Laboratory of Economics and Management
Sant'Anna School of Advanced Studies

Via Carducci, 40 - I-56127 PISA (Italy)
Tel. +39-050-883-341 Fax +39-050-883-344
Email: lem@sssup.it Web Page: http://lem.sssup.it

LEM
Working Paper Series

Engel Curves Specification in an Artificial Model of Consumption Dynamics with Socially Evolving Preferences

Giorgio Fagiolo[*]

[*] *St.Anna School of Advanced Studies, Pisa, Italy*

2001/16 **November 2001**

ISSN (online) 2284-0400

Engel Curves Specification in an Artificial Model of Consumption Dynamics with Socially Evolving Preferences

Giorgio Fagiolo

Sant'Anna School of Advanced Studies
Laboratory of Economics and Management
Pisa, Italy

This Version, November 2001

Abstract

In studying individual consumption behavior, an important issue is the analysis of the relation between commodity expenditure and income (or total expenditure). In this paper we firstly review the more recent theoretical and empirical literature attempting to: (i) derive theory-consistent demand systems models which are able to account for empirically observed non-linearities in total expenditure; (ii) find out whether there exist necessary and sufficient conditions on the across-households distributions such that empirically obtained demand functions still preserve a strong consistency between micro and macro parameters (e.g. consumption-income elasticities). We then apply the techniques discussed in the first part of the paper to the data generated by a computer-simulated model of consumption dynamics presented in Aversi et al. (1999). We find that the model, under a large range of parametrizations, is pretty well equipped to replicate most of the stylized facts displayed by empirically observed consumption patterns, both cross-section and across time.

Keywords: Household Behavior, Consumption, Demand, Engel Curves, Socially Evolving Preferences, Dynamics.

JEL Classification: C31, C32, D10, D12

1 Introduction

In studying individual consumption behavior, one of the main goal is to analyze the relationships between commodity expenditure and income or total expenditure (i.e. the well-known Engel curves)[1]. This broad area of research has recently displayed a strong interest in two related issues.

First, a lot of work has been done in deriving theory-consistent demand-systems models that are able to account for recently collected empirical evidence on the shape of Engel curves. More specifically, one is looking for systems of commodity *expenditure* (or *demand*) *functions* that (i) are coherent with a standard utility-based framework, i.e. are generated – and constrained – by households behaving as maximizers of some utility functions under budget constraints; and (ii) are able to account for recent empirical evidence – stemming for expenditure data-driven investigations – showing that standard linear logarithmic expenditure-share models[2] are robust in describing the observed behavior for certain classes of goods (such as food), but they should be generalized for other goods (such as alcohol and clothing) so as to allow for non-linearities in total expenditure.

Second, from a more data-driven point of view, many efforts have been made in: (i) developing *statistical demand functions* for (homogeneous groups of) commodities, e.g. relating the (nominal or real) expenditure of consumers or households for a given commodity (or homogeneous groups of them) to commodity prices and individual-specific variables as total expenditure or income, household size, etc.; and in (ii) finding out necessary and sufficient conditions on the across-households distributions of the relevant economic variables so that individual parameters are consistent with aggregate parameters. For instance, one of the most interesting questions which this kind of studies has attempted to answer has been the following: Are estimates of *households' income elasticity* based on cross-section micro-data consistent with the estimates of *global income elasticity* based on time-series macro-data ?

In these notes we will try to give a brief (and by no means exhaustive) description of the main findings of these two related streams of research. In Section 2 and 3, after having set out a common notation, we shall review some of the most interesting problems arising in assessing the shape of Engel curves and in accounting for aggregation problems. Next, in Section 4, we will discuss some

---

[1] Generally speaking, Engel curves studies take also into account household composition, while prices, consistently with the specification of Marshallian demand functions, are treated as fixed, which is appropriate if one employs budget survey of one year only. However, if panel data for a sufficient number of years are available, then the focus of the analysis can be shifted to price effects as well, cf. Bierens and Pott-Buter (1990), Banks et al. (1997).

[2] E.g. the Almost Ideal (AI) model of Deaton and Muellbauer (1980).

applications to the model presented in Aversi et al. (1999). Finally, in Section 5, we will draw some conclusions.

1.1 Notation

In order to illustrate in more details both empirical and theoretical results mentioned above, let us start by defining a common notation.

Say we are given a time-series of repeated (cross-sectional) panel data

$$\{\underline{x}_{h,t}\,,\ h=1,\dots,H_t;\ \underline{z}_t\ \}_{t\in T}, \qquad T\in \mathrm{IN},$$

where:

- h labels households (or individuals);
- $\underline{x}_{h,t}$ is a vector of households-specific variables, containing, e.g.,
  - $q_{h,t}^g$ = purchased quantity of goods g=1,…,G by household h in period t;
  - $c_{h,t}^g$ = nominal expenditure in goods g=1,…,G by household h in period t;
  - $y_{h,t}$ = nominal income – or total nominal expenditure[3] – of household h in period t;
  - $s_{h,t}$ = the size of household h in the period t;
- $\underline{z}_t$ is a vector of aggregate variables as:
  - current commodity prices, i.e. $\mathbf{p}_t = (p_t^1, p_t^2, \dots, p_t^G)$;
  - the current overall price-index, generally indicated as $a(\mathbf{p}_t)$;
  - the total (or mean) nominal income (or expenditure) Y_t, obtained averaging across households h= 1, 2, …, H_t.

Furthermore, let $w_{h,t}^g = c_{h,t}^g / y_{h,t}$ be the share of h's total expenditure going to good g, $q_{h,t}^g = c_{h,t}^g / a(\mathbf{p}_t)$ the household h's real consumption of good g, $m_{h,t} = y_{h,t} / a(\mathbf{p}_t)$ the real income – or total expenditure – of household h and M_t the average real income or total expenditure. Finally, assume that household h has an indirect utility function given by $V(y_{h,t}, \mathbf{p}_t)$.

[3] On the issue of using deflated income versus total expenditure in Engel-curve analyses cf. Bierens and Pott-Buter (1990), section 3.

2 Linear vs. Nonlinear Engel Curves

Since the seminal work of Engel (1895), the study of how commodity expenditure is affected by income, prices and other relevant economic variable (both household specific or not) has been characterized by the trade-off between *data-driven* and *theory-driven* approaches[4].

Indeed, as far as the problems of specification of the functional form and estimation of systems of individual demand (or expenditure) functions are concerned, one can single out two broad methodological strategies.

At one extreme, a totally empirical approach would imply fitting statistical models to cross-section or time-series data and finding the 'preferred' ones according to a battery of econometric tests involving functional form mispecification, normality, heteroscedasticity, etc. . By doing this, one can either assume an *a priori* functional form (parametric estimates) or not (non-parametric estimates). Moreover, one might also employ a different functional form for each commodity expenditure equation. In any case, very few restrictions are needed *ex ante*, so that, provided that the model does not display any evidence for mispecification and allows for meaningful testing, it is possible to check *ex post* the plausibility of any theory of consumer behavior by performing appropriate econometric test.

On the other hand, a *theory-driven* approach prescribes that the model employed in the estimation of separate (or systems of) commodity demand functions should be consistent, generally speaking, to some theory of household expenditure behavior. Specifically, as long as this theory is the standard *utility-based model of rational choice*, one requires that (i) the functional form of demand equations to be estimated is generated by constrained maximization of a well-defined utility function; (ii) the unknown parameters involved in the estimation satisfy all induced restrictions[5].

For instance, earlier attempts to estimate *separately* individual demand functions $q_{h,t}^{g} = f_{h,t}^{g}(y_{h,t}; \mathbf{p}_t)$ widely relied upon the desire of identify in a direct way income and price elasticities. As a consequence, many scholars have tried to fit cross-section data by standard OLS with the simple intuitive econometric model:

---

[4] For complete reviews of the issues involved in demand analysis, see the surveys by Deaton (1986) and Blundell (1988).

[5] The assumptions underlying rational choice of the utility-based model imply, firstly, that the commodity demand functions system must satisfy the *adding-up* property, i.e. as consumers maximize over a linear budget constraint, budget shares $w_{h,t}^{g}$ must add up to one. Secondly, each demand function must be *homogeneous* (absence of money illusion) i.e. $f_{h,t}^{g} (\theta\ y_{h,t}; \theta\ \mathbf{p}_t) = f_{h,t}^{g} (y_{h,t}; \mathbf{p}_t)$, any $\theta \in$ IR. Thirdly, there must be *symmetry* in cross-price derivatives of Hicksian demands. Finally, the Slutsky matrix must be *negative semi-definite*.

$$log\ q_{h,t}^{g} = \alpha_t^{g} + \beta_t^{0}\ log\ y_{h,t} + \sum_{k=1}^{G} \gamma_k^{g} \log p_{k,t} + \varepsilon_t^{g}\ . \qquad (1)$$

However, the correspondent deterministic specification of the log-linear functional form (1) – as well as many other empirical functional forms for Engel curves (see e.g. Prais and Houthakker (1955)) – does not completely satisfy the restrictions stemming from standard utility-based theory, as *adding up* is never satisfied, while others (e.g. *homogeneity*) may only be checked *ex-post* or imposed *a priori*.

It should be noted, in addition, that as long as demand functions for different commodities are estimated separately, only homogeneity really matters[6]. In this vein, Stone (1954a) employs an equivalent version of (1), enforcing homogeneity by appropriately restricting the coefficients to be estimated. On the other hand, while one attempts to model the complete demand system, *adding up*, *symmetry* and *negative semi-definiteness* of the Slutsky matrix become crucial as well in order to have a theory-coherent setup. These three restrictions can be either enforced algebraically (see the Linear Expenditure System developed by Stone (1954b)) or statistically (cf. the Rotterdam Model by Theil (1965) and Barten (1977)[7]), leading to theory-consistent models employing log-linear versions of (1). Moreover, one might take up a relatively milder approach – half a way between a theory-driven and a data-driven approach – by using 'flexible functional forms'[8], i.e. by approximate utility or cost functions by functional forms general enough 'to be regarded as a reasonable proxy for whatever the true unknown function may be' (Deaton and Muellbauer, 1980, p.74). In any cases, the estimation of both log-linear and trans-log demand systems lead to strong rejections of theory-induced restrictions, suggesting, at the very least, a strong incoherence between the functional forms employed in estimation and the prescriptions of the theory[9].

Another data-driven model that has been extensively used in a cross-section setting is the so-called Working-Leser share expenditure system (cf. Working (1943) and Leser (1963)), relating budget shares to the log of households' total expenditure:

$$w_{h,t}^{g} = \alpha_t^{g} + \beta_t^{g}\ log\ (y_{h,t})\ , \quad g = 1, \ldots , G\ , \qquad (2)$$

---

[6] Moreover, modeling expenditure functions equation by equation allows one to vary the functional form, as well as the explanatory variables, taking up this way a more data-driven approach.
[7] An interesting feature of the Rotterdam model is that theoretical restrictions apply directly to the parameters of the model (see Deaton and Muellbauer (1980)).
[8] See e.g. the trans-logarithmic model of Christensen et al. (1975). Cf. also Diewert (1971) and Barnett (1983).
[9] Cf. for instance Deaton (1986) and Deaton and Muellbauer (1980).

where α_t^g and β_t^g can in principle be functions of prices. This specification, unlike (1), can be seen to easily satisfy *adding up* provided that $\Sigma_g\alpha_t^g = 1$ and $\Sigma_g\beta_t^g = 0$. Furthermore, Muellbauer (1976) shows that a sufficient condition on individual preferences so as to yield expenditure shares $w_{h,t}^g$ as in (2) is that the log of the cost function is linear in utility, that is:

$$\log c_h (u, \mathbf{p}_t) = \log h(\mathbf{p}_t) + u \cdot \log k(\mathbf{p}_t) , \qquad (3)$$

for some functions $h(\mathbf{p}_t)$ and $k(\mathbf{p}_t)$, or, equivalently, that indirect utility is linear in $\log(y_{h,t})$. Preferences displaying this specification has been called Price-Independent, Generalized Logarithmic (PIGLOG).

However, in order to employ the Working-Leser form over repeated cross-section surveys (e.g. in assessing the impact of indirect tax changes) and/or to use it in time-series analyses, one has to give an equivalent formulation of (2) which takes explicitly into account the effects of relative prices and real expenditure. This has been done by Deaton and Muellbauer (1980)[10], who develop a simple but widely employed econometrically testable model based on PIGLOG preferences, the so-called Almost-Ideal (AI) demand system. They obtain an expenditure shares system linearly relating the share of total expenditure in good g by household h at time t ($w_{h,t}^g$ [11]) to the log of total real expenditure – or real income – of household h at time t (m_t) and the logs of commodity prices, i.e.

$$w_h^g = \alpha_g + \sum_{k=1}^{G}\gamma_{gk} \log p_k + \beta_g \log \left[\frac{y_h}{a(\mathbf{p})}\right], \qquad g = 1, \dots , G. \qquad (4)$$

where the price-index has the trans-log form:

$$\log a(\mathbf{p}) = \alpha_0 + \Sigma_k \alpha_k \log p_k + \tfrac{1}{2} \Sigma_k \Sigma_l \omega_{kl} (\log p_k) \cdot (\log p_l), \qquad (5)$$

Testable specifications (4) and (5) can be derived from (2) and (3) by choosing $h(\mathbf{p})$ and $k(\mathbf{p})$ to be:

$$\log h(\mathbf{p}) = \alpha_0 + \Sigma_k \alpha_k \log p_k + \tfrac{1}{2} \Sigma_k \Sigma_l \gamma_{kl} (\log p_k) \cdot (\log p_k)$$
$$\log k(\mathbf{p}) = \beta_0 \prod_{l=1}^{G} p_l^{\beta_l} \qquad (6)$$

[10] See also the exactly aggregable Trans-Log model of Jorgenson et al. (1982).

and by letting $\omega_{kl} = ½ (\gamma_{kl} + \gamma_{lk})$. This model has many important features. First, it assumes that expenditure shares are linear in the log of income alone. This implies that non-linear terms cannot play any role, so that Engel curves are monotonic in utility and hence in total expenditures, see Deaton and Muellbauer (1986). Second, provided that $\Sigma_g\alpha_g = 1$ and $\Sigma_g\beta_g = 0$, *adding up* is preserved as in the Working-Leser equations, see (2). Third, both *homogeneity* and *symmetry* of cross-price derivatives of the Hicksian demands are satisfied, as long as $\Sigma_k\ \omega_{kl} = \Sigma_l\ \omega_{kl} = 0$ and $\gamma_{kl} = \gamma_{lk}$. Finally, (4) is very close to be linear so that its statistical counterpart can be easily estimated by standard OLS techniques, unlike previous examples as the Rotterdam model.

Despite its theory-consistency, however, the AI model has displayed evidence of mispecification (e.g. omitted lagged variables). Moreover, and relatedly, homogeneity is almost always strongly rejected.

Furthermore, recent empirical Engel curves studies point out that, albeit linear logarithmic share models provide a robust description for certain classes of goods (e.g. food), further terms in income are required for some expenditure share equations and, in particular, that the square of log income always has a strong statistically significant role (cf. among others Hildenbrand (1994), Bierens and Pott-Buter (1990), Hausman et al. (1995), Blundell et al. (1993), Härdle and Jerison (1990), Banks et al. (1997)). This body of literature shares the common view that earlier attempts in estimating household expenditure relations have strongly relied on *a priori* functional form for utility or cost functions, often chosen on the basis of theoretical consistency and tractability rather than plausibility. As Hausman et al. (1995) put it, 'economic theory gives almost no general guidance in specification of Engel curves', other than imposing the restrictions on the parameters discussed above. Consequently, they suggest a thoroughly *data-driven* approach, in which not only the model is derived directly from the data, without restricting its functional form in the first place, but the aim is to avoid any mispecifications by using *non-parametric kernel approaches* in estimating expenditure functions (see Härdle (1990)).

The evidence stemming from these empirical contributions is nonetheless mixed. If, on the one hand, Bierens and Pott-Buter (1990) find that data supports the linear Engel model over a wide income range for the 1980 Dutch Budget Survey, Blundell and Ray (1984), on the other, reject that functional form by testing Linear Expenditure systems (Stone (1954a)) against a class of non-linear Engel curves. Moreover, Blundell et al. (1993) show that a simple quadratic extension of the Deaton and Muellbauer's (1980) 'Almost Ideal' model fits the cross-section micro data of the U.K. Family Expenditure Survey (FES) data 1970-84 very well. These findings are confirmed by the results of

---

[11] In the following, we suppress the time subscripts in order to keep the notation as simple as possible. Our expressions are intended to be valid for each time period $t \in T$.

Hausman et al. (1995), who estimate a non-linear regression between household budget shares and log of total expenditure, finding statistically significant coefficients for higher-order terms of $y_{h,t}$. Finally, Banks et al. (1997) exhibit strong evidence pointing out that 'although the linear formulation appears to provide a reasonable approximation for the food share curve, for some groups, in particular alcohol and clothing, distinct non-linear behavior is evident…' (Banks et al., 1997, p.528-9)[12].

As a result, further work has been recently done in order to develop demand systems which are coherent with the latter empirical evidence while displaying consistency with *standard* utility theory. Hence, they aim to move once again away from data-driven models, which, by definition, are not sufficiently supported by *a priori* economic insights (see among others Gorman (1981), Blundell et al. (1993), Banks et al. (1997)).

The starting point of these studies is the class of PIGLOG preferences (Muellbauer (1976)) which generate Engel curves for budget share looking like (2) and lead to the AI demand system defined by eqs. (4), (5) and (6). An obvious parsimonious generalization of this setup involves letting budget shares to be linearly affected by a finite number of smooth functions of log (m_t) (e.g. polynomials). A simple example of this extended class of expenditure shares functions is:

$$w_h^g = \alpha_g + b_{0g}(\mathbf{p}) + \sum_{j=1}^{L} b_{jg}(\mathbf{p})\, g_j(\log m_h), \quad g = 1, \dots, G \tag{7}$$

where $b_{jg}(\cdot)$, j=0,1,…,L are differentiable functions and $g_j(\cdot)$ are polynomials in log (m_h). Engel curves like (7) are called 'exactly nonlinearly aggregable' since, calling $\mu_{h,t} = m_{h,t} / M_t$ the total expenditure share of household h, one has that the $\mu_{h,t}$-weighted sum of budget shares $w_{h,t}^g$ yields an aggregate budget share:

$$w_t^g = \alpha_t^g + b_{0,t}^g(\mathbf{p}_t) + \sum_{j=1}^{L} b_{j,t}^g(\mathbf{p}_t) \sum_{h=1}^{H} \mu_{h,t}\, g_j(\log m_{h,t}), \tag{8}$$

---

[12] Another interesting related result is that obtained by Härdle and Jerison (1990), who study how real Engel curves (REC), relating quantity demanded and real total expenditure, vary over time. They find that: (i) REC for an aggregate commodity do change over time, in that they shifts in the direction opposite to the change in the relative price index for that aggregate commodity; (ii) the shapes of the non-parametrically estimated REC are remarkably stable over time.

so that the aggregate Engel curve has the same coefficients of the individual ones[13]. Moreover, they are sufficiently general to cover more specific cases as, e.g. Working-Leser and Deaton and Muellbauer (1980) expenditure shares functions. Finally, an important and striking result about demand systems (7) has been proved by Gorman (1981). He has shown that, if such equations are to be theory-consistent, then the rank of the G×(L+1) matrix whose generic row is $[\alpha_t^g + b_{0,t}^g(\mathbf{p}_t) : b_{1,t}^g(\mathbf{p}_t) : \ldots : b_{L,t}^g(\mathbf{p}_t)]$ must be no higher than 3. This basically means that 'the quadratic case is as general as we can go' (Deaton (1981, p.3)). In this connection, Hausman et al. (1995) and Banks et al. (1997) find empirical evidence agreeing with the Gorman's rank-3 result. By testing the simple case L=2 and $g_j(\log m_{h,t}) = \log m_{h,t}$, they are able to conclude that there would be little or no gain in adding extra terms to the equation:

$$w_{h,t}^g = \alpha_t^g + b_{0,t}^g(\mathbf{p}_t) + b_{1,t}^g(\mathbf{p}_t) \log m_{h,t} + b_{2,t}^g(\mathbf{p}_t) [\log m_{h,t}]^2 . \qquad (9)$$

More generally, Banks et al. (1997) show that all (exactly nonlinearly aggregable) demand systems as:

$$w_{h,t}^g = \alpha_t^g + b_{0,t}^g(\mathbf{p}_t) + b_{1,t}^g(\mathbf{p}_t) \log m_{h,t} + b_{2,t}^g(\mathbf{p}_t)\, g(m_{h,t}), \qquad (10)$$

that are derived by utility maximization, either have (i) rank less than 3 (i.e. $b_{2,t}^g(\mathbf{p}_t) \propto b_{1,t}^g(\mathbf{p}_t)$) or (ii) they have rank 3 – according to Gorman's result – and they are generated by indirect utility functions of the form:

$$\log V(y_{h,t}, p_t) = \left\{ \left[\frac{\log y_{h,t} - \log a(\mathbf{p_t})}{d(\mathbf{p_t})} \right]^{-1} + \lambda(\mathbf{p_t}) \right\} \qquad (11)$$

where $\lambda(\mathbf{p}_t)$ is a differentiable, homogeneous of degree 0 function[14]. Interestingly enough, Banks et al. (1997) also prove that: (a) all rank-3, exactly aggregable, utility derived demand systems (10) have $g(m_{h,t})=[\log (m_{h,t})]^2$ and (b) no rank-3, exactly aggregable, utility derived demand systems (10) exists that has both $b_{1,t}^g(\mathbf{p}_t)$ and $b_{2,t}^g(\mathbf{p}_t)$ independent of prices. The fact that the rank-3 condition forces $g(\cdot)$ to be $\log^2(\cdot)$, and that (10) can be derived by standard utility maximization, is of a great

[13] Hence, provided that one is able to construct the quantity $\Sigma_h \mu_{h,t} \cdot g_j (\log m_{h,t})$ at each t, the relevant estimates can be recovered by time-series data alone, as long as $g_j (\log m_{h,t})$ do not contain any unknown parameters.

[14] Notice that if $\lambda(\mathbf{p}_t)$ were 0, then the indirect utility function would be that of a PIGLOG demand system. Moreover, when $\lambda(\mathbf{p}_t)=\lambda$, then (4) is observationally equivalent to the PIGLOG class.

importance, as empirical evidence suggest that rank-3 is the case arising in practice for many commodities, while the linear case arises in the remaining situations. Finally, unlike previous results, the fact that quadratic expenditure shares *must* have price-dependent coefficients in order to be theory-consistent, permits goods to be luxuries at some income levels and necessities at others.

By choosing appropriate specifications for $\lambda(\mathbf{p}_t)$ and $d(\mathbf{p}_t)$, Banks et al. (1997) obtain a data-coherent and theory-consistent demand system (QUAIDS) which nest the AI model while simultaneously allowing for quadratic Engel curves. Indeed, given the form of V as in (11), and applying the Roy's identity, the share equations become:

$$\mathrm{w}_{h,t}^{g} = \alpha_t^g + \sum_{k=1}^{G} \gamma_k^g \log p_{k,t} + \beta_g \log m_{h,t} + \frac{\lambda_g}{d(\mathbf{p})} \left[\log m_{h,t}\right]^2 \qquad (12)$$

The estimation of (12)[15] employing the U.K. FES for the period 1970-86 shows that the QUAIDS model fits the data very well. First, no evidence for mispecification is found, suggesting that no further terms but the quadratic ones are really required. Second, one can capture both the linear and the quadratic shape of empirically observed Engel curves[16].

3 Individual vs. Aggregate Statistical Expenditure Functions

Another important issue in empirical studies of consumption concerns the explanation of observed inconsistencies between the cross-section estimates of the parameters underlying micro-equations modeling individual behavior, on the one hand, and the time-series estimates of the parameters characterizing macro-equations modeling aggregate variables, on the other (cf. Deaton (1992)).

The 'ideal' case often envisaged by standard economic theory is that of 'exact aggregation', where one can treat aggregate consumer behavior as it were the outcome of the consumption

---

[15] Given the conditional linearity of (12), an iterative two-stage procedure is employed by Banks et al. (1997), who also use a generalized method of moments in order to cope with endogeneity of expenditure, measurement error and non-normality. Moreover, they also take into account restrictions in parameters stemming from regularity conditions as Slutsky symmetry. Inequality constraints and negative semi-definiteness of the Slutsky matrix are also tested but they are not considered as restrictions in the estimation procedure.

[16] This results also have strong implications for welfare analysis. As Banks et al. (1997) put it, 'studies based on AI or translog preferences will badly specify the distribution of welfare losses by failing to model Engel curvature correctly'. Moreover, as in general both empirical and theoretical findings show that Engel curves are not monotonic in utility for many commodities, one cannot use the expenditure in such goods to measure welfare, as high or low-income households could in principle exhibit the same budget shares.

decisions of a representative individual. However, as we will see below, the required conditions for 'exact aggregation' are very stringent.

In more general terms, one can start by postulating that household h's consumption behavior, say expenditure for a given good in period t, is determined by both individual-specific variables (e.g. income, wealth, household size and other demographic characteristics) and economy-wide ones (e.g. relative prices). Let us rule out for simplicity the cases in which the past (i.e. lagged terms) and/or the future (i.e. expectations) influence today's behaviors. Then period-t household demand can be written as:

$$c_{h,t} = f_{h,t} (\underline{x}_{h,t} , \underline{z}_t ; \theta_t) , \quad h = 1, \dots, H_t \ , \tag{13}$$

where the shape of the expenditure function f is allowed to vary both through time and across households. Exact aggregation is possible if some function g_t exists such that, labeling with C_t and $\underline{X}_t$ the aggregate variable obtained by averaging out household-specific variables, it holds that:

$$C_t = \frac{1}{H_t} \Sigma_h f_{h,t} (\underline{x}_{h,t} , \underline{z}_t ; \theta_t) = g_t (\underline{X}_t , \underline{z}_t ; \theta). \tag{13'}$$

This requirement is so stringent that no general functions g_t will in general exist at all, so that one must restrict himself to special cases as the linear one or the 'Gorman polar form' (see Section 3.2 below for some more details).

The deterministic models (13) can be straightforwardly modified so as to allow for econometric testing and (cross-section) household panel data (where demand is replaced by actual consumption) may be employed to assess the impact that individual specific variables have on household consumption. As a simple example, one might test, at any t, a cross-section linear model such as:

$$c_{h,t} = f_t(\underline{x}_{h,t} ; \theta_t) = \theta_t' \ \underline{x}_{h,t} + \varepsilon_h , \quad h = 1, \dots, H_t \ , \text{ where } \varepsilon_h \sim N (0 , \sigma_\varepsilon^2). \tag{14}$$

On the other hand, one can also postulate that an aggregate relation holds between average (or total) consumption, average household-specific variables and economy-wide ones (say, prices):

$$C_t = G (\underline{X}_t , \underline{z}_t ; \theta), \qquad t = 1, 2, \dots . \tag{15}$$

Consequently, one might test a time-series model based on (15) to see whether (and to what extent) aggregates (or average) of individual-specific variables (e.g. average consumption for a certain good, average household income, average household size, etc.) are related to economy-wide variables (e.g. commodity prices, etc.). A straightforward example is the linear time-series model:

$$C_t = \theta' \underline{X}_t + \beta' \underline{z}_t + \eta_t \qquad t = 1, 2, \ldots \quad \text{, where } \eta_t \sim N\,(0\,, \sigma_\eta^{2}). \qquad (16)$$

In general, as the distributions of household specific variables ($c_{h,t}$; $\underline{x}_{h,t}$) change through time, one cannot expect $f_t \equiv G$ and $\theta_t \equiv \theta$, all t. Hence, if one attempts e.g. to estimate econometrically testable versions of (13) and (15) by imposing an *a priori* functional-form equivalence, mispecifications are likely to arise. Furthermore, even though the models are well-specified, any interpretation of the aggregate parameters as microeconomic parameters would be inappropriate, so that one might not in principle 'pool' cross-sectional and time-series data to develop statistical demand functions for any commodity. As an example, one might not be sure that estimates of θ_t and θ one gets by the cross-section regressions (14) and by the time-series regression (16) are mutually consistent.

To fix the ideas, let us focus on the simple case, widely analyzed in the literature, wherein family food consumption is considered. It is a standard assumption in this case to start with $\underline{x}_{h,t} = \{ y_{h,t} ; s_{h,t} \}$ and $\underline{z}_t = \mathbf{p}_t$. Since Tobin (1950), the principal aim of this body of research has been to develop an aggregate statistical demand function for food employing both (micro) cross-section and (macro) time-series data. In particular, Tobin (1950) gave a set of sufficient conditions under which: (i) aggregation leads to $f_t \equiv G$; (ii) long-run relationships can be estimated from cross-sectional data sets. Hence, under these conditions, macro-parameters (e.g. food consumption/income elasticity estimated on time series aggregates) can be interpreted as micro-economic parameters (e.g. cross-section estimates of household food consumption/income elasticity), as the two sets are theoretically the same.

However, Tobin (1950) simply assumed that these conditions were satisfied. More generally, one would like to have testable conditions under which aggregation leads to mutually consistent (and interpretable) cross-section (i.e. household specific) and time-series (i.e. aggregate) estimates. In the following, we will briefly recall some of the major causes that lead to such aggregation problems. The focus will be on the simple models relating (individual or average) real food consumption, (individual or average) real income, commodity prices and, possibly, (individual or average) household size. Hence, the basic question we shall address will concern sufficient conditions under which household's income elasticity (coming from cross-section estimation) and aggregate income elasticity (coming from time-series regressions) are the same.

3.1 Aggregation and Functional Forms

The first obvious problem arising in comparing cross-section and time-series analyses concerns the purported functional forms. Let us assume the following simple cross-section model relating real food consumption and real total expenditure holds:

$$q_{h,t} = f\,(\,m_{h,t}\,;\,\theta_t\,)\,, \qquad h = 1, \ldots, H_t\ . \tag{17}$$

The question here is whether the same specification relating consumption and income (e.g. linear, log-linear, etc.) will also hold in the time-series aggregate setup, i.e. whether one can write:

$$Q_t = f\,(\,M_t\,;\,\theta\,)\,, \qquad t = 1, 2, \ldots\ . \tag{18}$$

Again, two approaches are available. One the one hand, a data-driven procedure would suggest to find the best model in both cross-section and time-series settings, without imposing any *a priori* functional form (i.e. without assuming *a priori* that the same specification will hold, more generally, both in the household cross-section eq. (13) and in the aggregate time-series regression (15)). Given a well-specified econometric model for a cross-section analysis (i.e. significant variables, lags to be included, linear vs. nonlinear functional form, exogeneity assumptions, etc.), then, one cannot be sure that a similar model is also well-specified in a time-series setting and, accordingly, that e.g. micro and macro income elasticities will be the same. As noted by Deaton and Muellbauer (1980, p.148), '… it is not neither necessarily, nor necessarily desirable, that macroeconomic relations should replicate their microeconomic foundations so that exact aggregation is possible. Indeed, to force them to do so often prevents a satisfactory derivation of market relations at all."

Nevertheless, an alternative approach widely employed in literature is that of exploring the conditions under which a cross-section specification relating real consumption and real income will result in the same functional form in the aggregate. The obvious candidate to satisfy this property is the class of linear functional forms (see Stoker (1982) and (1980)), which however have not proved themselves to be satisfying explanations for empirically observed data (see Section 2).

More generally, any specifications such that in (7) will eventually lead to mutually consistent parameters in cross-section and time-series setups modeling weighted averages of budget shares, provided that the appropriate weighting system is available and one is able to construct the correspondent aggregate (weighted average) variables (cf. Blundell et al. (1993)). The conditions on the underlying individual utility functions so that this is possible (known as 'exact non-linear aggregation') are much weaker than those required for exact linear aggregation. Moreover, when the average representative expenditure level is independent of prices, one gets as a particular case the PIGLOG preferences introduced in Section 2, which, among the other points of strength, allow for straightforward and elegant modeling of demographic effects (see Deaton and Muellbauer (1980)).

On the contrary, simple time-series log-linear models such as:

$$\log Q_t = \alpha + \beta \log M_t \,, \qquad (19)$$

although successfully employed in many studies, do not in general result from aggregating log-linear cross-section household models:

$$\log q_{h,t} = \alpha_t + \beta_t \log m_{h,t} \,, \qquad (20)$$

cf. Lewbel (1992). Furthermore, Stoker (1986) shows that if cross-sectional functional forms have not the log-linear (or linear) specification, then the assumption that the time-series aggregate relationship also have the log-linear (or linear) specification (19) leads to biases due to omitted variables. The aggregate specification has in this case a functional form quite different from the micro one, as it usually contains other explanatory variables accounting for underlying changes in the moments of cross-sectional distributions and for additional lags. Hence, if one attempts to estimates (20) with time-series data, some further conditions are required in order to interpret the OLS estimate for $\hat{\beta}$ as the households (average) income elasticity.

3.2 Aggregation with Log-Linear Models: Mean Scaling

As pointed out by Stoker (1986), the practice to interpret aggregate parameters estimated on time-series regressions as individual-specific parameters estimated in cross-section analyses (i.e. the so-called *representative individual* or *per-capita* modeling approach) could lead to accurate descriptions of economic behavior only if all marginal reactions of individual agents coincide, or, alternatively, if all individuals' decisions are only affected by aggregate variables (i.e. exact linear aggregation holds).

On the one hand, if all individuals have identical marginal propensities to spend on each of the commodities (i.e. if aggregate demand for each good is not affected by changes in the cross-section income distribution which do not change its mean), then one can easily write a relation between average aggregate demand, average income and relative prices (parallel Engel curves). As shown by Gorman (1959) and (1961), this condition can be satisfied if and only if individual cost functions have the 'Gorman polar form'[17]:

$$c_h (u_h , \mathbf{p}_t) = k'(\mathbf{p}_t) + u_h \cdot k''(\mathbf{p}_t) . \qquad (21)$$

However, as Stoker (1986) points out, 'this situation of "equal marginal effects" is […] unrealistic for most macroeconomic problems, since it states that no individual differences will affect marginal decisions, whether such differences affect needs, initial economic positions, or whatever." (Stoker, 1986, p.764)[18]

On the other hand, in presence of social imitation and heterogeneity in the cross-section distributions of individual variables (and/or if the latter change through time), aggregation entails serious problems. At the very least, the estimates of the parameters in the aggregate time-series equations are not easy to interpret in terms of individual-specific ones.

Nevertheless, as far as log-linear functional forms are concerned, some (testable) restrictions on individual behaviors' distributions can be found so that elasticity estimates in the time-series model coincide with their micro cross-section counterparts. Indeed, Lewbel (1992) shows that if changes in the mean of the distribution of agents over time is independent of changes in the relative distribution of agents, then no aggregation errors in log-linear models can arise. This condition, known as 'mean scaling', includes as a special case the Hicks (1936) and Leontief (1936) 'composite commodity theorem' and can be formally stated in our case as follows. Let F ($m_{h,t}$; M_t , ζ) the distribution of real income (or real total expenditure) across agents at time t. Then F is 'mean scaled' if:

$$F (m_{h,t} \mid M_t , \zeta_t) = \frac{1}{M_t} F'(\frac{m_{h,t}}{M_t} \mid \zeta_t) , \qquad (22)$$

[17] Or, alternatively, preferences are 'quasi-homothetic', see Deaton and Muellbauer (1980).
[18] See also Deaton (1986), Section 5, for an appraisal of some further ways of circumvent this problem.

that is if changes in the parameters ζ_t are independent of M_t. This condition is also sufficient for zero-problems in aggregation of more complicated log-linear models such as:

$$log\, q_{h,t} = \beta_t\, log\, m_{h,t} + \sum_{j=2}^{n} \beta_{jt}\, log\, x_{jht} + r_{ht}\,, \tag{23}$$

where x_{jht} are other household-specific variables (e.g. family size and other demographic indicators) and r_{ht} is any combination of additional parameters and variables (including a constant). In this case, provided that R_t defined by:

$$log\, Q_t = \beta_t\, log\, M_t + \sum_{j=2}^{n} \beta_{jt}\, log\, X_{jt} + R_t\,, \tag{24}$$

is independent of X_t over time for any point in the parameter space (i.e. provided that log-linear aggregation exists), then if $F(m_{h,t}, \underline{x}_{h,t} \mid M_t, X_t, \xi_t) = F^*(m_{h,t}, \underline{x}_{h,t} \mid \xi_t)$, all aggregated elasticities coincide with micro elasticities.

3.3 Other Causes of Aggregation Errors with Log-Linear Models

Even though simple log-linear models as in (19) and (20) are imposed *a priori* in micro and macro regressions, and income is 'mean-scaled', there still can be other sources of aggregation errors leading to inconsistent estimates of income elasticity.

In very general terms, one can single out two types of such causes. First, having constrained our cross-section and time-series data to be explained by a log-linear model, various kinds of mispecifications might arise. Second, even though the models are well-specified, our data might display measurement errors which results in discrepancies between micro and macro elasticities (see Izan (1980)), or, alternatively, the micro and macro data are not compatible in that they measure different entities (see Tobin (1950)). In the following sub-sections, we will briefly recall some sources of aggregation errors belonging to the first class.

3.3.1 Non-Linearities

Suppose we find, in accordance with empirical evidence summarized in Section 2, that our preferred cross-section consumption function involves non-linear terms in (the log of real) household income. Then, as emphasized above, the correspondent preferred aggregate model relating, say, average consumption and average income, will only as a particular case maintain the same specification.

More in general, the presence of significant non-linearities in the micro-consumption function will imply by itself inconsistencies between micro and macro estimates of income elasticity, even if the income distribution appears to be 'mean-scaled'.

In this vein, Anderson and Vahid (1997) reanalyze Tobin's study (Tobin (1950)) finding strong evidence for omitted variables in the cross-section log-linear specification. In particular, their preferred model explaining household food consumption involves both the square of the log of income and a multiplicative term such as $(\log m_{h,t})\cdot(\log s_{h,t})$. Hence, they conclude that, at the very least, the 'best' aggregate time-series model would probably contain quadratic and higher order terms in log of income, as changes in the average log of income provides insufficient information about how the distribution of the log of income affects the log of food consumption. Indeed, the preferred long-run per capita and per-household equations are not log-linear, implying that the estimates for aggregate income elasticity strongly differs from the micro one (see also Blundell et al. (1993) for results along the same lines).

The lack of interpretability of a simple log-linear aggregate food consumption equation (or, more in general, of an aggregate equation which has the same specification of the cross-section household food expenditure function), raises the interesting theoretical question of what the implied functional form of the aggregate relationship will look like given a well-defined non-linear micro model. In this connection, it is possible to show that even if one starts from a simple 'mean scaled' distribution for individual variables (say, a Gamma distribution for income) and a simple micro-model involving both the log of income and its square, then the aggregate theoretical food schedule will involve additional non-linear terms and will in general be different from the micro expenditure equation (see Anderson and Vahid (1997)).

3.3.2 Dynamics

Another source of aggregation errors concerns mispecification due to omitted dynamics in the time-series model.

For instance, Tobin (1950), Chetty (1968) and Maddala (1971) estimated, with different procedures, a long-run time-series inverse demand function for food relating the log of price index $a(\mathbf{p}_t)$ to the log of food consumption, the log of aggregate disposable income at time t and at time t–1. However, Izan (1980) found strong evidence for auto-correlated residuals in all the above mentioned analyses, which is often the case when the dynamical structure of the model is mispecified. More generally, a well-specified micro model explaining the cross-section variability in households characteristics cannot be easily carried over in an aggregate time-series setup, as probably the latter would require some lagged variables too. Hence, a better strategy is again a data-driven one as that employed by Anderson and Vahid (1997) who estimated independently micro (cross-section) and aggregate (time-series) models finding the preferred ones and comparing elasticity measures. Anyway, their 'best' time-series specification does not involved lagged variables, but evidence for structural breaks unable the authors to find a constant-parameter, log-linear relationship between aggregate food consumption and aggregate income for the entire sample period considered (i.e. 1941-1972).

3.3.3 Income Related Heteroscedasticity

In Section 3.2 we have seen that 'mean scaling' is a sufficient condition for the deterministic log-linear micro model (23) to yield in the aggregate the log-linear macro model (24) with interpretable parameters, provided that the residual term R_t is independent of the explanatory variables (e.g. income).

Such kind of dependence can however arise in stochastic log-linear models when the moments of the errors of the micro-relationship depend on income. Indeed, assume the simple cross-section regression:

$$\log q_{h,t} = \alpha_t + \beta_t \log m_{h,t} + \varepsilon_{h,t} \ , \tag{25}$$

and suppose that there is income-dependent heteroscedasticity, i.e. that for some function κ, we have $\varepsilon_{h,t} \mid m_{h,t} \sim N(0, \kappa(\log m_{h,t}))$. For example, if κ is itself log-linear, i.e.

$$\kappa(\log m_{h,t}) = \omega_0 + \omega_1 \log m_{h,t} \quad , \omega_1 \neq 0 \ , \tag{26}$$

then the 'true' aggregate elasticity stemming from (25) is given by $\beta_t + \omega_1/2 \neq \beta_t$.

Hence, even though income distribution is mean scaled and the macro relation will still look log-linear, the micro and macro elasticities will differ as long as some income dependence in the moments of the cross-section noises. This source of aggregation problems in estimating income elasticity appear to be more serious than so far suspected, as Anderson and Vahid (1997) find strong evidence for income-related heteroscedasticity in the errors of the cross-section models based on the U.S. budget surveys in 1950, 1960 and 1972, as well as in the 1980 and 1988 Dutch surveys. After having controlled for this kind of heteroscedasticity, however, indication for non-linearities in income and household size led the authors to abandon the simple log-linear model (see also above). As a consequence, when they estimate their preferred cross-section and time-series models (involving nonlinearities in both the logs of income and household size), some interesting evidence arises. First, the data reject the hypothesis that cross-section individual income elasticity keeps constant across time. On the contrary, there is evidence that food consumption/income family elasticity has increased, whereas its aggregate counterpart has declined. Second, the observed departures from the simple log-linear model (in particular non-linearities and income-dependent heteroscedasticity) lead to a strong inconsistency between the estimated individual and aggregate income elasticity, the latter being in general less than the individual one. Finally, it appears that the dispersion of income elasticity across household in a cross-section sample has increased over time.

4 An Applications to the Aversi et al. (1999) Model

The model that presented in Aversi et al. (1999) might be considered as a sort of rudimentary reduced form of a theory of consumption where purchase acts socially co-evolve with preference structures and 'lifestyles', in a process ridden with decision inconsistencies and cognitive dissonance, and in a precarious balance between path-dependent reproduction of habits and exploration of novelties. Let us just sketch out here the basic elements of such reduced form (for a thoroughly discussion cf. Dosi et al. (1999)).

4.1 Social Adaptation, Consumption Innovation and their Collective Outcomes: a Sketch of a Model

The starting point are utterly simple agents whose (lexicographic) preference structure is represented through a modified version of *Genetic Algorithms* (GA).

In essence, a GA is based on the reproduction and modification of information coded on *strings* of finite length. In analogy with DNA coding, think of a sequentially ordered set of elements (genes in the biological interpretation, demanded goods in the Aversi et al. (1999) model). Each element can take two or more alternative forms (or *alleles*): in our model, straightforwardly, it can have two states, 0 or 1, that is the good is either demanded or not by any one consumer. Hence, for example, the string 01010 encodes the fact that the consumer is going to demand only - reading from left to right - the second and the fourth good.

Strings in GA's evolve through two operators, namely *crossover* and *mutation*. *Crossover* entails a recombination over two ‘parent’ strings. For example, given two strings, say,

01010

and

10011

a random draw of an integer K (in this case, $1 \leq K \leq 5$)determine, so to speak, the ‘cutting point’ (say, 3). In this case the recombined strings will consist of the first three alleles of the first one and the last two of the second one,

01011

and, vice-versa, for the second ‘child’:

10010

i.e. the first three of the second ‘parent’ and the last two of the first one.

Mutation involves the change of state of any one random element on the sequence (from 0 to 1 or vice-versa). In our model, *mutation* captures ‘innovative behaviors’ of each consumer (i.e. a new good appears or disappears in the desired consumption basket), while *crossover* applied to the strings of different agents is meant to account for the process of social imitation (… ‘I have come to like something that you like, too…’).

In the standard formulation, strings are in turn selected over time according to their relative ‘fitness’ as revealed by the environmental payoffs that they obtain. This is not so in the model described here. As already mentioned, there is no reason to think that some consumption pattern may be intrinsically ‘better’ than another one, and, in any case, there is no collective mechanism (thank God!!) to check it. Therefore, more technically, our GA’s evolve over a flat selection landscape, solely driven by crossover and mutation. The death process (of strings) in our model is only determined by the (time-lagged) effects of budget constraints (‘… once upon a time, I desired to have a villa at Cap Ferrat, five servants and caviar every day… however, I have now forgotten all that, and I am quite content with my little apartment and meat twice a week…’).

For our purposes, GA's provide a simple (albeit inevitably rough) account of an evolving lexicographic order over the desired commodities, whose structure is indeed a proxy for the 'lifestyle' of the consumer. Needless to say, the model of consumer behaviour that we propose is highly stylized and 'abstract' (although possibly as 'abstract' as the standard utility based model). However, the assumptions that it incorporates are radically different from the latter in that it tries to capture: (a) the social nature of preference formation; (b) the role of individual and collective history; (c) the formation (and change) of consumption habits; and (d) the permanent possibility of innovation. Contrary to the canonical decision model, we assume agents with extremely limited computational capabilities, but with the possibility of 'learning their preferences' through the very process by which they select their consumption patterns.

In our model, each consumer is fully defined by four strings, namely:

(i) the (hierarchically ordered) list of goods actually demanded;

(ii) the corresponding sequence of budget allocations;

(iii) the list of goods, if any, that one *would have* liked to acquire, but were not allowed to buy by the budget constraint;

(iv) the corresponding sequence of 'desired' budget allocations.

We label the latter two strings *frustrated memory* of the consumer (which decays – i.e. is forgotten – exponentially over time) and it captures the potential cognitive dissonance stemming from exploratory behavior and social imitation (both possibly leading to phenomena like '…I wish but I cannot…').

Goods are grouped into *categories*, metaphorically standing for different basic functions, so that – reading on the strings from left to right – one goes from 'basic' to more 'luxury' categories of expenditures. All goods are non-durable and there is no saving in the economy (except for some involuntary saving due to indivisibilities in the consumption basket). New goods arrive randomly throughout the history of the economy and might or might not be adopted by any consumer (via the *mutation* mechanism) and have a random price. Monetary incomes of consumers grow as a random walk with drift.

The history of the economy starts with a population of identically poor consumers with identical tastes. Consumption patterns endogenously diversify through time via income growth and additions of new items to the individual consumption basket.

The model is explored via simulation and the typical iteration runs as follows.

Having acquired the new level of income at time t, any consumer faces four stochastic alternative, namely:

a) leave unchanged the consumption basket in terms of expenditure shares;

b) access to the *frustrated memory* and try to achieve the 'desired' pattern of consumption;
c) change one or few elements of its basket by 'innovating' (i.e. via the GA's *mutation* operator);
d) change (part of) the consumption pattern by imitating another (randomly chosen) consumer belonging to its same income cohort or to a higher one.

If the consumer draws one of the latter three alternative, it might however be unable to undertake the target consumption, due to the budget constraint. In that case, it might try some local adjustment algorithm (basically involving the variation of the desired quantities or, with lower probability, the elimination of some pre-existing commodities in the basket). If also these adaptive adjustments fail to match the budget constraint, the desired but unbought commodities transit into the *frustrated memory*.

Despite its unrealistically simple structure, the model seems potentially able to capture some of the 'stylized facts' discussed above. Moreover, precisely due to its reduced-form nature, it makes for an easier exploration of its statistical properties under different parametrizations.

As already mentioned, the model *endogenously* generates differentiation in individual consumption patterns, and, at the same time, entails processes of social imitation which prevents such diversity from exploding. Although totally uniform initial conditions are assumed, as incomes stochastically grow, both patterns of consumption and 'preferences' evolve in ways that are *path-dependent* and *socially embedded*. Path dependency appears at two levels: first, the individual consumption patterns at any time depend also on the sequence of past 'preferences' and consumption acts; second, they indirectly depend on the whole *collective* history of the latter. Relatedly, the social embeddedness of the dynamics is straightforward, in that tastes and revealed purchasing patterns emerge from collective mechanisms of social imitation, which represent also *ordering mechanism*, possibly accounting for the relative predictability of aggregate patterns over time. Finally, the model allows the persistent exploration of new items of consumption – and, through that, an everlasting evolution of 'lifestyles'.

Given all that, an important 'exercise in plausibility' (although not a rigorous validation of the model itself) is the statistical analysis of the patterns of consumption generated by the model.

4.2 Purposes of the Analyses

The goal of the exercises presented in the next sub-sections is two-fold. First, we will test simple expenditure functions as (2) and (9) in a separate way for each commodity group g=1,2,…,5 – as well as demand systems – to see what kind of non-linearities (if any) are displayed by the computer-simulated data generated by our model. Second, we will look at aggregation problems discussed in

Section 3. In particular, by testing simple log-linear models as (20), we will explore whether (i) the distribution of log of income is mean-scaled; (ii) non-linear terms in the log of income are statistically significant; (iii) income-related heteroscedasticity arises in modeling log of expenditure for food-like commodities[19]. Finally, we will present some evidence about the evolution of individual income elasticities over time.

All exercises presented in the following are very preliminary in that they refer to single simulations and not to a Montecarlo sample of them (for a given parametrization). Then, they should be considered as 'spot' examples of the typical behavior displayed by our model, since additional results (not shown in this report) point out that the cross-section results summarized here for a sub-sample of time periods appear to be robust across time (i.e. in repeated cross-section analyses).

4.3 Cross-Section Analyses on Engel Curves Specifications

Following Banks et al. (1997), we firstly attempted to assess the shape of the Engel curve relationships for different levels of aggregation over commodities. We considered two different setups: (i) goods are aggregated in 5 commodity groups, g=1, 2, …, 5; (ii) goods are aggregated into 2 commodity groups (food, non-food). In each setup, we employed repeated cross-section, simulated data for the sample period set T={100, 150, 200, …, 500}.

We started by both a parametric and non-parametric qualitative description of the Working-Leser model. For every commodity group g, we considered cross-plots of expenditure shares vs. the log of total individual (real) expenditure. Moreover, we fitted standard OLS regressions (lines and polynomials) and non-parametric Kernel regressions (see Fig. 1(a)).

As a general pattern, one is likely to find a low correlation between budget shares and log of income, due to the high dispersion of the clouds of points in the regression space. Despite that, in both setups, inferior (respectively, superior) commodities tend to be negatively (respectively, positively) correlated with log of income, as expected. Strong evidence for nonlinearities is furthermore displayed by non-parametric Kernel regressions (in line with the results summarized in Banks et al. (1997)), suggesting the need for higher order terms in the Engel curve relationships.

---

[19] Recall that, in the Aversi et al. (1999) model, commodity group 1 can be considered as a metaphorical proxy for 'necessities', while higher-order groups contain goods that, being on the right-hand part of the consumption string, are, in probability, 'filled up' after more basic necessities have been satisfied. In the following, we will then interpret commodity group 1 (or aggregated data for groups 1 and 2) as 'food'. Accordingly, commodities groups 2, 3, 4 and 5 (or aggregated data for groups 3, 4 and 5) as 'non-food'.

In order to give quantitative support to the latter results, we performed standard OLS cross-section regressions by testing the two alternative specifications:

$$w_{h,t}^g = \alpha_t^g + b_{0,t}^g(\mathbf{p}_t) + b_{1,t}^g(\mathbf{p}_t) \log m_{h,t} + \varepsilon_{h,t}^g \quad (27)$$

$$w_{h,t}^g = \alpha_t^g + b_{0,t}^g(\mathbf{p}_t) + b_{1,t}^g(\mathbf{p}_t) \log m_{h,t} + b_{2,t}^g(\mathbf{p}_t) [\log m_{h,t}]^2 + \varepsilon_{h,t}^g \quad (28)$$

where $t \in T$, $h=1,\ldots,1000$ and $g=g_1,g_2$ (either $g_1=1$, g_2 obtained by aggregating $g=2,..,5$; or $g_1=1\cup 2$, $g_2=3\cup 4\cup 5$). An example of the typical results is reported in Table 1, where standard OLS estimates for equations (27) and (28) are summarized. Although the R^2s for all the cross-section regressions are very low, both food-like and non-food-like expenditure shares display non-linearities in the log of income. Tests for Autoregressive Conditional (ARCH) and Income Dependent Heteroscedasticity (F-Test, not reported) failed to find any evidence for heteroscedastic residuals. Nevertheless, functional form mispecification arise in all estimated log-linear models: both the equivalent Reset F-test and LM tests – performed to assess whether the variable $[\log m_{h,t}]^2$ has been omitted – strongly reject the null hypothesis. However, once the square of the log of income is introduced in the regression, no mispecifications are reported, even though the R^2s still remain very low. Finally, further nonlinear terms appear to be not significant in explaining budget shares.

The foregoing results – quite in tune with those obtained for empirical data by Banks et al. (1997) – suggest, first, that non-linear terms (especially the square of the log of income) do indeed matter in Engel curve specifications, and, secondly, that the Gorman's Rank 3 assumption should be satisfied by our computer-simulated data. This conjecture is indeed supported by jointly testing a demand system for 4 out of the 5 commodity groups (avoiding singularity of the dependent variables matrix) and employing χ^2 statistics to test non-linear restrictions implied by the determinants of the matrices of estimated parameters (not shown).

4.4 Cross-Section vs. Time-Series Analyses of Income Elasticity: Aggregation Errors

A second set of results in cross-section analyses concerns aggregation problems in the estimation of individual vs. aggregate income elasticity. As discussed at length in Section 3, if we start from the simple log-linear model (25), different sources of aggregation errors can independently result in a lack of interpretation of time-series estimates, which basically become meaningless as far as elasticities (i.e. Engel curves slopes) are concerned.

In order to explore whether the cross-section distributions generated by our model also display ‘aggregation errors’ as the empirically observed data do (see Anderson and Vahid (1997)), we performed three kinds of analyses. In the following, we will report about them separately.

4.4.1 Mean Scaling

To assess whether the distribution of the log of (real) income is mean-scaled or not, we performed two different kind of computations, as suggested by Lewbel (1992). Given a sufficiently long time-period sample T, let s_{qt} be the q-th quantile of the distribution of individual real income Y at time t. Next, define by $\overline{Y}_t$ and $\widetilde{Y}_t$, respectively, the arithmetic and geometric average of the time-t income distribution. Finally, let $\lambda_{qt}= s_{qt} / \overline{Y}_t$ and $\omega_t = \overline{Y}_t / \widetilde{Y}_t$. It is easy to show that the distribution of Y_t is mean-scaled if and only if λ_{qt} and $\overline{Y}_t$ are independent over time for every q. Moreover, the condition that ω_t and $\overline{Y}_t$ are independent over time is necessary for the distribution of Y_t to be mean-scaled.

In our computations, we considered T=10, 20, 30, …, 500 and q=0.05, 0.10, … , 0.95. Then, in order to test the independence over time of the pairs of time-series (λ_{qt} , $\overline{Y}_t$), for every q, and ($\overline{Y}_t$, $\widetilde{Y}_t$), we performed a t-test on the slope of the related linear regression (after having checked for mispecifications). The results reported in Table 2 strongly rejects mean-scaling. This is in line with the evidence reported by Lewbel (1992) about the income distribution in the U.S. Current Population Reports data 1947-83 and allows one to conclude that, even though a log-linear model relating consumption and income is assumed in cross-section regressions, the same specification cannot arise from aggregation.

4.4.2 Non-linearities

As already emphasized in Section 4.2, non-linearities appear to characterize Engel curve specifications also in the data generated by our model.

To further this analysis, we performed some cross-section regressions, estimating by OLS the log-linear specification:

$$log\, C_{h,t}^{g} = \alpha_t^{g} + \beta_t^{g}\, log\, Y_{h,t} + \varepsilon_{h,t}^{g}, \quad \varepsilon_{h,t}^{g} \sim N(0, \sigma^2)\,, \qquad (29)$$

for g=1,2,…,5 and t∈ {200, 250, 300, …, 500}. In Fig. 1(b) we show an example (period t=500) of the cross-plots ($log\ C_{h,t}^g$, $log\ Y_{h,t}$) for each commodity group. The shape of the cross-plots is robust across-time and through different levels of aggregation over commodities. Moreover, in Table 3 we report the results of the comparison of the regression for commodity group 1 (food) and that for all other groups aggregated (non-food). Income elasticities are all significant and the R^2 are very high. However, the widespread, strong, evidence for functional-form mispecifications (cf. the large value of the F test for omitted variables) suggests to include (at least) the square of log of income in the regressions. As to other kind of mispecifications (not reported in Table 3), one often finds evidence for non-normality.

After having introduced the additional explanatory variable $(log\ Y_{h,t})^2$ in the regression (29), RESET tests tend to fail to display functional form mispecifications – see Table 4. This is not completely true for Group 1, suggesting that, after all, the linear specification for food-like commodities Engel curves is not completely wrong (see Banks et al. (1997)).

4.4.3 Income-Dependent Heteroscedasticity

Another source of aggregation problems relies in income-dependent heteroscedasticity of the errors in the cross-section regressions. To see whether it is the case, we considered the regression (29) and we assumed that $\varepsilon_{h,t}^g \mid Y_{h,t} \sim N\ (\ 0\ ,\ \kappa\ (log\ Y_{h,t})\ )$. Then, we ran an auxiliary regression to test whether the specification:

$$\kappa\ (log\ m_{h,t}) = \omega_0 + \omega_1\ log\ Y_{h,t} + \omega_2\ (log\ Y_{h,t})^2 \qquad (30)$$

correctly explains the variance of the errors.

In general, income-related heteroscedasticity is often present both at different levels of aggregation over commodities and across time (see Table 5), but one is more likely to observe it for non-food-like commodities. This basic evidence is confirmed also when one performs the same exercise for each separate commodity groups g=2, 3, 4 and 5.

4.4.4 Evolution of income elasticities over time

As reported by Anderson and Vahid (1997), there appears to be significant evidence that the (average) individual income elasticity for food-like commodities has declined over time.

This empirical finding is robustly confirmed by the data generated by our model under different parametrizations (see Figure 2). Moreover, our results also display a general tendency for increasing intercepts over time. However, this pattern of behavior is not so clear for other (higher) commodity groups.

4.4.5 *Aggregated Expenditure Schedules and Shares Equations: Time Series Analyses*

The foregoing evidence on non-linearities, mean scaling and income-dependent heteroscedasticity suggests two related considerations. First, finding a stable, constant parameter, log-linear relationship between aggregate consumption or budget shares (at whatever level of aggregation over commodities) and aggregate income appear to be a very difficult task, due to the presence of basically all possible sources of aggregation errors in our simulated data (which is however quite in line with empirical findings). Second, even if such a stable time-series model indeed existed, the very interpretation of aggregate parameters in terms of the individual ones would be inappropriate.

Even though any would-be aggregate time-series regularities would not be able to bear any isomorphism with microscopic behavior, we have nevertheless attempted to explore whether significant Engel-type patterns of evolution do indeed appear over time in consumption for different commodity groups by performing two sets of analyses.

First, we estimated time-series models relating the levels of the (log of) consumption for each commodity group to simultaneous and lagged values of: (i) log of (real) total expenditure; (ii) commodity groups' price-indexes[20]; (ii) powers of the log of real total expenditure. Second, we considered as dependent variables budget shares of each commodity group.

In both cases, one has to face two important issues, namely that our data are characterized by (i) endogeneity of consumption and income (i.e. total expenditure approximately equals total income); and (ii) prices and income are exogenous (independent) stochastic processes, while, of course, the series $\log C_t^g$ are not, because consumption choices in our model are taken simultaneously. Therefore, one should model together the series $\{\log C_t^g, g=1, \ldots, 5\}$ and assume – by (ii) above – weak (and strong) exogeneity of both prices and (log of) income. However, because of the identity between total expenditure and total income, one can only model simultaneously up to four consumption series so as to avoid singularity of the matrices involved in the regressions. In the following, we have chosen to model $\log C_t^g$ and w_t^g for g= 1, …, 4 (our main results hold irrespective of that choice).

---

[20] Unlike previously employed notation, in order not to complicate the exposition, we will label by p_t^g the price-index of commodities belonging to group g (not price levels).

The results reported in what follows are an example concerning the formulation, selection and estimation of our preferred econometric models explaining the simulated data. Even though they refer to a single-parametrization, single-simulation setup, they should be notwithstanding read as an example of the general behavior displayed by the model under different parametrizations. Indeed, preliminary Montecarlo time-series analyses performed under diverse parametrizations have shown that the ensuing aggregate evidence is sufficiently robust to be understood as an 'emergent property' of our model (see Lane (1993)).

We first considered an 'as-general-as-possible' VAR model relating consumption series to their lagged values and other explanatory (exogenous, i.e. not explained in the model) variables of the form:

$$\mathbf{X}_t = \sum_{j=1}^{k} \Pi_j \mathbf{X}_{t-j} + \sum_{h=0}^{m} \Gamma_h \mathbf{Y}_{t-h} + \boldsymbol{\varepsilon}_t \qquad (31)$$

where $\mathbf{X}_t' = (log\ C_t^1, log\ C_t^2, log\ C_t^3, log\ C_t^4)$ and $\mathbf{Y}_t'$ which contain a subset of the variables ($log M_t$, $log^2 M_t$, $log^3 M_t$, $\Delta log M_t$, $\Delta log^2 M_t$, $\Delta log^3 M_t$ …, p_t^1, …, p_t^5, $\Delta\ p_t^1$, …, Δp_t^5, …).

As to non-stationarity, standard ADF test (both with constant and/or trend included) largely accept the null of unit-root for $\{log\ C_t^g$, g=1, …, 5}, see Table 6. On the other hand, $log M_t$ and $log Y_t$ are I(1) by assumption. Hence, all the analyses have been carried out by replacing $\mathbf{X}_t$ with $\Delta\mathbf{X}_t$, and employing as regressors $\Delta log M_t$ (or $\Delta log Y_t$).

In the Aversi et al. (1999) model, prices are generated by two alternative data generating processes, namely (i) a stationary stochastic process (i.e. a noise added to a constant mean); (ii) a non-stationary stochastic process (i.e. prices falling along with 'learning curves'). The results we present in this section are examples of the case in which price indexes are I(0).

As we have chosen to explain changes in the logs of nominal consumption with respect to changes in the log of real income, we must consider changes (not levels) in price indexes to avoid that the coefficients of p_t^g and p_{t-1}^g are equal in absolute value but opposite in sign. Indeed, as $Y_t = M_t \cdot a(\mathbf{p}_t)$, then the estimated coefficients of p_{t-1}^g tend to capture also the variation of the price index employed to deflate nominal income, and so to be equal to minus the coefficient of p_t^g, because *corr* $\{a(\mathbf{p}_t),\ p_t^g\} \sim 1$, all g. However, results similar to those presented here can be obtained by estimating a specification relating changes in log of consumption to changes in log of *nominal* income and price-indexes levels.

As a general outcome, testing a VAR specification like (31) allows us to verify the general structure of the data-generating process at an aggregate level. Indeed, our parsimonious-preferred model – given the I(1) specification for the log of consumption series – is a very simple one, relating changes in (log of) consumption levels to changes in price indexes and changes in (log of) real aggregate income, without any additional lagged variable being significant, i.e.:

$$\Delta log\ C_t^g = \alpha + \beta\ \Delta log M_t + \sum_{g'=1}^{5} \gamma^{g'} \Delta p_t^{g'} + \varepsilon_t \qquad (32)$$

The VAR model (see Table 7, (i) for estimation results) is the smaller one displaying no mispecifications and providing a good description of the underlying relationships among variables, as the battery of diagnostics tests (Table 7, (ii)) clearly shows. However, its explanatory power is very limited in that it does not add any interesting knowledge about the model's behavior. We notice, indeed, that the four consumption series are not related one to the other: LM tests for omitted variables do not reject the null that (lagged and simultaneous values of) $\Delta log C_t^{g'}$ significantly affects $\Delta log C_t^{g''}$, for $g' \neq g''$. Therefore, testing a VAR model is in this case equivalent to a single-equation analysis.

On the contrary, large part of the variation of changes in consumption is captured by a quasi unit-root in changes in income. This is not surprising as in our model there is no saving. All changes in consumption are positively correlated with changes in price-indexes. Moreover, a χ^2 test to test whether the sum of the coefficients of price-indexes is equal to 1 does not in general reject the null. The sign and size of price coefficients is probably due to a strong negative correlation between real outcome and the overall price index $a(\mathbf{p}_t)$. Indeed, as $corr\ \{a(\mathbf{p}_t), p_t^g\} \sim 1$, all g, and $Y_t = M_t \cdot a(\mathbf{p}_t)$, then $0 << corr\ \{ logC_t^g , log\ Y_t \} = corr\ \{ logC_t^g , M_t \cdot p_t^g \}$ so that $corr\ \{ logC_t^g , p_t^g \} >> 0$. This appears to be just a consequence of the kind of data employed in the model and does not help us in explaining any emergent properties of the model. Other specifications concerning real vs. nominal figures have been also estimated, but in all cases one cannot exactly separate real from nominal components.

To overcome these difficulties, we estimated single-equations (as well as VAR models) with budget shares as dependent variables. This has been done in order to assess whether long-term changes in budget coefficients emerge – driven by social innovation and imitation, jointly with stochastically (exogenously) growing incomes – despite the extreme (and unrealistic) hypothesis that, when consumers opt for the reproduction over time of their past consumption patterns, they do so in a way that amounts to assuming a homothetic demand with unitary price elasticity.

All five budget shares appear to be stochastically non-stationary in the selected sample simulation period (i.e. t=250, …, 500), as displayed by both Figure 3 (graphically) and standard

ADF tests (Table 8). Therefore, all subsequent analyses have been carried out on budget shares' differences. We estimated – both simultaneously and separately – models of the form:

$$\Delta w_t^g = \alpha_0 + \sum_{i=1}^{k}\sum_{g'=1}^{5}\alpha_i^{g'}\Delta w_{t-i}^{g'} + \sum_{j=0}^{n}\beta_j \Delta \log y_{t-j} + \sum_{h=0}^{m}\gamma_h \Delta p_{t-j} + [\text{other terms}] + \varepsilon_t. \quad (33)$$

As a general result, we get (as before) that in the equation for Δw_t^g neither lagged terms of Δw_t^g nor (contemporaneous and lagged) terms of $\Delta w_t^{g'}$, g'≠g, are statistically significant (i.e. $\alpha_i^{g'}=0$, all i and g'≠g). Therefore, as happens in the previous exercise, one can revert to a single-equation analysis, since both income and price indexes are (weakly and strongly) exogenous for the parameters to be estimated.

Following a 'general to specific' modeling strategy, we can in general select preferred models with no mispecifications displaying Engel-type patterns of evolution in the share of different product groups. This type of dynamics appear more evidently in the case of group 1 and group 5 (see figures on budget shares in the paper but also Fig.3 in this report). More rigorously, in Table 9 two examples of OLS estimates of (34) are reported for commodity groups 1 and 5. Our preferred expenditure schedules are of the form:

$$\Delta w_t^g = \alpha_0 + \sum_{i=1}^{k}\alpha_i^{g}\Delta w_{t-i}^{g} + \sum_{j=0}^{n}\beta_j \Delta \log m_{t-j} + \sum_{h=0}^{m}\gamma_h \Delta p_{t-j} + \varepsilon_t^g , \quad (34)$$

or, employing nominal income and price-indexes levels:

$$\Delta w_t^g = \alpha_0 + \sum_{i=1}^{k}\alpha_i^{g}\Delta w_{t-i}^{g} + \sum_{j=0}^{n}\beta_j \Delta \log y_{t-j} + \sum_{h=0}^{m}\gamma_h p_{t-j} + \varepsilon_t^g , \quad (35)$$

We found significant lagged values for both $\Delta log m_t$ (resp. $\Delta log y_t$) and Δp_t (resp. p_t) very far from time t (even for the lags t–j, j>20), although our extreme assumptions and the stationarity of the price generating process.

What is interesting is the significant effect of lagged incomes, yielding Engel-type patterns which are purely an aggregate *emergent property*, driven by the collective exploration of new consumption opportunities, together with the progressive relaxation of budget constraints.

As to the required dynamic specification, a dynamic analysis of the lag structure generally suggests that the choice of k≅10, n≅20 and m≅10 is the one which optimally trades off the goodness of fit and correct specification. Solving for the static long-run equations allows us to get statistically significant coefficients which have the 'right' expected sign. Moreover, a Wald test for the joint significance of all the variables (excluding the constant) in the long-run solution (see Table 9) strongly rejects the null, suggesting that in the long-run (i.e. when the means of the independent variables have remained at a constant level for long enough and the dependent one have reached its long-run solution) the influences of income and prices on budget shares are like the empirically observed ones.

However, even after the dynamics has reached its static long-run solution, in the short-run there appear to be a sort of cycles in the response of the change of budget shares to the impulses of a change in (the log of) real income and price indexes. This can be clearly seen if we take a look at the plots of the normalized lag weights (see Fig. 4 for an example concerning Group 1), which give the responses of the dependent variable at time t+1, t+2, etc., when one slightly perturbs the level of an explanatory variable at time t.

Tests on other simulation results conducted on the 'version 2' of price dynamics (i.e. price falling along with 'learning curves'), not shown here, show that, while in general Engel-type patterns continue to emerge, prices (both the price index of the group in question and of the others) appear to exert a significantly greater influence of the dynamics of budget shares (up to the fifth lag, and mostly but not always with the expected sign). However, note that, again, this should be considered as an emergent property which does not bear any isomorphism with microscopic behavior: in fact, by construction, individual agents either have unit price elasticities when acting 'business-as-usual' or do not look at all at prices when imitating or innovating – except insofar as prices affect budget constraints. Indeed, what appears in the aggregate as the dynamic influence of prices upon shares rests in fat on the process by which the fall in the former help relaxing budget constraints (a sort of dynamic version of an income effect) and that in turn makes easier innovation, imitation and fulfillment of 'frustrated' options.

Moreover, in empirical time prices, one often detects evidence of important and generalized structural breaks in the patterns of consumption within and across product groups (on the former, cf. Combris (1992) and Anderson and Vahid (1997); more generally on changes in consumption patterns, cf. Houttaker (1957) and (1965), Kuznets (1962), Gardes and Louvet (1986), Deaton and Muellbauer (1980)). Remarkably, notwithstanding our rather rudimentary behavioral assumptions, structural instability -–most often emerges when testing models as (34) and (35). When applying the usual tests for structural stability (Chow, CUSUM, CUSUMSQ), one generally finds (especially with regards to groups 1 and 5) significant structural changes, intertwined by rather long periods of structural stability. At the risk of some over-interpretation, these patterns might suggest the easy emergence of *punctuated discontinuities* in historically shaped, collectively shared, 'models of consumption', which, however, display a 'metastable' character (in the sense that they persist on time scales of orders of magnitude greater than those of the processes which generated them, but nonetheless tend to vanish with probability one as time goes on).

5 Conclusions

In these notes we firstly attempted to provide a brief (and by no means exhaustive) survey of the literature analyzing – both empirically and theoretically – the relationships between commodity expenditure and income (or total expenditure). In particular, we reviewed studies trying to derive theory-consistent demand-systems models that are able to account for recently collected empirical evidence on the shape of Engel curves. Moreover, we discussed efforts made in developing *statistical demand functions* for (homogeneous groups of) commodities and in finding out necessary and sufficient conditions on the across-households distributions of the relevant economic variables so that individual parameters are consistent with aggregate parameters.

Secondly, we employed data generated by the model presented in Aversi et al. (1999) to test whether this formalization is able to capture some well-established stylized facts in consumption behavior singled out by the above discussion. We find that he model generates patterns of diffusion of new commodities displaying the usual S-shape generally observed in empirical studies. Furthrmore, the dynamics of demand for individual commodities (and groups of them) shows the familiar negative elasticity to prices, notwithstanding the absence of any notional 'demand curve' within the decision algorithm of individual consumers. In addition to all that, budget shares of different commodity groups display Engel-type patterns of evolution over time, at *macroscopic* level, even if, again, nothing of that sort is inbuilt with the microscopic description of how agents evolve their consumption patterns over time. Finally, as in empirical time-series, also the statistics generated by our model present evidence of structural breaks in the patterns of consumption within and across product groups.

References

Anderson, H.M. and Vahid, F. (1997), “On the Correspondence Between Individual and Aggregate Food Consumption Functions: Evidence from the U.S.A. and the Netherlands”, *Journal of Applied Econometrics*, **12**: 477-507.

Aversi, R., Dosi, G., Fagiolo, G., Meacci, M. and Olivetti, C. (1999), “Demand Dynamics with Socially Evolving Preferences”, *Industrial and Corporate Change*, **8, 2**: 353-408. Also published as: Interim Report IR-97-081 / Oct, IIASA, Laxenburg (Austria), 1997.

Banks, J., Blundell, R.W. and Lewbel, A. (1997), “Quadratic Engel Curves and Consumer Demand”, *Review of Economics and Statistics*, **79**: 527-539.

Barnett, W.A. (1983), “New Indices of Money Supply and the Flexible Laurent Demand System”, *Journal of Economics and Business Statistics*, **1**: 17-23.

Barten, A.P. (1977), “The Systems of Consumer Demand Functions Approach: A Review”, *Econometrica*, **45**: 23-51.

Bierens, H.J. and Pott-Buter, H.A. (1990), “Specification of Household Engel Curves by Nonparametric Regressions”, *Econometric Reviews*, **9**: 123-184.

Blundell, R.W. (1988), “Consumer Behavior: Theory and Empirical Evidence. A Survey”, *Economic Journal*, **98**: 16-65.

Blundell, R.W. and Ray, R. (1984), “Testing for Linear Engel Curves and Additively Separable Preferences using a New Flexible Demand System”, *Economic Journal*, **94**: 800-811.

Blundell, R.W., Pashardes, P. and Weber, G. (1993), “What do we Learn about Consumer Demand Patterns from Micro Data ?”, *The American Economic Review*, **83**: 570-597.

Chetty, V.K. (1968), “Pooling of Time-Series and Cross-Section Data”, *Econometrica*, **36**: 279-290.

Christensen, L.R., Jorgenson, D.W. and Lau, L.J. (1975), “Transcendental Logarithmic Utility Functions”, *American Economic Review*, **65**: 367-383.

Combris, P. (1992), "Changements Structurels: le cas des consommations alimentaires en France de 1949 à 1988", *Economie et Prevision*, **102**: 221-245.

Deaton, A.S. (1986), “Demand Analysis”, in Griliches, Z. and Intrilligator, M.D. (Eds.), *Handbook of Econometrics, Vol. III*, Amsterdam, Elsevier.

Deaton, A.S. (1992), *Understanding Consumption,* Oxford, Clarendon Press.

Deaton, A.S. (Ed.) (1981), *Essays in the Theory and Measurement of Consumer Behavior in Honour of Sir Richard Stone*, New York, Cambridge University Press.

Deaton, A.S. and Muellbauer, J. (1980), “An Almost Ideal Demand System”, *American Economic Review*, **70**: 312-326.

Diewert, W.E. (1971), “An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function”, *Journal of Political Economy*, **79**: 481-507.

Dosi, G., Fagiolo, G., Aversi, R., Meacci, M. and Olivetti, C. (1999), “Cognitive Processes, Social Adaptation and Innovation in Consumption Patterns: from Stylized Facts to Demand Theory”, in Dow, S.C and Earl, P.E. (Eds.), *Economic Organizations and Economic Knowledge: Essays in Honour of Brian Loasby*, Cheltenham, Edward Elgar.

Gardes, F. and Louvet, Ph. (1986), *La Convergence Internationale des Structures de Consommation,* Paris, CREDOC.

Gorman, W.M. (1981), “Some Engel Curves”, in Deaton, A.S. (Ed.), *Essays in the Theory and*

Measurement of Consumer Behavior in Honour of Sir Richard Stone, New York, Cambridge University Press.

Hardle, W. (1990), *Applied Nonparametric Regression,* Cambridge, Cambridge University Press.

Hardle, W. and Jerison, M. (1990), "Cross Section Engel Curves over Time", CORE, *Discussion Paper*, 9016.

Hausman, J.A., Newey, W.K. and Powell, J.L. (1995), "Nonlinear Errors in Variables Estimation of Some Engel Curves ", *Journal of Econometrics*, **65**: 205-233.

Hicks, J.R. (1936), *Value and Capital,* Oxford, Oxford University Press.

Houthakker, M.S. (1957), "An International Comparison of Household Expenditure Patterns", *Econometrica*, **25**: 532-551.

Houthakker, M.S. (1965), "New Evidence on Demand Elasticities", *Econometrica*, **33**: 277-288.

Izan, H.Y. (1980), "To Pool or Not To Pool ? ", *Journal of Econometrics*, **13**: 391-402.

Jorgenson, D.W., Lau, L.J. and Stoker, T. (1982), "The Trascendental Logarithmic Model of Aggregate Consumer Behavior", in Basmann, R. and Rhodes, G. (Eds.), *Advances in Econometrics, Vol. 1*, Greenwich, JAI Press.

Leontief, W.W. (1936), "Composite Commodities and the Problem of Index Numbers", *Econometrica*, **4**: 39-59.

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

Lewbel, A. (1992), "Aggregation with Log-Linear Models", *Review of Economic Studies*, **59**: 635-642.

Maddala, G.S. (1971), "The Likelihood Approach to Pooling Cross-Section and Time-Series Data", *Econometrica*, **39**: 939-953.

Muellbauer, J. (1976), "Community Preferences and the Representative Consumer", *Econometrica*, **44**: 979-999.

Prais, S.J. and Houthakker, H.S. (1955), *The Analysis of Family Budgets,* Cambridge, Cambridge University Press.

Stoker, T. (1980), "Distribution Restrictions and the Form of Aggregate Functions", MIT Sloan School of Management, *Working Paper*, 1117-80.

Stoker, T. (1982), "The Use of Cross-Section Data to Characterize Macro Functions", *Journal of the American Statistical Association*, **77**: 369-380.

Stoker, T. (1986), "Simple Tests of Distributional Effects on Macroeconomic Equations", *Journal of Political Economy*, **94**: 763-795.

Stone, J.R.N. (1954a), "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand", *Economic Journal*, **64**: 511-527.

Stone, J.R.N. (1954b), *The Measurement of Consumers' Expenditure and Behaviour in United Kingdom, 1920-1938,* Cambridge, Cambridge University Press.

Theil, H. (1965), "The Information Approach to Demand Analysis", *Econometrica*, **33**: 67-87.

Tobin, J. (1950), "A Statistical Demand Function for Food in U.S.A.", *Journal of the Royal Statistical Society, Series A*, **113**: 113-141.

Working, H. (1943), "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, **38**: 43-56.

Table 1: OLS Regressions

1. Commodity Group g=1 vs. Other Goods

Modeling Commodity Group 1

Variable	Coefficient	Std.Error	t-value	t-prob
Constant	0.34038	0.032107	10.602	0.0000
LogY	-0.014446	0.0027304	-5.291	0.0000

R^2 = 0.0272853
F(1,998) = 27.995 [0.0000]
σ = 0.0245054
RSS = 0.5993157973 for 2 variables and 1000 observations
AR 1- 2 F(2,996) = 1.4293 [0.2400]
ARCH 1 F(1,996) = 0.11711 [0.7323]
Normality Chi^2(2)= 1.7256 [0.4220]
X_i^2 F(2,995) = 0.12428 [0.8831]
RESET F(1,997) = 4.8094 [0.0285] **

Variable	Coefficient	Std.Error	t-value	t-prob
Constant	-1.8393	0.99441	-1.850	0.0647
LogY	0.35625	0.16905	2.107	0.0353
LogYSq	-0.015752	0.0071827	-2.193	0.0285

R^2 = 0.031955
F(2,997) = 16.455 [0.0000]
σ = 0.0244588 RSS = 0.5964386538 for 3 variables and 1000 observations
AR 1- 2 F(2,995) = 1.3306 [0.2648]
ARCH 1 F(1,995) = 0.069725 [0.7918]
Normality Chi^2(2)= 1.3718 [0.5036]
X_i^2 F(4,992) = 0.075916 [0.9896]
RESET F(1,996) = 0.087293 [0.7677]

Modeling Other Goods

Variable	Coefficient	Std.Error	t-value	t-prob
Constant	0.65962	0.032107	20.545	0.0000
LogY	0.014446	0.0027304	5.291	0.0000

R^2 = 0.0272853
F(1,998) = 27.995 [0.0000]
σ = 0.0245054
RSS = 0.5993157992 for 2 variables and 1000 observations
AR 1- 2 F(2,996) = 1.4293 [0.2400]
ARCH 1 F(1,996) = 0.11711 [0.7323]
Normality Chi^2(2)= 1.7256 [0.4220]
X_i^2 F(2,995) = 0.12428 [0.8831]
RESET F(1,997) = 4.8102 [0.0285] *

Variable	Coefficient	Std.Error	t-value	t-prob
Constant	2.8393	0.99441	2.855	0.0044
LogY	-0.35625	0.16905	-2.107	0.0353
LogYSq	0.015752	0.0071827	2.193	0.0285

R^2 = 0.031955
F(2,997) = 16.455 [0.0000]
σ = 0.0244588
RSS = 0.5964386555 for 3 variables and 1000 observations
AR 1- 2 F(2,995) = 1.3306 [0.2648]
ARCH 1 F(1,995) = 0.069725 [0.7918]
Normality Chi^2(2)= 1.3718 [0.5036]
X_i^2 F(4,992) = 0.075916 [0.9896]
X_i*X_j F(5,991) = 0.40092 [0.8484]
RESET F(1,996) = 0.064755 [0.7992]

2. Food (Groups 1 and 2) vs. non-Food (Groups 3, 4 and 5)

Modeling Food

Variable	Coefficient	Std.Error	t-value	t-prob
Constant	0.55795	0.040546	13.761	0.0000
LogY	-0.017478	0.0034480	-5.069	0.0000

R^2 = 0.025101 F(1,998) = 25.696 [0.0000]
σ = 0.0309468 RSS = 0.9557908475 for 2 variables and 1000 observations

AR 1- 2 F(2,996) = 0.087195 [0.9165]
ARCH 1 F(1,996) = 1.3868 [0.2392]
Normality Chi^2(2)= 0.94589 [0.6232]
X_i^2 F(2,995) = 0.67247 [0.5107]
RESET F(1,997) = 5.3045 [0.0215] **

Variable	Coefficient	Std.Error	t-value	t-prob	JHCSE	PartR^2
Constant	-2.3321	1.2555	-1.858	0.0635	1.2495	0.0034
LogY	0.47404	0.21344	2.221	0.0266	0.21262	0.0049
LogYSq	-0.020886	0.0090685	-2.303	0.0215	0.0090425	0.0053

R^2 = 0.0302605
F(2,997) = 15.556 [0.0000]
σ = 0.0308803
RSS = 0.9507325206 for 3 variables and 1000 observations

AR 1- 2 F(2,995) = 0.078782 [0.9242]
ARCH 1 F(1,995) = 1.8712 [0.1716]
Normality Chi^2(2) = 1.2526 [0.5346]
X_i^2 F(4,992) = 0.64262 [0.6322]
X_i*X_j F(5,991) = 1.3313 [0.2485]
RESET F(1,996) = 2.5479 [0.1108]

Modeling non-Food

Variable	Coefficient	Std.Error	t-value	t-prob	JHCSE	PartR^2
Constant	0.44205	0.040546	10.903	0.0000	0.041830	0.1064
LogY	0.017478	0.0034480	5.069	0.0000	0.0035596	0.0251

R^2 = 0.025101
F(1,998) = 25.696 [0.0000]
σ = 0.0309468
RSS = 0.9557908475 for 2 variables and 1000 observations
AR 1- 2 F(2,996) = 0.087195 [0.9165]
ARCH 1 F(1,996) = 1.3868 [0.2392]
Normality $Chi^2(2)$= 0.94589 [0.6232]
X_i^2 F(2,995) = 0.67247 [0.5107]
RESET F(1,997) = 5.3046 [0.0215] **

Variable	Coefficient	Std.Error	t-value	t-prob	JHCSE	PartR^2
Constant	3.3321	1.2555	2.654	0.0081	1.2495	0.0070
LogY	-0.47404	0.21344	-2.221	0.0266	0.21262	0.0049
LogYSq	0.020886	0.0090685	2.303	0.0215	0.0090425	0.0053

R^2 = 0.0302605
F(2,997) = 15.556 [0.0000]
σ = 0.0308803
RSS = 0.9507325206 for 3 variables and 1000 observations
AR 1- 2 F(2,995) = 0.078782 [0.9242]
ARCH 1 F(1,995) = 1.8712 [0.1716]
Normality $Chi^2(2)$= 1.2526 [0.5346]
X_i^2 F(4,992) = 0.64262 [0.6322]
X_i*X_j F(5,991) = 1.3312 [0.2485]
RESET F(1,996) = 2.4372 [0.1188]

Table 2: Mean Scaling

a) Correlation Coefficients, Slopes and t-tests (with related 2-tail probabilities) in the linear regression between λ_{qt} and $\overline{Y}_t$

Quantile	Correlation Coefficients	Slope	Interc	t-test (H_0 : Slope=0)	t-prob
0.05	-0.48434	-0.02701	0.931281	-26.4454	2.87E-30
0.10	-0.49381	-0.0233	0.95172	-29.7429	1.4E-32
0.15	-0.48586	-0.02054	0.965446	-31.2597	1.44E-33
0.20	-0.48467	-0.01792	0.975194	-31.5544	9.4E-34
0.25	-0.50168	-0.01574	0.984421	-36.3622	1.36E-36
0.30	-0.5158	-0.01339	0.990982	-40.8192	6.27E-39
0.35	-0.50572	-0.0109	0.995992	-41.6818	2.36E-39
0.40	-0.5456	-0.00887	1.003407	-54.91	5.55E-45
0.45	-0.57435	-0.00667	1.008879	-43.11	4.88E-40
0.50	-0.64344	-0.00399	1.010857	-22.0775	8.71E-27
0.55	-0.7525	-0.0013	1.014056	-6.10536	1.73E-07
0.60	0.007196	0.001201	1.01882	4.644559	2.67E-05
0.65	0.362836	0.00375	1.024419	11.36428	3.27E-15
0.70	0.356507	0.006315	1.032558	14.48989	3.75E-19
0.75	0.430486	0.009965	1.036768	19.30651	2.87E-24
0.80	0.47293	0.014504	1.038847	23.58233	4.78E-28
0.85	0.500399	0.020081	1.040037	28.57797	8.63E-32
0.90	0.517477	0.029452	1.026954	37.13548	5.12E-37
0.95	0.534887	0.043081	1.015893	48.10711	2.84E-42
1.00	0.551493	0.107345	1.002358	27.31886	6.64E-31

b) Correlation Coefficient, Slope and t-Test in the linear regression between $\widetilde{Y}_t$ and $\overline{Y}_t$

	Correlation Coefficient	Slope	R^2	t-test (H_0 : Slope=0)	t-Prob
Statistics	-0.6008	-1.2629E-09	0.36	-5.1553	4.7404E-06

Table 3.

Output of the regressions: $log\ C_{h,t}{}^{g} = \alpha_t{}^{g} + \beta_t{}^{g}\ log\ Y_{h,t} + \varepsilon_t{}^{g}$.

Commodity Group 1	$\hat{\alpha}_t^g$	$\hat{\beta}_t^g$	Std.Error of $\hat{\beta}_t^g$	t-value H_o: $\hat{\beta}_t^g = 0$	R^2	LM Test For Omitted $(Log\ Y_t)^2$	
t=200	-0.2498	0.8293	0.3847	19.064 [0.0000] **	0.27	7.0917 [0.0079]	**
t=250	-0.1628	0.8446	0.0250	33.721 [0.0000] **	0.53	2.1374 [0.1441]	
t=300	-0.5754	0.8978	0.0162	55.316 [0.0000] **	0.75	2.2440 [0.1344]	
t=350	-1.0496	0.9446	0.0141	66.731 [0.0000] **	0.81	6.8328 [0.0091]	**
t=400	-1.0251	0.9461	0.0143	66.359 [0.0000] **	0.82	16.030 [0.0001]	**
t=450	-1.0058	0.9474	0.0152	62.354 [0.0000] **	0.80	8.4432 [0.0037]	**
t=500	-0.9558	0.9475	0.0149	63.196 [0.0000] **	0.81	3.3934 [0.0822]	*

Commodity Groups 2,…,5 (Aggregated)	$\hat{\alpha}_t^g$	$\hat{\beta}_t^g$	Std.Error of $\hat{\beta}_t^g$	t-value H_o: $\hat{\beta}_t^g = 0$	R^2	LM Test F(1,997) for Omitted $(Log\ Y_t)^2$	
t=200	-0.4585	1.0294	0.0096	107.66 [0.0000] **	0.92	4.7358 [0.0298]	*
t=250	-0.4774	1.0277	0.0052	198.02 [0.0000] **	0.97	4.8723 [0.0402]	*
t=300	-0.3919	1.0174	0.0033	308.54 [0.0000] **	0.99	4.6806 [0.0307]	*
t=350	-0.3400	1.0116	0.0028	356.32 [0.0000] **	0.99	7.0823 [0.0079]	**
t=400	-0.3414	1.0110	0.0027	367.72 [0.0000] **	0.99	16.033 [0.0001]	**
t=450	-0.3041	1.0081	0.0028	362.38 [0.0000] **	0.99	17.215 [0.0000]	**
t=500	-0.3411	1.0098	0.0027	367.92 [0.0000] **	0.99	4.6001 [0.0318]	*

Table 4

Reset Test for Functional Form Specification in the extended regression

$$log\ C_{h,t}^{g} = \alpha_t^{g} + \beta_t^{g}\ log\ Y_{h,t} + (log\ Y_{h,t})^{2} + \varepsilon_{h,t}^{g}$$

Time-Period	RESET Test F(1, 996) Commodity Group 1	Commodity Groups 2,3,4 and 5 (Aggregated)
200	0.37068 [0.5428]	0.29995 [0.5840]
250	0.67287 [0.4122]	0.15043 [0.6982]
300	0.00980 [0.9211]	0.00083 [0.9769]
350	5.95680 [0.0148] *	2.31770 [0.1282]
400	8.50660 [0.0036] **	2.67330 [0.1024]
450	4.76523 [0.0293] *	2.65432 [0.1035]
500	0.15582 [0.6931]	0.89528 [0.3443]

Table 5

χ^2 and F tests for Income Dependent Heteroscedasticity

(Auxiliary Regression: $\varepsilon_{h,t}^{g} = \omega_0 + \omega_1\ log\ Y_{h,t} + \omega_2\ (log\ Y_{h,t})^{2} + \nu_{h,t}^{g}$)

Time	Commodity Group 1		Commodity Groups 2,3,4 and 5 (Aggregated)	
	χ^2 (2) Test	F-Form F(2,995)	χ^2 (2) Test	F-Form F(2,995)
200	1.4007 [0.4964]	0.6978 [0.4979]	17.030 [0.0002] **	8.6190 [0.0002] **
250	0.1052 [0.9488]	0.0523 [0.9423]	14.965 [0.0006] **	7.5582 [0.0006] **
300	4.0997 [0.1288]	2.0480 [0.1295]	7.826 [0.0199]*	3.901 [0.0202]*
350	3.3612 [0.1863]	1.6778 [0.1873]	5.9344 [0.0514] *	2.9700 [0.0518] *
400	5.9352 [0.0514]*	2.9704 [0.0517]*	5.2625 [0.0720] *	2.6319 [0.0724] *
450	1.4154 [0.4928]	0.7051 [0.4943]	1.7359 [0.4198]	0.8651 [0.4213]
500	8.4578 [0.0146]*	4.2437 [0.0146]*	14.468 [0.0007] **	7.3034 [0.0007] **

Table 6
Stationarity vs. Stochastic Non-Stationarity of the series *log* C_t^g

Unit-root tests; Sample Period is from t=256 to t=491
Critical values: 5%=-3.43 1%=-4; Constant and Trend included

	t-adf	β	σ	lag	t-Test	t-Prob	F-Prob
LogC1	-1.7360	0.98067	0.0016726	5	1.5883	0.1136	
LogC1	-1.6633	0.98143	0.0016781	4	-0.81041	0.4185	0.1136
LogC1	-1.7092	0.98096	0.0016769	3	-2.0746	0.0391	0.2058
LogC1	-1.8478	0.97932	0.0016888	2	0.44875	0.6540	0.0600
LogC1	-1.8269	0.97963	0.0016859	1	0.80525	0.4215	0.1065
LogC1	-1.7877	0.98011	0.0016846	0			0.1416
LogC2	-1.7080	0.95354	0.0018160	5	-1.6429	0.1018	
LogC2	-2.0915	0.94417	0.0018227	4	1.2688	0.2058	0.1018
LogC2	-1.8738	0.95092	0.0018251	3	-0.25230	0.8010	0.1176
LogC2	-1.9770	0.94951	0.0018214	2	0.00777	0.9938	0.2258
LogC2	-2.0336	0.94956	0.0018175	1	0.040983	0.9673	0.3591
LogC2	-2.0876	0.94980	0.0018136	0			0.4969
LogC3	-3.2396	0.89898	0.0017289	5	-0.37166	0.7105	
LogC3	-3.3830	0.89672	0.0017257	4	-0.67679	0.4992	0.7105
LogC3	-3.6111*	0.89237	0.0017236	3	0.95407	0.3410	0.7432
LogC3	-3.4898*	0.89826	0.0017233	2	0.83900	0.4023	0.6828
LogC3	-3.3932	0.90306	0.0017222	1	-0.54188	0.5884	0.6995
LogC3	-3.5922*	0.89981	0.0017196	0			0.7776
LogC4	-0.74438	0.98897	0.0016310	5	-0.00089	0.9993	
LogC4	-0.75344	0.98897	0.0016275	4	-0.80004	0.4245	0.9993
LogC4	-0.87384	0.98734	0.0016262	3	-0.23810	0.8120	0.7275
LogC4	-0.91981	0.98684	0.0016229	2	-0.48071	0.6312	0.8746
LogC4	-0.99945	0.98587	0.0016202	1	-0.77989	0.4362	0.9210
LogC4	-1.1176	0.98436	0.0016188	0			0.9099
LogC5	-1.3024	0.98215	0.0013638	5	1.6495	0.1004	
LogC5	-1.1024	0.98495	0.0013689	4	0.55503	0.5794	0.1004
LogC5	-1.0486	0.98579	0.0013669	3	-0.24757	0.8047	0.2219
LogC5	-1.0895	0.98538	0.0013641	2	0.68036	0.4970	0.3796
LogC5	-1.0138	0.98652	0.0013625	1	-1.2706	0.2051	0.4712
LogC5	-1.1708	0.98452	0.0013643	0			0.3987

Table 7

Estimation Results and Diagnostic Tests on the VAR model relating changes in log of nominal consumption for each commodity group to changes in log of real income and price indices (tail probabilities in square brackets).

(i) Estimated Coefficients (Standard Errors in round brackets)

$$\begin{bmatrix} \Delta \log C_t^1 \\ \Delta \log C_t^2 \\ \Delta \log C_t^3 \\ \Delta \log C_t^4 \end{bmatrix} = \begin{matrix} \underset{(0.0034)}{0.9845} \Delta \log M_t + \underset{(0.0085)}{0.1659} \Delta P_t^1 + \underset{(0.0084)}{0.1855} \Delta P_t^2 + \underset{(0.0074)}{0.1931} \Delta P_t^3 + \underset{(0.0070)}{0.1905} \Delta P_t^4 + \underset{(0.0070)}{0.2674} \Delta P_t^5 \\ \underset{(0.0037)}{0.9939} \Delta \log M_t + \underset{(0.0090)}{0.2046} \Delta P_t^1 + \underset{(0.0089)}{0.1748} \Delta P_t^2 + \underset{(0.0078)}{0.1975} \Delta P_t^3 + \underset{(0.0073)}{0.2042} \Delta P_t^4 + \underset{(0.0090)}{0.2635} \Delta P_t^5 \\ \underset{(0.0035)}{1.0065} \Delta \log M_t + \underset{(0.0086)}{0.1712} \Delta P_t^1 + \underset{(0.0085)}{0.1996} \Delta P_t^2 + \underset{(0.0075)}{0.1780} \Delta P_t^3 + \underset{(0.0070)}{0.1861} \Delta P_t^4 + \underset{(0.0087)}{0.2379} \Delta P_t^5 \\ \underset{(0.0032)}{0.9991} \Delta \log M_t + \underset{(0.0081)}{0.1840} \Delta P_t^1 + \underset{(0.0080)}{0.1690} \Delta P_t^2 + \underset{(0.0070)}{0.1878} \Delta P_t^3 + \underset{(0.0066)}{0.2088} \Delta P_t^4 + \underset{(0.0081)}{0.2543} \Delta P_t^5 \end{matrix}$$

(ii) Diagnostic Tests

Fitting Statistics:

σ = 0.00150022	RSS = 0.0005311548359	loglik = 6304.6863
Log|Ω| = -52.1048	|Ω| = 2.35046e-023	T = 242
Log|Y'Y/T| = -43.7033	R^2(LR) = 0.999775	R^2 (LM) = 0.281764

F-test on all regressors except unrestricted: F(24,814) = 343.08 [0.0000] **

F-tests on retained regressors, F(4, 233):

ΔLogM	227441	[0.0000] **	ΔP_1	1247.97	[0.0000] **
ΔP_2	1287.66	[0.0000] **	ΔP_3	1784.30	[0.0000] **
ΔP_4	2211.54	[0.0000] **	ΔP_5	2458.94	[0.0000] **

Equations	Diagnostic Tests					
	Portmanteau (12 Lags)	AR 1-2 F(2,234)	Normality Chi^2(2)	ARCH 1 F(1,234)	X_i^2 F(12,223)	$X_i * X_j$ F(27,208)
$\Delta log C_t^1$	17.471	0.7094 [0.4930]	1.6181 [0.4453]	0.6469 [0.4221]	0.8829 [0.5651]	0.8454 [0.6885]
$\Delta log C_t^2$	13.713	0.5546 [0.5751]	0.0492 [0.9757]	0.5065 [0.4774]	0.8078 [0.6423]	0.7346 [0.8281]
$\Delta log C_t^3$	4.3074	0.1188 [0.8880]	1.2494 [0.5354]	0.9476 [0.3313]	1.1682 [0.3073]	0.7950 [0.7559]
$\Delta log C_t^4$	9.3214	0.2119 [0.8092]	2.8726 [0.2378]	0.6739 [0.4125]	1.0201 [0.4312]	1.1956 [0.2407]
VAR	174.63	1.0124 [0.4495]	4.1483 [0.8435]	-	0.9667 [0.5846]	0.97063 [0.9274]

Table 8

Stationarity vs. Stochastic Non-Stationarity of the series W_t^g

Unit-root tests; Sample Period is from t=253 to t=491
Critical values: 5%=-2.874 1%=-3.459; Constant included

	t-adf	β	σ	lag	t-Test	t-prob	F-prob
W1	0.50766	1.0015	0.00026168	2	0.55747	0.5777	
W1	0.55740	1.0016	0.00026129	1	0.75130	0.4532	0.5777
W1	0.61719	1.0018	0.00026105	0			0.6466
W2	-0.26652	0.99730	0.00030707	2	-0.89313	0.3727	
W2	-0.37172	0.99627	0.00030694	1	0.13438	0.8932	0.3727
W2	-0.35780	0.99644	0.00030630	0			0.6655
W3	-1.1284	0.99020	0.00032129	2	-0.35660	0.7217	
W3	-1.1461	0.99008	0.00032070	1	-0.59432	0.5529	0.7217
W3	-1.1795	0.98982	0.00032026	0			0.7872
W4	-1.4144	0.98690	0.00028880	2	-0.19794	0.8433	
W4	-1.4288	0.98681	0.00028821	1	-0.73246	0.4646	0.8433
W4	-1.4680	0.98648	0.00028793	0			0.7510
W5	-0.29730	0.99899	0.00034315	2	1.4889	0.1379	
W5	-0.25012	0.99915	0.00034404	1	-2.2129	0.0279	0.1379
W5	-0.32907	0.99887	0.00034685	0			0.0297

Unit-root tests; Sample Period is from t=253 to t=491
Critical values: 5%=-3.43 1%=-4; Constant and Trend included

	t-adf	β	σ	lag	t-Test	t-prob	F-prob
W1	-1.9450	0.97632	0.00025973	2	0.67013	0.5034	
W1	-1.9051	0.97689	0.00025943	1	0.86694	0.3869	0.5034
W1	-1.8501	0.97762	0.00025929	0			0.5500
W2	-2.1372	0.95243	0.00030443	2	-0.55185	0.5816	
W2	-2.2821	0.95016	0.00030398	1	0.47871	0.6326	0.5816
W2	-2.2355	0.95214	0.00030348	0			0.7663
W3	-3.7804*	0.87969	0.00031341	2	0.44783	0.6547	
W3	-3.7832*	0.88291	0.00031287	1	0.22129	0.8251	0.6547
W3	-3.8438*	0.88449	0.00031224	0			0.8828
W4	-0.27694	0.99675	0.00028826	2	-0.45330	0.6507	
W4	-0.35175	0.99593	0.00028778	1	-0.94940	0.3434	0.6507
W4	-0.49300	0.99436	0.00028772	0			0.5766
W5	-1.6823	0.97660	0.00034188	2	1.6218	0.1062	
W5	-1.5450	0.97852	0.00034306	1	-2.0700	0.0395	0.1062
W5	-1.7433	0.97571	0.00034544	0			0.0326

Table 9
Estimation Results for Budget Shares of Group 1 and 5

Modelling ΔW^1 by OLS. The present sample is: 250 to 491

Variable	Coefficient	Std.Error	t-value	t-prob	PartR^2
Constant	0.014399	0.017400	0.828	0.4091	0.0041
DW1_1	0.067477	0.080796	0.835	0.4048	0.0042
DW1_2	0.072266	0.078136	0.925	0.3564	0.0052
DW1_3	-0.13778	0.079565	-1.732	0.0852	0.0178
DW1_4	-0.12425	0.078185	-1.589	0.1139	0.0151
DW1_5	0.0081782	0.079610	0.103	0.9183	0.0001
DW1_6	0.041917	0.081683	0.513	0.6085	0.0016
DW1_7	-0.093529	0.079630	-1.175	0.2419	0.0083
DW1_8	-0.11975	0.077689	-1.541	0.1251	0.0142
DW1_9	0.11576	0.076248	1.518	0.1309	0.0138
DW1_10	-0.037119	0.077013	-0.482	0.6305	0.0014
DLogY	-0.0079119	0.034163	-0.232	0.8171	0.0003
DLogY_1	0.0055317	0.034025	0.163	0.8711	0.0002
DLogY_2	0.020731	0.033492	0.619	0.5368	0.0023
DLogY_3	-0.018563	0.034391	-0.540	0.5901	0.0018
DLogY_4	0.045341	0.034535	1.313	0.1910	0.0103
DLogY_5	0.077241	0.034523	2.237	0.0266	0.0294
DLogY_6	0.039277	0.035487	1.107	0.2700	0.0074
DLogY_7	-0.051060	0.034684	-1.472	0.1429	0.0130
DLogY_8	-0.0069982	0.035748	-0.196	0.8450	0.0002
DLogY_9	0.074726	0.034452	2.169	0.0315	0.0277
DLogY_10	0.057994	0.034924	1.661	0.0987	0.0164
P1	-0.0040681	0.0022155	-1.836	0.0681	0.0200
P1_1	-0.0014405	0.0022903	-0.629	0.5303	0.0024
P1_2	-0.0049606	0.0023072	-2.150	0.0330	0.0273
P1_3	0.0038150	0.0023237	1.642	0.1025	0.0161
P1_4	-0.0037049	0.0023212	-1.596	0.1124	0.0152
P1_5	0.0012277	0.0023123	0.531	0.5962	0.0017
P1_6	-0.00018341	0.0023001	-0.080	0.9365	0.0000
P1_7	-0.0029585	0.0023352	-1.267	0.2070	0.0096
P1_8	-0.0027406	0.0024003	-1.142	0.2552	0.0078
P1_9	0.0018748	0.0023725	0.790	0.4305	0.0038
P1_10	-0.00032785	0.0023565	-0.139	0.8895	0.0001
P2	0.00085988	0.0022112	0.389	0.6979	0.0009
P2_1	0.0026362	0.0021606	1.220	0.2242	0.0089
P2_2	0.0028655	0.0022212	1.290	0.1988	0.0100
P2_3	-3.1874e-005	0.0022014	-0.014	0.9885	0.0000
P2_4	-0.0019596	0.0022060	-0.888	0.3757	0.0048
P2_5	0.0030341	0.0022523	1.347	0.1798	0.0109
P2_6	0.0025096	0.0022884	1.097	0.2744	0.0072
P2_7	-0.00037745	0.0022475	-0.168	0.8668	0.0002
P2_8	-0.0041533	0.0022574	-1.840	0.0676	0.0201
P2_9	0.0031807	0.0022966	1.385	0.1679	0.0115
P2_10	-0.00064465	0.0022645	-0.285	0.7762	0.0005
P3	-0.00075375	0.0021648	-0.348	0.7281	0.0007
P3_1	-0.0012626	0.0020942	-0.603	0.5474	0.0022
P3_2	0.0031266	0.0021020	1.487	0.1388	0.0132
P3_3	0.0034611	0.0021262	1.628	0.1055	0.0158
P3_4	0.00095185	0.0021348	0.446	0.6563	0.0012
P3_5	-0.0010036	0.0020987	-0.478	0.6331	0.0014
P3_6	-0.0017712	0.0020423	-0.867	0.3871	0.0045

P3_7	-0.00038312	0.0020183	-0.190	0.8497	0.0002
P3_8	0.0025405	0.0020090	1.265	0.2078	0.0096
P3_9	-0.0013308	0.0019754	-0.674	0.5015	0.0027
P3_10	0.0018484	0.0020330	0.909	0.3646	0.0050
P4	-0.0032225	0.0019425	-1.659	0.0990	0.0164
P4_1	-0.0011197	0.0019668	-0.569	0.5699	0.0020
P4_2	-0.00061173	0.0019426	-0.315	0.7532	0.0006
P4_3	-0.0026184	0.0019550	-1.339	0.1823	0.0108
P4_4	0.0019880	0.0019309	1.030	0.3047	0.0064
P4_5	0.00047779	0.0018993	0.252	0.8017	0.0004
P4_6	-0.0029956	0.0019146	-1.565	0.1196	0.0146
P4_7	-0.0054100	0.0019310	-2.802	0.0057	0.0454
P4_8	-0.0020209	0.0019950	-1.013	0.3126	0.0062
P4_9	0.0020014	0.0019600	1.021	0.3087	0.0063
P4_10	-0.0015133	0.0019722	-0.767	0.4440	0.0036
P5	0.0040710	0.0024694	1.649	0.1011	0.0162
P5_1	0.00015334	0.0025030	0.061	0.9512	0.0000
P5_2	-0.0025588	0.0024199	-1.057	0.2919	0.0067
P5_3	0.00064062	0.0023429	0.273	0.7849	0.0005
P5_4	-7.1316e-005	0.0023095	-0.031	0.9754	0.0000
P5_5	-0.0026849	0.0023655	-1.135	0.2580	0.0077
P5_6	0.0010634	0.0022633	0.470	0.6391	0.0013
P5_7	0.00091247	0.0022652	0.403	0.6876	0.0010
P5_8	-0.00093153	0.0022951	-0.406	0.6854	0.0010
P5_9	-0.0060717	0.0023336	-2.602	0.0101	0.0394
P5_10	-0.00094319	0.0024006	-0.393	0.6949	0.0009

R^2 = 0.374521 F(76,165) = 1.3 [0.0839] σ = 0.000248291 DW = 1.98
RSS = 1.017200123e-005 for 77 variables and 242 observations

AR 1- 2 F(2,163) = 1.1594 [0.3162]
ARCH 1 F(1,163) = 0.32107 [0.5717]
Normality $Chi^2(2)$= 3.053 [0.2173]
X_i^2 F(152, 12) = 0.13954 [1.0000]
RESET F(1,164) = 1.0579 [0.3052]

Solved Static Long Run equation (Std. Err. in parentheses)

ΔW^1 = +0.01193 (0.01485) -0.1958 ΔLogY (0.09559) -0.01116 P^1 (0.006759)

+0.006562 P^2 (0.005806) +0.004494 P^3 (0.005267) -0.01247 P^4 (0.006793)

-0.00532 P^5 (0.007467)

WALD test on the joint significance of the regressors in the static long-run equation:

$Chi^2(6)$ = 12.14 [0.0589]*

Modelling ΔW^5 by OLS; The present sample is: 250 to 491

Variable	Coefficient	Std.Error	t-value	t-prob	PartR^2
DW5_1	-0.082185	0.077074	-1.066	0.2878	0.0068
DW5_2	0.071646	0.076488	0.937	0.3503	0.0053
DW5_3	-0.0096307	0.077876	-0.124	0.9017	0.0001
DW5_4	0.055801	0.077361	0.721	0.4717	0.0031
DW5_5	0.033120	0.078369	0.423	0.6731	0.0011
DW5_6	-0.051880	0.076937	-0.674	0.5011	0.0027
DW5_7	-0.035467	0.076054	-0.466	0.6416	0.0013
DW5_8	0.00098872	0.076691	0.013	0.9897	0.0000
DW5_9	-0.12282	0.075094	-1.636	0.1038	0.0159
DW5_10	0.022530	0.075979	0.297	0.7672	0.0005
DLogY	0.012063	0.047117	0.256	0.7983	0.0004
DLogY_1	0.030212	0.046187	0.654	0.5139	0.0026
DLogY_2	-0.043489	0.046183	-0.942	0.3477	0.0053
DLogY_3	-0.036847	0.047890	-0.769	0.4427	0.0036
DLogY_4	0.029431	0.047730	0.617	0.5383	0.0023
DLogY_5	0.079272	0.048522	1.634	0.1042	0.0158
DLogY_6	-0.036480	0.048421	-0.753	0.4523	0.0034
DLogY_7	0.050134	0.047122	1.064	0.2889	0.0068
DLogY_8	-0.024313	0.047179	-0.515	0.6070	0.0016
DLogY_9	0.010077	0.046147	0.218	0.8274	0.0003
DLogY_10	-0.041547	0.046244	-0.898	0.3703	0.0048
P1	-0.0027548	0.0029610	-0.930	0.3535	0.0052
P1_1	0.0026985	0.0029956	0.901	0.3690	0.0049
P1_2	0.0036776	0.0030099	1.222	0.2235	0.0089
P1_3	0.0013651	0.0030325	0.450	0.6532	0.0012
P1_4	0.0011860	0.0030519	0.389	0.6981	0.0009
P1_5	-0.0055964	0.0030561	-1.831	0.0689	0.0198
P1_6	0.0025663	0.0030818	0.833	0.4062	0.0042
P1_7	0.00062041	0.0030890	0.201	0.8411	0.0002
P1_8	0.0017759	0.0031212	0.569	0.5701	0.0019
P1_9	-0.00069218	0.0031232	-0.222	0.8249	0.0003
P1_10	0.00086999	0.0030791	0.283	0.7779	0.0005
P2	-0.0032273	0.0030453	-1.060	0.2908	0.0067
P2_1	-0.0023747	0.0029514	-0.805	0.4222	0.0039
P2_2	0.0010171	0.0029625	0.343	0.7318	0.0007
P2_3	-0.0036754	0.0029815	-1.233	0.2194	0.0091
P2_4	0.0021103	0.0030270	0.697	0.4867	0.0029
P2_5	-0.0048765	0.0031438	-1.551	0.1228	0.0143
P2_6	1.3506e-005	0.0032066	0.004	0.9966	0.0000
P2_7	-0.0033518	0.0031482	-1.065	0.2886	0.0068
P2_8	-0.0040231	0.0031632	-1.272	0.2052	0.0097
P2_9	0.00076303	0.0031769	0.240	0.8105	0.0003
P2_10	-0.0020661	0.0030912	-0.668	0.5048	0.0027
P3	0.0015409	0.0029322	0.525	0.5999	0.0017
P3_1	0.0028598	0.0027970	1.022	0.3081	0.0063
P3_2	0.0029338	0.0027934	1.050	0.2951	0.0066
P3_3	0.0018003	0.0028435	0.633	0.5275	0.0024
P3_4	-0.0041426	0.0028621	-1.447	0.1497	0.0125
P3_5	-0.00081055	0.0028010	-0.289	0.7726	0.0005
P3_6	-0.0031370	0.0027469	-1.142	0.2551	0.0078
P3_7	-0.0016872	0.0028079	-0.601	0.5487	0.0022
P3_8	-0.0026356	0.0027974	-0.942	0.3475	0.0053
P3_9	0.0059293	0.0027877	2.127	0.0349	0.0265
P3_10	-0.0030227	0.0028883	-1.047	0.2968	0.0066

P4	-0.00096590	0.0025289	-0.382	0.7030	0.0009
P4_1	0.0020119	0.0025429	0.791	0.4300	0.0038
P4_2	0.0061854	0.0025650	2.411	0.0170	0.0338
P4_3	0.0027090	0.0026503	1.022	0.3082	0.0063
P4_4	-0.0015558	0.0026208	-0.594	0.5536	0.0021
P4_5	0.0023941	0.0026267	0.911	0.3634	0.0050
P4_6	0.0010877	0.0026446	0.411	0.6814	0.0010
P4_7	0.00078059	0.0026301	0.297	0.7670	0.0005
P4_8	0.0020296	0.0026642	0.762	0.4473	0.0035
P4_9	-0.0030548	0.0026582	-1.149	0.2521	0.0079
P4_10	0.0030617	0.0026875	1.139	0.2563	0.0078
P5	0.0012626	0.0033587	0.376	0.7075	0.0009
P5_1	-0.0030755	0.0033403	-0.921	0.3585	0.0051
P5_2	0.0032968	0.0032282	1.021	0.3086	0.0062
P5_3	-0.00097235	0.0031370	-0.310	0.7570	0.0006
P5_4	0.0029049	0.0030984	0.938	0.3498	0.0053
P5_5	-0.0044821	0.0031865	-1.407	0.1614	0.0118
P5_6	0.0011213	0.0031122	0.360	0.7191	0.0008
P5_7	0.00068764	0.0030605	0.225	0.8225	0.0003
P5_8	-0.0017455	0.0030233	-0.577	0.5645	0.0020
P5_9	0.00081573	0.0030821	0.265	0.7916	0.0004
P5_10	-0.00097498	0.0031339	-0.311	0.7561	0.0006

R^2 = 0.346397 σ = 0.000343694
DW = 1.99RSS = 1.960882911e-005 for 76 variables and 242 observations
AR 1- 2 F(2,164) = 0.077785 [0.9252]
ARCH 1 F(1,164) = 1.4758 [0.2262]
Normality $Chi^2(2)$ = 0.9086 [0.6349]
X_i^2 F(152, 13) = 0.085042 [1.0000]
RESET F(1,165) = 0.23458 [0.6288]

Solved Static Long Run equation (Std. Err. in parentheses)

$$\Delta W^5 = +0.02551\ \Delta LogY\ (0.1359) \quad +0.005113\ P^1\ (0.008066) \quad -0.01761\ P^2\ (0.007964)$$

$$-0.0003325\ P^3\ (0.007537) \quad +0.01313\ P^4\ (0.008583) \quad -0.001039\ P^5\ (0.0095)$$

WALD test on the joint significance of the regressors in the static long-run equation:

$Chi^2(6)$ = 20.006 [0.0028] **

Appendix 2

In order to try to assess the shape of cross-section Engel curve specification, we have firstly performed a descriptive analysis of the Working-Leser model:

$$w_i^g = \alpha^g + \beta_t^g \log m_i^{\,g}$$

where w_i^g is the budget share of agent i in commodity group g and m_i is total real income (expenditure) of agent i. As to aggregation in commodity groups, we considered two different setups, namely (i) goods are aggregated into the original 5 groups, g=1,… ,5; and (ii) goods are aggregated into 2 groups, i.e. g={B,L}, where B=(1∪2) stands for 'basics' and L=(3∪4∪5) stands for 'luxury'. For every commodity group, and for different points in time, we carried out cross-plots of w_i^g vs. (log m_i^g) and we performed both parametric (OLS) and non-parametric (Kernel) regressions.

As a general pattern (see Fig. 1), one is likely to find a low correlation between budget shares and log of income. Despite that, irrespective of the aggregation setup, budget shares and log of individual incomes seem to be correlated with the expected signs (cf. panels (a) and (b) with (c) and (d)). Moreover, quite in line with the results of Banks et al. (1997), non-linearities appear throughout, suggesting the need for higher order terms of log m_i^g in cross-section Engel curves.

We then estimated by standard OLS in both aggregation setups the alternative specifications:

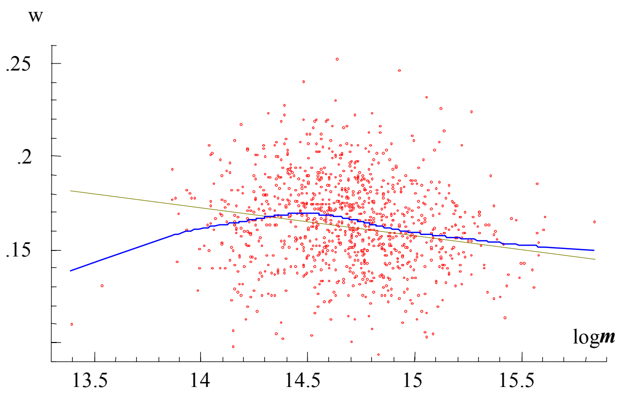
$$\text{M1:}\quad w_i^g = \alpha^g + \beta_{1t}^g \log m_i^{\,g} + \beta_{1t}^g (\log m_i^{\,g})^2 + \varepsilon_i^g$$
$$\text{M2:}\quad w_i^g = \alpha^g + \beta_t^g \log m_i^{\,g} + \varepsilon_i^g$$

Estimation results for a paradigmatic case are reported in Table 2 below. Although the R^2s for all the cross-section regressions are very low, both food-like and non-food-like expenditure shares display non-linearities in the log of income. Tests for Autoregressive Conditional (ARCH) and Income Dependent Heteroscedasticity (F-Test, not reported) failed to find any evidence for heteroscedastic residuals. Nevertheless, functional form mispecification arise in all estimated log-linear models: both the equivalent Reset F-test and LM tests – performed to assess whether the variable $[\log m_{h,t}]^2$ has been omitted – strongly reject the null hypothesis. However, once the square of the log of income is introduced in the regression, no mispecifications are reported, even though the R^2s still remain very low. Finally, further nonlinear terms appear to be not significant in explaining budget shares.

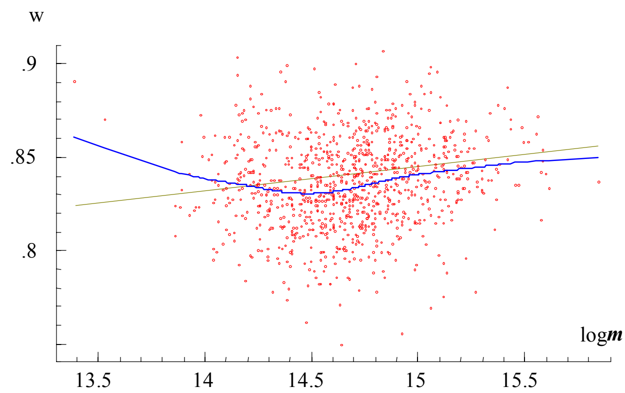
The foregoing results – quite in tune with those obtained for empirical data by Banks et al. (1997) – suggest, first, that non-linear terms (especially the square of the log of income) do indeed matter in Engel curve specifications, and, secondly, that the Gorman's Rank 3 assumption should be satisfied by our computer-simulated data. This conjecture is indeed supported by jointly testing a demand system for 4 out of the 5 commodity groups (avoiding singularity of the dependent variables matrix) and employing χ^2 statistics to test non-linear restrictions implied by the determinants of the matrices of estimated parameters (not shown).

Figure 1

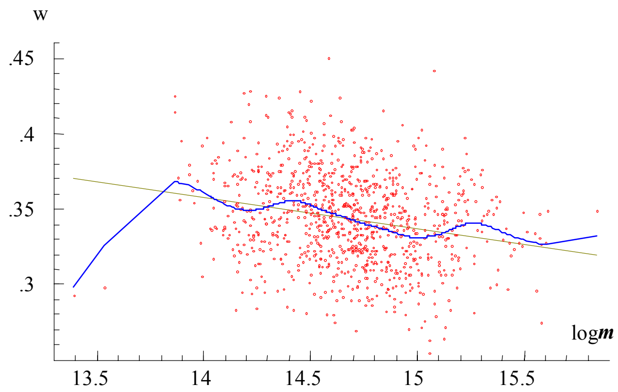
Linear Regression (Thin Lines) and Non Parametric - Kernel Estimates of Engel Curves



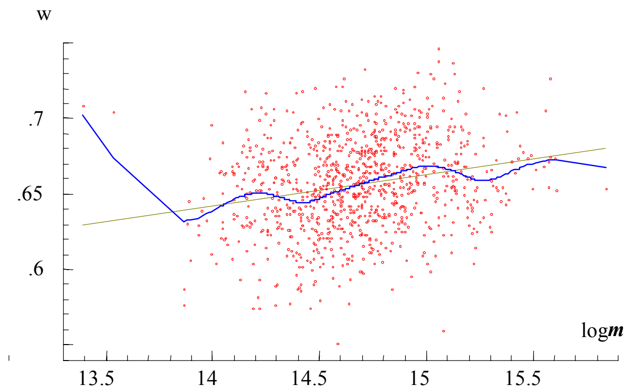
(a) Commodity Group g=1



(b) Commodity Group g=5



(c) Commodity Group g=B (1∪2∪3)



(d) Commodity Group g=L (4∪5)

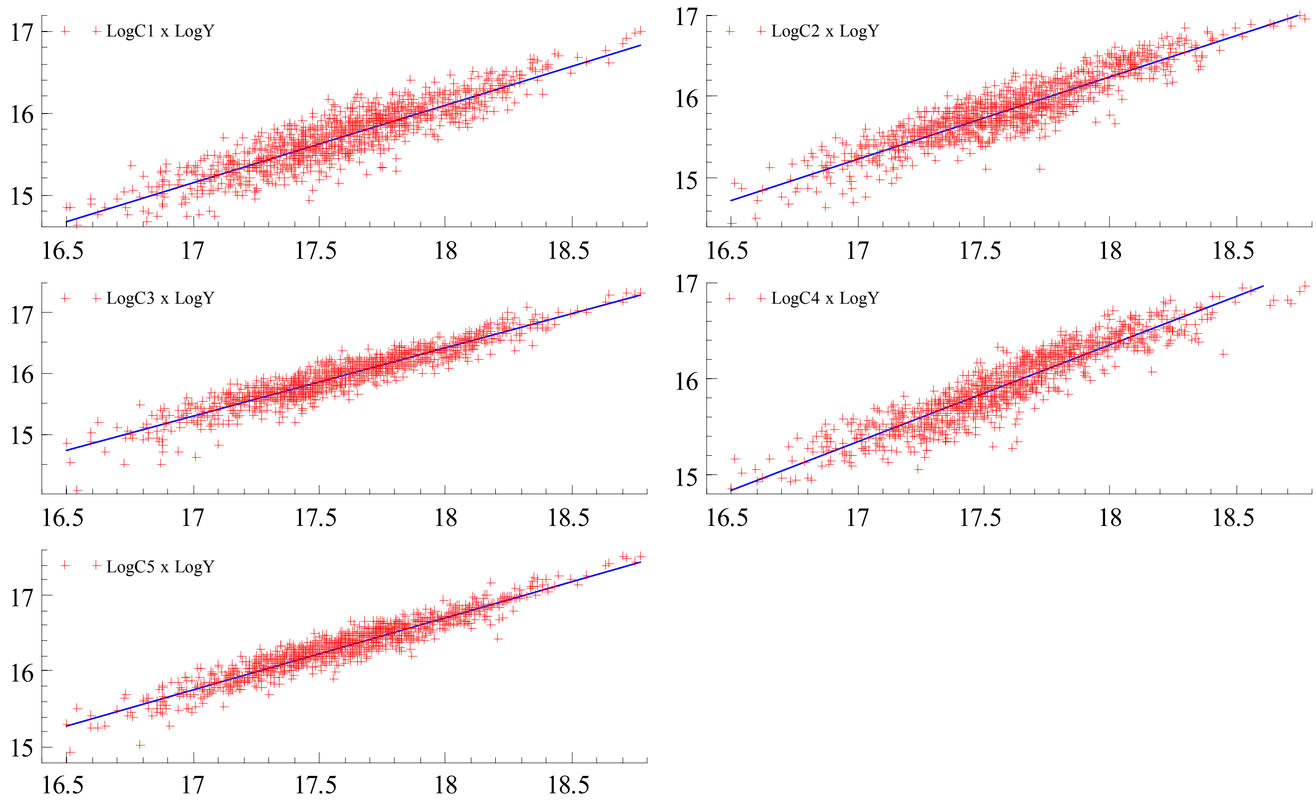
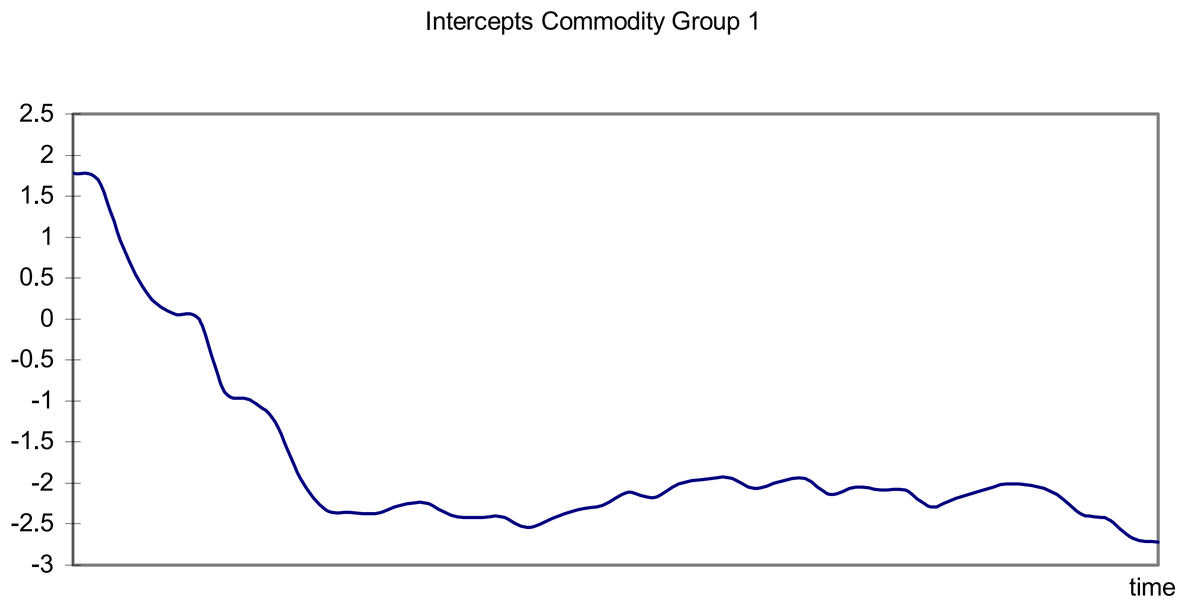


Figure 1(b)

Cross-Plots of Log (Cg) vs. Log (Y) at time t=500.

Figure 2
Evolution over Time of Income Elasticity and Intercepts for Commodity Group 1



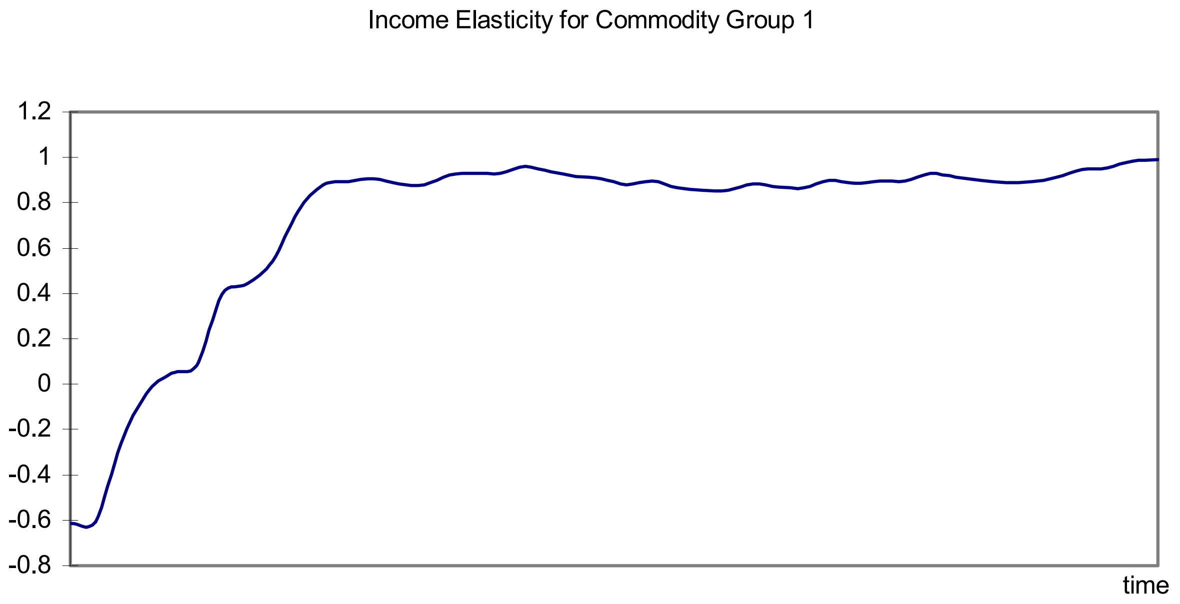
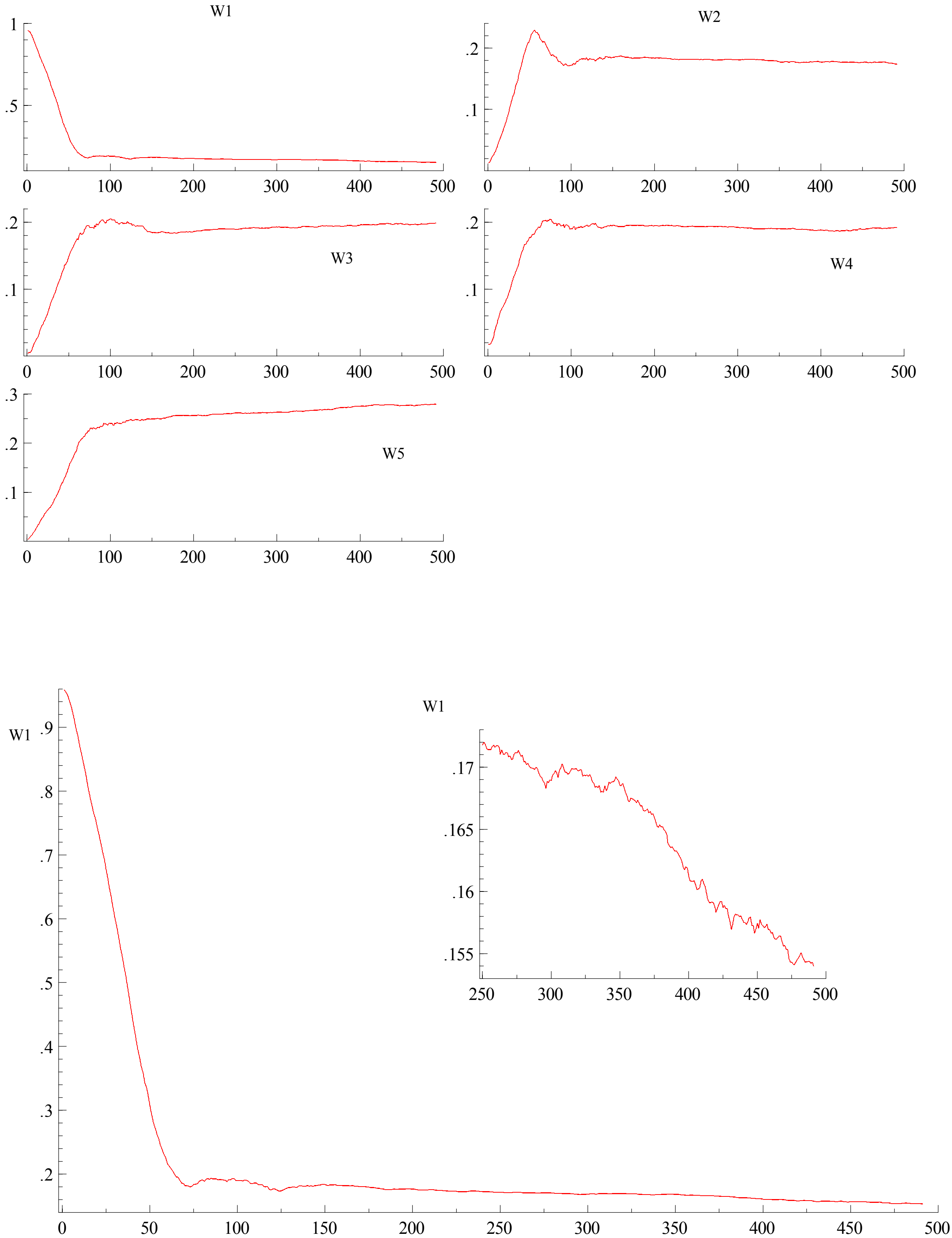


Figure 3
Budget Shares Time Series



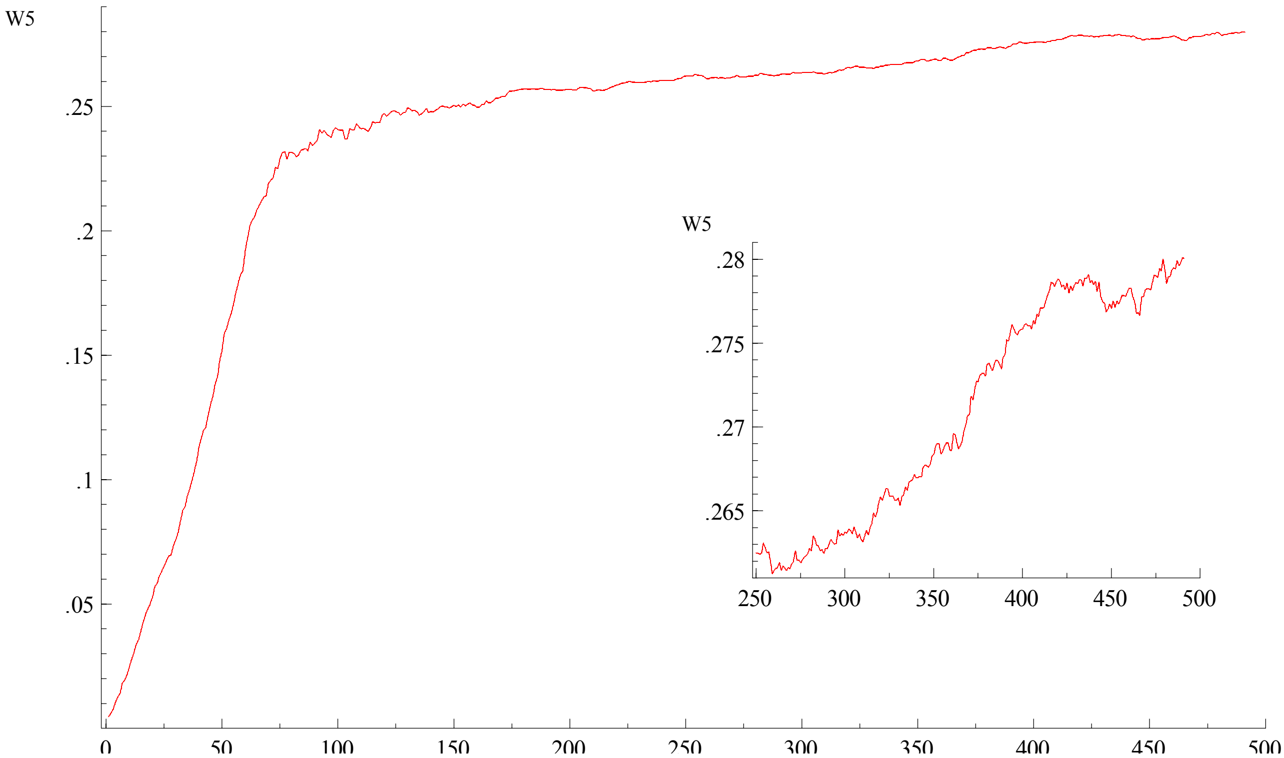

W5
.25
.2
.15
.1
.05
0
50
100
150
200
250
300
350
400
450
500
W5
.28
.275
.27
.265
250
300
350
400
450
500

Figure 4

Plots of the Normalized Lag Weights (lags t+1, t+2, … on the x-axis) for the regression:

$$\Delta w_t^g = \alpha_0 + \sum_{i=1}^{k} \alpha_i^g \Delta w_{t-i}^g + \sum_{j=0}^{n} \beta_j \Delta \log y_{t-j} + \sum_{h=0}^{m} \gamma_h p_{t-j} + \varepsilon_t^g$$

(Commodity Group 1)