The effects of a ”fat tax” on the nutrient intake of French households

Olivier Allais, Patrice Bertail, Véronique Nichèle

Unité de recherche 1303

alimentation et sciences sociales

THE EFFECTS OF A "FAT TAX" ON THE NUTRIENT INTAKE OF FRENCH HOUSEHOLDS

ALLAIS Olivier, BERTAIL Patrice, NICHELE Véronique

JUIN 2008

Working Paper ALISS 2008-03

INRA UR 1303 ALISS
65, Bd de Brandebourg
94202 Ivry-sur–Seine Cedex
France

http://www.paris.inra.fr/aliss

THE EFFECTS OF A "FAT TAX" ON THE NUTRIENT INTAKE OF FRENCH HOUSEHOLDS

ALLAIS Olivier[1], BERTAIL Patrice [2], NICHELE Véronique[1]

[1] *INRA, UR1303 ALISS, F-94200 Ivry-sur-Seine, France*

[2] *Université Paris 10, MODAL'X. Modélisation aléatoire de Paris X, Paris, France*

Abstract: This article assesses the effects of a "fat tax" on the nutrient intake of French households across different income groups using a method that estimates the nutrient elasticities of French households. The incompleteness of our data requires the use of a cohort model. We estimate a complete demand system by aggregating an individual demand system over cohorts. We find that a "fat tax" would have ambiguous and extremely small effects on the nutrient intake of French households, and its associated economic welfare costs would be similarly weak.

Keywords: HOUSEHOLD SURVEY DATA ; COHORT ; DEMAND SYSTEM ; NUTRIENT ELASTICITIES ; MISSING DATA ; FAT TAX POLICY

Codes JEL : D12, C33

French food policy authorities are questioning the efficacy of a special tax in foods high in calories, fat, or sugar on household purchases and nutrient intake. This tax is generally called the "fat tax" or "junk food tax". It would probably decrease their consumptions, or at least function as a disincentive to unhealthy eating, and generate revenue earmarked to support health measures: improving diet by subsidizing healthful foods, increasing physical activity, obesity prevention, nutrition education, etc. In this paper, we study the question of the efficacy of a fat tax by estimating a complete food demand system using French household data collected by TNS Worldpanel between 1996 and 2001. Prices and nutrient elasticities are then calculated to determine whether a fat tax can substantially alter French household intake of saturated fat and cholesterol as well as the consequences of such a policy on the intake of other nutrients, and how much it affects household economic welfare in the short term. The study period was chosen to evaluate the effects of the fat tax immediately prior to the implementation of the first national nutrition and health program in France.[1] Thus, the effects of informational programs are removed from our analysis.

The estimation of a complete food demand system allows us to take into account all of the effects of prices and income on consumption and on nutrient intake when evaluating food and health policies (see Beatty and LaFrance 2005, and Huang 1996). Most of applied research on French household consumption, however, has been based on incomplete food demand systems using either some strong separability assumptions, as in Bertail and Caillavet (2008), or a methodology developed by LaFrance and Hanemann (1989). One of the main reasons for using incomplete systems is the difficulty of obtaining complete information -including prices, expenditures, and budget shares-on a large set of food products for all households sampled. Indeed, expenditures, quantities, and supply prices for some food categories for any given household are not recorded in TNS Worldpanel data.

To solve the problem of data incompleteness, we propose a cohort model obtained by aggregating the Almost Ideal Demand System (AIDS) over cohorts. Such methodology was suggested a long time ago by Deaton (1985). The cohort construction was devised to compensate for the lack of panel data, but, in our case and in related studies, such as Cardoso and Gardes (1996) and Gardes (2005), the cohort construction serves to solve the problem of missing data.[2] The crucial step in this approach is to estimate the unobserved variables using a pre-model over cohorts. One of the contributions of this paper is to show how the aggregation process leads to induced bias and heteroscedasticity.

In this study, the general method is used to estimate nutrient elasticities for 32 nutrients in response to changes in 24 food category prices for the period 1996-2001. In particular, households are segmented into six regions of residence, four income groups, and three age classes. Then, the methodology of Huang (1996), which consists of applying a nutrient conversion matrix to the price elasticities, is used to derive nutrient elasticities. Finally, the latter are used to examine whether a fat tax would change household nutrient intake. In similar works, Chouinard et al. (2005) and Kuchler, Tegene, and Harris (2005) assessed the impact of fat taxes on U.S. household food purchases but not their nutrient intake. These assessments were based on an incomplete demand system, so only some of the impacts of food policy reforms on household behavior, and particularly on the substitutions between products encouraged by these reforms, could be assessed.

The household food acquisition data contained in the TNS Worldpanel database have advantages over the results of individual food surveys more commonly used to study the relationship between food and health in France. Households respond over a longer period of time (an average of 4 years in the TNS Worldpanel survey), which enables the observation of long-term behaviors and avoids the well-known biases of individual food surveys. In the individual surveys, respondents may over- or under-report their consumption of certain foods of high or low nutritional value, respectively, either because they wish to lie or because they did in fact increase or reduce their consumption deliberately for the short period of the survey. However, these data do not take into account the effects of food purchased away from home and do not reach the level of individual choice. Chesher (1997) and Allais and Tressou (2008) developed non-parametric methods to decompose a series of household quantities into individual quantities. But, in light of the incompleteness of our database and the errors of approximation that follow from decomposition methods, as underscored by Allais and Tressou (2008), we chose not to use these methods in the present study.

We find fat tax to have ambiguous and extremely small effects on household nutrient intake. The effects are ambiguous because increasing the price of a food to reduce intake of nutrients deemed less desirable for health generally reduces intake of other nutrients deemed good for health. For example, reducing saturated fats and cholesterol by taxing dairy products like cheeses and butter reduces household intake of calcium and potassium. Even more importantly, we find nutrient price elasticities to be remarkably inelastic, as in Huang and Lin (2000) and Beatty and LaFrance (2005). These results call into question the effectiveness of tax policies intended to alter nutrient intake, and are consistent with the conclusions of the Chouinard et al. (2005) and Kuchler, Tegene, and Harris (2005) for household food purchases. Tax policies, however, offer the advantage of raising revenue that can be used to finance health measures. Moreover, we find the welfare costs associated with a fat tax on dairy products to be quite weak, and they do not vary much across income class, contrary to the results of Chouinard et al. (2005).

This paper is organized as follows. First, the data and the cohort construction are presented. Second, the estimation over cohort of the unobserved data is described. Then, we lay out the aggregation procedure for the AIDS and highlight the resulting estimation problems. Fourth, the estimation results and the demand and nutrient elasticities are provided. Fifth, the fat tax policy is assessed based on changes in the intake of several nutrients, the level of revenue raised by the tax, and household welfare costs. Finally, we conclude on wondering how would the food industry respond to a fat tax policy, and whether a fat tax could be rather used as a threat to urge on voluntary approaches by food industry to reduce saturated fat in food products.

Data and Cohort Construction

This section begins with a presentation of the data used and an explanation of why the construction of cohorts is needed to estimate a complete food demand system in France.

The Data

The TNS Worldpanel is the principal source of information on food purchases in France. Each annual survey contains weekly food acquisition data of approximately 5,000 households, with an annual rotation of 1/3 of the participants. The households are selected by stratification according to several socioeconomic variables and remain in the survey for a mean period of 4 years. All participating households register grocery purchases through the use of EAN bar codes (Universal Product Code), allowing their purchases to be categorized under such heading as cereals, dairy products, cheese, eggs, sugar, and pastries. To register grocery purchases without a bar code, households are assigned to two groups to alleviate the workload. Each group (half of the survey) is requested to register its "at home" purchases for a restricted set of products: meat, fish, and wine for the first group and fresh fruits and vegetables for the second group. Hence, each group covers a different set of products. Although the two lists together include nearly all possible food products at a very disaggregated level, this method means that purchasing information is never complete for a given household. In other words, expenditures, quantities, and supply prices are missing for some food categories for any given household. This has strong implications for micro-econometric studies of food consumption in France: it means that complete demand systems cannot be estimated for the country. It is then impossible to estimate the total impact of nutritional or price reforms on household behavior, particularly the product substitution that these reforms seek to encourage.

To solve the problem of data availability resulting from data structuring, we follow the methodology of Deaton (1985) by using cohorts to estimate the demand system on TNS Worldpanel data for the period 1996 to 2001. We considered carefully how to categorize different food products. Thus, to facilitate the estimation procedure and to reduce the number of parameters to be estimated, the food items are grouped into 24 categories based on similarities in the nutritional content of the products. Our study, therefore, focuses on the following categories of goods: cereal-based products or grain products, including bread, rolls, biscuits, pastries, pasta, rice, wheat flour, and cereals; potatoes; fresh vegetables; processed vegetables; fresh fruits; processed fruits; nuts; beef and veal; pork; poultry; delicatessen (meat products); eggs; other meats; fish and sea foods; mixed dishes (pizza, choucroute, cassoulet, etc.); dairy products; cheese and butter; oils; sugars; chocolate and sweets; mineral and spring waters; non-alcoholic beverages (fruit juices and soda); alcoholic beverages (including wine); coffee and tea.

Potatoes are not included in fresh or processed vegetables since their nutrient content is very different from that of other vegetables. Moreover, we have chosen to group butter with cheese rather than oils because we wanted to distinguish vegetable fats from animal fats. All the quantities and prices of these categories of goods are expressed in the same unit (kilogram, and French franc per kilogram) to ensure that the demand model used to estimate elasticity is "Closed Under Unit Scaling" (CUUS), meaning that the estimated economic effects are invariant to a simultaneous change in unit, as stressed by Alston, Chalfant, and Piggott (2001).

We then determined the quantities of the 32 nutrients in the 24 food categories, based on consultations with nutritionists and the composition tables of food products developed by Favier et al. (1995). The nutrients of interest are energy (measured in food calories); fat, subdivided into saturated (red meat, egg, whole milk, etc.), monounsaturated (olive oil, canola oil, peanut oil, etc.), and polyunsaturated (oils from corn, soybean, safflower, cottonseed, fish, etc.); cholesterol and alcohol; proteins, subdivided into vegetable and animal protein; carbohydrates; dietary fibres;

micronutrients such as vitamin A (retinol and beta-carotene), B vitamins (1, 2, 3, 5, 6, 9, 12), vitamin C, vitamin D, and vitamin E; and minerals (calcium, iron, magnesium, sodium, phosphorus, potassium).

Cohort Construction

The population is split into homogeneous cohorts based on the following three variables:

(1) a geographical variable that indicates the region of residence of the household. Adjacent regions where traditions and food purchasing patterns show significant similarities are grouped together. French regions are compared according to whether they over- or under-consume the 24 homogeneous categories of goods relative to the national average. For example, the North-Pas de Calais, Picardy, and East regions are aggregated into one region since all three show over-consumption of potatoes, butter, animal fats, and meat products and under-consumption of fresh fruits and oils. This approach leads to six regional modalities: Paris and its suburbs; North-Pas de Calais, Picardy, and East regions (Lorraine, Alsace, Champagne-Ardenne); South-East regions (Provence, Alps, Côte d'Azur); South-West regions (Poitou-Charente, Aquitaine, Midi-Pyrenees, Languedoc-Roussillon) with Limousine and Auvergne; Brittany, Western Loire, and Normandy; and the Centre region (Bourgogne, Franche-Comté, Rhône-Alpes, and the Savoy).

(2) A socioeconomic classification of the households constructed by TNS Worldpanel. This classification scheme comprises four modalities. The first modality contains the households with the highest levels of income; the second includes households whose income is above the national average; the third comprises the households whose income is below the national average; and the fourth contains the households with low income levels.

(3) An age variable indicating the age of the head of the household. The modalities of the variable were chosen to reflect changes in total energy expenditure (TEE) and basal metabolic rate (BMR) with age, assuming a male with normal weight (defined

as body mass index 18.5-25.0 $kg\ m^{-2}$). Roberts and Dallal (2007) show that the TEE is relatively similar up to 60 years of age, beyond which it decreases moderately. Furthermore, Poehlman (1992) find that the decline in BMR with age may not be linear, with a breakpoint occurring around 40 years. For the present study, the age of participating heads of household in 1996 is divided into three modalities: less than 40; between 41 and 54; and above 55. This split was made to ensure that the number of households in a cell is never below 20, the importance of which was emphasized by Verbeek and Nijman (1992).

This set of variables enables us to detect the likely differences in dietary intake patterns across regions of residence, income, and age. The effects of the region of residence and age are available upon request. We get 72 cells, which represent typical households for a given region, income level, and age of the head of household, and each cell contains sufficient number of households, as table 1 illustrates.

Estimation of the Unobserved Data

The problem of unobserved data is addressed using cohorts. Unobserved consumption and expenditures in the 24 food categories defined above are estimated for 13 four-week periods over six years. The estimates are based on the mean consumption and price values over all households in a given cohort. We first define the cohorts, and then we detail the estimation procedure.

Estimation of Unobserved Quantities and Expenditures

In the following discussion, $\mathcal{S}_k,\ k = 1, 2, 3$ designates, respectively, the sub-panel corresponding to group 1 (meat, fish, and wine), the sub-panel corresponding to group 2 (fresh fruits and vegetables), and, finally, the sub-panel of products that are registered by all households. Notice that one household appears in both sub-panels S_1 and S_3 or in sub-panels S_2 and S_3. Moreover, we denote $\mathcal{P}(i) = \mathcal{S}_l$ if $i \in \mathcal{S}_l$. So writing $h \in \mathcal{P}(i)$ (resp. $h \in \mathcal{P}(i)^c$) means that the household h has (resp. has not) registered the product i.

In the TNS Worldpanel dataset, we never observe the full consumption of any given household basket. Thus, items are aggregated at the cohort level using the following procedures. For a given cohort $c \in \{1, ..., C = 72\}$, at time $t \in \{1, ..., T = 78\}$, we observe $N_{c,t}$ households in the corresponding cohort denoted $H_{c,t}$. Let Y_{iht} be some variable of interest in a food category i of a household h at period t which is observed only in the subpanel $\mathcal{S}_l$, $l = 1, 2, 3$. The unobserved value for a household h at period t who belongs to the cell $H_{c,t}$ but who does not register the product i is predicted by the mean $\overline{Y}_{ict}$, over the households in a cell c at period t to whom product i is registered.

Estimation of Unobserved Prices

Similar to the situation with food quantities consumed, not all supply prices are captured in the TNS Worldpanel database. Generally, food prices are approximated by unit values obtained by dividing expenditures by quantities purchased for a given good. In the present study, however, unit values cannot be calculated for each household since we do not observe all the expenditures and quantities purchased for any given household. Second, the unit value is not the supply price of a good, as it reflects both its average market price and consumer choices of food quality: two different households subject to the same pricing scheme may well exhibit different unit values because food items purchased by households have different qualities.

The first problem is addressed by approximating the unobserved quantities and expenditures for any given household using the cohort method described above. Yet contrary to unobserved quantities and expenditures, the unit values are constructed across regions to capture variations in market prices induced by transportation costs. This means that unit values within regions are constant. This aggregation process also attenuates the second problem. The unit values that it provides are used below as prices for households.

The second problem is further addressed using a procedure similar to that of Park and Capps (1997). Prices are quality-adjusted by regressing the log of unit values on total household food expenditure and household characteristics that may affect the choice of food quality, such as income level, household composition and size, and the education level of the principal household earner. We find significant food quality effects for each food group at the 5% percent level, with the exception of processed vegetables and non-alcoholic beverages. However, these effects are much smaller than those obtained by Huang and Lin (2000) since our study aggregated data over a longer period of time (four weeks vs. seven days) and for a larger group of consumers (aggregated across cohort vs. no aggregation). Estimation results are not presented here to save space, but complete results are available upon request.

Aggregating AIDS Model: A Cohort Model

The total household food expenditure cannot be directly calculated for a given household in the TNS Worldpanel database. As a consequence, this variable must be extrapolated for each cell. In the following discussion, we describe the AIDS model and propose a simple model for estimating the total household food expenditure of each household as well as the shares that will be compatible with the aggregation of the AIDS system. The consequences of the aggregation for the estimation, in terms of bias and heteroscedasticity, are carefully examined in the last subsection.

The AIDS Model

We focus, here, on a standard AIDS model developed by Deaton and Muellbauer (1980a). Quadratic AIDS models (see Banks, Blundell, and Lewbel 1997) are more flexible, but the non-linear quadratic term in these models makes them difficult to aggregate and estimate when considering cohorts. However, Banks, Blundell, and Lewbel (1997) show that the AIDS is unlikely to be rejected for most food items. Other models, particularly in the framework of an incomplete demand system, have been proposed by LaFrance (1990), LaFrance et al. (2000), and Beatty and LaFrance (2005) based on the work of LaFrance and Hanemann (1989). The purpose of these models is essentially to propose incomplete demand models consistent with standard microeconomic theory. To avoid some complications induced by the non-linearities in their models (see, for instance, the box-cox transformation model proposed in LaFrance et al., 2000), we do not apply any non-linear transformation to our data prior to analyzing it. Another reason for not using this kind of model is that our database contains gaps in the unit values for all the households in our database. We now recall a few facts about the AIDS model.

In the framework of the household production model, the consumption behavior at the household level during period t can be described with an AIDS by replacing unit values for prices. As such, in this framework, the budget share w_{iht}, for product i, household h, and time t is given by

$$w_{iht} = \mu_{ih} + \sum_{j=1}^{N} \gamma_{ij} \ln v_{jht} + \beta_i \left[\ln x_{ht} - \ln a(v_{ht})\right] + u_{iht}, \tag{1}$$

for $i = 1, .., N$ food categories and $h = 1, .., H$ households, where $a(v_{ht})$ stands for the price index given by $\ln(a(v_{ht})) = \mu_0 + \sum_{i=1}^{N} \mu_{ih} \ln v_{iht} + \frac{1}{2} \sum_{i,j=1}^{N} \gamma_{ij} \ln v_{iht} \ln v_{jht}$. The variable v_{iht} stands for the unit value of a category of goods i for household h at period t. The variable x_{ht} stands for the total expenditure of household h at period t; and α_i, γ_i, and β_i are the parameters to be estimated. To take into account the heterogeneity of behavior, the parameter μ_{ih} is modelled as a linear form $\mu_{ih} = \alpha_{i0} + Z_h \alpha_i$, where $Z_h = (Z_{kh},\ k = 1, ...K)$ is a vector $(1, K)$ of household characteristics. We denote as $\mathcal{I}_{iht} = \{(v_{jht})_{j=1,...,N},\ \ln(x_{ht}),\ Z_h\}$ the set of all explanatory variables for the share w_{iht}. It may be proved that this system is derived from some cost minimization if it satisfies the restrictions imposed by the properties of demands i.e., additivity, homogeneity of degree zero in prices and total household food expenditure together, and the symmetry of Slutsky's matrix. This implies the well-known additivity constraints $\sum_{i=1}^{N} \alpha_{i0} = 1$, $\sum_{i=1}^{N} \alpha_i = 0$, $\sum_{i=1}^{N} \gamma_{ij} = 0$, for all j and $\sum_{i=1}^{N} \beta_i = 0$ and the homogeneity and symmetry constraints $\sum_{j=1}^{N} \gamma_{ij} = 0$, and $\gamma_{ij} = \gamma_{ji}$ for all i and j. All the shares add up to one, giving us $\sum_{i=1}^{N} w_{iht} = 1 + \sum_{i=1}^{N} u_{iht} = 1$. It follows that the u_{iht}

perturbations are not independent. However, it is assumed, that if we drop one share then the u_{iht} perturbations are independent conditionally to the whole information $\mathcal{I}_{iht}$. Of course, this is a strong assumption that can be tested in future works. Even if these residuals are independent, they may not be identically distributed because of product or temporal effects.

The Underlying Cell Models

The AIDS model is based on budget share, which we cannot calculate at the household level in the present study because our database does not capture total household food expenditure. Nevertheless, we have sufficient information in our dataset to predict budget share by aggregating over cells.

We assume that the expenditure of a household h to purchase a product i at time t essentially depends on the characteristics of the cell to which the household belongs

$$(2) \qquad x_{iht} = x_{ict} + \varepsilon_{iht}^{(1)}, \text{ for } i = 1, ..., N, \;\; h \in H_{ct}, \; c = 1, ..., C, \; t = 1, ..., T.$$

Here, the x_{ict} should be seen as the parameters of the model, i.e., the quantities to be predicted. To simplify, we assume that the expenditures of a household are not correlated to each product. We also assume that the expenditures of any household in a given cell are not correlated over time. This is partially true insofar as the households which that belong to H_{ct} are generally not the same and are independent from the ones in $H_{ct'}$, for t' and t far-distant enough. We take into account that the partial correlation in a short period would make the estimation procedure more difficult. For a given household, $\varepsilon_{ht}^{(1)} = [\varepsilon_{iht}^{(1)}]_{1 \leq i \leq N}$ has variance $V(\varepsilon_{ht}^{(1)}) = \Omega_t$, where $\Omega_t = [\omega_{ijt}]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}$ is a $N \times N$ full-rank matrix. For a given product i and a time t, an estimator of x_{ict} is given by $\overline{x}_{ict} = \frac{1}{N_{ict}} \sum_{h \in H_{ct} \cap \mathcal{P}(i)} x_{iht}$, where N_{ict} is the number of households in a cell c at time t for which product i is registered. So, the best predictor of x_{iht} of a household that belongs to H_{ct} and for which we do not observe the expenditure, $h \in \mathcal{P}(i)^c$ [equivalently $h \notin \mathcal{P}(i)$], is $\widehat{x}_{iht} = \overline{x}_{ict}$. It follows that total expenditure for a given household is predicted in an unbiased manner by

$$\widehat{x}_{ht} = \sum_{i=1}^{N} x_{iht} I_{h \in H_{ct} \cap \mathcal{P}(i)} + \sum_{j=1}^{N} \overline{x}_{jct} I_{h \in H_{ct} \cap \mathcal{P}(i)^c}, \tag{3}$$

where I_A stands for the indicator function of event A. Similarly, we define the total predicted expenditure over a cell as $\widehat{x}_{ct} = \sum_{i=1}^{N} N_{ct}\overline{x}_{ict}$, where N_{ct} is the number of households in the cell c at period t.

Finally, the predicted household shares are given by $\widehat{w}_{iht} = \frac{x_{iht}}{\widehat{x}_{ht}}$ if $h \in \mathcal{P}(i) \cap H_{ct}$, and $\widehat{w}_{iht} = \frac{\overline{x}_{ict}}{\widehat{x}_{ht}}$ if $h \in \mathcal{P}(i)^c \cap H_{ct}$; the predicted shares over cells are given by $\widehat{w}_{ict} = \frac{N_{ct}\overline{x}_{ict}}{\widehat{x}_{ct}}$ which clearly satisfy the share equation $\sum_{i=1}^{N} \widehat{w}_{ict} = 1$.

Aggregation of AIDS Model over Cells

For aggregating the model, it is better to write the shares over cells $\widehat{w}_{ict}$ as the weighted sums of household estimated shares. A simple calculation shows that $\widehat{w}_{ict} = \sum_{h \in H_{ct}} \widehat{\theta}_{hct}\widehat{w}_{iht}$, where $\widehat{\theta}_{hct} = \frac{\widehat{x}_{ht}}{\widehat{x}_{ct}}$ and satisfies $\sum_{h \in H_{ct}} \widehat{\theta}_{hct} = 1$. Gardes et al. (2005) propose the same aggregation process, but the main difference in their approach is that, in our study, the total household food expenditure is not known and $\widehat{\theta}_{hct}$ estimates the true share θ_{hct}. By aggregating model (1) over cells, i.e., by reweighting the shares with the estimated values $\widehat{\theta}_{hct}$, for $h \in H_{ct}$, we get

$$\begin{aligned} \widetilde{w}_{ict} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} w_{iht} &= \alpha_{i0} + \overline{Z}_{c,t}\alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln \overline{v}_{irt} \\ &+ \beta_i \left(\overline{\ln x}_{ct} - \overline{\ln(a(v_{rt}))} \right) + \overline{u}_{ict}, \end{aligned} \tag{4}$$

where $\overline{Z}_{ct} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} Z_h$ is the weighted mean characteristic of a cell. By recalling the constancy of unit values within regions, the log unit value of product j over a cell c in region r, we have $\ln \overline{v}_{jct} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} \ln \overline{v}_{irt} = \ln \overline{v}_{irt}$. The weighted mean total log-expenditure of a cell is equal to $\overline{\ln x}_{ct} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} \ln x_{ht}$, and, similarly, the weighted mean price index over a cell c in region r is equal to

$$(5) \qquad \overline{\ln(a(v_{rt}))} = \mu_0 + \sum_{i=1}^{N} \left(\alpha_{i0} + \overline{Z}_{ct}\alpha_i\right) \ln \overline{v}_{irt} + \frac{1}{2} \sum_{i,j=1}^{N} \gamma_{ijt} \ln \overline{v}_{irt} \ln \overline{v}_{jrt},$$

In the end, we note that, since $\overline{u}_{ict} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} u_{iht}$, we have $E\overline{u}_{ict} = 0$ and $V(\overline{u}_{ict}) = \sum_{h=1}^{N_{ct}} E\widehat{\theta}_{hct}^2 V(u_{iht})$.

Estimation of the Aggregated Model

The aggregated AIDS model is estimated using the iterated least squares estimator developed by Blundell and Robin (1999). It amounts to iterating a series of ordinary least squares regressions until convergence on the parameters is reached. Within each iteration, the estimation is performed equation by equation while imposing the constraints of additivity, homogeneity, and symmetry. Thus, in contrast to the approach of Blundell and Robin (1999), the symmetry constraint is directly imposed in the present study.[3]

The main problems in the estimation step are bias and heteroscedasticity, both of which result from the use of estimated variables instead of the true ones, as well as the potential endogeneity of total household food expenditure, as stressed by Blundell and Robin (1999), and Lecocq and Robin (2006). These problems are solved as follows:

(1) Induced bias: Estimating budget shares and expenditure cause a bias problem that should be taken into account. Recall that $\widetilde{w}_{ict} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} w_{iht}$ is unknown and is replaced by the predictor $\widehat{w}_{ict}$. More precisely, if we define $\varepsilon_{iht}^{(2)} = \widehat{w}_{iht} - w_{iht}$, we get over each cell $\widetilde{w}_{ict} = \widehat{w}_{ict} - \overline{\varepsilon_{ict}^{(2)}}$, where $\overline{\varepsilon_{ict}^{(2)}} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} \varepsilon_{iht}^{(2)}$. This creates a bias term -as well as an additional source of heteroscedasticity (see below)-that can be approximated to the first order (see Appendix 1) by

$$(6) \qquad E\overline{\varepsilon_{ict}^{(2)}} \approx \sum_{h \in \mathcal{P}(i)^c} \theta_{hct} \; N_{ict}^{-1} \sum_{k,j} \omega_{kjt} I_{h \in \mathcal{P}(k)^c \cap \mathcal{P}(j)^c \cap H_{ct}}.$$

Notice that this term is not zero if k and j are both in $\mathcal{S}_l$, for $l = 1, 2$; otherwise, the bias is null. This bias effect, which is small if N_{ict} is large, can be seen as a cross effect of the cell and the subpanel. It is directly corrected by introducing the estimated values of $E\overline{\varepsilon_{ict}^{(2)}}$ into the model.

(2) Heteroscedasticity: Due to the aggregation process, the new model becomes heteroscedastic. Notice that, if the u_{iht} is i.i.d., then the variance of the aggregated residual is $V(\overline{u}_{ict}) = V(u_{iht}) \sum_{h=1}^{N_{ct}} E\widehat{\theta}_{hct}^2$. Since $\sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} = 1$, we have from the preceding computations $E\widehat{\theta}_{hct} = O(1/N_{ct})$ and $E\widehat{\theta}_{hct}^2 = O(1/N_{ct}^2)$, so that $V(\overline{u}_{ict}) = O(1/N_{ct})$. In this case, it is possible to correct for most of the heteroscedasticity simply by multiplying each variable defined at the cell level by the square root of the size of the cell. However, if the residual can be decomposed into some fixed effects and a mixed effect, say $u_{iht} = u_i^* + u_t^* + u_{it}^* + u_{iht}^*$, where the components are centered and independent, then, by aggregation, we get $\overline{u}_{ict} = u_i^* + u_t^* + u_{it}^* + \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} u_{iht}^*$ with $V(\sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} u_{iht}^*) = O(1/N_{ct})$ and fixed variances for the other components. In addition, it is also worth noting that $\overline{\varepsilon_{ict}^{(2)}}$, the error introduced by using predictors instead of the true values may be interpreted precisely as a cross effect of the form u_{it}^*. It follows that, if the products and temporal effects are large, then the heteroscedasticity should not be corrected by multiplying the equations by $\sqrt{N_{ct}}$ because this would lead to even greater heteroscedasticity. In these circumstances, for reasons that are unclear, the aggregation process tends to reduce the variance inside the cell (the intra-variance), but the imputation process tends to increase the variance of the predictor of the total household food expenditure and, therefore, of the shares. Since we want to obtain robust estimates, we will essentially use standard two-step methods and generalized least square estimators to correct for the heteroscedasticity.

(3) Endogeneity of total household food expenditure: The log total household food expenditure variable $\overline{\ln x_{ct}}$ and regression residuals $\overline{u}_{ct}$ may be correlated for at least one of the following two reasons: first, either because of simultaneity of the determination of total household food expenditure and budget shares since common shocks

may both determine taste and total household food expenditure changes, and/or second, because of unobserved heterogeneity. Following Blundell and Robin (1999), the first likely source of correlation is usually controlled for by means of instrumental variable techniques, using income as an instrument for total household food expenditure. In particular, we augment the AIDS specification with the residuals υ_{ct} of the regression of the total household food expenditure $\overline{\ln x_{ct}}$ on socio-demographic variables $\overline{Z}_{ct}$, prices $\ln \overline{v}_{irt}$, and the logged income of cohort c at period t, denoted by $\overline{\ln y_{ct}} = \sum_{h=1}^{N_{ct}} \widehat{\theta}_{hct} \ln y_{ht}$.

The likely second source of correlation is corrected as in Lecocq and Robin (2006). Following Mundlak (1978), they show that unobserved heterogeneity can be fully taken into account by integrating the means of the log of income and the log total household food expenditure for each cell c in the set of socio-demographic variables $\overline{Z}_{ct}$, i.e., $\overline{\ln y_{c\bullet}} = \frac{1}{T}\sum_{t=1}^{T} \overline{\ln y_{ct}}$, and $\overline{\ln x_{c\bullet}} = \frac{1}{T}\sum_{t=1}^{T} \overline{\ln x_{ct}}$ respectively. Testing for the absence of $\overline{\ln y_{c\bullet}}$, and $\overline{\ln x_{c\bullet}}$ in the regressions allows direct testing to detect biases due to unobserved heterogeneity.

Finally, we estimate the following aggregated AIDS model over cells, in region r

$$\widehat{w}_{ict} = \alpha_{i0} + \overline{Z}^{*}_{c,t}\alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln \overline{v}_{jrt} + \beta_i \left(\overline{\ln x_{ct}} - \overline{\ln(a(v_{rt}))} \right) + \overline{u}_{ict}. \tag{7}$$

$\overline{Z}^{*}_{ct}$ is composed of two sets of variables: i) a set of variables to overcome the estimation problem described above: the estimation of $\overline{E\varepsilon^{(2)}_{ict}}$ to correct the induced bias, and υ_{ct}, $\overline{\ln y_{c\bullet}}$, and $\overline{\ln x_{c\bullet}}$ to correct the likely endogeneity of total household food expenditure; ii) a set of socio-demographic factors that may influence consumer food choices. Socio-demographic variables include the actual or former occupation category of the household head (self-employed person, white collar worker, *blue collar worker*, no activity); whether (s)he is a retiree; the education level of the principal household earner (no diploma, low degree of diploma, level of bac, *bac, and higher degree*); urbanization (rural, small city less than 10,000 inhabitants, city less than

50,000 inhabitants, *city less than 200,000 inhabitants*, big city, and Paris and its suburbs); the proportion of households in the cell that have a garden, a cellar, and own a home; and the composition of children in the household.[4] The child household composition is characterized by 4 groups: children for age groups 0-5, 6-10, 11-15, and 16-18. We also include the proportion of households in the cell that have at least one child younger than 18. Finally, four-week and annual dummies are introduced in the model. Table 2 displays some descriptive statistics for these variables. These variables are then aggregated over cohorts.

Estimation and Elasticities

Estimation of the AIDS

Table 3 presents some descriptive statistics for expenditure shares and unit values in French francs per kilogram (1 euro was taken to equal 6.55957 French francs). Table 3 also displays the standard errors of the regressions, denoted by RMSE. The close correspondence between simulated values and sample observations indicates that our estimated AIDS is reliable for use in estimating demand elasticities. The goodness-of-fit appears to be satisfactory in the standard of analyzing household survey data, with R^2 value in a range of 0.82 to 0.09. In addition, biases due to unobserved heterogeneity are detected. Table 3 shows that $\overline{\ln x_{c\bullet}}$ is significant for all food items except fresh fruits, beef and veal, meat products, fish, sugars, sweets, water, non-alcoholic beverages, alcoholic beverages, and coffee and tea; and $\overline{\ln y_{c\bullet}}$ is significant for all groups except grain products, poultry, eggs, dairy products, sweets, and coffee and tea. As Lecocq and Robin (2006) also suggests, these results show that the usual instrumentation by income, proposed by Blundell and Robin (1999), is not sufficient on its own to control fully for the endogeneity of total food expenditure in the AIDS.

The parameter estimates of socio-demographic variables are not reported in this paper, but they are available upon request. The results indicate that the larger city households are, the smaller the fraction of their total food expenditure that they allocate to grain products, fresh vegetables and fruits, and poultry. However, they spend a significantly greater fraction of their total food expenditure on beef, meat products, fish, mixed dishes, dairy products, mineral and spring waters, and alcohol. Moreover, we find that the higher the education level of the head of the household, the more they allocate their food budgets to grain products, fresh vegetables and fruits, and poultry and the less they allocate it to beef, meat products, mixed dishes, and alcohol. Similar results were obtained for occupational categories. Having a child younger than 18 contributes significantly to the purchase of more grain products, processed vegetables and fruits, meat products, dairy products, sweets, non-alcoholic beverages, but to fewer purchases of fresh vegetables and fruits, fish, and alcoholic beverages. However, if we focus our attention on households with children aged 0-5, our estimates show that they allocate significantly more of their food expenditure to fresh vegetables and fruits, dairy products, sweets, and less to processed vegetables and fruits, meat products, and non-alcoholic beverages.

Demand Elasticities

Following the approach of Banks, Blundell, and Lewbel (1997), we calculate demand elasticities at the average point. Uncompensated price elasticities are equal to $e_{ij,c,t} = -\imath_{ij} + \widehat{w}_{ict}^{-1}\left[\gamma_{ij} - \beta_i\left(\mu_{ic} + \sum_{j=1}^{n}\gamma_{ij}\ln\overline{v}_{jrt}\right)\right]$, where $\imath_{ij}$ equals one when $i = j$ and zero otherwise for $i, j = 1, ..., N$, and for a cell $c = 1, ..., C$ in region $r = 1, ..., R$. They are all negative (displayed across income only). Figure 1 presents *significant absolute* own-price elasticities across income class, at the 5% level, but they are all negative. The figure shows that low-income households are significantly more sensitive to an own-price change than high-income households for fish, sugar, mineral and spring waters but less sensitive for fresh vegetables and fruits, and alcohol.

Nutrient Elasticities

The nutrient elasticities are calculated following the approach of Huang (1996) and using the demand elasticities calculated above. He shows that combining demand elasticities with the values of the nutrient shares of each composite good category, the nutrient price and total household food expenditure elasticities can be calculated easily. One advantage of Huang's procedure is that a change in a particular food price or total household food expenditure will affect all food quantities demanded through the interdependent demand relationships and, thus, cause the levels of consumer nutrient availability to change simultaneously.

Figures 2a and 2b summarize our results. They display the highest absolute nutrient prices elasticities (all negative) for high-income and low-income households for saturated fat, protein, carbohydrate, fiber, vitamin C, carotenoid, calcium, and iron. The crucial finding is that nutrient price elasticities are inelastic, as Huang and Lin (2000) and Beatty and LaFrance (2005) also found. However, we identify disparities across income class for carbohydrate, fiber, vitamin C, carotenoid, and iron. For example, regarding the effects of food prices on carotenoid intake, figure 2b (panel B) shows that low-income households are less sensitive to variations in fresh vegetables and fruits prices (a 1% decrease of fresh vegetable and fruit prices increases carotenoid intake for high-income and low-income households by 0.33% and 0.11% and 0.17% and 0.07%, respectively), but they are less sensitive to variations in processed vegetable prices.

Simulation of a Fat Tax

In this section, we examine whether a fat tax policy can alter French household intake of saturated fat and cholesterol as well as the consequences of such a policy on the intake of other nutrients. As Chouinard et al. (2005) pointed out, the assessment of the impact of a fat tax policy impacts is relevant only if we assume that the percentage change in targeted food prices is exactly equal to the tax rate. Below, its impact is assessed by calculating (1) the change in nutrient quantities caused by a price variation in a specified food category, (2) the level of revenue raised per household and for the French population, (3) the welfare cost of a fat tax in terms of equivalent variation in total household food expenditure. All the values are calculated at the average point over time for a 1% fat tax policy. However, the effects for other fat tax percentages can easily be calculated since changes in quantity and revenue raised are proportional to the fat tax rate.

Recommendations Versus Facts for Fats

Carbohydrates, fats, and proteins provide the energy in food.[5] To ensure an adequate daily energy supply and lower the risk of chronic diseases, the National Academic of Sciences recommends that 20-35% of calories in a diet should come from fats and no more than 10% from saturated fats.

However, analysis of the nutrient shares of different foods, which refers to the proportion of nutrient $i = 1, ..., 32$ contributed by food category $j = 1, ..., 24$, indicates that the main source of energy is provided by fats: fats and saturated fats contribute, on average, 43.56% and 16.84% of the total caloric intake of high-income and low-income households, respectively.[6] This analysis also indicates that saturated fats provide 42% of total fat intake for high-income households and 40% for low-income households. The individuals nutrient survey INCA, conducted in France in 1999 by AFSSA,[7] see AFSSA (1999), showed that fats contribute, on average, 38.5% of the total caloric intake. This discrepancy can be explained by the over-representation of children (1/3 of the sample), the problem of infrequent consumptions, as well as over- or under-recording of the consumption of certain foods of high or low nutritional value, respectively, in INCA.

Which food categories should be taxed to reduce household intake of saturated fat and cholesterol?

The analysis of the nutrient shares of different foods across income levels reveals interesting differences in the sources of fat across income and may, therefore, provide insightful information to food policymakers. For both high- and low-income households, the main source of saturated fat is the cheese-butter category; foods from this group account for 42.36% and 36.40% of total saturated fat intake for the two classes of household, respectively. The two household groups differ in their main source of cholesterol. For high-income households, it is the dairy products and cheese-butter categories, which provide 21.63% of total cholesterol intake, compared to 19.16% for low-income households. The main cholesterol source for low-income households is eggs, accounting for 19.55% of total cholesterol intake compared to 19.54% for high-income households.

Figure 2a (panel A) provides further support for the results found above by illustrating that, if we want to reduce the amount of saturated fat in household diets, increasing the prices of dairy products, cheese, and butter may be an effective strategy. Surprisingly, we find that grain product prices can be an effective instrument for modifying household fat intake. However, these instruments also affect calcium. For example, a 1% increase in the price of the cheese-butter category (or in the dairy products category) decreases calcium intake by 0.20% (0.44%) in high-income households and by 0.14% (0.51%) in low-income households (figure 2b, panel C). This same increase also reduces sodium intake by 0.10% and 0.12%, respectively.

Reducing cholesterol intake in household diets also has ambiguous effects. A 1% increase in the prices of eggs, dairy products, and the cheese-butter category reduces cholesterol intake by 0.08%, 0.22%, and 0.14%, respectively. However, as shown for saturated fat, a 1% increase in the price of the cheese-butter category reduces calcium intake. Thus, increasing the price of eggs may be the solution to reduce cholesterol in household diets, but it may be less effective than increasing the price of the cheese-butter category for decreasing saturated fat.

The Effects of a Fat Tax

The effects of imposing a fat tax on dairy products and the cheese-butter category on household behavior were also assessed by Marshall (2000), Chouinard et al. (2005), Kuchler, Tegene, and Harris (2005) and Mytton et al. (2007) for Great Britain and the U.S. To the best of our knowledge, the present study is the first time that such an assessment has been carried out in France.

(1) Impact on nutrient intake. The change in nutrient quantity n caused by a price variation in food category i, ΔQ_n, is calculated as in Chouinard et al. (2005),

such as: $\Delta Q_n = \tau Q_n Nut_{n,i}$, where τ is the ad valorem tax rate, Q_n is the average intake of the nutrient n, and Nut stands for the $(l \times N)$ matrix of nutrient elasticities, showing the effects on $l = 32$ nutrients in response to changes in $N = 24$ food prices. If the tax affects a subset I of food categories, then $\Delta Q_n = \tau Q_n \sum_{i \in I} Nut_{n,i}$.

Table 4 reports the average effects of a 1% tax on cheese-butter category on the intake of specific nutrients across income. The calculations suggest that the tax induces very small dietary changes, as also reported by Chouinard et al. (2005) and Kuchler, Tegene, and Harris (2005). In particular, the tax would reduce, on average, saturated fat intake over a four-week period by 3.27 grams and 4.87 grams among high- and low-income households, respectively. To provide a sense of the magnitude of this effect, the average saturated fat intake for high- and low-income households is 1630 grams and 2607 grams, respectively. An additional and positive effect of this tax would be to reduce the quantity of sodium in household diets. However, the tax would also reduce the intake of calcium and phosphorus, especially in low-income households, as seen in the table 4. Implementing a fat tax on cheese-butter and dairy products produces larger effects, but they are still quite small and ambiguous (see table 4).

(2) Revenue raised. Despite the small impact on nutrient intake, the two taxes generate substantial revenue equal to $\tau \ln \overline{v}_n \sum_{i \in I} Q_n \left(1 + \tau Nut_{n,i}\right)$, where $\ln \overline{v}_n$ is the average price over time and regions of the food category n. We find that the tax on the cheese-butter category (or the two categories of cheese-butter and dairy products together) raises an average of 0.18 (0.44) euros per household per four-week period. Chouinard et al. (2005) found that households pay slightly less than \$0.17 per four-week period if a 1% tax on dairy products together is implemented. Given that the 1999 census counted 23.8 million households in France, this corresponds to 4.2 and 10.47 million euros per four-week period for a tax on the butter-cheese category and on both the butter-cheese and dairy products categories, respectively.

(3) The impact on short-run welfare. The short-run welfare cost is defined as the

fall in total household food expenditure that a household living in an environment with no tax is willing to accept while remaining indifferent to living in an environment with a tax. This definition means that the welfare assessment does not include the long-term effects of the tax on household physical health. Its measurement for the aggregated AIDS is developed in Appendix 2. The costs are weak, and total household food expenditure falls to the same extent for both high- and low-income households. We estimate that a low- and high-income household would be willing to accept on average a total household food expenditure reduction of 0.12 (0.25) euros and 0.11 (0.24) euros per four-week period, respectively, instead of facing a tax on the butter-cheese category or on both categories of butter-cheese and dairy products. Contrary to the results of Chouinard et al. (2005), the welfare costs do not vary much across income class. To compare our results with those of Chouinard et al. (2005), we simulate the welfare effects of a 10% tax applied to dairy products together. We get that the average household is willing to accept a total household food expenditure reduction of 35.68 euros per year, while Chouinard et al. (2005) found a reduction of \$22.11. The welfare cost is higher than those reported by Chouinard et al. (2005), but their equivalent variation results are obtained by using an incomplete demand model over dairy products only.

Conclusion

This paper questioned the relevance of a fat tax policy in influencing households' nutrient intakes by estimating a complete demand model.

We developed a cohort model by aggregating AIDS over cohorts, and we precisely analyzed how the aggregation process affected estimations in terms of bias and heteroscedasticity. Especially as the number of data sources available to researchers increases, the cohort method developed here may be useful for combining information obtained from two or more samples drawn from the population. It should be particularly relevant when there is no single sample that contains all relevant variables, as in our case and in many other cases when economists want to combine administrative data sets.

Our general approach was applied to the French sub-panels of the TNS Worldpanel for the period 1996-2001. Following the methodology of Huang (1996), demand elasticities were used to estimate the implied nutrient elasticities across income groups. We identify informative disparities in nutrient price elasticities for carbohydrate, fiber, vitamin C, carotenoids, and iron. However, we find that price nutrient elasticities are highly inelastic, as Huang and Lin (2000) and Beatty and LaFrance (2005) also found for the U.S. We conclude that a fat tax policy is unsuitable for *substantially* affecting the nutrient intake of French households.

All assessments of fat tax policy so far have assumed a fixed set of food products, thereby excluding the possibility of changes in the food industry in response to a fat tax policy. If a tax is implemented, how would the food industry hedge the tax? Would the food industry change the nutritional quality of the taxed products to smooth retail prices and avoid a decrease in sales? Would the food industry modify the composition of the taxed products by substituting them for more expensive components and/or implementing new industrial production processes, thereby making the innovative product less affordable for low-income households? These likely strategies would aggravate socio-economic disparities in the nutritional quality of food selection and may have major implications for health since nutrition is related to the development of certain chronic diseases. Thus, food policymakers need to keep in mind that a fat tax policy may exacerbate nutritional disparities among consumers

Finally, we wonder whether a fat tax could be used as a credible threat to urge on voluntary approaches by food industries to reduce saturated fat in food products. We calculate that if saturated fat in dairy products together is voluntary reduced by 1%, the saturated fat intake would fall on average by 11.51 grams per household per four-week period, all else equal (particularly prices[8] and average quantities consumed). It is more three times efficient than increasing the prices of dairy products together: the prices should increase by 3.52% to get a similar variation.

Acknowledgements: The authors would like to thank Boizot, C., Caillavet, F., Darmon, N., Etilé, F., Lecocq, S, and Soler, LG. This research has benefited from a grant from INRA-INSERM program "A.T.C. Nutrition".

Notes

1) The first program was implemented in 2001.

2) Another advantage of considering cohorts of households is that we never observe null mean consumption for the categories of products considered.

3) In Blundell and Robin (1999), the symmetric restricted parameters are obtained in the second step of the estimation using a minimum distance estimator.

4) The reference modality for each socioeconomic variable is in italics.

5) The energy yield per gram is as follows: Carbohydrate - 4 kcal, Fats - 9 kcal and Protein - 4 kcal.

6) These statistics do not include alcohol.

7) AFSSA is a French public independent organism contributing through monitoring, alert, research and research instigation to the protection and improvement of public health, animal health and welfare, vegetal and environmental health.

8) Constant prices are not so unrealistic regarding the assumed weak reduction in saturated fat.

References

AFSSA. 1999. *Enquète INCA*, J. L. Volatier, ed. Tec & Doc.

Allais, O., and J. Tressou. 2008. "Using decomposed household food acquisitions as inputs of a Kinetic Dietary Exposure Model." *Statistical Modelling* 8:in press.

Alston, J., J. Chalfant, and N. Piggott. 2001. "Incorporating demand shifters in the Almost Ideal demand system." *Economics Letters* 70:73–78.

Banks, J., R. Blundell, and A. Lewbel. 1997. "Quadratic Engel Curves and Consumer Demand." *The Review of Economics and Statistics* 79:527–539.

Beatty, T., and J. LaFrance. 2005. "United States Demand for Food and Nutrition in the Twentieth Century." *American Journal of Agricultural Economics* 87:1159–1166.

Bertail, P., and F. Caillavet. 2008. "Fruit and Vegetables Consumption Patterns: a Segmentation Approach." *American Journal of Agricultural Economics* 90:in press.

Blundell, R., and J. Robin. 1999. "Estimation in Large and Disaggregated Demand Systems: An Estimator for Conditionally Linear Systems." *Journal of Applied Econometrics* 14:209–232.

Cardoso, N., and F. Gardes. 1996. "Estimation de lois de consommation sur un pseudo-panel d'enquêtes de l'INSEE (1979, 1984, 1989)." *Economie et Prévision* 5:111–125.

Chesher, A. 1997. "Diet Revealed?: Semiparametric Estimation of Nutrient Intake-Age Relationships." *Journal of the Royal Statistical Society A* 160:389–428.

Chouinard, H., D. Davis, J. LaFrance, and J. Perloff. 2005. "The Effects of a Fat Tax on Dairy Products." Working paper, Working Paper 1007, Department of Agricultural and Resource Economics and Policy, University of California, Berkeley.

Deaton, A. 1985. "Panel data from time series of cross-sections." *Journal of Econometrics* 30:109–26.

Deaton, A., and J. Muellbauer. 1980a. "An Almost Ideal Demand System." *The American Economic Review* 70:312–326.

—. 1980b. *Economics and Consumer Behavior*. Cambridge University Press.

Favier, J., J. Ireland-Ripert, C. Toque, and M. Feinberg. 1995. *REpertoire Général des Aliments (REGAL): Tables de Composition*. INRA editions, CNEVA-CIQUAL, Paris.

Gardes, F. 2005. "L'Apport de l'Econométrie des Panels et des Pseudo-Panels à l'Analyse de la Consommation." *Economie et Statistique* 324–325:157–162.

Gardes, F., G. Duncan, P. Gaubert, M. Gurgand, and C. Starzec. 2005. "Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption: The Case of US and Polish Data." *Journal of Business & Economic Statistics* 23:242–254.

Huang, K. 1996. "Nutrient Elasticities in a Complete Food Demand System." *American Journal of Agricultural Economics* 78:21–29.

Huang, K., and B. Lin. 2000. "Estimation of Food Demand and Nutrient Elasticities from Household Survey Data." Working paper, USDA- ERS.

Kuchler, F., A. Tegene, and J. Harris. 2005. "Taxing Snack Foods: Manipulating Diet Quality or Financing Information Programs?" *Review of Agricultural Economics* 27:4–20.

LaFrance, J. 1990. "Incomplete Demand Systems and Semilogarithmic Demand Models." *Australian Journal of Agricultural Economics* 34:118–131.

LaFrance, J., T. Beatty, B. Pope, and G. Agnew. 2000. "US Income Distribution and Gorman Engel Curves for Food." Working paper, IIFET 2000 Proceedings.

LaFrance, J., and W. Hanemann. 1989. "The Dual Structure of Incomplete Demand Systems." *American Journal of Agricultural Economics* 71:262–274.

Lecocq, S., and J. Robin. 2006. "Estimating Demand Response with Panel Data." *Empirical Economics* 31:1043–1060.

Marshall, T. 2000. "Exploring a fiscal food policy: the case of diet and ischaemic heart disease." *British Medical Journal* 320:301–305.

Mundlak, Y. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46:69–85.

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

Park, J.L., and O. Capps. 1997. "Demand for Prepared Meals by U.S. Households." *American Journal of Agricultural Economics* 79:814–824.

Poehlman, E. 1992. "Energy Expenditure and Requirements in Aging Humans." *Journal of Nutrition* 122:2057–2065.

Roberts, S., and G. Dallal. 2007. "Energy requirements and aging." *Public Health Nutrition* 8:1028–1036.

Verbeek, M., and T. Nijman. 1992. "Can cohort data be treated as genuine panel data?" *Empirical Economics* 17:9–23.

Appendix 1: Control of the Induced Bias

Using estimated variables instead of the true ones introduces endogeneity. To control the bias induced by this imputation step, we need to calculate expectation of ratios, given that $E\left(\frac{X}{Y}\right) = E\left(\frac{X}{EY(1+\frac{Y-EY}{EY})}\right) \approx \frac{EX}{EY} - \frac{cov(X,Y)}{(EY)^2}$. Applying the formula both to $w_{iht} = \frac{x_{iht}}{x_{ht}}$ and $\widehat{w}_{iht} = \frac{\hat{x}_{iht}}{\widehat{x}_{ht}}$, and recalling that $E(\widehat{x}_{ht}) = E(x_{ht}) = x_{ct}$, and $\varepsilon^{(2)}_{iht} = \widehat{w}_{iht} - w_{iht}$, we have

$$E\varepsilon^{(2)}_{iht} \approx \begin{cases} -\frac{cov(x_{iht}, \widehat{x}_{ht} - x_{ht})}{(x_{ct})^2} = 0, \text{ if } h \in \mathcal{P}(i) \cap H_{ct} \\ -\frac{cov(\overline{x}_{ict}, \widehat{x}_{ht} - x_{ht})}{(x_{ct})^2} \approx \sum_{i,j} N^{-1}_{ict} \omega_{i,j,t} I_{h \in \mathcal{P}(i)^c \cap \mathcal{P}(j)^c \cap H_{ct}} \text{ else} \end{cases}$$

which is small either if everybody has given an information on the product i in the cell or if $\sum_{i,j} \omega_{i,j,t} I_{h \in \mathcal{P}(i)^c \cap \mathcal{P}(j)^c \cap H_{ct}}$ (which is of order N^2) is small compared to N^{-1}_{ict}. This bias is essentially linked to the way the TNS Worldpanel database is constructed.

Appendix 2: Welfare Measurement

Following Deaton and Muellbauer (1980b), we know that the AIDS model is derived from the expenditure function $\ln c(u, \ln \overline{v_r}, \overline{\ln x}_c) = \overline{\ln(a(v_r))} + u\beta_0 \prod_{k=1}^{N} \overline{v}^{\beta_k}_{kr}$, where $\ln \overline{v_r}$ stands for the vector of aggregated price over the region r, $\overline{\ln(a(v_r))}$ is defined in equation (5), $\overline{v}_{k,r} = \exp(\ln \overline{v}_{kr})$, and u a given value of utility. Given our definition of the welfare cost, the equivalent variation in total household food expenditure for cohort c, denoted $\overline{\Delta x}$, is given by the following equality

$$\{\overline{\ln x}_c - \overline{\ln(a(v_{r,1}))}\} \prod_{k=1}^{N} \overline{v}^{-\beta_k}_{kr,1} = \{\overline{\ln(x_c + \Delta x)} - \overline{\ln(a(v_{r,0}))}\} \prod_{k=1}^{N} \overline{v}^{-\beta_k}_{kr,0}$$

where $\overline{\ln(a(v_{r,0}))}$, and $\overline{\ln(a(v_{r,1}))}$ ($\overline{v}_{kr,0}$ and $\overline{v}_{kr,1}$) are the price index (price of food category k) in region r before and after the implementation of the tax respectively. Finally, we find for a given tax τ affecting a subset I of food categories

$$\overline{\Delta x} = \exp\left[\left(\overline{\ln x}_c - \overline{\ln(a(v_{r,1}))}\right)(1+\tau)^{-\sum_{i \in I} \beta_i} + \overline{\ln(a(v_{r,0}))}\right] - \overline{x}_c$$

where $\overline{x}_c = \exp(\overline{\ln x}_c)$, and recalling that the price of the taxed food category i is such that $\overline{v}^{\beta_i}_{ir,1} = [(1+\tau)\overline{v}_{ir,0}]^{\beta_i}$.

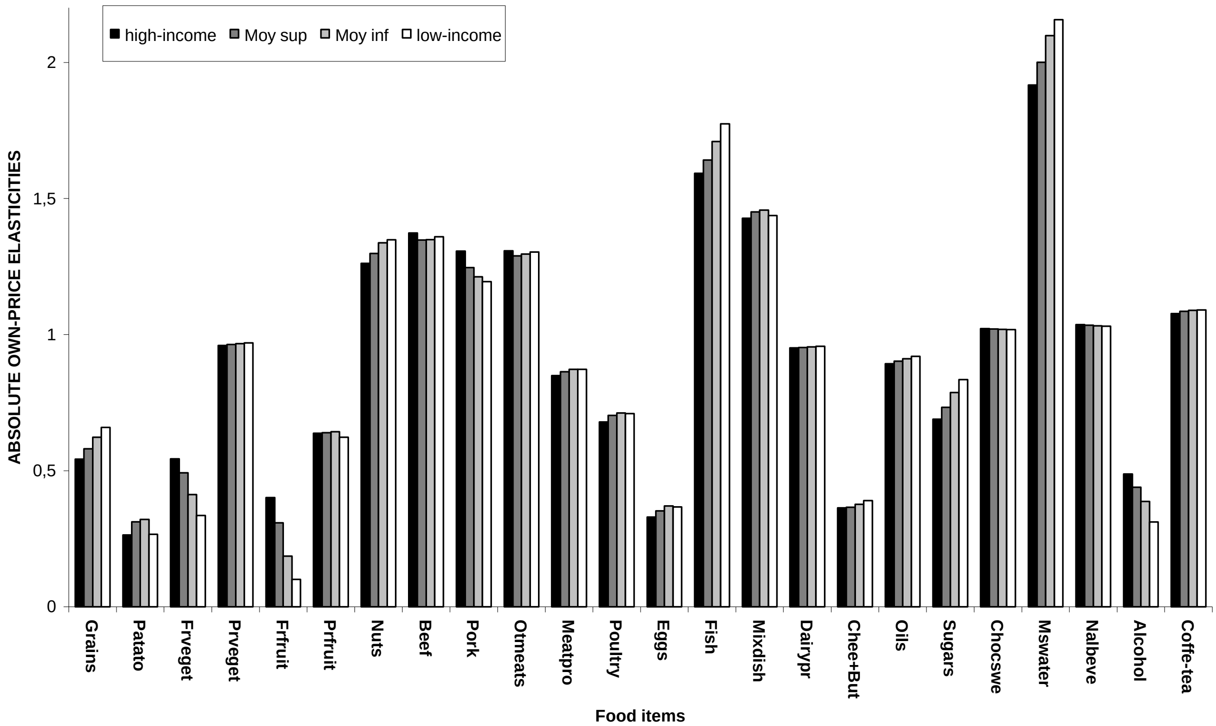

Figure 1: **Absolute own-price elasticities across income class (all negative)**

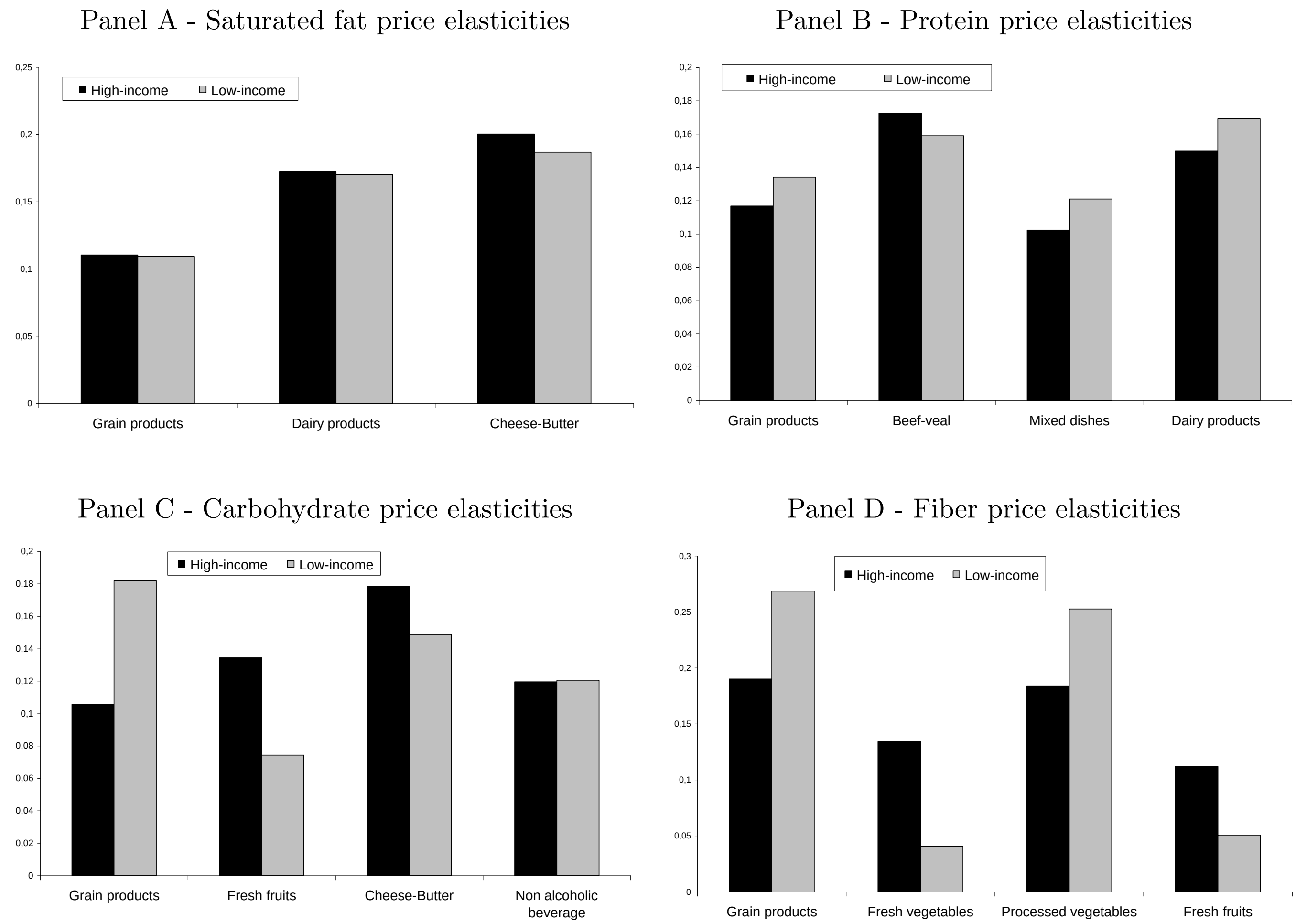

Figure 2a: Highest absolute nutrient price elasticities (all negative)

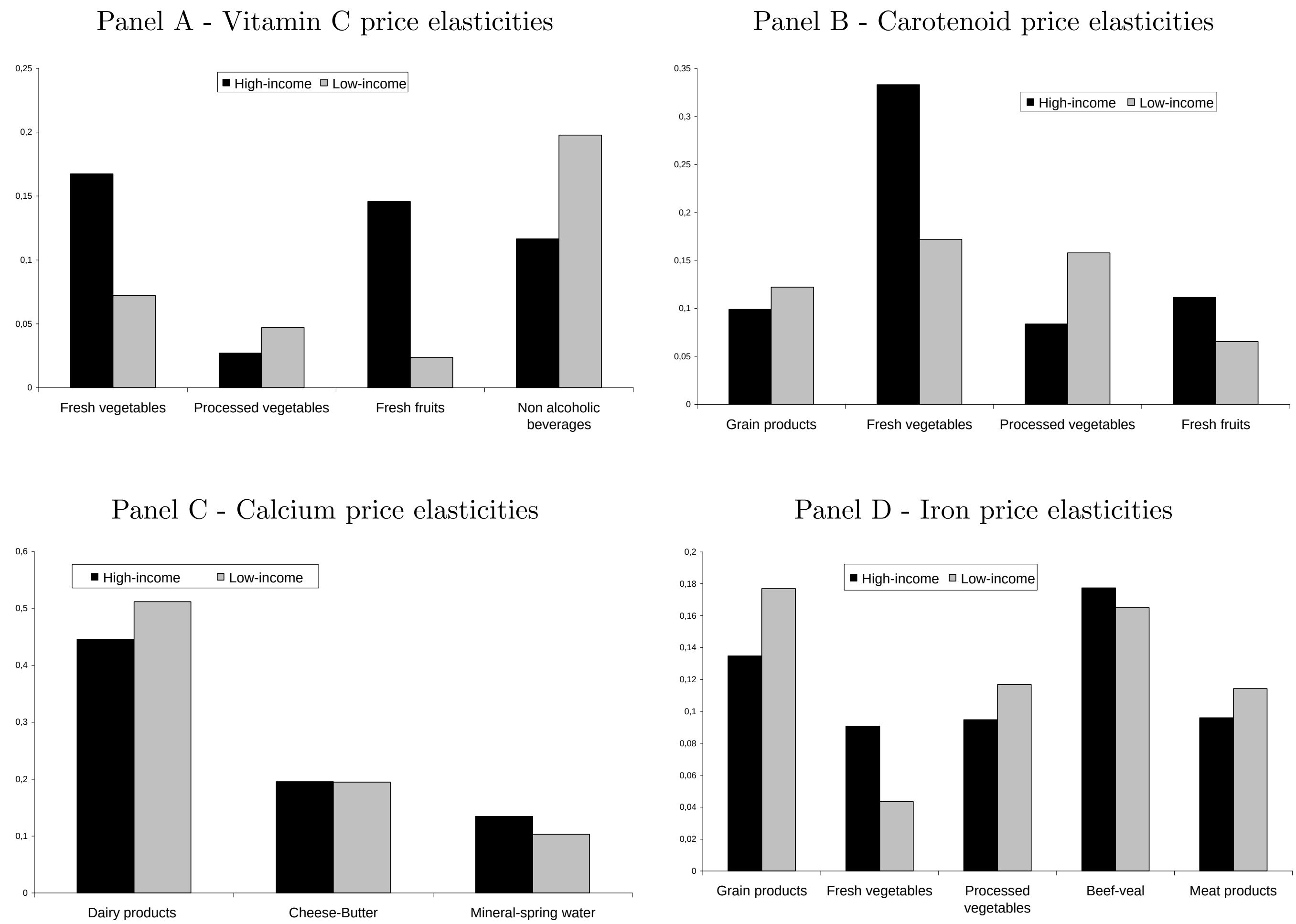

Figure 2b: Highest absolute nutrient price elasticities (all negative)

Table 1: **Descriptive statistics of the number of households in cells**

	Observations	Mean	Standard deviation	Min	Max
N_{ct}	5616	89.38	47.31	26	267

Table 2: **Proportion of households for each sociodemographic variables**

Sociodemographic variables	Mean	Standard error
Occupation category of the household head		
Self-employed persons	0.079	0.001
White collar workers	0.374	0.003
Blue collar workers	0.518	0.003
No activity	0.029	0.001
Level of education of the principal household earner		
No diploma	0.159	0.180
Low degree of diploma	0.351	0.169
Level of Bac	0.176	0.093
Bac and Higher degree	0.314	0.253
Urbanization		
Rural city	0.242	0.143
Small city less than 10,000 inhabitants	0.120	0.070
City less than 50,000 inhabitants	0.128	0.080
City less than 200,000 inhabitants	0.144	0.088
Big city	0.226	0.166
Paris and its suburb	0.140	0.316
Child household composition		
Children for age group 0-5	0.181	0.252
Children for age group 6-10	0.210	0.268
Children for age group 11-15	0.248	0.252
Children for age group 16-18	0.160	0.171
Proportion of households that have at least a child (less 18)	0.418	0.337
Proportion of households with a garden	0.680	0.174
Proportion of households with a cellar	0.749	0.115
Proportion of home owners	0.653	0.173

Table 3: **Sample mean shares and unit values (in Francs by kilo, 1 euro equals 6.55957 French Francs), estimation summary statistics and tests for existing biases due to unobserved heterogeneity; * significant at 5 percent level**

Food categories	Shares	Unit values	RMSE	R^2	$\overline{\ln x_{c\bullet}}$	$\overline{\ln y_{c\bullet}}$
Grain products	0.076	22.87	0.910	0.824	-0.027*	0.002
Potatoes	0.007	4.58	0.506	0.089	0.007*	-0.002*
Fresh vegetables	0.048	11.82	1.320	0.513	0.051*	-0.014*
Processed vegetables	0.020	18.62	0.370	0.498	0.006*	-0.0023*
Fresh fruits	0.054	10.66	1.558	0.531	0.007	-0.008*
Processed fruits	0.005	16.09	0.174	0.378	-0.002*	0.001*
Nuts, dried fruits	0.006	37.57	0.225	0.287	-0.003*	-0.001*
Beef and veal	0.065	61.83	1.682	0.326	0.002	-0.009*
Pork	0.023	37.49	0.796	0.354	0.006*	-0.004*
Other meats	0.019	52.78	0.087	0.384	0.010*	-0.009*
Meat products	0.084	58.79	1.270	0.373	-0.004	0.007*
Poultry	0.037	36.16	0.099	0.229	0.027*	-0.000
Eggs	0.011	17.23	0.020	0.307	0.004*	-0.000
Fish	0.058	52.63	1.577	0.558	0.010	0.009*
Mixed dishes	0.063	31.98	1.174	0.491	-0.079*	0.012*
Dairy products	0.090	9.75	0.970	0.755	-0.024*	-0.002
Cheese and butter	0.083	44.01	0.834	0.423	0.013*	-0.004*
Oils	0.014	17.51	0.272	0.546	0.006*	-0.004*
Sugars	0.006	9.40	0.206	0.588	0.000	-0.002*
Sweets, chocolate,...	0.058	51.47	0.888	0.690	-0.001	-0.002
Waters	0.020	1.87	0.439	0.450	0.003	-0.001*
Non alcoholic beverages	0.031	6.94	0.546	0.745	0.001	0.001*
Alcoholic beverages	0.096	24.74	2.960	0.355	-0.017	0.027*
Coffee, tea (reference share)	0.024	81.29	2.815	0.373	0.003	0.004

ALISS Working Papers

2008

2008-03 Allais, O. ; Bertail, P. ; Nichèle, V. **The effects of a "Fat Tax" on the nutrient intake of French Households,** *Aliss Working Paper 2008-03,* Juin 2008, 36 p.

2008-02 Etilé, F. **Food Price Policies and the Distribution of Body Mass Index: Theory and Empirical Evidence from France,** *Aliss Working Paper 2008-02,* Juin 2008, 52 p.

2008-01 Boizot-Szantaï, C., Etilé, F. **Le prix des aliments et la distribution De l'Indice de Masse Corporelle des Français,** *Aliss Working Paper 2008-01*, Mai 2008, 19 p.

Table 4: **Effect of a 1 percent dairy products and/or cheese-butter tax on specific nutrient quantities (in gram if not specified)**

		Reduction in 4-weeks quantity per household	
Tax Base	Nutrient	High-income	Low-income
Cheese-butter			
	Energy (kcal)	-112.38	-164.42
	Saturated Fat	-3.27	-4.87
	Cholesterol	-0.20	-0.30
	Calcium	-0.65	-1.04
	Phosphorus	-0.63	-0.97
	Sodium	-0.80	-1.20
Cheese-butter and dairy Products			
	Energy (kcal)	-165.01	-255.60
	Saturated Fat	-6.08	-9.31
	Cholesterol	-0.52	-0.79
	Calcium	-2.22	-3.78
	Phosphorus	-1.80	-3.02
	Sodium	-1.29	-2.08