The regional price of junk foods relative to healthy foods in the UK: indirect estimation of a time series, 1997-2009

Sara Capacci*, Mario Mazzocchi, Bhavani Shankar

Sara Capacci and Mario Mazzocchi, University of Bologna; Bhavani Shankar, School of Oriental and African Studies

Contributed Paper prepared for presentation at the 86th Annual Conference of the Agricultural Economics Society, University of Warwick, United Kingdom

16 - 18 April 2012

*Corresponding author: Sara Capacci: sara.capacci@unibo.it, Department of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, 40126 Bologna, Italy.

Abstract

The paper aims at indirectly estimating a time series of food prices from household expenditure data, focusing on foods considered as ‘junk’ relative to healthy foods. The “big 6” among the HFSS (high in fats, sugar and salt) foods identified by the Food Standard Agency have been selected to compose a target ‘unhealthy’ basket, compared to a ‘healthy’ benchmark aggregate including fruit and vegetables. UK data from the National Food Surveys, the Household Expenditure Surveys and the Living Costs and Food Survey were harmonized and merged to compose a set of household level unit values from 1997 to 2009 for the healthy and unhealthy aggregates. Well-established techniques are then employed to estimate regional prices by disentangling the quality component from unit values. The analysis provides information on the geographical and time heterogeneity of estimated price of the unhealthy basket relative to fruit and vegetable prices.

Keywords Price, Unit value, Junk food, Fruit and vegetable, Demand, Elasticity, Healthy diet

JEL code D11, Q11

1. Introduction

The unfavourable evolution of relative retail prices is among the potential determinants of unhealthy food choices investigated in the wide literature on increasing obesity rates and diffusion of unhealthy diets in western countries (Mazzocchi et al., 2009). The declining prices of certain energy-dense food items (due for example to technological innovations in food production and transportation, see (Cutler et al., 2003) have been blamed for the deep worsening of diet quality, although the extent of this effect is unclear. Yet, recent enforcement of fiscal measures as policy instruments for healthy eating promotion in Europe (see Capacci et al. 2012) put prices at the centre of the debate, requiring an accurate evaluation of consumption response and potential tax revenues.

The objective of the present work is the indirect estimation of time series of prices for “unusual” food aggregates, whose nutritional characteristics make them relevant for the investigation of eating habits. The estimation procedure applies to data on expenditures and purchased quantities sometimes recorded within household budget surveys (HBS) and is aimed at controlling for the quality choice component embedded in the observed data and the time-varying heterogeneity of food items composing the consumption aggregates. The analysis focuses on a junk food basket, composed of a set of food items high in fat, sugar and salt, but any aggregation is feasible using the same estimation procedure. As a result an aggregate junk food price index may be observed over time with regard to a healthy benchmark (here assumed to be the fruit and vegetables aggregate). Moreover geographical variables in HBS may allow estimation of regional prices in countries where they are not provided by statistical offices, like the UK.

The analysis is based on UK data from the National HBS over a period of 13 years (1997-2009). Harmonization of two different survey structures due to the transition from the National Food Survey (NFS) to the Expenditure and Food Survey (EFS) was needed to achieve a sufficiently long time span for price trend analysis.

The paper is structured as follows. In section 2 the theoretical background is briefly sketched out, and Section 3 illustrates the method for price estimation. Section 4 presents the data and the harmonization of two main data structures (4.1), together with the definition of the healthy and unhealthy food aggregates (4.2). The application to the UK case is illustrated in Section 5, and results are shown in section 6. Some conclusions are drawn in section 7.

2. Economic background

National household budget surveys (HBS) are a broad source of information for consumers and demand analysis and under certain conditions they can be extremely useful also to draw information on prices of a large amount of items, specifically foods. The first condition is that both expenditures and purchased quantities are recorded. As a result unit values can be computed, as the ratio between expenditures and quantities. Unit values have been extensively employed in the literature in place of prices when price observations were not available. In many cases they have been used in place of prices to estimate demand models and price elasticities. Despite this protracted (and continuing) use, there is a major difference between the two variables and employing unit values instead of prices has some precise implications and risks. Despite unit values depend on market prices, they also embed a quality choice component which can be substantial for prices of quality heterogeneous goods like foods. While prices are exogenous to the single consumer, unit values are observed after the consumer choice has performed and, together with the purchased quantity, they are the output of consumer choice. When choosing their purchase basket, consumers simultaneously choose the amount and the quality level of a given food item. As different quality levels of the same item have different prices, by choosing the quality level, the price to be paid is also chosen. Formally, this quality choice issue can be easily reduced to a pure aggregation matter. If products of different quality were treated as different products, and their different prices could be observed, quality choice would vanish and consumers' choices would be easily described by classical economic theory as a matter of quantity chosen in response to a price level. In the present work we refer to quality as an exact synonymous of aggregation, under the assumption that in a maximum disaggregation scenario no quality differentiation exists (Deaton, 1997). Quality differentiation, which under this framework is any differentiation within the same aggregate item leading to differences in prices, might come from a set of alternative sources: different brands, different purchasing points, different purchased quantities. Hence, a quality component is embedded in any possible price observation not at the highest disaggregation level, and even disaggregation up to the brand level would be not enough, considering point-of-purchase heterogeneity, or offers and discounts. It is interesting to notice that even Consumer or Retail Price Indices (CPIs or RPIs) which are commonly provided by National Statistical offices are also affected by an aggregation problem to some extent.

According to this approach, when unit values are employed instead of prices as explanatory variables to model demanded quantities or share expenditures, two aspects of the same choice are regressed one on another, and a potentially serious endogeneity problem is generated. Ignoring these problems leads to incorrect evaluations of price responses. Price elasticity estimated employing observed unit values might overestimate actual price elasticity because of the reallocation of quality choices within the same consumption aggregate. As a response to a price increase consumers might reduce their purchases of a given food group, but they might also choose lower quality items (hence lower unit values compared to the actual price change) within the same aggregate. Given that the observed increase in unit values is smaller than the actual price increase, a larger responsiveness of demand is estimated. The same mechanism leads to overestimation of elasticities in case of price reductions, as consumers may react by reallocating their basket towards goods with higher unit values.

This topic is well covered in the literature, and mainly refers to several works by Deaton, who has provided a model for the simultaneous choice of quality and quantity (see Deaton 1987, 1988, 1990, 1997) with the aim of estimating price elasticities from unit values[1]. In the present work, we apply a basic procedure for cleaning unit values from their choice component to provide an estimate of actual prices of a food aggregate characterized by specific nutritional characteristics.

3. Method

Drawing from Deaton 1997, we define quality as the value of a specific composition of a food aggregate. Changes in the proportion of the food items composing the aggregate leads to a change

[1] On the same topic see also Crawford, Laisney, and Preston (2003)

in the quality of the aggregate as a whole. As discussed, we assume unit values deviate from market prices because of the quality choice they incorporate. We also presume the purchased amount of a food item affects them inversely. A larger quantity of the same good is likely to cost less. Following Crawford et al. 2003, we model unit values of good *G* according to the following specification:

$$ln\, v_{Ght} = \alpha_G + \beta_G\, ln\, x_{ht} + \sum_l^L \gamma_G z_{lht} + ln\, \pi_{Ght} + \theta_G ln q_{Gt} + u_{Ght} \quad (1)$$

where logarithm of the unit value *v* of the good *G* for the *h*-th household at time *t* is a function of quality choices (reasonably depending on *L* household socio-demographic characteristics *z* and income *x*), purchased quantity of good *G* (q_G) and its (unobserved) market price π_G, plus a normally distributed household specific error term.

As in Deaton and Crawford procedures, a key assumption on market price variability is required to enable estimation of the non-price parameters in equation (1). It is assumed that consumer within the same geographical area at a given time period face the same prices ($\pi_{Ght} = \pi_{Gt}$), while prices are allowed to vary across regions and time periods. It follows that among consumers within the same geographical region and time period, departures from the mean unit value are explained solely by heterogeneity in household characteristics and purchased amounts, so that a regression of mean unit values on household characteristics and quantities allows disentangling the non-price component.

Under this hypothesis and considering each geographical areas at a point in time as a cluster, equation (1) at the cluster mean has the following form:

$$ln\, v_G^c = \alpha_G + \beta_G\, ln\, x^c + \sum_l^L \gamma_G z_l^c + ln\, \pi_G^c + \theta_G ln q_G^c + u_G^c \quad (2)$$

where the upper script *c* indicates the cluster mean, and the price term remains the same of equation (1), because $\pi_{Ght} = \pi_G = \pi_G^c$ within the same cluster.

By subtracting (2) from (1), we obtain a within-cluster equation of deviations from the cluster means, where the constant and the price terms are canceled since they are cluster-specific:

$$(ln\, v_{Gh} - ln\, v_G^c) = \beta_G (ln\, x_h - ln\, x^c) + \sum_l^L \gamma_G (z_{lh} - z_l^c) + \theta_G (ln q_{Gh} - ln q_G^c) + u_{hG} - u_G^c \quad (3)$$

Equation (3) is estimated on all available observations by two-stage least squares (2SLS) to address endogeneity between unit values and quantities and estimates of the non-price coefficients ($\hat{\beta}$, $\hat{\gamma}$ and $\hat{\theta}$'s) are obtained.

Using between-cluster information only, and assuming that cluster-specific measurement errors u_G^c are negligible, it is now possible to obtain an estimate of cluster prices plus a non-identifiable good-specific constant:

$$\hat{p}_G^c = ln v_G^C - \hat{\beta}_G\, ln\, x^c - \sum_l^L \hat{\gamma}_G z_l^c - \hat{\theta}_G ln q_G^c = \alpha_G + ln \pi_G^C \quad (4)$$

where $\hat{p}_G$ represents (the logarithm of) a price index for the good *G*, computed on a monthly-regional basis and cleaned of the quality and quantity components affecting unit values. These price estimates can be fed as explanatory variables into a demand system and they can safely be treated as exogenous variables for aggregate food groups. Employing estimated prices in a classical Almost Ideal Demand System (AIDS) setting for junk food and fruit and vegetable demands enables corrected uncompensated price elasticity estimation.

4. Data

4.1 Harmonization of different survey structures

The procedure is applied on UK household budget data. The aim of observing price variability over a period of time which is long enough to capture differences in trends between different aggregates has prompted us to merging two different survey structures. The actual Living Costs and Food Survey (LCS), known as the Expenditure and Food Survey (EFS) from 2001 to 2007, was combined with the National Food Survey (NFS) which existed before 2000, in order to get a multiple cross-sections of 13 years, from 1997 to 2009. The Family Expenditure Survey, also available before 2000, was not considered because purchased quantities were not recorded in the diary. Since the LCS (and EFS) and NFS have different structures, data had to be harmonized. Harmonization factors provided by the Department for Environment Food and Rural Affairs (DEFRA), which is responsible for the food module of the UK HBS, have been applied to NFS data in order to make aggregate expenditures estimates compatible with those from EFS data. Measurement units of purchased quantities have been turned from the Imperial System (in NFS) into the Metric System (EFS) and a further adjustment has been made to account for different reference periods between the NFS (one week) and the LCS (two weeks). Moreover, because of the unavailability of adjustment factors for eating out observations, we have excluded eating out purchases from the whole dataset. In order to check for the correctness of the harmonization process of food data, we have exactly reproduced aggregated DEFRA expenditures and purchased quantities estimates over the 13 years. Some further minor adjustments were applied to demographic variables for harmonization. These concerned household composition, geographic location and household income.

The resulting multiple cross-sections dataset contains unit values for more than 250 food items, recorded on about 6000 households each year, over the 13 years time span (from 1997 to 2009).

According to the region of residence of each household and the month in which the interview took place the household observations were grouped into regional cluster observed at the same point in time. A total of 144 clusters per year (12 regions, and 12 months) was obtained.

4.2 Definition of the junk food basket and the healthy benchmark

The inclusion criteria for an unhealthy food aggregate (the ‘junk food’ basket) may be a very controversial exercise, which goes beyond the scopes of this work. Frist, single food items can hardly be blamed to be unhealthy per se, as healthiness should rather refer to the diet as a whole. Hence, nutritional composition of foods should be accompanied at least by information on the frequency of consumption before being employed as discriminating criteria for detecting its healthiness. Moreover even if the UK household budget survey includes information on nutritional composition of foods at a certain level of disaggregation, nutrient contents vary a lot among different types or brands of food belonging to the same food group, and this important variability is not observable through the available HBS data.

For our study, we refer to an existing definition, which relates to the “big 6” criterion by the Food Standard Agency. According to this definition, individual food brands are categorized as high in fats, sugar and salt (HFSS) or non-HFSS for the purposes of enforcing the OFCOM advertising regulation. Confectionery, soft drinks, crisps/savoury snacks, fast food, pre-sugared breakfast cereals and pre-prepared convenience foods are the “big 6” categories.

Expenditures of the big 6 have been aggregated to compute a comprehensive junk food expenditure variable, and a junk food unit value is computed for each household by dividing expenditures and purchased quantities.

For the sake of simplicity and for the uncontroversial positive role they have in diets, we have adopted fruit and vegetables (excluding potatoes) as the benchmark healthy aggregate.

The choice of the food aggregates may affect results at least in absolute levels if not in trends. Yet, other aggregation strategies might be applied to investigate different food groups, both for testing robustness of our ‘junk food’ exploration, or to investigate price trends for specific nutrients dense foods (e.g. food high in saturated fat).

5. Application to UK expenditures data

Corrected prices for a set of food items and for the junk food and the fruit and vegetable (F&V) aggregates have been estimated according to (3) and (4). Twelve UK regions[2] have been employed as geographical units for the definition of monthly clusters, for a total of 144 clusters per 13 years. Together with purchased quantity and total household income, the following household characteristics are employed as relevant to the quality choice: age of the household head, the number of female members and a set of dummies for the family composition (one adult only, two or more adults only, adult/s plus children). Total household size has been used as instrument in the 2SLS estimation of the unit value equation, to address the endogenous choice of unit values and quantities.

Estimated prices for the junk and the healthy aggregates have been fed into a two-goods demand system specified as a linearised AIDS, following Deaton and Muellbauer 1980:

$$w_{Gh} = a_{0G} + \sum_{l=1}^{L} a_{1Gl} z_{lh} + \sum_{H} g_{GH} ln\hat{p}_{H} + \mathrm{b_G}(ln\frac{x_h}{P^*}) + e_{Gh} \tag{5}$$

where the *h*-th household’s budget shares *w* of each food aggregate (*G*=2, junk food and F&V) are function of household characteristics *z*, the estimated prices $\hat{p}$ and total food expenditure (*x*) deflated by the Stone price index (P^*)[3]. No time subscript is included in equation (5), as each household only participates to the survey at one point in time, so that prices vary across households if their observations belong to different regions and/or time periods. For comparison purposes, equation (5) was also estimated employing unit values in place of estimated prices to obtain uncompensated direct price elasticities (see Green and Alston, 1990).

[2] Nine Government Office Regions (GORs, defined by the UK Office for National Statistics) plus Wales, Scotland and Northern Ireland.

[3] The estimated demand system is conditional on food expenditure. Since NFS data do not include total household expenditures, unconditional demand system cannot be estimated on data before 2001.

6. Results

Estimates from equation (3) for the healthy and the unhealthy baskets are reported in Table 1. Average variability in unit values within the same cluster seems to be well captured by household composition both for fruit and vegetables and for junk food. Better-off households pay on average higher prices of both aggregates, and the quantity effect is higher for F&V than for junk food, as it seems reasonable.

Table 1 Results of within cluster estimation of unit values quality and quantity components

	Fruit and vegetables		**Junk food basket**	
	Coeff.	Prob.	Coeff.	Prob.
N. of female adults	0.088	0.001	0.037	0.001
Family of 2 or more adults only (dummy)	0.113	0.001	0.046	0.001
Family with children (dummy)	0.063	0.001	-0.016	0.340
Age of household head	-0.001	0.001	-0.002	0.001
Household income	0.006	0.001	0.005	0.001
Purchased quantity	-0.247	0.001	-0.184	0.001

The effectiveness of the correction procedure in disentangling the quality component embedded in unit values can be observed in Table 2 where we report the correlations between the original unit values and the corresponding quality-corrected prices for a set of food aggregates characterized by different level of aggregation. Apples, fruit, fruit and vegetables and the total food aggregate are considered. According to the different heterogeneity of their compositions, the quality component embedded in their unit values should weight differently. As a result, the quality correction of unit values is expected to be less relevant in relatively homogenous aggregates, while discrepancies between unit values and corrected prices should be more noticeable for complex aggregates.

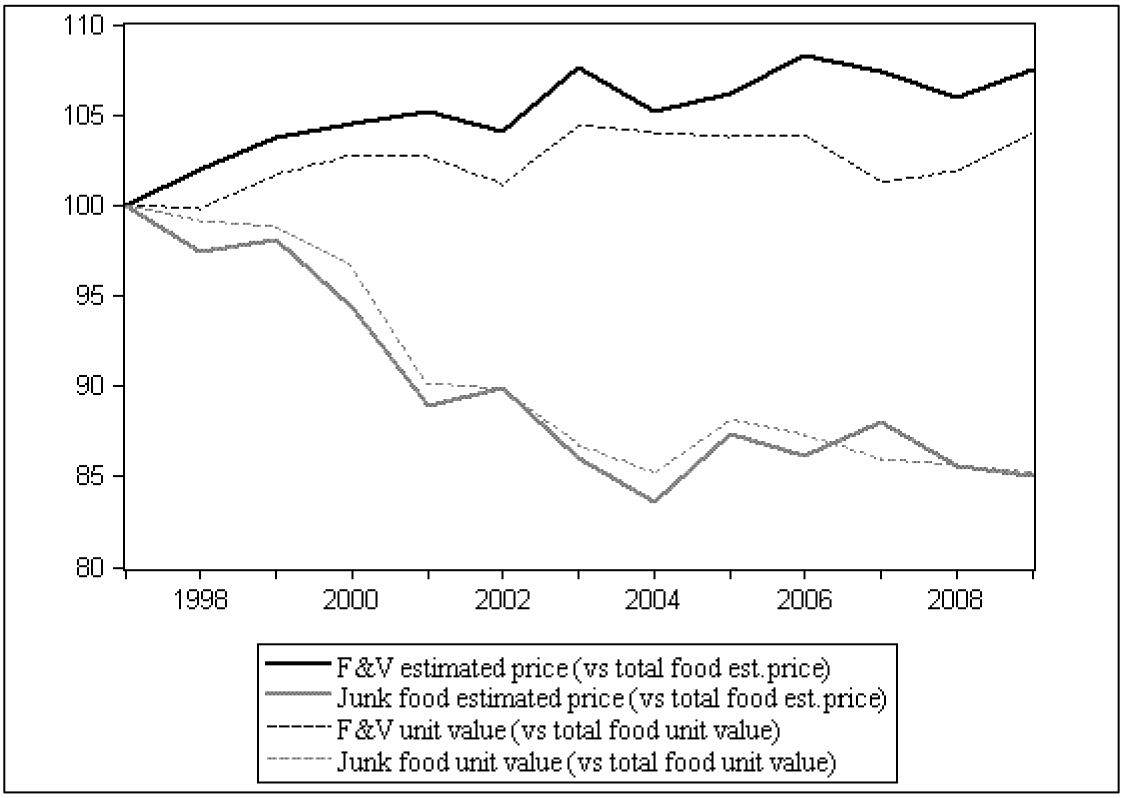
Table 2. Correlation between unit values and corrected prices of different food aggregates.

	Apples	Fruit	Fruit and Vegetables	Food
Apples	0.915			
Fruit		0.871		
Fruit and Vegetables			0.834	
Food				0.711

The correlations in Table 2 confirm that the adjustment procedure gains relevance at higher aggregation levels. Trends in observed unit values and estimated prices for the junk and the F&V bundles (relative to the price of the total food aggregate) are shown in Figure 1.

While increasing by about 7% over 13 years, the average (relative) cost of fruit and vegetables has been generally steadier compared to the unhealthy basket, whether considering estimated prices or unit values. According to our estimates, the relative price of F&V has grown more than what is detected by unit values, as consumers have reacted by adjusting their basket towards cheaper products.

Figure 1. Fruit and vegetables and junk food, unit values and corrected prices relative to total food

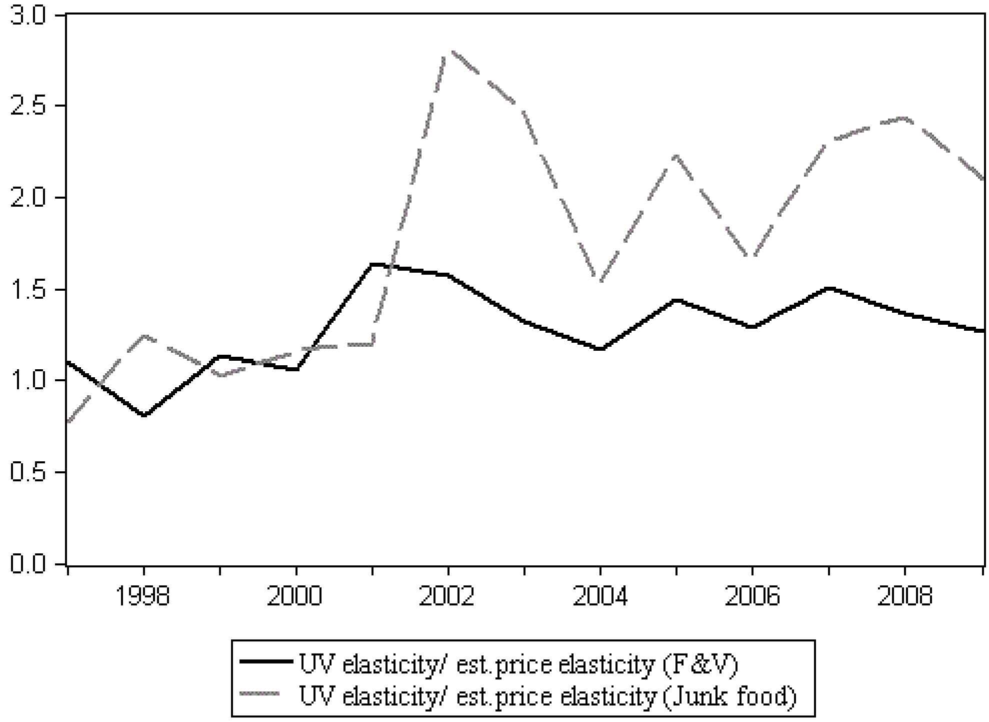


On the opposite, junk food price has become substantially cheaper (-15%) relative to the general food aggregated and the quality compensation represented by the distance between the solid and the dotted lines has been much smaller and less evident over time. Thus the diverging trends of the two bundles prices are underrated by unit values, as the latter incorporate compensating quality choices adopted by consumers.

As mentioned in Section 2, demand elasticities based on unit values are likely to overestimate the real demand response to price changes which would have been observed without compensating strategies by consumer, i.e. reallocation within the same food aggregate. This is our case for both aggregates, as shown in Figure 2, where the ratio of the elasticities estimated respectively from unit values and estimated prices is reported. Values above 1 indicate that elasticities based on unit values are higher than those based on corrected prices, which is the case for all years after 1999 for both baskets. The size of the correction for the ‘junk food’ aggregate is striking.

As illustrated in previous sections, the correction procedure produces an estimate of the average cluster price plus a non-identifiable good specific constant term, which, according to our initial assumption about price variability represents an estimate of the average price everyone faces in that geographic area in a given point in time.

Figure 2. Unit values and price elasticity ratios, for junk food and FV.



A set of regional price time series for the two considered aggregates becomes then available. Figure 3 shows the quality-adjusted price of junk food and F&V relative to the total food price aggregate in London and Scotland between 1997 to 2009. Geographical variability is evident for both baskets. F&V are on average more expensive with respect to food in general, both in London and in Scotland, though distances are more pronounced in London. Instead, junk food price trends show less evident differences between the two areas. As one would expect, F&V prices are more likely to be region-specific, because of distance from production areas and transportation costs. On the other hand, the junk food aggregate is exclusively composed by processed food, and its distribution costs are likely to be more homogeneous across the country.

Figure 3. Estimated prices of junk food and fruit and vegetables in London and Scotland

7. Conclusion

We provide an indirect estimate of a time series of food price indices based on household expenditure data, with a focus on a comparison between foods classified as ‘junk’ and a healthy benchmark represented by fruit and vegetables. The analysis has shown the relevance of a quality correction when unit values are used in place of prices in food demand analysis to investigate demand response to price changes. When food aggregates are considered, without such quality adjustment one would ignore any compensating (reallocation) behavior adopted by consumers in response to price changes and elasticity estimates would be upward biased. Such a bias would be especially serious when the objective of demand studies is the simulation of fiscal interventions to encourage or discourage purchases of specific goods. A trivial example is the possibility that consumers might react to a general increase in junk food prices due to a “fat tax” by decreasing the quality level of their junk food baskets, i.e. moving towards cheaper options within the same aggregate, with a paradoxical deterioration of their diet quality, also generating lower tax revenues compared to those estimated by simulations based on unit values. This issue is especially serious if little substitution occurs between healthy and unhealthy aggregates, as several studies suggest.

The work has two main empirical outcomes. First, it enables the investigation of prices for aggregate food groups whose composition reflects specific nutritional content or healthiness properties. Second, it provides an estimate of geographical variability of prices, which is not provided by the National Statistical offices through regional price indexes.

Reference List

Capacci,S. et al. 2012. "Policies to promote healthy eating in Europe: a structured review of policies and their effectiveness." *Nutrition Reviews* 70(3):188-200.

Crawford,I., F.Laisney, and I.Preston June 2003. "Estimation of household demand systems with theoretically compatible Engel curves and unit value specifications." *Journal of Econometrics* 114(2):221-41.

Cutler,D.M., E.L.Glaeser, and J.M.Shapiro 2003. "Why have Americans become more obese?" *Journal of Economic Perspectives* 17(3):93-118.

Deaton,A. 1987. "Estimation of own-and cross-price elasticities from household survey data." *Journal of Econometrics* 36(1-2):7-30.

----- 1988. "Quality, quantity, and spatial variation of price." *The American Economic Review* 78(3):418-30.

----- 1990. "Price elasticities from survey data: Extensions and Indonesian results." *Journal of Econometrics* 44(3):281-309.

----- 1997. *The analysis of household surveys: a microeconometric approach to development policy.* Baltimore: Johns Hopkins University Press.

Deaton,A. and J.Muellbauer 1980. "An almost ideal demand system." *The American Economic Review* 70(3):312-26.

Green,R. and J.M.Alston 1990. "Elasticities in AIDS models." *American Journal of Agricultural Economics* 72(2):442-5.

Mazzocchi,M., B.Traill, and J.F.Shogren 2009. *Fat economics: nutrition, health, and economic policy.*Oxford University Press, USA.