Food Price Policies and the Distribution of the Body Mass Index

Fabrice Etilé*

(INRA-ALISS and Paris School of Economics)

3rd February 2009

Abstract

This paper uses French food-expenditure data to examine the effect of the local prices of 23 food-product categories on the distribution of Body Mass Index (BMI) in a sample of French adults. It is shown that the identification of price effects is in general not granted when individual physical activity is unobserved. However, identification is possible, when body weight is stationary, and the prices are orthogonal to both physical activity and the control variables correlated with the latter. Using quantile regressions, unconditional BMI distributions can then be simulated for various price policies. In the preferred scenario, increasing the price of alcohol, soft drinks, breaded proteins, pastries and desserts, snacks, and ready meals by 10%, and reducing the price of fruit and vegetables in brine by 10% would reduce the prevalence of overweight and obesity by 3.8 and 3.9 percentage points respectively.

*Address: Fabrice Etilé, etile@ivry.inra.fr; INRA, UR 1303 - ALISS, 65 rue de Brandebourg, F-94205 Ivry-sur-Seine. I am grateful to Christine Boizot-Szantai for research assistance, and to Olivier Allais, Arnaud Basdevant, Andrew Clark, Pierre Dubois, Sébastien Lecocq, David Madden and Anne-Laure Samson for discussions and suggestions. I also thank seminar participants at the 2005 EAAE Congress (Copenhagen), INRA-IDEI (Toulouse), INRA-GAEL (Grenoble), the York Seminar in Health Econometrics, INRA-EC (Blois), JESF 2007 (Lille), SFER conference 2007 (Paris), the Erasmus School of Economics (Rotterdam), AUHE seminar (Leeds), and the 17th European Workshop on Health Economics and Econometrics (Coimbra, 2008) for helpful comments on various versions of this paper.

1 Introduction

In 2002, 37.5% of French adults were overweight and 9.4% were obese, compared to figures of 29.7% and 5.8% respectively in 1990 (OECD Health Data, 2005).[1] These trends have become a major public health concern, as reflected in the goal of the National Plan for Nutrition and Health (PNNS 2006-2010) to reduce the prevalence of adult obesity by 20%. In this perspective, this paper asks whether appropriate food price policies may help to attain this public health objective.[2]

The mechanism underlying the food price/body weight relationship is well known. When food prices are lower, ingesting food calories becomes cheaper, and calorie intake is likely to rise. Body weight then increases to restore the metabolic equilibrium between calorie intake and calorie expenditure. Cutler *et al.* (2003) note that the full price of calorie intake has fallen over the past forty years, because there has been a revolution in the mass preparation of food. The cost of primary food products, food preparation and processed food have declined while, owing to the rise in real wages, the opportunity cost of home cooking has risen. The consumption of mass produced food has increased, but technical progress has been biased in favour of energy-dense food. As a result, the cost of a healthy and well-balanced diet is now relatively much higher than that of an energy-dense diet (Drewnowski and Darmon, 2004). This suggests that taxing food calories to raise their price may change trends in overweight and obesity.

However, the main economic rationale for taxation would not be any public health goal, but rather the existence of externalities. In France, the medical cost of obesity was about 2.6 Billion Euros in 2002 and 3.6 Billion Euros in 2006 (Emery *et al.*, 2007; IGF-IGAS, 2008). Taxing food would further help solve the *ex ante* moral hazard problem that arises from the inability of Social Security to charge individuals fairly (Strnad, 2005). In this perspective the consumer is held responsible for his/her health, and the primary goal of the tax is to raise revenue. A smaller elasticity of food demand (and therefore of body weight) may even be desirable (Chouinard *et al.*, 2007). Public health concerns and economics can be reconciled if less classic normative goals are considered, for instance correcting the "internalities" that ensue from rationality failures, or from a Senian standpoint, making healthy food products (typically fruit and vegetables) more affordable, in order to help consumers to adjust their food habits. Here, the environment is held responsible, to some extent, not the consumer. The point of the policy is to change behaviour, which is feasible only if the elasticity of calorie demand is high enough, and taxes and subsidies are transmitted to consumer prices.

A tax on calories may not be very efficient, as monitoring costs would be substantial - the recipes of food producers change constantly -, and human energy requirements are heterogenous (Leicester and Windmeijer, 2004). As such, considering taxes or subsidies for specific product categories is more interesting. Taxes on confectionery, carbohydrated drinks or snacks already

[1] According to the World Health Organisation's international standards, a Body Mass Index (BMI: weight in kg divided by height in meters squared) over 25 signals overweight. Beyond 30, the individual is obese.

[2] Perhaps surprisingly, while public health authorities remain cautious about price policies, French consumer associations have taken a firm position in favour of a tax on snacks, carbohydrated drinks and confectionery (*Cf.* the lobbying campaign by the federation of consumer associations "UFC Que Choisir?" in favour of a nutritional VAT, www.quechoisir.org). In the U.S., consumer associations have traditionally been opposed to taxes (Sheu, 2006), while price interventions have long been advocated by the public health sector (see for instance Brownell K., 1994, "Get slim with higher taxes", New York Times, 15/12/1994, A29.).

exist in a number of U.S. states, although not for nutritional reasons (Jacobson and Brownell, 2000). Beyond these products, if the policy objective is to shift the BMI distribution to the left, then all energy-dense products might be covered *a priori* by a tax: soft drinks as well as foie gras, whatever their cultural legitimacy. As such, I here try to identify the effects of the prices of *23 food product categories,* which cover the entire diet*,* on the *BMI distribution* of French adults. Given any normative objective, the results may provide clues for the choice of a relevant tax base.[3]

Empirical work on the price-BMI relationship is relatively scarce. Using time and regional variations in food taxes in the US, Lakdawalla and Philipson (2002) find that the fall in food-at-home price explains 41% of the growth in BMI between 1981 and 1994. Chou *et al.* (2004) report a negative and significant correlation between the price of food-at-home and BMI in repeated cross-sections from the Behavioural Risk Factor Surveillance System. Sturm and Datar (2005) present evidence that lower fruit and vegetables prices predict smaller increases in body weight between the kindergarten and the third grade for American children. Asfaw (2006) relies on a single cross-section of a household survey to study the relationship between the prices of nine food groups and average BMI in Egyptian women. He finds, as expected, significant negative effects for energy-dense products, and positive effects for less-dense products. Using seven repeated cross-sections of the Monitoring The Future survey (1997-2003), Powell *et al.* (2007) report positive, albeit not significant, effects of the price of fruits and vegetables on the BMI of American adolescents. The current paper also uses repeated cross-sections and spatial price variation to identify the price-BMI relationship, but improves on previous work by considering the prices of all food product categories, in order to control the pattern of substitution between products.

I use the French TNS WorldPanel household survey, which provides socio-demographic information at the household level, and household scanner data of food-at-home expenditures throughout the year. The BMI of all household members was self-reported annually between 2002 and 2005; I focus here on adults. One challenge posed by the use of scanner data is that they do not provide truly exogenous prices, but rather unit values computed by dividing expenditures by quantities. These unit values are endogenous, as they reflect households' tastes for quality, which are unobserved and may be correlated with BMI. Empirical inference relies on the spatial price variations that are generated by the particular spatial structure of the French retail market. I adjust unit values for quality effects to construct exogenous local price indices that capture these variations.

An empirical specification, which links body weight to prices and income, is derived from the energy-balance equation, which states that body weight is an adjustment variable in the equilibrium between calorie intake and calorie expenditure. One important consequence of the energy-balance equation is that the coefficients on the right-hand side variables depend structurally on the individual Physical Activity Level (PAL) of the individual. Since the latter is unobserved, there is slope heterogeneity in the price-BMI relationship. It is shown, however, that price elasticities are identified as long as PAL and food prices are independent conditional on the other control variables, especially income, and food prices are also independent of these

[3] By working at a relatively disaggregated level, I hope to identify a feasible price intervention, since opposition from numerous pressure groups may be encountered.

control variables. These conditions are met here, because the prices are constructed by adjusting unit values for the effects of a number of variables, including income.

Price elasticities of the *whole* BMI distribution are then estimated by quantile regressions, and various price scenarios are simulated by using a counterfactual method. In my preferred scenario I find that increasing the prices of alcohol, soft drinks, breaded proteins, pastries and desserts, snacks, and ready meals by 10%, and reducing the prices of fruit and vegetables in brine by 10% would reduce the prevalence of overweight and obesity by 3.8 and 3.9 percentage points respectively, and the medical cost of obesity by 960 Million Euros. Although there is some statistical fragility in the quantile estimates, I show that these results are fairly robust to misspecification biases.

The remainder of the paper is structured as follows. Section 2 presents the data. Section 3 explains in detail how prices are constructed. Section 4 sets up the empirical framework and discusses identification. Section 5 reports the main results, and Section 6 simulates the impact of several price scenarios. Section 7 then checks the robustness of the results, and Section 8 concludes.

2 Data

I use five waves of data drawn from the TNS French household panel survey (2001-2005), which is nationally representative. This has a number of specific features that to an extent limit the empirical analysis. Each year, up to 8000 households are observed. Each household leaves the panel after four years, during which all purchases with a barcode are recorded.[4] Additional information is provided on the products' labelled characteristics. For instance, the fat content of cheese is specified, but not the calorie content of a ready meal. For the purchases of products without barcodes (e.g. fresh meat bought at the butchers), the panel is split into two sub-panels. The first sub-panel is dedicated to fresh meat and fish, and the second to fresh fruit and vegetables. Hence, information on household purchases is not exhaustive. Individual food consumption and food-away-from-home intakes are not observed.

The causal structural chain that connects prices to BMI is as follows. First, households and/or individuals buy products for food-at-home or food-away-from-home consumption. Food prices play a role at this stage. Second, purchases for food-at-home are shared between household members. A large part is consumed, and the remainder is wasted. Third, individual consumption is converted into calories. Then, given a level of physical activity, calorie intake determines body weight. Obviously, the structure of the TNS data set does not allow us to identify the structural chain which links prices to household purchases, individual consumption, calories and ultimately weight, without strong statistical assumptions and without ignoring food-away-from-home. As a consequence, this paper focusses directly on the relationship between food prices and *individual* BMI.

The starting sample (Sample 1: $N = 21407$ individual-year observations) consists of observations without missing values, and for which it was possible to assign food prices. I also drop observations in the first and 99^{th} percentiles of the BMI distribution for robustness. Descriptive

[4] The barcodes themselves (Universal Product Codes) are not provided with the data.

statistics for all variables are presented in Appendix B, Table B1. These statistics, *as are all those in the paper* (including the regression results and the simulations), have been adjusted for yearly sampling weights at the household level.[5]

2.1 Body Mass Index

From 2002 to 2005, the BMIs of all household members were self-reported. I am not able to correct for declaration biases, as correction equations that are valid for the whole population are not available.[6] While overweight and obesity are probably underestimated, this may be less of a concern here than, for instance, in the OECD Health Data (2005). According to the latter, there were 9.4% of obese adults in France in 2002. In Sample 1, the corresponding figure is 11.1%. The sample prevalence of overweight is 44.1%, as against 37.5% in the OECD data. The left side of Figure B1 in Appendix B plots the distribution of the BMI in Sample 1. This distribution is not Gaussian according to standard statistical tests. The skewness is strongly positive, since the distribution has an elongated right tail. Applying a logarithmic transformation does not eliminate skewness. The empirical estimation takes this issue seriously by using quantile regressions.

Last, body weight remains stable for 86.5% of those individuals who can be followed over two consecutive years (14576 transitions are observed). This stability will be exploited in the econometric analysis.

2.2 23 product categories

Purchases are first classified into food products, the definition of which takes into account nutritional information that is labelled, and therefore available to the consumer. For instance, there is a distinction between mid-fat Brie cheese (fat content between 30 and 59%) and full-fat Brie cheese (fat content over 60%). Likewise, I distinguish diet/light sodas from standard sodas. Overall, there are 340 food products.

Food products are then sorted into 23 product categories, for which I would like to construct exogenous prices: mineral water, alcohol, soft drinks, vegetables in brine, fruit in brine, processed vegetables, processed fruit, cereals, meat in brine and eggs, seafood in brine, processed seafood, cooked meat, breaded proteins, yoghurt and fresh uncured cheese, cheese, milk, animal fats and margarine, oils, sugar and sweets, pastries and desserts, sweet and fatty snacks including breakfast cereals, salty and fatty snacks, and ready meals. Each category is made up of between

[5] These weights are rescaled to sum to the yearly number of observations, and therefore account for their relative representativity.

[6] Body weight is measured with errors that can be decomposed into two parts. First, there are deliberate declaration biases. It has been found in a company cohort of French middle-aged subjects that weight is systematically underestimated and height is systematically overestimated, leading to an underestimation of BMI that is larger for women ($-0.44\ kg/m^2$) than for men ($-0.29\ kg/m^2$). Overweight status, age, education and occupation are significantly correlated with this declaration bias (Niedhammer *et al.*, 2000). Second, there are errors due to rounding to the nearest integer value, heaping and digit preferences. Beyond these reporting biases, BMI is not an accurate measure of fatness, and misclassifies some individuals as obese, overweight or non-overweight. Burkhauser and Cawley (2008) show that it would be preferable to use better measures of obesity that distinguish fat from fat-free mass. Unfortunately, such accurate measure of fatness are not yet available in consumer survey data.

1 and 77 food products. Appendix B, Table B2, provides more details on the food categories with some examples.

This classification is intended to depict collective representations of food products' healthiness, as any taxation policy will require some support from public opinion. This concern, as well as advice from health professionals, leads me to differentiate breaded meat and cooked meat from meat in brine, processed sea products from sea products in brine, and fruit or vegetables in brine (even canned or frozen) from fruit and vegetables that are prepared with additives, such as fats or syrups. Some food products are not classified in their "natural" category, when their nutritional quality may have been profoundly altered by the production process. For example, breakfast cereals are considered as sweet and fatty snacks, and not as cereals. The next section explains how I construct prices for these 23 categories.

2.3 Control variables

A number of socio-demographic effects and potential confounders will be controlled for in the analysis: income,[7] age, household structure, education (which arguably proxies information effects), the home-production of fruit and vegetables, recent pregnancy/birth in the household, and responsibility for household food expenditure (the meal planner may be better able to control her or his weight through food choices if she or he is not prone to impulse buying). Last, regional and time differences in tastes are controlled for via a set of dummies for region (which groups together several "departements"), the type of residential area, and the calendar year.

3 The construction of local food prices

By dividing, for each household and for any given food product, yearly expenditure by the quantity purchased, a household-specific unit value can be constructed. Unfortunately, unit values are not exogenous, as they also reflect households' tastes for quality. We can imagine that households with higher average BMIs are more likely to buy, in a given food category, products that are more energy-dense, and the latter generally have lower unit values. To construct exogenous prices from unit values, I first suppose that the law of one price holds at the level of spatio-temporal clusters c (following Deaton, 1988). They are defined as follows: two households belong to the same cluster if their purchases are observed over the same calendar year t, and they live in the same or adjacent "departement" (roughly the size of a US county), and the same or similar type of residential area.[8] This paper therefore relies on spatial and time variations in prices to identify the price-BMI relationship.

[7] Income is equivalenced and deflated by the yearly Consumer Price Index provided by the National Statistical Office (INSEE) for households, according to their position in the income distribution (reference: 2004 Euros). Unit values were also deflated by this CPI before being used for the construction of the price indices.

[8] There are 94 departements in Metropolitan France (Corsica is not covered by the survey), and each departement has between two and nine neighbours. There are eight types of residential area, from "rural" to "urban units with more than 20000 inhabitants (excluding Greater Paris)" and "Greater Paris". These residential area are ranked according to their size so that it is easy to define closeness. For instance, in a given year t, a household living in a urban unit of between 2000 and 4999 inhabitants is close to households in the same or adjacent departements who live in urban units of between 5000 and 9999 residents or in rural areas. These belong to the same cluster c. Lecocq and Robin (2005) use the same criteria.

A number of authors then construct cluster-specific prices by computing cluster-averages of unit values (see, for instance, Lecocq and Robin, 2005; Asfaw, 2006). However, this requires the undesirable assumption that the distribution of tastes be similar across clusters, so that between-cluster differences in average unit values reflect only differences in supply prices. I therefore implement a second approach, which involves two steps. First, household Paasche price indices at the level of food categories are constructed from unit values computed for food products. Second, for each food category, I regress the price index on observable household characteristics that are likely to capture quality effects, and a set of cluster fixed effects. The latter represent my measure of local prices.

3.1 Procedure

3.1.1 Food categories as Hicks aggregates

The data set provides the quantities q_{lj}^{hc}, and unit values ν_{lj}^{hc} associated with the yearly expenditure of household h in cluster c on food product j in category l. Each food product aggregates items of different qualities, which are unobserved. Following Deaton (1988), I assume that the relative prices of items within each product category l are fixed everywhere. Hence, product categories are treated as Hicks aggregates.[9] As such, if $\mathbf{p}_l$ is the vector of unobserved prices for items in product category l, there exists a scalar λ_l^c such that $\mathbf{p}_l = \lambda_l^c \mathbf{p}_l^0$, where $\mathbf{p}_l^0$ is the relative-price structure. Denote by λ_l^c the linear homogeneous price level for category l in cluster c, so that differences in λ_l^c between clusters reflect spatial and time heterogeneity in supply prices. The goal here is to construct a measure of λ_l^c.

Let $\mathbf{p}_{lj}$ and $\mathbf{p}_{lj}^0$ be vectors extracted from $\mathbf{p}_l$ and $\mathbf{p}_l^0$ that collect together the prices of all different qualities of food product j, and $\mathbf{q}_{lj}^{hc}$ be the corresponding vector of unobserved quantities purchased by household h. The average unit value of food product j in category l for household h is:

$$\nu_{lj}^{hc} = \frac{\mathbf{p}_{lj}\prime \mathbf{q}_{lj}^{hc}}{\mathbf{1}\prime \mathbf{q}_{lj}^{hc}} = \lambda_l^c \frac{\mathbf{p}_{jl}^0\prime \mathbf{q}_{lj}^{hc}}{\mathbf{1}\prime \mathbf{q}_{lj}^{hc}} \tag{1}$$

where all vectors are column vectors, $\mathbf{x}\prime$ is the transpose of $\mathbf{x}$, $\nu_{lj}^{hc,0} = \frac{\mathbf{p}_{lj}^0\prime \mathbf{q}_{lj}^{hc}}{\mathbf{1}\prime \mathbf{q}_{lj}^{hc}}$ can be considered as a quality index (Deaton, 1988), $\mathbf{1}$ is a vector of 1's, and $\mathbf{1}\prime.\mathbf{q}_{lj}^{hc} = q_{lj}^{hc}$.

3.1.2 Local Paasche price indices

In order to weight the unit values by the household's structure of consumption, local Paasche indices are computed at the level of each category for each household:

$$P_l^{hc} = \frac{\sum_{j=1}^{J_l} q_{lj}^{hc} \nu_{lj}^{hc}}{\sum_{j=1}^{J_l} q_{lj}^{hc} \nu_{lj}^0} \tag{2}$$

where J_l is the number of food products in l and ν_{lj}^0 is a reference unit value for food product j. Here, the reference unit values are average unit values for purchases made in 2004 in Paris and surrounding departements and, accordingly, the price levels in this cluster are normalised, *i.e.* $\lambda_l^0 = 1$.

[9] The treatment of product categories as Hicks aggregates is fairly standard in consumption economics.

3.1.3 Adjusting prices for quality effects

Using (1) and (2):

$$\ln\left(P_l^{hc}\right) = \ln(\lambda_l^c) + \underbrace{\ln\left(\frac{\sum_{j=1}^{J_l} q_{lj}^{hc}\nu_{lj}^{hc,0}}{\sum_{j=1}^{J_l} q_{lj}^{hc}\mathbf{E}\left(\nu_{lj}^{h'c,0}\left|h' \in \{paris, 2004\}\right.\right)}\right)}_{\rho_l^{hc}} \tag{3}$$

where ρ_l^{hc} is a quality index for category l, and $\mathbf{E}(.\,|.)$ is the conditional expectation operator. Were food products to be perfectly homogenous in quality accross clusters and households, then $\rho_l^{hc} = 1$. Following Cox and Wohlgenant (1986), a widely-used method of constructing a proxy measure of $\ln(\lambda_l^c)$ is to specify ρ_l^{hc} as a function of a vector of observable variables $\mathbf{Z}^{hc}$ and an error-term $\widetilde{\mu}_l^{hc}$:

$$\ln(\rho_l^{hc}) = \kappa_l \mathbf{Z}^{hc} + \widetilde{\mu}_l^{hc} \tag{4}$$

implying:

$$\ln\left(P_l^{hc}\right) = \ln(\lambda_l^c) + \kappa_l \mathbf{Z}^{hc} + \widetilde{\mu}_l^{hc} \tag{5}$$

κ_l is estimated by an OLS regression of $\ln\left(P_l^{hc}\right)$ on $\mathbf{Z}^{hc}$ after a within-cluster transformation of (5). Then, $\ln(\lambda_l^c)$ is identified by computing the cluster mean of the residuals:

$$\widehat{\ln(\lambda_l^c)} = \mathbf{E}\left\{\ln\left(P_l^{hc}\right) - \widehat{\kappa}_l \mathbf{Z}^{hc}\left|h \in c\right.\right\} \tag{6}$$

This will be my price index for product category l. It is (up to an additive constant) an unbiased measure of $\ln(\lambda_l^c)$ as long as $\mathbf{E}\left\{\widetilde{\mu}_l^{hc}\left|h \in c, \mathbf{Z}^{hc}\right.\right\} = 0$: conditional on the control variables in $\mathbf{Z}$ (especially income and education), the average value of unobservable factors that affect quality choice must not systematically differ between clusters. This assumption is briefly discussed in Section 7.

The estimation of the quality effect in (5) controls for the following variables: real equivalent income; education, age and occupation of the meal planner; household structure; self-production of fruits and vegetables; region of residence; ownership of a micro-wave and a freezer, and the size of the latter. Since having a car and facing a larger food market potentially expand the choice set, I also introduce a dummy for car ownership, and a measure of the surface (in m^2) of supermarkets and hypermarkets in a radius of 20 km around the city of residence. The latter is constructed from exhaustive annual geocoded data.

The regression results show that unit values always depend positively on income (with elasticities between 0.065 for fresh fruits and 0.210 for sea products in brine). The characteristics of the meal planner, the size of the local retail market and the region of residence also have some influence.

3.2 Comments

3.2.1 Source of price variations

Descriptive statistics show that, in the estimation sample, between-individual standard deviations in prices are slightly higher than within-time standard deviations, so that the identification of price effects will rely more on spatial than time variation in prices. There are also very few outlying values, in the sense that the maxima and minima are generally close to the means ± two standard deviations.

A key question is whether the variance is produced by actual variations in supply prices. There is an ongoing debate over the level of retail prices in France, as compared to other EU countries. A number of reports have emphasised that appropriate zoning regulations would benefit consumers, by introducing more competition in local markets and thus lowering prices (see *inter alia*, Canivet, 2004). Descriptive work has shown that the structure of retail distribution is largely characterised by a lack of spatial competition. In about 60% of the 630 consumption areas, a single national retail group has more than 25% of the market share, with the second firm lying at least 15 points behind (ASTEROP, 2008). Analysis of the price of a well-defined consumer basket confirms that there are significant spatial variations in price, even for supermarkets belonging to the same retail group.[10] As a result, I suppose that the variance in food prices is largely due to the structure of the food retail market

3.2.2 Other remarks

A number of other comments are in order. First, there are potentially 3008 clusters in the analysis (94 departements times 8 types of residential area times four years). Clusters with less than 25 households were dropped from the sample for greater precision, and household sampling weights were used everywhere. Second, some indices can be computed in one sub-panel only. These prices are then imputed to individuals in the other sub-panel, by matching on the variables that identify clusters. Third, expenditure on food away-from-home is not observed and, therefore, its price can not be constructed. A number of papers have found empirical evidence of the role of the food-away-from-home sector in the U.S. (Chou *et al.*, 2004; Rashad *et al.*, 2006, Powell *et al.*, 2007). Since the prices of food-at-home are likely to be positively correlated with the prices of food away-from-home, and the effects of the latter and the former on BMI are likely to work in the same direction, the price elasticities may be biased away from zero.[11]

4 Econometric modelling

This section presents the empirical model and discusses the identification of the price effects.

4.1 The energy balance equation

In Physiology, body weight is an adjustment variable in the balance equation between calorie intake and expenditure. This equation results from the law of conservation of energy applied to the human body: the net flow of energy intake, which equals the flow of energy supplied minus

[10] See the study by the consumer association "UFC Que-Choisir?", published in the magazine *Que Choisir?*, 455, January 2008. As an illustration, compared to the national average, the price of a basket of national-brand products bought in a store owned by the retail group Carrefour is more expensive in the 14th district of Marseilles (south-east of France, +2.5%) and Drancy (near Paris, +0.6%), and cheaper in the 8th district of Marseilles (-3.5%) and Lille (North of France, -0.1%).

[11] There is a lack of published evidence on this point, with the exception of Sturm and Datar (2005). The magnitude of the effects they estimate for the prices of food-at-home products (fruit and vegetables, and meat) is very robust to the inclusion of fast-food prices.

the flow of energy expended, is stored in fat cells, which generates body weight variations (see, *inter alia*, Kozusko, 2001).

Let K be calorie intake (the flow of energy supplied). This is produced exclusively by food consumption. Following the Physiology literature, the flow of calorie expenditure is expressed as a multiple E (> 1) of the Basal Metabolic Rate (BMR), where E is a normalised index for physical activity level (see AFSSA, 2001; or UNU/WHO/FAO, 2004). Instantaneous changes in body weight W are described by a differential equation:

$$\dot{W} = \gamma \left[K - E * BMR \right] \tag{7}$$

where γ is a constant for the conversion of calories into Kgs per time unit. The World Health Organisation recommends specifying BMR as a linear function of weight:

$$BMR = \alpha + \beta W \tag{8}$$

where the parameters α and β depend on age and gender (UNU/WHO/FAO, 2004). For any well-defined physical activity (e.g. walking one hour at a speed of 3km/h), calorie expenditure $E * BMR$ increase with body weight. The energy-balance equation reflects the equilibrium between calorie intake and expenditure: $K = E * BMR = \alpha E + \beta E * W \iff \dot{W} = 0$. The stationary equilibrium weight, W^*, is then:

$$W^* = \frac{1}{\beta} \frac{K}{E} - \frac{\alpha}{\beta} \tag{9}$$

Since β is positive, W^* increases with calorie intakes and decrease with calorie expenditure. Relative trends in the full price of intake and expenditure therefore explain trends in the prevalence of overweight and obesity.[12]

4.2 Empirical specification

To derive an empirical specification for the price-BMI relationship, I assume that K in equation (9) for individual i is a linear semi-logarithmic function of income and food prices p_{lc}: $K_i = \sum_{l=1}^{L} \delta_l \ln(p_{l,c}) + \delta_I \ln(I_i) + \delta_0$, where I_i is the real income per unit of consumption, and $p_{l,c}$ is the price of food category l for i living in cluster c. Following Section 3, $\ln(p_{l,c})$ is measured by $\widehat{\ln(\lambda_l^c)}$.

Dividing each side of the equation by $H_i^2 = height^2$ (in meters squared), adding a set of control variables Z_i and a residual $\widetilde{\epsilon}_i$ then produces a specification for the body mass index, BMI_i^*:

$$BMI_i^* = \sum_{l=1}^{L} \theta_l^* P_{l,c}^* + \theta_I^* I_i^* + \theta_0^* \tfrac{1}{H_i^2} + \theta_Z^* Z_i + \widetilde{\epsilon}_i \tag{10}$$

[12] As outlined in the introduction, trends in food prices are now well documented. Evidence on calorie expenditure is scarcer and more mixed. Cutler *et al.* (2003) note that, in the developed world, the majority of the shift away from highly-active jobs occurred in the 1960s and 1970s, before the major rise in obesity. However, using US microdata, Lakdawalla and Philipson (2006) uncover empirical evidence of a relationship between the fall in job-related exercise and the increase in BMI over 1982-2000. There has also been an increase in leisure-time physical exercise, which concerns essentially the better-educated (Sturm, 2004).

where $I_i^* = \frac{\ln(I_i)}{H_i^2}$, $P_{l,c}^* = \frac{\ln(\widehat{\lambda_l^c})}{H_i^2}$, $\theta_l^* = \frac{\delta_l}{\beta E}$, $\theta_I^* = \frac{\delta_I}{\beta E}$, $\theta_0^* = \frac{\delta_0}{\beta E} - \frac{\alpha}{\beta}$. The set of control variables in Z_i is listed in Section 2. This equation explains BMI at the end of year t by prices averaged over year t. Unfortunately, it is not possible to know when, in a calendar year, the anthropometric measures (height and weight) are collected. Hence, the dependent variable will be BMI measured in year $t+1$.

This equation will first be estimated by Ordinary Least Squares (OLS). OLS tells us how the conditional mean BMI changes when prices vary, but has at least two drawbacks. First, OLS is not robust to outliers, e.g. individuals with very high or low BMI (although the distribution has been trimmed). Second, the BMI distribution is not Gaussian. Hence, price elasticities of the conditional mean may not accurately characterise changes in the conditional BMI distribution in response to price interventions, especially for those who are in the right-hand tail, which is the most interesting for public health. I here follow a number of papers in the field, by using quantile regressions to obtain a more complete picture (see, *inter alia*, Kan and Tsaï, 2004). The conditional quantile q of BMI_i^* that will be estimated is:

$$Q_q\left(BMI_i^* | P_{l,c}^*, I_i^*, H_i, Z_i\right) = \sum_{l=1}^{L} \theta_l^*(q) P_{l,c}^* + \theta_I^*(q) I_i^* + \theta_0^*(q) \frac{1}{H_i^2} + \theta_Z^*(q) Z_i \tag{11}$$

where all parameters are free to vary across quantiles.

4.3 Identification issues

I now discuss the identification of the price effects in the above econometric framework. The discussion here is based on a rational choice model developed in Appendix A, which includes the physiological constraints in equations (7) and (8) into the economic framework proposed by Arnade and Gopinath (2006). A consumer i has to choose, under a static budget constraint, the consumption basket that maximises the hedonic pleasure derived from her food intake, while taking into account its potential impact on future well-being through changes in body weight. Her physical activity level $E_{i,t}$ is supposed to be pre-determined over year t. Appendix A shows that, when indirect utility is a quadratic function of the logarithms of food prices and income, BMIs at time $t+1$ and t are linked by the following relationship:

$$BMI_{i,t+1} = BMI_{i,t}\rho(E_{i,t}) + \left[1 - \rho(E_{i,t})\right] \zeta(\mathbf{P}_c^*, I_{i,t}^*, E_{i,t}) \tag{12}$$

where $\rho(E_{i,t})$ is a conservation factor, $\mathbf{P}_c^*$ is the vector of height-adjusted food prices $P_{l,c}^*$ in the spatio-temporal cluster c (i.e. local prices at time t), $I_{i,t}^*$ is the height-adjusted income and $\zeta(\mathbf{P}_c^*, I_{i,t}^*, E_{i,t})$ is:

$$\zeta(\mathbf{P}_c^*, I_{i,t}^*, E_{i,t}) = \left\{ \sum_{l=1}^{L} \theta_l^*\left(E_{i,t}\right) P_{l,c}^* + \theta_I^*\left(E_{i,t}\right) I_{i,t}^* + \theta_0^*\left(E_{i,t}\right) \frac{1}{H_i^2} \right\} \tag{13}$$

$\zeta(\mathbf{P}_c^*, I_{i,t}^*, E_{i,t})$ is the stationary equilibrium weight that would pertain in the absence of shocks to prices, income and PAL. This equilibrium is stable as long as the marginal effect of body weight on calorie intake (through food choices) is lower than its marginal effect on calorie expenditure (through the basal metabolism). The analogy between (13) and (10) is clear: $BMI_i^* = \zeta(\mathbf{P}_c^*, I_{i,t}^*, E_{i,t}) + \theta_Z^* Z_i + \widetilde{\epsilon}_i$, and the theoretical model provides a structural foundation for the empirical specification.

Equations (10) and (13) show explicitly that physical expenditure ($E_{i,t}$) affects the slope of the price-BMI relationship. To capture the intuition behind this result, imagine two individuals who adapt their calorie intakes similarly in response to a price change, but have different PAL because, for instance, they are in different occupations. Then, the body weight of the individual with the higher PAL will be less affected, because she burns a greater fraction of any calories ingested. The flow of energy to be stored, as described by equation (7), will be lower, and so will be the change in her equilibrium body weight.

An important consequence is that OLS *a priori* identifies the average price effects $\theta_l^* = \mathbf{E}\left(\theta_l^*\left(E_{i,t}\right)\right)$ in equation (10) only if the PAL is conditional mean independent of the covariates. Similarly, the quantile elasticities are identified only if there is conditional quantile independence between PAL and the covariates, which is stronger than conditional mean independence. How credible are these identifying restrictions? Although the PAL is probably uncorrelated with food prices, independence from the control variables, especially income, is less obvious. Related evidence is scarce due to a lack of data, but there may be a positive income gradient in PAL for women.[13] It is worth noting however that, here, the food prices are by construction orthogonal to many control variables, in particular income, as these were used in the quality-adjustment regressions of Section 3. As such, the correlations between the PAL and these variables do not affect the estimates of the price effects. The average price effects are thus identified, because the prices are independent from the PAL and from the control variables (Appendix A.4. discusses these identifying restrictions more formally).

The structural model is dynamic in essence, while the empirical specification assumes that individuals are, at least approximately, at a stationary equilibrium. Were this not to be the case, the estimates would suffer from the omission of lagged BMI. I choose to stick with a stationary specification for two reasons. First, the dynamic model in (12) could be estimated by using appropriate moment conditions (*cf.* Blundell and Bond, 1998). However, the identification of average dynamic price effects $\mathbf{E}\left(\left[1-\rho(E_{i,t})\right]\theta_l^*\left(E_{i,t}\right)\right)$ would then require that $E_{i,t}$ be conditional mean independent from $BMI_{i,t}$, since $BMI_{i,t}$ is likely to be correlated with contemporaneous food prices $P_{l,c}^*$ (at least because there is autocorrelation in the latter). This is very unlikely, and the estimated price effects would thus be biased. Second, self-reported body weight is stable between t and $t+1$ for most individuals who are observed over two consecutive periods (see Section 2.1). Stability formally implies that body weight is at a stationary equilibrium.[14] To ensure the

[13]Regarding the income-PAL gradient, descriptive statistics for Europe actually show that leisure-time PALs are on average significantly lower in the first quartile of the income distribution, and do not differ over the remaining quartiles (Rütten and Abu-Omar, 2004). In developed countries, the gradient between PAL and SES tends to be flat for men. In lower social classes, on-the-job physical activity is more important and often offsets the deficit in leisure-time physical activity. Even when individuals are unemployed, they tend to walk more because they use public transportation rather than cars. For women, on-the-job activity may not be more demanding in lower social classes, so that the SES gradient in PAL for them may be positive. For both men and women, the gradient is also flatter when SES is measured by income rather than education or social class, and when we consider both work and leisure-time PALs (Dowler, 2001; IARC, 2002; Gidlow *et al.*, 2006).

[14]The apparent stability of BMI over time may reflect measurement error. However, to counterbalance this point, it is worth noting that empirical longitudinal observations by physiologists have shown that individual body weight variance is generally very small over periods of several weeks to several years. There are strong biological and cognitive control mechanisms that prevent body weight from moving away from its habitual (and *de facto* stationary) level. Hence, stability of body weight over several years may simply mean that consumers are at an

consistency between the empirical specification and the structural model, the estimation sample consists of those individuals for whom $BMI_{i,t+1} = BMI_{i,t}$ (and therefore $BMI_i^* = BMI_{i,t+1}$). This is denoted Sample 2, and the descriptive statistics in Table B1 show that the related socio-demographic characteristics do not differ from those in Sample 1, but stable individuals are likely to have a lower BMI (see the comparison of the BMI distributions for stable and unstable individuals on the right side of Figure B1). Section 7 provides evidence that selection has only a minor impact on the estimates.

5 Empirical results

This section presents results of conditional regressions for the mean and a number of quantiles of the BMI distribution. The regressions were run separately for women and men for three reasons. First, as shown in Figure B2, men have on average higher BMI, although the prevalence of obesity in men and women is about the same. Second, the parameters α and β in the basal metabolic rate equation (8) depend on sex. Third, food habits differ. For instance, men consume more alcohol and meat, and less fruit and vegetables.

The tables present the following quantile elasticity (QE) computed at the sample medians of the right-hand side variables ($\overline{X}^{50}$):

$$\hat{\varepsilon}_{BMIp_l}(\tau) = \frac{1}{\hat{Q}_\tau\left(BMI|\overline{X}^{50}\right)} \frac{\partial \hat{Q}_\tau\left(BMI|\overline{X}^{50}\right)}{\partial \ln(p_l)} = \frac{\widehat{\theta}_l(\tau)}{\overline{H^2}^{50} \times \widehat{\theta}(\tau)'\overline{X}^{50}} \tag{14}$$

where $\widehat{\theta}(\tau)$ is the estimated coefficient vector. The elasticities computed at $\overline{X}^{50}$ are also reported for the OLS regressions.

Individuals are observed over a certain period of time. Confidence intervals are thus constructed by bootstraping the quantile estimates so that, at each replication, individuals rather than individual-year observations are drawn with replacement (1500 draws)

5.1 Main results

Tables C1 and C2 in Appendix C show the results for women and men respectively. Each table has 8 columns of numbers. These present the results from OLS regressions, and quantile regressions for the median and deciles above it, and the sample quantiles corresponding to overweight and obesity in the unconditional BMI distributions (as shown by the values of τ in the column title). Each cell consists of a point estimate of the elasticity with clustered standard errors. Food categories for which these elasticities are significant at the 5% level in at least one of the regressions are in bold; italics indicate that significance is reached at only the 10% level in at least one of the regressions. I focus here, for illustrative purposes, on what happens at the overweight and obesity quantiles only.

There are three important methodological results. First, a number of elasticities are not significant, and those that are significant are not always the same for men as for women. Second,

equilibrium that remains stable in the absence of major shocks. In this context, self-reported body weight should be interpreted as a habitual or reference body weight rather than an imperfect measure of "true" body weight (see, *inter alia*, Cabanac, 2001, and Herman and Polivy, 2003).

price elasticities for the conditional mean and at the overweight and obesity quantiles distribution are often of the same sign, but not of the same significance. They also vary between means and quantiles, and between quantiles, for a number of product categories, although the statistical differences are generally not significant. For instance, for women, the elasticity of the conditional mean is significant for oils (−0.304), while quantile elasticities are of the same magnitude but not significant. The price elasticity is negative and significant for cheese at the overweight quantile (−0.625), but becomes insignificant at the obesity quantile, albeit still large (−0.454). Elasticities with respect to the price of sugar and confectionery are positive at the overweight quantile, but turn out to be negative at the obesity quantile. However, they are all insignificant. Third, distinguishing between processed food and food made at home from raw ingredients matters, as shown by the results for fruit and vegetables.

For men, negative elasticities are found for soft drinks (at the overweight quantile only: −0.160), breaded proteins (−0.066 at the overweight quantile and −0.121 at the obesity quantile), milk (only at the overweight quantile: −0.220) and ready meals (only at the overweight quantile: −0.113). For women, elasticities are negative for cheese, pastries and desserts (−0.209 at the overweight quantile and −0.309 at the obesity quantile), and ready meals (but only at the overweight quantile: −0.192). Income elasticities are shown at the bottom of Tables C1 and C2. These are small, negative and become significant only for women when BMI increases (−0.055 at the obesity quantile). The results for the other control variables Z are available upon request. These show that self-producing fruit and vegetables is negatively correlated with women's BMI, whilst the correlation is positive for men. Being responsible for food expenditure is positively related to BMI for women, with the opposite result for men. There is also a positive and concave age effect (with a peak around 60/70 years old), and a negative education-BMI gradient, which may reflect information and efficiency effects. Some of the regional dummies are significant.

5.2 Interpretation

Regarding the price effects, note that we have no clear-cut predictions about their sign. When there are many food groups, as shown by Schroeter *et al.* (2008), the effect of a change in the price of one food group depends on the own- and cross-price elasticities of consumption, and on their relative share in total energy intakes (see Appendix A.5. for a formal argument). To illustrate this point, consider an increase in the price of some but not all energy-dense products (e.g. snacks but not pastries). Imagine that individuals substitute the former by the latter, and that the cross-price elasticity is strongly positive, while the own-price elasticity is small. It could then result that the fall in calories from snacks be more than compensated by an increase in calories provided by pastries. The results must therefore be interpreted in the light of this analysis and current knowledge about energy intake by product category and elasticities of quantities for food-at-home purchases. The information here is drawn from Allais *et al.* (2008).[15]

For both men and women, positive elasticities are found for (bottled) water. While water

[15] I am indebted to Olivier Allais, who provided me with estimates of Marshallian elasticities of household purchases for food-at-home and proportion of calorie intakes. The estimates in Allais *et al.* (2008) were computed using the same data set and a pseudo-panel approach that helps overcome the problem of unobservability outlined in the Data section. In comparison to the work here, Allais *et al.* work with a slightly different nomenclature, and predictions about the price effects have to be made in terms of individual consumption elasticities, not household purchase elasticities.

brings no calories, increasing its price increases the consumption of energy-dense food such as starches and dairy products, which explains the result. Positive elasticities are also found for fruits in brine, but for women only, although these should rather be negative according to Allais *et al.*'s estimates. The results for soft drinks (respectively dairy products and fats) may be explained by both strong own-price elasticities, and the negative cross-price elasticity of alcohol (respectively cereals and meat) to the price of soft drinks (respectively dairy products and fats). Allais *et al.* find that increasing the price of ready meals and snacks is associated with lower expenditure on meat, and usually leaves expenditures on dairy products, cereals and fats unaffected. Hence, the BMI elasticity to the price of ready meals and snacks should rather be negative. This is the case only for ready meals, while elasticities are positive, but not significant, for snacks.[16]

6 Food price policies and the distribution of BMI

Only relatively few elasticities were found to be significant in the regressions. These seem to be of small size, which is in line with previous empirical findings. Chou *et al.* (2004) find that the elasticity of BMI to food-at-home price is −0.039. In Powell *et al.* (2008), the OLS price elasticity of fruit and vegetables is small and insignificant (0.012). To our knowledge, only Asfaw (2006) has found empirical evidence of significant price effects, but for a developing country (Egypt) with perhaps greater spatial price variation: the BMI elasticities of energy-dense products (bread, sugar, oil and rice) range between −0.1 .and −0.2, while the BMI elasticity of fruit is significantly positive (+0.09), as is that on milk and eggs (+0.141). I will now show that small price elasticities may produce large price effects on the BMI distribution, when the prices of several product categories vary at the same time.

6.1 Simulation method

Table C3 translates naively the estimated elasticities in weight changes for a 1.70 meter tall woman, and 1.80 meter tall man. For instance, a 10% *decrease* in the price of fruit and vegetables in brine would reduce a man's weight by 1.2 kg, if his initial weight was about 81 kg (at the overweight quantile), and by 1.5 kg if his initial weight was 97.2 kg (at the obesity quantile). A policy that increased the prices of soft drinks, pastries and desserts, snacks and ready meals by 10%, and reduced the price of fruit and vegetables in brine by 10% produces weight losses of 2.6 kg at the overweight quantile and 3.6 kg at the obesity quantile for men. These numbers are respectively −3.2 kg and −2.9 kg for women.

However, these simulations are naive, because price elasticities at conditional quantiles are not price effects on unconditional quantiles. They do not fully profit from the advantage of quantile regression over OLS, as the former also provide information on how the *unconditional*

[16] However, Bellisle (2004) recalls that, in France, snacking is associated with greater total energy intake only for obese individuals. Non-obese individuals tend to consume snacks that have better nutritional properties than standard meals. If snacking and going to a full-meal restaurant are substitutes, then raising the price of the former may increase total energy intake in non-obese individuals. The lack of information about substitution between food-at-home and food-away weakens any prediction that could have been made on the sole basis of expenditure on the food-at-home exploited in Allais *et al.*

distribution of BMI is affected by price changes, which is more interesting for the simulation of price policies. The method proposed by Machado and Mata (2005) is applied to simulate the marginal densities *implied by the conditional quantile model* under a given price regime. The procedure consists of five steps:

1. Draw a random sample of B numbers from a uniform distribution on $[0,1]$: τ_1, $\tau_2, .., \tau_B$. Each number represents a quantile of the distribution. Here, $B = 1500$.

2. For each quantile τ_b, estimate using the actual data the quantile regression model (11). This generates B quantile regression parameters $\widehat{\theta}(\tau_b)$, that can be used to simulate the predicted conditional distribution.

3. Generate a random sample of size B by drawing with replacement observations in the actual data set (i.e. from the rows of X): this generates a sample of size B with typical observation X_i.

4. Then $\{BMI_i^* = \widehat{\theta}(\tau_b)'X_i\}$ is a random sample of the BMI distribution integrated over the covariates X, *i.e.* the unconditional distribution of BMI that is consistent with the conditional quantile regressions results.

5. Construct a hypothetical data set from the actual data set, by replacing actual prices by their desired levels (for instance increase the price of vegetables by 10%. Repeat step 3 to obtain a new sample of size B, and step 4 to obtain the marginal distribution that would prevail under the new price regime. Comparing the latter to the actual (predicted) distribution draws a precise picture of the effect of price policies on the prevalence of overweight and obesity.[17]

Confidence intervals can in theory be constructed by repeating these five steps. However, given that the procedure is highly time-consuming, I here focus on point estimates of the effect of hypothetical policy reforms.

6.2 Simulation results

I now compare the simulated distribution of BMI in the current price regime, and the distribution that would prevail under five different price scenarios. In scenario 1, the price of soft drinks and snacks increases by 10%, while the prices of fruit and vegetables in brine decrease by 10%, as suggested by a recent official report (IGF-IGAS, 2008). Scenario 2 *adds* to scenario 1 a 10% increase in the price of alcohol, breaded proteins, pastries and desserts, and ready meals, but does not reduce the prices of fruit and vegetables. In scenario 3, the latter again fall by 10%. Scenario 4 imagines that the prices of fats, sugar and confectionery also increase by 10%. In scenario 5, dairy products, especially cheese, which is at the heart of French gastronomy, also enter in the tax base.

Table C4 reports the results. The first and second lines presents the food categories that are subject to a price increase or decrease. There are three blocks of results for, respectively, men, women, and the entire French population. In each block, the first and second lines give for each scenario, and before and after the implementation of the policy, the prevalences of obesity

[17] The results are robust to the choice of different seeds.

(BMI>30). These prevalences are based on unconditional BMI distributions, using the gender-specific quantile regression results in the previous section and the procedure describe above. Lines 3 to 6 display the corresponding prevalence for two other outcomes: a BMI of over 27, and a BMI between 25 and 30, which corresponds to being overweight. Results for the French population were obtained by adding up the gender-specific results weighted by the census proportions of men and women in the French population.

Emery *et al.* (2000) estimate that the extra medical costs associated with obesity vary between 506 Euros and 648 Euros per obese individual. The lower bound considers only a limited set of medical conditions and individuals with BMI over 30 (population P1 in Table C4), while the upper bound extends the set of medical conditions and takes into account all those with BMI of over 27 (population P2 in Table C4). Hence, by extrapolating the percentages simulated in the lower block to the 48.5 Million French adults (in 2004), I obtain a point estimate of the expected reduction in health-care expenditure. This evaluation is shown in the last two lines of Table C4. It does not take into consideration any statistical uncertainty in the estimated conditional quantiles.

The minimum reduction in health care expenditure varies between 534 million Euros (Scenario 1 - population P1) and 2498 million Euros (Scenario 5 - population P2). Given the regression results, it is not surprising that the larger the tax base, the greater the expected effects. Ideally, I would like a tax base that is not so wide as to produce a sizeable coalitions of opponents and to override the collective representation of food products' healthiness, but not so narrow as to be inefficient. Scenario 3 seems to be a good design, since the tax base does not include symbolic products such as cheese or olive oil, and has sizeable effects. The prevalence of adult obesity here would fall by 3.8 percentage points. Figures C1 and C2 in Appendix C present the results. Figure C1 shows for women and men separately the non-parametric estimates of the BMI distributions before and after the price changes. The distribution of BMI in the population is clearly more favourable in a public-health sense. Figure C2 plots the expected change in BMI against pre-policy BMI. This is negative for most individuals, whatever their initial BMI. As a result, the prevalence of "risky overweight" (BMI>27) in the simulated sample drops from 28.6% to 21.8%. The fall in health care expenditure for population P2 would represent about 1.39% of total public health spending in 2004.

7 Robustness checks

This section first examines the statistical robustness of the quantile regressions. It then checks the sensitivity of the results to the absence of fixed effects and to the sample choice.

7.1 Statistical robustness

The conditional quantile function (11) is well-identified if it is monotonic in τ (Koenker, 2005, section 2.6.).[18] This may not be the case, especially in finite samples, when there are a lot of covariates. Following Machado and Mata (2005), problems with monotonicity can be evaluated

[18] In the structural model, the signs of the derivatives $\partial\theta^*_{\cdot}(E_t)/\partial E_t$ do not vary with E_t. As $\alpha > 0$ and $\beta > 0$, $sign\left[\partial\theta^*_X(E_t)/\partial E_t\right] = sign\left[\pi_X\right]$ (with $X = l, I$) and $sign\left[\partial\theta^*_0(E_t)/\partial E_t\right] = -sign\left[\frac{\gamma(\kappa_{KW}+\lambda_{WW}\gamma)}{\kappa_{KK}}\right]$ where $\frac{(\kappa_{KW}+\lambda_{WW}\gamma)}{\kappa_{KK}}$ is a taste factor defined in Appendix A. This factor equals $\partial K_\tau/\partial W_\tau$ at the optimum, and is therefore positive. Hence, the quantile function is in theory monotonic.

by estimating conditional quantiles at a number of equally-spaced points $\tau \in [0.05, 0.95]$, and by seeing whether, for particular values of the covariates X^0, there are frequent monotonicity violations. There is crossing when $\hat{Q}_\tau\left(BMI_{t+1}|X^0\right) < \max\left\{\hat{Q}_t\left(BMI_{t+1}|X^0\right); t < \tau\right\}$, i.e. the predicted median is smaller than the predicted first quartile. When the design point X^0 sets all log prices to their mean+one standard deviation, and other variables to their median (as in Machado and Mata), violations are observed for 36.5% and 34.5% of the quantiles for women and men respectively. These results are illustrated in Figures C3 and C4 in Appendix C, which depict $\Delta(\tau) = \max\left\{\hat{Q}_t\left(BMI_{t+1}|X^0\right); t < \tau\right\} - \hat{Q}_\tau\left(BMI_{t+1}|X^0\right)$ for women and men respectively, as a function of $\tau \in [0, 1]$: a negative value of $\Delta(\tau)$ indicates a monotonicity violation. These violations are frequent but generally small: less than 0.131 and 0.080 points of BMI for women and men respectively in 90% of cases. These are larger in the extreme quantiles (above 0.95). Although this robustness check only rarely appears in empirical papers, it is useful as it reveals the reliability of the empirical inference. Here, it calls for caution in the use of the results.

However, the lack of monotonicity may reflect the large number of covariates in the model. To investigate this point, complementary results (available upon request) were obtained by aggregating food categories into nine broad food groups: water, beverages other than water, fruit and vegetables, meat and seafood, dairy products, fats, sugar and confectionery, snacks and ready meals, and constructing appropriate price indices. The stationary price-BMI equation was re-estimated for Sample 2 *via* quantile regressions. The results were unchanged from those in Section 5, and statistical inference was more robust with fewer variables. The rate of violations (at the design point X^0) is smaller at 23% and 18.5%, for men and women respectively.

7.2 Specification checks

I now examine the sensitivity of the results to sample selection. The comparison of columns 1 and 2 (for women), and 4 and 5 (for men), in Table C5 shows that changing from Sample 1 to Sample 2 does not significantly alter the OLS estimates. Two striking exceptions are, for women, the price elasticities of oils and cheese, which are much lower in Sample 1 than in Sample 2 (respectively -0.047 and -0.126 in Sample 1, as against -0.386 and -0.306 in Sample 2). The quantile regression results for Sample 1 are available upon request. Fortunately, these food categories were not subject to price changes in Scenario 3. The other differences are minor, although their sum could conceivably produce substantial changes in the expected effect of public policies. Table C6 investigates this issue, by providing simulation results for Scenario 3, using OLS and quantile regressions in Sample 1 and Sample 2. For each outcome (obesity, risky overweight, strict overweight), and each gender, this shows the actual sample prevalence, the simulated pre-policy prevalence and the simulated post-policy prevalence. Four points deserve attention. First, quantile regressions provide simulated pre-policy prevalences that are generally closer to the actual sample prevalences, than those generated by OLS.[19] For instance, for women in Sample 2, the simulated prevalence of strict overweight is 24.3% from quantile regressions, as against 25.7% from OLS. The actual prevalence is 23.54. Second, and as a result, the expected cost reduction for population P1 is smaller from OLS regressions (661 Million Euros vs 960 Million Euros). Quantile regressions are actually much more precise for the evaluation of the

[19] For the OLS simulations, note that I would like to simulate the unconditional probability $Pr(BMI > T)$. This can be done by drawing with replacement B observations in the actual data set (i.e. from the rows of X), and computing $\frac{1}{B}\sum_{b=1}^{B}\left(1 - \Phi\left(\frac{T - \hat{\theta}X_b}{\hat{\sigma}}\right)\right)$, where $\hat{\theta}$ and $\hat{\sigma}$ are estimated by the OLS regression of BMI on X.

tails of the distribution. Here, OLS regressions underestimate the expected benefit of price policies. The advantage of quantile regressions is likely of interest to policy makers especially when they have accurate data on the relationship between health care costs and BMI for the obese individuals. Third, the expected fall in health-care costs, which is shown in the last line, is practically unaffected by the sample choice. This reduction is smaller in the regression results from Sample 1 - 834 Million Euros as against 960 Million Euros in Sample 2 for population P1 -, although the difference is unlikely to be significant. Fourth, it is possible to compute bootstrapped confidence intervals for the OLS simulations, which show that the expected fall in public-health expenditures is significantly different from 0. The 95% confidence interval is [384, 910] Million Euros. This sensitivity analysis also suggests that statistical robustness in the quantile regressions is not a major issue.

A RESET test checks to see that no variable has been omitted from the specification, even if it may be of only limited power (Wooldridge, 2002, p. 125). This reveals potential bias in the OLS regressions for men ($p - value : 0.062$), but not for women ($p - value : 0.1858$). The omission of variables uncorrelated with food prices are of only little importance. Following Section 3, prices are orthogonal to omitted variables only if the unobserved factors that affect quality choices do not systematically differ by cluster (see Section 3.1.). It is possible, for instance, that ethnicity be correlated with quality choice (through cultural foodways), BMI (through gene expression), and that the racial mix in some clusters depart significantly from the national average (in the suburbs of big cities for instance). One way of checking the robustness of results to this assumption is to introduce a fixed effect in the OLS regressions. Table C5 compares OLS and fixed-effect estimates for Sample 1. The fixed-effect results were obtained using Blundell-Bond's (1998) GMM estimator. Since identification relies essentially on individuals whose BMI changes from one period to another, Sample 1 was used instead of Sample 2. The results in columns 2 and 3, and in columns 5 and 6 do not significantly differ, suggesting that the main results are robust.

8 Conclusion

Could appropriate taxes and subsidies reduce the prevalence of obesity? To answer this question, I have estimated whether and how the distribution of BMI in the French adult population is affected by the prices of 23 food products.

Data drawn from an exhaustive household survey on food-at-home purchases, the French TNS-World Panel Survey, were used to investigate this relationship. Negative elasticities were found for soft drinks, cheese, milk, oils, breaded/fried meat and fish, pastries and desserts and ready meals; positive elasticities for water, and fruit and vegetables in brine. These results differ by gender, and by quantiles. Based on the regression results, and using Monte-Carlo simulation techniques, a number of policy scenarios were analysed. A 10% fall in the price of fruit and vegetables in brine, together with a similar increase in the prices of alcohols, soft drinks, breaded proteins, pastries and desserts, snacks and ready meals may reduce the prevalence of obesity by about 3.8 percentage points, with a corresponding reduction in health-care costs of about 960 Million Euros. These are point estimates and, given the width of the estimated confidence intervals for elasticities, should be taken with caution. This is all the more true that micro- and macro-nutrient needs differ according to job requirements, age, gender and health status. Hence, optimal obesity taxes may not be so if we consider other health/nutritional outcomes, or specific socio-demographic groups.

To conclude, it is worth outlining the main methodological findings for operational research: (i) the elasticities of BMI to food prices can be identified even if physical activity is unobserved; and (ii) quantile regression techniques have a clear advantage over simple OLS regressions only if precise data on the relationship between health-care costs and BMI are available

References

[1] AFSSA (2001), *Apports Nutritionnels Recommandés pour la Population Française*, Paris: Editions Tec&Doc.

[2] Allais O., Bertail P. and Nichèle V. (2008), "The effects of a fat tax on French households' nutrient intakes", INRA-ALISS Working Paper 2008-03, http://www.paris.inra.fr/aliss/publications_working_papers/working_papers/aliss_working_papers.

[3] Arnade C. and Gopinath M. (2006), "The Dynamics of Individuals' Fat Consumption", *American Journal of Agricultural Economics*, 88, 836-850.

[4] Asfaw A. (2007), "Do Government Food Price Policies Affect the Prevalence of Obesity? Empirical Evidence from Egypt", *World Development*, 35, 687-701.

[5] ASTEROP (2008), *Etude Loc@lEnseignes*, summary available online at http://www.asterop.com/fr/etudes/localenseignes.htm. Paris: ASTEROP.

[6] Bellisle F. (2004), "Impact of the daily meal pattern on energy balance", *Scandinavian Journal of Nutrition*, 48, 114-118.

[7] Blundell R. and Bond S. (1998), "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87, 115-143

[8] Burkhauser R. and Cawley J. (2008), "Beyond BMI: The value of more accurate measures of fatness and obesity in social science research", *Journal of Health Economics*, 27, 519-529.

[9] Cabanac M. (2001), "Regulation and the ponderostat", *International Journal of Obesity*, 25, S7-S12.

[10] Canivet G. (2004), *Rapport au Ministre des Finances du groupe d'expert constitué sur les rapports entre industrie et commerce* [Report to the Ministry of Finance on the relationships between producers and retailers], 18 Octobre 2004, Paris.

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

[12] Chouinard H., Davis D., LaFrance J. and Perloff J. (2007), "Fat Taxes: Big Money for Small Change", *Forum for Health Economics & Policy*, Vol. 10: Iss. 2 (Obesity), Article 2, http://www.bepress.com/fhep/10/2/2

[13] Cox T. and Wohlgenant M. (1986), "Price and Quality Effects in Cross-Sectional Demand Analysis", *American Journal of Agricultural Economics*, 68, 908-919.

[14] Cutler D., Glaeser E. and Shapiro J.M. (2003), "Why have Americans become more obese?", *Journal of Economic Perspectives*, 17, Summer 2003, 93-118.

[15] Deaton A. (1988), "Quality, Quantity and Spatial Variation of Price", *American Economic Review*, 78, 418-430.

[16] Dowler E. (2001), "Inequalities in diet and physical activity", *Public Health Nutrition*, 4(2B), 701-709.

[17] Drewnowski A. and Darmon N. (2004), "Replacing Fats and Sweets With Vegetables and Fruits – A Question of Cost", *American Journal of Public Health*, 94, 1555-1559.

[18] Emery C., Dinet J., Lafuma A., Sermet C., Khoshnood B. and Fagnagni F. (2007), "Évaluation du coût associé à l'obésité en France", *La Presse Médicale*, 36, 832-840.

[19] Gidlow C., Johnston L.H., Crone D., Ellis N. and James D. (2006), "A systematic review of the relationship between socio-economic position and physical activity", *Health Education Journal*, 65, 338-367.

[20] Herman C.P. and Polivy J. (2003), "Dieting as an Exercise in Behavioral Economics", in *Time and Decision* (Loewenstein G., Read D., Baumeister R.F. eds), New York: Russell Sage Foundation,459-489.

[21] IARC (2002), *IARC Handbook of Cancer Prevention, Vol. 6 Weight control and physical activity*, IARC Press: Lyon.

[22] IGF-IGAS [Inspection général des Finances/Inspection Générale des Affaires Sociales] (2008), "Rapport sur la pertinence et la faisabilité d'une taxe nutritionnelle", Paris: La Documentation Française.

[23] Jacobson M.F. and Brownell K. (2000), "Small Taxes on Soft Drinks and Snack Foods to Promote Health", *American Journal of Public Health*, 90, 854-857.

[24] Kan K. and Tsaï W-D (2004), "Obesity and Risk Knowledge in Taïwan: A Quantile Regression Analysis", *Journal of Health Economics,* 23, 907-934.

[25] Koenker R. (2005), *Quantile Regression*, Cambridge: Cambridge University Press.

[26] Kozusko, F.P. (2001), "Body Weight Setpoint, Metabolic Adaption and Human Starvation", *Bulletin of Mathematical Biology*, 63, 393-404.

[27] Lakdawalla D. and Philipson T. (2006), "Labour Supply and Weight", *Journal of Human Resources*, 42, 85-116.

[28] Lecocq S. and Robin J-M. (2005), "Estimating Demand Response with Panel Data", *Empirical Economics*, 31, 1043-1060.

[29] Leicester A. and Windmeijer F (2004), "The 'fat tax': economic incentives to reduce obesity", *Institute for Fiscal Studies Briefing Note* n°49.

[30] Martin-Houssart G. and Tabard N (2002), "Les équipements publics mieux répartis sur le territoire que les services marchands", in *France: portrait social 2002-2003*, Paris: INSEE, La Documentation Française.

[31] Machado J. and Mata J. (2005), "Counterfactual decomposition of changes in wage distributions using quantile regression", *Journal of Applied Econometrics*, 20, 445-465.

[32] Niedhammer I., Bugel I., Boenfant S., Goldberg M. and Leclerc A. (2000), "Validity of self-reported weight and height in the French GAZEL cohort", *International Journal of Obesity*, 24, 1111-1118.

[33] Powell L.M., Auld M.C., Chaloupka F.J., Johnston L.D., O'Malley P. (2007), "Access to fast food and food prices: the relationship with fruit and vegetable consumption and overweight status among adolescents", *Advances in Health Economics and Health Services Research*, 17, 23-48.

[34] Rütten A. and Abu-Omar K. (2004), "Prevalence of physical activity in the European Union", *Social and Preventive Medicine*, 49, 281-289.

[35] Schroeter C., Lusk J. and Tyner W. (2008), "Determining the impact of food price and income changes on body weight", *Journal of Health Economics*, 27, 45-68.

[36] Sheu W (2006), "The Evolution of the Modern Snack Tax Bill: From World War I to War Against Obesity", Harvard Law School, mimeo available at http://www.law.harvard.edu/faculty/hutt/table_of_contents_2002.html.

[37] Strnad J. (2005), "Conceptualizing the "Fat Tax": The Role of Food Taxes in Developed Economies", *Southern California Law Review*, 78, 1221–1326.

[38] Sturm R. (2004), "The Economics of Physical Activity. Societal Trends and Rationales for Interventions", *American Journal of Preventive Medicine*, 27 (3S), 126-135.

[39] Sturm R. and Datar A. (2005), "Body mass index in elementary school children, metropolitan area food prices and food outlet density", *Public Health*, 119, 1059-1068.

[40] UNU/WHO/FAO (2004), *Human energy requirements*, Report of a Joint FAO/WHO/UNU Expert Consultation 17-24 October 2001, Rome: UNU/WHO/FAO.

[41] Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, London: The MIT Press.

A Model

Using the modelling framework proposed by Arnade and Gopinath (2006), we here propose a dynamic model that simultaneously captures the health and hedonic aspects of the consumer's weight-control problem.

A.1 Set-up

Time The consumer is time-consistent and forward-looking. Time is continuous and divided up into periods (e.g. years). Each period is indexed by $t \in \{0, 1, 2, ...\}$ and, for the empirical analysis, we mostly consider changes occurring between t and $t + 1$.

Budget constraint and choice set At each moment $\tau \in [t, t+1[$, the consumer has to allocate her consumption budget between a numeraire good, y_τ, and a diet made up of L food items, which is represented by a vector of consumptions $\mathbf{c}_\tau$. Let I_t be the consumption budget at τ. This is considered to be exogenously predetermined. The vector of food prices $\mathbf{p}_t$ is also constant over period t. Further, expectations are static, *i.e.* prices and income are expected to remain constant over all future periods. The budget constraint is:

$$\forall \tau \in [t, t+1[, \ \mathbf{p}_t'\mathbf{c}_\tau + y_\tau = I_t \tag{15}$$

Physiology of weight production Food consumption at τ is converted into calorie intake K_τ via a simple linear equation:

$$K_\tau = \mathbf{A}'\mathbf{c}_\tau \tag{16}$$

where $\mathbf{A}$ is a vector of energy densities. Information about the latter is assumed to be perfect. Using equation (8) in Section 4 for the basal metabolic rate, equation (7) becomes:

$$\forall \tau \in [t, t+1[, \ \dot{W}_\tau = \gamma K_\tau - E_t\beta\gamma W_\tau - E_t\alpha\gamma \tag{17}$$

Physical activity level Only calorie intake is endogenised, and the index E_τ for PAL is treated as pre-determined (constant over t):

$$\forall \tau \in [t, t+1[, E_\tau = E_t. \tag{18}$$

While this assumption is likely to hold for work- and commuting-related energy expenditures, it may be less accurate for leisure-time physical activity. A recent general population survey of the health behaviour of the French ("Enquête Conditions de Vie des Ménages", INSEE, 2001) shows that 69.1% of the population do not exercise at least once a week. Only 5.8% exercise explicitly to slim. The barriers to exercise are tastes (36.9%), lack of time (31.9%), impairments to health (21.7%), and "other reasons" (9.4%), which may include prices. Regarding the latter, access to community facilities in France is heavily subsidised and the prevalence of local exercise facilities does not differ notably between low- and high-income areas (Martin-Houssart and Tabard, 2002). The endogenisation of the choice of leisure-time physical activity thus mainly requires us to take the consumer's time constraint into account. This is left for future research.

Preferences Instantaneous preferences are represented by the following utility function:

$$U_\tau = u(\mathbf{c}_\tau, y_\tau; W_\tau, E_t) \tag{19}$$

The utility function has the usual properties.

A.2 The consumer-decision problem

In order to derive the Bellman equation associated with the decision problem of a rational consumer, it is useful to express the latter in a discrete time framework, with arbitrarily small time periods of length $\Delta\tau$. Let $V(W_\tau; \mathbf{p}_t, I_t, E_t)$ be the value function of the consumer at time $\tau \in [t, t+1[$. Between any date $\tau \in [t, t+1[$ and $\tau + \Delta\tau$, the consumer's utility flow is

$U(\mathbf{c}_\tau, y_\tau; W_\tau, E_t)\Delta\tau$. The expected value function for the consumer at $\tau+\Delta\tau$ is $V(W_{\tau+\Delta\tau}; p_t, I_t, E_t)$. Consequently, if σ is the subjective discount rate, the following Bellman equation holds:

$$V(W_\tau; \mathbf{p}_t, I_t, E_t) = Max_{\mathbf{c}_\tau, y_\tau} u(\mathbf{c}_\tau, y_\tau; W_\tau, E_t)\Delta\tau + \frac{1}{1+\sigma\Delta\tau} V(W_{\tau+\Delta\tau}; \mathbf{p}_t, I_t, E_t) \tag{20}$$

under the budget constraint (15). Let $F(\tau) = V(W_\tau; \mathbf{p}_t, I_t, E_t)$, then, assuming that $V(.)$ is C^1, we have by a Taylor expansion:

$$\begin{aligned} V(W_{\tau+\Delta\tau}; \mathbf{p}_t, I_t, E_t) &= F(\tau + \Delta\tau) && (21) \\ &= F(\tau) + \frac{dF(\tau)}{d\tau}\Delta\tau + o(\Delta\tau) \\ &= V(W_\tau; \mathbf{p}_t, I_t, E_t) + V_W(W_\tau; \mathbf{p}_t, I_t, E_t)\dot{W}_\tau\Delta\tau + o(\Delta\tau) \end{aligned}$$

Since $V(W_\tau; \mathbf{p}_t, I_t, E_t)$ does not depend on the control variables $\{\mathbf{c}_\tau, y_\tau\}$, the above approximations can be used to rewrite the Bellman equation as:

$$\sigma V(W_\tau; \mathbf{p}_t, I_t, E_t)\Delta\tau = Max_{\mathbf{c}_\tau, y_\tau} \begin{array}{l} \{u(\mathbf{c}_\tau, y_\tau; W_\tau, E_t)\Delta\tau \\ +V_W(W_\tau; \mathbf{p}_t, I_t, E_t)\dot{W}_\tau\Delta\tau + o(\Delta\tau)\Big\} \end{array} \tag{22}$$

Divide each side by $\Delta\tau$ and let $\Delta\tau \to 0$. Since $\lim_{\Delta\tau\to 0}\left(\frac{o(\Delta\tau)}{\Delta\tau}\right) = 0$, this produces the following Bellman equation:

$$\sigma V(W_\tau; \mathbf{p}_t, I_t, E_t) = Max_{\mathbf{c}_\tau, y_\tau} \Big\{u(\mathbf{c}_\tau, y_\tau; W_\tau, E_t) + V_W(W_\tau; \mathbf{p}_t, I_t, E_t)\dot{W}_\tau\Big\} \tag{23}$$

$$\left| \begin{array}{c} \mathbf{p}_t'\mathbf{c}_\tau + y_\tau = I_t \\ \dot{W}_\tau = \gamma K_\tau - E_t\beta\gamma W_\tau - E_t\alpha\gamma \end{array} \right. \tag{24}$$

where σ is the discount rate. The left-hand term represents the "annuity" from optimal investment decisions, and can be decomposed into the instantaneous stream of utility plus the marginal change in well-being produced by a small change in W.

A.3 Solution

To solve the decision problem, a two-step approach is used (see Arnade and Gopinath). Note that, at time τ, given an optimal path for W, or equivalently K, the consumer would like to choose $\{\mathbf{c}_\tau, y_\tau\}$ in order to maximise $u(\mathbf{c}_\tau, y_\tau; W_\tau, E_t)$. Hence, the first-step of the maximisation procedure consists in finding $\Psi(K_\tau; W_\tau, \mathbf{p}_t, I_t, E_t)$ such that:

$$\begin{aligned} \Psi(K_\tau; W_\tau, \mathbf{p}_t, I_t, E_t) &= Max_{\mathbf{c}_\tau}\{u(\mathbf{c}_\tau, I_\tau - \mathbf{p}_t'\mathbf{c}_\tau; W_\tau, E_t)\} && (25) \\ \text{and } K_\tau &= \mathbf{A}\mathbf{c}_\tau \end{aligned}$$

This decision program yields demand functions $\mathbf{c}_\tau$ that are conditional on a fixed level of calory intake K_τ.

Then, in a second step, the consumer maximises a Bellman equation, in which the control variable is calorie intake K_τ. Replacing $\dot{W}_\tau$ from equation (17) yields:

$$\sigma V(W_\tau;\mathbf{p}_t,I_t,E_t) = Max_{K_\tau} \quad \begin{array}{l}\{\Psi(K_\tau;W_\tau,\mathbf{p}_t,I_t,E_t)\\ +V_W(W_\tau;\mathbf{p}_t,I_t,E_t)\left(\gamma K_\tau - E_t\beta\gamma W_\tau - E_t\alpha\gamma\right)\}\end{array} \tag{26}$$

The first-order condition for the above maximisation problem is:

$$\Psi_K + \gamma V_W = 0 \tag{27}$$

Consider the following quadratic local approximations for the indirect utility functions,

$$\begin{array}{rl} V(W_t;\mathbf{p}_t,I_t,E_t) = & \begin{array}{c}\frac{1}{2}\lambda_{WW}W_t^2 + \sum_{l=1}^{L}\lambda_{p_lW}\ln(p_{lt})W_t \\ +\lambda_{IW}\ln(I_t)W_t + \lambda_{EW}E_tW_t + h^V\left(p_t,I_t,E_t\right)\end{array} \\ \Psi(K_t;W_t,\mathbf{p}_t,I_t,E_t) = & \begin{array}{c}\frac{1}{2}\kappa_{KK}K_t^2 + \kappa_{KW}K_tW_t + \sum_{l=1}^{L}\kappa_{p_lW}\ln(p_{lt})K_t \\ +\kappa_{IK}\ln(I_t)K_t + \kappa_{EK}E_tK_t + h^\Psi(W_t,p_t,I_t,E_t)\end{array} \end{array} \tag{28}$$

then equation (27) implies:

$$\begin{aligned} K_\tau \;=\; & -\frac{1}{\kappa_{KK}}\left[\left(\kappa_{KW}+\lambda_{WW}\gamma\right)W_\tau + \right. \\ & \left.\sum_{l=1}^{L}\left(\kappa_{p_lK}+\lambda_{p_lW}\gamma\right)\ln(p_{lt}) + \left(\kappa_{IW}+\gamma\lambda_{IW}\right)\ln(I_t) + \left(\kappa_{EK}+\gamma\lambda_{EW}\right)E_t\right] \end{aligned} \tag{29}$$

and replacing K_τ in equation (17) then yields:

$$\dot{W}_\tau = -\left[E_t\beta\gamma + \frac{\gamma\left(\kappa_{KW}+\lambda_{WW}\gamma\right)}{\kappa_{KK}}\right]W_\tau + \left[E_t\beta\gamma + \frac{\gamma\left(\kappa_{KW}+\lambda_{WW}\gamma\right)}{\kappa_{KK}}\right]\zeta(\mathbf{p}_t,I_t,E_t) \tag{30}$$

where:

$$\zeta(\mathbf{p}_t,I_t,E_t) = \frac{-\frac{\gamma}{\kappa_{KK}}\left[\sum_{l=1}^{L}\left(\kappa_{p_lK}+\lambda_{p_lW}\gamma\right)\ln(p_{lt}) + \left(\kappa_{IW}+\overline{\gamma}\lambda_{IW}\right)\ln(I_t)\right] - \gamma E_t\left(\alpha + \frac{\left(\kappa_{EK}+\gamma\lambda_{EW}\right)}{\kappa_{KK}}\right)}{\left[E_t\beta\gamma + \frac{\gamma\left(\kappa_{KW}+\lambda_{WW}\gamma\right)}{\kappa_{KK}}\right]} \tag{31}$$

The dynamics of W_τ are given by a first-order linear differential equation, whose solution is:

$$W_\tau = \left[W_t - \zeta(\mathbf{p}_t,I_t,E_t)\right]\exp\left(-\left[E_t\beta\gamma + \frac{\gamma\left(\kappa_{KW}+\lambda_{WW}\gamma\right)}{\kappa_{KK}}\right](\tau - t)\right) + \zeta(\mathbf{p}_t,I_t,E_t) \tag{32}$$

>From which we have an explicit specification for W_{t+1}:

$$W_{t+1} = \left[W_t - \zeta(\mathbf{p}_t,I_t,E_t)\right]\exp\left(-\left[E_t\beta\gamma + \frac{\gamma\left(\kappa_{KW}+\lambda_{WW}\gamma\right)}{\kappa_{KK}}\right]\right) + \zeta(\mathbf{p}_t,I_t,E_t) \tag{33}$$

Dividing both sides of the equation by height squared produces a specification for the BMI. A stable stationary equilibrium exists iff:

$$-\left[E_t\beta\gamma + \frac{\gamma\left(\kappa_{KW}+\lambda_{WW}\gamma\right)}{\kappa_{KK}}\right] < 0 \tag{34}$$

To interpret this condition, consider the effect of a small change in W_t on the optimal choice of K_τ, holding the environmental variables and E_t constant. Implicitly differentiating equation (27) yields:

$$\frac{dK_\tau}{dW_t} = -\frac{(\kappa_{KW} + \lambda_{WW}\gamma)}{\kappa_{KK}} \tag{35}$$

Hence, stability requires that:

$$\frac{dK_\tau}{dW_t} < \frac{d\,(BMR_\tau E_\tau)}{dW_t} \tag{36}$$

which leads to the following proposition:

Proposition 1. *A unique stable equilibirum exists if and only if the marginal effect of body weight on calorie intake is lower than the marginal effect on calorie expenditure*

Clearly, were this condition not to hold, individuals would continue to eat more without any adequate counterbalance in terms of energy expenditure, and body weight would grow indefinitely. The stationary weight is:

$$W^* = \zeta(\mathbf{p}_t, I_t, E_t) \tag{37}$$

The weight-production equation then becomes:

$$W_{t+1} = W_t\rho(E_t) + [1 - \rho(E_t)]\,\zeta(\mathbf{p}_t, I_t, E_t) \tag{38}$$

where $\rho(E_t) = \underbrace{\exp(-\gamma\beta E_t)}_{\rho^0(E_t)}\left(-\gamma\left[\frac{(\kappa_{KW}+\lambda_{WW}\gamma)}{\kappa_{KK}}\right]\right) = \rho^0(E_t) \times \exp\left(\gamma\frac{dK_\tau}{dW_t}\right)$ is a conservation factor. The depreciation of body weight is greater as the individual is more active (since then $\rho^0(E_t)$ decreases), and as the marginal impact of body weight on optimal calorie intake falls. Note also that $W_{t+1} = W_t$ implies that the individual is at the stationary equilibrium specified by (37).

A.4 Identification in the stationary model

I here discuss the identification of the structural model. Dividing both sides of equation (37) by height squared (H^2), adding an orthogonal residual $\tilde{\epsilon}$ and a vector of control variables Z produces an empirical specification for the stationary body mass index BMI^*:

$$BMI^* = \sum_{l=1}^{L}\theta_l^*(E_t)\tfrac{p_l}{H^2} + \theta_I^*(E_t)\tfrac{I}{H^2} + \theta_0^*(E_t)\tfrac{1}{H^2} + \theta_Z^* Z + \tilde{\epsilon} \tag{39}$$

where E_t is the only variable that is not observed in the data. Let p be the row vector of L food prices p_l, $\boldsymbol{\theta}_P^*(E_t)$ the column vector of L coefficients $\theta_l^*(E_t)$, and X the row vector that stacks the variables I/H^2,$1/H^2$ and Z. The structural model of interest in compact form is:

$$\mathbf{E}\left(BMI^*|E_t, p, X\right) = \tfrac{1}{H^2}p\boldsymbol{\theta}_P^*(E_t) + X\boldsymbol{\theta}_X^*(E_t) \tag{40}$$

where $\mathbf{E}\left(.|.\right)$ is the conditional expectation operator. My first identifying restriction is that physical activity E is independent of food prices, but may depend on some of the X variables, other than H.

Assumption 1.

$$\begin{aligned}
\mathbf{E}\left(\boldsymbol{\theta}^*_P(E_t)|p,X\right) &= \boldsymbol{\theta}^0_P + \mathbf{h}_P(\widetilde{X}) \qquad (41)\\
\mathbf{E}\left(\boldsymbol{\theta}^*_X(E_t)|p,X\right) &= \boldsymbol{\theta}^0_I + \mathbf{h}_X(\widetilde{X})\\
\mathbf{E}\left(\mathbf{h}_P(\widetilde{X})|H\right) &= \mathbf{E}\left(\mathbf{h}_X(\widetilde{X})|H\right) = \mathbf{0} \qquad (42)
\end{aligned}$$

where $\widetilde{X} = \{I, Z\}$, $\mathbf{h}_P(\widetilde{X})$ and $\mathbf{h}_X(\widetilde{X})$ are column vectors of dimension L, whose elements are unknown functions of $\widetilde{X}$.

The normalisation of expectation $\mathbf{E}\left(\mathbf{h}_P(\widetilde{X})\right)$ to 0 implies that $\boldsymbol{\theta}^0_P = \mathbf{E}\left(\boldsymbol{\theta}_P(E_t)\right)$ is the column vector of average price effects that I would like to identify. Taking the expectation of (40) conditional on p and X produces, by the law of iterated expectations, the following model:

$$\mathbf{E}(BMI^*|p,X) = \frac{1}{H^2}p\boldsymbol{\theta}^0_P + X\boldsymbol{\theta}^0_I + p\frac{1}{H^2}\mathbf{h}_P(\widetilde{X}) + X\tfrac{1}{H^2}\mathbf{h}_X(\widetilde{X}) \qquad (43)$$

As suggested by Wooldridge (2002, p. 638-639), as long as, conditional on the control variables X, the first two moments of $\frac{1}{H^2}p$ do not depend on E_t, the unconditional average price effect θ^0_P is identified. The main empirical problem is to approximate correctly the control functions $p\frac{1}{H^2}\mathbf{h}_P(\widetilde{\mathbf{X}})$, and $X\frac{1}{H^2}\mathbf{h}_X(\widetilde{\mathbf{X}})$, for instance by introducing polynomials of I, p, Z and $1/H^2$. However, this is not feasible in the quantile regressions as it would require the addition of too many variables, with ensuing effects on statistical robustness. Since I would like to compare the OLS and quantile regression results, the model estimated in the empirical section is simply:

$$\mathbf{E}(BMI^*|p,X) = \frac{1}{H^2}p\boldsymbol{\mu}_P + X\mu_X \qquad (44)$$

The identification problem then arises from the omission of the control functions $p\frac{1}{H^2}\mathbf{h}_P(\widetilde{\mathbf{X}})$, and $X\frac{1}{H^2}\mathbf{h}_X(\widetilde{\mathbf{X}})$ in the regression. However, if p is orthogonal to X, then $\boldsymbol{\mu}^{OLS}_P$ is an asymptotically unbiased estimator of the average price effect $\boldsymbol{\theta}^0_P$. The intuition is that $\frac{1}{H^2}\mathbf{h}_P(\widetilde{\mathbf{X}})$ will be uncorrelated with p, so that the interaction term $p\frac{1}{H^2}\mathbf{h}_P(\widetilde{\mathbf{X}})$ is essentially noise, and does not include information on BMI^*

More formally, let $\mathbf{P}$ be the $N\times L$ matrix of height-adjusted prices $p\frac{1}{H^2}$, where N is the number of observations, $\mathbf{X}$ the matrix of observed values for X, and $\mathbf{BMI}^*$ the vector of observed BMIs. The OLS estimate of $\boldsymbol{\mu}_P$ is (by the Frisch-Waugh-Lovell theorem):

$$\boldsymbol{\mu}^{OLS}_P = (\mathbf{P'M}_X\mathbf{P})^{-1}\mathbf{P'M}_Q BMI^* \qquad (45)$$

where $\mathbf{M}_X = \mathbf{I}_N - \mathbf{X}(\mathbf{X\prime X})^{-1}\mathbf{X}'$ is the orthogonal projection matrix onto the space orthogonal to $\mathbf{X}$ ($\mathbf{I}_N$ being the identity matrix of dimension N). Since the true model is given by (43), we have the usual decomposition of the bias:

$$\boldsymbol{\mu}^{OLS}_P = \boldsymbol{\theta}^0_P + (\mathbf{P'M}_X\mathbf{P})^{-1}\mathbf{P'M}_X\boldsymbol{\Delta}_P(\mathbf{P},\widetilde{\mathbf{X}}) + (\mathbf{P'M}_X\mathbf{P})^{-1}\mathbf{P'M}_X\boldsymbol{\Delta}_X(\mathbf{X},\widetilde{\mathbf{X}}) \qquad (46)$$

where $\boldsymbol{\Delta}_P(\mathbf{P},\widetilde{\mathbf{X}})$ (respectively $\boldsymbol{\Delta}_X(\mathbf{X},\widetilde{\mathbf{X}})$) is a $N\times 1$ vector with typical element $p\frac{1}{H^2}\mathbf{h}_P(\widetilde{X})$ (respectively $X\frac{1}{H^2}\mathbf{h}_X(\widetilde{X})$). Note that $(\mathbf{P'M}_X\mathbf{P})^{-1}\mathbf{P'M}_X\boldsymbol{\Delta}_X(\mathbf{X},\widetilde{\mathbf{X}}) = \mathbf{0}$ because $\boldsymbol{\Delta}_X(\mathbf{X},\widetilde{\mathbf{X}})$ is in the space spanned by $\mathbf{X}$, so that $\mathbf{M}_X\boldsymbol{\Delta}_X(\mathbf{X},\widetilde{\mathbf{X}}) = \mathbf{0}$. Taking the probability limit gives the asymptotic bias:

$$\operatorname*{plim}_{n\to\infty}\left(\boldsymbol{\mu}^{OLS}_P\right) = \boldsymbol{\theta}^0_P + \operatorname*{plim}_{n\to\infty}\left(\tfrac{1}{n}\mathbf{P'M}_X\mathbf{P}\right)^{-1}\operatorname*{plim}_{n\to\infty}\left(\tfrac{1}{n}\mathbf{P'M}_X\boldsymbol{\Delta}_P(\mathbf{P},\widetilde{\mathbf{X}})\right) \qquad (47)$$

Let $\underset{n\to\infty}{\text{plim}}\left(\frac{1}{n}\mathbf{P}'\mathbf{M}_X\mathbf{\Delta}_P(\mathbf{P},\widetilde{\mathbf{X}})\right) = \Omega$, where Ω is a $L\times 1$ vector with k^{th} element w_k. Note that $w_k = \mathbf{E}\left(\sum_{l=1}^{L} cov(\frac{p_l}{H^2}h_{Pl}(\widetilde{X}), p_j\frac{1}{H^2}|X)\right)$, where $h_{Pl}(\widetilde{X})$ is the l^{th} element of $\mathbf{h}_P(\widetilde{X})$: the remaining bias depends on the covariance between $\frac{p_l}{H^2}h_{Pl}(\widetilde{X})$ and $p_j\frac{1}{H^2}$ for all $l,k\in\{1,..,L\}^2$ conditional on the control variables X. The second identifying restriction is that the vector of prices p does not depend on X

Assumption 2.

$$\mathbf{E}(p|X) = \mathbf{E}(p) \tag{48}$$

Under assumption 2, $\forall\, l,k\in\{1,..,L\}^2$, $cov(p_l,p_k|X) = cov(p_l,p_k) = \rho_{lk}$ and $cov(\frac{p_l}{H^2}h_{Pl}(\widetilde{X}), p_j\frac{1}{H^2}|X) = h_{Pl}(\widetilde{X})\frac{1}{H^4}\rho_{lk}$. As a consequence, $\mathbf{w}_k = \sum_{l=1}^{L}\mathbf{E}\left(cov(\frac{p_l}{H^2}h_{Pl}(\widetilde{X}), p_j\frac{1}{H^2}|X)\right) = \rho_{lk}\sum_{l=1}^{L}\mathbf{E}\left(h_{Pl}(\widetilde{X})\frac{1}{H^4}\right) = \rho_{lk}\sum_{l=1}^{L}\mathbf{E}\left(h_{Pl}(\widetilde{X})\right)\mathbf{E}\left(\frac{1}{H^4}\right) = 0$, where the assumption of independence between $\mathbf{h}_P(\widetilde{X})$ and H (see assumption 1) has been used. The asymptotic bias is zero: $\underset{n\to\infty}{\text{plim}}\left(\boldsymbol{\mu}_P^{OLS}\right) = \boldsymbol{\theta}_P^0$.

To sum up, average price effects are identified if: (i) food prices are independent of the control variables (especially income) and physical activity; and (ii) conditional on income, physical activity is independent of food prices and height.

The identification of price effects in the quantile regressions may well require more stringent conditions. However, this issue is more difficult to investigate as little is known about misspecified quantile regressions.

A.5 Price effects

As daily calorie expenditures are constant over the period $[t,t+1[$, a general solution to (17) is:

$$W_{t+1} = \exp\left(-\beta\gamma E_t\right)W_t + \int_t^{t+1}\left[\gamma K_\tau \exp\left(-E_t\beta\gamma((t+1)-\tau)\right)\right]d\tau - \frac{\alpha}{\beta}\left[1-\exp\left(-\beta\gamma E_t\right)\right] \tag{49}$$

This equation expresses body weight at the beginning of period $t+1$ as a function of body weight at the beginning of period t and the optimal calorie intake stream (represented by the integral). Prices act implicitly in this equation by determining eating behaviour and hence the path followed by K_τ. The conservation factor $\exp\left(-E_t\beta\gamma((t+1)-\tau)\right)$ moderates the way in which calorie intake is transformed into body weight.

Equation (16) implies that $K_\tau = \sum_l a_l c_{\tau l}$, where $c_{\tau l}$ is the consumption of food product l at time τ and a_l is its per-unit caloric content. Denote its price by p_{lt}, and consider the effect of a change in the price p_{1t} of c_{1t} by differentiating (49). Under the usual regularity conditions regarding the function K_τ, we have:

$$\begin{aligned}\frac{\partial W_{t+1}}{\partial p_{1t}} &= \int_t^{t+1}\gamma\frac{\partial K_\tau}{\partial p_{1t}}\rho^0(E_t,\tau)d\tau \\ &= \int_t^{t+1}\gamma\left\{\sum_{l=1}^{L}a_l\frac{\partial c_{\tau l}}{\partial p_{1t}}\right\}\rho^0(E_t,\tau)d\tau \end{aligned} \tag{50}$$

This price effect cannot be signed without further assumptions on the own- and cross-price elasticities of consumption and the relative densities of each food item. The price effect is positive if $\forall\tau\in[t,t+1[$, $\sum_{l=1}^{L}a_l\frac{\partial c_{\tau l}}{\partial p_{1t}} > 0$. When there are only two product categories, this condition holds when category 1 and category 2 have the same densities ($a_1/a_2 \approx 1$), and when the own-price elasticity of category 1 is small while the cross-price elasticity is positive and fairly high (i.e. categories 1 and 2 are strong substitutes).

Appendix B. Descriptive statistics.

Figure B1. BMI distributions – Sample 1 and Sample 2.

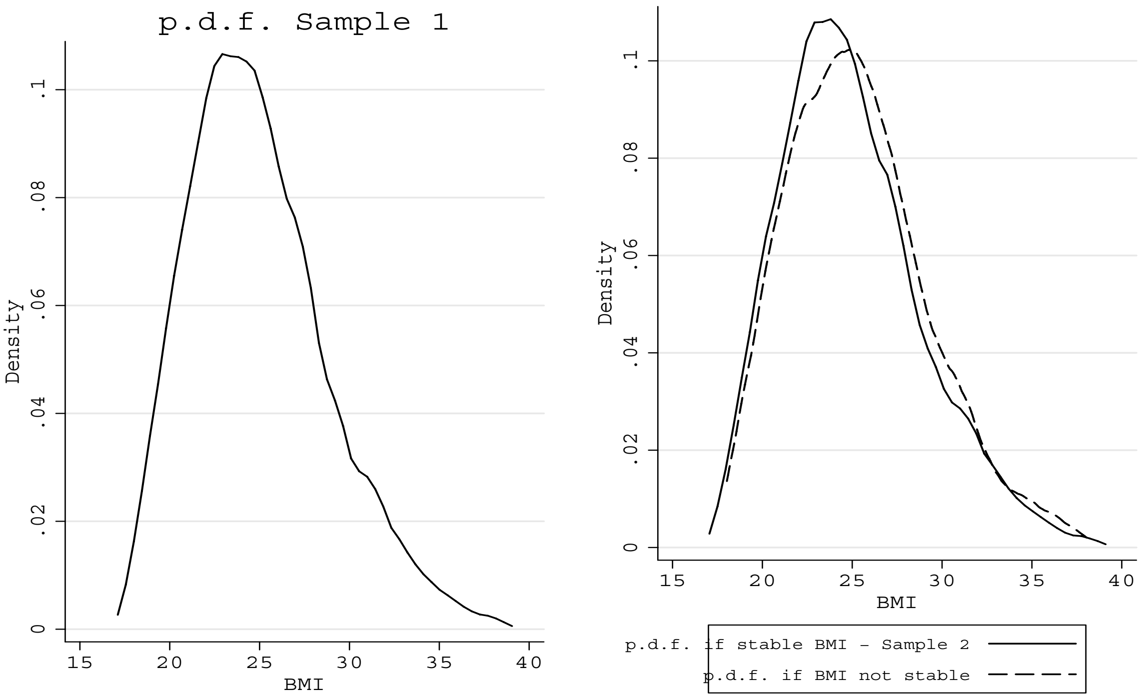

Figure B2. BMI distributions by gender – Sample 2

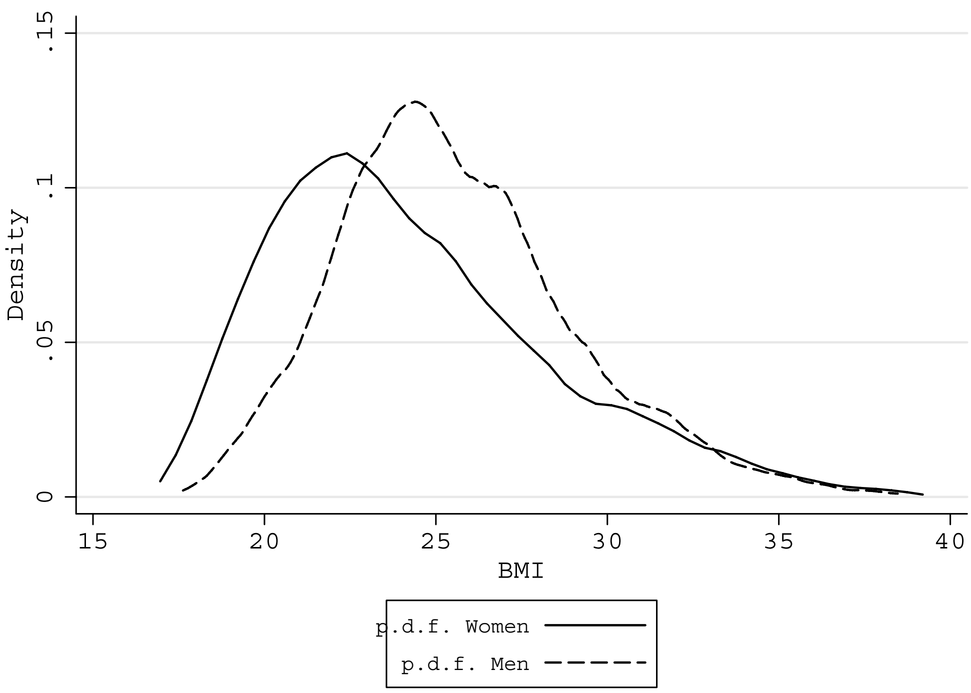

Table B1. Variable definition and descriptive statistics

Variable Name	Variable definition	Sample 1 (N=21407)	Sample 2 (N=12608)
BMI	Body Mass Index (Weight in Kg divided by height in squared meters)	24.91 (3.86)	24.92 (3.86)
CONTHEIGHT	[Height (in meters)]$^{-2}$	0.359 (0.039)	0.359 (0.038)
INCOME	Logarithm of real household income per unit of consumption (Oxford scale, 2004 Euros)	1137.1 (613.3)	1142.2 (617.1)
DEG1	No qualification or primary school	15.5%	16.2%
DEG2	Short vocational or technical qualification	6.8%	6.8%
DEG3	First cycle of secondary school (BEPC)	30.5%	30.7%
DEG4	Baccalaureat (general, vocational or technical)	19.0%	18.7%
DEG5	Baccalaureat + 2 years	10.6%	10.4%
DEG6 (reference)	Baccalaureat + 3 years or more	17.6%	17.2%
SEX	=1 for male, 0 otherwise	47.1%	47.2%
AGE	Age	50.40 (15.73)	51.65 (15.67)
BABYWOMAN	=1 for women with a baby aged under one year.	1.3%	1.2%
BABYMAN	=1 for men with a baby aged under one year.	1.4%	1.3%
COUPLE (reference)	Couples (reference)	66.6%	65.9%
SINGLE	Single without children	17.9%	18.0%
OTHHHOLD	Other household structure	15.6%	16.1%
NBIND	Number of person in the household	2.63 (1.27)	2.61 (1.27)
FRUITSORVEG	Household produces fruit or vegetables	19.0%	18.8%
MEALPLANNER	=1 if the individual is responsible for food-at-home expenditures	30.6%	30.5%
UNIT1	Lives in a rural residential area	30.3%	30.3%
UNIT2	Lives in an urban unit of between 2000 and 4999 residents	6.4%	6.7%
UNIT3	Lives in an urban unit of between 5000 and 9999 residents	2.7%	2.9%
UNIT4	Lives in an urban unit of between 10000 and 19999 residents	2.6%	2.4%
UNIT5	Lives in an urban unit of between 20000 and 49999 residents	3.7%	3.3%
UNIT6	Lives in an urban unit of between 50000 and 99999 residents	5.6%	5.6%
UNIT7	Lives in an urban unit of between 100000 and 199999 residents	4.2%	3.6%
UNIT8	Lives in an urban unit of 200000 residents or more, or in Ile-de-France outside Paris	22.8%	23.6%
REGION1 (reference)	Ile-de-France	23.5%	23.2%
REGION2	Picardie, Normandie	15.0%	14.9%
REGION3	Nord	8.6%	8.7%
REGION4	Champagne-Ardennes, Alsace, Lorraine	8.9%	9.1%
REGION5	Bretagne, Pays de Loire, Centre	15.8%	16.4%
REGION6	Limousin, Aquitaine, Poitou-Charente	6.1%	5.4%
REGION7	Bourgogne, Franche-Comté, Rhône-Alpes, Auvergne, Midi-Pyrénées, Languedoc	12.6%	12.9%
REGION8	Provence – Alpes - Côte d'Azur	9.4%	9.5%
YR2002	Calendar year = 2002 (for the dependent variable)	26.9%	_
YR2003	Calendar year = 2003 (for the dependent variable)	27.8%	38.0%
YR2004 (reference)	Calendar year = 2004 (for the dependent variable)	23.5%	30.5%
YR2005	Calendar year = 2005 (for the dependent variable)	21.8%	31.6%

Table B2. Classification of Food Products

Product category	**Food products (examples)**	**Comments**
Water	Fizzy or still, mineral or not.	
Alcohol	All kind of wines, cocktails, beers, ciders, liquors etc.	Products are aggregated according to their average alcohol content.
Soft drinks	Fruit juice, soda and other carbohydrated drinks (lemonade, syrups etc.). Flavoured waters were dropped.	Products are distinguished according to their sugar or fat content when available.
Fruit in brine	All fresh fruit and fruit canned/frozen in brine	
Processed Fruit	Fruits canned in syrup etc.	Products light in sugar are distinguished.
Vegetables in brine	All fresh vegetables plus vegetables canned/frozen in brine	
Processed vegetables	Cooked frozen vegetables, vegetables and soups canned/frozen with additives.	
Cereals	Dried vegetables, potatoes, beans except fresh green or yellow beans, pasta, rice, bread, flour, chestnuts, oat flakes, couscous.	
Meat in brine and eggs	Fresh/raw meat: beef, veal, snails, chicken, eggs…	
Seafood in brine	All fish, shellfish, frogs etc. in brine	
Processed sea products	Fish canned in oil, smoked salmon, marinated haddock, rollmops	Canned
Cooked meat	Sausages, ham, pâté, foie gras, bacon, smoked pork	
Breaded proteins	Breaded/fried fish or meat	
Yoghurt and fresh uncured cheese	Natural yoghurt, milk, fresh uncured cheese (fromage blanc ou frais)	Products are distinguished according to their fat content when available. Products without explicit fat content were dropped.
Cheese	All cheese except fromage blanc and fromage frais.	
Milk	All milk (soja milk was dropped).	
Animal fat and margarine	Butter, fresh cream.	
Oils	Oils. Sauces were dropped.	
Sugar and confectionery	Lump/caster sugar, honey, jam marmalades.	
Pastries and desserts	Milk desserts, croissants, cakes, fresh or frozen pastries	
Sweet and fatty snacks	Breakfast cereals, cereal bars, chocolate bars, most chocolate products, biscuits, ice creams	“sweet” here means either simple or complex carbohydrates.
Salty and fatty snacks	Crackers, pop-corn, peanuts, most appetisers, olives	
Ready-meals	All ready meals including sandwiches, and canned/frozen recipes of vegetables and cereals (ratatouille, etc.).	

Appendix C. Results

Table C1. Conditional mean and quantile regression results - 23 food categories – Elasticities - Sample 2 - Women – N=6633

Estimator	**OLS**	**Quantile Regressions**						
Conditional moment	**Mean**	**τ=0.5 median**	**τ=0.6 6th decile**	**τ=0. 7 7th decile**	**τ=0.8 8th decile**	**τ=0.9 9th decile**	**τ=0.637 "overweight quantile"**	**τ=0.898 "obesity quantile"**
				Price Elasticities				
Water	0.065* (0.038)	0.075 (0.052)	0.075 (0.053)	0.093 (0.060)	0.152*** (0.059)	0.099 (0.065)	0.062 (0.056)	0.098 (0.061)
Alcohol	0.058 (0.047)	0.058 (0.060)	0.058 (0.063)	0.069 (0.074)	0.040 (0.079)	0.036 (0.081)	0.059 (0.066)	0.035 (0.078)
Soft drinks	-0.055 (0.068)	-0.055 (0.088)	-0.080 (0.097)	-0.023 (0.113)	-0.104 (0.123)	-0.137 (0.133)	-0.030 (0.100)	-0.141 (0.129)
Fruit in brine	0.134* (0.060)	0.155** (0.077)	0.184** (0.081)	0.140 (0.097)	0.209* (0.110)	0.091 (0.113)	0.166* (0.087)	0.093 (0.111)
Processed Fruit	0.018 (0.053)	0.098 (0.067)	0.081 (0.070)	0.016 (0.087)	-0.050 (0.089)	-0.006 (0.108)	0.040 (0.077)	-0.008 (0.109)
Vegetables in brine	-0.025 (0.080)	-0.050 (0.107)	0.013 (0.117)	0.100 (0.132)	-0.030 (0.130)	-0.010 (0.152)	0.049 (0.121)	-0.002 (0.150)
Processed vegetables	-0.002 (0.060)	0.045 (0.071)	0.027 (0.080)	-0.041 (0.097)	-0.079 (0.106)	-0.001 (0.123)	0.017 (0.086)	0.009 (0.120)
Cereals	0.045 (0.073)	-0.009 (0.097)	-0.061 (0.096)	-0.084 (0.107)	-0.020 (0.124)	0.065 (0.131)	-0.064 (0.102)	0.061 (0.133)
Meat in brine and eggs	-0.006 (0.088)	-0.016 (0.108)	0.013 (0.128)	0.023 (0.135)	0.109 (0.145)	0.082 (0.172)	0.009 (0.131)	0.085 (0.166)
Seafood in brine	-0.007 (0.035)	-0.027 (0.044)	-0.021 (0.047)	-0.024 (0.058)	-0.016 (0.061)	0.055 (0.065)	-0.050 (0.049)	0.049 (0.065)
Processed sea products	-0.019 (0.050)	-0.031 (0.067)	-0.033 (0.071)	-0.065 (0.080)	-0.098 (0.087)	-0.087 (0.096)	-0.067 (0.072)	-0.081 (0.092)
Cooked meat	0.135* (0.079)	0.098 (0.099)	0.110 (0.109)	0.102 (0.121)	0.172 (0.137)	0.185 (0.147)	0.182 (0.114)	0.185 (0.145)
Breaded proteins	0.005 (0.032)	0.030 (0.045)	0.007 (0.045)	0.016 (0.051)	0.027 (0.048)	-0.021 (0.053)	0.011 (0.048)	-0.018 (0.053)
Yoghurt and fresh uncured cheese	0.134 (0.112)	0.050 (0.151)	0.089 (0.163)	0.139 (0.177)	0.298 (0.189)	0.190 (0.193)	0.150 (0.164)	0.194 (0.193)
Cheese	-0.386**	-0.585**	-0.784***	-0.464	-0.591*	-0.455	-0.625**	-0.454

	(0.193)	(0.251)	(0.270)	(0.321)	(0.356)	(0.330)	(0.276)	(0.329)
Milk	-0.020 (0.100)	-0.049 (0.132)	0.032 (0.136)	0.043 (0.157)	-0.001 (0.165)	0.013 (0.179)	-0.043 (0.141)	0.001 (0.180)
Animal fat and margarine	-0.021 (0.055)	-0.009 (0.073)	-0.031 (0.076)	-0.120 (0.085)	-0.062 (0.092)	-0.040 (0.104)	-0.102 (0.075)	-0.041 (0.101)
Oils	-0.304** (0.154)	-0.206 (0.203)	-0.338 (0.221)	-0.287 (0.255)	-0.226 (0.266)	-0.258 (0.276)	-0.284 (0.227)	-0.258 (0.276)
Sugar and confectionery	0.046 (0.093)	0.220* (0.123)	0.204 (0.133)	0.101 (0.149)	0.107 (0.148)	-0.213 (0.171)	0.183 (0.143)	-0.207 (0.167)
Pastries and desserts	-0.182** (0.078)	-0.220** (0.103)	-0.251** (0.105)	-0.128 (0.132)	-0.223* (0.135)	-0.317** (0.142)	-0.219** (0.106)	-0.309** (0.140)
Sweet and fatty snacks	0.136 (0.100)	0.119 (0.126)	0.178 (0.125)	0.226 (0.146)	0.122 (0.166)	0.201 (0.205)	0.176 (0.133)	0.190 (0.208)
Salty and fatty snacks	-0.042 (0.088)	-0.034 (0.115)	0.023 (0.113)	-0.074 (0.132)	0.002 (0.135)	0.054 (0.154)	0.034 (0.117)	0.051 (0.150)
Ready-meals	-0.058 (0.068)	-0.114 (0.093)	-0.162* (0.092)	-0.160* (0.094)	-0.089 (0.117)	-0.031 (0.139)	-0.192** (0.092)	-0.038 (0.137)
Income Elasticities								
Income	-0.030*** (0.009)	-0.018 (0.011)	-0.013 (0.013)	-0.031** (0.015)	-0.040*** (0.015)	-0.055*** (0.016)	-0.019 (0.014)	-0.055*** (0.015)
Other control variables	*CONTHEIGHT, DEG1-DEG6, (AGE/10), $(AGE/10)^2$, BABYWOMAN, COUPLE, SINGLE, OTHHHOLD, NBIND, FRUITSORVEG, MEALPLANNER, UNIT1-UNIT8, REGION1-REGION8, YR2003-YR2005*							

Note: For each regression and each variable, the point estimates of elasticities at median values of the control variables are shown, with their robust standard deviations in parenthesis; * = significant at the 10% level. ** = at the 5% level. *** = at the 1% level.

Table C2. Conditional mean and quantile regression results - 23 food categories – Elasticities - Sample 2 - Men – N=5975

Estimator	**OLS**	**Quantile Regressions**						
Conditional moment	**Mean**	**τ=0.5 median**	**τ=0.6 6th decile**	**τ=0. 7 7th decile**	**τ=0.8 8th decile**	**τ=0.9 9th decile**	**τ=0.484 "overweight quantile"**	**τ=0.888 "obesity quantile"**
Price Elasticities								
Water	0.092*** (0.031)	0.095*** (0.037)	0.094** (0.037)	0.112*** (0.042)	0.118** (0.055)	0.125** (0.062)	0.092*** (0.036)	0.127** (0.064)
Alcohol	-0.013 (0.044)	-0.027 (0.047)	-0.034 (0.051)	0.021 (0.067)	0.071 (0.083)	0.115 (0.084)	-0.026 (0.046)	0.132 (0.089)
Soft drinks	-0.113* (0.059)	-0.162** (0.077)	-0.161** (0.076)	-0.100 (0.086)	-0.093 (0.109)	-0.115 (0.123)	-0.160** (0.076)	-0.108 (0.122)
Fruit in brine	0.045 (0.050)	0.055 (0.064)	0.042 (0.066)	0.099 (0.078)	0.138* (0.083)	0.121 (0.096)	0.044 (0.064)	0.142 (0.096)
Processed Fruit	0.022 (0.047)	0.000 (0.056)	0.034 (0.059)	0.044 (0.070)	0.094 (0.078)	0.164* (0.093)	-0.007 (0.057)	0.175* (0.090)
Vegetables in brine	0.027 (0.073)	0.102 (0.085)	0.000 (0.085)	0.086 (0.107)	0.043 (0.135)	-0.038 (0.150)	0.105 (0.091)	0.011 (0.154)
Processed vegetables	0.016 (0.054)	0.041 (0.064)	0.059 (0.065)	0.029 (0.080)	-0.065 (0.095)	-0.125 (0.096)	0.036 (0.061)	-0.111 (0.097)
Cereals	0.009 (0.059)	-0.042 (0.076)	-0.064 (0.074)	-0.141 (0.094)	-0.031 (0.116)	0.151 (0.114)	-0.036 (0.076)	0.111 (0.121)
Meat in brine and eggs	0.078 (0.079)	0.035 (0.092)	0.000 (0.097)	0.018 (0.122)	0.066 (0.151)	0.033 (0.155)	0.049 (0.087)	0.036 (0.156)
Seafood in brine	-0.022 (0.030)	-0.053 (0.040)	-0.032 (0.040)	-0.062 (0.047)	-0.039 (0.055)	0.013 (0.057)	-0.043 (0.038)	0.021 (0.057)
Processed sea products	0.068 (0.044)	0.059 (0.052)	0.074 (0.055)	0.031 (0.066)	0.053 (0.077)	0.045 (0.078)	0.049 (0.051)	0.029 (0.081)
Cooked meat	-0.042 (0.067)	0.013 (0.083)	-0.010 (0.090)	0.011 (0.102)	0.021 (0.114)	-0.039 (0.122)	0.006 (0.078)	-0.043 (0.121)
Breaded proteins	-0.047 (0.029)	-0.073** (0.033)	-0.041 (0.034)	-0.055 (0.044)	-0.090 (0.057)	-0.120* (0.062)	-0.066** (0.033)	-0.121* (0.063)
Yoghurt and fresh uncured cheese	0.057 (0.094)	0.033 (0.115)	0.068 (0.117)	0.086 (0.139)	0.263 (0.170)	0.177 (0.167)	0.053 (0.109)	0.231 (0.164)
Cheese	-0.135 (0.166)	-0.268 (0.206)	-0.130 (0.187)	-0.234 (0.226)	-0.303 (0.301)	-0.135 (0.321)	-0.283 (0.198)	-0.140 (0.314)
Milk	-0.185**	-0.229**	-0.252**	-0.185	-0.333**	-0.185	-0.220**	-0.156

	(0.085)	(0.102)	(0.107)	(0.133)	(0.161)	(0.160)	(0.101)	(0.167)
Animal fat and margarine	0.032 (0.047)	0.057 (0.061)	0.060 (0.059)	0.025 (0.068)	0.034 (0.081)	-0.051 (0.086)	0.051 (0.058)	-0.063 (0.084)
Oils	-0.060 (0.127)	0.008 (0.152)	-0.018 (0.171)	-0.085 (0.206)	-0.133 (0.221)	-0.183 (0.225)	-0.017 (0.153)	-0.195 (0.224)
Sugar and confectionery	0.018 (0.078)	0.044 (0.096)	0.035 (0.101)	-0.034 (0.121)	-0.001 (0.136)	-0.068 (0.140)	0.035 (0.091)	-0.110 (0.146)
Pastries and desserts	-0.041 (0.071)	-0.030 (0.084)	-0.049 (0.086)	0.045 (0.102)	-0.064 (0.124)	-0.155 (0.133)	-0.035 (0.084)	-0.211 (0.138)
Sweet and fatty snacks	-0.018 (0.083)	0.036 (0.094)	-0.031 (0.103)	-0.025 (0.122)	-0.025 (0.165)	0.273 (0.182)	0.031 (0.091)	0.234 (0.187)
Salty and fatty snacks	0.033 (0.074)	0.116 (0.086)	0.158* (0.090)	0.122 (0.104)	0.073 (0.135)	-0.210 (0.142)	0.104 (0.088)	-0.148 (0.145)
Ready-meals	-0.033 (0.063)	-0.121* (0.072)	-0.130* (0.077)	-0.075 (0.097)	-0.091 (0.117)	0.037 (0.126)	-0.113* (0.068)	0.015 (0.127)
Income Elasticities								
Income	-0.008 (0.008)	-0.009 (0.010)	-0.014 (0.010)	-0.013 (0.012)	-0.007 (0.014)	-0.013 (0.014)	-0.005 (0.010)	-0.019 (0.014)
Other control variables	*CONTHEIGHT, DEG1-DEG6, (AGE/10), $(AGE/10)^2$, BABYMAN, COUPLE, SINGLE, OTHHOLD, NBIND, FRUITSORVEG, MEALPLANNER, UNIT1-UNIT8, REGION1-REGION8, YR2003-YR2005*							

Note: For each regression and each variable, the point estimates of elasticities at median values of the control variables are shown, with their robust standard deviations in parenthesis; * = significant at the 10% level. ** = at the 5% level. *** = at the 1% level.

Table C3. Illustrations of the price effects for typical individuals.

	Women		Men	
BMI	25	30	25	30
Height	1.70	1.70	1.80	1.80
Weight	72.25	86.70	81.00	97.20
Effect (in kg) of a 10 % price decrease				
Fruits and vegetables in brine	-1.55	-0.79	-1.21	-1.49
Effect (in kg) of a 10 % price increase				
Soft drinks	-0.22	-1.22	-1.31	-1.05
Pastries, deserts, snacks and ready-meals	-1.45	-0.92	-0.11	-1.07
Total Effect	-3.22	-2.93	-2.62	-3.61

Note: approximate effects using Tables C1 and C2' results.

Table C4. Five simulated policy scenarios – based on quantile regression results obtained in Sample 2

Scenario		**1**	**2**	**3**	**4**	**5**
Price increase: 10%		Softs, Snacks	Softs, Snacks, **Alc**, **Br. P., Past. & D., R-Meals**	Alc., Softs, Br. P., Past. & D., Snacks, R-Meals	Alc., Softs, Br. P., Past. & D., Snacks, R-Meals, **Fats, S&C**	Alc., Softs, Br. P., Past. & D., Snacks, R-Meals, Fats, S&C, **Dairies**
Price decrease: 10%		F & V in B	_	**F & V in B**	F & V in B	F & V in B.
				Women		
Prevalence of obesity (BMI ≥ 30)	Pre-policy			10.9%		
	Post-policy	8.9%	8.9%	7.1%	5.9%	6.4%
% BMI ≥ 27	Pre-policy			24.3%		
	Post-policy	21.3%	20.9%	18.4%	18.2%	17.0%
Prevalence of overweight (BMI ≥ 25 & BMI < 30)	Pre-policy			26.8%		
	Post-policy	25.8%	23.4%	23.1%	27.1%	21.8%
				Men		
Prevalence of obesity (BMI ≥ 30)	Pre-policy			12.5%		
	Post-policy	10.3%	9.7%	8.6%	8.4%	8.9%
% BMI ≥ 27	Pre-policy			33.3%		
	Post-policy	28.2%	28.3%	25.5%	28.0%	24.6%
Prevalence of overweight (BMI ≥ 25 & BMI < 30)	Pre-policy			42.3%		
	Post-policy	40.1%	39.8%	38.2%	41.9%	33.2%
				All		
Prevalence of obesity (BMI ≥ 30)	Pre-policy			11.7%		
	Post-policy	9.5%	9.3%	7.8%	7.1%	7.6%
% BMI ≥ 27	Pre-policy			28.6%		
	Post-policy	24.6%	24.5%	21.8%	23.0%	20.7%
Prevalence of overweight (BMI ≥ 25 & BMI < 30)	Pre-policy			34.3%		
	Post-policy	32.7%	31.4%	30.4%	34.3%	27.3%
Reduction in health care expenditure (million Euros)	Obese (P1)	534	603	960	1131	1004
	BMI ≥ 27 (P2)	1257	1302	2133	1781	2498

Note: Alc = alcohol; Softs = soft drinks; Br. P. = breaded proteins; Past. & D. = pastries and deserts; Snacks = either sweet and fatty or salty & fatty R-meals = ready-meals; F& V in B = Fruit and Vegetables in brine; Fats = animal fats + oils; S & C = sugar and confectionery; Proc F & V = processed Fruit and vegetables; Dairies = yogurt, cheese & milk.

Figure C1. Scenario 3 – pre/post BMI distributions – Sample 2

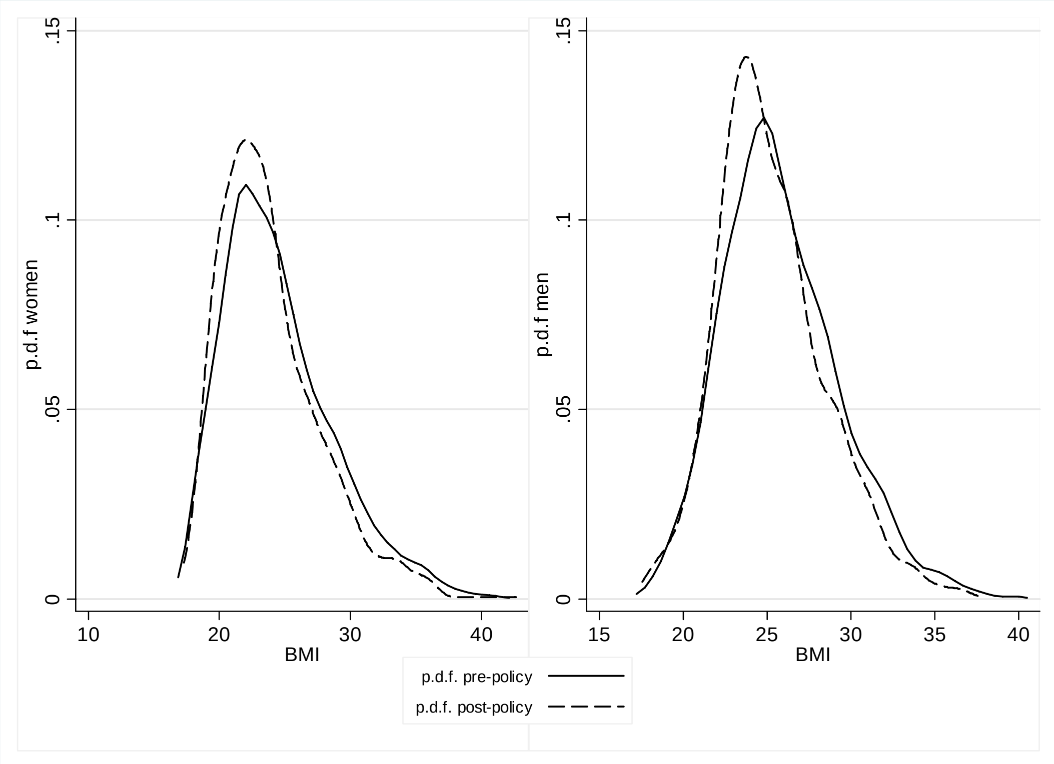

Figure C2. Scenario 4 – Change in BMI vs pre-policy BMI – Sample 2

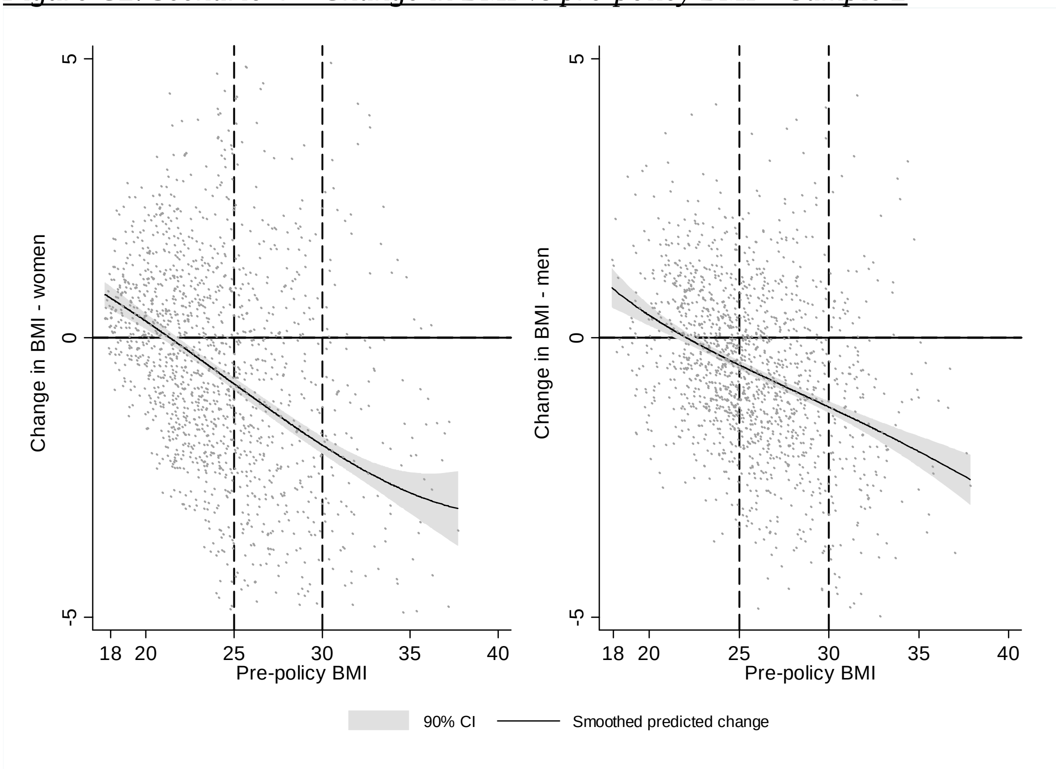

Figure C3. Monotonicity test - Δ(τ)- Women – 23 food categories.

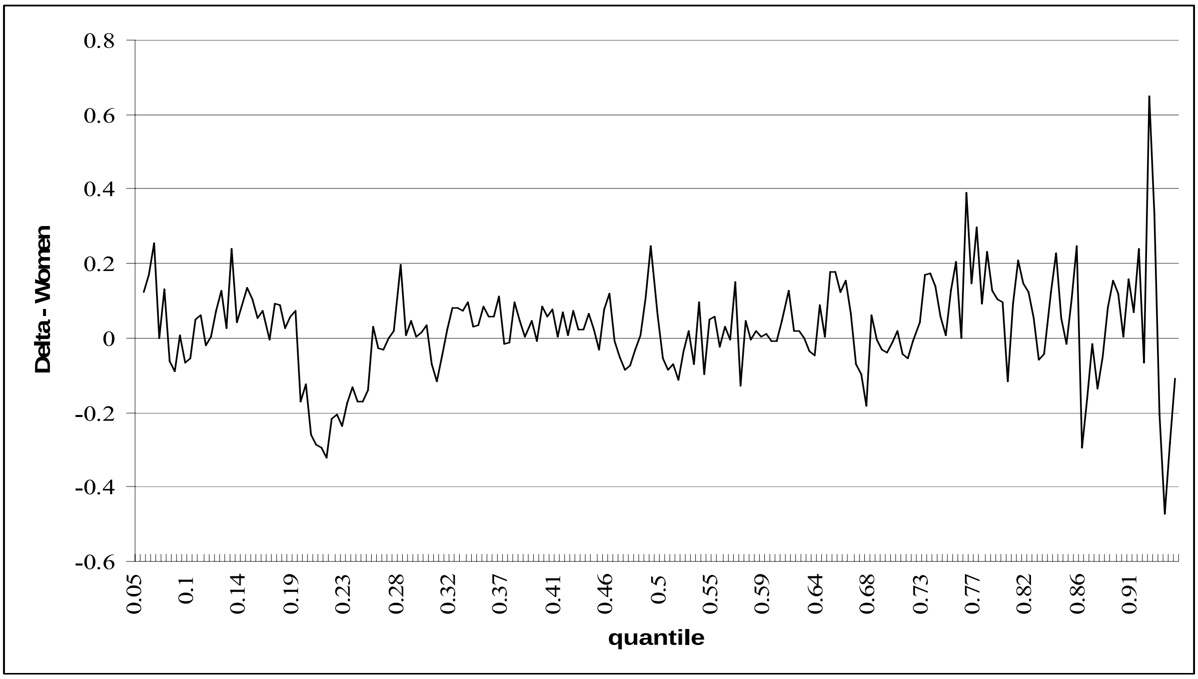

Figure C4. Monotonicity test - Δ(τ) – Men – 23 food categories.

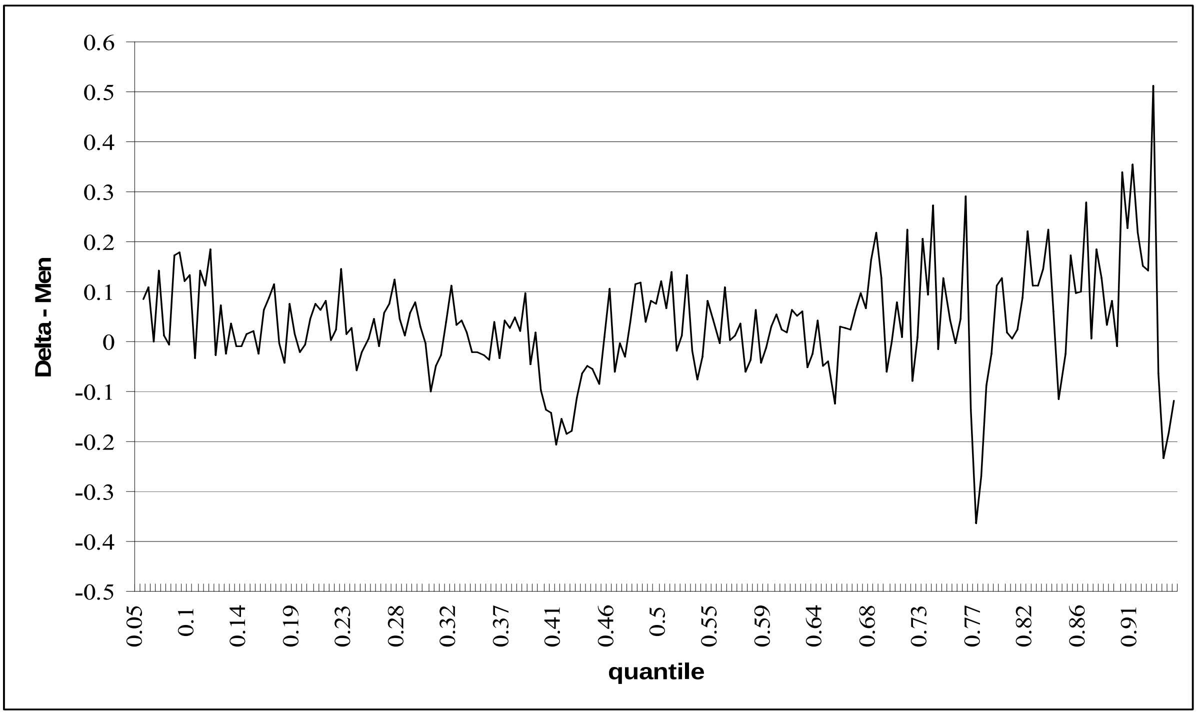

Table C5. Robustness checks for the conditional mean regressions - 23 food categories – Elasticities

Gender	***Women***			***Men***		
Estimator	**OLS**	**OLS**	**Fixed effects**	**OLS**	**OLS**	**Fixed effects**
Sample	Sample 2	Sample 1	Sample 1	Sample 2	Sample 1	Sample 1
Number of observations	6633	11245	11245	5975	9802	9802
Number of individuals	3654	4619	4619	3268	4077	4077
Conditional moment	Mean	Mean	Mean	Mean	Mean	Mean
Price Elasticities						
Water	0.065* (0.038)	0.011 (0.030)	0.010 (0.024)	0.092*** (0.031)	0.052** (0.025)	0.044** (0.021)
Alcohol	0.058 (0.047)	0.042 (0.030)	0.028 (0.022)	-0.013 (0.044)	-0.002 (0.028)	-0.002 (0.021)
Soft drinks	-0.055 (0.068)	-0.109** (0.050)	-0.082** (0.038)	-0.113* (0.059)	-0.118** (0.048)	-0.096*** (0.036)
Fruit in brine	0.134* (0.060)	0.110*** (0.041)	0.080*** (0.029)	0.045 (0.050)	0.076* (0.039)	0.050* (0.028)
Processed Fruit	0.018 (0.053)	0.011 (0.033)	-0.002 (0.025)	0.022 (0.047)	0.002 (0.031)	0.003 (0.024)
Vegetables in brine	-0.025 (0.080)	-0.033 (0.062)	-0.015 (0.047)	0.027 (0.073)	-0.027 (0.058)	-0.019 (0.044)
Processed vegetables	-0.002 (0.060)	-0.017 (0.039)	-0.022 (0.030)	0.016 (0.054)	0.017 (0.037)	0.015 (0.029)
Cereals	0.045 (0.073)	0.088 (0.055)	0.065 (0.041)	0.009 (0.059)	-0.010 (0.046)	-0.003 (0.034)
Meat in brine and eggs	-0.006 (0.088)	0.046 (0.061)	0.034 (0.046)	0.078 (0.079)	0.052 (0.058)	0.041 (0.043)
Seafood in brine	-0.007 (0.035)	-0.003 (0.025)	-0.005 (0.019)	-0.022 (0.030)	-0.017 (0.024)	-0.014 (0.017)
Processed sea products	-0.019 (0.050)	-0.009 (0.030)	-0.005 (0.022)	0.068 (0.044)	0.018 (0.029)	0.015 (0.021)
Cooked meat	0.135* (0.079)	0.088* (0.050)	0.070* (0.037)	-0.042 (0.067)	0.002 (0.046)	-0.006 (0.033)
Breaded proteins	0.005 (0.032)	-0.020 (0.024)	-0.020 (0.018)	-0.047 (0.029)	-0.032 (0.023)	-0.021 (0.017)
Yoghurt and fresh uncured cheese	0.134 (0.112)	-0.001 (0.073)	0.004 (0.055)	0.057 (0.094)	0.025 (0.065)	0.026 (0.049)

Cheese	-0.386** (0.193)	-0.047 (0.111)	-0.034 (0.082)	-0.135 (0.166)	-0.123 (0.107)	-0.080 (0.079)
Milk	-0.020 (0.100)	-0.092 (0.072)	-0.071 (0.055)	-0.185** (0.085)	-0.172*** (0.063)	-0.129*** (0.048)
Animal fat and margarine	-0.021 (0.055)	0.027 (0.036)	0.020 (0.026)	0.032 (0.047)	0.002 (0.032)	0.001 (0.024)
Oils	-0.304** (0.154)	-0.126 (0.092)	-0.099 (0.071)	-0.060 (0.127)	0.057 (0.086)	0.031 (0.065)
Sugar and confectionery	0.046 (0.093)	0.114* (0.066)	0.091* (0.050)	0.018 (0.078)	0.092 (0.060)	0.068 (0.045)
Pastries and desserts	-0.182** (0.078)	-0.092* (0.054)	-0.061* (0.037)	-0.041 (0.071)	-0.085* (0.050)	-0.064* (0.034)
Sweet and fatty snacks	0.136 (0.100)	0.054 (0.061)	0.047 (0.045)	-0.018 (0.083)	0.009 (0.056)	0.004 (0.042)
Salty and fatty snacks	-0.042 (0.088)	-0.036 (0.044)	-0.020 (0.032)	0.033 (0.074)	0.042 (0.042)	0.031 (0.031)
Ready-meals	-0.058 (0.068)	-0.089* (0.051)	-0.070* (0.040)	-0.033 (0.063)	-0.019 (0.047)	-0.017 (0.036)
Income Elasticities						
Income	-0.030*** (0.009)	-0.025*** (0.007)	-0.024*** (0.007)	-0.008 (0.008)	-0.007 (0.007)	-0.007 (0.006)
Other control variables	*CONTHEIGHT, DEG1-DEG6, (AGE/10), $(AGE/10)^2$, BABYWOMAN or BABYMAN, COUPLE, SINGLE, OTHHHOLD, NBIND, FRUITSORVEG, MEALPLANNER, UNIT1-UNIT8, REGION1-REGION8, YR2003-YR2005*					

Note: For each regression and each variable, the point estimates of elasticities at median values of the control variables are shown, with their robust standard deviations in parenthesis; * = significant at the 10% level. ** = at the 5% level. *** = at the 1% level.

Table C6. Sensitivity of Scenario 3's expected results

Scenario		3			
Sample		**Sample 2**		**Sample 1**	
Estimation method		**Quantile regressions**	**OLS**	**Quantile regressions**	**OLS**
		Women			
Prevalence of obesity (BMI ≥ 30)	Actual	10.9%		11.0%	
	Pre-policy (simulated)	10.9%	10.5%	10.5%	8.3%
	Post-policy (simulated)	7.1%	6.9%	6.8%	6.3%
% BMI ≥ 27	Actual	23.5%		23.6%	
	Pre-policy (simulated)	24.3%	25.7%	21.8%	25.8%
	Post-policy (simulated)	18.4%	21.1%	16.5%	21.2%
Prevalence of overweight (BMI ≥ 25 & BMI < 30)	Actual	26.1%		26.2%	
	Pre-policy (simulated)	26.8%	35.2%	27.5%	35.3%
	Post-policy (simulated)	23.1%	31.3%	21.7%	31.5%
		Men			
Prevalence of obesity (BMI ≥ 30)	Actual	11.2%		10.8%	
	Pre-policy (simulated)	12.5%	10.5%	10.8%	10.2%
	Post-policy (simulated)	8.6%	6.9%	7.7%	7.1%
% BMI ≥ 27	Actual	31.1%		30.8%	
	Pre-policy (simulated)	33.3%	34.5%	30.4%	34.2%
	Post-policy (simulated)	25.5%	26.6%	23.7%	27.1%
Prevalence of overweight (BMI ≥ 25 & BMI < 30)	Actual	40.5%		41.0%	
	Pre-policy (simulated)	42.3%	46.4%	42.5%	46.3%
	Post-policy (simulated)	38.2%	41.0%	36.8%	41.5%
		All			
Prevalence of obesity (BMI ≥ 30)	Pre-policy (simulated)	11.7%	9.3%	10.6%	9.2%
	Post-policy (simulated)	7.8%	6.5%	7.2%	6.6%
% BMI ≥ 27	Pre-policy (simulated)	28.6%	30.0%	26.0%	29.8%
	Post-policy (simulated)	21.8%	23.7%	20.0%	24.1%
Prevalence of overweight (BMI ≥ 25 & BMI < 30)	Pre-policy (simulated)	34.3%	40.6%	34.7%	40.6%
	Post-policy (simulated)	30.4%	36.0%	29.0%	36.4%
Reduction in health care expenditure (in million Euros)	Obese (P1)	960	661 [384, 910]	834	659 [384, 926]
	BMI ≥ 27 (P2)	2133	1894 [1064, 2677]	1864	1887 [1067, 2718]

Note: In the "OLS" columns, the two last lines give the mean reduction in health care expenditure with the associated 95% confidence interval; both statistics are computed by bootstrap (1500 draws).