to be no strong difference between girls and boys for children and adolescents.

2002-2003			N	Average BMI (std.dev.)	% Obese	% Overweight
All			22202	23.12 (4.89)	8.72	25.82
Adults	All		16819	24.71 (4.20)	10.82	30.94
	Male	All	7633	25.27 (3.68)	10.42	38.02
		20-60 years old	5572	24.87 (3.67)	8.67	34.58
		more than 60	2061	26.35 (3.48)	15.14	47.31
	Female	All	9186	24.25 (4.54)	11.16	25.06
		20-60 years old	6493	23.71 (4.57)	10.07	20.34
		more than 60	2693	25.55 (4.19)	13.78	36.43
Children	All		5383	18.16 (3.34)	2.17	9.81
and adolescents		Male	2788	18.15 (3.34)	2.62	9.76
less than 20		Female	2595	18.18 (3.34)	1.70	9.87

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 low obesity rate of 4.4% whereas for male farmers, this rate is 10.6%.

2002 - 2003		N	Average BMI (std.dev.)	% Obese	% Overweight
Male	Farmers	142	25.60 (3.70)	10.56	40.85
	Self employed	207	25.76 (3.21)	13.04	41.55
	Senior executive	556	25.06 (3.30)	7.73	37.77
	Middle manager	1124	25.18 (3.36)	7.47	38.17
	White collar	993	24.80 (3.60)	8.86	31.92
	Blue collar	1819	25.22 (3.81)	10.67	36.72
	Retired and without professional activity	2792	25.49 (3.82)	12.32	40.62
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	2946	24.01 (4.60)	11.20	22.03
	Blue collar	337	24.36 (4.46)	13.35	20.47
	Retired and without professional activity	4420	24.72 (4.60)	12.44	29.59

Table 2: Adults BMI and obesity by professional category

3 From Household to Individual Consumption

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 with a set of characteristics denoted by vector x_{ipt}, the individual food calories consumption y_{ipt} is

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

where u_{ipt} is an unknown deviation for this person's consumption. Then, the household consumption y_{it} is

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

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

Remark that assuming $E(u_{ipt}|x_{ipt}) = 0$ would be a way to define the function β without imposing any restriction. However, we do the following assumption that we will maintain throughout the paper:

Assumption 1: For all p, i, t :

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

Assumption 1 implies that

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

This implies that β is non parametrically identified and even overidentified[3] by the natural additive structure between individual consumptions imposed on total household consumption: $E\left(y_{it}|x_{i1t},..,x_{iP(i)t}\right) = \sum_{p=1}^{P(i)} \beta(x_{ipt})$. Assumption 1 thus implies some testable implications of separability of the form

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

It has to be noted that if the separability assumption does not hold exactly, then $E\left(y_{it}|x_{i1t},..,x_{iP(i)t}\right) - \sum_{p=1}^{P(i)} \beta(x_{ipt})$ is a separability error that appears in (2). This error can be due to the intra-household correlation of individual deviations u_{ipt} from average consumption. For example, if $E(u_{ipt}|x_{ipt}) = 0$ but $E(u_{ipt}|x_{ipt}, x_{ip't}) \neq 0$, which means that the mean of u_{ipt} given person p characteristics depends on

[3]Note that assumption 1 is a sufficient assumption but that we could assume a weaker assumption like an additive separability of $E\left(\varepsilon_{it}|x_{i1t},..,x_{iP(i)t}\right) = \sum_{p=1}^{P(i)} h(x_{ipt})$ or $E\left(u_{ipt}|x_{i1t},..,x_{iP(i)t}\right) = \sum_{p=1}^{P(i)} g(x_{ipt})$.to be able to define $\beta(.)$. What matters is that the sum of the u_{ipt} within the household or each of the u_{ipt} be have a conditional expectation on the x_{ipt} of the household members to be additively separable across household members' characteristics.

another person characteristics in the household. Then, we have

$$y_{it} = \sum_{p=1}^{P(i)} \beta(x_{ipt}) + \left[E\left(y_{it} | x_{i1t}, .., x_{iP(i)t}\right) - \sum_{p=1}^{P(i)} \beta(x_{ipt}) \right] + \varepsilon_{it}$$

where $\varepsilon_{it} = \sum_{p=1}^{P(i)} u_{ipt}$ and $E(u_{ipt}|x_{ipt}) = 0$ for all $p = 1, .., P(i)$.

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 straightforwardly this methodology to other data sets from Italy and the Czech Republic.

The crucial point in the choices of covariates x is that the separability assumption of the conditional mean of household consumption y_{it} must hold. Thus with 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) $E(y_{it}|x_{it}) = \sum_{p=1}^{P(i)} \beta(x_{ipt})$ and (B) $E(y_{it}|x_{it}, z_{it}) = \sum_{p=1}^{P(i)} \delta(x_{ipt}, z_{ipt})$ are not equivalent and none is more general than the other. Actually, it could be that (B) is true but not (A) since (B) implies $E(y_{it}|x_{it}) = \sum_{p=1}^{P(i)} E(\delta(x_{ipt}, z_{ipt})|x_{it})$ 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}. Moreover, (A) can be true and not (B), for example if $E(y_{it}|x_{it}, z_{it}) = \sum_{p=1}^{P(i)} \delta(x_{ipt}, z_{ipt}, z_{i1t})$ where $\frac{\partial \delta(.,.,.)}{\partial z_1} \neq 0$ and $E(\delta(x_{ipt}, z_{ipt}, z_{i1t})|x_{it}) = \beta(x_{ipt})$ which will be the case if z_{i1t} is independent of x_{i1t}.

Chesher (1998) uses gender and age as conditioning covariates while we choose to add the BMI as an additional covariate. As mentioned above, none approach is more general than the other. Of course, separability of the conditional mean of household consumption can be tested, but such non parametric tests are expected to have relatively low power. However, we choose to use also the BMI in our conditioning set because there are plausible circumstances under which estimating average consumption conditionally on age and gender only is likely to be biased (because of the violation of the separability assumption).

Equivalently, in order to get consistent estimates of average individual consumptions for an individual of a given age and gender, one needs to assume that household level deviations ε_{it} which are the sum of individual level deviations are not correlated with cross effects (across household members) in the demographic composition of the household in terms of age and gender. This assumption will be invalid when controlling for age and gender only for example if individual deviations from average consumption (for example related to BMI) are on average larger for men than women and correlated with the partner's age.

As the consumption of two individuals of the same age and gender but with different anthropometric measures is likely to be still very heterogenous in particular along the dimension related to the body mass index, we study average consumptions conditional in particular on an anthropometric measure z_{ipt}, that will be the body mass index.

Then, we specify the individual and household consumption as:

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

and

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

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

Specification

Although the function $\beta(.)$ is non parametrically identified, we choose to specify it 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 $\delta^g\left(x^1_{ipt}\right) = 1_{\left\{x^1_{ipt} \leq 13\right\}}\delta^g_1 + 1_{\left\{13 < x^1_{ipt} < 20\right\}}\delta^g_2 + 1_{\left\{x^1_{ipt} \geq 20\right\}}\delta^g_3$, $\overline{z}_{a,g}$ 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 old is the maximum age in the population). 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

We obtain consistent estimates of the model parameters using ordinary least squares in equation (4). Chesher (1998) introduced a smoothing technique to smooth across ages, which amounts to estimate β (the vector of β^g_a for $a = 1, .., 100$, $g = 1, 2$) of equation (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 2×2 and A is the following matrix of size 98×100 :

$$A = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}$$

Measurement errors

Measurement errors are likely to affect household consumption as measured in the survey because households may forget to scan some food items, or because of the lack of observation of wasted food or of food intakes away from home. Assuming that all these errors, denoted ς_{it}, are uncorrelated with individual characteristics x_{ipt}^1, x_{ipt}^2, z_{ipt} of household members, we denote $\widetilde{y}_{it} = y_{it} + \varsigma_{it}$ the observed household food consumption.

Remark that, although for single households, the household consumption is a consistent measure of the individual consumption ($E\left(\widetilde{y}_{it}|P(i)=1\right) = \beta\left(x_{ipt}^1, x_{ipt}^2, z_{ipt}\right)$), it is not necessarily more precise than the estimated $\widehat{\beta}\left(x_{ipt}^1, x_{ipt}^2, z_{ipt}\right)$ because of the measurement error ς_{it}. Actually, assuming that $P(i)$ is independent of ς_{it} (for simplicity but the same result is also obtainable in general), the variances of each estimator are $V\left(\widetilde{y}_{it}|x_{ipt}^1, x_{ipt}^2, z_{ipt}, P(i)=1\right) = V\left(\varsigma_{it}|x_{ipt}^1, x_{ipt}^2, z_{ipt}, P(i)=1\right) = V\left(\varsigma_{it}\right)$ and $V(\widehat{\beta}(x_{ipt}^1, x_{ipt}^2, z_{ipt})|P(i)=1) = \frac{V(\varepsilon_{it}+\varsigma_{it})}{card\left\{i \mid x_{ipt}^1=a, x_{ipt}^2=g, z_{ipt}=z\right\}}$. The second will in general be lower than $V\left(\varsigma_{it}\right)$ if the number of observations such that $x_{ipt}^1 = a$, $x_{ipt}^2 = g$ and $z_{ipt} = z$ is large enough[4].

3.2 Empirical Tests and Estimates

We now present the results of the empirical estimation of individual food consumptions using the previous method.

We apply this method to the total calories 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 apply this method to measures of nutrients like proteins, lipids, or carbohydrates in Bonnet, Dubois and Orozco (2009).

Separability and Specification Choice

As presented in the previous sub-section, the choice of covariates used to compute the conditional mean of household consumption and obtain individual level average consumptions is not innocuous. Conditioning on gender and age only as in Chesher (1998), or on gender, age and BMI relies on different assumptions. As we said before in subsection 3.1, the separability assumption of $E\left(y_{it}|x_{it}\right)$, across the different individual characteristics $x_{i1t}, .., x_{iP(i)t}$ is crucial but its test on its second order derivatives estimated non parametrically leads to a test that clearly lacks of power.

Nevertheless, we have seen previously that when using gender and age only, the separability assump-

[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.

tion is likely to be wrong if another covariate (for example the BMI) is affecting consumption and if such covariate for an individual is correlated with other members' covariates. 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 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 gender, age and BMI.

Empirical Results

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 from the 2004 representative survey about the average number of meals taken out at the household level because this information is not available for 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 again apply the method of Chesher (1998) on this number of meals taken out.

Figure 1 presents the graphs of the estimated function $\beta(.)$ (with penalization parameter $\lambda = 300$). The vertical axis corresponds to kilocalories, while the right oriented horizontal axis is for age and the left oriented horizontal axis is for BMI. These graphs show that the individual calories 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 men but larger and positive for adult men. For women, the slope of calories intake with respect to BMI is more clearly increasing with BMI at all ages.

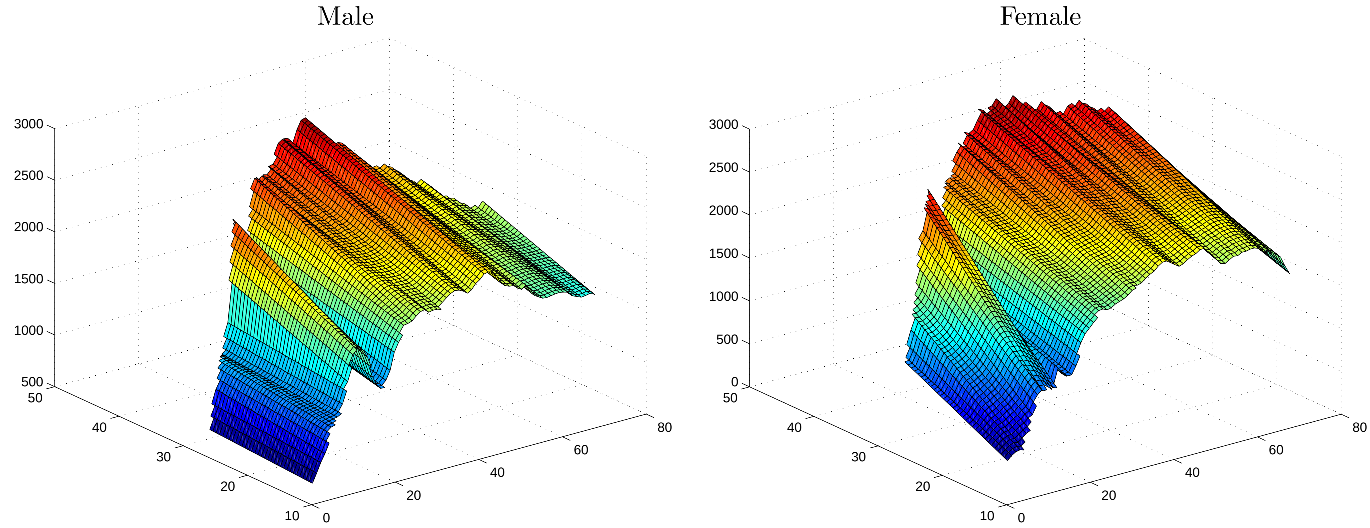

Figure 1: Estimated Individual Calorie Consumption $\beta(age, gender, BMI)$ in kcal per day

Looking at the projection on the age hyperplane of the function β is also interesting to examine the age profile of individual calories consumption. Figure 2 presents such projections for three chosen categories of individuals defined as obese, overweight or underweight and normal[5]. These graphs show mainly that the calories intake is increasing until 18 years old for both girls and boys (with a stagnation for boys between 8 and 11). Then the calories consumption decreases until 25 and increases again until 70 for women and 60 for men. Finally, we can see that even if the age profiles of calories consumption have similar shapes across the three categories of individuals defined as "normal weight", overweight and obese, the consumption of calories is clearly higher for obese than for overweight and for overweight than for "normal". However, it seems that overweight and obese people do consume more calories specially during the periods of life of highest consumption.

Figure 2: Estimated Individual Consumption of Calories

Figure 2 also shows that the differences between obese and non obese people in terms of calories

[5] In appendix, Figure 3 presents individual calories consumption with 95% confidence intervals.

consumption are higher for women than for men. Boys under 10 have very similar consumption of total calories. The same is not true for girls. More obese girls do consume more than less obese girls. Bonnet, Dubois and Orozco (2009) shows that 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. Obese people also eat more proteins and carbohydrates but in a less important relative difference.

Energy Intakes by Category

Implementing now the same decomposition of energy intakes by category of food products, we can evaluate how individual food intakes are distributed across categories such as meat, fish, fruits and vegetables, dairy products, starchy food, junk food, fats (butter, oil, ..). Results are presented in Tables 3 and 4. Tables 3 shows that more than one fourth of energy is on average provided by products in the junk food category (snacks, pizzas, chips, chocolate bars, candies, sodas and soft drinks, ..). The other important categories are fats, starchy food, dairy products and meat, each category providing around 10% of energy intakes, then comes the fruits and vegetables (around 7%) and fish (less than 2%).

Average % (std. dev.)	N	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
Total	22202	9.79	1.65	6.59	14.20	10.44	28.84	15.93	12.56
		(1.64)	(0.71)	(2.62)	(2.06)	(1.18)	(5.84)	(3.17)	(1.98)
Income level									
High	2105	10.47	2.02	7.76	13.79	10.10	26.20	16.91	12.76
		(1.24)	(0.64)	(2.49)	(1.18)	(0.98)	(4.77)	(3.05)	(1.57)
Medium-High	5556	10.18	1.84	7.23	13.95	10.23	27.43	16.51	12.63
		(1.46)	(0.69)	(2.61)	(1.68)	(1.04)	(5.38)	(3.12)	(1.75)
Medium-Low	10697	9.62	1.56	6.31	14.32	10.50	29.44	15.70	12.55
		(1.69)	(0.70)	(2.57)	(2.22)	(1.19)	(5.91)	(3.18)	(2.06)
Low	3844	9.34	1.40	5.80	14.43	10.76	30.66	15.22	12.38
		(1.72)	(0.68)	(2.44)	(2.36)	(1.32)	(5.86)	(3.01)	(2.21)

Table 3: Share of calories intakes by group and by income level

Table 3 shows how the distribution of energy intakes varies depending on the income levels of households. It shows that the lowest socio economic classes eat less energy in fruits and vegetables, meat and fish but a larger share of their energy comes from "junk food" and a little bit from dairy products.

Average % (std. dev.)	N	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
Total	22202	19.47	5.36	10.91	14.57	3.59	21.97	4.26	19.87
		(2.76)	(2.11)	(4.26)	(3.53)	(0.74)	(7.65)	(0.82)	(6.71)
Income level									
High	2105	20.16	6.41	12.52	13.23	3.35	18.75	4.13	21.44
		(2.08)	(1.93)	(3.97)	(2.38)	(0.65)	(6.18)	(0.74)	(4.82)
Medium-High	5556	19.90	5.91	11.83	13.77	3.46	20.18	4.19	20.75
		(2.46)	(2.08)	(4.24)	(2.99)	(0.68)	(7.02)	(0.78)	(5.71)
Medium-Low	10697	19.27	5.12	10.51	14.90	3.65	22.67	4.29	19.59
		(2.88)	(2.06)	(4.20)	(3.69)	(0.75)	(7.74)	(0.83)	(7.04)
Low	3844	19.05	4.64	9.77	15.52	3.78	24.36	4.37	18.51
		(3.02)	(1.98)	(4.10)	(3.90)	(0.78)	(7.86)	(0.87)	(7.60)

Table 4: Share of expenses by group and by income level

Table 4 shows the share of expenses by category of products and by income level of households. It shows that low income households also spend a larger share of their food expenses in junk food and a lower share in fruits and vegetables. It also shows that fats and starchy food represent a small share of expenses while providing a substantial share of calories (Table 3).

Table 5 presents these statistics across weight categories of individuals.

Average % (std. dev.)	N	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
Total	22202	9.79	1.65	6.59	14.20	10.44	28.84	15.93	12.56
		(1.64)	(0.71)	(2.62)	(2.06)	(1.18)	(5.84)	(3.17)	(1.98)
BMI category									
Normal-underweight	14533	9.27	1.51	6.31	14.49	10.72	30.42	14.90	12.38
		(1.64)	(0.74)	(2.80)	(2.19)	(1.19)	(5.98)	(2.55)	(2.10)
Overweight	5732	10.64	1.91	7.10	13.74	9.98	26.02	17.74	12.87
		(1.07)	(0.62)	(2.24)	(1.58)	(0.98)	(4.27)	(3.11)	(1.71)
Obese	1937	11.24	1.90	7.14	13.38	9.71	25.31	18.33	13.00
		(1.05)	(0.50)	(1.91)	(1.80)	(0.86)	(3.73)	(3.77)	(1.52)

Table 5: Share of calories intakes by group and by class of BMI

Table 5 shows that the differences of the spread of energy intakes across categories is less significant across BMI categories. However, overweight and obese people still eat more of their energy from fats and less from the junk food even if Table 7 confirms that they eat more calories of everything.

Average % (std. dev.)	N	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
Total	22202	19.47	5.36	10.91	14.57	3.59	21.97	4.26	19.87
		(2.76)	(2.11)	(4.26)	(3.53)	(0.74)	(7.65)	(0.82)	(6.71)
BMI category									
Normal	14533	19.11	4.92	10.31	15.20	3.80	23.74	4.27	18.66
		(2.94)	(2.12)	(4.50)	(3.82)	(0.76)	(8.10)	(0.86)	(7.35)
Overweight	5732	20.18	6.17	11.89	13.41	3.23	18.77	4.28	22.07
		(2.20)	(1.89)	(3.62)	(2.56)	(0.51)	(5.56)	(0.72)	(4.69)
Obese	1937	20.16	6.26	12.47	13.23	3.07	18.14	4.20	22.46
		(2.33)	(1.54)	3.05	2.39	0.43	4.33	0.79	3.69)

Table 6: Share of expenses by group and by class of BMI

Table 6 shows the share of expenses of each food category according to the BMI of individuals.

Average (kcal) (std. dev.)	N	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
All	22202	215.2	37.1	148.0	294.9	218.5	598.2	345.5	269.9
		(76.2)	(20.1)	(79.3)	(57.5)	43.9	134.0	116.5	84.1
BMI category									
Normal	14533	190.0	32.1	133.7	279.5	208.9	588.3	300.5	247.5
		(71.8)	(20.1)	(81.3)	(56.8)	(44.1)	(138.8)	(99.3)	(81.4)
Overweight	5732	252.2	45.5	170.3	318.3	232.1	604.4	415.9	303.2
		(54.9)	(16.5)	(67.1)	(45.1)	(35.5)	(115.4)	(89.5)	(68.7)
Obese	1937	295.0	50.1	189.7	342.2	250.4	654.6	474.7	339.4
		(62.0)	(15.3)	(66.6)	(45.3)	(38.4)	(134.3)	(106.6)	(74.6)

Table 7: Calories per day by food group and class of BMI

Table 8 shows expenditures of each food category for obese and non obese people. It confirms that overweight and obese people eat more calories of all categories and spend more for nutrition in all types of food.

Average (€) (std. dev.)	N	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
All	22202	0.91	0.28	0.58	0.61	0.16	0.90	0.19	1.01
		(0.42)	(0.18)	(0.40)	(0.21)	(0.06)	(0.29)	(0.07)	(0.58)
BMI category									
Normal	14533	0.80	0.24	0.51	0.56	0.15	0.85	0.17	0.88
		(0.42)	(0.19)	(0.41)	(0.21)	(0.07)	(0.29)	(0.07)	(0.60)
Overweight	5732	1.08	0.35	0.68	0.69	0.17	0.95	0.22	1.22
		(0.32)	(0.15)	(0.34)	(0.16)	(0.05)	(0.25)	(0.05)	(0.42)
Obese	1937	1.22	0.39	0.79	0.77	0.19	1.07	0.25	1.38
		(0.33)	(0.14)	(0.34)	(0.16)	(0.05)	(0.31)	(0.05)	(0.40)

Table 8: Expenses per day by food group and class of BMI

4 The Causal Effect of Food Intakes on Body Mass Index

4.1 Identification

Using the information on individual body mass index, as well as the information on individual average food intakes and household food intakes, we want to identify the causal relationship between individual food consumption and individual measures of obesity, taking into account the unobserved heterogeneity of individuals. Denoting b_{ipt} the body mass index of person p in household i at the end of period t, we consider the following specification

$$b_{ipt} = \eta_{ip} + \mu y_{ipt} + \xi_{ipt} \tag{5}$$

where y_{ipt} is the individual calories consumption, μ is a coefficient, η_{ip} is an individual specific unobserved effect and ξ_{ipt} is an unobserved random shock mean independent of y_{ipt} ($E\left(\xi_{ipt}|y_{ipt}\right)=0$) which can be considered as a "production function" shock on BMI.

The specification choice of (5) implies that all the unobserved heterogeneity varying over time that affects the body mass index of the individual like their level of physical activity is included in ξ_{ipt}, while that their natural propensity to be heavy can be considered as fixed over time and capture by η_{ip} . We will introduce later some heterogeneity in the parameter μ without changing the spirit of the identification strategy.

Identification of the causal effect of unobserved individual food intake on BMI

Remark that the identification of the effect of food intake on BMI is not straightforward since we do not observe y_{ipt} in (5). Instead, we have shown in the previous section how to identify average individual food intakes (conditional on some characteristics) from household level food consumption.

However, using (1), we can define $\widehat{y}_{ipt}$ and $\widetilde{y}_{ipt}$ as

$$\widehat{y}_{ipt} \equiv y_{ipt} - u_{ipt} = \beta\left(x_{ipt}\right)$$

or

$$\widetilde{y}_{ipt} \equiv y_{it} - \sum\nolimits_{p' \neq p} \beta(x_{ip't}) = y_{ipt} + \sum\nolimits_{p' \neq p} u_{ip't} = \widehat{y}_{ipt} + \sum\nolimits_{p'=1}^{P(i)} u_{ip't}$$

where $\beta\left(x_{ipt}\right) = \beta\left(x_{ipt}^{1}, x_{ipt}^{2}, z_{ipt}\right)$ is identified using $E\left(y_{it}|x_{i1t},..,x_{iP(i)t}\right) = \sum_{p=1}^{P(i)} \beta\left(x_{ipt}\right)$ with the specification (4), as shown in section 3.

Then, replacing y_{ipt}, the equation (5) becomes

$$b_{ipt} = \eta_{ip} + \mu\widehat{y}_{ipt} + \xi_{ipt} + \mu u_{ipt} \tag{6}$$

and

$$b_{ipt} = \eta_{ip} + \mu\widetilde{y}_{ipt} + \xi_{ipt} - \mu\sum\nolimits_{p' \neq p} u_{ip't} \tag{7}$$

In both cases, potential endogeneity problems of $\widehat{y}_{ipt}$ and $\widetilde{y}_{ipt}$ will prevent the identification of μ.

Remind that with Assumption 1, we have the following mean orthogonality $E\left(u_{ipt}|x_{i1t},..,x_{iP(i)t}\right)=0$ which implies that $E\left(u_{ipt}|x_{ipt}\right)=0$ and $cov(u_{ipt},\widehat{y}_{ipt})=0$. Thus, the potential endogeneity problems in (6) can come from the correlation between unobserved BMI shocks ξ_{ipt} and individual deviations of food intakes u_{ipt}. Whereas the potential endogeneity problems in (7) exist whatever the distributional properties of ξ_{ipt}. Before precisely determining the identification conditions, we make the following additional assumption:

Assumption 2: For all i, t and $p' \neq p$

$$cov(\xi_{ipt}, u_{ip't}) = 0$$

Assumption 2 means that unobserved shocks to BMI for a given individual are uncorrelated with individual deviations of food intakes of other members of the household[6].

Then, we have that

$$\frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{var(\widehat{y}_{ipt})} = \mu - \frac{cov\left(\xi_{ipt}, u_{ipt}\right)}{var(\widehat{y}_{ipt})}$$

and

$$\frac{cov\left(b_{ipt}, \widetilde{y}_{ipt}\right)}{var(\widetilde{y}_{ipt})} = \mu \left[1 - \frac{var\left(\sum_{p' \neq p} u_{ip't}\right)}{var(\widetilde{y}_{ipt})}\right] < \mu$$

Since we can conjecture that physical activity will be negatively correlated with shocks ξ_{ipt} to the BMI of the individual and positively correlated to individual deviations u_{ipt} of food intake from the average, unobserved physical activity is likely to generate a negative correlation between ξ_{ipt} and u_{ipt}

Then, if $cov\left(\xi_{ipt}, u_{ipt}\right) \leq 0$,

$$\frac{cov\left(b_{ipt}, \widetilde{y}_{ipt}\right)}{var(\widetilde{y}_{ipt})} \leq \mu \leq \frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{var(\widehat{y}_{ipt})} \tag{8}$$

with the particular case that

$$\mu = \frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{var(\widehat{y}_{ipt})} \tag{9}$$

if $cov\left(\xi_{ipt}, u_{ipt}\right) = 0$.

Equation (9) shows that one can point identify μ if $cov\left(\xi_{ipt}, u_{ipt}\right) = 0$. However, this identification relies on the fact that no unobserved physical activity or other shocks generate a correlation between unobserved BMI shocks ξ_{ipt} and individual food intakes deviations (appetite) u_{ipt}. Remark, that the time period being a year, this uncorrelation is much less unlikely that if the time period was shorter because the whole source of correlation coming from fixed (over time) unobserved heterogeneity is taken into account by the fixed effect η_{ip}. Nonetheless, we could still imagine that individuals change permanently of physical activity level from one year to the next and thus that ξ_{ipt} and u_{ipt} are negatively correlated. Then, the inequalities (8) show that we can still set-identify μ by obtaining a lower bound and an upper bound. Moreover, the upper and lower bounds rely simply on OLS regressions with individual fixed effects. The informativeness of the set identified will then depend on the difference between the upper and lower bounds.

[6] Remark that a sufficient condition for the next result to be true would simply be that $cov\left(\xi_{ipt}, \sum_{p' \neq p} u_{ip't}\right) = 0$.

Taking into account measurement errors

However, we also would like to consider the robustness of this result if one takes into account the measurement errors on $\beta(.)$ that appear in $\widehat{y}$.

Actually, with measurement error ς_{ipt}, we have

$$\widehat{y}_{ipt} \equiv \beta\left(x_{ipt}\right) - \varsigma_{ipt} = y_{ipt} - u_{ipt} - \varsigma_{ipt}$$

and

$$\widetilde{y}_{ipt} \equiv y_{it} - \sum\nolimits_{p' \neq p} \left[\beta(x_{ip't}) - \varsigma_{ipt}\right] = y_{ipt} + \sum\nolimits_{p' \neq p} \left[u_{ip't} + \varsigma_{ip't}\right] = \widehat{y}_{ipt} + \sum\nolimits_{p'=1}^{P(i)} \left[u_{ip't} + \varsigma_{ip't}\right]$$

Then, (6) and (7) become

$$b_{ipt} = \eta_{ip} + \mu \widehat{y}_{ipt} + \xi_{ipt} + \mu u_{ipt} + \mu \varsigma_{ipt}$$

and

$$b_{ipt} = \eta_{ip} + \mu \widetilde{y}_{ipt} + \xi_{ipt} - \mu \sum_{p' \neq p} \left[u_{ip't} + \varsigma_{ip't}\right]$$

This implies that

$$\frac{cov\left(b_{ipt}, \widetilde{y}_{ipt}\right)}{var(\widetilde{y}_{ipt})} = \mu \left[1 - \frac{var(\sum_{p' \neq p} u_{ip't} + \varsigma_{ip't})}{var(\widetilde{y}_{ipt})}\right] < \mu$$

but also that

$$\frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{var(\widehat{y}_{ipt})} = \mu \left[1 - \frac{var\left(\varsigma_{ipt}\right)}{var(\widehat{y}_{ipt})}\right] - \frac{cov\left(\xi_{ipt}, u_{ipt}\right)}{var(\widehat{y}_{ipt})}$$

so that we cannot determinate whether it is larger or smaller than μ if $cov\left(\xi_{ipt}, u_{ipt}\right) < 0$.

However, if we consider the two-stage least squares parameter obtained by regressing b_{ipt} on $\widehat{y}_{ipt}$ instrumented by $\widetilde{y}_{ipt}$, we have the following

$$\frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{cov(\widehat{y}_{ipt}, \widetilde{y}_{ipt})} = \mu - \frac{cov\left(\xi_{ipt}, u_{ipt}\right)}{var(\widehat{y}_{ipt}) - var(\varsigma_{ipt})}$$

Remark that $cov(\widehat{y}_{ipt}, \widetilde{y}_{ipt}) = var(\widehat{y}_{ipt}) - var\left(\varsigma_{ipt}\right)$ is identified, so that $var\left(\varsigma_{ipt}\right)$ is also identified. We can now consider that inference on the sign (and the value) of $var(\widehat{y}_{ipt}) - var\left(\varsigma_{ipt}\right)$ is known.

Assuming that the variance of measurement errors ς_{ipt} is not too large ($var(\varsigma_{ipt}) < var(\widehat{y}_{ipt})$), these results show that we can set identify μ with the following inequalities

$$\frac{cov\left(b_{ipt}, \widetilde{y}_{ipt}\right)}{var(\widetilde{y}_{ipt})} \leq \mu \leq \frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{cov(\widehat{y}_{ipt}, \widetilde{y}_{ipt})}$$

that replace (8) when measurement errors are taken into account, with again the particular case that μ is point identified if $cov\left(\xi_{ipt}, u_{ipt}\right) = 0$ with the two-stage least squares parameter

$$\mu = \frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{cov(\widehat{y}_{ipt}, \widetilde{y}_{ipt})}$$

Remark that the "reverse" two-stages parameter obtained by regressing b_{ipt} on $\widetilde{y}_{ipt}$ instrumented by $\widehat{y}_{ipt}$, is

$$\frac{cov\left(b_{ipt},\widetilde{y}_{ipt}\right)}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}=\mu\left[1+\frac{var(u_{ipt}+\varsigma_{ipt})}{var(\widehat{y}_{ipt})-var(\varsigma_{ipt})}\right]$$

which provides another upper bound for μ if we maintain that measurement errors are such that $var(\widehat{y}_{ipt}) > var(\varsigma_{ipt})$. Remark also that it provides an upper bound for μ which does not depend on the correlation between shocks to BMI ξ_{ipt} and individual food intake deviations u_{ipt}.

To summarize (see proof and other case in Appendix 8.3), when $0 < var(\varsigma_{ipt}) < var(\widehat{y}_{ipt})$, we have the following identification results (point or set-identification) depending on $cov\left(\xi_{ipt},u_{ipt}\right)$:

$cov\left(\xi_{ipt},u_{ipt}\right)=0$	$\mu=\frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})}$
$cov\left(\xi_{ipt},u_{ipt}\right)>0$	$\max\left[\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{var(\widetilde{y}_{ipt})},\frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})}\right]\leq\mu\leq\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}$
$cov\left(\xi_{ipt},u_{ipt}\right)<0$	$\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{var(\widetilde{y}_{ipt})}\leq\mu\leq\min\left[\frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})},\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}\right]$

The case where $var(\widehat{y}_{ipt}) > var\left(\varsigma_{ipt}\right)$ will happen to be empirically the more relevant one since we can identify $cov(\widehat{y}_{ipt},\widetilde{y}_{ipt}) = var(\widehat{y}_{ipt}) - var\left(\varsigma_{ipt}\right)$ and the empirical estimates is found to be positive, statistically significant and equal to 0.24.

Moreover, we will assume that $cov\left(\xi_{ipt},u_{ipt}\right) < 0$ because of unobserved physical activity, so that the ordinary least squares and two stage least squares coefficients provide us the following inequality

$$\frac{cov\left(b_{ipt},\widetilde{y}_{ipt}\right)}{var(\widetilde{y}_{ipt})}\leq\mu\leq\min\left[\frac{cov\left(b_{ipt},\widehat{y}_{ipt}\right)}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})},\frac{cov\left(b_{ipt},\widetilde{y}_{ipt}\right)}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}\right]\tag{10}$$

This set identification is robust to any sort of unobserved heterogeneity that would lead the endogeneity of calorie consumption in the BMI production function because of any negative correlation between unobserved true caloric intake (u_{ipt}) and the unobserved shock on BMI (ξ_{ipt}). It thus allows unobserved physical activity to reduce the BMI of individuals while increasing their food consumption at the same time. Of course the lower bound on μ will be further away from μ the higher the absolute value of the correlation between these unobserved terms. However, this interval can give robust values for the minimum and maximum effects of calorie consumption on BMI.

Overidentification with external instrumental variables allowing individual unobserved heterogeneity

We study now the case where external variables would allow us to point identify our parameter of interest μ. We consider the identification conditions using either measure of food consumption $\widehat{y}_{ipt}$ or

$\widetilde{y}_{ipt}$:

$$b_{ipt} = \eta_{ip} + \mu \widehat{y}_{ipt} + \xi_{ipt} + \mu u_{ipt} \tag{11}$$

and

$$b_{ipt} = \eta_{ip} + \mu \widetilde{y}_{ipt} + \xi_{ipt} - \mu \sum\nolimits_{p' \neq p} u_{ip't} \tag{12}$$

The endogeneity problem in (11) and (12) when using $\widehat{y}_{ipt}$ or $\widetilde{y}_{ipt}$ for the individual food intake can also be solved with additional external instruments. The existence of such instruments will lead to overidentification of μ. One needs instruments w_{ipt} correlated with the individual deviation of food intake from the average individual of the same gender, age and BMI, but not with the individual shock ξ_{ipt} on the body mass index "production function".

Remark that the validity of instruments requires always that $E\left(\xi_{ipt}|w_{ipt}\right) = 0$ in addition to different conditions required for $\widehat{y}_{ipt}$ and $\widetilde{y}_{ipt}$ requiring that $E\left(u_{ipt}|w_{ipt}\right) = 0$ for $\widehat{y}_{ipt}$ and that $\sum_{p' \neq p} E\left(u_{ip't}|w_{ipt}\right) = 0$ for $\widetilde{y}_{ipt}$. These conditions are quite different. Results may thus differ according to which identifying assumption is used. Moreover, we need that the instruments be correlated with the endogenous variable that is $\frac{\partial}{\partial w_{it}} E\left(\widehat{y}_{ipt}|\eta_{ip}, w_{ipt}\right) \neq 0$ and $\frac{\partial}{\partial w_{it}} E\left(\widetilde{y}_{ipt}|\eta_{ip}, w_{ipt}\right) \neq 0$.

However $E\left(\widehat{y}_{ipt}|\eta_{ip}, w_{ipt}\right) = E\left(\beta(x^1_{ipt}, x^2_{ipt}, z_{ipt})|\eta_{ip}, w_{ipt}\right)$ may vary with w_{ipt} only if characteristics x^1_{ipt}, x^2_{ipt}, z_{ipt} vary with time while the source of variation of $E\left(\widetilde{y}_{ipt}|\eta_{ip}, w_{ipt}\right)$ with respect to w_{ipt} seems more informative since $E\left(\widetilde{y}_{ipt}|\eta_{ip}, w_{ipt}\right) = E\left(\beta\left(x^1_{ipt}, x^2_{ipt}, z_{ipt}\right)|\eta_{ip}, w_{ipt}\right) + E\left(u_{ipt}|\eta_{ip}, w_{ipt}\right)$ (because $\sum_{p' \neq p} E\left(u_{ip't}|\eta_{ip}, w_{ipt}\right) = 0$) depends on individual deviations of food intakes u_{ipt}. Thus, any variable w_{ipt} correlated with u_{ipt} will be a good instrument for $\widetilde{y}_{ipt}$. Such variables will be used in practice. Then, given the identification assumptions for instruments w_{ipt} the coefficient μ will be point identified.

4.2 Empirical estimates

Identified interval of μ

We first present the more robust results that only provide set identification. Actually, estimating the values of the upper and lower bound whose theoretical expressions are given in (10) and which simply correspond to OLS and 2SLS parameters, we find

$$\underset{(0.00002)}{0.00053} \leq \widehat{\mu} \leq \underset{(0.00027)}{0.00947}$$

where standard errors are in parentheses.

This shows that our interval of identification of μ is large but has a lower bound which is statistically significant and strictly positive, proving that food calories intakes do increase BMI (which is not a surprise), even if one allows for unobserved heterogeneity of individuals

We now turn to the case where we do stronger identifying assumptions in order to get point identification, by using external instrumental variables.

Point identification of μ

Using individual or household characteristics such as living in rural vs. urban area, socioeconomic classes, income levels, internet access, computer ownership or shoe size interacted with indicators of individual obesity status as instruments, we use a two stage least squares estimator with fixed effects and the results of the estimation of (**??**) are presented in Table 9. These instruments are supposed to be correlated with variations on preferences for "ideal" weight and calories consumption across gender, age and BMI but not with production shocks or individual specific metabolism of individuals ξ_{ipt} that are not fixed over time, since fixed effects η_{ip} already capture these unobserved individual physiologic or metabolic specificities. In the specification chosen, the coefficient μ is allowed to depend on some characteristics of the individual. In particular, we will also allow different effects of the calories consumption if the individual is initially of "normal" weight, overweight or obese.

IV FE Estimates	Men	Women	Children
	Coef. (std. err.)	Coef. (std. err.)	Coef. (std. err.)
Calories×underweight-normal (μ_0)	0.010 (0.003)	0.010 (0.004)	0.009 (0.003)
Calories×overweight (μ_1)	0.011 (0.002)	0.011 (0.004)	0.010 (0.003)
Calories×obese (μ_2)	0.012 (0.002)	0.013 (0.003)	0.012 (0.002)
F test that all $\eta_{ip} = 0$ (p value)	34.04 (0.00)	36.38 (0.00)	12.29 (0.00)
Obs.	7520	8786	5068
Sargan test statistic (df)	4.192 (4)	0.707 (4)	8.149 (5)

Table 9: 2SLS estimation of the effect of calories on BMI with individual fixed effects

Results of Table 9 show that μ_0, μ_1, μ_2 are statistically significant and that the magnitude of these coefficients is increasing with BMI. Calories consumption has an impact on overweight and obesity of adults and children. For children, a 100 kcal increase per day during a year would lead to an increase of between 0.9 and 1.2 kg/m^2 (unit of BMI), that is between 1.7 and 2.4 kgs for a 10 year old child who is around 140 cm. These results justify the particular attention that should be given to the nutrition of children. For adult women, the rate of increase is such that 100 kcal increase per day during a year would lead to an increase of between 1 and 1.3 kg/m^2 depending on the current level of her BMI, that is between 2.6 and 3.5 kgs for a 163 cm woman. For adult men, the rate of increase is lower in absolute value such that 100 kcal increase per day during a year would lead to an increase of between 1 and 1.2 kg/m^2 that is between 3 and 3.6 kgs for an adult of 173 cm depending on the current level of his BMI.

These estimates show that even after taking into account the unobserved individual specific level of BMI (through η_{ip}), additional calories consumption do lead to a permanent increase of BMI. Estimates are based on continuous food consumption over two years and thus can explain individual level time variations of weight but not the cross-sectional variation of BMI because the unobserved component of BMI (η_{ip}) is likely correlated with calories intake too. Comparing these results with those obtained without including fixed effects shows that without fixed effects the correlation between calories intakes and BMI is much lower in absolute value than it is after removing the individual level heterogeneity. This is also consistent with the descriptive statistics of calories as a function of BMI obtained in section 3. The estimated effects of additional calories on BMI are not valid for comparisons across individuals but for the individual time variation of BMI. Thus, calories intakes can lead to obesity but the heterogeneity is also important in terms of the propensity to be obese and in terms of metabolic assimilation of calories. Individuals may have different levels of η_{ip} that put them more or less close to obesity but calorie consumption always cause weight gain. The inclusion of individual unobserved heterogeneity η_{ip} allows to look at their correlation with characteristics of the individuals and the household by a simple OLS regression in Table 10.

Dep. Variable: η_{ip} (OLS)	Men	Women	Children
Demographic Characteristics			
BMI	0.414*** (0.005)	0.340*** (0.006)	0.539*** (0.006)
Age	-0.123*** (0.001)	-0.189*** (0.002)	-0.087*** (0.005)
Household size	-0.077*** (0.026)	0.176*** (0.039)	0.141*** (0.025)
Number of Children	0.020 (0.024)	0.027 (0.038)	-0.264*** (0.024)
Wealth			
Income	-0.037*** (0.015)	-0.151*** (0.019)	-0.040*** (0.014)
Socioeconomic Level	-0.099*** (0.050)	-0.386*** (0.067)	-0.129*** (0.048)
Other Household Characteristics			
Fruit tree	-0.041 (0.043)	0.017 (0.065)	-0.020 (0.038)
Vegetable garden	-0.028 (0.045)	-0.151 (0.069)	0.005 (0.0411)
TV	0.026 (0.019)	-0.008 (0.028)	-0.033** (0.018)
Internet	-0.058 (0.044)	0.444 (0.066)	-0.030 (0.037)
Computer	-0.369*** (0.044)	-0.433*** (0.066)	0.026 (0.040)
Geographical Indicators			
Rural Areas	0.027 (0.043)	0.142*** (0.066)	0.082*** (0.038)
East region	-0.158*** (0.075)	-0.166 (0.109)	0.118** (0.070)
North region	-0.287*** (0.075)	-0.183** (0.108)	0.035 (0.070)
West region	-0.163*** (0.063)	-0.264*** (0.092)	0.021 (0.062)
Centre West region	-0.233*** (0.079)	-0.169 (0.113)	0.190*** (0.072)
Centre East region	-0.243*** (0.067)	-0.223*** (0.095)	0.020 (0.061)
Southeast region	0.008 (0.071)	-0.060 (0.101)	0.022 (0.071)
Southwest region	-0.070 (0.077)	-0.164 (0.111)	0.047 (0.074)
Const.	-3.193*** (0.194)	3.828 (0.351)	-8.046*** (0.174)
Obs.	7633	8868	5251

Table 10: Correlation of η_{ip} with household and individual characteristics

Table 10 shows that the unobserved heterogeneity factors η_{ip} that explain the individual non time varying part of the body mass indices are negatively correlated with age for men, women and children. They are also positively correlated with the body mass index and negatively correlated with income.

5 The Price Elasticity of Food Consumption

5.1 Prices of Calories

Given the data at hand, the precise observation of purchases allows us to use quantities and expenses by product or category of products to compute the price of food products and thus the cost of calories for each product. The cost of energy obtained is on average of 0.184 € per 100 kcal. Equivalently, the cost of energy for the purchases of this representative panel of the french population is of 3.7 € for 2000 kcal. We can also compute the cost of energy per category of products by dividing the total expenses for a product category by the total calories of the quantities purchased of such category. These household level prices of calories are presented in Table 11.

We also take advantage of our disaggregation method to compute individual level prices of calories. Actually, we define the individual level prices (by category) by the ratio of the average individual level expenditures (by category) with the average individual level quantities of calories. This will allow us to distinguish household level elasticities from individual level elasticities.

We denote respectively the true cost paid by the household i at period t for a calorie as p_{it} (it corresponds to the total expenditure divided by the number of calories purchased) and the individual level cost of a calorie p_{ipt}. Dividing all food products into K groups, one can define prices and calories consumption per food group and denote them p_{ipt}^k and y_{ipt}^k respectively for person p in household i, and p_{it}^k and y_{it}^k for household level variables. Remark that the within household variation of food prices can be explained by the variation in composition of food items eaten by each individual in the household which can be explained by different tastes and by within household heterogeneity of price elasticities.

Our choice of food groups responds either to a natural definition (like meat, fish, fruits and vegetables or starchy food) or have distinctive calories content in terms of densities. From that point of view, the definition of the group "junk food" which is rather vague shows that such food items exhibit high energy density in addition to provide low cost per calorie. Table 11 also reports the density of food products (its average and the 10 and 90 percent quantile) for each category. It shows that food categories can be distinguished on "objective" criteria using both these two characteristics of cost of calories and energy density (remark that here we refer to the energy density of food items as eaten, which means in particular once they are cooked for starchy food such as rice and pasta whose density is much higher if considered raw).

The distinction between household level and individual level elasticities is important for policy purpose since obesity and overweight problems can be different across members of a given household and thus elasticities of individual consumption are important to assess the effect of prices on individual level food intakes. Moreover, the average price of a calorie paid by the household can be different from the price of those calories paid by different individuals of the household. We thus will use estimated individual level prices of calories for the estimation of individual level elasticities.

	Costs of Calories (0.01€/100kcal)				Density
	Household Level		Individual Level		(kcal/100g)
All	Mean	23.25	20.73		
	Std. Dev.	6.79	6.22		
Food Group:				Food Group:	
Meat	Mean	46.69	40.28	Mean	206.3
	Std. Dev.	44.85	8.39	10%	146.8
	Median	44.44	41.32	90%	245.8
Fish	Mean	80.03	67.99	Mean	131.0
	Std. Dev.	32.02	17.42	10%	82.0
	Median	76.48	70.72	90%	191.5
Fruits and Vegetables	Mean	40.73	34.32	Mean	135.3
	Std. Dev.	10.55	10.93	10%	87.2
	Median	40.29	35.54	90%	168.5
Dairy Products	Mean	21.74	20.25	Mean	175.0
	Std. Dev.	5.32	3.87	10%	138.7
	Median	20.95	20.27	90%	207.5
Starchy Food	Mean	8.01	6.98	Mean	217.2
	Std. Dev.	3.03	1.95	10%	199.7
	Median	7.61	7.03	90%	249.6
Junk Food	Mean	16.04	14.80	Mean	345.5
	Std. Dev.	4.69	3.02	10%	289.1
	Median	15.50	15.35	90%	403.6
Fat (oil, butter, ..)	Mean	6.04	5.35	Mean	381.2
	Std. Dev.	2.76	0.93	10%	158.5
	Median	5.71	5.18	90%	499.8
Others	Mean	42.30	34.01	Mean	155.2
	Std. Dev.	19.47	15.18	10%	110.0
	Median	38.94	38.75	90%	218.1

Table 11: Cost of Calories per Group

Table 11 gives the household and individual level prices of calories and shows that calories have very different prices according to which food products category they belong to. One can order the food groups from the most expensive in terms of energy to the least expensive as follows: fish, meat, fruits and vegetables, dairy products, junk food, starchy food, fats (oil, butter, ..). This shows that for a given energy requirement, variations in the source of calories across groups can have important budget effects.

Table 11 also shows that individual level prices are always lower than household level prices which shows that there is a positive correlation within the household between individual calories intakes and the price of the calories consumed (unless expenditures and purchased quantities exhibit strange patterns of measurement errors). Actually one can easily show for a two persons household that the ratio of total expenditures to total quantity is always larger than the average of the individual ratios of expenditures to quantities if the individual that consumes more also pays proportionately more[7].

5.2 Estimation of Price Elasticities of Nutrients Consumption

We now turn to the estimation of price elasticities of energy both at the household and individual levels using our measures of calories consumption and costs.

Household level elasticities:

As a first simple model, we suppose that the true cost paid by the household i at period t for a calorie is denoted by p_{it} and we estimate the following reduced form equation by OLS

$$\ln y_{it} = \alpha + \beta \ln p_{it} + \varepsilon_{it}$$

and find $\widehat{\beta} = -0.51$ (with a standard error of 0.017). Looking at food expenditures, Blundell, Browning and Meghir (1994) find a price elasticity of -0.6 at the household level in the UK.

For the estimation of price elasticities of calorie consumption by food group, we use observed expenses and quantities y_{it}^k of calories by group to define prices p_{it}^k and estimate for each food group $l = 1, ..., K$

$$\ln y_{it}^l = \alpha_i^l + \sum_{k=1}^{K} \beta_l^k \ln p_{it}^k + \delta_l \ln y_{it} + \varepsilon_{it}^l \tag{13}$$

However, the estimation of (13) may suffer from an endogeneity bias because of the measurement error on prices of calories obtained using total expenses by food groups and the total quantities of calories purchased by food group. In order to instrument the prices, we use the information on the retailer characteristics at which the household makes purchases which provides some price variation that is not supposed to be correlated with measurement errors on quantities of calories.

Remark that we obtained distances from the household to each supermarket by matching the households addresses information with an exhaustive dataset on the location of all supermarkets in France collected by the firm LSA. Moreover, we compute the share of food expenses done at each retailer chain

[7] Actually, given $q_1 \leq q_2$ we can show the equivalence

$$\frac{y_1}{q_1} \leq \frac{y_2}{q_2} \Leftrightarrow \frac{y_1 + y_2}{q_1 + q_2} \geq \frac{1}{2}\left(\frac{y_1}{q_1} + \frac{y_2}{q_2}\right)$$

for each household. As retailer chains have different prices for the same products, the individual level market share of each retailer chain in the household purchases during each year will sometimes be used to instrument prices in the next regressions.

Table 12 presents the results of the estimation by two stage least squares of (13) using different sets of instruments[8] obtained by combination of the distances from the household to three different types of supermarkets (hypermarkets, supermarkets, hard discounters) and of household level shares of expenditures at these different retailers interacted with household characteristics like its size, the number of children, the income or the region of residence (OLS estimations with and without fixed effects are presented in Tables 15 and 16 in appendix 8.2). The fixed effect estimates α_{ip}^{l} are not shown in this table but an F-test testing if they are jointly statistically different from zero rejects their nullity. The estimation results of these household price elasticities seem to show that the junk food and dairy products are the more elastic food categories while fruits and vegetables or starchy foods are less elastic even if elasticities are still significant. Concerning cross-price elasticities, most of them are relatively imprecisely estimated but generally not large in absolute values anyway.

FE-IV (Std. Errors)	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat (Oil, butter, .)
$\ln p_{it}^{Meat}$	-1.61 (0.65)	2.73 (1.55)	-0.06 (0.72)	-0.004 (0.61)	0.48 (0.48)	-0.07 (0.84)	-0.80 (0.91)
$\ln p_{it}^{Fish}$	0.18 (0.45)	-0.85 (0.66)	-0.002 (0.39)	0.08 (0.26)	0.06 (0.38)	-0.25 (0.45)	0.36 (0.60)
$\ln p_{it}^{FruitsVeg.}$	0.01 (0.76)	0.41 (1.24)	-1.37 (0.71)	0.76 (0.54)	-0.72 (0.41)	-0.48 (0.9)	-1.52 (0.89)
$\ln p_{it}^{Dairy}$	-0.17 (0.79)	0.38 (2.09)	0.25 (0.77)	-0.98 (0.59)	1.05 (0.96)	1.61 (1.38)	2.75 (1.12)
$\ln p_{it}^{Starchyfood}$	0.24 (0.61)	-1.23 (1.60)	0.12 (0.40)	0.47 (0.35)	-0.74 (0.44)	0.35 (0.90)	1.09 (0.94)
$\ln p_{it}^{Junkfood}$	-0.40 (0.71)	-2.19 (1.80)	0.55 (1.07)	-0.31 (0.52)	-0.90 (0.44)	-2.00 (1.03)	-1.96 (1.1)
$\ln p_{it}^{Fat}$	0.15 (0.36)	0.57 (1.35)	-0.03 (0.57)	-0.34 (0.39)	-0.091 (0.48)	0.42 (0.56)	-1.19 (0.57)
$\ln p_{it}^{Others}$	-0.44 (0.57)	-0.03 (0.92)	0.20 (0.40)	0.40 (0.32)	0.14 (0.40)	0.02 (0.6)	0.65 (0.34)
$\ln y_{it}$	1.42 (0.33)	1.29 (0.47)	0.29 (0.72)	1.11 (0.32)	0.67 (0.23)	1.11 (0.75)	2.06 (1.07)
Const	14.04 (2.95)	3.01 (4.49)	10.41 (3.47)	6.16 (2.57)	8.76 (2.50)	10.07 (3.02)	6.92 (3.11)
F Test ($\alpha_i^l = 0, \forall i$)	3.41 (0.00)	2.27 (0.00)	4.12 (0.00)	4.59 (0.00)	4.00 (0.00)	2.47 (0.00)	1.71 (0.00)
Obs.	8332	8332	8332	8332	8332	8332	8332
Sargan stat. (df)	13.47 (16)	9.29 (11)	3.65 (12)	20.47 (17)	13.07 (16)	4.97 (4)	8.51 (9)

Table 12: Fixed Effects - Instrumental Variables Estimation of Household Price Elasticities

Then, as these household level price elasticities could hide different individual patterns of elasticities of substitution, we look at the individual level behavior thanks to the methodology developed earlier to disaggregate consumption.

Individual level elasticities:

[8] The number of excluded instruments represents the degrees of freedom of the Sargan test of overidentifying restrictions. This number is varying because not all instruments are always used for each group among the list of distances to supermarkets and the shares of total expenses of the household in each type of supermarket chain.

We now turn to the estimation of individual level elasticities using the following model for each food group $l = 1, .., K$,

$$\ln y_{ipt}^{l} = \alpha_{ip}^{l} + \sum_{k=1}^{K} \beta_{l}^{k} \ln p_{ipt}^{k} + \delta_{l} \ln y_{ipt} + \varepsilon_{ipt}^{l} \tag{14}$$

where p_{ipt}^{k} and y_{ipt}^{l} are respectively the price and the calorie consumption per food group for a person p in a household i and α_{ip}^{l} are individual fixed effects. We first estimate (14) in the case where $\alpha_{ip}^{l} = \alpha^{l}$ by OLS. The results are in Table 17 in appendix 8.2 and show very implausible elasticities. Table 18 in appendix 8.2 shows the results of (14) using fixed effects to account for the heterogeneity of α_{ip}^{l}. Again, the results are implausible. In Table 18, the F test presents the p-value of the test that all fixed effects are zero. It is likely that measurement errors on the individual level prices of calories by groups are biasing the previous results. In order to take into account this endogeneity, we then introduce some instrumental variable estimator where we instrument individual level prices by interactions between indicators of the obesity status of the individual with household level prices of calories and geographical and socioeconomic class indicators. Using these sort of instruments together with individual fixed effects, we obtain more reasonable results presented in Table 13. All Sargan tests do not reject overidentifying restrictions and own price elasticities are always negative.

FE-IV	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
$\ln p_{ipt}^{Meat}$	**-0.90** (0.44)	1.18 (0.80)	-0.35 (0.42)	1.27 (0.63)	0.38 (0.51)	1.10 (1.04)	0.52 (0.42)	1.36 (28.9)
$\ln p_{ipt}^{Fish}$	-0.29 (0.36)	**-1.65** (0.73)	-0.45 (0.39)	0.23 (0.45)	-0.52 (0.67)	0.14 (1.09)	0.11 (0.39)	1.37 (10.4)
$\ln p_{ipt}^{FruitsVeg.}$	-0.02 (0.14)	0.00 (0.34)	**-0.31** (0.15)	0.44 (0.15)	-0.05 (0.29)	-0.24 (0.32)	0.03 (0.17)	1.44 (3.78)
$\ln p_{ipt}^{Dairy}$	-0.37 (1.01)	4.33 (3.02)	3.89 (0.83)	**-3.24** (1.39)	2.90 (1.91)	0.07 (1.57)	1.42 (0.94)	-5.79 (25.6)
$\ln p_{ipt}^{Starchyfood}$	0.62 (0.42)	-0.05 (0.67)	0.31 (0.23)	-0.17 (0.41)	**-1.15** (0.48)	1.70 (0.72)	-1.49 (0.37)	-1.29 (13.8)
$\ln p_{ipt}^{Junkfood}$	0.89 (0.53)	-0.85 (0.83)	-0.37 (0.69)	-0.26 (0.68)	-0.14 (0.82)	**-2.19** (0.96)	2.62 (0.60)	-1.61 (7.04)
$\ln p_{ipt}^{Fat}$	-0.27 (0.67)	2.38 (1.64)	-0.07 (0.55)	1.16 (1.06)	0.78 (0.76)	1.54 (1.42)	**-1.76** (0.83)	3.00 (7.63)
$\ln p_{ipt}^{Others}$	-0.00 (0.10)	-0.23 (0.32)	0.12 (0.22)	-0.21 (0.22)	0.08 (0.09)	-0.24 (0.22)	-0.51 (0.17)	**-0.60** (4.19)
$\ln y_{ipt}$	0.08 (0.09)	0.01 (0.20)	0.00 (0.09)	0.02 (0.09)	0.01 (0.10)	0.11 (0.21)	0.01 (0.10)	0.48 (2.71)
F Test (α_{ip}^{l})	22.43 (0.00)	10.34 (0.00)	20.30 (0.00)	8.33 (0.00)	10.40 (0.00)	6.39 (0.00)	16.64 (0.00)	1.59 (0.00)
Obs.	22064	22069	22064	22064	22056	22056	22056	22064
Sargan stat.(df)	7.50 (17)	15.37 (16)	2.52 (22)	5.53 (13)	1.79 (11)	4.53 (7)	13.53 (17)	0.01 (1)

Table 13: Fixed Effects - Instrumental Variables Estimation of Individual Price Elasticities

The results show that own price elasticities are between 0 and -3.24 for all food groups. All elasticities are significantly different from zero except for the category "others". Individual own-price elasticities are larger than household price elasticities only for fruits and vegetables, the "junk food" category and the fat category. Other elasticities are lower at the individual level than at the household level. This shows that it is important to take into account the intra-household distribution of food. These elasticities are actually crucial for policy implications aiming at reducing calories intake by modifying the system of prices. The more elastic product category is the one of the family of oils and butter, followed closely by the "junk food" category. The less elastic category of food products is the fish, dairy and starchy foods.

6 Simulations of Taxation Policies

Public policy intervention is often called upon to reduce the obesity epidemic. Let's now consider the simulation of a taxation policy and the question of evaluating its impact on obesity. The system of equations linking consumption to prices and body mass index to consumption for an individual p of household i at time t is given by:

$$\begin{aligned} b_{ipt} &= \eta_{ip} + \mu_c y_{ipt} + \xi_{ipt} \text{ (body mass index equation)} \\ y_{ipt} &= \sum\nolimits_{l=1}^{K} y_{ipt}^{l} \text{ (calories intake across categories)} \\ \ln y_{ipt}^{l} &= \alpha_{ip}^{l} + \sum\nolimits_{k=1}^{K} \beta_l^k \ln p_{ipt}^{k} + \varepsilon_{ipt}^{l} \text{ for } l = 1, .., K \text{ (calories demand equations)} \end{aligned}$$

where b_{ipt} is the individual body mass index, y_{ipt} is his total calories consumption, y_{ipt}^l is the calories consumption of each food group l, μ_c is the coefficient of transformation of calories into BMI and p_{ipt}^k is the price of calories of food group k for this individual. The body mass index equation gives how calorie consumption affects the BMI of individuals. The second equation is simply an accounting equation of calories across the K food categories. The calories demand equations explain how the prices of calories of the different food groups affect the individual demand for calories of each food category. Combing these equations, we obtain the following relationship between BMI and prices of food categories:

$$b_{ipt} = \eta_{ip} + \mu_c \left(\sum\nolimits_{l=1}^{K} \exp \left(\alpha_{ip}^{l} + \sum\nolimits_{k=1}^{K} \beta_l^k \ln p_{ipt}^{k} + \varepsilon_{ipt}^{l} \right) \right) + \xi_{ipt} \tag{15}$$

Given this structural equation, we can remark that a reduced form of the conditional expectation of BMI conditional on prices

$$E\left(b_{ipt} | p_{ipt}^1, .., p_{ipt}^k, .., p_{ipt}^K\right) = H_{ip}\left(p_{ipt}^1, .., p_{ipt}^k, .., p_{ipt}^K\right)$$

is of an unknown functional form because of the possible endogeneity of prices and because of unobserved individual heterogeneity in both the price elasticities and the energy elasticity. Non parametric instrumental variables estimation of such equation would require the need for instruments for prices that are not correlated with taste shocks ε_{ipt}^l and preferences α_{ip}^l and with unobserved shocks related to energy expenditure (ξ_{ipt}) and to the individual propensity to gain weight (η_{ip}).

Such estimation is unfeasible while the two step structural procedure implemented in sections 4 and 5 has the advantage of allowing unobserved heterogeneity in tastes and energy intakes to be accounted for with weaker assumptions on instrumental variables.

Given this system of equations, one can simulate the effect of some taxation policy on food demands and body mass index changes. Actually, these equations imply that the change of prices p_{ipt}^k leads to the following changes in BMI:

$$\begin{aligned}\Delta b_{ipt} &= \mu_c \Delta y_{ipt} = \mu_c \sum\nolimits_{l=1}^{K} \Delta y_{ipt}^l = \mu_c \sum\nolimits_{l=1}^{K} \Delta \left[\exp\left(\alpha_{ip} + \sum\nolimits_{l=1}^{K} \beta_l^k \ln p_{ipt}^k + \delta_l \ln y_{ipt} + \varepsilon_{ipt}^l\right)\right] \\ &= \mu_c \sum\nolimits_{l=1}^{K} \left[y_{ipt}^l \left[1 - \exp - \left(\sum\nolimits_{l=1}^{K} \beta_l^k \Delta \ln p_{ipt}^k\right)\right]\right]\end{aligned}$$

Given the parameter estimates, we can simulate the effects of some price increase or price decrease of different categories on the individual BMI of the population.

Simulations		% of Overweight			% of Obese		
		Before	After	Change	Before	After	Change
1: Junk Food: price increase of 10%							
	Children	9.81	7.06	-28.03%	2.17	1.62	-25.34%
	Adult Males	38.02	36.29	-4.55%	10.42	10.34	-0.01%
	Adult Females	25.06	23.44	-6.46%	11.16	9.60	-13.98%
2: Fruit and Vegetables: price decrease of 10%							
	Children	9.81	9.04	-7.85%	2.17	1.90	-12.44%
	Adult Males	38.02	36.60	-3.73%	10.42	9.14	-12.28%
	Adult Females	25.06	24.38	-2.71%	11.16	10.11	-9.41%
3: Junk Food (+10%) & Fruit and Vegetables (-10%)							
	Children	9.81	6.51	-33.64%	2.17	1.50	-30.88%
	Adult Males	38.02	34.68	-8.78%	10.42	9.26	-11.13%
	Adult Females	25.06	22.14	-11.65%	11.16	8.86	-20.61%

Table 14: Tax policy simulations

Table 14 shows the results of the simulations. It shows that taxing the "junk food" category actually reduces the prevalence of overweight and obesity dramatically. It is to be noted that this effect is due to the fact that the price elasticity of the "junk food" category is quite important and to the fact that this category is also providing a substantial share of total energy intakes by individuals. Note that this category of products is of high density in calories while providing cheap calories but that alternative cheap calories in less dense products are also available (pasta, rice, ..). Using dietary history interview data, Darmon et al. (2004) find that adjusting for energy intakes, energy-dense diets cost less than energy dilute diets. Taxing the junk food category only is thus important for consumers to switch from high to low density products without changing dramatically the price of calories.

Moreover, even if all own-price elasticities are negative, the effect of a price increase is not always to decrease total calories consumption because of the positive cross price elasticities that can generate some substitution between calories intake in some category by calories intakes with others. It is thus interesting to note that a price decrease of fruits and vegetables would not generate an increase of caloric intake but rather a decrease (thanks to substitution with other categories) and thus a decrease of BMI in the population or of the rate of overweight or obese people. Our simulations show that a price decrease of 10% of fruits and vegetables would involve a reduction of obesity of 9 to 12 percent for children and adults. A combination of a price increase of junk food and a price decrease of fruits and vegetables would not have very different results than a price decrease of fruits and vegetables except for adult women who would be much more responsive.

7 Conclusion

We have shown how household purchase data do allow to estimate individual food consumption taking into account precise demographic and anthropometric characteristics. This enables us to study the link between individual consumption and individual obesity. We have shown how we can still identify some interval of values for the effect of calories on BMI allowing for example for unobserved physical activity. Thanks to the use of two years of data, we are able to show statistically significant effects of food calories on the weight of individuals. We have shown that calories intakes do explain obesity and that heterogeneity is also very important in terms of the propensity to be obese and in terms of metabolic assimilation of calories. Then, we estimate price elasticities of calorie consumption per category of food. Public policy based on taxation aiming at reducing obesity or overweight is shown to be possible. Actually, price elasticities at the individual level are quite significant and taxing high density and cheap energy categories of food like the one usually said as "junk food" appears to be an effective way to change consumption patterns and reduce obesity and overweight.

8 Appendix

8.1 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 which consists of the following:

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}

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

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 food consumption of category k by the household 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 which it is observed and the sample for which it is not. 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 those of "similar" households for which it is observed. Minimizing the remaining heterogeneity by conditioning on as many variables as possible will make the 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 high dimension of W_{it}.

However, the previous conditional independence implies that (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 rely on the conditioning on a unidimensional variable which is the propensity score and thus solve the dimensionality problem of conditioning on a large set of variables.

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 equal to p ($\in [0,1]$) with the average household consumption of category k food products with propensity score p.

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

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

Therefore, 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.

One difficulty is that if $W_{it} = \left(W_{it}^1, .., W_{it}^L\right)$, the assumption

$$\omega_{it}^k \perp S_{it}^k | W_{it}^1, .., W_{it}^L$$

does not imply that

$$\omega_{it}^k \perp S_{it}^k | W_{it}^1, .., W_{it}^{L-1}$$

Thus, reducing the dimensionality of the conditioning set may lead to inconsistencies in the estimates.

Assuming that W_{it}' is a sub-set of the vector W_{it}, one can also use the fact that

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

then

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

and

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

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

We also have that

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

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.

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 in the gender and activity status of the household member doing most food purchases, indicators of the 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, and 8 dummy variables for urban, rural and municipality population size.

The estimations of the propensity score was done using a probit model, they are not shown for brevity but are available upon request.

8.2 Other Tables and Graphs

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

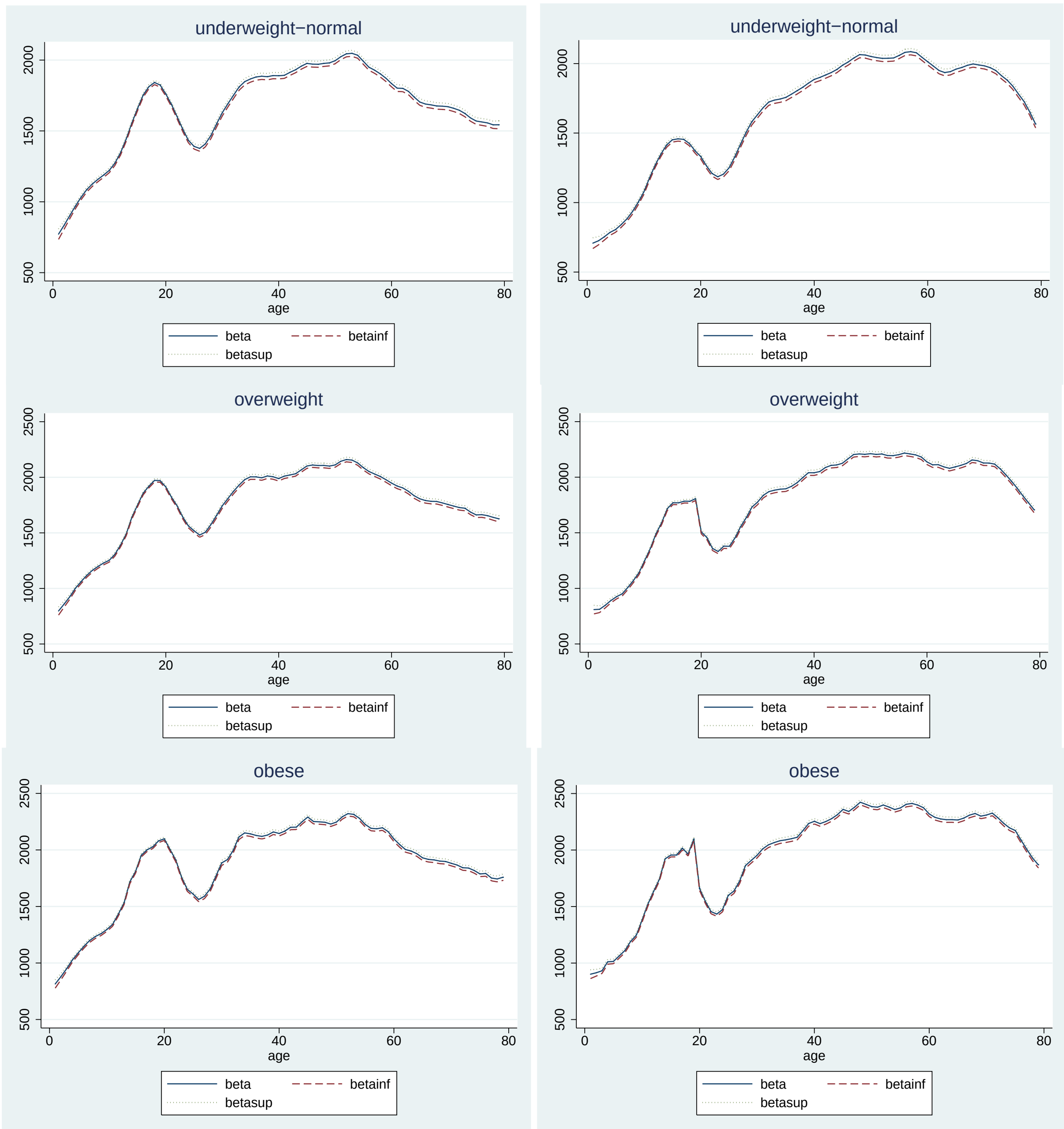

Figure 3: Estimated individual calories consumption by category per day

OLS	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat
$\ln p_{it}^{Meat}$	**-0.39 (0.02)**	-0.19 (0.03)	0.07 (0.02)	-0.02 (0.02)	-0.08 (0.02)	-0.05 (0.01)	0.10 (0.024)
$\ln p_{it}^{Fish}$	0.18 (0.02)	**0.22 (0.02)**	-0.09 (0.01)	0.03 (0.01)	-0.03 (0.015)	-0.04 (0.011)	0.08 (0.02)
$\ln p_{it}^{FruitsVeg.}$	-0.03 (0.025)	-0.01 (0.04)	**-0.32 (0.02)**	0.01 (0.02)	0.02 (0.023)	-0.25 (0.02)	0.08 (0.03)
$\ln p_{it}^{Dairy}$	-0.27 (0.03)	0.1 (0.048)	0.3 (0.029)	**-0.39 (0.029)**	-0.04 (0.032)	-0.17 (0.02)	0.02 (0.04)
$\ln p_{it}^{Starchyfood}$	-0.26 (0.02)	-0.19 (0.031)	0.09 (0.019)	-0.12 (0.019)	**-0.61 (0.02)**	-0.01 (0.015)	-0.45 (0.02)
$\ln p_{it}^{Junkfood}$	0.03 (0.03)	-0.01 (0.04)	-0.26 (0.024)	-0.39 (0.02)	-0.43 (0.03)	**-0.77 (0.02)**	-0.56 (0.032
$\ln p_{it}^{Fat}$	-0.19 (0.02)	-0.19 (0.03)	-0.1 (0.02)	0.12 (0.015)	0.09 (0.02)	0.09 (0.01)	**-0.44 (0.02**
$\ln p_{it}^{Others}$	-0.08 (0.02)	0.12 (0.03)	0.02 (0.01)	-0.22 (0.01)	-0.17 (0.02)	-0.27 (0.01)	-0.031 (0.02
$\ln y_{it}$	1.2 (0.01)	1.1 (0.02)	0.73 (0.01)	1 (0.01)	0.94 (0.01)	1.1 (0.01)	1.1 (0.01)
Const	5.6 (0.12)	1.4 (0.17)	6 (0.1)	6.8 (0.1)	7.5 (0.11)	8.9 (0.08)	6 (0.14)
N	8332	8332	8332	8332	8332	8332	8332
R^2	0.563	0.322	0.396	0.590	0.568	0.736	0.513

Table 15: OLS estimation of household price elasticities of calories by food group

FE	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat
$\ln p_{it}^{Meat}$	**-0.10 (0.03)**	0.16 (0.04)	-0.011 (0.02)	-0.03 (0.022)	-0.11 (0.03)	0.16 (0.02)	-0.04 (0.0
$\ln p_{it}^{Fish}$	0.14 (0.02)	**0.36 (0.02)**	-0.05 (0.01)	0.02 (0.012)	-0.032 (0.016)	-0.03 (0.01)	0.12 (0.0
$\ln p_{it}^{FruitsVeg.}$	-0.01 (0.03)	0.02 (0.05)	**-0.27 (0.03)**	0.14 (0.02)	0.09 (0.03)	-0.08 (0.02)	0.03 (0.0
$\ln p_{it}^{Dairy}$	-0.24 (0.04)	0.06 (0.06)	0.32 (0.04)	**0.38 (0.03)**	-0.04 (0.042)	-0.07 (0.03)	0.29 (0.0
$\ln p_{it}^{Starchyfood}$	-0.07 (0.03)	-0.03 (0.04)	0.14 (0.02)	0.02 (0.021)	**-0.31 (0.03)**	-0.005 (0.02)	-0.24 (0.0
$\ln p_{it}^{Junkfood}$	0.11 (0.04)	-0.14 (0.05)	-0.16 (0.03)	-0.35 (0.029)	-0.2 (0.04)	**-0.53 (0.03)**	-0.42 (0.0
$\ln p_{it}^{Fat}$	-0.01 (0.02)	-0.05 (0.035)	-0.08 (0.02)	-0.003 (0.018)	-0.04 (0.02)	-0.03 (0.02)	**-0.3 (0.0**
$\ln p_{it}^{Others}$	-0.05 (0.02)	-0.10 (0.032)	-0.04 (0.02)	-0.088 (0.02)	-0.083 (0.02)	-0.11 (0.01)	-0.11 (0.0
$\ln y_{it}$	1.14 (0.04)	1.18 (0.05)	0.74 (0.03)	0.89 (0.03)	0.98 (0.03)	1.07 (0.02)	1.02 (0.0
Const	3.5 (0.28)	-0.59 (0.38)	5.5 (0.23)	4.3 (0.2)	5.8 (0.26)	6.3 (0.18)	5.2 (0.35
N	8332	8332	8332	8332	8332	8332	8332
F Test ($\alpha_i = 0$)	6.3	6.7	6.6	8.9	6.4	7.7	4.6

Table 16: Fixed effects estimation of household price elasticities of calories by food group

OLS	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
$\ln p_{ipt}^{Meat}$	**0.03 (0.01)**	0.20 (0.01)	0.05 (0.01)	0.01 (0.00)	0.05 (0.00)	0.06 (0.00)	0.10 (0.01)	0.04 (0.00)
$\ln p_{ipt}^{Fish}$	-0.06 (0.01)	**-0.04 (0.01)**	0.12 (0.01)	-0.01 (0.00)	0.09 (0.00)	0.01 (0.01)	-0.05 (0.01)	-0.02 (0.00)
$\ln p_{ipt}^{FruitsVeg.}$	-0.56 (0.01)	-0.38 (0.01)	**-0.33 (0.01)**	-0.21 (0.00)	-0.40 (0.01)	-0.60 (0.01)	-0.57 (0.01)	-0.45 (0.01)
$\ln p_{ipt}^{Dairy}$	0.08 (0.02)	2.39 (0.03)	1.95 (0.02)	**0.14 (0.01)**	-0.21 (0.02)	-0.57 (0.02)	0.44 (0.03)	-0.61 (0.02)
$\ln p_{ipt}^{Starchyfood}$	0.73 (0.01)	-0.75 (0.02)	-0.16 (0.01)	-0.08 (0.01)	**0.37 (0.01)**	1.12 (0.01)	0.85 (0.02)	0.43 (0.01)
$\ln p_{ipt}^{Junkfood}$	1.54 (0.02)	1.82 (0.03)	1.03 (0.02)	0.47 (0.01)	0.74 (0.01)	**0.38 (0.02)**	1.00 (0.02)	0.92 (0.02)
$\ln p_{ipt}^{Fat}$	-1.59 (0.02)	-1.66 (0.02)	-0.73 (0.02)	-0.19 (0.01)	-0.54 (0.01)	-0.27 (0.02)	**-2.40 (0.02)**	-0.64 (0.01)
$\ln p_{ipt}^{Others}$	0.12 (0.00)	0.34 (0.01)	0.21 (0.00)	0.08 (0.00)	-0.07 (0.00)	-0.32 (0.00)	0.11 (0.01)	**0.37 (0.00)**
Constant	3.69 (0.05)	-5.13 (0.06)	-2.77 (0.04)	4.65 (0.03)	5.11 (0.03)	8.02 (0.05)	5.40 (0.05)	5.22 (0.04)
Obs.	22473	22473	22473	22473	22473	22473	22473	22473
R^2	0.78	0.86	0.91	0.49	0.61	0.55	0.68	0.79

Table 17: OLS estimation of individual price elasticities of calories by food group

FE	Meat	Fish	Fruits and Vegetables	Dairy Products	Starchy Food	Junk Food	Fat	Others
$\ln p_{ipt}^{Meat}$	**-0.03 (0.00)**	0.08 (0.00)	0.02 (0.00)	-0.02 (0.00)	0.01 (0.00)	-0.10 (0.00)	-0.02 (0.00)	0.02 (0.00)
$\ln p_{ipt}^{Fish}$	-0.04 (0.00)	**0.04 (0.01)**	0.06 (0.00)	0.02 (0.00)	0.04 (0.00)	-0.03 (0.01)	-0.01 (0.00)	-0.05 (0.00)
$\ln p_{ipt}^{FruitsVeg.}$	-0.35 (0.01)	-0.23 (0.01)	**-0.15 (0.01)**	-0.14 (0.01)	-0.22 (0.01)	-0.4 (0.01)	-0.26 (0.01)	-0.33 (0.01)
$\ln p_{ipt}^{Dairy}$	0.12 (0.03)	1.1 (0.04)	1.25 (0.02)	**0.36 (0.02)**	-0.04 (0.02)	-0.38 (0.03)	0.03 (0.03)	0.61 (0.02)
$\ln p_{ipt}^{Starchyfood}$	0.58 (0.01)	-021 (0.02)	0.09 (0.01)	0.17 (0.01)	**0.27 (0.01)**	0.71 (0.02)	0.47(0.01)	0.27 (0.01)
$\ln p_{ipt}^{Junkfood}$	0.23 (0.02)	1.03 (0.02)	0.14 (0.01)	-0.19 (0.01)	0.41 (0.01)	**0.28 (0.02)**	0.25 (0.02)	-0.12 (0.01)
$\ln p_{ipt}^{Fat}$	-0.79 (0.02)	-0.41 (0.03)	-0.24 (0.02)	-0.49 (0.01)	-0.54 (0.01)	-0.32 (0.02)	**-0.77 (0.02)**	-0.45 (0.02)
$\ln p_{ipt}^{Others}$	-0.03 (0.00)	0.03 (0.01)	0.05 (0.00)	-0.02 (0.00)	-0.03 (0.00)	-0.16 (0.00)	-0.07 (0.00)	**0.13 (0.00)**
Constant	5.87 (0.08)	-1.60 (0.11)	0.67 (0.06)	5.96 (0.05)	5.21 (0.05)	8.12 (0.09)	6.44 (0.07)	4.89 (0.06)
Obs.	22473	22473	22473	22473	22473	22473	22473	22473
F Test ($\alpha_i = 0$)	44.42 (0.00)	36.57 (0.00)	54.73 (0.00)	41.89 (0.00)	42.53 (0.00)	24.51 (0.00)	63.23 (0.00)	37.7 (0.00)

Table 18: Fixed effect estimation of individual price elasticities of calories by food group

8.3 Identification results

Remark first that the proof is complete using the following result when $cov\left(\xi_{ipt}, u_{ipt}\right) > 0$:

$$\frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{var(\widehat{y}_{ipt})} = \left[1 - \frac{var\left(\varsigma_{ipt}\right)}{var(\widehat{y}_{ipt})}\right] \frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{cov(\widehat{y}_{ipt}, \widetilde{y}_{ipt})} \leq \frac{cov\left(b_{ipt}, \widehat{y}_{ipt}\right)}{cov(\widehat{y}_{ipt}, \widetilde{y}_{ipt})} = \mu - \frac{cov\left(\xi_{ipt}, u_{ipt}\right)}{var(\widehat{y}_{ipt}) - var(\varsigma_{ipt})} \leq \mu$$

Then, when there are no measurement errors ($var(\varsigma_{ipt}) = 0$), then $\frac{cov(b_{ipt},\widehat{y}_{ipt})}{var(\widehat{y}_{ipt})} = \frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})}$.and:

$cov\left(\xi_{ipt}, u_{ipt}\right) = 0$	$\mu = \frac{cov(b_{ipt},\widehat{y}_{ipt})}{var(\widehat{y}_{ipt})}$
$cov\left(\xi_{ipt}, u_{ipt}\right) > 0$	$\max\left[\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{var(\widetilde{y}_{ipt})}, \frac{cov(b_{ipt},\widehat{y}_{ipt})}{var(\widehat{y}_{ipt})}\right] \leq \mu \leq \frac{cov(b_{ipt},\widetilde{y}_{ipt})}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}$
$cov\left(\xi_{ipt}, u_{ipt}\right) < 0$	$\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{var(\widetilde{y}_{ipt})} \leq \mu \leq \min\left[\frac{cov(b_{ipt},\widehat{y}_{ipt})}{var(\widehat{y}_{ipt})}, \frac{cov(b_{ipt},\widetilde{y}_{ipt})}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}\right]$

If measurement errors are such that $0 < var(\widehat{y}_{ipt}) < var(\varsigma_{ipt})$, then :

$cov\left(\xi_{ipt}, u_{ipt}\right) = 0$	$\mu = \frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})}$
$cov\left(\xi_{ipt}, u_{ipt}\right) > 0$	$\max\left[\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{var(\widetilde{y}_{ipt})}, \frac{cov(b_{ipt},\widetilde{y}_{ipt})}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}, \frac{cov(b_{ipt},\widehat{y}_{ipt})}{var(\widehat{y}_{ipt})}\right] \leq \mu \leq \frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})}$
$cov\left(\xi_{ipt}, u_{ipt}\right) < 0$	$\max\left[\frac{cov(b_{ipt},\widetilde{y}_{ipt})}{var(\widetilde{y}_{ipt})}, \frac{cov(b_{ipt},\widehat{y}_{ipt})}{cov(\widehat{y}_{ipt},\widetilde{y}_{ipt})}, \frac{cov(b_{ipt},\widetilde{y}_{ipt})}{cov(\widetilde{y}_{ipt},\widehat{y}_{ipt})}\right] \leq \mu$

References

Abay A., (2006) “Do government Food Prices Policies Affect the Prevalence of Obesity? Empirical Evidence from Egypt” IFPRI, Washington DC, USA

Allais O. and J. Tressou (2005) "Risque alimentaire et consommation de produits de la mer : évaluation de l’exposition individuelle de long terme au méthylmercure en France", *INRA Sciences Sociales* 2005

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

Anderson, P. M., K. F. Butcher and P. B. Levine (2003a) “Economic Perspectives on Childhood Obesity”, *Economic Perspectives*, 3Q, 30-48.

Anderson, P. M., K. F. Butcher and P. B. Levine (2003b) "Maternal Employment and Overweight Children”, *Journal of Health Economics*, 22, 477-504.

Asfaw A. (2006) "The Role of Food Price Policy in Determining the Prevalence of Obesity: Evidence from Egypt", *Review of Agricultural Economics*, 28, 3, 305-312

Bhattacharya J. and M. K. Bundorf (2005) "The Incidence of the Healthcare Costs of Obesity" NBER Working Paper 11303

Bhattacharya J., J. Currie, and S. Haider (2004) "Poverty, food insecurity, and nutritional outcomes in children and adults" *Journal of Health Economics,* 23, 839–862

Blundell, R., M. Browning and C. Meghir (1994) “Consumer Demand and the Life-Cycle Allocation of Household Expenditures”, *Review of Economic Studies* 61, 57-80.

Boizot-Szantaï C. and F. Etilé (2005) "The Food Prices / Body Mass Index Relationship: Theory and Evidence from a Sample of French Adults", *mimeo*

Bonnet C., P. Dubois and V. Orozco (2009) "Household Food Consumption, individual Caloric Intake and Obesity in France", Working Paper 542, IDEI, Toulouse School of Economics

Brownell K., Jacobson M. (2000) "Small taxes on soft drinks and snack foods to promote health", *American Journal of Public Health*,

Caraher M. and G. Cowburn (2005) "Taxing Food: Implications for Public Health Nutrition", *Public Health Nutrition*, 8(8), 1242-1249

Cash S.B., Sunding D. L. and Zilberman D. (2005) "Fat taxes and thin subsidies: prices, diet, and health outcomes", Acta Agriculturae Scand. Section C: *Food Economics*, 2:167-174.

Cawley J. (2006) "Contingent Valuation Analysis of Willingness to Pay to reduce Childhood Obesity". NBER WP 12510.

Chesher A. (1997) "Diet Revealed?: Semiparametric Estimation of Nutrient Intake - Age Relationships", *Journal of the Royal Statistical Society. Series A (Statistics in Society), Vol. 160, No. 3. p. 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

Chou S., I. Rashad, M. Grossman (2005) "Fast-Food Restaurant Advertising on Television and Its Influence on Childhood Obesity", NBER Working Paper 11879

Chou S., M. Grossman, H. Saffer (2004) "An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System", *Journal of Health Economics* 23 (2004) 565–587

Chouinard H. H., D.E. Davis, J.T. La France and J.M. Perloff (2005) "The effects of a fat tax on dairy products", WP University of California, Berkeley.

Classen T., C. Hokayem (2005) "Childhood influences on youth obesity", *Economics and Human Biology*, 3, 165-187

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

Cutler D. And E. Glaeser (2005) "What explains differences in smoking, drinking and other health-related behaviors?", mimeo Harvard University

Darmon N., A. Briend and A. Drewnowski (2004) "Energy-dense diets are associated with lower diet costs: a community study of French adults" *Public Health Nutrition* , 7:21-27 Cambridge University Press

De Agostini P. (2005) "The relationship between food consumption and socioeconomic status: Evidence among British youths", ISER working paper

Emery et al. (2007) "Cost of obesity in France" *Presse Médicale* 36:832-40, June 2007.

Epstein L.H. , K.K. Dearing, R.A. Paluch, J.N. Roemmich, D. Cho (2007) "Price and maternal obesity influence purchasing of low- and high-energy-dense foods", *American Journal of Clinical Nutrition*, Volume 86, Pages 914-922

Gandal N. and A. Shabelansky (2009) "Obesity and Price Sensitivity at the Supermarket", *mimeo* Tel Aviv University

Grossman M. and I. Rashad (2004) “The Economics of Obesity”, *The Public Interest*, 156, 104-112

Holt D., Ohlhausen M. and Zywicki T. (2004) "Obesity and advertising policy", George Mason law and economic research paper n°45

Kucheler F., Abebayehu T., Harris J. M. (2004) "Taxing snacks foods: manipulating diet quality or financing information programs?" *Review of Agricultural Economics*, vol. 27

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*, May 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

Mytton O., A. Gray, M. Rayner, H. Rutter (2007) "Could targeted food taxes improve health?" *Journal of Epidemiology and Community Health* 61:689-694

Nelson M. , Dyson P. A., Paul A. A. (1985) "Family food purchases and home food consumption: comparison of nutrient contents" *British Journal of Nutrition*, vol. 54, no2, pp. 373-387

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* Vol. 124 No. 9_Suppl September 1994, pp. 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 W7423.

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

Rashad, I. (2006) "Structural Estimation of Caloric Intake, Exercise, Smoking, and Obesity", NBER Working Paper 11957

Rosenbaum, P.R. and D.B. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 1, 41-55.

Smith T.G. (2002) "The MacDonald's equilibrium: advertising, empty calories and the endogenous determination of dietary preferences", *mimeo* UCSB.

Smith T.G. (2003) "Obesity and Nature's thumbprint: how modern waistlines can inform economic theory", mimeo.

Smith P., Bogin B. and D. Bishai (2005) "Are time preference and body mass index associated? Evidence from the national longitudinal survey of youth", *Economics and Human Biology*, 3:259-270.

Srinivasan C.S. and I. X. Shankar (2006) "An Assessment of the Potential Consumption Impacts of WHO Dietary Norms in OECD countries", *Food Policy*, 31: 53-77.

Stewart H. and J.M. Harris (2004) "Obstacles to overcome in promoting dietary variety: the case of vegetables", *Review of Agricultural Economics*, Vol 27, 1,

Sunder M. (2005) "Toward generation XL: Anthropometrics of longevity in late 20th-century United States", *Economics and Human Biology*, 3, 271-295

Vasdekis, V.G.S. and A. Trichopoulou (2000). Non parametric estimation of individual food availability along with bootstrap confidence intervals in household budget surveys. *Statistics and Probability Letters* 46, 337–345.

Variyam, J. N. and J. Cawley (2006) "Nutrition Labels and Obesity", NBER Working Paper 11956

Wansink B. and P. Chandon (2006) "Can low fat nutrition labels lead to obesity", INSEAD WP 2006/25/MKT.

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