Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMIA E STATISTICA AGROALIMENTARE

Ciclo XXII

Settore scientifico-disciplinare di afferenza: SECS-P/02 Politica Economica

TITOLO TESI

Quantitative evaluation of household nutrition patterns: an econometric assessment of the UK 5-a-day impact on fruit and vegetable consumption

Presentata da: **Sara Capacci**

Coordinatore Dottorato

Prof. Roberto Fanfani

Relatore

Dott. Mario Mazzocchi

Esame finale anno 2010

ACKNOWLEDGEMENT

The present thesis would not have been possible without the outstanding supervision I have been receiving from Dott. Mario Mazzocchi. The work also benefitted a lot from the time I spent at the Department of Agricultural and Food Economics at the University of Reading, UK. Thanks are also due to the colleagues and friends of the coloured and funny working environment of *Via Zamboni 18*.

CONTENTS

LIST OF TABLES

1. Introduction

Healthy food habits have been shown to be the first prevention tool against most of non-communicable chronic diseases, including heart diseases, stroke and cancer (WHO, 2003).
Adequate consumption of fruit and vegetables (F&V) represents a crucial element of healthy eating. Unfortunately, actual intakes are still largely below the recommended level of 5 portions of F&V per day (about 400 grams per person per day) almost everywhere in western countries (Naska et al., 2000).
During the last ten years, the World Health Organization has explicitly asked countries to set out effective health communication programmes to improve people dietary choices, in order to reduce the risk of deaths from chronic disease.
The question of unhealthy diets has become a new policy priority for the government of most western countries, and the debate about which intervention might affect effectively people food habits is much-discussed.
The “5-a-day” campaign to increase fruit and vegetable intakes towards the WHO recommendation of 5 portions (or 400 grams) per day was first introduced in 1991 in the US and subsequently it has been adopted by several other countries (Stables et al., 2002). Today it represents one of the most widespread public interventions in the field of healthy eating. In the UK it has started as a nationwide communication campaign in 2003.
The demand for unambiguous evaluations of public interventions effectiveness and the need for detecting the proper policy instruments in the nutrition field has increasingly involved economists, as it has already happened for other public policy debates (e.g. environment) (Mazzocchi et al., 2009).
Under the economic perspective consumers make their decisions on purchases (including food purchases) based on their preferences, market price levels for different goods and disposable income. Economically optimal decision requires that consumer hold full information when making their choices, including knowledge on the health implications of alternative consumption bundles.
Since the full information assumption may not be reflected in the market, one of the main objectives of public intervention is to fill information gaps. In economic terms, the absence of perfect information leads to market failure. People may develop unhealthy food habits unwillingly, because they do not have adequate information about the health consequences of their food choices. Either they do not have adequate information about the nutrient contents of

different food items, or about the proper portions to be consumed in a day. This lack of information might create a discrepancy between rational and welfare optimising choices for consumers. For instance, people can gain utility from intangibles like health, but might fail to maximize their utility because asymmetric or incomplete information on nutrient food content prevents them to achieve the desired nutrient intake levels. According to mainstream economics, governments must intervene to address market failures and restore the adequate environment for free market choices.

Yet, this is not the sole possible scenario. Perfectly informed individuals may have different preference structures and attach different importance to pleasure and health. Thus, well-informed people might consciously choose to have an unhealthy diet, preferring short-term gratification with regard to possible long-term health risks (Mazzocchi et al., 2009).[1]

Moreover, even in a perfectly informed market, people wishing to maximize their demand of health through food choices can be hindered by their budget constraint, or by price levels of the healthier food.

Public intervention aimed at taking part in this mechanism should be planned considering all the acting forces. In fact even if information is a precondition of good diet, policies addressing information problems are not necessarily the most effective ones for improving diets. Adopting a pure public health perspective leads evaluators and policy makers to ignore possible interaction with those market forces which normally are considered in the economic field.

This thesis adopts an empirical economics perspective, with the aim of providing an ex-post quantitative evaluation of the UK 5-a-day programme impact on fruit and vegetable consumption of British households. Microeconomic models as demand systems are employed here to stylize consumption behaviour which results from alternative acting forces, and the estimate of the counterfactual scenario (without the policy) is based on econometric methods which are expected to disentangle the policy impact from potentially conflicting market forces dynamics.

We estimate a demand model for fruit and vegetables based on the Quadratic Almost Ideal Demand System (QUAIDS), allowing for demographic effects, and controlling for potential endogeneity of prices and total food expenditure. The model, estimated for the baseline period (prior to the intervention), is then projected to estimate the counterfactual demand for fruit and vegetables over the years following the information campaign for which data are available. The coefficients of the demand model which are estimated on pre-campaign data represent the reference behavioural parameters for assessing consumer response to the campaign over the following years. The difference between the post campaign consumption

[1] Researchers have shown that when risks are well known public information campaigns are ineffective in changing behaviours (Rindfleisch et al., 1999). In other words when perfect information is granted market forces act to determine consumers' choices.

(forecasted through post-campaign estimated model) and the model-projected consumption is an estimate of demand response to the campaign, after controlling for price and expenditure variations over the years.
The work is organized as follows: the next section offers a brief description of the 5-a-day programmes implemented and evaluated by western countries in the last 15 years. Chapter 2 presents the theoretical tools employed for counterfactual estimation and ex-post evaluation with a particular focus on the demand system specification. In Chapter 3 the application to the UK case is described. Finally results are summarized and conclusions drawn.

1.1 Effectiveness of the 5-a-day interventions in western countries.

Five-a-day campaigns are currently in action in the US ("5-a-day for better health" promoted by the National Cancer Institute), in Western Australia ("Go for 2&5 campaign"), in Spain ("5 al dia" programme), in Portugal ("Programa 5 ao dia"), in Denmark ("6-a-day"), in Poland, in Sweden (driven by the Swedish supermarket chain) and in the UK[2].
Because of the employed communication tools and strategies, 5-a-days programmes can be defined as social marketing interventions. According to Andreasen's definition

> Social marketing is the application of commercial marketing technologies to the analysis, planning, execution and evaluation of programs designed to influence the voluntary behaviour of target audiences in order to improve their personal welfare and that of society (Andreasen, 1995).

The literature above effectiveness of social marketing interventions in nutrition field is quite rich. The balance of evidence is that normally interventions are effective in raising awareness, increasing knowledge and self-efficacy, and changing attitudes, but they are less effective in changing behaviours (Mazzocchi et al., 2009).
A detailed literature review of ex-post impact assessment studies for policies promoting fruit and vegetable consumption in Europe and US has shown that the average effect on consumption is roughly between +0.2 and +0.6 portion per day (Pomerleau et al., 2005).
Western Australian Go for 2&5 evaluation has been based on a pre-post survey and has highlighted an increasing of 0.8 servings/day for adults during the program period, with a following decrease of 0.3 portions after the end of the campaign (Pollard et al., 2008).

[2] This information has been collected within the project "Interventions to Promote Healthy Eating Habits: Evaluation and Recommendations – EATWELL" funded by the European Commission (7th Framework Programme, EC Grant Agreement 226713).

In the US, the impact assessment of the worksite-based interventions included in the 5 a day for Better Health campaign has shown a pre-post increasing of 19% of consumption levels, reflecting a difference of one half serving (Sorensen et al., 1992)

The evaluation of the whole US 5 a day program (which is made up of multiple initiatives, designed at national level and locally implemented) through two nationally representative pre and post surveys (in 1991 and in 1997) has shown a statistically significant improvement in consumption level from baseline to follow-up survey (from 3.75 to 3.98 per day for the total population). However the adjusted analysis revealed that the positive change was probably attributable to demographic changes between the two survey years, correlated with vegetable and fruit consumption. Nonetheless the same study points out that the program awareness (measured as the percentage of people aware of the 5 a day message) has significantly changed among the total population and all the demographic subgroups (Stables et al., 2002).

The next chapter will focus on the evaluation strategy employed for impact assessment of the UK 5-a-day.

2. Methodology

2.1 Evaluation

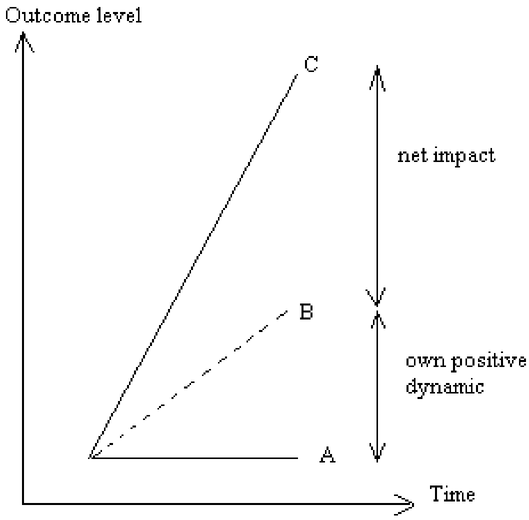
The objective of the present work is an ex-post evaluation of the effect of the 5-a-day intervention on UK consumers. Some preliminary definitions are needed.

With reference to the policy evaluation literature, the 5-a-day campaign here considered is what is called the *treatment*, i.e. a policy, or intervention explicitly directed to a group of units (individuals, firms, etc.) and explicitly pursuing a change in some dimension of that group. The group of units to which the treatment is directed is the *target group*.

Some variables can be found to effectively describe those dimensions of the target group the treatment should affect (e.g. psychological traits, behaviours, etc.); they can be defined as *outcome variables*, or outcomes. The 5-a-day campaign is meant to affect at least two dimensions of its target group (UK consumers): attitudes (i.e. mostly knowledge and awareness of the importance of eating more fruit and vegetable) and behaviour (i.e. actual consumption) and ultimately they are expected to have a positive effect on health and life expectancy. The aim of the present work is to assess the effect of the campaign on behaviour. Thus, fruit and vegetable consumption is the evaluated outcome.

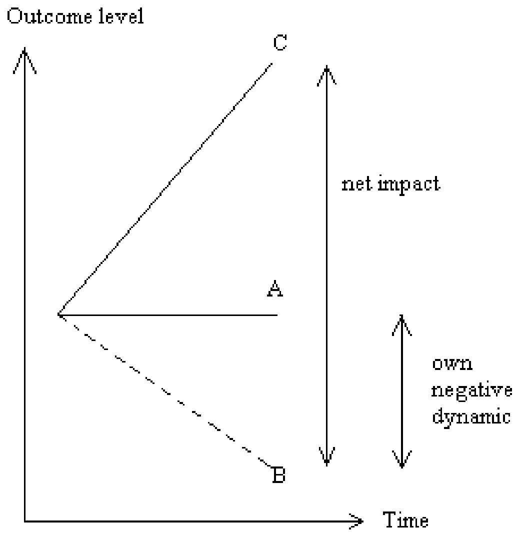
The effect or the impact of the intervention is often erroneously confounded with the simple difference between the outcome level observed *before* the treatment on the target units and the outcome level observed *after* the treatment on the same units. This difference ignores the possible outcome's own dynamic due to many factors other than the intervention. Consumption patterns normally evolve over time, even in absence of specific public policies. As it will be discussed later, consumption can be strongly affected by economic forces other than by changes in information brought by a social marketing intervention. Sometimes even in opposite directions (see Figure 2.1 and Figure 2.2 for an effective representation). If the outcome variable has its own positive trend, the change due to the intervention (BC in Figure 2.1) should be disentangled from the change due to this positive inherent dynamic (AB). Otherwise the treatment effect would be overestimated (AB+CB). Similarly, if the outcome variable had a negative own trend the net impact should be computed as the difference between the new outcome level (C in Figure 2.2) and the lower level which would be reached because of the negative dynamic (B). The simple *before and after* difference would underestimate the treatment effect (CA in Figure 2.2).

Figure 2.1 Impact measurement on an outcome variable with positive trend.



Source: (Martini, 1997)

Figure 2.2 Impact measurement on an outcome variable with negative trend.



Source: (Martini, 1997)

The impact of the intervention should be intended as the difference between the outcome level observed after the treatment (C in both figures) and the outcome level that would have been observed at the same time without any treatment (B in both figures). The second term of the difference is of course hypothetical, thus non-observable; it is called *counterfactual*.

Following Caliendo and Hujer (2006), C_t^1 is the outcome level observed at time t on those subjects exposed to the intervention, C_t^0 is the outcome level observed at time t in absence of the intervention and D is a binary variable which reflects participation to the intervention (D is equal to 1 if the individual participates to the intervention and equal to 0 otherwise). The individual treatment effect is defined as:

$$\Delta_i = C_i^1 - C_i^0 \tag{2.1}$$

It is not directly computable, since for each individual:

$$C_i = D_i C_i^1 + (1 - D_i) C_i^0 \tag{2.2}$$

The second term of (2.1) has to be estimated.

Normally averaged population outcomes rather than individual treatment effects are investigated. When one considers the total population two scenarios occur. The first is the case in which a treated group and a non-treated group exist. The Average Treatment Effect (ATE) can thus be computed:

$$\Delta_{ATE_t} = E(C_t^1) - E(C_t^0) \tag{2.3}$$

ATE is the average effect on the overall population, and requires that individuals are assigned randomly to the treated group or to the non-treated group, so that there is no selection bias. From this condition follows that the non-exposed group is not systematically different from the treated group and can be considered as a *control group*. This is the case of experimental data.

In the second case, when all individuals in the sample are exposed to the intervention – which is the most reasonable case for our study – the assumption above falls and C_t^0 cannot be observed. Thus, an alternative definition is adopted, which focuses on the average treatment effect on the treated subjects (ATT, Average Treatment on Treated), i.e. a measure based directly and exclusively on those exposed to the intervention:

$$\Delta_{ATT_t} = E(C_t^1 \mid D = 1) - E(C_t^0 \mid D = 1) \tag{2.4}$$

The second term of (2.4) – which is the *counterfactual* – reflects the outcome that would have been observed on those subject to the intervention in absence of the intervention, it cannot be computed and requires some estimation strategy.

Different strategies to estimate the unknown counterfactual component of the above equation represent alternative evaluation methods.

In the following section a possible strategy to estimate the counterfactual individual and average consumption levels is shown.

2.2 Estimation of counterfactual demand through a demand model

An estimate of the counterfactual consumption for each individual can be obtained by resorting to economic consumer theory, presuming that some sort of demand function exists and is able to explain the demanded (if not the consumed) quantity of a good.

We assume that individuals allocate their consumption bundle based on exogenous prices, income and preferences. This follows from the classical framework where consumers maximize a direct utility function $U(\mathbf{q})$ which incorporates their preferences and depends on the consumption of a vector of n quantities of goods $\mathbf{q} = (q_1, q_2, .., q_n)$, given a level of total expenditure $x(\mathbf{p},\mathbf{q})$[3] and a vector of n prices $\mathbf{p} = (p_1, p_2, ..., p_n)$ (Deaton et al., 1980b). The solution of the following constraint optimization problem:

$$\max_{\mathbf{q}} U(\mathbf{q}) \qquad s.t. \ (\mathbf{p},\mathbf{q}) = x \tag{2.5}$$

is the well known set of Marshallian (uncompensated) demand functions:

$$\mathbf{q} = g(x,\mathbf{p}) \tag{2.6}$$

where demanded quantities are a function of prices and income. The maximised utility is called the indirect utility function $U(x,\mathbf{p})$ and is the maximum level of utility obtainable by a consumer, given his budget constraint.

This primal approach (demanded quantities as solution of a constrained maximisation problem) has an alternative, but equivalent, dual representation.

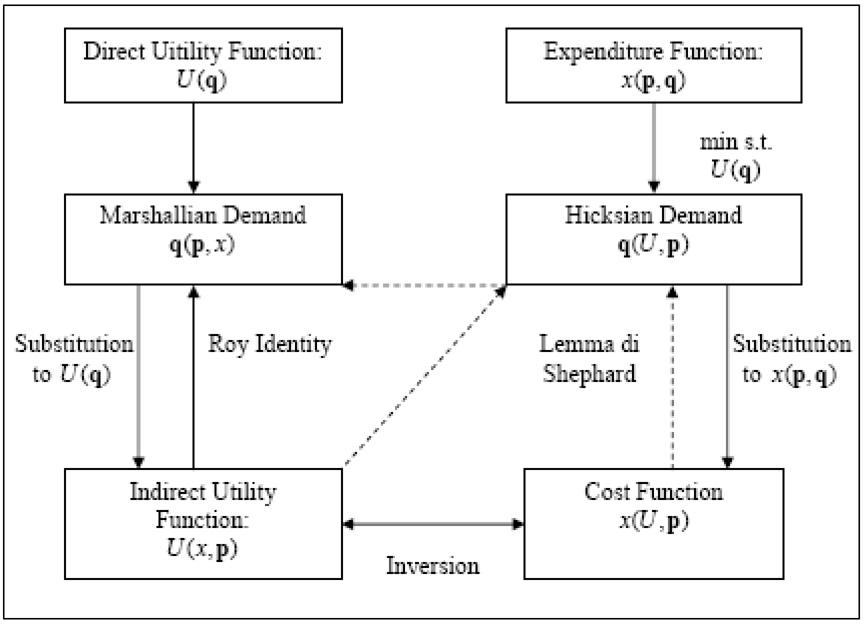
The same optimal demanded quantity is also given by the minimization of the expenditure function $x(\mathbf{p},\mathbf{q})$ subject to a given level of the direct utility function $U(\mathbf{q})$. The solution for this formulation of the optimization problem generates a system of demand functions known as Hicksian (compensated) demand functions, where quantities are dependent on utility and prices:

$$\mathbf{q} = h(U,\mathbf{p}) \tag{2.7}$$

Substituting these quantities back into the original problem (the expenditure function) gives the cost function $c(U,\mathbf{p})$, i.e. the minimum cost of having a utility level U, given the vector of prices $\mathbf{p}$. Price derivatives of the cost function are the Hicksian demand functions (this is known as Shephard's Lemma). Primal and dual scenarios of the consumers' optimization problem are of course strictly connected as it is clearly shown in Figure 2.3 below.

[3] In the classical approach non-satiation characterizes utility functions, i.e. given two different consumption bundles q^A and q^B, where $q^A \geq q^B$, q^A is strictly preferred to q^B. In other words larger bundles are preferred to smaller bundles. Moreover saving is not considered. It follows that the total expenditure x can be considered in place of total income and the optimal choice will be *on the boundary* of the budget constraint.

Figure 2.3 The dual problem of consumer's optimization



Source: (Deaton et al., 1980b), our elaboration. See Figure 2.8 and Figure 2.10 in (Deaton et al., 1980b)

Inversion of the cost function gives the indirect utility function. By substituting the indirect utility function in the Hicksian demand functions, the Marshallian demand functions are obtained. Symmetrically, Hicksian demand functions can be obtained by substituting the cost function into the Marshallian demand function.

Both Hicksian and Marshallian demands show some properties which derive from the structure of consumer preferences and the characteristics of the optimization problems[4]. Those properties have strict consequences on the econometric specification of demands:

- the demanded quantities times their respective prices sum up to the total expenditure (*adding-up*)
- Hicksian demands are homogeneous of degree zero in prices and Marshallian demands are homogenous of degree zero in price and total expenditure (*homogeneity*)
- the cross-price derivatives of the Hicksian demands are symmetric (*symmetry*)
- the Slutsky matrix (or substitution matrix, i.e. the matrix of the second order price derivatives of the cost function) is negative semidefinite. It follows that compensated price responses (i.e. in Hicksian demands) are non-negative (*non-negativity*).

When estimating demand functions, analysis of demands' responsiveness to price and income changes is of main interest. Elasticities measure the percentage change in demand of good *i* per (marginal) percentage change in the price of good *i* (direct price elasticity) or good *j*

[4] Demand properties directly derive from the consumer's preferences structure and the axioms characterizing it (reflexivity, completeness, transitivity, continuity, non-satiation and convexity) and from the budget constraint, which in this framework is assumed to be linear (Moro, 2004)

(cross-price elasticity) or income (Mas-Colell et al., 1995). Table 2.1 displays compensated and uncompensated elasticities.

Table 2.1 Compensated and uncompensated elasticities.

	Direct price elasticity	Cross-price elasticity	Income elasticity
Marshallian (uncompensated) Demand	$e_{ii} = \frac{\partial q_i(x,\mathbf{p})}{\partial p_i}\frac{p_i}{q_i}$	$e_{ij} = \frac{\partial q_i(x,\mathbf{p})}{\partial p_j}\frac{p_j}{q_i}$	$e_i = \frac{\partial q_i(x,\mathbf{p})}{\partial x}\frac{x}{q_i}$
Hicksian (compensated) Demand	$\mu_{ii} = \frac{\partial q_i(u,\mathbf{p})}{\partial p_i}\frac{p_i}{q_i}$	$\mu_{ij} = \frac{\partial q_i(u,\mathbf{p})}{\partial p_j}\frac{p_j}{q_i}$	

The relation between compensated and uncompensated demand is represented by the Slutsky equation, which comes from the following identity where q is the optimal demanded quantity expressed both in terms of Hicksian and Marshallian function:

$$q_i = h_i(U,\mathbf{p}) = g_i(x,\mathbf{p}) \tag{2.8}$$

Derivation of (2.8) with respect to p_j gives[5]:

$$s_{ij} = \frac{\partial h_i}{\partial p_j} = \frac{\partial g_i}{\partial x}q_j + \frac{\partial g_i}{\partial p_j} \tag{2.9}$$

The effect (s_{ij}) on demanded quantity q_i of a price change of good j can be split into the uncompensated change in demand ($\partial g_i / \partial p_j$) and the compensation $(\partial g_i / \partial x)q_j$, which measures the income effect of price changes. The Slutsky equation can also be expressed in terms of elasticities:

$$e_{ij} = \mu_{ij} - e_i w_i \tag{2.10}$$

where e_{ij} and μ_{ij} are respectively uncompensated and compensated elasticities.

Usually Marshallian demands are used for estimation purposes (in alternative specifications), since Hicksian demands include the utility term which is not empirically observable.

In the following evaluation procedure we adopt the Almost Ideal Demand System (AIDS) specification for a set of Marshallian demand functions (Deaton et al., 1980a) and some subsequent evolution of it.

2.3 Information and demand function

When resorting to demand functions for the estimation of the effect of an intervention like an information campaign we assume that individuals allocate their consumption bundle based on

[5] Note that in $g_i(x,\mathbf{p})$, x can be expressed in terms of u and p as the minimized cost function of the dual problem. For the *chain rule* (see Chiang (1974)) $\frac{\partial g_i(x,\mathbf{p})}{\partial p_j} = \frac{\partial g_i(c(u,\mathbf{p}),\mathbf{p})}{\partial p_j} = \frac{\partial g_i}{\partial x}\frac{\partial c(u,\mathbf{p})}{\partial p_j} + \frac{\partial g_i(\mathbf{p})}{\partial p_j} = \frac{\partial g_i}{\partial x}q_j + \frac{\partial g_i(\mathbf{p})}{\partial p_j}$.

exogenous prices, income and preferences and most importantly that information alters the structure of preferences.

Consumer preferences vary depending on quantities and on characteristics of goods. These characteristics embody physical attributes (e.g. the nutritional content of a food), but also perceived attributes (e.g. the subjective nutritional value and health content associated with a food) and the latter can be altered by information (Nayga et al., 1999). This theoretical framework translates into a set of Marshallian demand functions where information enters the utility function through a vector **r** of goods' attributes:

$$U(x_1, x_2, ...x_g, \theta(\mathbf{r})) \tag{2.11}$$

It follows that an information campaign should result in a modification of consumer preferences, i.e. a shift in the demand functions, while the income and price coefficients (i.e. the behavioural parameters) remain stable, at least in the short-run. Under this assumption, a change in information acts as a demand shifter (Piggott et al., 2004).

It might be argued that in presence of rigid supply, information might also affect prices through increased aggregate demand. However, for individual households prices can be safely assumed to be exogenous. Furthermore in the spirit of the Lucas' critique, this approach might be generalised to allow for other behavioural parameters to change in response to information release (i.e. time varying price and expenditure coefficients). However, this would require data with a longer time span enabling to capture smooth structural changes with some further difficulties in disentangling the individual effects. The stylisation suggested here channels the impact of information on the level of the demand curve rather than its inclination and provides a reasonable short-run approximation.

2.4 The Almost Ideal Demand System (AIDS) specification and its linear version (LA/AIDS).

Deaton and Muellbauer derive their well-known specification of a system of Marshallian demand functions from a logarithmic cost function known as PIGLOG[6] (Deaton et al., 1980a):

$$\log c(U, \mathbf{p}) = \log a(\mathbf{p}) + b(\mathbf{p})U \tag{2.12}$$

where $\log a(\mathbf{p})$, price function homogeneous of degree one, has the translog form:

$$\log a(\mathbf{p}) = \alpha_o + \sum_{i=1}^{n} \alpha_i \log p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \log p_i \log p_j \quad i, j = 1, ..., n \tag{2.13}$$

and $b(\mathbf{p})$ term, homogeneous of degree zero, has the following Cobb-Douglas form:

[6] Demand function where budget shares are linear in log total expenditure have been called Price-Independent Generalized Logarithmic (PIGLOG) by (Muellbauer, 1976). They embody indirect utility functions themselves linear in log total expenditure.

$$b(p)=\prod_{i=1}^{n} p_i^{\beta_i} \tag{2.14}$$

where *n* is the number of goods in the consumption bundle and p_i and p_j are respectively the price of the *i*-th and *j*-th good.

Strictly following Deaton and Muellbauer (1980a) price derivation of (2.12) gives a set of Hicksian demands (see in Figure 2.3, the dotted arrows indicate the steps followed by Deaton and Muellbauer in their derivation of Marshallian demand functions). By inversion of the above defined cost function, the indirect utility function is derived and then substituted into the Hicksian function to generate the following Marshallian demand system expressed in budget shares (w_{ih}):

$$w_i^h=\alpha_i+\sum_{j=1}^{n}\gamma_{ij}\log p_j^h+\beta_i\log(x^h/P) \quad i:1,...,n \tag{2.15}$$

where w_i^h is the share of total expenditure allocated by the *h*-th consumer's to good *i*. This quantity varies as a function of prices faced by the consumer (p_j^h) and his total expenditure (x^h), deflated by P which is the non-linear price index $a(\mathbf{p})$ defined in (2.13).

Consistency with economic theory (see paragraph 2.2) requires some testable restrictions to hold. In particular the *adding-up* property of demand functions requires that:

$$\sum_{i=1}^{n}\alpha_i=1 \qquad \sum_{i=1}^{n}\gamma_{ij}=0 \qquad \sum_{i=1}^{n}\beta_i=0 \tag{2.16}$$

The *symmetry* property requires that:

$$\gamma_{ij}=\gamma_{ji} \tag{2.17}$$

and *homogeneity* requires that:

$$\sum_{j=1}^{n}\gamma_{ij}=0 \tag{2.18}$$

A fourth condition on the negative semi-definitiveness of the Slutsky matrix is commonly replaced by the broader requirement that own-price elasticities are negative.

In their original work Deaton and Muellbauer (1980a) suggest a linear specification of (2.15) by substituting the non linear P with the linear Stone's price index P^*:

$$\log P^*=\sum_{k=1}^{n}w_k\log p_k \tag{2.19}$$

They found that the above specification could be a good approximation of the true non linear price index when prices are closely collinear.

Note that in the non-linear specification of the AIDS model the parameter α_0 appears, although this cannot be identified at the estimation stage. Economic interpretation of this parameter is the minimum outlay required for a minimal standard of living (Deaton et al., 1980a), having scaled prices to one. Assigning a reasonable value to α_0 can overcome the identification problem and is the easiest and most frequent choice in estimation of non linear AIDS model (see Chapter 3).

2.4.1 AIDS and LA/AIDS elasticities

By applying the elasticity formulas displayed in Table 2.1 to (2.15), the equations for the uncompensated price and income elasticities for AIDS model are given (see Table 2.2).[7] When adopting the linear specification of the AIDS model, some differences in price elasticities arise due to computational problems in differentiating the linear price index (P^*) with respect to the *i*-th price (Green et al., 1990). In the analysis of the LA/AIDS model we assume

$$\frac{\partial \log(P^*)}{\partial p_j} = w_j \tag{2.20}$$

following Chalfant (1987). The resulting elasticity formulas for AIDS and LA/AIDS are shown in Table 2.2.

Table 2.2 Elasticity formulas in LA/AIDS and AIDS models

	Uncompensated Price elasticities	Uncompensated Income elasticity
AIDS	$e_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \beta_i \frac{\alpha_j}{w_i} - \frac{\beta_i}{w_i} \sum_k^n \gamma_{kj} \log p_k$	$e_i = \frac{\beta_i}{w_i} + 1$
LA/AIDS	$e_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \beta_i \frac{w_j}{w_i}$	$e_i = \frac{\beta_i}{w_i} + 1$

Source: Green and Alston (1990)

Note that in price elasticity formulas δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$ for $i = j$; $\delta_{ij} = 0$ for $i \neq j$).

2.5 Differences in demand responsiveness to income changes: the Quadratic Almost Ideal Demand System (QUAIDS)

The AIDS model has been largely used in consumer literature, and some extensions of it have been developed in the past and have become as popular as the original specification. The quadratic extension of the AIDS is one of the most accepted developments and allows the original form to adjust to different income responsiveness of demand. Several specifications

[7] When computing elasticities on demands expressed in terms of budget shares with logarithmic prices and expenditure the following relations hold: $e_{ij} = \frac{\partial w_i}{\partial \log p_j} \frac{1}{w_i} - \delta_{ij}$ $\quad e_i = \frac{\partial w_i}{\partial \log x} \frac{1}{w_i} + 1$

for the quadratic extension to the AIDS model have been developed, in this section the QUAIDS model by Banks, Blundell and Lewbel (1997) is considered.
PIGLOG preferences (see paragraph 2.4) always give rise to Engel curves[8] of the following form:

$$w_i = \alpha_i + \beta_i \log x \tag{2.21}$$

where budget shares are linear to the logarithmic of outlay. This specification is known as Working-Leser specification (Leser, 1963;Working, 1943). The original AIDS model embodies this kind of linear Engle curves and integrates them with consumer theory.
Yet, income varies considerably among individuals, and demand responsiveness to income is likely to vary for people in different point of income distribution. Empirical analysis of Engel curves shows that for some commodities the linear relation among income and expenditure shares fails to capture real individual behaviours (Banks et al., 1997). For some goods some further terms in income are required for expenditure shares equation in order to capture the real nature of the goods. For a consumer at a certain point of income distribution a good can be a luxury ($\beta_i > 0$) whereas for people at other points of income distribution the same good can be necessary or inferior ($\beta_i < 0$). The AIDS model (and all the demand specification belonging to the PIGLOG class) embeds linear Engel curves, and is not flexible enough to allow for differences in income responsiveness of demands.
A new class of demand system starting from the AIDS has been introduced by Banks, Blundell and Lewbel (1997). The new model is called Quadratic Almost Ideal Demand System (QUAIDS) and includes an additional higher order income term. This new specification preserves consistency with consumer theory while allowing a more flexible specification of Engel curves.
In Banks, Blundell and Lewbel (1997) the indirect utility function of a PIGLOG demand system:

$$\log u = \left[\frac{\log x - \log a(\mathbf{p})}{\log b(\mathbf{p})} \right] \tag{2.22}$$

has been generalized by adding an extra term $\lambda(\mathbf{p})$ (differentiable, homogeneous function of degree zero of prices):

$$\log u = \left\{ \left[\frac{\log x - \log a(\mathbf{p})}{\log b(\mathbf{p})} \right]^{-1} + \lambda(\mathbf{p}) \right\}^{-1} \tag{2.23}$$

By setting:

$$\lambda(\mathbf{p}) = \sum_i \lambda_i p_i \qquad \text{where} \qquad \sum_i \lambda_i = 0 \tag{2.24}$$

and applying Roy's Identity the QUAIDS model is given:

[8] Engel curves express the relation between commodity expenditure shares and income.

$$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \log\left[\frac{x}{a(\mathbf{p})}\right] + \frac{\lambda_i}{b(\mathbf{p})}\left\{\log\left[\frac{x}{a(\mathbf{p})}\right]\right\}^2 \qquad (2.25)$$

where $a(\mathbf{p})$ is the translog price function in (2.13) and $b(\mathbf{p})$ is the Cobb-Douglas price aggregator in (2.14) used in the original AIDS.

Since it is derived as a generalization of PIGLOG preferences the QUAIDS model preserves all the characteristics of the linear AIDS (as proved by Banks, Blundell and Lewbel (1997)). Note that in this framework the original AIDS by Deaton and Muellbauer is a special case of (2.25) where $\lambda(\mathbf{p})$ is set to zero. The uncompensated price elasticities for the QUAIDS model as derived in Banks, Blundell and Lewbel (1997) are:

$$e_{ij} = \frac{1}{w_i}\left[\gamma_{ij} - \left[\beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left(\log\frac{x}{a(\mathbf{p})}\right)\right]\left(\alpha_j + \sum_k \gamma_{jk} \log p_k\right) - \frac{\lambda_i\beta_j}{b(\mathbf{p})}\left[\log\frac{x}{a(\mathbf{p})}\right]^2\right] - \delta_{ij} \qquad (2.26)$$

And the uncompensated income elasticity is:

$$e_i = \frac{1}{w_i}\left[\beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left(\log\frac{x}{a(\mathbf{p})}\right)\right] + 1 \qquad (2.27)$$

Income elasticity changes at different point of the x distribution. In (Banks et al., 1997) empirical Engel curves have been explored for some goods, using data from the UK Family Expenditure Survey. Engel curves estimated for clothes and alcohol present a positive β_i and a negative λ_i. According to this findings, and applying (2.27) income elasticity turns to be greater than one at low levels of x and lower than one at high level of x. Thus alcohol and clothes can be considered luxuries at low levels of total expenditure and necessities at high levels. The QUAIDS specification as a generalization of the AIDS proves to be able to account for goods whose demands react differently to income changes at different income levels. In the following sections results of the QUAIDS estimation will be explored in comparison with the original AIDS' ones.

2.6 Heterogeneous preferences: demographics in the AIDS model

The original AIDS models demand levels for the average representative consumer. Although very likely, no heterogeneity in preferences among individuals is considered in the AIDS model. Yet, estimating a demand system allowing for heterogeneous preferences among consumers is quite difficult. One of the possible approach to the heterogeneity problem is the assumption that differences in preferences can be to some extent connected to (and explained by) some socio-demographic characteristics of individuals. Belonging to different geographic areas, or to specific socio-economic group, being an adult or a child are likely to be important determinants of individual preferences.

In the following application the unity of analysis is not the individual consumer, but the household, as it often happens in applied demand analysis, as data from household budget surveys are exploited. When considering the aggregate behaviour at the household level, some characteristics concerning its composition (number of household members, age and gender of the household members) cannot be ignored in order to explain consumption choices. Consumption behaviour of households with different demographic characteristics is likely to be systematically different.

The effects of demographic characteristics on consumption patterns have been deeply explored in the past. In particular, the literature on the introduction of demographic effects into coherent demand systems is quite large. Three main approaches exist: demographic scaling, demographic translating and the Gorman procedure. Demographic scaling consists in modifying the arguments of the cost function, so that prices and total expenditure are scaled to reflect heterogeneity in household demographics. Scaling can be interpreted as adjusting prices and total expenditure to reflect equivalence scales (Lewbel, 1985;Pollak et al., 1981). This results in a demand system where the price and income coefficients depend on demographics. Demographic translating consist in allowing the constant term in a demand equation to depend on demographics, so that only preferences are allowed to vary according to household characteristics, while the other behavioural parameters (the price and expenditure coefficients) are constant across households. The Gorman procedure is basically a combination of these two approach (see also Blundell, Pashardes, and Weber (1993)).

Intermediate specifications exist. Here we describe Moro and Sckokai (2000) approach, where the demographic variables are incorporated as shifter of the intercept and the expenditure terms, while the price coefficient are kept constant to avoid overparametrisation problems. Assuming that heterogeneity in preferences is related to some socio-demographic characteristics of the household, we consider the introduction of a vector $\mathbf{z}^h = \left[z_1^h, z_2^h, \ldots z_k^h \right]$ of K characteristics of the *h*-eth household into the original AIDS model.

Following Moschini and Rizzi (1998) the translog and Cobb-Douglas price aggregators of the AIDS model ($a(\mathbf{p})$ and $b(\mathbf{p})$) are allowed to vary with the household-*h* characteristics:

$$\log a(\mathbf{p}, \mathbf{z}^h) = \alpha_o + \sum_{i=1}^{n} \alpha_i(\mathbf{z}^h) \log p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \log p_i \log p_j \tag{2.28}$$

$$\log b(\mathbf{p}, \mathbf{z}^h) = \sum_{i=1}^{n} \beta_i(\mathbf{z}^h) \log p_i \tag{2.29}$$

where according to Blundell, Pashardes, and Weber (1993) $\alpha_i(\mathbf{z}^h)$ and $\beta_i(\mathbf{z}^h)$ can be specified as follows[9]:

$$\alpha_i(\mathbf{z}^h) = \alpha_i + \sum_{k=1}^{K} \alpha_{ik} z_k^h \tag{2.30}$$

$$\beta_i(\mathbf{z}^h) = \beta_i + \sum_{k=1}^{K} \beta_{ik} z_k^h \tag{2.31}$$

The generalized PIGLOG preferences are then represented through the following cost function:

$$\log c(U, \mathbf{p}, \mathbf{z}^h) = \log a(\mathbf{p}, \mathbf{z}^h) + b(\mathbf{p}, \mathbf{z}^h)U \tag{2.32}$$

that is an extension of (2.12). The resulting expenditure share equations have the following form:

$$w_i^h = \alpha_i(\mathbf{z}^h) + \sum_{j=1}^{n} \gamma_{ij} \log p_{jh} + \beta_i(\mathbf{z}^h) \log\left(\frac{x^h}{a(\mathbf{p}, \mathbf{z}^h)} \right) \tag{2.33}$$

Some further restrictions with respect to (2.16), (2.17) and (2.18) are needed in order to guarantee the adding-up, symmetry and homogeneity conditions:

$$\sum_i \alpha_{ik} = 0 \tag{2.34}$$

$$\sum_i \beta_{ik} = 0 \tag{2.35}$$

Since in this specification price coefficients do not depend on demographic parameters, price elasticity formulas are not different from the AIDS' ones, while uncompensated income elasticity is defined as:

$$e_i \equiv \frac{\partial w_i^h}{\partial \log x^h} \frac{1}{w_i^h} + 1 = \frac{\beta_i(\mathbf{z}^h)}{w_i} + 1 \tag{2.36}$$

2.7 Demographic effects in the Quadratic Almost Ideal Demand System

The two generalization of the AIDS model described in the previous sections can be jointly applied to the original specification. Strictly following Moro and Sckokai (2000) the expenditure share equation system derived by Banks, Blundell, and Lewbel (1997) can be modified by allowing the constant term and the income coefficients to vary across different households. Another function of household characteristics $\lambda_i(\mathbf{z}^h)$ will enter the coefficient of the quadratic income term. Similarly to (2.30) and (2.31) this function is defined as:

[9] The set of additional deterministic time-dependent variables T_k which are included in the specification by Blundell, Pashardes, and Weber (1993) is not considered in this formulation which follows Moschini and Rizzi (1998).

$$\lambda_i(\mathbf{z}^h) = \lambda_i + \sum_k^K \lambda_{ik} z_k^h \tag{2.37}$$

And the consequent demand system in budget shares form is:

$$w_i = \alpha_i(\mathbf{z}^h) + \sum_{j=1}^n \gamma_{ij} \log p_j + \beta_i(\mathbf{z}^h) \log\left[\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right] + \frac{\lambda_i(\mathbf{z}^h)}{b(\mathbf{p},\mathbf{z}^h)} \left\{\log\left[\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right]\right\}^2 \tag{2.38}$$

Homogeneity and symmetry conditions are granted by (respectively) (2.17) and (2.18) while adding-up conditions requires bedsides (2.16) the following restrictions:

$$\sum_{i=1}^n \lambda_i = 0 \qquad \sum_i \alpha_{ik} = 0 \qquad \sum_i \beta_{ik} = 0 \qquad \sum_{i=1}^n \lambda_{ik} = 0 \tag{2.39}$$

In the following application a slightly different specification of the quadratic almost ideal demand system has been chosen. To overcome the estimation problems associated with the excessive number of parameters, the demand system can be simplified as follows using the intercept *translating* approach described in Lewbel (1985), where demographics only enter the $a(\mathbf{p})$ term:

$$w_i = \alpha_i(\mathbf{z}^h) + \sum_{j=1}^n \gamma_{ij} \log p_j + \beta_i \log\left[\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right] + \frac{\lambda_i}{b(p)} \left\{\log\left[\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right]\right\}^2 \tag{2.40}$$

Under this specification, the demographic variables only enter the intercept term and the Cobb-Douglas price index of the demand system; differently from Moschini and Rizzi (1998) and Moro and Sckokai (2000), where the income coefficients are also allowed to depend on **z** vector. This form of demographic translation also correspond to the one chosen by – among others – Dhar, Chavas, and Gould (2005).

Table 2.3 Uncompensated price elasticities in AIDS, LA/AIDS, AIDS with demographic effects, QUAIDS and QUAIDS with demographic effects

	Uncompensated Price elasticities
AIDS	$e_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \beta_i \frac{\alpha_i}{w_i} - \frac{\beta_i}{w_i} \sum_k^n \gamma_{kj} \log p_k$
LA/AIDS	$e_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \beta_i \frac{w_j}{w_i}$
AIDS with demographic effects	$e_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \beta_i \frac{\alpha_j}{w_i} - \frac{\beta_i}{w_i} \sum_k^n \gamma_{kj} \log p_k$
QUAIDS	$e_{ij} = \frac{1}{w_i}\left[\gamma_{ij} - \left[\beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left(\log\frac{x}{a(\mathbf{p})}\right)\right]\left(\alpha_j + \sum_j \gamma_{jk} \log p_k\right) - \frac{\lambda_i \beta_j}{b(\mathbf{p})}\left[\log\frac{x^h}{a(\mathbf{p})}\right]^2\right] - \delta_{ij}$
QUAIDS with demographic effects	$e_{ij} = \frac{1}{w_i}\left[\gamma_{ij} - \left[\beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left(\log\frac{x}{a(\mathbf{p},\mathbf{z}^h)}\right)\right]\left(\alpha_j(\mathbf{z}^h) + \sum_k \gamma_{jk} \log p_k\right) - \frac{\lambda_i \beta_j}{b(\mathbf{p})}\left[\log\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right]^2\right] - \delta_{ij}$

Elasticity formulas for the QUAIDS elasticities need to be adjusted to reflect introduction of $\mathbf{z}^h$ as shifters of the intercept. Table 2.3 and Table 2.4 summarize the elasticity formulas for the AIDS, LA/AIDS, QUAIDS and QUAIDS with demographic effects.

Table 2.4 Uncompensated income elasticities in AIDS, LA/AIDS, AIDS with demographic effects, QUAIDS and QUAIDS with demographic effects

	Uncompensated Income elasticity
AIDS, LA/AIDS, AIDS with demographic effects	$e_i = \frac{\beta_i}{w_i} + 1$
QUAIDS	$e_i = \frac{1}{w_i}\left[\beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left(\log\frac{x}{a(\mathbf{p})}\right)\right] + 1$
QUAIDS with demographic effects	$e_i = \frac{1}{w_i}\left[\beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left(\log\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right)\right] + 1$

2.8 Price and expenditure endogeneity

When Marshallian demand functions are estimated, demanded quantity is expressed as function of prices and total household expenditure. Yet, in empirical demand analysis conditional demand system for sub-groups of goods are estimated, assuming that consumer choice proceeds at different stages with separability between different groups of good (see e.g. Edgerton (1997)). For example, food expenditure can be modelled separately to expenditure in clothing, housing, etc. In this case, the expenditure term for the demand model is the total outlay for food. In this case – but also when total expenditure is considered – it is unlikely that a change in the allocation of food expenditure across different types of food does not influence total food expenditure, because of the different prices associated with each food. This generates an endogeneity problem. Even prices may be endogenous, especially when supply is rigid.

In order to account for potential endogeneity when using cross-sections, one approach is the augmentation of the demand system with additional equations where total (food) expenditure and prices are the dependent variables, which can be explained by a set of truly exogenous variables. This corresponds to instrumenting expenditure and prices in the estimation process. As it will be clear in the following paragraphs, while accounting for endogeneity may be relevant to obtain consistent estimates from the data, the use of the augmenting equations to project demand has strong implications on the policy evaluation process and – in our view – it should be avoided.

2.9 Estimation of the campaign impact through the counterfactual demand.

Once the demand model has been estimated using pre-intervention data, it can be projected over the time periods (t) after the intervention to obtain the counterfactual budget shares for each good and household (w_{tih}^{0}). This projection is obtained by applying estimated pre-intervention coefficients to post-intervention data (prices and total expenditure). When prices and total expenditure are taken as endogenous (as discussed in the previous paragraph) this complicates the policy evaluation process. In fact, projections over future periods using the augmented demand system generate a new set of price and expenditure levels which ignore any exogenous shock that may have affected the demand determinants. For example, an exogenous supply shock which may have raised prices or simply overall inflationary dynamics would enter the estimation of current demand, but would be ignored in counterfactual demand, so that they would be ascribed to the policy effect. For these reasons price and total expenditure will be treated as exogenous in the projection procedure (see section 3.8).

The coefficients of the demand model estimated on pre-intervention data reflect consumer preferences and responsiveness to prices (and total expenditure), thus they represent the reference behavioural parameters for assessing consumer response to the campaign over the following years. In other words, the estimated model enables us to build the counterfactual scenario, which is an estimate of the consumption level which would have been observed if the 5-a-day campaign had not taken place. A possible estimation of demand response to the campaign is thus achieved.

The impact on preferences of the additional information provided through the information campaign can be estimated by computing an individual treatment effect on the treated (ITT) for each household through the difference between actual consumption for each good (observed outcome) and projected (model-predicted counterfactual) consumption:

$$\Delta_{ITT_{tih}} = C_{tih}^{1} - C_{tih}^{0} \tag{2.41}$$

With $C_{tih}^{0} = \dfrac{w_{tih}^{0}}{p_{tih}} x_{tih}$.

The average ATT follows directly from (2.41) by averaging the ITT across the sample using the appropriate survey weights.

Since the policy evaluation approach used here derives the regression method for the estimation of treatment effects when evaluating policies with non-experimental data (see (Blundell (2000) and Caliendo and Hujer (2006)), the correct ATT estimation procedure for the model-based technique require that the factual outcome is also re-estimated based on the model. Thus, the ITT estimate is obtained by substituting the first term of the difference in (2.41), C_{tih}^{1}, with the estimated consumption at time *t*, based on the model estimated at time *t*.

In this case, the ITT would be computed as the difference between two estimated consumption levels at time *t*: the first term is computed by modelling time *t* data, the second trough projection of pre-intervention model on time *t* data.

3. Application

3.1 5-a-day campaign in the UK

In the UK the "5-a-day" program is promoted by the Department of Health. The campaign includes a National Food Scheme (which entitles every child aged 4-6 to a free piece of fruit each school day, by 2004), a series of local initiatives to increase access to fruit and vegetables within disadvantages communities, involvement of food industry, and an intense communication program. The beginning date for the national UK program can be set at 25 March 2003, when the official logo was launched. Some of the preliminary actions – including local 5-a-day pilot projects – were actually launched in late 2001, but these initiatives were on a very small scale, whereas the logo launch had wide press coverage and also implied initial licensing to over 550 organisations and 700 fruit and vegetables product.
The central message of the program is to stimulate people to eat at least 5 portions of mixed fruit and vegetable in a day (excluding potatoes and including only one fruit juice per day), according to WHO guidelines. The campaign can be considered as a huge communication and information programme aimed at awaken families to the importance of eating as a health prevention tool. In particular its objective deals with informing people about the exact recommended quantities and training them to count portions in order to quantify their usual personal fruit and vegetable intake and possibly improve it.

3.2 Previous evaluations of the 5-a-day campaign

To our knowledge, existing evaluations of the 5-a-day campaign impact in terms of F&V intake have been based on comparisons of consumption levels across the years.
Official evaluations of the 5-a-day program cover the School Fruit and Vegetable Scheme and the 5-a-day local community initiative. The latter is quite close to our purpose since it concerns the overall population and not a specific group (as children in school). It has been carried out trough a pre and post-intervention survey (Bremner et al., 2006). The Pre Test Survey has been administered in 2003, before the beginning of the intervention. A target and a control group have been identified. The same groups were interviewed in 2005 (Post Test Survey) in order to measure changes in consumption, attitude and knowledge. This procedure was aimed at evaluating the effect of community initiatives including home delivery services, improving transport to local markets, voucher schemes, media campaigns, etc. Yet, the

coverage of the 5-a-day campaign has been extensive and went definitely beyond the initiatives carried out at community level. TV and other media allowed a nationwide diffusion of 5-a-day messages. Thus also the control group, although not involved in specific local initiatives, has been reached by the intervention somehow. F&V consumption is measured through a 5-a-day index representing the number of portions consumed in a typical day. The pre-post survey highlighted an increase of the index from 3.36 to 3.64 for the programme areas and a slighter increase from 3.49 to 3.64 for the control areas. Although the change in the group involved in the local initiatives is not significantly different from the change in the control group, the overall positive change (across the entire group) is statistically significant (Bremner et al., 2006).

Beyond this ad hoc survey, some useful secondary data exist and have been used for assessing the impact of the campaign.

The Health Survey for England (HSE) is a yearly survey commissioned by the NHS Information Centre for health and social care on the health of people living in England, and includes questions on fruit and vegetable consumption (since 2001). The questionnaire has been designed to asses F&V consumption in the context of the 5-a-day programme. Consumption is measured in terms of number of portions, and participants aged 5 and over are asked about any fruit and vegetable consumed on the day before the interview.

Reports on the HSE (Aresu et al., 2009) account for a positive trend of consumed F&V portions from 2003 to 2007[10].

The other important source for impact evaluation of the 5-a-day campaign is the yearly Expenditure and Food Survey (Burgon, 2007). It collects data on purchases of fruit and vegetable, thus providing only an indication of consumption. Yet, it is probably the most reliable source of information for evaluating purchases trends in the UK. Thus, accounting for a percentage of wastage (normally 10% in the case of fruit and vegetables (DEFRA, 2007)) purchased quantities can be considered a good approximation of consumed quantities. The official evaluation report of the 5-a-day programme itself refers to EFS data, reporting an increase in F&V consumption from 2002-03 to 2003-04 (Bremner et al., 2006) to confirm the effectiveness of the campaign.

3.3 Data: Expenditure and Food Survey

The present analysis is based on data from the Expenditure and Food Survey (EFS) over the period 2002/3 to 2005/2006.

[10] Average of per capita F&V portions among adults increases of 0.1 portions per year from 2003 to 2006. It remains stable at 3.8 portions in 2006 and 2007 and decreases of 0.1 portions in 2008 (Aresu et al., 2009).

The timing of the campaign launch (end of March 2003[11]) is synchronised with the EFS survey period (1 April-31 March), thus allowing to consider EFS data as a sort of natural experiment for the 5-a-day policy.
Until 2005/2006, the EFS survey has covered the period running from 1 April to 31 March (fiscal year). From 2006 onwards the survey coverage has moved to the calendar year, in preparation for its inclusion to the Integrated Household Survey (IHS). From January 2008, the EFS questionnaire has become the Living Costs and Food (LCF) module of the HIS (DEFRA, 2008).
EFS data are collected from a sample of household in the UK using self-reported diaries of all purchases, including food, over a 2-weeks period (Burgon, 2007). Data include expenditure values and quantities, which are recorded where possible, and otherwise estimated. A diary of all personal expenditure is kept by each adult for two weeks, and a simplified diary is also kept by children aged 7 to 15 years for two weeks. Data on food consumption and nutrition are responsibility of DEFRA.
Goods and services are coded according to the United Nations Statistical Commission's Classification of Individual Consumption by Purpose (COICOP) developed further by Eurostat. For the food survey MAFF codes are used[12]. They allow a great level of disaggregation for food items and they are more detailed than COICOP ones. Data are collected using MAFF codes, and DEFRA supply cleaned data on food classified by MAFF codes[13].
Data derived from food diary are given by DEFRA in aggregated form. Total household expenditure for each food item in the two weeks period is provided together with purchased quantities (in grams).

3.4 Some descriptive statistics: average fruit and vegetable consumption

When computing fruit and vegetables consumption, we aggregate different fruit and vegetables (F&V) items. In the aggregation procedure we follow DEFRA choices (DEFRA, 2007) which make our results comparable with DEFRA reports on EFS data. In the “vegetables” category we include: fresh green vegetables (fresh cabbages, fresh brussels sprouts, etc.), other fresh vegetables (fresh carrots, fresh onions, fresh tomatoes, etc), processed vegetables excluding processed potatoes (canned tomatoes, canned or bottled peas, canned beans, etc). In the “fruit” category we include fresh fruit (fresh oranges, apples, pears,

[11] The Department of Health’s 5-a-day logo was launched in March 2003. This date has been chosen as the official start of the campaign.
[12] MAFF stands for Ministry of Agriculture, Food and Fishery, now DEFRA.
[13] Food data are supplied as a separate dataset by DEFRA with MAFF classification. Yet, data on food expenditure (translated into COICOP classification) can also be found in the main body of the expenditure dataset among the non food data.

etc.), processed fruit and fruit product (tinned peaches, pears, pineapples, etc..), nuts and edible seeds and pure fruit juices. This aggregation may carry some problems. Inclusion of nuts and edible seeds in the fruit category is incoherent with the 5-a-day recommendations, but this is probably of minor importance in terms of consumed quantity with respect to the total amount. Furthermore, our estimates include all the purchase of fruit juice, as opposed to the first 80 grams allowed by the 5-a-day norms. EFS data do not distinguish the first fruit juice of the day from the following ones[14]. Despite these issues, the above aggregation strategy has granted important consistency with DEFRA interpretation of food data from the EFS.

Table 3.1 shows the average per capita fruit and vegetable consumption over time, disaggregated by income quartile. Average per capita consumption is computed for every income quartile as a weighted average of per capita consumption of each household, using the sampling weights supplied with EFS data.

[14] "One 150 ml glass of unsweetened 100% fruit or vegetable juice can count as a portion. But only one glass counts, further glasses of juice don't count toward the total 5 A DAY portions" (http://www.nhs.uk/Livewell/5ADAY/Pages/FAQs.aspx)

Table 3.1 Average per capita fruit and vegetable purchases by income quartile (grams per week).

Income Quartiles	2002-03	2003-04	2004-05	2005-06	*% Variation 2002-2006*
Lowest					
Fruit consumption	1033	1027	998	1112	*+7.75*
Vegetable consumption	1059	964	1005	1073	*+1.29*
Total F&V consumption	2092	1991	2003	2185	*+4.47*
(Standard Deviation)	*(1865)*	*(1687)*	*(1751)*	*(1778)*	
Medium-low					
Fruit consumption	1236	1194	1222	1298	*+5.01*
Vegetable consumption	1142	1132	1157	1149	*+0.55*
Total F&V consumption	2379	2326	2379	2447	*+2.87*
(Standard Deviation)	*(1835)*	*(1855)*	*(1794)*	*(1811)*	
Medium-high					
Fruit consumption	1381	1414	1324	1477	*+6.94*
Vegetable consumption	1210	1265	1251	1301	*+7.53*
Total F&V consumption	2591	2680	2575	2778	*+7.22*
(Standard Deviation)	*(1983)*	*(2164)*	*(1913)*	*(2145)*	
Highest					
Fruit consumption	1768	1663	1663	1749	*-1.07*
Vegetable consumption	1357	1354	1387	1409	*+3.78*
Total F&V consumption	3125	3017	3050	3157	*+1.04*
(Standard Deviation)	*(2333)*	*(2227)*	*(2359)*	*(2217)*	
Total population					
Fruit consumption	1355	1325	1302	1409	*+4.04*
Vegetable consumption	1192	1179	1200	1233	*+3.41*
Total F&V consumption	2547	2504	2501	2642	*+3.75*
(Standard Deviation)	*(2049)*	*(2032)*	*(2005)*	*(2030)*	

Source: our processing on EFS data.

People in lower income quartiles consume less fruit and vegetable, a common finding in the literature (Pollard et al., 2008). In the 2002-03 baseline year, prior to the campaign kick-off, average F&V consumption for households in the richest quartile was 49% higher than for those in the lowest quartile. In 2004-05 the gap is still about 52%, with a decrease to 44% in 2005-06.

Assuming 80g per portion, and allowing 10% for wastage an estimate of the number of F&V portions consumed individually in a day has been computed (Table 3.2). Differences among income quartiles are quite evident. The richest quartile of the population seems to be *on average* already in line with the recommendation of 5 portions per day (5.07 in 2005-06)[15]. However the lowest income quartile is well-below the recommended threshold (3.51 portions in 2005-06).

Table 3.2 Average number of F&V portions per day by income quartile (per-capita).

Income Quartiles	**2002-03**	**2003-04**	**2004-05**	**2005-06**
Lowest				
Fruit consumption	1.66	1.65	1.60	1.79
Vegetable consumption	1.70	1.55	1.62	1.72
Total F&V consumption	3.36	3.20	3.22	3.51
Medium-low				
Fruit consumption	1.99	1.92	1.96	2.09
Vegetable consumption	1.84	1.82	1.86	1.85
Total F&V consumption	3.82	3.74	3.82	3.93
Medium-high				
Fruit consumption	2.22	2.27	2.13	2.37
Vegetable consumption	1.94	2.03	2.01	2.09
Total F&V consumption	4.16	4.31	4.14	4.46
Highest				
Fruit consumption	2.84	2.67	2.67	2.81
Vegetable consumption	2.18	2.18	2.23	2.26
Total F&V consumption	5.02	4.85	4.90	5.07
Total population				
Fruit consumption	2.18	2.13	2.09	2.26
Vegetable consumption	1.92	1.89	1.93	1.98
Total F&V consumption	4.09	4.02	4.02	4.25

Another route to estimating average per capita consumption can be pursued by computing the ratio between average (weighted) household consumption and average (weighted) number of household members per each income quartile, as shown in Table 3.3. These are exactly the estimates provided by DEFRA (DEFRA, 2007) in its 2007 report on EFS food data. Yet, since our following analysis will be focused also on consumption differences among income

[15] The issue is not of minor importance. As noted in Mazzocchi, Traill and Shogren (2009), assuming a symmetric distribution of fruit and vegetable consumption among the population the target of *an average* of 5 portions per day means that half of the population would be below the threshold. The 5-a-day message seems to require that everyone should consume five 80 grams portions per day, it follows that the average should be well-above.

quartiles, the first procedure turns out to be more consistent with the choice of defining quartiles using per capita income for each individual household[16].

Table 3.3 Average per capita fruit and vegetable purchases (grams per week), DEFRA computation strategy.

	2002-03	2003-04	2004-05	2005-06
Fruit and vegetables	2307	2273	2275	2449
Fruit	1207	1192	1171	1294
Vegetables	1100	1081	1104	1155

Source: our processing on EFS data.

As described in the previous sections, prices are likely to play an important role in affecting consumption trends even in presence of changes in information conditions. Thus prices are crucial elements of the following analysis of the 5-a-day impact.

Only a few expenditure surveys collect data on prices, asking directly for the prices faced by each respondent at the time of purchases. More commonly, price information are deduced by the knowledge of expenditures and quantities expressed in a common unit. These implicit prices are referred to as unit values. This is the case of the EFS, where information on quantities and expenditures are collected. Unit values are then computed as the ratio between expenditure and quantity purchased. When actual prices are not available (as it often happens) unit values may be used for demand estimation purposes. Unit values reflect both heterogeneity in purchase prices faced by the households (which can depend on different geographical location or time of the year), quality of chosen foods and the composition of aggregate food groups in terms of individual specific foods. This means that an increase in unit value might be a consequence of a higher quality purchase or a reallocation within the group which involves a larger quantity of relatively more expensive foods. For our purpose it looks reasonable to assume that within the same year heterogeneity in unit values across households mainly reflects different quality choices, while variations in aggregated unit values over time across groups of households (e.g. income quartiles) are likely to indicate changes in the price levels. However, a strategy to estimate quality-adjusted unit-values will be explored later.

[16] Weights are applied to per capita income when quartiles are computed.

Table 3.4 Fruit and vegetable unit values by income quartile (pound per kilogramme) and retail price changes.

	2002-03		2003-04		2004-05		2005-06		Avg. Variation 2002-06	
	Fruit	Veg	Fruit	Veg	Fruit	Veg	Fruit	Veg	Fruit	Veg
1st quartile	1.149	1.367	1.207	1.443	1.281	1.492	1.294	1.551	+3.01	+3.22
2nd quartile	1.231	1.450	1.308	1.636	1.407	1.606	1.397	1.638	+3.21	+3.10
3rd quartile	1.341	1.766	1.351	1.816	1.452	1.808	1.440	1.756	+1.80	-0.14
4th quartile	1.511	2.128	1.521	2.297	1.632	2.183	1.692	2.278	+2.88	+1.72
Tot. pop.	1.308	1.678	1.347	1.798	1.443	1.772	1.456	1.806	+2.71	+1.85
Retail price increase compared to previous year			*+2.30%*	*+8.35%*	*-4.12%*	*-2.56%*	*+2.19%*	*+2.54%*	*+0.06%*	*+2.00%*

Source: our processing on EFS data and Office for National Statistics (retail prices).

As shown in Table 3.4 unit values rise as income increases, this suggests major differences in choice for quality, composition of the fruit and vegetable basket and possibly different points of purchase. Unit values in Table 3.4 are average unit values computed using sampling weights. Considering aggregate price trends as reflected by retail price indices (RPI), on average vegetable prices have increased by 2% per year over the same period, while the level of fruit prices is steady. However, a major RPI increase was observed in the fiscal year following the 5-a-day campaign (+8.3% for vegetables, +2.3% for fruit), with a slowdown over the two subsequent years. Considering unit values and income classes, the average price increase is higher for the two lowest quartiles and especially in 2003-04 a sharp rise in unit values is observed. Unit values increase for all income quartiles, with an average rise for fruit (vegetables) of +2.7% (+1.9%) per annum over the period covered by the analysis. Again, the increase in unit values is more relevant for the two lowest quartiles (with an average rise above 3% for both fruit and vegetables). The last row of Table 3.4 shows the yearly changes in RPIs of F&V, as observed at the retail level, which confirms a noticeable price increase for 2003-04 (well above an inflation rate around 1.3%), especially for vegetables, a reduction in both prices in 2004-05, and an increase in line with a 2.3% inflation rate for 2005-06. At the population level, it is reasonable to assume that the increase in unit values over time mainly reflects an actual price increase rather than quality choices.

Finally, to complete the description of data Engel curves for fruit and vegetables have been drawn. The curves show expenditure of fruit and vegetable as a function of total household expenditure in food. They have been estimated nonparametrically using the Lowess smoothing technique as in Dhar and Foltz (2005) for each year. Figure 3.1 and Figure 3.2 report respectively Engel curves for fruit and vegetables. Note that these Engel curves are drawn with respect to total expenditure in food and not to total expenditure or income. In Banks, Blundell and Lewbel, (1997) nonparametric exploration of food Engel curves shows that the linear formulation provides a reasonable approximation for the food share curve. Our application involving fruit and vegetable with respect to total food expenditure depicts an increasing relation, and non linearity cannot be clearly rejected.

Figure 3.1 Nonparametric Engel curves for fruit expenditure (years 2002-2005)

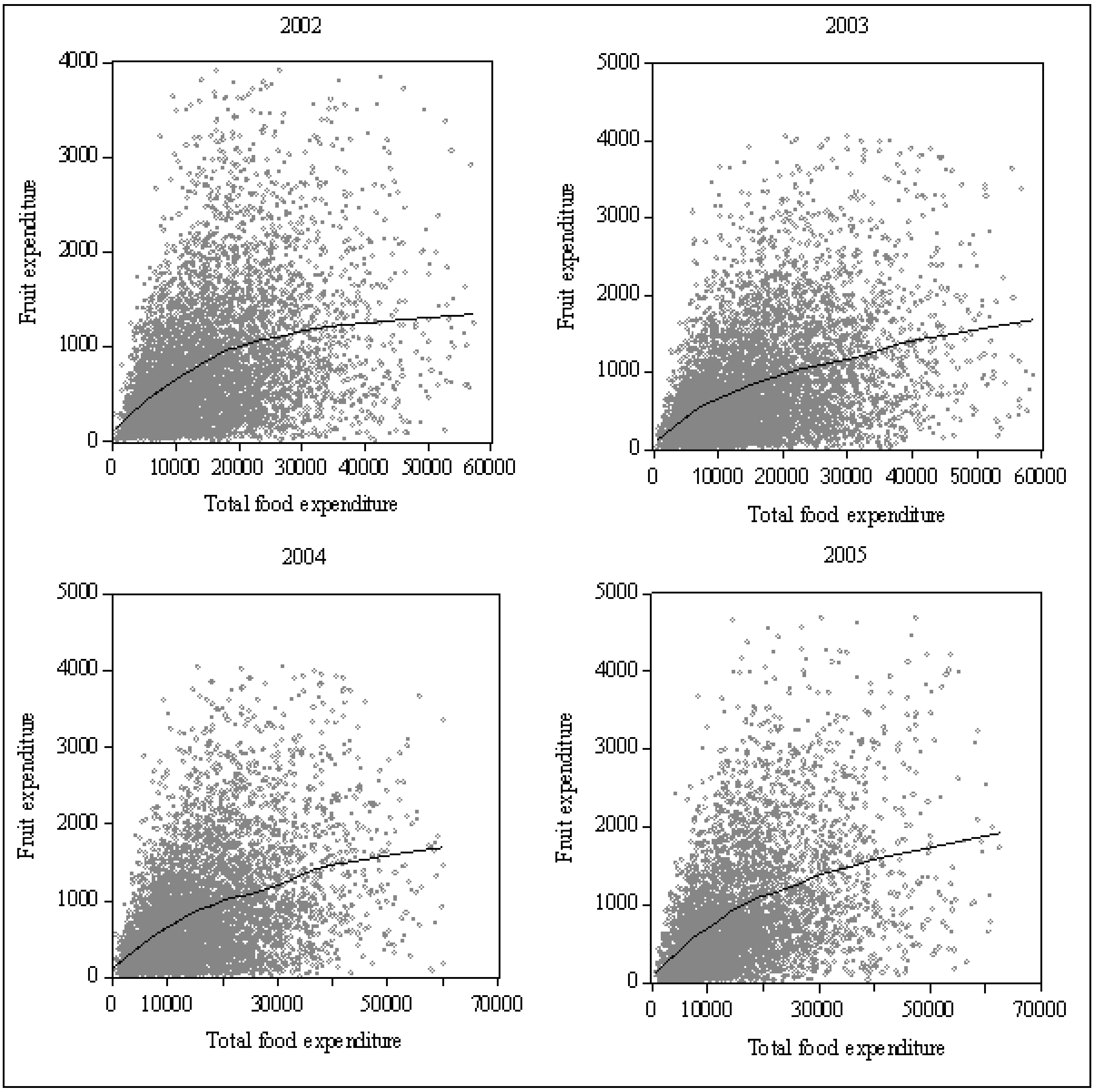
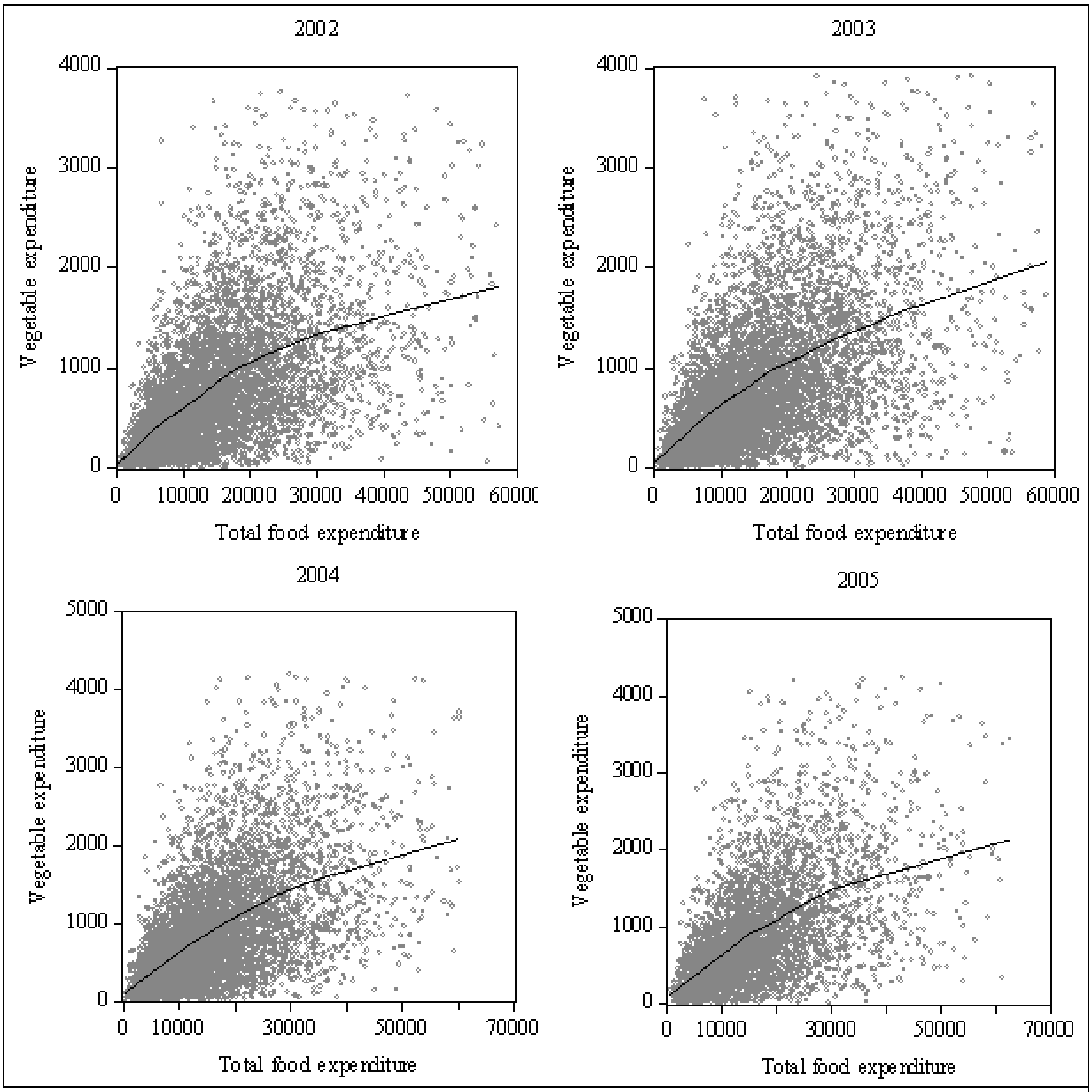


Figure 3.2 Nonparametric Engel curves for vegetable expenditure (years 2002-2005)



Finally some descriptive statistics of the variables involved in the demand system estimation have been computed by per-capita income quartile and are reported in the Appendix (see Table A.1- Table A.4).

3.5 Pre-post comparison of consumption levels.

As mentioned in section 3.2 existing (official) evaluation of the 5-a-day programme are based on comparisons of consumption levels across the years, mainly between a baseline pre-intervention period and a post intervention period. Simple pre-post percentage difference in F&V consumption levels are shown by income quartiles in Table 3.5.

Table 3.5 Change in per capita fruit and vegetable consumption with respect to baseline period (2002-03) by income quartile.

	2003-04	2004-05	2005-06
Low quartile	-4.80%	-4.25%	+4.47%
Medium-low quartile	-2.21%	-0.01%	+2.87%
Medium-high quartile	+3.41%	-0.65%	+7.22%
High quartile	-3.44%	-2.40%	+1.04%
Total Population	-1.69%	-1.78%	+3.75%

Sources: our processing on EFS data

For the whole population a drop in purchases has been registered in the two years after the beginning of the campaign , while a noticeable increase occurs in 2005-06.

Under the perspective of the simplistic pre-post comparison, these figures would demonstrate to a large extent the ineffectiveness of the campaign for all income levels, at least over the first two years after the intervention. Some positive impacts would emerge only starting from 2005-06.

As already mentioned consumption is affected by market forces besides changes in information structures. In particular prices affect consumption, and price trend should be taken into account when evaluating changes in consumption patterns.

Considering jointly Table 3.4 and Table 3.5 suggests that the impact of the information campaign over the first year could have been limited by the price dynamics, especially for the medium-low income households who suffered a higher price increase according to the unit value statistics.

Our case is probably close to the example in Figure 2.2 where the outcome variable (F&V consumption) has its own negative dynamics (due to price increases) and the intervention affects it positively by neutralizing (partially) its negative pattern. Estimation of counterfactual levels enables us to disentangle the effects of enhanced information from the potentially conflicting effect of price increases and assess the campaign impact accounting for price patterns.

3.6 Model specification.

3.6.1 Unit values and quality effects.

Food policy research often makes use of household budget cross-sectional data. With the exception of some commercial data, official household budget surveys do not collect information on retail prices directly, but only report expenditures and – sometimes - purchased quantities. As already mentioned unit values are then computed by dividing expenditures by the corresponding quantities to overcome the lack of information on actual prices. Yet, the use of unit values in place of prices might raise several issues when demand functions are estimated, in particular with reference to price elasticities. Unit values in fact may reflect actual differences in prices (within a given country and year prices can vary according to the geographical location and season), but also differences in the quality of the chosen food or the composition of aggregated food groups[17]. Thus the causes of variability other than price variation, in particular the effect of quality choices on unit values should be purged from the data in order to correctly interpret the effect of prices in consumption choices.

Commodity aggregation is often linked to the quality effect issue on prices. Cramer's distinction of goods and commodities is useful when managing the heterogeneity problem (Cramer, 1973). Specific varieties or brands of an item, sold at a single price, are defined as *goods*. When quantities of these goods can be added together (from the consumers' viewpoint) they belong to the same *commodity* (Cox et al., 1986). The classic consumer theory assumes homogeneous goods with a single price. When different goods are aggregated into a single commodity, the commodity price (computed as the average of prices of the individual goods) depends on the quantities of the single good which are purchased. Of course the bigger the heterogeneity of the aggregated commodity, the worse is the potential distortion of aggregation and quality effects on the aggregated price.

Commodity unit values are thus strictly dependent on quality choices. These quality choices can be associated by some characteristics of the household (e.g. the presence of children, the number of household members, etc.). According to this approach which is well described in Cox and Wohlgenant (1986) unit values can be adjusted for the impact of quality effects induced by household characteristics. In the present work we follow Cox and Wohlgenant (1986) and use quality-adjusted unit values to estimate commodity demand functions. The procedure consists in two steps: the prior estimation of prices from observed unit values and the subsequent estimation of the demand system based on estimated prices. The rational of the procedure is that quality choices occur before the commodity quantity choice. In other word consumers first decide the quality of their purchases (in terms of selection of specific components of the aggregate commodity) and then they choose the quantity of commodity for purchase. It follows that quality choices can be modelled independently of (i.e. before) the quantity decision.

The empirical procedure for price estimation is the following. First, *n* clusters depending on the month of purchase and the region where the household resides have been identified, and

[17] Actually, as noted in Cox and Wohlgenant (1986) the issue of price variability raises whenever cross-sectional data on purchases are used, since cross-section are usually characterized by sensitive variability of prices while consumer theory assumes prices to be constant among individuals.

unit value means for each commodity g are computed ($\overline{UV_n}$) for each cluster[18]. Quality effects can then be identified as the difference between the unit value faced by h-th household and the corresponding cluster average unit value. This difference is then explained through a vector of j household characteristics (D_{hj}) plus the quantity of commodity g consumed by the household (q_h^g):

$$UV_h^g - \overline{UV}_n^g = \sum_j^m b_{1j} D_{hj} + b_2 q_h^g + e_h^g \qquad (3.1)$$

Note that differently from Cox and Wohlgenant (1986) and following Crawford, Laisney and Preston (2003) average consumed quantity of commodity g in cluster n has been added. The residuals generated form each regression are then added to the average unit value $\overline{UV_n}$ and quality adjusted prices are then obtained:

$$p_h^g = \overline{UV}_n^g + \hat{e}_h^g \qquad (3.2)$$

In the present application the quality adjusted prices are estimated using the following household characteristics: household size, type of family dummies, gross current income, age and education of the household reference person, presence of female components, a dummy for households of pensioners only.

3.6.2 The demand system

The counterfactual estimation is obtained using a Quadratic Almost Ideal Demand System with demographic effects, after controlling for price and total food expenditure endogeneity.

We estimate a conditional food demand system considering three aggregate food groups, fruit, vegetables and a residual (numeraire) aggregate including all other foods. While food quantities are all measured in grams, the inclusion of other categories than food is not feasible, since other aggregates (e.g. clothing) are too heterogeneous in terms of quantity measurement (e.g. number of items) to allow an acceptable estimation of unit values. The numeraire category was dropped to overcome the singularity problem, so that the demand system is a bivariate system, with one equation for fruit and one equation for vegetables, while total expenditure x relates to total food purchases[19].

To capture eventual nonlinearity of fruit and vegetable Engel curves the quadratic specification of the AIDS model as introduced by Banks, Blundell and Lewbel (1997) is adopted. The non linearity of the original demand system defined by Deaton and Muellbauer

[18] The same approach has been followed by Lazaridis (2003).

[19] The system is invariant to which equation is deleted and the parameters of the dropped equation (other foods) are derived from the adding up conditions.

(1980a) is maintained. Thus the Cobb-Douglas price index A(**p**) is used (see equation 2.13). As mentioned in Chapter 2 when estimating the non linear demand system, α_0 in 2.13 has to be a priori defined.

If we consider the hypothetical situation in which all prices are unity (one pence in our case) the price index A(**p**) would simply be equal to α_0. Since real expenditure in good *i* (and w_i) is required to be positive, according to the AIDS specification, the minimum level of the logarithm of real total expenditure observed (*x*) places an upper bound on $\ln A(\mathbf{p})$ and therefore on α_0 (Banks et al., 1997). In other word this parameter can be interpreted as the outlay necessary for a minimum standard of living (Deaton et al., 1980a) given the price levels. In our application α_0 represents the total expenditure in food required for the above minimum standard and is set equal to the minimum of $\log x$ [20].

In order to account for heterogeneity in preferences, demographic translating is employed. Heterogeneity is supposed to be related to some socio-demographic characteristic of the household.

The QUAIDS specification first introduced by Banks, Blundell and Lewbel (1997) is generalized allowing for the inclusion of a set of demographics (**z**). Translating implies that the intercept of the model is let to be a function of some demographic characteristics. For computational reasons only the intercept is specified as function of demographic variables, differently from Moro and Sckokai (2000) and Moschini and Rizzi (1998) who apply a scaling rather than translating technique where the intercept and the income coefficients depend on household characteristics. In principle, the price coefficients might also be allowed to vary with demographics, but there is a clear trade-off between the number of coefficients allowed to vary with demographic characteristics and the number of unknown parameters in the model (see Moro and Sckokai (2000)). Thus, in the model specification vector **z** appears in the intercept and in the price index A(**p**). The translating approach is followed among others by Dhar, Chavas and Gould (2005).The final demand system is then specified as follows:

$$w_i = \alpha_i(\mathbf{z}^h) + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \log\left[\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right] + \frac{\lambda_i}{b(p)}\left\{\log\left[\frac{x^h}{a(\mathbf{p},\mathbf{z}^h)}\right]\right\}^2 \qquad (3.3)$$

where the set of demographic variables (**z**) includes: number of adults and number of children in the household, the current gross income of the household, the age and the level of education (in terms of years of education) of the household reference person, the presence of female household members (as a binary variable), a dummy for household of pensioners only.

[20] This choice is also recommended in Buse and Chan (2000). In that study through a Monte Carlo simulation it is also shown that conditioning on a plausible prior value of α_0, even if not exact, does not cause a substantial bias.

To control for potential price and expenditure endogeneity the augmented regression approach discussed in Blundell and Robin (1999) to overcome the total expenditure endogeneity problem is used and extended to the price endogeneity issue, according to the experimental approach illustrated by Dhar, Chavas and Gould (2003) and Dhar and Foltz (2005). Three price equations and a total expenditure equation are added to the system. The reduced form price equation is specified as follows:

$$p_i = \theta_i + \sum_k^{13} \theta_{ik} g_k + \sum_l^{3} \theta_{il} q_l \qquad (3.4)$$

where g_k are 13 dummies for the geographic areas and q_l is the quarter in which data have been collected. The regional dummies are defined according to the GORs (Government Office Regions) classification, which in 1996 became the primary classification for the presentation of regional statistics in England. In fact GORs only occur in England, but Scotland, Wales and Northern Ireland are regarded as an equivalent for statistical purposes. Thus, the areas considered are: North East, North West, Merseyside, Yorkshire and the Humber, East Midlands, West Midlands, Eastern, London, South East, South West, Wales, Scotland and Northern Ireland.

Similarly the reduced total food expenditure function is specified as follows:

$$x = c_1 + c_2 INC + c_3 TOTEXP + c_4 TOTEXP^2 + c_5 HHSIZE \qquad (3.5)$$

where INC is the household gross current income, TOTEXP is the total household expenditure and HHSIZE is the number of household components.

Thus the system adopted to estimate the demand level for fruit and vegetables comprehends two share equations (for fruit and for vegetables, according to the adding-up restriction “other food” is considered as the numeraire item) (3.3), three reduced form price equations (fruit price, vegetable price and other food price) (3.4) controlling for price endogeneity and one food expenditure equation (3.5) controlling for expenditure endogeneity. Homogeneity and symmetry constraints have been imposed to fulfil theoretical requirements.

3.7 Empirical results.

The system is non linear and is estimated using Full Information Maximum Likelihood (FIML) procedures, separately on data from the 2002-03, 2003-04, 2004-05 and 2005-06 EFS.

Table 3.6 provides parameter estimates for the four estimated demand systems. Among the demographic variables, attributes of the household head (education and age), and the presence of female members affect both fruit and vegetable consumption across all the years (coefficients are all significant at 1% significance level, except for the presence of females parameter which has a lower significance level for fruit demand in 2003 and 2005). All the other demographic variables are significant at 5% level of significance for the fruit or the vegetable share equation, at least for one year.

The parameter measuring how consumption changes with total food expenditure is significant for fruit and for the vegetable (5% level) for all the observed years, except for 2002-03 vegetable consumption. The quadratic term for total food expenditure results constantly significantly different from zero for fruit (10% level), except for 2002-03 data, while it is not significant for vegetables in the same period.

Table 3.6 Demand system regression results (years 2002-03, 2003-04, 2004-05, 2005-06).

	2002-03		2003-04		2004-05		2005-06	
	Fruit	**Veg.**	**Fruit**	**Veg.**	**Fruit**	**Veg.**	**Fruit**	**Veg.**
Intercept	0.114	0.039	0.224	0.155	0.264	0.163	0.206	0.203
	(0.016)	(0.372)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Demographics								
Number of adults	-2.363	0.321	-1.552	-2.379	-4.198	-0.189	-0.698	-0.439
	(0.071)	(0.761)	(0.212)	(0.034)	(0.001)	(0.860)	(0.646)	(0.731)
Number of children	-0.181	-2.126	0.443	-2.103	0.090	-1.280	1.249	-0.649
	(0.854)	(0.007)	(0.644)	(0.008)	(0.923)	(0.116)	(0.255)	(0.489)
Gross current income	0.004	0.004	0.003	0.001	0.005	0.003	0.011	0.003
	(0.001)	(0.019)	(0.041)	(0.889)	(0.001)	(0.013)	(0.001)	(0.102)
Age of household head	0.963	0.392	0.937	0.461	0.849	0.443	0.656	0.286
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Education of household head	3.698	2.606	3.277	2.260	1.863	1.550	1.235	1.265
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Presence of female members	6.730	11.745	3.193	15.714	5.703	12.230	3.436	10.825
	(0.001)	(0.001)	(0.079)	(0.001)	(0.002)	(0.001)	(0.109)	(0.001)
Household of pensioners	-8.496	-2.696	-4.431	-0.847	-2.128	-1.772	-2.884	-1.503
	(0.001)	(0.145)	(0.047)	(0.693)	(0.400)	(0.456)	(0.384)	(0.625)
Prices								
Fruit price	-0.040	0.025	0.001	0.009	0.006	-0.013	-0.013	0.025
	(0.014)	(0.004)	(0.991)	(0.261)	(0.580)	(0.070)	(0.534)	(0.006)
Vegetable price	0.025	0.022	0.009	0.036	-0.013	0.041	0.025	0.042
	(0.004)	(0.003)	(0.261)	(0.001)	(0.070)	(0.001)	(0.006)	(0.001)
Other food price	0.015	-0.047	-0.009	-0.045	0.007	-0.028	-0.011	-0.066
Food expenditure								
Total food expenditure	-0.030	0.011	-0.056	-0.025	-0.062	-0.025	-0.034	-0.026
	(0.039)	(0.415)	(0.001)	(0.012)	(0.001)	(0.059)	(0.053)	(0.091)
Total food expenditure (quadratic)	0.001	-0.002	0.003	0.001	0.003	0.001	0.001	0.001
	(0.415)	(0.042)	(0.012)	(0.435)	(0.059)	(0.639)	(0.091)	(0.615)
Adj R^2	0.101	0.573	0.165	0.102	0.147	0.110	0.121	0.084

Notes: P-values in brackets. All demographic coefficients have been multiplied by 1,000.

The homogeneity and symmetry restrictions have been tested before being imposed. Since both homogeneity and symmetry condition affect the coefficients of the system through linear constraints, the Wald test is used. In Table3.7 results of the Wald test on 2002-03 demand system are reported.

Table 3.7 Test of symmetry and homogeneity constraints for 2002-03 demand system.

Wald test	Degrees of freedom	χ^2	P-value
Homogeneity and symmetry constraints	3	40.895	0,000
Homogeneity constraint	2	22.176	0,000
Symmetry constraint	1	0.213	0,645

Symmetry restriction is not rejected (P-value is about 65%), thus the symmetry condition results consistent with the data. Still, homogeneity is rejected, a common finding in empirical demand analysis with large cross-section samples (see Keuzenkamp and Barten (1995)). Here we opted for the consistency with economic theory and imposed both constraints for the simulation over all the years.

Following Moro and Sckokai (2000) the overall significance of demographic effects and the QUAIDS versus AIDS specification have been tested for the 4 systems. Since the considered restrictions are linear, the Wald test is used. Results of the hypothesis testing are reported in Table 3.8. The first Wald test tests the null hypothesis that both the quadratic expenditure coefficients are zero. Through this test a comparison between two nested models (respectively AIDS in the QUAIDS) is performed (see paragraph 3.7.2.1 for the description of the criteria used for nested models comparison). The rejection of the null hypothesis in the first 3 systems (at 5% significance level, except for 2002-03) leads to reject the AIDS model in favour of the QUAIDS specification. AIDS specification is not rejected in the 2005-06 model.

The second test on demographics verifies the hypothesis of total absence of demographic effects on demanded quantities. The test implies a comparison of the QUAIDS specification without demographics and the QUAIDS specification with demographic effects (in which the first is nested). Absence of overall demographic effects is strongly rejected (1% significance level) for all the years.

Table 3.8 Wald tests for quadratic specification and demographic effects (for all years)

		χ^2	Degrees of freedom	P value
2002-03	Quadratic specification	5.306	2	0.070
	Demographic effects	702.843	14	0.001
2003-04	Quadratic specification	6.328	2	0.042
	Demographic effects	656.406	14	0.001
2004-05	Quadratic specification	8.862	2	0.012
	Demographic effects	1280.786	14	0.001
2005-06	Quadratic specification	0.481	2	0.786
	Demographic effects	524.157	14	0.001

Finally, parameter estimates for the reduced form price and food expenditure equations are reported in the Appendix (see Table A.5- Table A.8).

3.7.1 Elasticity estimates.

Price and expenditure elasticities are estimated according to the formulas in Chapter 2 and computed at the sample means, using the mean values of the regressors[21]. Estimates are reported in Table 3.9.

Table 3.9 Direct and cross price elasticities and food expenditure elasticities (years 2002-03, 2003-04, 2004-05, 2005-06).

	2002-03		2003-04		2004-05		2005-06	
	Fruit	Vegetables	Fruit	Vegetables	Fruit	Vegetables	Fruit	Vegetables
Price:								
Fruit	-1.909	0.292	-1.666	-0.297	-1.647	-0.707	-1.487	0.111
Vegetables	0.526	-0.577	-0.145	-0.608	-0.518	-0.514	0.155	-0.558
Food Expenditure	0.513	1.192	0.078	0.584	-0.031	0.586	0.493	0.559

Total food expenditure elasticities have the expected positive sign, except for fruit demand in 2004-05, which means that a positive change in consumption is expected against a one percent positive change in price. Direct price elasticities have the expected negative sign both for fruit and vegetable across all the observed years. In general fruit elasticity seems quite

[21] The elasticities sample estimates (h_j) converge to the true elasticities (e_{ij}), evaluated at the limiting values of the sample means of the regressors (Greene, 2008).

high with respect to other published results.[22] Cross-price elasticities are positive in 2002-03 and in 2005-06, when fruit and vegetables seems to be perceived as substitute goods. In 2003-04 and 2004-05 the opposite occurs, and cross-price elasticities are negative (fruit and vegetables behave as complementary goods).

3.7.2 Selection of the model.

3.7.2.1 Model comparison criteria: nested and non-nested models

The specification analysis is the process of detecting the appropriate functional form of a model (i.e. the functional form which enables the model to represent the reality and foresee future agents' behaviour). In econometric terms the appropriate model is the one correctly representing the unknown data generating process.

The first step in the analysis of different functional specifications is the distinction between nested and non nested models.

Two models are nested if one can be derived as a particular case of the other. Formally, two models, M_0 and M_1, defined by the following likelihood functions:

$$\begin{aligned} L_{M_o}(\theta_0) &= f_{M_o}(x_1, x_2, ..., x_T, \theta_o) \\ L_{M_1}(\theta) &= f_{M_1}(x_1, x_2, ..., x_T, \theta) \end{aligned} \tag{3.6}$$

are nested (M_0 is nested in M_1) if M_1 can be redefined so that

$$L_{M_1}(\theta) = f_{M_1}(x_1, x_2, ..., x_T, \theta_0, \lambda) \tag{3.7}$$

and M_0 is obtained by imposing the set of constraints $\lambda = g(\theta) = 0$ on parameters (Gardini et al., 2000). Two nested models can be compared by testing the following generic hypothesis on parameters:

$$H_0 : g(\theta) = 0 \tag{3.8}$$

The above hypothesis can be verified in different ways. In the following work among the others the Wald test and the Likelihood Ratio test will be considered[23].

The Wald test requires only the estimation of the unrestricted model. After computing an estimate of parameters ($\hat{\theta}$) the hypothesis that $g(\hat{\theta})$ is sufficiently close to zero to avoid rejection of the null hypothesis is verified. Under the null hypothesis the Wald test statistic is distributed as a χ^2 with *J* degrees of freedom (where *J* is the number of restrictions imposed

[22] Durham and Eales (2006) employs weekly data from several retail stores on fruit prices and sales in the US to estimate elasticities of individual fruits. An interesting exercise is shown: by employing different elasticity formulas commonly used in the literature values between -0.208 and -1.32 have been found.

[23] A third procedure is often followed. It is the Lagrange Multiplier test, and it is based on the estimation of the restricted model as opposed to the Wald test (Greene, 2008).

by the null hypothesis). Following Greene (2008) the Wald statistic is the distance measure for the degree to which the unrestricted estimator fails to satisfy restrictions.
The Likelihood Ratio test requires the estimation of both the unrestricted and the restricted model and is based on the comparison of the maximum likelihoods. If H_0 is true the maximum likelihood of the restricted model should not be significantly lower than the unrestricted one. The following ratio is computed:

$$\delta = \frac{\max L \mid H_o}{\max L} \in [0,1] \tag{3.9}$$

The numerator of (3.9) is a restricted optimum and cannot be lower than the unrestricted one (the denominator), thus δ ranges from 0 to 1. When δ is close to 1 the restriction outlined by the null hypothesis does not cause a huge decreasing of the likelihood and H_0 is not rejected. When δ is close to 0, the opposite occurs and H_0 is rejected. Under regularity conditions and in large samples it can be demonstrated that $-2\log\delta$ is distributed as a chi-squared with J degrees of freedom as above.
Thus, $-2\log\delta$ is used as test statistics (and it is commonly referred to as the Likelihood Ratio). The bigger is the likelihood reduction caused by the restrictions, the closer the test statistic is to 0.
Note that asymptotically the Wald test and the Likelihood Ratio test are equivalent, but in finite samples they can lead to contrasting results.
When two models are not nested they can be compared through some measures of their goodness of fit[24]. Among the others the most known are the adjusted R^2 and the so called information criteria.
The adjusted R^2 is a measure of the quote of variability of the dependent variable explained by the model, adjusted for degrees of freedom. For the linear model R^2 and R^2adjusted are defined respectively as:

$$R^2 = 1 - \frac{\sum_i^N \hat{u}_i^2}{\sum_i^N \left(y_i - \overline{y}\right)^2} = \frac{\sum_i^N \left(\hat{y} - \overline{\hat{y}}\right)^2}{\sum_i^N \left(y_i - \overline{y}\right)^2} \qquad R^2_{adj} = R^2 \frac{n-1}{n-K} - \frac{K-1}{n-K} \tag{3.10}$$

where u_i is the *i-th* residual, *n* is the observation number and *K* is the variables number.
Since it can be shown algebraically that R^2 increases if any variable is added to the model, R^2adjusted has been introduced. The inclusion of a further variable does not imply an automatic increase in R^2 adjusted.

[24] Many other strategies can be used to compare non-nested models: methods based on auxiliary models (through an "artificial" nesting of the two models), bayesian methods, and the encompassing strategy (Gardini et al., 2000).

The information criteria include a set of measures combining the goodness of fit of the model with the need of a small number of parameter.
The most known information criterion is the Akaike Information Criterion (AIC), defined as:

$$AIC = -\frac{2}{n}\log L_{M_i} + \frac{2}{n}K_i \tag{3.11}$$

Where L_{M_i} is the maximum of M_i model likelihood function and K_i is the number of parameters to be estimated in the *i-th* model. The model with the lower AIC index should be chosen.
Another well-known information criterion is the Schwarz criterion (SC), computed as:

$$SC = -\frac{2}{n}\log L_{M_i} + \frac{\log n}{n}K_i \tag{3.12}$$

As for the AIC measure, the model with lower SC should be chosen. Since it includes explicitly the number of observations it normally leads to prefer smaller model as regards to the AIC index.
The strategies explained in this section will be used in the following analysis to compare different AIDS specification for the model- based estimation of the counterfactual.

3.7.2.2 Alternative specifications and selection of the model.

A series of different possible demand systems have been estimated using baseline data (2002-03 food data from the EFS) before selecting the QUAIDS specification with demographic effects augmented with price and expenditure equations. Alternative functional forms (based on some extension of the original AIDS model as described in the previous chapter) have been evaluated in terms of predicting ability and other criteria in order to select the most convenient specification. In Table 3.10 description of 5 alternative models is reported. The first two models control only for endogeneity of food expenditure and differ for the specification of food expenditure equation (one is linear and the other has a quadratic total expenditure term). The last three models include price equations controlling for price endogeneity and differ for the food expenditure function specification and the set of demographic variables included in the demand system.

Table 3.10 Description of five alternative model specifications.

	Regressors of the food expenditure equation	Price equations	Demand system demographic regressors
Model 1	Gross current income, total expenditure, number of adults, number of children	No	Household size, family types [a], gross current income, age and education of the household head, presence of female members, household of pensioners only
Model 2	Gross current income, total expenditure, total expenditure (quadratic) number of adults, number of children	No	Household size, family types, gross current income, age and education of the household head, presence of female members, household of pensioners only
Model 3	Gross current income, total expenditure, number of adults, number of children	Yes	Household size, family types, gross current income, age and education of the household head, presence of female members, household of pensioners only
Model 4	Gross current income, total expenditure, number of adults, number of children	Yes	Number of adults, number of children , gross current income, age and education of the household head, presence of female members, household of pensioners only
Model 5	Gross current income, total expenditure, total expenditure (quadratic) number of household members	Yes	Number of adults, number of children , gross current income, age and education of the household head, presence of female members, household of pensioners only

Notes: (a) dummies for three family types have been included: household of single members, household of two or more adults with children, and household of two or more adults without children.

The five alternative models[25] have been compared according to the criteria exposed in the previous paragraph. They are all non-nested models, and Table 3.11 shows AIC, BIC and Adjusted R^2 values.

Besides, significance levels of the quadratic expenditure term coefficient for fruit and vegetables equations are reported.

AIC and BIC values drop dramatically when price equations are added to control for price endogeneity. Homogeneity is rejected and symmetry is accepted for all the specifications. Model 5 has the lowest AIC and BIC values, and it shows a significant quadratic total expenditure term (for vegetables) differently from the other models. For these reasons the Model 5 specification has been chosen for the estimation and projection procedure.

[25] Estimation output of the five alternative models can be found in the Appendix (see Table A.9-Table A.13).

Table 3.11 Model comparison

	Model 1	Model 2	Model 3	Model 4	Model 5
Quadratic expenditure term in fruit share equation (coefficient P-value)	0.705	0.800	0.600	0.822	0.417
Quadratic expenditure term in vegetable share equation (coefficient P-value)	0.153	0.135	0.137	0.104	0.072
Adj R^2 – fruit equation	0.172	0.171	0.164	0.164	0.163
Adj R^2 – vegetable equation	0.104	0.104	0.100	0.560	0.100
AIC	13.550	13.490	6.030	6.030	6.000
BIC	13.560	13.530	6.120	6.060	6.030
Homogeneity constraint (Wald test P-value)	0.002	0.002	0.001	0.001	0.001
Symmetry constraint (Wald test P-value)	0.360	0.351	0.633	0.640	0.645
Homogeneity and symmetry constraints (Wald test P-value)	0.001	0.001	0.001	0.001	0.001

3.8 Model projections and counterfactual scenarios

As described in Chapter 2 the QUAIDS model estimated on 2002-03 is employed to project consumption levels over the following years, conditional on current prices and expenditures, but based on pre-intervention preferences (*projected quantities*). Technically, coefficients of the QUAIDS model estimated on 2002-03 data representing pre-intervention preferences are employed on prices and expenditure levels of the following years. Assuming that the 5-a-day campaign affects consumers' preferences, projected consumption levels can be interpreted as the counterfactual scenario, i.e. the fruit and vegetables quantities which would have been demanded at time $t+1$ ($t+1$ prices and expenditures are employed) in absence of the intervention (applying pre-intervention (time *t*) preferences).

The augmented approach to account for endogeneity employed for model estimation allows consistent estimates of the parameters. However, projecting consumption with endogenous prices and expenditure would imply that any exogenous shocks (e.g. a supply or import shock) is not taken into account and all consumption changes are imputed to the 5-a-day campaign. For this reason projection is computed only on expenditure shares, considering prices and total expenditure as exogenous variables (in spite of the 6 equations system in which prices and total expenditure appears as dependent variables).

Projected outcome is then compared to the factual outcome in order to measure the effect of the intervention across the years. For consistency with the model-estimated counterfactual consumption level the factual consumption level at time *t* has been re-estimated using the time *t* model (*estimated quantities*).

Table 3.12 Impact of the 5-a-day campaign on fruit and vegetable consumption (quantity of fruit and vegetables consumed per person per two weeks).

2003	Projected fruit quantity (counterfactual)	Estimated fruit quantity	Impact	Impact (%)	Projected veg. quantity (counterfactual)	Estimated veg. quantity	Impact	Impact (%)
1st quart.	2735	2678	-56.791	-2.08	2418	2452	34.064	1.41
	(2634)	(2435)	(452.52)		(1875)	(1899)	(150.43)	
2nd quart.	3149	3101	-48.364	-1.54	2759	2776	16.802	0.61
	(3475)	(3241)	(431.09)		(3147)	(3101)	(179.62)	
3rd quart.	3497	3415	-81.510	-2.33	3313	3321	8.791	0.27
	(2626)	(2579)	(470.7)		(12346)	(12733)	(448.63)	
4th quart.	4552	4393	-159.159	-3.50	3529	3449	-80.101	-2.27
	(5760)	(5097)	(837.37)		(9398)	(9153)	(338.45)	
Tot.pop.	3501	3414	-87.199	-2.49	3011	3005	-5.631	-0.19
	(3929)	(3582)	(577.60)		(8014)	(8089)	(308.55)	
2004	Projected fruit quantity (counterfactual)	Estimated fruit quantity	Impact	Impact (%)	Projected veg. quantity (counterfactual)	Estimated veg. quantity	Impact	Impact (%)
1st quart.	2714	2620	-94.175	-3.47	2514	2541	27.236	1.08
	(2478)	(2035)	(822.22)		(3694)	(3476)	(429.12)	
2nd quart.	3060	3006	-53.454	-1.75	2819	2868	49.213	1.75
	(2231)	(2071)	(563.78)		(4108)	(4082)	(312.77)	
3rd quart.	3412	3310	-101.745	-2.98	3020	3024	3.770	0.12
	(2167)	(2018)	(628.78)		(2427)	(2394)	(301.69)	
4th quart.	4193	4023	-169.554	-4.04	3438	3335	-102.984	-3.00
	(2734)	(2549)	(785.2)		(2684)	(2565)	(344.86)	
Tot.pop.	3375	3269	-106.134	-3.14	2962	2955	-7.797	-0.26
	(2478)	(2249)	(707.52)		(3297)	(3195)	(354.14)	
2005	Projected fruit quantity (counterfactual)	Estimated fruit quantity	Impact	Impact (%)	Projected veg. quantity (counterfactual)	Estimated veg. quantity	Impact	Impact (%)
1st quart.	2716	2735	19.686	0.72	2520	2583	62.649	2.49
	(2353)	(2051)	(672.81)		(2394)	(2305)	(432.79)	
2nd quart.	3264	3270	5.563	0.17	2838	2876	37.999	1.34
	(2341)	(2027)	(713.12)		(2340)	(2210)	(401.46)	
3rd quart.	3521	3575	53.770	1.53	3145	3167	22.192	0.71
	(2355)	(2158)	(543.47)		(4202)	(4067)	(348.53)	
4th quart.	4335	4454	118.513	2.73	3308	3271	-37.000	-1.12
	(4233)	(4207)	(875.84)		(2439)	(2641)	(521.389)	
Tot.pop.	3491	3542	51.223	1.47	2965	2985	20.019	0.68
	(3027)	(2877)	(715.5)		(2974)	(2928)	(433.20)	

Notes: standard deviations in brackets.

Average per-capita *estimated* and *projected* demanded quantities are computed using sample weights. Figures represent per-capita weekly fruit and vegetables consumption in grams. The same quantities are also computed per per-capita income quartile (Table 3.12)[26]. Differences of *estimated* and *projected* quantities are reported as an estimate of the ATT of equation (2.4). Percentage differences are also provided.

As variability measure of the impact standard deviation of the differences are computed (in brackets in Table 3.12 and in Table 3.13).

Some major results emerge from the simulation. First, very slight or no impact is observed over the first two years neither for fruit nor for vegetables. In 2005-06 some positive effect for fruit consumption (+1.47) and a weak increase in vegetable consumption (+0.68) are observed.

Second during the first two years there is a noticeable difference between fruit and vegetables, as on average a deep negative change in fruit consumption is observed across the income quartiles compared to some slight increases (or very slight decreases) for vegetables.

Third, there are some differences across income distribution. In regard to vegetables, the richest quartile of the population shows the worse reaction to the intervention. Negative changes in vegetables consumption are recorded only for the fourth quartile of the population across all the years. On the other hand with regard to fruit in 2005-06 when some effects occur the average increase of 1.5% in consumption levels appears highly dragged by the highest income quartiles (+1.53 for the 3rd quartile and +2.73 for the 4th quartile).

When the aggregated category of fruit and vegetable is considered (Table 3.13), a negative impact is shown for all income quartiles till 2005, when a first positive trend in F&V consumption appears.

[26] Per capita income quartiles are computed on each year using sampling weights.

Table 3.13 Impact of the 5-a-day campaign on aggregated fruit & vegetables consumption (quantity of fruit and vegetables consumed per person per two weeks).

2003	Projected F&V quantity (counterfactual)	Estimated F&V quantity	Impact	Impact %
1st income quartile	5170	5155	-15.286	-0.296
	(3607)	(3489)	(484.61)	
2nd income quartile	5922	5895	-26.448	-0.447
	(5188)	(5005)	(450.76)	
3rd income quartile	6865	6795	-70.009	-1.020
	(13636)	(14009)	(657.9)	
4th income quartile	8105	7872	-232.806	-2.872
	(11492)	(10953)	(908.26)	
Total population	6555	6466	-88.398	-1.349
	(9629)	(9559)	(661.54)	

2004	Projected F&V quantity (counterfactual)	Estimated F&V quantity	Impact	Impact %
1st income quartile	5283	5219	-64.151	-1.214
	(5043)	(4568)	(1037.46)	
2nd income quartile	5907	5900	-6.818	-0.115
	(5251)	(5166)	(605.28)	
3rd income quartile	6420	6321	-98.782	-1.539
	(3635)	(3551)	(675.62)	
4th income quartile	7619	7347	-272.922	-3.582
	(4299)	(4113)	(819.23)	
Total population	6361	6246	-115.160	-1.810
	(4655)	(4438)	(800.87)	

2005	Projected F&V quantity (counterfactual)	Estimated F&V quantity	Impact	Impact %
1st income quartile	5325	5405	79.861	1.500
	(4002)	(3601)	(1077.78)	
2nd income quartile	6099	6139	40.837	0.670
	(3968)	(3494)	(1098.47)	
3rd income quartile	6686	6756	69.923	1.046
	(5313)	(5063)	(799.21)	
4th income quartile	7709	7784	75.059	0.974
	(5378)	(5433)	(1254.93)	
Total population	6505	6572	66.243	1.018
	(4826)	(4605)	(1071.37)	

Notes: standard deviations in brackets.

4. Conclusions

The simulation described in the present thesis is far from providing an exhaustive evaluation of the 5-a-day campaign in the UK. The issues to be explored with this approach are numerous, in particular with regard to the model employed to estimate the counterfactual demand level. Nonetheless the results of this work if intended as intermediate are encouraging. In fact, the estimation of a model projected demand system offers an interesting research perspective for the evaluation of those interventions which are intended to affect consumption and for which control groups do not exist or are not artificially detectable (e.g. with matching procedures).

The modelling of consumption data over the years after the beginning of the intervention has shown some interesting features. First, direct price elasticity of fruit demand seems to be generally high and to decrease over time, differently from vegetables elasticity which simply fluctuates around -0.5. Moreover an unclear oscillation of food expenditure elasticity (both for fruit and for vegetable demand) is also observable.

In principle a public information campaign with the objective of encouraging consumption of a particular good (for health purpose) might have the effect of reducing the price elasticity of demand of such good. In general, from 2003-04 our estimations show a clear push toward inelasticity, which in principle could be consistent with behavioural effects of 5-a-day, although these changes and the extent to which they can be ascribed to the campaign would need further analysis and different instruments.

According to our results, the policy seems to be ineffective, with a negative sign for the ATT of fruit consumption for all income quartiles till 2005, and a positive, but small ATT for vegetables consumption across the years.

The high price elasticity for fruit seems to suggest that any impact on preferences through information would play a small role compared to price changes. In fact the model seems unable to detect any preference shift – at least in the first two years of the analysis.

Conversely, some minor positive effect is detected in all years for vegetable consumption (and smaller price elasticity is estimated for vegetable demand).

There is little evidence on distributional effects of the policy. If one considers the only year which returns a positive ATT (2005-06), it would seem that impact by income follows opposite directions for fruit and vegetables, where the former shows higher effectiveness for high-incomes and the latter for low-incomes. Again, this is consistent with the model-based price elasticities, which suggest that the price-factor is more relevant for fruit than for vegetables.

Apart from the distribution of the policy effectiveness, data clearly return inequalities in intakes across income groups, where the higher income quartile seems to be quite close to the recommended quantities, *on average*, while the lower quartiles are well-below.

Finally, as shown in Table 4.1 the results of the model-estimated impact evaluation are quite different from those of the sole pre-post comparison of consumption levels. Regardless of the sign, the impact resulting from our simulation seems generally lower with respect to the pre-post difference. One may conclude that market forces accounted for in the estimation procedure mitigate the effect computed by comparing pre-post intervention consumption levels.

Table 4.1 Model-estimated impact of the 5-a-day campaign and pre-post consumption comparison.

	Model-estimated impact evaluation			**Comparison of pre-post outcome levels**		
	2003-04	2004-05	2005-06	2003-04	2004-05	2005-06
Low quartile	-0.30%	-1.21%	1.50%	-4.80%	-4.25%	4.47%
Medium-low quartile	-0.45%	-0.12%	0.67%	-2.21%	-0.01%	2.87%
Medium-high quartile	-1.02%	-1.54%	1.05%	3.41%	-0.65%	7.22%
High quartile	-2.87%	-3.58%	0.97%	-3.44%	-2.40%	1.04%
Total Population	*-1.35%*	*-1.81%*	*1.02%*	*-1.69%*	*-1.78%*	*3.75%*

Some limits of the work have emerged and offer important avenue for further in depth research. First, there is some evidence that the employed models (both before and after the intervention) tend to produce overestimates of actual quantity purchases. This problem might be due to some non-normality in expenditure shares and would need further investigation.

Moreover the problem of zero expenditure and the infrequency of purchases should be analysed more in depth.

Since the distribution issue seems to be crucial in this kind of impact evaluation (as partially shown also by the present work), future investigation should focus on separate model estimation and projection for different groups (e.g. low-income, low- F&V intake, etc.). Behavioural parameters may be very different (for example price elasticity).

With regard to the estimation of quality adjusted prices trough observed unit values, joint estimation of prices and demand system should be experimented. Finally, a different demographic specification of the demand system may be explored (e.g. demographic variables entering income coefficients, and equivalence scales employed for impact measurement).

The basic idea of this thesis is that market forces, price in particular, strongly affect consumers behaviour and information can affect consumption choices only partially, in particular when actors are close to be perfectly informed. It follows that information campaigns like the 5-a-day programme in the UK, even if rarely raise opposition, risk to have very limited impact. Quantitative evaluation of these sorts of public interventions should not avoid accounting for possible contrasting or mitigating factors and the approach shown in this work is meant to offer a possible framework for future impact evaluations.

Appendix

Table A. 1 Descriptive statistics of the variable involved in the demand system, per per-capita income quartile (year 2002-03).

	1st income quartile		2nd income quartile		3rd income quartile		4th income quartile		Total population	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of adults	1.723	0.800	1.816	0.697	1.974	0.806	1.826	0.718	1.836	0.762
Number of children	0.938	1.259	0.617	0.950	0.472	0.812	0.200	0.554	0.546	0.958
Number of components	2.661	1.535	2.433	1.256	2.445	1.196	2.026	0.950	2.382	1.265
Education level of the ref. person	16.028	2.226	15.859	2.026	16.674	2.544	18.562	3.233	16.833	2.794
Age of the ref. person	49.849	18.689	56.020	18.378	52.855	16.526	46.846	14.706	51.285	17.419
Per capita gross current income	68.788	27.702	137.940	20.814	225.771	32.340	485.545	298.628	236.494	223.844
Per capita food expenditure	48.111	31.159	57.356	29.190	72.961	35.915	99.455	54.188	70.350	44.004
Per capita fruit expenditure	2.885	3.015	3.281	2.992	4.151	3.952	5.699	5.190	4.076	4.099
Per capita vegetable expenditure	2.833	2.651	3.217	2.580	3.997	3.320	5.691	4.558	3.991	3.590
Per capita expenditure	91.991	75.071	115.836	61.893	161.614	109.367	249.918	156.129	157.537	124.691

Table A. 2 Descriptive statistics of the variable involved in the demand system, per per-capita income quartile (year 2003-04).

	1st income quartile		2nd income quartile		3rd income quartile		4th income quartile		Total population	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of adults	1.704	0.760	1.837	0.738	1.955	0.776	1.784	0.681	1.820	0.745
Number of children	0.878	1.200	0.635	0.967	0.450	0.783	0.189	0.552	0.537	0.941
Number of components	2.582	1.466	2.472	1.330	2.405	1.161	1.974	0.942	2.357	1.261
Education level of the ref. person	16.106	2.264	15.978	2.211	16.826	2.551	18.470	3.124	16.852	2.752
Age of the ref. person	49.586	18.010	56.214	17.483	51.966	15.946	47.067	14.450	51.194	16.867
Per capita gross current income	73.621	30.761	147.583	20.948	240.421	34.855	519.096	308.808	245.923	231.167
Per capita food expenditure	50.519	30.832	61.221	31.580	74.352	37.716	103.060	55.368	72.379	44.758
Per capita fruit expenditure	2.912	3.024	3.670	3.533	4.298	4.078	5.600	4.975	4.145	4.105
Per capita vegetable expenditure	3.016	3.048	3.488	3.257	4.179	3.518	5.937	4.793	4.165	3.885
Per capita expenditure	97.663	81.123	126.392	96.300	168.753	119.478	258.746	159.254	163.129	132.769

Table A. 3 Descriptive statistics of the variable involved in the demand system, per per-capita income quartile (year 2004-05).

	1st income quartile		2nd income quartile		3rd income quartile		4th income quartile		Total population	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of adults	1.730	0.806	1.869	0.777	1.945	0.780	1.847	0.725	1.849	0.775
Number of children	0.905	1.236	0.629	0.984	0.440	0.799	0.210	0.567	0.538	0.957
Number of components	2.635	1.537	2.498	1.353	2.385	1.160	2.057	0.971	2.387	1.285
Education level of the ref. person	16.663	6.233	16.157	3.425	16.938	3.333	18.612	3.156	17.122	4.300
Age of the ref. person	50.380	18.030	56.085	17.130	52.919	15.972	46.740	13.953	51.453	16.655
Per capita gross current income	79.180	30.905	156.472	22.103	247.487	34.240	526.100	337.732	257.734	244.511
Per capita food expenditure	52.080	31.318	61.462	32.054	75.393	39.004	103.330	58.595	73.690	46.395
Per capita fruit expenditure	3.082	3.505	3.607	3.383	4.266	4.076	5.792	5.454	4.239	4.346
Per capita vegetable expenditure	3.154	3.091	3.560	3.183	4.357	3.538	5.926	4.652	4.291	3.845
Per capita expenditure	101.485	82.573	126.906	81.463	164.121	89.288	253.506	151.173	163.361	121.089

Table A. 4 Descriptive statistics of the variable involved in the demand system, per per-capita income quartile (year 2005-06).

	1st income quartile		2nd income quartile		3rd income quartile		4th income quartile		Total population	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of adults	1.732	0.871	1.782	0.712	1.930	0.777	1.844	0.736	1.823	0.779
Number of children	0.907	1.216	0.594	0.982	0.446	0.803	0.202	0.552	0.529	0.949
Number of components	2.639	1.575	2.376	1.328	2.376	1.179	2.046	0.965	2.353	1.291
Education level of the ref. person	16.663	5.808	16.395	5.411	16.952	2.612	18.734	3.181	17.214	4.526
Age of the ref. person	48.791	17.316	57.447	17.526	52.961	15.993	47.087	14.025	51.530	16.723
Per capita gross current income	78.119	32.335	158.847	22.149	257.593	36.375	551.386	315.858	266.986	243.743
Per capita food expenditure	53.627	32.480	64.786	34.153	76.573	39.815	107.365	68.717	76.225	50.849
Per capita fruit expenditure	3.330	3.397	4.046	3.942	4.548	3.801	6.591	6.097	4.685	4.645
Per capita vegetable expenditure	3.296	2.887	3.627	3.125	4.539	4.453	6.199	4.756	4.454	4.078
Per capita expenditure	106.633	96.681	130.528	80.197	169.427	91.490	272.579	191.982	171.735	140.251

Table A. 5 Estimation results for the price and expenditure equations (year 2002-03)

	Coefficient	P-Value		Coefficient	P-Value
Fruit Price Equation			**Other food price equation**		
Regional dummies			Regional dummies		
East Midlands	-2.280	0.001	East Midlands	-1.603	0.001
Eastern	-2.223	0.001	Eastern	-1.551	0.001
London	-2.176	0.001	London	-1.525	0.001
Merseyside	-2.289	0.001	Merseyside	-1.621	0.001
North East	-2.321	0.001	North East	-1.661	0.001
Northern Ireland	-2.173	0.001	Northern Ireland	-1.564	0.001
North West	-2.269	0.001	North West	-1.610	0.001
Scotland	-2.224	0.001	Scotland	-1.611	0.001
South East	-2.161	0.001	South East	-1.522	0.001
South West	-2.229	0.001	South West	-1.566	0.001
Wales	-2.215	0.001	Wales	-1.667	0.001
West Midlands	-2.238	0.001	West Midlands	-1.641	0.001
Yorkshire and the Humber	-2.274	0.001	Yorkshire and the Humber	-1.645	0.001
Quarters			Quarters		
First Quarter	0.003	0.443	First Quarter	-0.061	0.001
Second Quarter	0.032	0.001	Second Quarter	-0.058	0.001
Third Quarter	-0.015	0.001	Third Quarter	-0.048	0.001
Adj R^2	0.320		Adj R^2	0.393	
Vegetable price equation			**Food Expenditure function**		
Regional dummies			Intercept	784.034	0.004
East Midlands	-2.074	0.001	Gross current Income	2.805	0.001
Eastern	-2.002	0.001	Total expenditure	26.703	0.001
London	-1.879	0.001	Total expenditure (quadratic)	-0.004	0.001
Merseyside	-2.074	0.001	Number of household members	2198.752	0.001
North East	-2.178	0.001			
Northern Ireland	-1.914	0.001	Adj R^2	0.558	
North West	-2.053	0.001			
Scotland	-1.963	0.001			
South East	-1.896	0.001			
South West	-2.033	0.001			
Wales	-2.074	0.001			
West Midlands	-2.070	0.001			
Yorkshire and the Humber	-2.098	0.001			
Quarters					
First Quarter	0.083	0.001			
Second Quarter	0.085	0.001			
Third Quarter	0.005	0.138			
Adj R^2	0.606				

Table A. 6 Estimation results for the price and expenditure equations (year 2003-04).

	Coefficient	P-Value		Coefficient	P-Value
Fruit Price Equation			**Other food price equation**		
Regional dummies			Regional dummies		
East Midlands	-2.253	0.001	East Midlands	-1.615	0.001
Eastern	-2.205	0.001	Eastern	-1.606	0.001
London	-2.112	0.001	London	-1.512	0.001
Merseyside	-2.266	0.001	Merseyside	-1.594	0.001
North East	-2.230	0.001	North East	-1.685	0.001
Northern Ireland	-2.109	0.001	Northern Ireland	-1.555	0.001
North West	-2.236	0.001	North West	-1.588	0.001
Scotland	-2.166	0.001	Scotland	-1.606	0.001
South East	-2.133	0.001	South East	-1.537	0.001
South West	-2.154	0.001	South West	-1.554	0.001
Wales	-2.206	0.001	Wales	-1.620	0.001
West Midlands	-2.204	0.001	West Midlands	-1.627	0.001
Yorkshire and the Humber	-2.235	0.001	Yorkshire and the Humber	-1.596	0.001
Quarters			Quarters		
First Quarter	-0.036	0.001	First Quarter	-0.033	0.001
Second Quarter	0.029	0.001	Second Quarter	-0.051	0.001
Third Quarter	-0.052	0.001	Third Quarter	-0.061	0.001
Adj R^2	0.331		Adj R^2	0.314	
Vegetable price equation			**Food Expenditure function**		
Regional dummies			Intercept	769.737	0.005
East Midlands	-2.032	0.001	Gross current Income	1.410	0.001
Eastern	-1.858	0.001	Total expenditure	31.775	0.001
London	-1.830	0.001	Total expenditure (quadratic)	-0.008	0.001
Merseyside	-2.059	0.001	Number of household members	2200.964	0.001
North East	-2.051	0.001			
Northern Ireland	-1.884	0.001	Adj R^2	0.560	
North West	-1.950	0.001			
Scotland	-1.926	0.001			
South East	-1.816	0.001			
South West	-1.951	0.001			
Wales	-2.041	0.001			
West Midlands	-1.994	0.001			
Yorkshire and the Humber	-2.023	0.001			
Quarters					
First Quarter	0.026	0.001			
Second Quarter	0.115	0.001			
Third Quarter	-0.002	0.698			
Adj R^2	0.573				

Table A. 7 Estimation results for the price and expenditure equations (year 2004-05).

	Coefficient	P-Value		Coefficient	P-Value
Fruit Price Equation			**Other food price equation**		
Regional dummies			Regional dummies		
East Midlands	-2.244	0.001	East Midlands	-1.589	0.001
Eastern	-2.133	0.001	Eastern	-1.534	0.001
London	-2.096	0.001	London	-1.478	0.001
Merseyside	-2.183	0.001	Merseyside	-1.577	0.001
North East	-2.213	0.001	North East	-1.607	0.001
Northern Ireland	-2.154	0.001	Northern Ireland	-1.532	0.001
North West	-2.174	0.001	North West	-1.546	0.001
Scotland	-2.162	0.001	Scotland	-1.611	0.001
South East	-2.100	0.001	South East	-1.501	0.001
South West	-2.145	0.001	South West	-1.519	0.001
Wales	-2.153	0.001	Wales	-1.610	0.001
West Midlands	-2.200	0.001	West Midlands	-1.574	0.001
Yorkshire and the Humber	-2.170	0.001	Yorkshire and the Humber	-1.530	0.001
Quarters			Quarters		
First Quarter	-0.039	0.001	First Quarter	-0.017	0.001
Second Quarter	0.064	0.001	Second Quarter	-0.075	0.001
Third Quarter	-0.043	0.001	Third Quarter	-0.050	0.001
Adj R^2	0.285		Adj R^2	0.374	
Vegetable price equation			**Food Expenditure function**		
Regional dummies			Intercept	768.874	0.013
East Midlands	-2.052	0.001	Gross current Income	1.700	0.001
Eastern	-1.977	0.001	Total expenditure	31.961	0.001
London	-1.850	0.001	Total expenditure (quadratic)	-0.008	0.001
Merseyside	-2.017	0.001	Number of household members	2198.214	0.001
North East	-2.056	0.001			
Northern Ireland	-1.939	0.001	Adj R^2	0.555	
North West	-1.998	0.001			
Scotland	-1.976	0.001			
South East	-1.891	0.001			
South West	-1.967	0.001			
Wales	-2.013	0.001			
West Midlands	-2.020	0.001			
Yorkshire and the Humber	-2.014	0.001			
Quarters					
First Quarter	0.129	0.001			
Second Quarter	0.130	0.001			
Third Quarter	0.004	0.269			
Adj R^2	0.514				

Table A. 8 Estimation results for the price and expenditure equations (year 2005-06).

	Coefficient	P-Value		Coefficient	P-Value
Fruit Price Equation			**Other food price equation**		
Regional dummies			Regional dummies		
East Midlands	-2.215	0.001	East Midlands	-1.571	0.001
Eastern	-2.111	0.001	Eastern	-1.497	0.001
London	-2.051	0.001	London	-1.478	0.001
Merseyside	-2.146	0.001	Merseyside	-1.636	0.001
North East	-2.194	0.001	North East	-1.614	0.001
Northern Ireland	-2.070	0.001	Northern Ireland	-1.452	0.001
North West	-2.140	0.001	North West	-1.598	0.001
Scotland	-2.174	0.001	Scotland	-1.529	0.001
South East	-2.078	0.001	South East	-1.492	0.001
South West	-2.137	0.001	South West	-1.497	0.001
Wales	-2.180	0.001	Wales	-1.617	0.001
West Midlands	-2.161	0.001	West Midlands	-1.591	0.001
Yorkshire and the Humber	-2.164	0.001	Yorkshire and the Humber	-1.555	0.001
Quarters			Quarters		
First Quarter	-0.033	0.001	First Quarter	-0.030	0.001
Second Quarter	0.005	0.183	Second Quarter	-0.030	0.001
Third Quarter	-0.041	0.001	Third Quarter	-0.037	0.001
Adj R^2	0.320		Adj R^2	0.393	
Vegetable price equation			**Food Expenditure function**		
Regional dummies			Intercept	769.452	0.028
East Midlands	-1.959	0.001	Gross current Income	3.443	0.001
Eastern	-1.882	0.001	Total expenditure	28.193	0.001
London	-1.833	0.001	Total expenditure (quadratic)	-0.006	0.001
Merseyside	-1.942	0.001	Number of household members	2198.094	0.001
North East	-2.121	0.001			
Northern Ireland	-1.829	0.001	Adj R^2	0.558	
North West	-1.976	0.001			
Scotland	-1.933	0.001			
South East	-1.837	0.001			
South West	-1.925	0.001			
Wales	-2.026	0.001			
West Midlands	-1.992	0.001			
Yorkshire and the Humber	-2.001	0.001			
Quarters					
First Quarter	0.050	0.001			
Second Quarter	0.147	0.001			
Third Quarter	0.006	0.186			
Adj R^2	0.606				

Table A. 9 Model 1 estimation output (year 2002-03).

	Coefficient	P-value		Coefficient	P-value
Demand system:					
Fruit equation					
Intercept	0.151	0.005	Presence of female members	0.012	0.001
Household size	-0.001	0.382	Household of pensioners	-0.003	0.159
Family type 1	0.002	0.571	Fruit price	0.012	0.108
Family type 3	0.002	0.668	Vegetable price	0.026	0.001
Family type 4	0.005	0.176	Other food price	-0.018	0.008
Gross current income	0.001	0.028	Food expenditure	0.004	0.773
Age of household head	0.001	0.001	Food expenditure (quadratic)	-0.002	0.153
Education of household head	0.004	0.001			
Presence of female members	0.007	0.001	Adj R^2	0.104	
Household of pensioners	-0.009	0.001			
Fruit price	0.013	0.134	**Total food expenditure equation:**		
Vegetable price	0.022	0.001	Intercept	266.353	0.370
Other food price	-0.009	0.309	Gross current income	0.881	0.001
Food expenditure	-0.018	0.242	Total expenditure	18.232	0.001
Food expenditure (quadratic)	0.001	0.705	Number of adults	4387.122	0.001
			Number of children	1680.573	0.001
Adj R^2	0.172		Adj R^2	0.558	
Vegetables equation					
Intercept	0.086	0.073			
Household size	-0.002	0.111			
Family type 1	0.005	0.120			
Family type 3	0.007	0.020			
Family type 4	0.006	0.038			
Gross current income	0.001	0.035			
Age of household head	0.001	0.001			
Education of household head	0.003	0.001			

Table A. 10 Model 2 estimation output (year 2002-03).

	Coefficient	P-value		Coefficient	P-value
Demand system					
Fruit equation					
Intercept	0.181	0.001	Presence of female members	0.012	0.001
Household size	-0.002	0.255	Household of pensioners	-0.003	0.163
Family type 1	0.002	0.633	Fruit price	0.012	0.103
Family type 3	0.001	0.831	Vegetable price	0.026	0.001
Family type 4	0.005	0.228	Other food price	-0.019	0.007
Gross current income	0.001	0.001	Food expenditure	0.004	0.749
Age of household head	0.001	0.001	Food expenditure (quadratic)	-0.002	0.135
Education of household head	0.004	0.001			
Presence of female members	0.007	0.001	Adj R^2	0.103	
Household of pensioners	-0.008	0.001			
Fruit price	0.012	0.176			
Vegetable price	0.022	0.001	**Total food expenditure equation**		
Other food price	-0.007	0.424	Intercept	-445.397	0.135
Food expenditure	-0.027	0.068	Gross current income	2.469	0.001
Food expenditure (quadratic)	0.001	0.800	Total expenditure	24.629	0.001
			Number of adults	3562.101	0.001
Adj R^2	0.171		Number of childre	1525.620	0.001
			Total expenditure (quadratic)	-0.004	0.001
Vegetables equation					
Intercept	0.085	0.072	Adj R^2	0.584	
Household size	-0.002	0.109			
Family type 1	0.005	0.118			
Family type 3	0.007	0.018			
Family type 4	0.006	0.038			
Gross current income	0.001	0.046			
Age of household head	0.001	0.001			
Education of household head	0.003	0.001			

Table A. 11 Model 3 estimation output (year 2002-03).

Demand system	Coefficient	P-value		Coefficient	P-value
Fruit equation			**Vegetable price equation**		
Intercept	0.208	0.002	East Midlands	-2.074	0.001
Household size	-0.001	0.405	Eastern	-2.002	0.001
Family type 1	0.002	0.551	London	-1.879	0.001
Family type 3	0.002	0.635	Merseyside	-2.075	0.001
Family type 4	0.005	0.166	North East	-2.178	0.001
Gross current income	0.001	0.029	Northern Ireland	-1.916	0.001
Age of household head	0.001	0.001	North West	-2.054	0.001
Education of household head	0.004	0.001	Scotland	-1.964	0.001
Presence of female members	0.007	0.001	South East	-1.896	0.001
Household of pensioners	-0.009	0.001	South West	-2.033	0.001
Fruit price	0.001	0.999	Wales	-2.074	0.001
Vegetable price	0.019	0.212	West Midlands	-2.070	0.001
Other food price	0.050	0.001	Yorkshire and the Humber	-2.098	0.001
Food expenditure	-0.015	0.319	1st Quarter	0.084	0.001
Food expenditure (quadratic)	-0.001	0.600	2nd Quarter	0.086	0.001
Adj R^2	0.164		3rd Quarter	0.006	0.091
			Adj R^2	0.405	
Vegetables equation					
Intercept	0.152	0.007	**Other food equation**		
Household size	-0.002	0.114	East Midlands	-1.603	0.001
Family type 1	0.005	0.117	Eastern	-1.551	0.001
Family type 3	0.007	0.018	London	-1.526	0.001
Family type 4	0.007	0.035	Merseyside	-1.622	0.001
Gross current income	0.001	0.038	North East	-1.662	0.001
Age of household head	0.001	0.001	Northern Ireland	-1.568	0.001
Education of household head	0.002	0.001	North West	-1.611	0.001
Presence of female members	0.012	0.001	Scotland	-1.613	0.001
Household of pensioners	-0.003	0.169	South East	-1.523	0.001
Fruit price	0.035	0.173	South West	-1.566	0.001
Vegetable price	0.021	0.095	Wales	-1.668	0.001
Other food price	-0.001	0.963	West Midlands	-1.641	0.001
Food expenditure	0.005	0.716	Yorkshire and the Humber	-1.645	0.001
Food expenditure (quadratic)	-0.002	0.137	1st Quarter	-0.060	0.001
Adj R^2	0.100		2nd Quarter	-0.057	0.001
			3rd Quarter	-0.047	0.001

Total food expenditure equation			Adj R^2	0.405
Intercept	291.268	0.327		
Gross current income	0.896	0.001		
Total expenditure	18.076	0.001		
Number of adults	4397.455	0.001		
Number of childre	1686.582	0.001		
Adj R^2	0.558			
Fruit price equation				
East Midlands	-2.281	0.001		
Eastern	-2.222	0.001		
London	-2.176	0.001		
Merseyside	-2.290	0.001		
North East	-2.322	0.001		
Northern Ireland	-2.175	0.001		
North West	-2.270	0.001		
Scotland	-2.226	0.001		
South East	-2.162	0.001		
South West	-2.229	0.001		
Wales	-2.216	0.001		
West Midlands	-2.238	0.001		
Yorkshire and the Humber	-2.275	0.001		
1st Quarter	0.003	0.317		
2nd Quarter	0.033	0.001		
3rd Quarter	-0.014	0.001		
Adj R^2	0.333			

Table A. 12 Model 4 estimation output (year 2002-03).

Demand system	Coefficient	P-value		Coefficient	P-value
			Vegetable price equation		
Fruit					
Intercept	0.243	0.001	East Midlands	-2.074	0.001
Number of adults	-0.003	0.114	Eastern	-2.002	0.001
Number of children	0.001	0.946	London	-1.879	0.001
Gross current income	0.001	0.040	Merseyside	-2.075	0.001
Age of household head	0.001	0.001	North East	-2.178	0.001
Education of household head	0.004	0.001	Northern Ireland	-1.916	0.001
Presence of female members	0.007	0.001	North West	-2.054	0.001
Household of pensioners	-0.008	0.001	Scotland	-1.964	0.001
Fruit price	-0.001	0.971	South East	-1.896	0.001
Vegetable price	0.020	0.197	South West	-2.033	0.001
Other food price	0.052	0.001	Wales	-2.074	0.001
Food expenditure	-0.026	0.096	West Midlands	-2.070	0.001
Food expenditure (quadratic)	0.001	0.822	Yorkshire and the Humber	-2.098	0.001
Adj R^2	0.164		1st Quarter	0.084	0.001
			2nd Quarter	0.086	0.001
Vegetables			3rd Quarter	0.006	0.092
Intercept	0.146	0.010	Adj R^2	0.560	
Number of adults	0.001	0.957			
Number of children	-0.002	0.004	**Other food equation**		
Gross current income	0.001	0.027	East Midlands	-1.603	0.001
Age of household head	0.001	0.001	Eastern	-1.551	0.001
Education of household head	0.002	0.001	London	-1.526	0.001
Presence of female members	0.012	0.001	Merseyside	-1.622	0.001
Household of pensioners	-0.002	0.198	North East	-1.662	0.001
Fruit price	0.035	0.169	Northern Ireland	-1.567	0.001
Vegetable price	0.021	0.100	North West	-1.611	0.001
Other food price	0.001	0.969	Scotland	-1.613	0.001
Food expenditure	0.008	0.586	South East	-1.523	0.001
Food expenditure (quadratic)	-0.002	0.104	South West	-1.566	0.001
Adj R^2	0.100		Wales	-1.668	0.001
			West Midlands	-1.641	0.001
Total food expenditure equation			Yorkshire and the Humber	-1.645	0.001
Intercept	285.139	0.338	1st Quarter	-0.060	0.001
Gross current income	0.847	0.001	2nd Quarter	-0.057	0.001
Total expenditure	17.633	0.001	3rd Quarter	-0.047	0.001
Number of adults	4488.753	0.001	Adj R^2	0.405	

Number of children	1825.022	0.001
Adj R^2	0.558	
Fruit price equation		
East Midlands	-2.280	0.001
Eastern	-2.222	0.001
London	-2.176	0.001
Merseyside	-2.290	0.001
North East	-2.322	0.001
Northern Ireland	-2.175	0.001
North West	-2.270	0.001
Scotland	-2.226	0.001
South East	-2.162	0.001
South West	-2.229	0.001
Wales	-2.216	0.001
West Midlands	-2.238	0.001
Yorkshire and the Humber	-2.275	0.001
1st Quarter	0.003	0.317
2nd Quarter	0.033	0.001
3rd Quarter	-0.014	0.001
Adj R^2	0.333	

Table A. 13 Model 5 estimation output (year 2002-03).

Demand system	**Coefficient**	**P-value**	**Vegetable price equation**	**Coefficient**	**P-value**
Fruit			East Midlands	-2.074	0.001
Intercept	0.268	0.001	Eastern	-2.002	0.001
Number of adults	-0.003	0.044	London	-1.879	0.001
Number of children	0.001	0.709	Merseyside	-2.075	0.001
Gross current income	0.001	0.001	North East	-2.178	0.001
Age of household head	0.001	0.001	Northern Ireland	-1.916	0.001
Education of household head	0.004	0.001	North West	-2.054	0.001
Presence of female members	0.007	0.001	Scotland	-1.964	0.001
Household of pensioners	-0.008	0.001	South East	-1.896	0.001
Fruit price	-0.002	0.938	South West	-2.033	0.001
Vegetable price	0.020	0.180	Wales	-2.074	0.001
Other food price	0.052	0.001	West Midlands	-2.070	0.001
Food expenditure	-0.034	0.019	Yorkshire and the Humber	-2.098	0.001
Food expenditure (quadratic)	0.001	0.417	1st Quarter	0.083	0.001
Adj R^2	0.163		2nd Quarter	0.086	0.001
			3rd Quarter	0.006	0.100
Vegetables			Adj R^2	0.560	
Intercept	0.144	0.008			
Number of adults	0.001	0.874	**Other food price**		
Number of children	-0.002	0.004	East Midlands	-1.603	0.001
Gross current income	0.001	0.040	Eastern	-1.550	0.001
Age of household head	0.001	0.001	London	-1.526	0.001
Education of household head	0.002	0.001	Merseyside	-1.622	0.001
Presence of female members	0.012	0.001	North East	-1.662	0.001
Household of pensioners	-0.002	0.200	Northern Ireland	-1.567	0.001
Fruit price	0.035	0.164	North West	-1.611	0.001
Vegetable price	0.021	0.102	Scotland	-1.613	0.001
Other food price	-0.001	0.924	South East	-1.522	0.001
Food expenditure	0.008	0.539	South West	-1.566	0.001
Food expenditure (quadratic)	-0.002	0.072	Wales	-1.668	0.001
Adj R^2	0.100		West Midlands	-1.641	0.001
			Yorkshire and the Humber	-1.645	0.001
Total food expenditure equation			1st Quarter	-0.060	0.001
Intercept	900.401	0.001	2nd Quarter	-0.057	0.001
Gross current income	2.849	0.001	3rd Quarter	-0.047	0.001
Total expenditure	26.891	0.001	Adj R^2	0.405	
Total expenditure (quadratic)	-0.004	0.001			

Household size	2121.875	0.001
Adj R^2	0.573	
Fruit price equation		
East Midlands	-2.280	0.001
Eastern	-2.222	0.001
London	-2.176	0.001
Merseyside	-2.290	0.001
North East	-2.322	0.001
Northern Ireland	-2.175	0.001
North West	-2.270	0.001
Scotland	-2.226	0.001
South East	-2.162	0.001
South West	-2.229	0.001
Wales	-2.216	0.001
West Midlands	-2.238	0.001
Yorkshire and the Humber	-2.275	0.001
1^{st} Quarter	0.003	0.342
2^{nd} Quarter	0.033	0.001
3^{rd} Quarter	-0.015	0.001
Adj R^2	0.333	

Reference List

Andreasen, A. R. (1995). *Changing Behaviour to Promote Health, Social Development, and the Environment*. Jossey-Bass Publishers.

Aresu, M., Becares, L., Brage, S., Chaudhury, M., Doyle-Francis, M., Esliger, D., Fuller, E., Gunning, N., Hall, J., Hirani, V., Jotangia, D., Mindell, J., Moody, A., Ogunbadejo, T., Pickup, D., Reilly, N., Robinson, C., Roth, M., and Wardle, H. (2009). "Health Survey for England 2008. Physical activity and fitness". The NHS Information Centre for Health and Social Care.

Banks, J., Blundell, R., and Lewbel, A. (1997). "Quadratic Engel curves and consumer demand". *Review of Economics and Statistics* 79(4):527-539.

Blundell, R. (2000). "Evaluation methods for non-experimental data". *Fiscal Studies* 21(4):427-468.

Blundell, R., Pashardes, P., and Weber, G. (1993). "What do we learn about consumer demand patterns from micro data?". *American Economic Review* 83(3):570-597.

Blundell, R. and Robin, J. M. (1999). "Estimation in large and disaggregated demand systems: An estimator for conditionally linear systems". *Journal of Applied Econometrics* 14(3):209-232.

Bremner, P., Dalziel D., and Evans, L. (2006). "Evaluation of the 5 a day programme. Final report.". Evaluation & Research Team. Big Lottery Fund.

Burgon, C. (2007). "Introduction to the Expenditure and Food Survey". *Nutrition Bulletin* 32(3):283-286.

Buse, A. and Chan, W. H. (2000). "Invariance, price indices and estimation in almost ideal demand systems". *Empirical Economics* 25(3):519-539.

Caliendo, M. and Hujer, R. (2006). "The microeconometric estimation of treatment effects. An overview". *Allgemeines Statistisches Archiv* 90(1):199-215.

Chalfant, J. A. (1987). "A globally flexible, almost ideal demand system". *Journal of Business & Economic Statistics* 5(2):233-242.

Chiang, A. C. (1974). *Fundamental methods of mathematical economics*. New York: McGraw-Hill.

Cox, T. L. and Wohlgenant, M. K. (1986). "Prices and quality effects in cross-sectional demand analysis". *American Journal of Agricultural Economics* 68(4):908.

Cramer, J. S. (1973). "Interaction of income and price in consumer demand". *International Economic Review* 14(2):351-363.

Crawford, I., Laisney, F., and Preston, I. (2003). "Estimation of household demand systems with theoretically compatible Engel curves and unit value specifications". *Journal of Econometrics* 114(2):221-241.

Deaton, A. and Muellbauer, J. (1980a). "An Almost deal Demand System". *American Economic Review* 70(3):312-326.

Deaton, A. and Muellbauer, J. (1980b). *Economics and consumer behavior*. Cambridge Univ Pr.

DEFRA. (2007). "Family food in 2005-06". "A National Statistics Publication".

DEFRA. (2008). "Family food in 2006". "A National Statistics Publication".

Dhar, T., Chavas, J. P., and Gould, B. W. (2003). "An empirical assessment of endogeneity issues in demand analysis for differentiated products". *American Journal of Agricultural Economics* 85(3):605-617.

Dhar, T. and Foltz, J. D. (2005). "Milk by any other name ... Consumer benefits from labeled milk". *American Journal of Agricultural Economics* 87(1):214-228.

Durham, C. and Eales, J. (2006). "Demand Elasticities for Fresh Fruit at the Retail Level". 2006 Annual meeting, July 23-26, Long Beach, CA . American Agricultural Economics Association.

Edgerton, D. L. (1997). "Weak separability and the estimation of elasticities in multistage demand systems". *American Journal of Agricultural Economics* 79(1):62.

Gardini, A., Cavaliere, G., Costa, M., Fanelli, L., and Paruolo, P. (2000). *Econometria. Volume primo.* Franco Angeli.

Green, R. and Alston, J. M. (1990). "Elasticities in Aids Models". *American Journal of Agricultural Economics* 72(2):442-445.

Greene, W. H. (2008). *Econometric analysis*. Prentice Hall.

Keuzenkamp, H. A. and Barten, A. P. (1995). "Rejection without falsification: on the history of testing the homogeneity condition in the theory of consumer demand". *Journal of Econometrics* 67(1):103-127.

Lazaridis, P. (2003). "Household meat demand in Greece: a demand systems approach using microdata". *Agribusiness* 19(1):43-59.

Leser, C. E. V. (1963). "Forms of Engel functions". *Econometrica* 31(4):694-703.

Lewbel, A. (1985). "A unified approach to incorporating demographic or other effects into demand systems". *The Review of Economic Studies* 52(1):1-18.

Martini, A. (1997). "Valutazione dell'efficacia di interventi pubblici contro la povertà: questioni di metodo e studi di caso". *Collana della Commissione di Indagine sulla povertà e l'Emarginazione Sociale, Roma, Presidenza del Consiglio dei Ministri*.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic theory*. Oxford university press New York.

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

Moro, D. (2004). *Analisi della domanda. Teoria e metodi.* Milano: Franco Angeli.

Moro, D. and Sckokai, P. (2000). "Heterogeneous preferences in household food consumption in Italy". *European Review of Agricultural Economics* 27(3):305-323.

Moschini, G. and Rizzi, P. L. (1998). "La struttura dei consumi delle famiglie in Italia". *Staff General Research Papers*.

Muellbauer, J. (1976). "Community preferences and the representative consumer". *Econometrica* 44(5):979-999.

Naska, A., Vasdekis, V. G. S., Trichopoulou, A., Friel, S., Leonhauser, I. U., Moreiras, O., Nelson, M., Remaut, A. M., Schmitt, A., Sekula, W., Trygg, K. U., and Zajkas, G. (2000). "Fruit and vegetable availability among ten European countries: how does it compare with the 'five-a-day' recommendation?". *British Journal of Nutrition* 84(4):549-556.

Nayga, R. M., Tepper, B. J., and Rosenzweig, L. (1999). "Assessing the importance of health and nutrition related factors on food demand: a variable preference investigation". *Applied Economics* 31(12):1541-1549.

Piggott, N. E. and Marsh, T. L. (2004). "Does food safety information impact US meat demand?". *American Journal of Agricultural Economics* 86(1):154-174.

Pollak, R. A. and Wales, T. J. (1981). "Demographic variables in demand analysis". *Econometrica: Journal of the Econometric Society*:1533-1551.

Pollard, C. M., Miller, M. R., Daly, A. M., Crouchley, K. E., O'Donoghue, K. J., Lang, A. J., and Binns, C. W. (2008). "Increasing fruit and vegetable consumption: success of the Western Australian Go for 2&5 campaign". *Public health nutrition* 11(3):314-320.

Pomerleau, J., Lock, K., Knai, C., and Mckee, M. (2005). "Interventions designed to increase adult fruit and vegetable intake can be effective: A systematic review of the literature.". *Journal of Nutrition* 135(10):2486-2495.

Rindfleisch, A. and Crockett, D. X. (1999). "Cigarette smoking and perceived risk: A multidimensional investigation.". *Journal of Public Policy & Marketing* 18(2):159-171.

Sorensen, G., Morris, D. M., Hunt, M. K., Hebert, J. R., Harris, D. R., Stoddard, A., and Ockene, J. K. (1992). "Work-Site Nutrition Intervention and Employees Dietary Habits - the Treatwell Program 1". *American Journal of Public Health* 82(6):877-880.

Stables, G. J., Subar, A. F., Patterson, B. H., Dodd, K., Heimendinger, J., Van Duyn, M. A. S., and Nebeling, L. (2002). "Changes in vegetable and fruit consumption and awareness among US adults: results of the 1991 and 1997 5 A Day for Better Health Program surveys". *Journal of the American Dietetic Association* 102(6):809-817.

WHO. (2003). "Diet, nutrition and the prevention of chronic disease.". 916. World Health Organization: Geneva. "WHO Technical Report Series".

Working, H. (1943). "Statistical laws of family expenditure". *Journal of the American Statistical Association* 38(221):43-56.