EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1061 / JUNE 2009

THE DISTRIBUTION OF HOUSEHOLDS CONSUMPTION-EXPENDITURE BUDGET SHARES

by Matteo Barigozzi,
Lucia Alessi,
Marco Capasso
and Giorgio Fagiolo

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1061 / JUNE 2009

THE DISTRIBUTION OF HOUSEHOLDS CONSUMPTION-EXPENDITURE BUDGET SHARES [1]

by Matteo Barigozzi [2], Lucia Alessi [3], Marco Capasso [4] and Giorgio Fagiolo [5]

In 2009 all ECB publications feature a motif taken from the €200 banknote.

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=1413689.

1 The views expressed in this paper remain those of the authors and do not necessarily reflect those of the ECB.

2 European Center for Advanced Research in Economics and Statistics (ECARES) – Université Libre de Bruxelles and Sant'Anna School of Advanced Studies, Laboratory of Economics and Management, Piazza Martiri della Liberta, 33, Pisa, I-56127, Italy; e-mail: matteo.barigozzi@gmail.com

3 Corresponding Author: European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; e-mail: lucia.alessi@ecb.europa.eu

4 Urban and Regional research centre Utrecht (URU), Faculty of Geosciences, Utrecht University, and Tjalling C. Koopmans research Institute (TKI), Utrecht School of Economics, Utrecht University, P.O. Box 80.115, 3508 TC Utrecht, the Netherlands; e-mail: marco.capasso@gmail.com

5 Sant'Anna School of Advanced Studies, Laboratory of Economics and Management, Piazza Martiri della Liberta, 33, Pisa, I-56127, Italy; e-mail: fagiolo@sssup.it

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Website
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

ISSN 1725-2806 (online)

CONTENTS

Abstract

This paper explores the statistical properties of household consumption-expenditure budget share distributions —defined as the share of household total expenditure spent for purchasing a specific category of commodities— for a large sample of Italian households in the period 1989-2004. We find that household budget share distributions are fairly stable over time for each specific category, but profoundly heterogeneous across commodity categories. We then derive a parametric density that is able to satisfactorily characterize household budget share distributions and: (i) is consistent with the observed statistical properties of the underlying levels of household consumption-expenditure distributions; (ii) can accommodate the observed across-category heterogeneity in household budget share distributions. Finally, we taxonomize commodity categories according to the estimated parameters of the proposed density. We show that the resulting classification is consistent with the traditional economic scheme that labels commodities as necessary, luxury or inferior.

Keywords: Household Consumption Expenditure, Budget Shares, Sum of Log-Normal Distributions.

JEL Classification D3, D12, C12.

Nontechnical summary

The study of household budget allocation – i.e., how the budget of a household is allocated to buy different commodities – is one of the most traditional topics in economics. Household budget shares, defined as the share of total household resources spent for purchasing a specific class of goods, contain useful information to shed light on this issue.

In the last decades, this topic has received a lot of attention by applied economists, however most of the empirical literature on consumption is characterized by a theory-driven approach. Indeed, the parametric specifications that are employed in the estimation of each specific demand function are generally taken to be consistent with some underlying theory of household-expenditure behavior, which very often is the standard model based on utility maximization undertaken by fully-rational agents. Furthermore, the estimation of demand systems or Engel curves compresses household heterogeneity to the knowledge of the first two moments (at best) of household expenditure-level or budget-share distribution for the commodity category under study.

This approach may be problematic for a number of reasons: (i) heterogeneity of household consumption expenditure patterns is widely considered as a crucial feature and to fully characterize such heterogeneity, one should perform distributional analyses that carefully investigate how the *shape* of household consumption expenditure and budget share distributions change over time and between different commodity categories; (ii) understanding heterogeneity may be important to build sound micro-founded, macroeconomic, consumption models that go beyond the often disputable representative-agent assumption; (iii) a data-driven approach focused on distributional analysis may help to discover fresh stylized facts related to how households allocate their consumption expenditures.

In this paper we explore the statistical properties of household budget share distributions employing data from the "Survey of Household Income and Wealth" provided by the Bank of Italy for a sequence of 8 waves from 1989 to 2004. The sample comprises about 8000 households. We focus on four expenditure categories: nondurable goods and its subcategory food, durable goods and insurance premia (which are rarely studied in the literature).

We look for a unique, parsimonious, closed-form density family that: (i) is able to satisfactorily fit observed unconditional household budget share distributions, so as to accommodate the existing heterogeneity emerging across households, among different consumption categories and over time; (ii) is consistent with the statistical properties of the (observed) household consumption expenditure distributions employed to compute budget share distributions; (iii) features economically-interpretable parameters that, once estimated, can help to build economically meaningful taxonomies of commodity categories.

We begin with a descriptive analysis aimed at empirically exploring the stability of household budget share distributions over time: indeed, estimated sample moments show that the shape of budget share distributions did not dramatically change over the time interval considered. However, for any given wave, there emerges a lot of heterogeneity across commodity categories. We also show that the underlying household consumption expenditure distributions are well-proxied by log-normal distributions.

We then derive an original family of densities, defined over the unit interval, which is consistent with the detected log-normality of household consumption expenditure distributions. We find that for all the waves under study, according to simple measures of goodness-of-fit (e.g., the Average Absolute Deviation), the proposed density family is able to accommodate the existing shape-heterogeneity that characterizes household budget share distributions across different commodity categories. To benchmark our results, we also fit household budget share distributions with Beta variates, which are in principle very flexible densities defined over the unit interval but lack any any consistency requirements with respect to the underlying shape of household consumption expenditure distributions. Indeed, the density we derive outperforms the Beta in fitting observed household budget share distributions for the majority of cases. Furthermore, the estimated parameters of the proposed density allow to reproduce an economically-meaningful taxonomy of commodity categories, which interestingly maps into the traditional classification of commodities among necessary, luxury or inferior goods.

1 Introduction

The study of household budget allocation —i.e., how the budget of a household is allocated to buy different commodities— is one of the most traditional topics in economics (Prais and Houthakker, 1955). Household budget shares contain useful information to shed light on this issue. Indeed, the household budget share for a given commodity category g is defined as the ratio between the expenditure for the commodity category g and total household resources —as measured by, e.g., total expenditure or total income.

In the last decades, this topic has received a lot of attention by applied economists. In particular, many efforts have been devoted to develop statistical demand functions for homogeneous groups of commodities, e.g. by relating the expenditure of consumers or households for a given commodity category to prices and individual-specific variables as total expenditure or income, household size, head-of-household age, and so on.[1]

Such a research program has been mostly characterized by a theory-driven approach (Attanasio, 1999). In fact, the parametric specifications that are employed in the estimation of each specific demand function are in general taken to be consistent with some underlying theory of household expenditure behavior, which very often is the standard model based on utility maximization undertaken by fully-rational agents.[2] Furthermore, no matter whether parametric or non-parametric techniques are employed, the estimation of demand systems or Engel curves compresses household heterogeneity —for any given income or total expenditure level— to the knowledge of the first two moments (at best) of household expenditure level or budget share distribution for the commodity category under study.[3]

This of course is fully legitimate if the aim of the researcher is to empirically validate a given theoretical model, or if there are good reasons to believe that the distribution under analysis can be fully characterized by its first two moments. However, from a more data-driven perspective, constraining in this way the exploration of the statistical properties of the observed household expenditure patterns may be problematic for a number of reasons.

[1]The first study of this kind was made by Engel (1857), who empirically studied the relation between German households' total income and expenditure for different commodities (Moneta and Chai, 2005). For a comprehensive appraisal of this huge body of literature, see e.g. Deaton (1992); Blundell (1988) and references therein.

[2]See Aitchison and Brown (1954); Prais (1952); Banks et al. (1997); Blundell et al. (2007) among others.

[3]An Engel curve describes how the expenditure for a given commodity category varies as household's total resources, see Lewbel (2008).

First, heterogeneity of household consumption-expenditure patterns is widely considered as a crucial feature because, as Pasinetti (1981) notices: "At any given level of per capita income and at any given price structure, the proportion of income spent by each consumer on any specific commodity may be very different from one commodity to another". This suggests that, in order to fully characterize such heterogeneity, one should perform distributional analyses that carefully investigate how the shape —and not only the first two moments— of household consumption expenditure and household budget share distributions change over time and between different commodity categories. Second, understanding heterogeneity may be important to build sound micro-founded, macroeconomic, consumption models that go beyond the often disputable representative-agent assumption (Kirman, 1992; Hartley, 1997; Gallegati and Kirman, 1999).[4] Third, adopting a more theory-free approach focused on distributional analysis may help to discover fresh stylized facts related to how households allocate their consumption expenditures across different commodity categories. In fact, theory-free approaches aimed at searching for stylized facts are not new in economics and econometrics (see *inter alia* Kaldor, 1961; Hendry, 2000). More recently, this perspective has been revived in the field of econophysics, where the statistical properties of many interesting micro and macro economic variables (e.g., firm size and growth rates, industry and country growth rates, wealth and personal income, etc.) have been successfully characterized by using parametric techniques.[5] These studies show that, despite the turbulence typically detected at the microeconomic level (e.g., entry and exit of firms; positive and negative persistent shocks to personal income, etc.), there exists an incredible high level of regularity in the shape of microeconomic cross-section distributions, both across years and countries.

Notwithstanding such successful results, similar distributional analyses have not been extensively performed, so far, on consumption-related microeconomic variables such as household consumption expenditures and budget shares, for which reliable and detailed cross-section data are also available. This is somewhat surprising because —as Attanasio (1999) notices— under-

[4]For example, Caselli and Ventura (2000) show that models based on the representative-agent assumption impose almost no restrictions on household consumption expenditure and budget share distributions. On the contrary, Forni and Lippi (1997) demonstrate that heterogeneity is crucial when aggregating individual behavior in macro models. Furthermore, Ibragimov (2005) provides support to the insight that higher-than-two moments can have a relevant impact on the dynamics of macro models. Additional perspectives on the importance of heterogeneity in consumption and demand may be found in Hildenbrand (1994).

[5]See among others Chatterjee et al. (2005), Clementi and Gallegati (2005), Axtell (2001), Bottazzi and Secchi (2006) and Fagiolo et al. (2008).

standing consumption is crucial to both micro- and macro-economists, as it accounts for about two thirds of GDP and it decisively determines (and measures) social welfare.

There are only two exceptions —to the best of our knowledge— to this lack of distributional studies on household consumption indicators. In a recent contribution, Battistin et al. (2007) employ expenditure and income data from U.K. and U.S. surveys and show that total household consumption expenditure distributions are well-approximated by log-normal densities (or, as they put it, are "more log-normal than income").[6] In a complementary paper (Fagiolo et al., 2009), we argue that log-normality is valid only as a first approximation for Italian total household consumption expenditure distributions, while a refined analysis reveals asymmetric departures from log-normality in the tails of the distributions.

Both contributions focus on characterizing the dynamics of household consumption expenditure aggregate distributions only and nothing is said on the statistical properties of household consumption expenditure or budget share distributions disaggregated among commodity categories. This paper is a preliminary attempt to fill this gap. To do so, we employ data from the "Survey of Household Income and Wealth" (SHIW) provided by the Bank of Italy to study household consumption expenditure and budget share distributions for a sequence of 8 waves between 1989 and 2004. We focus on four commodity categories: nondurable goods, food, durable goods, and insurance premia (which are rarely studied in the literature).[7]

We aim at empirically investigating the statistical properties of *unconditional* household budget share distributions (and consumption expenditure distributions) of these four commodity categories and their dynamics with a parametric approach.[8] More specifically, we look for a unique, parsimonious, closed-form density family that: (i) is able to satisfactorily fit observed unconditional household budget share distributions, so as to accommodate the existing heterogeneity emerging across households, among different commodity categories and over time; (ii) is consistent with the statistical properties of the (observed) household consumption expenditures distributions employed to compute budget share distributions; (iii) features economically-interpretable parameters that, once estimated, can help one to build economically-meaningful

[6] Log-normality of household consumption expenditure distributions in U.K. is confirmed by another early study in the econophysics domain, see Hohnisch et al. (2002). See also Mizuno et al. (2008) for a study of the distributional properties of individual purchases in Japanese convenience stores.

[7] Food is actually a subcategory of nondurable goods, but for its intrinsic importance we consider it as a separate commodity category throughout the paper.

[8] By *unconditional* distributions we mean here not conditioned to total household resources, i.e. income or total expenditures. More on this point in Section 2.

taxonomies of commodity categories.

We begin with a descriptive analysis aimed at empirically exploring the stability of household budget share distributions over time. Estimated sample moments show that the shape of the household budget share distribution of each given commodity category did not dramatically change over the time interval considered. However, for any given wave, there emerges a lot of across-commodity heterogeneity in the observed shapes of household budget share distributions. We also show that the underlying household consumption expenditures distributions —for any given wave and commodity category— are well-proxied by log-normal distributions (with very different parameters). We then derive an original family of densities, defined over the unit interval, which is consistent with the detected log-normality of household consumption expenditures distributions. The precise formulation of the closed-form density can be shown to depend on the chosen approximation for the random variable defined as the sum of (possibly correlated) log-normal distributions. In the literature there exist two possible approximations, namely the log-normal and the inverse-Gamma, which we both fit to our data. To benchmark our results, we also fit household budget share distributions with Beta variates, which are in principle very flexible densities defined over the unit interval but lack any consistency with the shape of the random variables which household budget shares stem from.

We find that in Italy, for all the waves under study and for all the commodity categories, the proposed density family —using either approximation— outperforms the Beta in fitting observed household budget share distributions for the majority of cases. Indeed, according to simple measures of goodness-of-fit (e.g., the average absolute deviation), the proposed density family is able to better accommodate the existing shape-heterogeneity that characterizes household budget share distributions across different commodity categories. Furthermore, the estimated parameters of the proposed density allow to reproduce an economically-meaningful taxonomy of commodity categories, which interestingly maps into the traditional classification of commodities among necessary, luxury or inferior goods.

The paper is structured as follows. In Section 2 we describe the database that we employ in the analysis and we discuss some methodological issues. Section 3 presents a preliminary descriptive analysis of household consumption expenditure and budget share distributions. In Section 4 we derive the proposed family of theoretical densities. Section 5 presents fitting results obtained with that density family, and compares them with Beta variates. Section 6 briefly reports on some interpretations of our exercises in terms of commodity category taxonomies. Finally, Section 7 concludes.

2 Data and Methodology

The empirical analysis below is based on the "Survey of Household Income and Wealth" (SHIW) provided by the Bank of Italy. The SHIW is one of the main sources of information on household income and consumption in Italy. Indeed, the quality of the SHIW is nowadays very similar to that of surveys in other countries like France, Germany and the U.K..[9]

The SHIW was firstly carried out in the 1960s with the goal of gathering data on income and savings of Italian households. Over the years, the survey has been widening its scope. Households are now asked to provide, in addition to income and wealth information, also details on their consumption behavior and even their preferred payment methods. Since then, the SHIW was conducted yearly until 1987 (except for 1985) and every two years thereafter (the survey for 1997 was shifted to 1998).

The present analysis focuses on the period 1989-2004. We therefore have 8 waves. The sample used in the most recent surveys comprises about 8000 households (about 24000 individuals distributed across about 300 Italian municipalities). The sample is representative of the Italian population and is based on a rotating panel targeted at 4000 units. Available information includes data on household demographics (e.g. age of household head, number of household components, geographical area, etc.), disposable income, consumption expenditures, savings, and wealth. In this study, we employ yearly data on (nominal) aggregate, household consumption expenditures and on the following disaggregated commodity categories: nondurable goods (N), durable goods (D), and insurance premia (I). Nondurable goods include also food (F), which we consider as a separate (sub-)category of commodities. According to the definition of the Bank of Italy, expenditures for nondurable goods correspond to all spending on both

[9]SHIW data are regularly published in the Bank's supplements to the Statistical Bulletin and made publicly available online at the URL `http://www.bancaditalia.it/statistiche/indcamp/bilfait`. We refer the reader to Brandolini (1999) for a detailed overview on data quality and main changes in the SHIW sample design, and to Battistin et al. (2003) on the general issues with recall consumption data. Notwithstanding the generally better quality of consumption data based on diaries, the SHIW remains one of the surveys with the highest coverage, carried out for a fairly long time with a high degree of accuracy. Moreover, the literature exploiting SHIW data already comprises more than 300 papers - of which, about 40 study consumption issues.

food and non-food items, excluding expenses for durable goods and insurance, maintenance, mortgage and rent payments. The expenditures for food include spending on food products in shops and supermarkets, and spending on meals eaten regularly outside home.[10] Household expenditures for durable goods correspond to items belonging to the following categories: precious objects, means of transport, furniture, furnishings, household appliances, and sundry articles. Finally, the commodity category labeled as "insurances" includes the following forms of insurance: life insurance, private or supplementary pensions, annuities and other forms of insurance-based saving, casualty insurance, and health insurance policies. A detailed description of the commodity categories under study is provided in the appendix, where we report the corresponding questions from the SHIW questionnaire.

The SHIW database includes a variable recording household aggregate expenditure. This quantity does not correspond to the sum of expenditures of the four consumption categories considered here, but it is obtained by aggregating household expenditure for durable and non-durable goods, plus non monetary income integrations and rent payments.[11] Forms of insurance are not included in the definition of household aggregate expenditure, as indeed these might be considered forms of savings. However, we consider insurances as a commodity category for basically two reasons: (i) insurance forms might also be seen as consumption goods, insomuch as they cover actual expenses which are borne by the household (e.g. for pharmaceutical products); (ii) the insurance itself might be considered —and indeed appears in theoretical models— as a good, which an agent might or might not purchase: the only difference with respect to a traditional consumption good stands in the fact that the degree of risk aversion influences the amount of insurance purchased. The sum of household consumption expenditure for non-durable goods, durable goods and insurances makes up on average 80% of total expenditures. Notice that, as total household consumption is just a proxy for total household resources, it is irrelevant whether total consumption is built by including other commodity categories than the four we focus on.

Household budget shares are computed as ratios of nominal yearly quantities. More formally,

[10]Expenditures for meals eaten regularly outside home are included in the food category only from 1995 on. In order to achieve intertemporal comparability, reported food expenditures for 1989, 1991 and 1993 have been complemented by using an annual index of expenditure for food outside home over total food expenditure - around 0.25 in the considered years - obtained by calculations on data from the Italian statistical office (ISTAT).

[11]If a household dwelling is neither owned nor rented, but occupied in usufruct or free of charge, the total consumption expenditure for that household will include an imputed rent, i.e. an amount corresponding to the rent that could be charged for such a dwelling.

our data structure consists of the distribution of yearly household budget shares defined as

$$B_i^{h,t} = \frac{C_i^{h,t}}{C^{h,t}}, \tag{1}$$

where $t \in T$={1989, 1991, 1993, 1995, 1998, 2000, 2002, 2004} are survey waves, $i \in I$={N, F, D, I} are the four commodity categories, $C_i^{h,t}$ is the (nominal) consumption expenditure of household $h = 1, \ldots, H_t$ for the commodity category i, and $C^{h,t}$ is the (nominal) total consumption expenditure of household h. All household consumption expenditure observations have been preliminary weighted using appropriate sample weights provided by the Bank of Italy. Outliers —defined as observations greater than 10 standard deviations from the mean— have been removed. Since in each wave there were some cases of unrealistic (e.g., zero or negative) aggregate consumption expenditure figures, we dropped such observations and we kept only strictly-positive ones. We also dropped households for which yearly expenditures for at least one commodity was larger or equal to total expenditure (as reported in the SHIW). Since we rule out borrowing, $B_i^{h,t} \in (0, 1)$. Finally, we excluded zero observations from the analysis, for the following reasons: (i) it is not clear whether zero entries for nondurable goods mean a null consumption expenditure or rather they are due to mistakes in data collection; (ii) the decision whether to buy a durable good or an insurance or not (depending for example on whether household income exceeds some threshold) is different from and precedes the decision on the budget share possibly allocated to this good. We focus on the second step of the decisional process; (iii) the degree of bunching at zero one typically finds in the distributions of durable goods and insurance expenditures is influenced by factors which are likely to vary over the business cycle; (iv) sample moments computed including zeroes would be poorly informative, given the large proportion of zero observations. Therefore, we ended up with a changing (but still very large) number of households in each wave H_t (see Table 3).

Two important points deserve to be discussed. First, we use total expenditures instead of income to proxy household total resources and compute budget shares. This is primarily done in order to separate the problem of allocating total consumption to various commodities from the decision of how much to save out of current income. Notice that this is common practice in the relevant literature. Indeed, due to the relatively higher reliability of expenditure data (as compared to income ones), most of empirical studies typically use household consumption expenditures even if theoretical models are originally developed in terms of total income (see, e.g., Banks et al., 1997). Since income is available in the SHIW database, we replicated our exercises by defining household budget shares in terms of household-income ratios without any appreciable differences in the results as far as descriptive analyses were concerned. Quantile-quantile plots comparing household income and total consumption expenditure distributions, reported in Figure 1 for wave 2004, show that as income increases the proportion of income devoted to total consumption increases.[12]

Second, as already mentioned, this study is not explicitly concerned with the estimation of Engel curves, either with parametric or non-parametric approaches (Engel and Kneip, 1996; Chai and Moneta, 2008). Conversely, we treat household budget shares as agnostic variables that have an economic meaning 'per se'. Moreover, note that Engel curves describe the relationship between conditional averages of household consumption expenditures (or budget shares) for a particular commodity category and levels of income or total consumption expenditure, where averages are computed conditional to levels of income or total consumption expenditure, and possibly other explanatory variables. In this paper, we begin instead to study the statistical properties of unconditional household budget share distributions, that is —for any commodity category and wave— we pool together households irrespective of their income or total consumption expenditure, and we consequently study the shape of the ensuing distributions and their dynamics. In other words, we do not compress the overall across-household heterogeneity existing for each commodity category and wave, as done in Engel-curve studies.[13] This is because the goal of the paper is simply to characterize the distributional shape of *unconditional* household budget share distributions and not how they change with household total budget.

As we briefly recall in the concluding Section, two possible extensions of this work come easily to the mind. In the first place, one might condition household budget share distributions, for each commodity category and wave, to total household resources and investigate how conditional household budget share distributions change as income or total consumption expenditure

[12]Data, scripts, and additional results are available from the Authors upon request. All statistical exercises were performed using MATLAB®, version 7.4.0.287 (R2007a).

[13]If one plots the cloud of points $(C_i^{h,t}, C^{h,t})$ for a given (i,t) in the expenditure-budget plane, the budget share of household h is simply the slope of the line connecting the origin of the plane with the point $(C_i^{h,t}, C^{h,t})$. Engel-curve exercises try to fit this cloud of points with some conditional-expectation relation of the form $E(C_i^{h,t}|C^{h,t}, \cdots)$ and study the shape of this object as a function of $C^{h,t}$ (or income). This paper focuses instead on the distributional properties of the whole cloud of points, and how these properties change across different (i,t).

increase.[14] Furthermore, one might extend the univariate approach employed here to a multivariate perspective. Indeed, the present study of household budget share distributions can be considered as a first approximation to the more thorough analysis of the statistical properties of the multivariate distribution $(C_1^{h,t}, \ldots, C_K^{h,t})$, where K is the overall number of commodity categories which total household consumption expenditures are disaggregated among. Given that we do not have data for all the K components of the multivariate variable, we employ here a univariate approach. Notice, however, that at least in principle some information about the missing commodity categories can be recovered from the study of household budget share distributions, as the denominator of household budget shares contains information on the correlations between C_i and all the unobserved components of the multivariate variate $(C_1^{h,t}, \ldots, C_K^{h,t})$.

3 Statistical Properties of Household Budget Share Distributions: Descriptive Analysis

In this Section, we begin with a descriptive analysis of Italian household consumption expenditure and budget share distributions, mainly focused on investigating whether such distributions —and their correlation structure— exhibit structural changes over time.

Let us start with household consumption expenditure distributions. Figure 2 shows kernel density estimates of the logs of household consumption expenditure distributions for waves 1989, 1993, 1998, and 2002.[15] A visual inspection of the four panels indicates that, with the exception of insurance premia, the shape of any given household consumption expenditure distribution is fairly constant over time. This evidence seems to be confirmed by Table 1, where we report estimated sample moments for logged household consumption expenditure distributions, and by Figure 3, which shows their evolution over time. Notice that sample means show a positive trend in time because we are considering nominal quantities. However, insurance expenditure distributions display a more pronounced trend, which is probably due to the observed structural increase in expenditure for insurance premia from the late 90s also in real terms. Note also that, for any given wave, sample moments are fairly similar across commodity categories, meaning

[14] As discussed above, this task is fairly more general than studying Engel curves only.

[15] Here and in what follows we show these four reference waves for the sake of exposition. Similar results hold also for the other waves. Note also that the kernel-density estimator is a non-parametric estimator. Therefore, at this stage, we are not imposing any a-priori parametric assumption on the density of the observed data.

that the across-commodity heterogeneity in the shape of nominal expenditure levels is not that relevant. Furthermore, Table 2 shows sample correlations and p-values for the null hypothesis of no correlation between household consumption expenditure distributions for different commodities in 2004.[16] As expected, the correlations among household consumption expenditure distributions are all strongly positive and significant.

We turn now to a descriptive analysis of household budget share distributions. Figure 4 shows the plots of kernel-density estimates for 1989, 1993, 1998, and 2002.[17] We immediately see that they are fairly stable over time. Conversely, as expected, their shapes differ significantly across commodity categories. Household budget share distributions for nondurable goods and food are relatively bell shaped and a large mass of observations is shifted towards the right of the unit interval. Kernels of durable goods and insurance premia are instead much more right-skewed and monotonically decreasing. Note also that insurance-premia kernels exhibit a relevant irregularity in the right tail, due to a small sample-size problem. The strong across-commodity heterogeneity that clearly emerges in the shape of household budget share distributions suggests that in order to find a unique, parsimonious, parametric statistical model able to satisfactorily fit the data, one would require a very flexible density family.

Estimated sample moments of household budget share distributions are reported in Table 3. On average, 68% of total household expenditures is related to nondurable goods, while food accounts for 33% of the total. Much less is spent on durable goods and insurance premia, as they respectively represent —on average— 13% and 5% of total household consumption expenditures.[18]

Figure 5 plots the time evolution of the first four sample moments of household budget share distributions. In general, moments are relatively stable over time. The exception is represented again by insurance premia, which display highly increasing moments from 1989 on. In particular, skewness and kurtosis exhibit big jumps in 1995, and then move to higher levels. Instead, standard deviation steadily increases from 1989 on. Notice also that skewness signs do not change over time: they are always negative for nondurable goods and always positive

[16]Correlations are computed only for households with non-zero expenditure for all commodity categories. This means using a sample size of about 1000 households for each wave.

[17]Kernel density estimation is performed so as to avoid possible biases due to the fact that budget shares are defined over a bounded set (i.e., the unit interval).

[18]No appreciable differences are found by replacing the denominator of household budget shares with the sum of nondurables, durables and insurance expenditure (i.e., by replacing the variable total expenditure provided by the Bank of Italy with the sum of the four commodity categories of expenditure employed in this study).

for all other categories (see Table 3). All this is good news if the aim is to look for a unique family of probability densities that are able to satisfactorily accommodate the heterogeneity observed across-commodity and over-time. In fact, the main message coming from the foregoing descriptive analysis is that household budget share distributions did not dramatically change their structural properties over time, notwithstanding many households did probably move back and forth across income quantiles. This is a strong result also in light of the introduction of the Euro in 2001.

Furthermore, we turn to study household budget share distributions' correlation structure. Table 4 reports the correlation matrix for 2004, together with the p-values for the null hypothesis of no correlation. Figure 6 plots instead the time evolution of the correlations between the distributions of nondurable goods, durable goods and insurance premia.[19] Note that all correlations are fairly stable over time and exhibit signs consistent with the economic intuition. Indeed, nondurables are negatively correlated with durables —the average correlations being -0.54%— which in turn are negatively correlated with food (here the average correlation coefficient is -0.3%). Negative correlations indicate that when households increase their relative expenditure for durable goods, they tend to reduce their relative expenditure for non-durable goods, including food. Notice also that the correlation between insurance and all other categories is statistically non significant.

Finally, as discussed in Battistin et al. (2007), notice that consumption and income data generally suffer from under reporting (especially in the tails) and outliers, and Italian data are not an exception (Brandolini, 1999). In order to minimize the effect of gross errors and outliers, we have employed robust statistics to estimate the moments of household consumption expenditure and budget share distributions (Huber, 1981). More specifically, we have used median and mean absolute deviation as robust estimators for location and scale parameters. Moreover, we have estimated the third moment with quartile skewness (Groeneveld and Meeden, 1984) and kurtosis using Moors's octile-based robust estimator (Moors, 1988). Results confirm, overall, our previous findings: robust moments for (logged) household consumption expenditure and budget share distributions are stable over time, with the same exceptions found before.

[19] As in the case of household consumption expenditure distributions, the correlation between household budget share distributions are computed only for households with non-zero expenditure for all the commodities (about 1000 households for each wave).

4 A Parametric Model for Budget Share Distributions

The main aim of this work is to determine a parsimonious, parametric, model able to satisfactorily fit household budget share distributions (in statistical terms). We look for a family of densities, defined on the unit interval, which holds at least the following three desirable features. First, the family of densities fitting household budget shares should be consistent with the statistical properties of the underlying household consumption expenditure distributions employed to compute budget shares. Second, it should be flexible enough to accommodate —for each wave— the observed across-commodity heterogeneity in the shape of household budget share distributions. Third, the parameters of the density should embody some economic meaning and allow one to taxonomize commodity categories according to their (high, low) level.

Let us begin with the first point. The existing literature shows that aggregate household consumption expenditure distributions are typically log-normally distributed.[20] Figure 7 indicates that a log-normal density provides reasonable fits also for our household consumption expenditure distributions disaggregated across our four commodity categories. Table 1 confirms this finding, as the logs of disaggregated household consumption expenditure distributions exhibit skewness and kurtosis values very close to what would be expected if the original distributions were log-normal (i.e. 0 and 3 respectively).[21] Robust estimators for the third and fourth moments (see previous section) also support log-normality of household consumption expenditure distributions. In fact, according to standard bootstrap tests, robust skewness and kurtosis of logged household consumption expenditure distributions are often close to their expected values in normal samples (0 and 1.233, respectively).

Let $C_1, \ldots, C_K$ be the expenditure levels of a given household in a representative time period, where K is the number of commodity categories considered.[22] The household budget

[20]Battistin et al. (2007) find that result for U.K and U.S. total household consumption expenditure distributions. Furthermore, in Fagiolo et al. (2009) we show that, as a first approximation, similar evidence is true also for the Italian total household consumption expenditure. We also find, however, that a much better fit can be obtained if one employs a more highly-parameterized density family (i.e., the asymmetric exponential power), which is able to accommodate the existing asymmetry in tail fatness.

[21]As already mentioned, we do not have data for all commodity categories. Therefore, we can not exactly check the assumption that *all* household consumption expenditure distributions disaggregated into commodity categories are log-normally distributed. Nevertheless, further analysis shows that also the distribution of $C^{h,t} - \sum_{i=1}^{4} C_i^{h,t}$ —that is, the remaining average 20% of consumption expenditures— can be reasonably fitted by a log-normal. In what follows, we shall use this fifth composite commodity category in order to keep the identity $C^{h,t} \equiv \sum_{i=1}^{4} C_i^{h,t} + C_5^{h,t}$.

[22]In our case $K = 5$, as we consider four commodity categories plus the composite commodity $C_5^{h,t} = C^{h,t} - \sum_{i=1}^{4} C_i^{h,t}$.

share of commodity category i is defined as

$$B_i = \frac{C_i}{C} = \frac{1}{1 + \frac{\sum_{j \neq i} C_j}{C_i}} = \frac{1}{1 + \sum_{j \neq i} Z_j(i)} = \frac{1}{1 + S_i} \tag{2}$$

where S_i is the sum of the $K-1$ random variables $Z_j(i)$, each being equal to the ratio between C_j and C_i, with $j = 1, \ldots, i-1, i+1, \ldots, K$. Obviously, $B_i \in (0,1)$ as required. From equation (2), it follows that the cumulative distribution function (cdf) of B_i reads:

$$F_{B_i}(x) = \text{Prob}\{B_i < x\} = \text{Prob}\left\{1 + S_i > \frac{1}{x}\right\} = 1 - F_{S_i}\left(\frac{1}{x} - 1\right) \ , \tag{3}$$

where $x \in (0,1)$ and F_{S_i} is the cdf of S_i. Therefore, the probability density function (pdf) of B_i is given by:

$$f_{B_i}(x)\mathrm{d}x = \frac{1}{x^2} f_{S_i}\left(\frac{1}{x} - 1\right)\mathrm{d}x, \tag{4}$$

where f_{S_i} is the pdf of S_i. This means that characterizing the distribution of B_i requires studying the distribution of $S_i = \sum_{j \neq i} C_j / C_i = \sum_{j \neq i} Z_j(i)$. Given the empirical evidence above, there are good reasons to assume that expenditure levels C_i are all log-normally distributed, at least as a first approximation. This implies that the ratios $Z_j(i)$ are also log-normally distributed, as:

$$\text{Prob}\{Z_j(i) < z\} = \text{Prob}\{\log(C_j) - \log(C_i) < \log(z)\} = \text{Prob}\{D_j(i) < \log(z)\}. \tag{5}$$

Since $\log(C_j)$ and $\log(C_i)$ are normally distributed (and possibly correlated), their difference $D_j(i)$ will also be normal. Hence $\exp(D_j(i))$ will be log-normally distributed.[23]

As a result, the shape of household budget share distribution B_i fully depends on the shape of the sum of the $K-1$ log-normal variates $Z_j(i)$s. Notice that in general $Z_j(i)$ will not be uncorrelated. Indeed, the C_is may be correlated because of household preferences. This seems to be the case from our empirical evidence, as we have already noticed statistically-significant correlations between household consumption expenditure distributions (see Table 2).[24] The

[23]More generally, if X and Y are log-normally distributed with parameters (μ_X, σ_X) and (μ_Y, σ_Y), and covariance σ_{XY}, then $D = \log(X) - \log(Y)$ is a $N(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}})$. Thus $X/Y = \exp(D)$ is a log-normal with the same parameters as D.

[24]Another reason why the $Z_j(i)$ may be in general correlated is that the sum of all expenditures cannot exceed household total expenditure. This source of correlation may be washed away, however, by considering only the first $K-1$ commodities.

significant correlation between household consumption expenditure distributions thus implies that $Z_j(i)$ —as well as household budget share distributions— will not be independent.

According to the literature, there does not exist a closed form for the pdf of a sum of log-normal (correlated or uncorrelated) random variables and only approximations are available.[25] The baseline result is that the distribution of S_i can be well approximated by a log-normal distribution, whose parameters depend in a non-trivial way on the parameters of the log-normals to be summed up and their covariance matrix.[26]

The log-normal proxy to the sum of log-normals is not, however, the only approximation available. Indeed Milevski and Posner (1998) show that when $K \to \infty$ then S_i converges in distribution to an an inverse-Gamma (InvΓ) density, which performs well in approximating the sum also for very small K.[27] Therefore there may be some gains in considering an InvΓ proxy to S_i instead of a log-normal one. Of course, the extent to which either approximation is to be preferred is an empirical issue. For this reason, we shall consider both proxies in our empirical application below.

In the case S_i has a log-normal pdf with parameters (m, s), then:

$$f_{S_i}(x; m, s) = \frac{1}{xs\sqrt{2\pi}} \exp\left[-\frac{(\log(x) - m)^2}{2s^2}\right]. \tag{6}$$

Using (4), we get:

$$f_{B_i}(x; m, s) = \frac{1}{x(1-x)s\sqrt{2\pi}} \exp\left[-\frac{(\log(1-x) - \log(x) - \log(m))^2}{2s^2}\right] \tag{7}$$

In what follows we shall refer to density (7) as the LN-B density. Note that the LN-B is already a pdf given that its integral over $[0, 1]$ is one. In Figure 8 we show a variety of shapes derived from (7) for selected values of the parameters m and s. If $m > 0$ ($m < 0$) the distribution is right-skewed (left-skewed), if $m = 0$ it is symmetric. If $0 < s \leq 1.5$ the distribution is bell-shaped, if $1.5 < s \leq 2.5$ it is bimodal, while if $s > 2.5$ it is U-shaped. This seems to confirm that

[25] See Beaulieu et al. (1995) for the case of independent summands and Mehta et al. (2006) for the case of correlated summands.

[26] Many methods are available to find approximations to the parameters of the resulting log-normal distribution, see e.g. Fenton (1960), Schwartz and Yeh (1982), and Safak and Safak (1994). We are not interested here in this issue because we can directly estimate the parameters of the resulting distribution for B_i via maximum likelihood.

[27] The InvΓ random variable is simply defined as the inverse of a Γ random variable, i.e. if $X \sim \Gamma(\eta, \theta^{-1})$ then $X^{-1} \sim \text{Inv}\Gamma(\eta, \theta)$.

despite its parsimony, the density (7) is sufficiently flexible to accommodate different shapes for household budget share distributions.

On the other hand, if we assume an InvΓ approximation for the distribution of a sum of log-Normals, then the distribution of S_i depends on two parameters (θ, p) and its pdf reads:

$$f_{S_i}(x;\theta,p) = \frac{\theta^p}{\Gamma(p)} x^{-p-1} \exp\left[-\frac{\theta}{x}\right] \tag{8}$$

Once again, using (4) we obtain the pdf of B_i (henceforth, InvΓ-B), which reads:

$$f_{B_i}(x;\theta,p) = \frac{\theta^p}{x^2\Gamma(p)} \left(\frac{1}{x} - 1\right)^{-p-1} \exp\left[-\theta\frac{x}{1-x}\right]. \tag{9}$$

Figure 9 shows the shape of the density (9) for selected values of θ and p. We immediately see that (9) is always an asymmetric distribution, as $f_{B_i}(1;\theta,p) = 0$ for any values of the parameters, while if $p > 1$ $f_{B_i}(0;\theta,p) = 0$ but if $p \leq 1$ $f_{B_i}(0;\theta,p) > 0$. The interpretation of the two parameters is less straightforward than in the previous case. Notice that for small values of p the function is monotonically decreasing, while as p increases a rightward-shifting maximum emerges. When $p < \theta$ ($p > \theta$) the maximum is attained for $x < 0.5$ ($x > 0.5$), while if $p = \theta$ the maximum is around $x = 0.5$: this is the most symmetric case we can model with this distribution. Even if the proxy (9) seems to be less flexible than (7), we shall retain it in our fitting exercises for the sake of comparison.

5 Measuring the Goodness of Fit

In the previous section, we have derived two alternative, parsimonious, approximations of budget-share distributions, which appear —at least in principle— flexible enough to accommodate the observed shape heterogeneity and are consistent with the empirically-detected log-normality of household consumption expenditure distributions.

To check how well the foregoing approximations fit the data, we firstly estimate the parameters of (7) and (9) via maximum likelihood. Results are reported in Table 5, together with asymptotic standard deviations for the parameters of LN-B and InvΓ-B. We shall comment parameter estimates in Section 6, where they will be employed to classify the commodity categories under study. In the rest of this Section, we focus instead on goodness-of-fit considerations.

To evaluate the performance of the two proposed proxies in fitting household budget share distributions as compared to alternative distributions, we choose as a benchmark the Beta density (Evans et al., 2000), whose pdf reads:

$$b(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{BE(\alpha,\beta)}, \tag{10}$$

where $x \in [0,1]$ and BE is the Beta function. Notice that the Beta also depends on only two parameters and typically is flexible enough to accommodate many alternative shapes.[28] However, it lacks any consistency requirements with respect to the underlying shape of household consumption expenditure distributions, because in general it cannot be derived as the density of household budget shares stemming from log-normally distributed expenditure levels.

We firstly employ a battery of goodness-of-fit tests based on empirical distribution function statistics. More specifically, we run four widely-used tests: Kolmogorov-Smirnov (KS; Massey, 1951; Owen, 1962), Kuiper (KUI; Kuiper, 1962), Cramér - von Mises (CvM; Pearson and Stephens, 1962) and Quadratic Anderson-Darling (AD2; Anderson and Darling, 1954). Notice that we do not expect the observations to be actually drawn from the parametric density defined in (7) or (9) or (10). We are only interested in understanding whether the theoretical distributions we suggest are able to proxy the empirical data better than plausible alternatives, because of their consistency with the findings above on household consumption expenditure distributions. Furthermore, as our samples are always very large, the unrealistic assumption of empirical data drawn from a theoretical distribution would bring the traditional goodness-of-fit tests, which evaluate the deviation of the empirical CDF from the theoretical one at each single observation, to reject the null hypothesis at any probability level, despite only minor discrepancies.[29] Therefore, we shall evaluate in what follows the goodness of fit of our competing densities after having grouped logged observations among equally-spaced bins and computed test statistics over such bins. In this way, the effect of a discrepancy between theoretical and empirical probability density will strongly affect the statistic only if the same discrepancy is recurring frequently within a particular interval (bin), while the existence of sporadic outliers

[28] An alternative, less parsimonious, benchmark to the Beta is the Generalized Beta, see e.g. Mauldon (1957) and Sepanski and Kong (2007).

[29] See e.g. Bentler and Bonett (1980).

will only have a minor effect on the test statistic.[30]

Table 6 reports the goodness-of-fit statistics only for KS and AD2 tests, as in general all the four tests agree on whether the benchmark is outperformed or underperformed by either specification of the proposed density. According to the statistics for all the four tests, the Beta fit is never the best one. With respect to the LN-B and the InvΓ-B densities, the latter seems to deliver a better fit in the case of nondurable and durable goods (and according to KUI and CvM tests, also in the case of food). However, the levels of the statistics do not take into account the different influences exerted by the sample size on the results of the binning procedure for different distributions. Moreover, a density may perform relatively better than another one even though both provide a very bad description of the empirical sample. In order to perform a more statistically-sound comparison, we have therefore proxied via simulation the distributions of the test statistics when using our procedure of grouping the observations into bins, and we have computed the relevant p-values of the empirical values obtained before, i.e. the probability mass to the right of the observed statistics.[31] The picture given by the p-values is generally consistent with that given by the statistics, however in a dozen cases the p-values for the Beta density are at least as good as those for the LN-B or the InvΓ-B densities, notwithstanding lower values of the statistics. Still, in terms of p-values, the LN-B or the InvΓ-B densities outperform the Beta density in the 75%, 78%, 59% and 78% of the cases according to KS, KUI, CvM and AD2, respectively.

Together with standard goodness-of-fit tests, we employ also the average absolute deviation as a simple measure of goodness-of-fit (see, e.g., Bottazzi et al., 2008, for an economic application). The average absolute deviation represents an alternative, additional, measure of agreement between the empirical and the theoretical frequencies. For any given commodity category i and wave t (labels are suppressed for the sake of simplicity), the average absolute

[30] This procedure also moderates the problems coming from the generation of pseudorandom numbers distributed as in (7) or (9). This is not a trivial task and the issue is one of the main points of our research agenda.

[31] To proxy the distribution of a test statistic for a given theoretical density and a commodity category of household budget shares of size n, we use the following procedure: (i) generate via a bootstrap-with-replacement method a random sample of observations of the same size n as the observed sample, then group the observations into L bins; (ii) on the randomly-extracted sample, re-estimate the parameters of the theoretical frequency by maximum likelihood and compute the test statistic; (iii) repeat this procedure a large number of times m to get the proxy for the distribution of the test statistic. Of course the foregoing steps should be repeated for any given empirical sample, i.e. for any wave and commodity category considered, and for any of the three densities studied. In what follows, we have set $m = 1000$ and we have considered L=100 bins.

deviation is defined as:

$$AAD = \frac{1}{L}\sum_{l=1}^{L} |\phi_{B_i}(x_l) - f_{B_i}(x_l; \bullet, \bullet)|, \tag{11}$$

where L is the number of bins in which we group the empirical observations, and each class is identified by its midpoint x_m, in correspondence of which we compute the empirical frequency ϕ_{B_i} and the theoretical frequency f_{B_i}. The latter is obtained using equations (7), (9), or (10), when parameters are replaced by their maximum-likelihood estimates.

According to the values obtained for the average absolute deviation (see Table 7), in 66% of the cases the Beta distribution is outperformed by either the LN-B or the InvΓ-B density. More precisely, in 34% of the cases the LN-B seems to deliver a better fit, whereas in 28% of the cases the InvΓ-B approximation fits better the data and in one case they have exactly the same performance. In the remaining 34% of the cases either the LN-B or the InvΓ-B density (or even both) show the same goodness-of-fit as the Beta , i.e. the proposed density always manages to do at least as well as the benchmark. According to p-values in Table 7, the Beta distribution fits the data better than the other two densities only in 19% of the cases. Conversely, the LN-B density provides a better fit in 50% of the cases, whereas in 19% of the cases the InvΓ-B approximation wins the competition. In five cases the proposed density in either specification and the benchmark have the same performance, while in the remaining couple of cases the LN-B or the InvΓ-B densities do equally better than the Beta. It is interesting to note that, according to these results, the InvΓ-B provides good fits for nondurable goods budget shares, while the LN-B works better for durable goods and insurance premia budget shares. Food budget shares seem to be well described by either the Beta or the LN-B . Notice also that both average absolute deviations and p-values are often very similar, thus empirically it seems that in some cases all alternatives may provide equally good fits. However, the distributions that we have proposed should be in our view preferred to the Beta because of their statistical consistency with the underlying household consumption expenditure distributions.

A graphical analysis of the goodness-of-fit for two waves (2000 and 2004) is provided in Figures 10 and 11. The LN-B density provides better fits for the left tail of nondurables budget shares and, more generally, for insurances and durables. In these latter cases, however, none of the distributions considered is able to account for the few observations lying on the extreme right of the support. The InvΓ-B performs well only on the right tail of nondurable goods budget share distributions.

6 Towards a Taxonomy of Commodity Categories

The foregoing analysis suggests that the LN-B and InvΓ-B densities are a statistically-satisfactory parametric model for Italian household budget share distributions, one that is able to accommodate the existing heterogeneity in the shape of the distributions and is consistent with the statistical properties of the underlying household consumption expenditure distributions. In this Section, we shall attempt to draw some economic implications stemming from estimated parameters in order to show that the family of density that we have proposed can also be employed to meaningfully classify commodity categories.

To begin with, notice that it is very hard to taxonomize our four commodity categories on the basis of the estimated parameters of household consumption expenditure distribution. Indeed, the sample moments reported in Table 1 are similar for all commodity categories. However, inspection of Table 5 reveals that estimated parameters for LN-B and InvΓ-B —as happened also for sample moments— feature a much higher heterogeneity. This difference between household budget share distributions and consumption expenditure distributions is not surprising, as household budget share distributions contain more information than household consumption expenditure distributions, namely the information about household-budget allocation behavior, which is itself the factor that can allow one to classify the commodity categories.

Therefore, it is tempting to employ the information coming from estimated parameters of both LN-B and InvΓ-B densities in order to build a taxonomy of the four commodity categories. More precisely, we shall employ the study of the shape of the LN-B and InvΓ-B densities performed in Section 4 to classify our commodity categories with respect to the high/low values of their estimated parameters (m, s) and (p, θ). Since these estimates are relatively stable across time (see again Table 5), we shall use averages of estimates across all the waves. As far as the LN-B is concerned, we shall discriminate between commodity categories exhibiting (average) estimates for $m \lesseqgtr 0$ and $s \lesseqgtr 1$, whereas for the InvΓ-B density we will differentiate between commodity categories with (average) estimates for $p \lesseqgtr 1$ and $\theta \lesseqgtr 1$. The two resulting taxonomies are shown in Table 8. Note that durable goods and insurance premia have similar characteristics, i.e. they have low dispersion and are right-skewed, while the budget share distributions of nondurable goods are more disperse and left-skewed. Food budget share distributions are similar to the latter in that are quite disperse, but are right-skewed.

Notice that, although the parameters of both the LN-B and the InvΓ-B densities cannot be easily traced back to the moments of the associated random variables, a clear-cut relation seems to exist between the taxonomies in Table 8 and estimated sample moments of household budget share distributions. Indeed, suppose to classify now commodity categories on the base of estimated sample moments only (i.e., without fitting household budget share distributions with any parametric model). In particular, suppose to focus on estimates of the mean (μ), the median (med), standard deviation (σ), skewness (ξ) and kurtosis (κ). Let us take the number of observations outside the estimated interval $[\mu-\sigma, \mu+\sigma]$ as a measure of dispersion of household budget share distributions: the larger this number the higher the dispersion around the mean. Let us also say that a household budget share distribution has low (high) mean if the latter is lower (higher) than the median. Finally, let us discriminate between left-skewed ($\xi < 0$) and right-skewed ($\xi > 0$) distributions; and call a distribution fat-tailed if $\kappa >> 3$.

Given this setup, one gets the two taxonomies of Table 9. Notice first that apart from the position of non-durables in the right taxonomy (the one involving kurtosis and standard deviation), both taxonomies reproduce the ones obtained using estimated parameters. More specifically, durables and insurances budget share distributions have mean lower than the median ($\mu/med < 1$), low dispersion, they are highly right-skewed ($\xi > 0$) and fat-tailed ($\kappa >> 3$). Non-durable budget share distributions display instead a mean similar to the median ($\mu/med \simeq 1$), are left-skewed ($\xi < 0$), and have thinner, but still thicker than a normal, tails ($\kappa \geq 3$). This taxonomy has a rather interesting economic meaning, related with Engel's classification of commodities. Indeed, we do not expect many extreme observations, and therefore a higher kurtosis, when dealing with the consumption of nondurable goods (more likely to be related to necessary goods), while exceptional events are more common when dealing with durable goods (category which includes luxury goods).

This simple exercise has one main implication. It shows that the proposed density family, in addition to its other appealing properties, can be easily employed —via the evaluation of the estimated parameters— to build classifications of commodity categories, which are also consistent with other taxonomies developed on the basis of estimated sample moments. In our view, the classification built using estimated parameters of LN-B and InvΓ-B densities (Table 8) should be preferred to the one based on sample moments (Table 9) for at least two reasons. First, it is more parsimonious, as it entails the estimation of only two parameters. Second, it is obtained through a statistically-sound parametric model of the whole household budget share distribution, and hence —unlike that based on sample moments— is based on a full description of the sample.

7 Conclusions

In this paper we have explored the statistical properties of household consumption expenditure and budget share distributions for a large sample of Italian households in the period 1989-2004.

A preliminary descriptive analysis has shown that the shapes of such distributions are relatively stable across time but display a lot of across-commodity heterogeneity. We have then derived a family of parsimonious parametric models (densities) for household budget share distributions that are consistent with the statistical properties of observed household consumption expenditure distributions (which household budget share distributions are computed from) and are able to satisfactorily fit the observed data while accommodating the existing shape heterogeneity. Finally, we have shown that the estimated parameters of such densities can be employed to build economically-meaningful taxonomies of commodity categories, which partly map into the well-known Engel's classification of goods into necessary, luxury or inferior.

Given its preliminary nature, the present work allows for many possible extensions. First, the foregoing exercises can be replicated on similar databases of other countries, possibly at different levels of commodity category disaggregation. This may help in assessing the robustness and generality of our findings.

Second, as already discussed in Section 2, one may consider to link more closely the approach pursued here with that employed in Engel-curve-related works (Lewbel, 2008). More specifically, instead of focusing only on unconditional budget share distributions, one might think to study the shape (and the moments) of household budget share distributions conditional to household income or total expenditures, age and cohort of household's head, and other relevant household- or commodity-specific variables. The idea here is to go beyond standard parametric or non-parametric Engel-curve studies and look not only at how the first (and maybe second) moment of such conditional distributions changes with household income or total expenditure, but also at how the whole shape of conditional household budget share distributions is affected by increasing income levels (and across different commodity categories).

Finally, in a similar perspective, the univariate approach employed in this study may be replaced by a multivariate one, where the statistical properties of the K-dimensional household budget share distribution is studied, either parametrically via e.g. multivariate extensions of our LN-B and InvΓ-B approximations, or non parametrically via multivariate kernel analyses. This might allow one to fully incorporate into the study the underlying across-budget share correlation structure, which at the moment is embodied in the estimates of density parameters and cannot be elicited to address issues related to substitution between different commodities.

References

Kuiper, N.H., "Tests concerning random points on a circle", *Proc. Kon. Ned. Akad. van Wet. A* (1962) 63, 38–47.

Anderson, T.W. and Darling, D.A., "A test for goodness-of-fit", *Journal of the American Statistical Association* (1954) 49, 765–769.

Massey, F.J., "The Kolmogorov-Smirnov test for goodness of fit", *Journal of the American Statistical Association* (1951) 46, 68–78.

Owen, D.B., *A Handbook of Statistical Tables*, (Addison-Wesley, Reading, 1962).

Pearson, E.S. and Stephens, M.A., "The goodness-of-fit tests based on Wk and Uk", *Biometrika* (1962) 49, 397–402 (1962).

Aitchison, J. and J. A. C. Brown, "A synthesis of Engel curve theory", *The Review of Economic Studies* (1954) 22, 35–46.

Attanasio, O., "Consumption Demand", in J. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics* (Amsterdam: Elsevier Science, 1999).

Axtell, R., "Zipf Distributions of U.S. Firm Sizes", *Science* (2001) 293, 1818–1820.

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

Battistin, E., R. Blundell and A. Lewbel, "Why is Consumption More Log Normal Than Income? Gibrat's Law Revisited", IFS Working Papers WP08/07, London, U.K., The Institute for Fiscal Studies (2007).

Battistin, E., R. Miniaci and G. Weber, "What Can We Learn From Recall Consumption Data?", *Journal of Human Resources* (2003) 38, 354–385.

Beaulieu, N. C., A. A. Abu-Dayya and P. J. McLane, "Estimating the distribution of a sum of independent lognormal random variables", *IEEE Trans. Commun.* (1995) 43, 2869–2873.

Bentler, P.M. and D.G. Bonett, "Significance tests and goodness of fit in the analysis of covariance structures", *Psychological Bulletin* (1980) 88, 588–606.

Blundell, R., "Consumer Behaviour: Theory and Empirical Evidence–A Survey", *The Economic Journal* (1988) 98, 16–65.

Blundell, R., X. Chen and D. Kristensen, "Semi-nonparametric IV estimation of shape-invariant Engel curves", *Econometrica* (2007) 75, 1613–1699.

Bottazzi, G., G. Dosi, G. Fagiolo and A. Secchi, "Sectoral and Geographical Specificities in the Spatial Structure of Economic Activities", *Structural Change and Economic Dynamics* (2008) 19, 189–202.

Bottazzi, G. and A. Secchi, "Explaining the Distribution of Firm Growth Rates", *RAND Journal of Economics* (2006) 37, 235–256.

Brandolini, A., "The Distribution of Personal Income in Post–War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality", Temi di Discussione (Economics Working Papers), No. 350, Bank of Italy, Rome (1999).

Caselli, F. and J. Ventura, “A Representative Consumer Theory of Distribution”, *American Economic Review* (2000) 90, 909–926.

Chai, A. and A. Moneta, “Satiation, Escaping Satiation, and Structural Change. Some Evidence from the Evolution of Engel Curves”, Max Planck Institute for Economics, Jena, Germany (2008).

Chatterjee, A., S. Yarlagadda and B. Chakrabarti (Eds.), *Econophysics of Wealth Distributions* (Milan: Springer–Verlag Italia, 2005).

Clementi, F. and M. Gallegati, “Pareto’s Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States”, in A. Chatterjee, S. Yarlagadda and B. Chakrabarti (Eds.), *Econophysics of Wealth Distributions* (Milan: Springer–Verlag Italia, 2005).

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

Engel, E., “Die Produktions und Consumptions Verhaltnisse des Konigreichs Sachsen”, *Zeitschrift des Statistisehen Bureaus des Koniglich Sachsischen Ministerium des Innern* (1857) 22.

Engel, J. and A. Kneip, “Recent approaches to estimating Engel curves”, *Journal of Economics* (1996) 63, 187–212.

Evans, M., N. Hastings and B. Peacock, *Statistical Distributions* (New York: Wiley, 2000).

Fagiolo, G., L. Alessi, M. Barigozzi and M. Capasso, “On the distributional properties of household consumption expenditures. The case of Italy”, *Empirical Economics* (2009), forthcoming.

Fagiolo, G., M. Napoletano and A. Roventini, “Are Output Growth–Rate Distributions Fat–Tailed? Some Evidence from OECD Countries”, *Journal of Applied Econometrics* (2008) 23, 639–669.

Fenton, L. F., “The sum of lognormal probability distributions in scatter transmission systems”, *IRE Trans. Commun. Syst.* (1960) CS–8, 57–67.

Forni, M. and M. Lippi, *Aggregation and the Microfoundations of Dynamic Macroeconomics* (Oxford: Oxford University Press, 1997).

Gallegati, M. and A. P. Kirman (eds.), *Beyond the Representative Agent* (Aldershot and Lyme: Edward Elgar, 1999).

Groeneveld, R. and G. Meeden, “Measuring Skewness and Kurtosis”, *The Statistician* (1984) 33, 391–399.

Hartley, J. E., *The Representative Agent in Macroeconomics* (London, New York: Routledge, 1997).

Hendry, D. F., *Econometrics: Alchemy Or Science? : Essays in Econometric Methodology* (Oxford: Oxford University Press, 2000).

Hildenbrand, W., *Market Demand: Theory and Empirical Evidence* (Princeton: Princeton University Press, 1994).

Hohnisch, M., S. Pittnauer and M. Chakrabarty, "Empirical Regularities in Distributions of Individual Consumption Expenditure", *International Journal of Modern Physics C* (2002) 13, 541–549.

Huber, P., *Robust Statistics* (New York: John Wiley Sons, 1981).

Ibragimov, R., "On the robustness of economic models to heavy-taildness assumptions", Bundesbank, November 2005 Conference (2005).

Kaldor, N., "Capital Accumulation and Economic Growth", in F. Lutz and D. Hague, (eds.), *The Theory of Capital* (New York: St. Martin's Press, 1961).

Kirman, A. P., "Whom or What Does the Representative Individual Represent?", *Journal of Economic Perspectives* (1992) 6, 117–136.

Lewbel, A., "Engel Curves" in S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*, Second Edition (London, New York, Palgrave Macmillan, 2008).

Mauldon, J. G., "A generalization of the Beta distribution", *The Annals of Mathematical Statistics* (1957) 30, 509–520.

Mehta, N. B., J. Wu and J. Zhang, "Approximating the sum of correlated lognormal or lognormal-Rice random variables", *IEEE International Conference on Communications (ICC)* (2006) 4, 1605–1610.

Milevski, M. A. and S. E. Posner, "Asian Options, the Sum of Log-normals, and the Reciprocal Gamma Distribution", *Journal of Financial and Quantitative Analysis* (1998) 33, 409–422.

Mizuno, T., M. Toriyama, T. Terano and M. Takayasu, "Pareto law of the expenditure of a person in convenience stores", *Physica A* (2008) 387, 3931–3935.

Moneta, A. and A. Chai, "At the origins of Engel curves estimation", Papers on economics and evolution, no. 0802, Max Planck Institute of Economics, Jena (2005).

Moors, J., "A Quantile Alternative to Kurtosis", *The Statistician* (1988) 37, 25–32.

Pasinetti, L. L., *Structural Change and Economic Growth* (Cambridge: Cambridge University Press, 1981).

Prais, S. J., "Non-linear estimates of the Engel curves", *The Review of Economic Studies* (1952) 20, 87–104.

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

Safak, A. and M. Safak, "Moments of the Sum of Correlated lognormal Random Variables", *Vehicular Technology Conference, IEEE 44th* (1994) 1, 140–144.

Schwartz, S. and Y. Yeh, "On the distribution function and moments of power sums with lognormal components", *Bell Syst. Tech. J.* (1982) 61, 1441–1462.

Sepanski, J. H. and L. Kong, "A family of generalized Beta distributions for income", arXiv: 07104614v1 [stat.ME] (2007).

Appendix

We report here the list of the questions in the 2004 SHIW questionnaire, which correspond to the four commodity categories we study. Questionnaires for previous waves are very close to the 2004 version as regards these four commodity categories.

- Durable consumer goods

  1. During 2004 did you (or your household) buy ...
  2. If yes, what is the total value of the objects bought? (Even if they were not paid for completely)

     – precious objects (jewelery, old and gold coins, works of art, antiques including antique furniture)
     – means of transport (cars, motorbikes, caravans, motor boats, boats, bicycles)
     – furniture, furnishings, household appliances and sundry articles (furniture, furnishings, carpets, lamps, small household appliances, washing machines, dishwashers, vacuum cleaners, floor polishers, TVs, PCs, fridges, cookers, heaters, air conditioners, radios, tape recorders, CD players, HI-FI equipment, mobile phonesets, fax machines, cameras, camcorders, etc.)

- Nondurable consumer goods

  1. What was the monthly average spending of your household in 2004 on all consumer goods, in cash, by means of credit cards, cheques, Bancomat cards, etc? Consider all spending, on both food and non-food consumption, and exclude only:

     – purchases of precious objects;
     – purchases of cars;
     – purchases of household appliances and furniture;
     – maintenance payments;
     – other contributions received from relatives or friends;
     – extraordinary maintenance of your dwelling;
     – rent for the dwelling;
     – mortgage payments;
     – life insurance premiums;
     – contributions to private pension funds.

- Food

  1. What instead is the monthly average figure for just food consumption? Consider spending on food products in supermarkets and the like and spending on meals eaten regularly outside the home.

- Forms of insurance

  1. – In 2004 did you or another member of your household hold a life insurance policy?
     – In 2004 how many life insurance policies did you, or another member of your household, hold?
     – How much did your household pay in 2004 for each policy?
  2. – In 2004 did you or another member of your household have a private health insurance policy (covering accidents and sickness)?
     – How much did your household pay in 2004 for health insurance policies?

3. – In 2004 did you or another member of your household, individually or with the help of your, his or her employer, pay premiums for a private (or supplementary) pension, an annuity or simply to receive a lump sum in the future (e.g. under children's saving plans)?
   – In 2004 how many private/supplementary pensions, annuities and other forms of insurance-based saving life insurance policies did you, or another member of your household, hold?
   – How much did your household pay in 2004 for each private/supplementary pension?
4. – In 2004 did you or another member of your household pay premiums for a policy or policies covering accidents, theft, fire, hail, third-party liability, etc. (exclude compulsory automobile liability insurance - RCA)?
   – How much did your household pay in 2004 for these premiums?

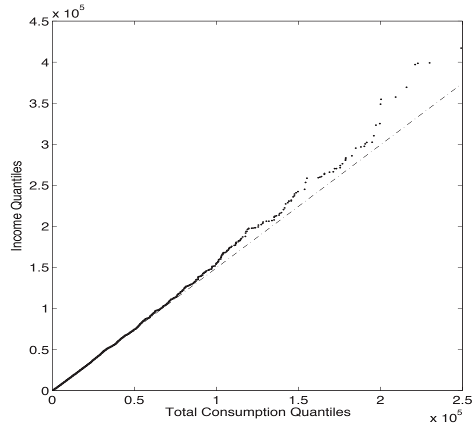

Figure 1: Quantile-quantile plot comparing income and total consumption expenditure distributions. Wave: 2004. Y-Axis: income quantiles; X-Axis: consumption quantiles.

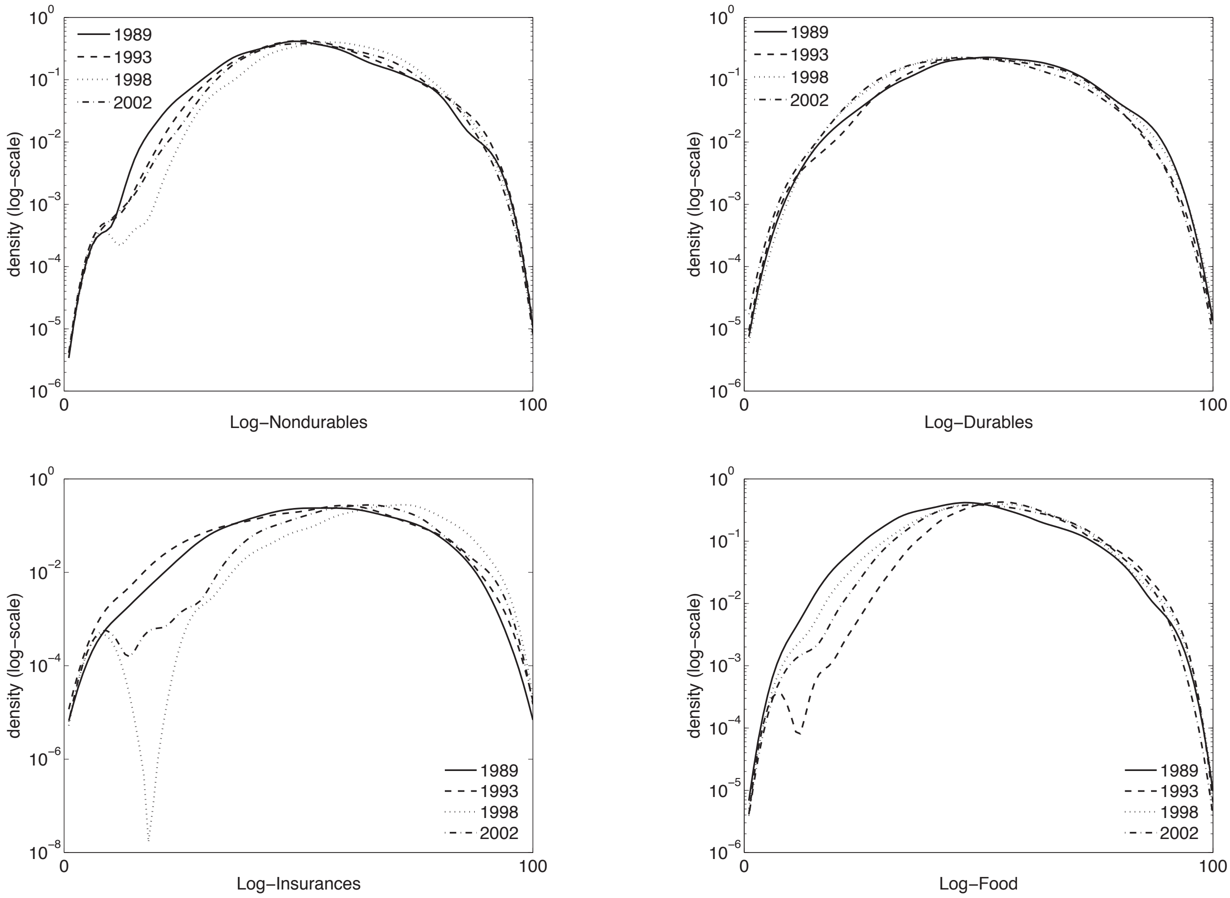

Figure 2: Kernel-density estimates of logged household consumption expenditure distributions and their evolution over time.

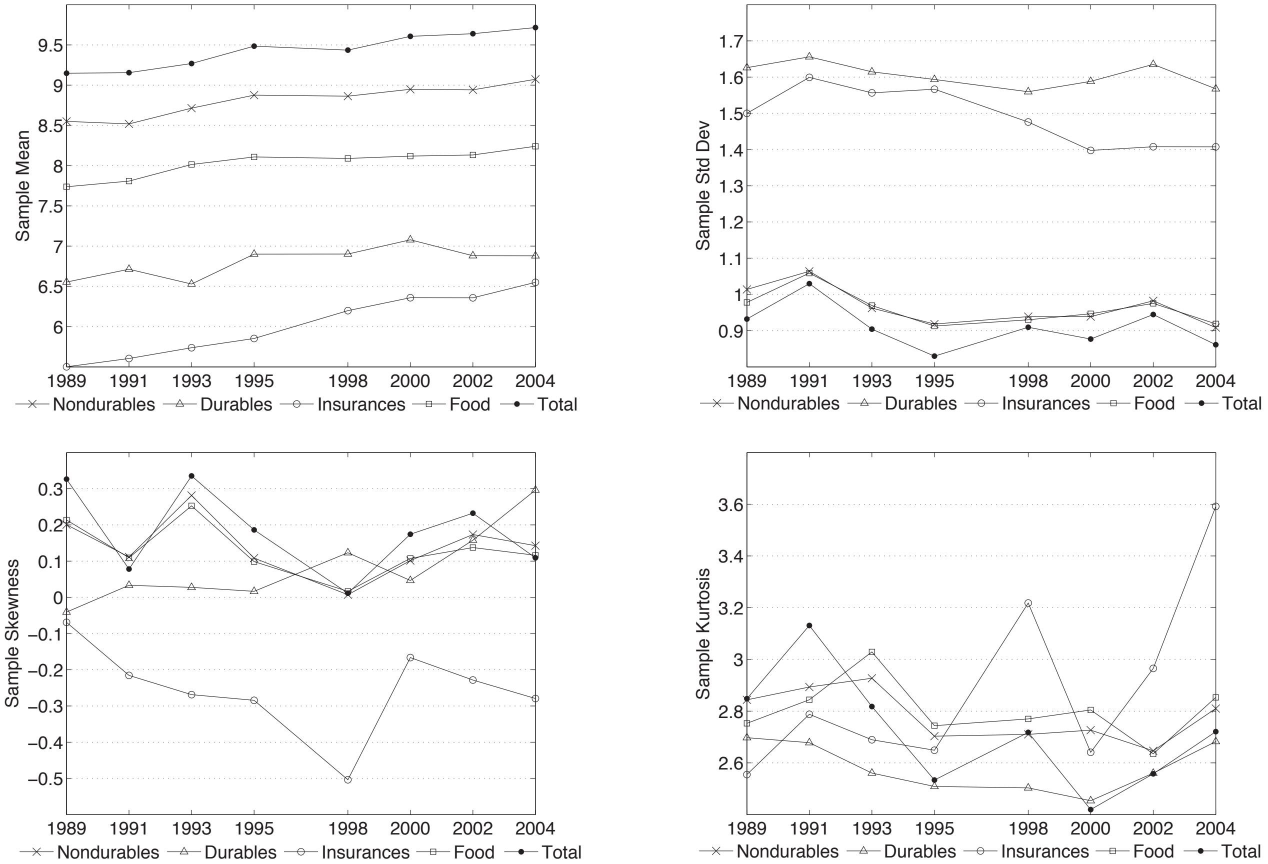

Figure 3: Sample moments of logged household consumption expenditure distributions and their evolution over time.

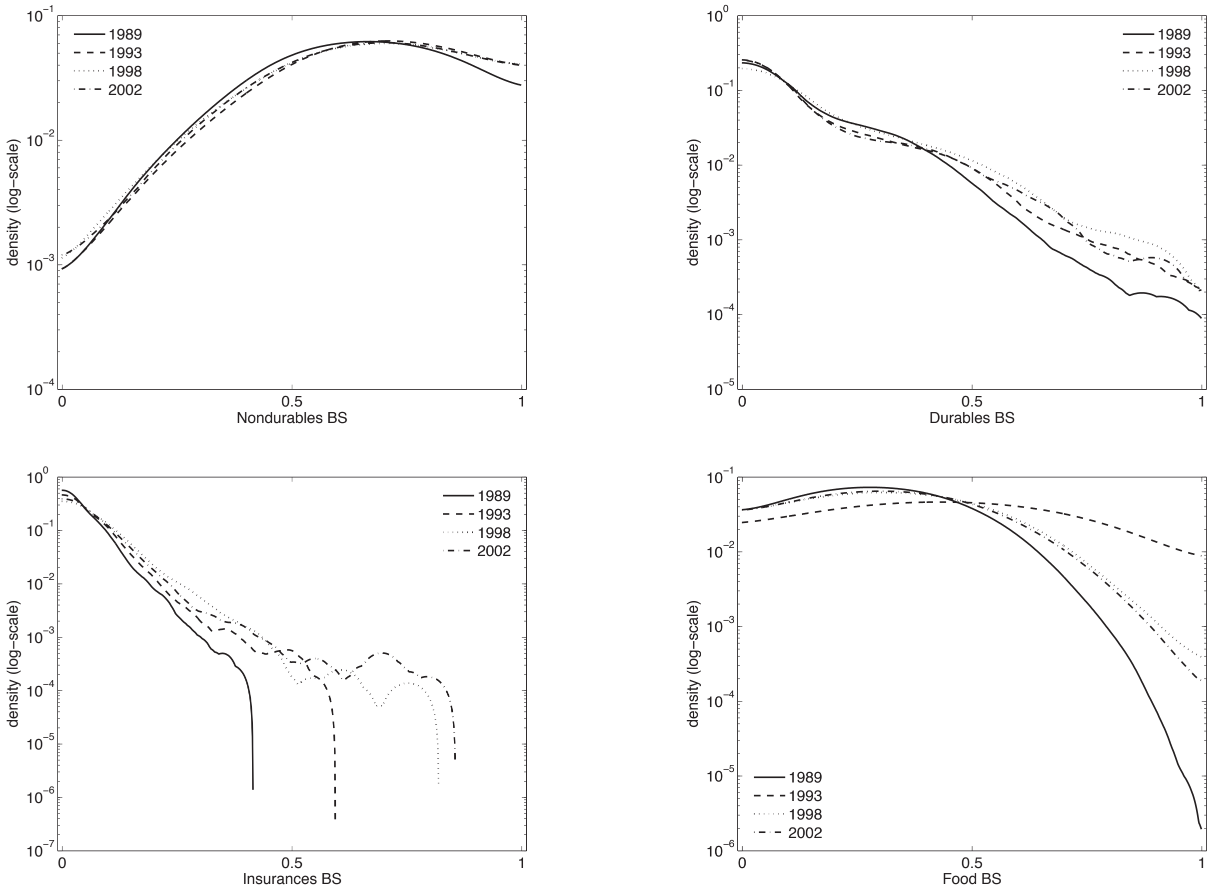

Figure 4: Kernel-density estimates of household budget share distributions and their evolution over time.

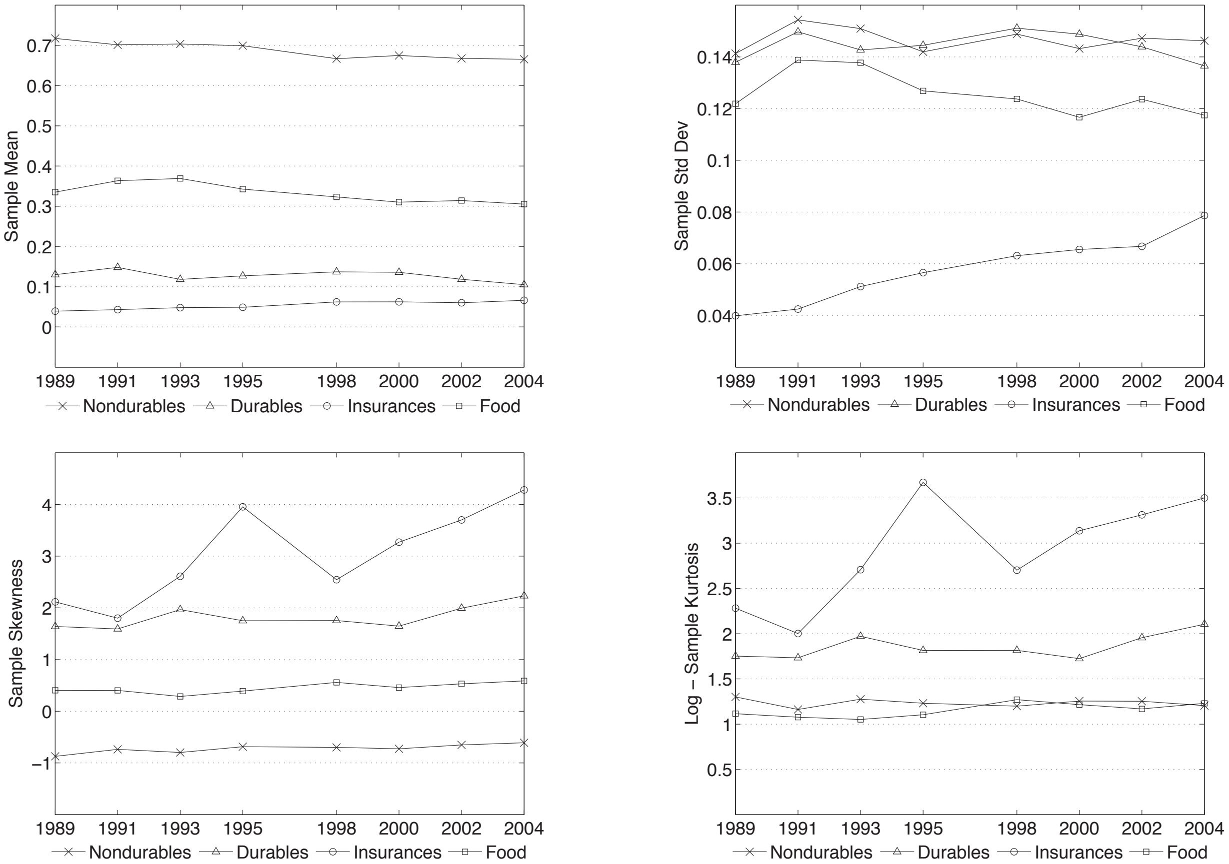

Figure 5: Sample moments of household budget share distributions and their evolution over time.

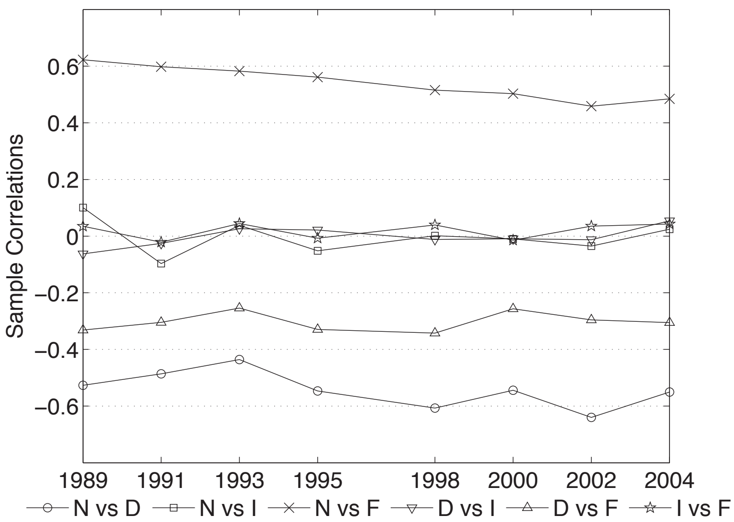

Figure 6: Correlations between household budget share distributions and their evolution over time. N = Nondurables; D = Durables; I = Insurances; F = Food.

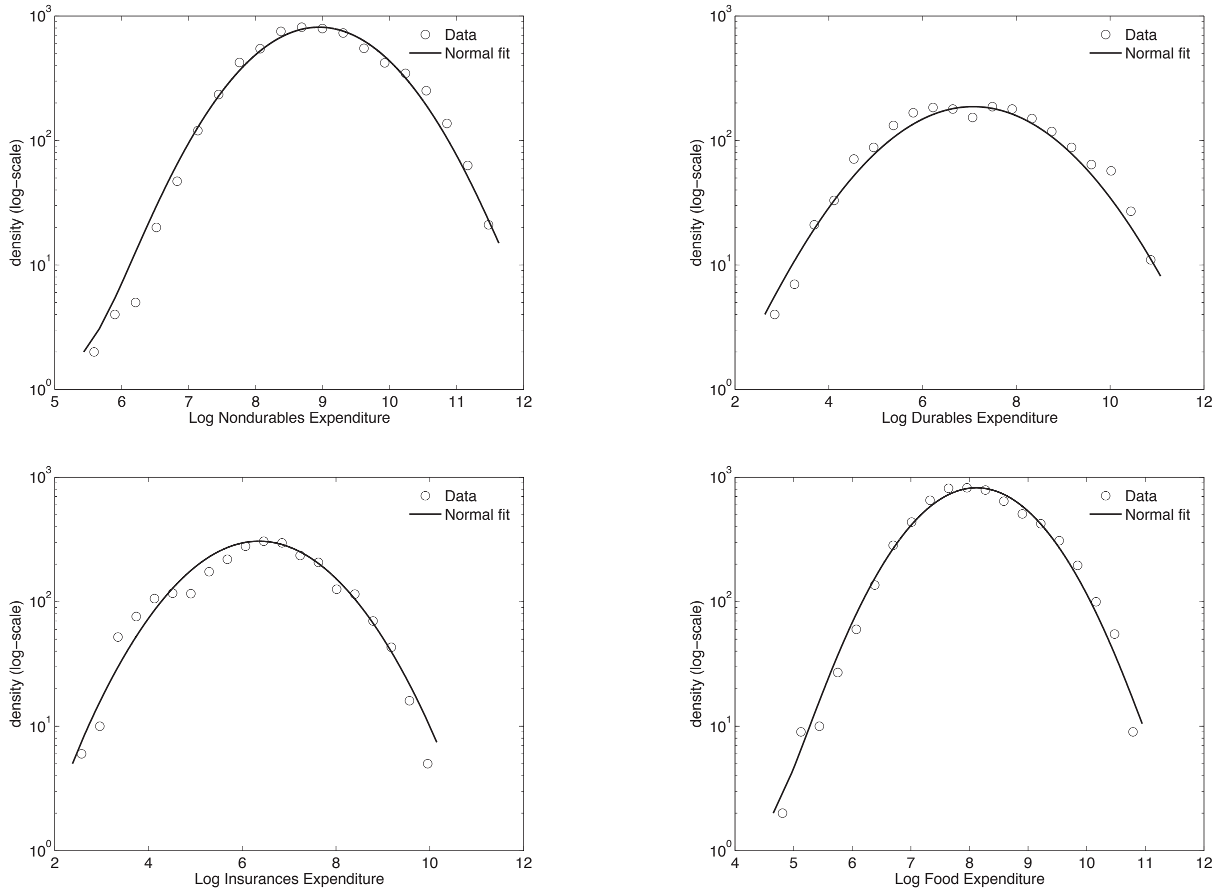

Figure 7: An example of normal fits to logged household consumption expenditure distributions. Wave: 2000.

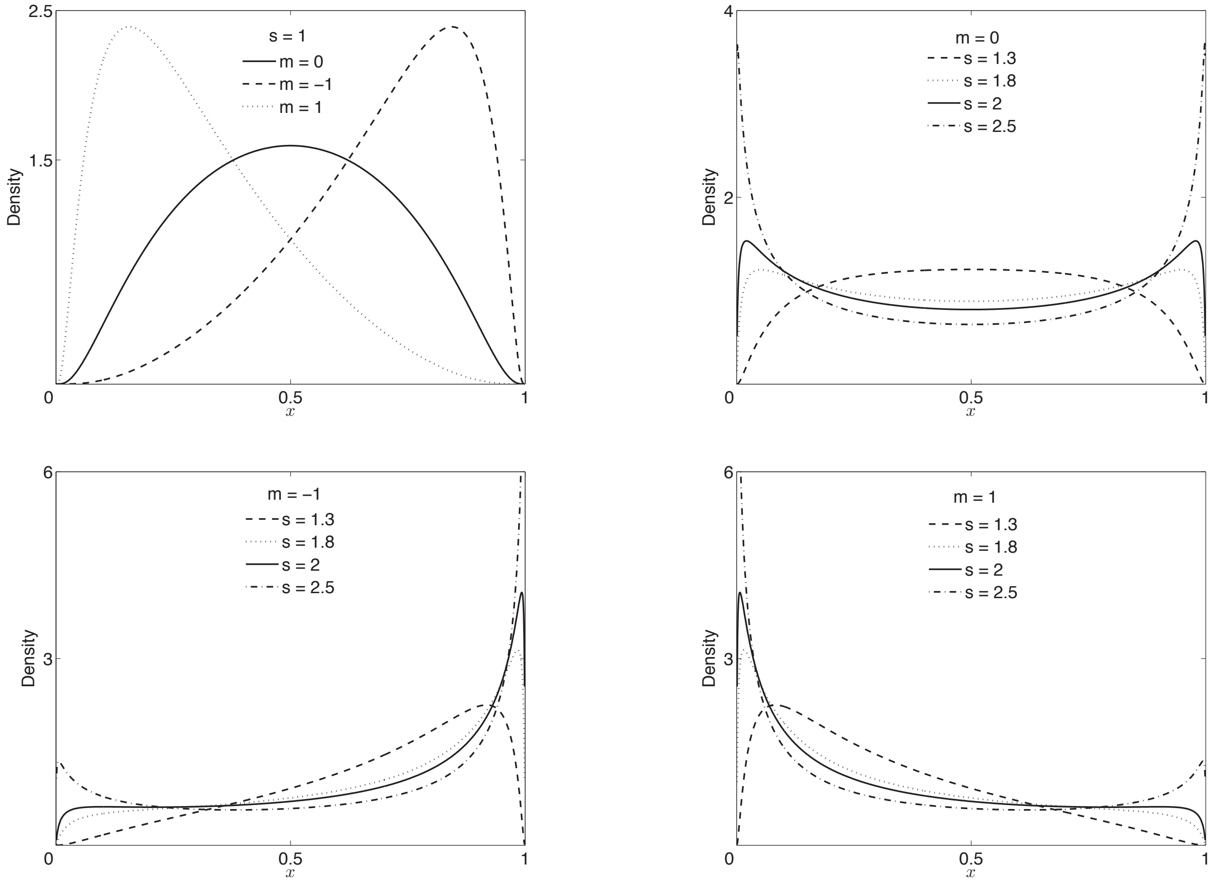

Figure 8: The LN-B approximation to household budget share distributions. Different shapes of f_{B_i} as parameters m and s change.

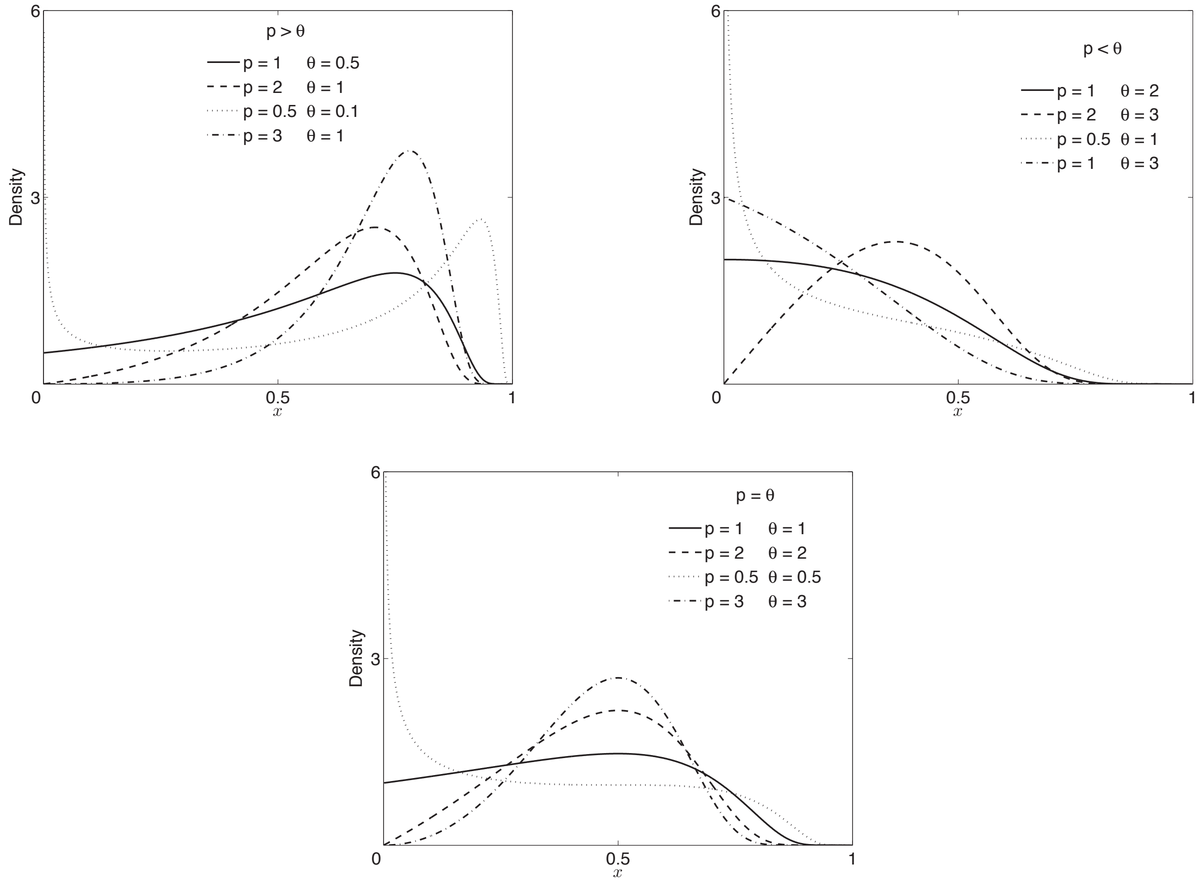

Figure 9: The InvΓ-B approximation to household budget share distributions. Different shapes of f_{B_i} as parameters p and θ change.

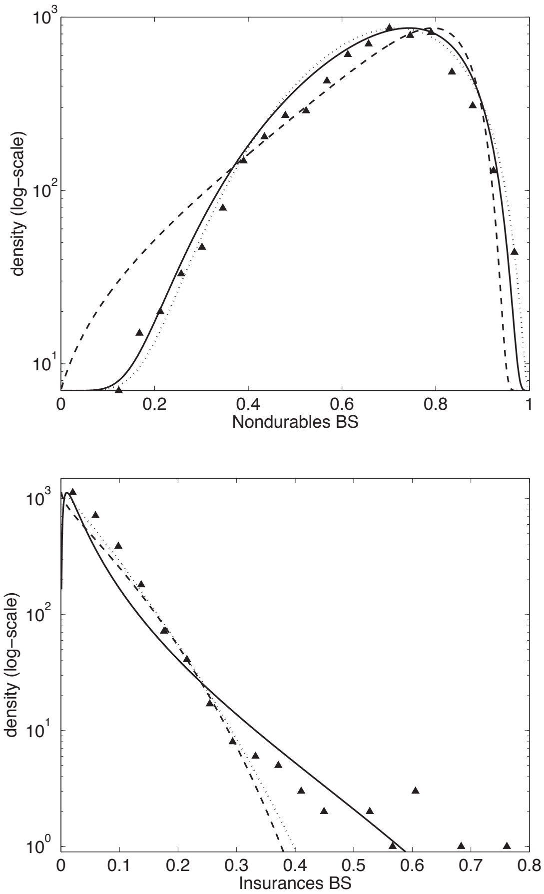

Figure 10: Fitting alternative densities to household budget share distributions. Wave: 2000. Beta: dotted line. LN-B : solid line. InvΓ-B: dashed line.

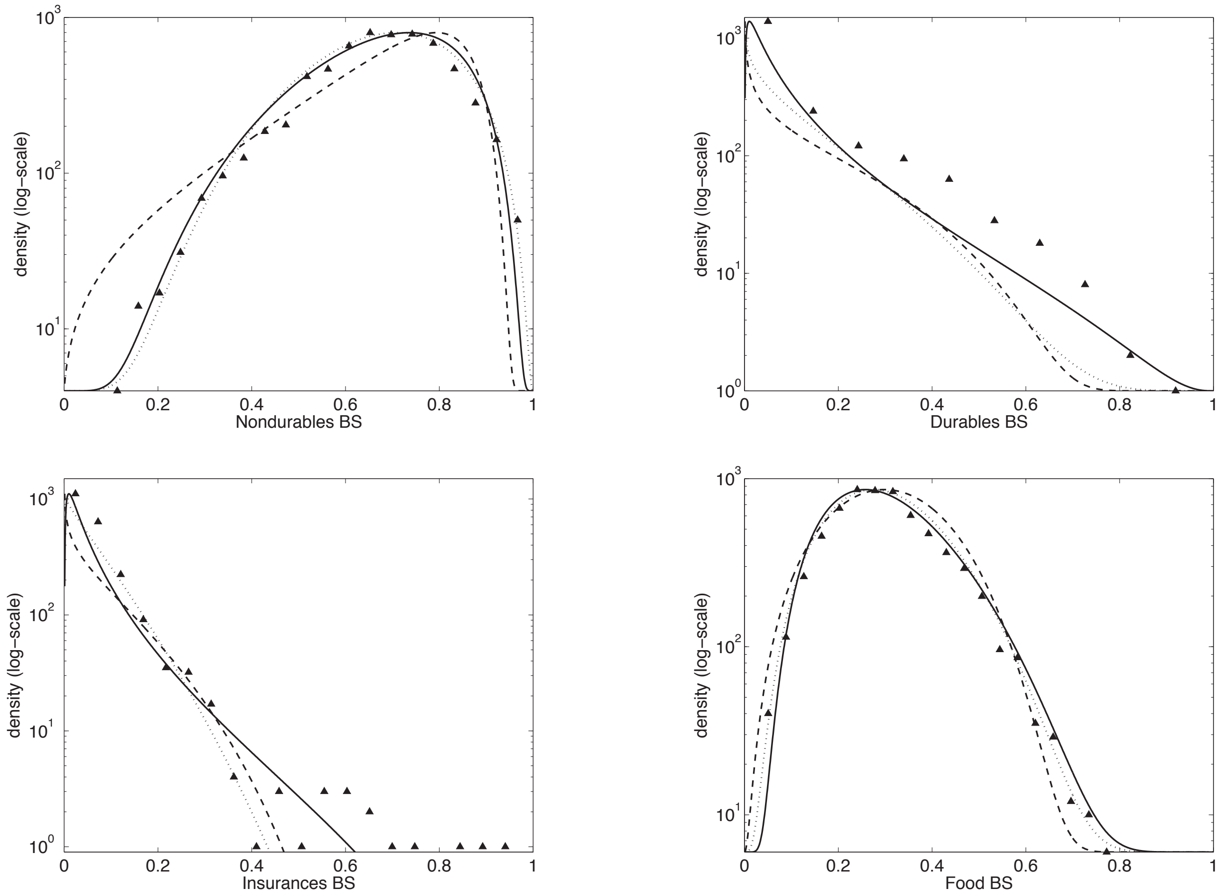

Figure 11: Fitting alternative densities to household budget share distributions. Wave: 2004. Beta: dotted line. LN-B : solid line. InvΓ-B: dashed line.

		Waves								
	Stats	**1989**	**1991**	**1993**	**1995**	**1998**	**2000**	**2002**	**2004**	**Avg**
N	**N Obs**	7409	7209	6223	6258	5588	6277	6361	6281	6451
	Mean	8.551	8.518	8.715	8.876	8.863	8.949	8.942	9.073	8.811
	Std Dev	1.014	1.064	0.962	0.918	0.939	0.939	0.982	0.909	0.966
	Skewness	0.201	0.111	0.281	0.108	0.007	0.102	0.174	0.143	0.141
	Kurtosis	2.844	2.893	2.927	2.703	2.710	2.727	2.647	2.810	2.783
D	**N Obs**	2534	2352	2082	1856	2091	1920	1833	1961	2079
	Mean	6.554	6.713	6.529	6.900	6.902	7.078	6.881	6.879	6.805
	Std Dev	1.626	1.656	1.615	1.593	1.560	1.588	1.635	1.568	1.605
	Skewness	-0.041	0.033	0.028	0.017	0.123	0.047	0.158	0.296	0.083
	Kurtosis	2.698	2.678	2.561	2.509	2.503	2.454	2.560	2.683	2.581
I	**N Obs**	1780	1928	2257	2961	2652	2575	2175	2164	2312
	Mean	5.501	5.604	5.737	5.853	6.198	6.359	6.358	6.551	6.020
	Std Dev	1.500	1.599	1.557	1.567	1.476	1.398	1.408	1.408	1.489
	Skewness	-0.069	-0.216	-0.269	-0.284	-0.504	-0.166	-0.228	-0.279	-0.252
	Kurtosis	2.555	2.788	2.689	2.649	3.218	2.641	2.965	3.592	2.887
F	**N Obs**	7409	7228	6235	6261	5596	6281	6366	6281	6457
	Mean	7.738	7.808	8.014	8.108	8.089	8.119	8.133	8.241	8.031
	Std Dev	0.978	1.059	0.969	0.913	0.930	0.947	0.975	0.919	0.961
	Skewness	0.214	0.109	0.253	0.099	0.017	0.108	0.138	0.116	0.132
	Kurtosis	2.753	2.844	3.029	2.744	2.770	2.805	2.635	2.853	2.804
N+D+I	**N Obs**	896	904	1099	1225	1310	1162	930	1016	1068
	Mean	9.148	9.156	9.269	9.484	9.435	9.607	9.640	9.716	9.432
	Std Dev	0.932	1.030	0.904	0.830	0.909	0.877	0.945	0.861	0.911
	Skewness	0.326	0.078	0.335	0.186	0.011	0.174	0.232	0.109	0.182
	Kurtosis	2.848	3.131	2.818	2.534	2.718	2.420	2.558	2.721	2.718
TC	**N Obs**	7416	7237	6245	6274	5598	6282	6370	6285	6463
	Mean	8.907	8.905	9.093	9.256	9.300	9.369	9.377	9.510	9.215
	Std Dev	1.028	1.095	0.978	0.925	0.934	0.939	0.975	0.908	0.973
	Skewness	0.230	0.150	0.240	0.120	0.077	0.132	0.262	0.236	0.181
	Kurtosis	2.828	2.865	2.855	2.690	2.659	2.704	2.659	2.866	2.766

Table 1: Moments of logged household consumption expenditure distributions vs. waves. Avg: Average values over the whole period. TC = Total Consumption; N = Nondurables; D = Durables; I = Insurances; F = Food. The figures labeled as N+D+I only refer to households with non-zero expenditure for each commodity category.

	N	D	I	F	TC
N	1.00	-	-	-	-
	-	-	-	-	-
D	0.40	1.00	-	-	-
	(0.00)	-	-	-	-
I	0.49	0.39	1.00	-	-
	(0.00)	(0.00)	-	-	-
F	0.87	0.29	0.44	1.00	-
	(0.00)	(0.00)	(0.00)	-	-
TC	0.92	0.58	0.51	0.80	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	-

Table 2: Correlations among household consumption expenditure distributions and p-values (in brackets) for the null hypothesis of no correlation. Wave 2004. TC = Total Consumption; N = Nondurables; D = Durables; I = Insurances; F = Food.

		Waves								
	Stats	**1989**	**1991**	**1993**	**1995**	**1998**	**2000**	**2002**	**2004**	**Avg**
N	**N Obs**	7409	7208	6223	6258	5588	6277	6361	6281	6451
	Mean	0.717	0.702	0.703	0.699	0.667	0.675	0.668	0.666	0.687
	Std Dev	0.141	0.154	0.151	0.142	0.149	0.143	0.147	0.146	0.147
	Skewness	-0.873	-0.736	-0.797	-0.686	-0.698	-0.726	-0.652	-0.610	-0.722
	Kurtosis	3.674	3.199	3.583	3.425	3.317	3.508	3.500	3.339	3.443
D	**N Obs**	2534	2352	2082	1856	2091	1920	1833	1961	2079
	Mean	0.130	0.148	0.118	0.127	0.137	0.136	0.118	0.105	0.127
	Std Dev	0.138	0.150	0.143	0.144	0.151	0.149	0.144	0.137	0.144
	Skewness	1.640	1.591	1.967	1.752	1.755	1.648	1.994	2.233	1.822
	Kurtosis	5.767	5.663	7.182	6.139	6.146	5.610	7.072	8.207	6.473
I	**N Obs**	1780	1928	2257	2961	2652	2575	2175	2164	2312
	Mean	0.039	0.043	0.048	0.049	0.062	0.062	0.060	0.066	0.054
	Std Dev	0.040	0.042	0.051	0.057	0.063	0.066	0.067	0.079	0.058
	Skewness	2.116	1.799	2.609	3.954	2.544	3.270	3.700	4.281	3.034
	Kurtosis	9.798	7.403	14.980	39.336	14.909	23.049	27.494	33.127	21.262
F	**N Obs**	7409	7228	6235	6261	5596	6281	6366	6281	6457
	Mean	0.335	0.364	0.369	0.343	0.323	0.310	0.314	0.305	0.333
	Std Dev	0.122	0.139	0.138	0.127	0.124	0.117	0.124	0.117	0.126
	Skewness	0.405	0.402	0.285	0.389	0.557	0.457	0.530	0.586	0.452
	Kurtosis	3.051	2.938	2.863	3.018	3.563	3.373	3.218	3.423	3.181
N+D+I	**N Obs**	896	904	1099	1225	1310	1162	930	1016	1069
	Mean	0.805	0.790	0.794	0.809	0.815	0.817	0.803	0.798	0.804
	Std Dev	0.153	0.159	0.175	0.150	0.151	0.158	0.142	0.172	0.157
	Skewness	-0.423	-0.129	-0.083	0.125	0.220	0.258	0.129	0.741	0.105
	Kurtosis	4.790	5.211	5.079	5.173	4.884	6.287	4.888	6.382	5.337

Table 3: Moments of household budget share distributions vs. waves. Avg: average values over the whole period. N = Nondurables; D = Durables; I = Insurances; F = Food. The figures labeled as N+D+I only refer to households with non-zero expenditure for each commodity category.

	N	D	I	F
N	1.00	-	-	-
	-	-	-	-
D	-0.55	1.00	-	-
	(0.00)	-	-	-
I	0.02	0.05	1.00	-
	(0.44)	(0.08)	-	-
F	0.48	-0.31	0.04	1.00
	(0.00)	(0.00)	(0.17)	-

Table 4: Correlations among household budget share distributions and p-values (in brackets) for the null hypothesis of no correlation. Wave 2004. N = Nondurables; D = Durables; I = Insurances; F = Food.

		Waves							
LN-B		**1989**	**1991**	**1993**	**1995**	**1998**	**2000**	**2002**	**2004**
N	**m**	-1.04 (0.009)	-0.98 (0.010)	-0.98 (0.010)	-0.95 (0.010)	-0.77 (0.010)	-0.81 (0.009)	-0.78 (0.009)	-0.77 (0.009)
	s	0.76 (0.006)	0.83 (0.007)	0.83 (0.007)	0.77 (0.007)	0.74 (0.007)	0.72 (0.006)	0.75 (0.007)	0.74 (0.007)
D	**m**	2.47 (0.026)	2.28 (0.027)	2.69 (0.031)	2.57 (0.033)	2.43 (0.030)	2.45 (0.031)	2.64 (0.032)	2.80 (0.030)
	s	1.33 (0.019)	1.33 (0.019)	1.42 (0.022)	1.40 (0.023)	1.37 (0.021)	1.37 (0.022)	1.35 (0.022)	1.34 (0.021)
I	**m**	3.74 (0.028)	3.66 (0.028)	3.61 (0.028)	3.61 (0.024)	3.25 (0.024)	3.22 (0.023)	3.26 (0.025)	3.17 (0.026)
	s	1.18 (0.020)	1.24 (0.020)	1.31 (0.020)	1.31 (0.017)	1.25 (0.017)	1.18 (0.016)	1.17 (0.018)	1.20 (0.018)
F	**m**	0.74 (0.007)	0.61 (0.008)	0.59 (0.008)	0.71 (0.008)	0.80 (0.008)	0.86 (0.008)	0.85 (0.008)	0.89 (0.008)
	s	0.59 (0.005)	0.66 (0.005)	0.66 (0.006)	0.61 (0.005)	0.62 (0.006)	0.60 (0.005)	0.62 (0.006)	0.60 (0.005)

		Waves							
InvΓ-B		**1989**	**1991**	**1993**	**1995**	**1998**	**2000**	**2002**	**2004**
N	**p**	1.79 (0.027)	1.45 (0.022)	1.38 (0.022)	1.52 (0.025)	1.93 (0.034)	1.96 (0.032)	1.74 (0.028)	1.81 (0.030)
	θ	0.47 (0.008)	0.37 (0.007)	0.34 (0.007)	0.41 (0.008)	0.67 (0.013)	0.66 (0.012)	0.58 (0.011)	0.62 (0.012)
D	**p**	0.66 (0.016)	0.68 (0.017)	0.57 (0.015)	0.63 (0.017)	0.64 (0.017)	0.64 (0.017)	0.61 (0.017)	0.58 (0.015)
	θ	3.09 (0.106)	2.70 (0.095)	2.78 (0.108)	3.08 (0.124)	2.81 (0.106)	2.83 (0.112)	3.12 (0.127)	3.26 (0.130)
I	**p**	0.99 (0.029)	0.95 (0.027)	0.86 (0.022)	0.78 (0.017)	0.92 (0.022)	0.93 (0.023)	0.91 (0.024)	0.71 (0.018)
	θ	23.14 (0.877)	20.14 (0.738)	15.95 (0.549)	13.37 (0.409)	12.69 (0.398)	12.50 (0.398)	12.46 (0.433)	7.30 (0.265)
F	**p**	3.07 (0.048)	2.38 (0.037)	2.52 (0.043)	2.86 (0.048)	2.73 (0.049)	3.02 (0.051)	2.77 (0.046)	3.00 (0.051)
	θ	5.41 (0.092)	3.51 (0.061)	3.68 (0.069)	4.82 (0.089)	5.00 (0.098)	6.02 (0.111)	5.34 (0.098)	6.09 (0.112)

Table 5: Estimated parameters and asymptotic standard deviations (in parentheses) of LN-B and InvΓ-B vs. waves. Avg: average values over the whole period. N = Nondurables; D = Durables; I = Insurances; F = Food.

	pdf	1989		1991		1993		1995		1998		2000		2002		2004	
		Waves															
N	KS																
	Beta	3.930	(0.36)	3.209	(0.34)	3.874	(0.33)	3.473	(0.36)	3.337	(0.41)	3.340	(0.45)	3.536	(**0.38**)	3.243	(**0.33**)
	LN-B	3.870	(0.36)	3.151	(0.34)	3.779	(0.33)	3.370	(0.37)	3.278	(**0.42**)	3.274	(**0.46**)	3.450	(**0.38**)	3.164	(0.32)
	InvΓ-B	**3.709**	(**0.37**)	**3.043**	(**0.37**)	**3.528**	(**0.35**)	**3.113**	(**0.40**)	**3.136**	(0.41)	**3.103**	(**0.46**)	**3.213**	(0.36)	**2.968**	(0.31)
	AD2																
	Beta	73.01	(0.33)	38.64	(0.33)	72.83	(0.32)	50.94	(0.34)	49.14	(**0.36**)	48.70	(**0.41**	60.60	(0.35)	45.25	(**0.30**)
	LN-B	65.13	(0.34)	34.86	(0.32)	63.86	(0.32)	44.29	(0.35)	45.96	(0.35)	45.04	(0.40)	55.26	(**0.36**)	41.56	(**0.30**)
	InvΓ-B	**42.55**	(**0.35**)	**24.07**	(**0.37**)	**36.05**	(**0.34**)	**26.42**	(**0.40**)	**31.77**	(0.36)	**31.05**	(0.21)	**33.73**	(0.34	**27.48**	(0.28)
D	KS																
	Beta	5.979	(0.26)	5.532	(0.32)	6.055	(0.28)	5.679	(0.41)	5.482	(0.38)	5.729	(0.28)	5.897	(0.36)	6.350	(0.26)
	LN-B	6.092	(**0.28**)	5.656	(0.33)	6.280	(**0.32**)	5.910	(**0.43**)	5.720	(**0.40**)	5.920	(**0.30**)	6.189	(0.38)	6.637	(**0.28**)
	InvΓ-B	**5.524**	(0.19)	**5.092**	(**0.37**)	**5.478**	(0.29)	**5.283**	(0.42)	**5.048**	(0.39)	**5.306**	(0.20	**5.427**	(**0.42**)	**5.786**	(0.18)
	AD2																
	Beta	193.28	(**0.26**)	149.61	(0.30)	189.78	(0.26)	148.09	(0.41)	134.88	(0.36)	161.90	(0.28)	168.69	(0.35)	224.49	(0.25)
	LN-B	150.16	(0.25)	121.19	(0.29)	153.64	(0.29)	124.44	(**0.43**)	116.48	(0.38)	133.49	(**0.29**)	146.18	(0.36)	186.59	(**0.27**)
	InvΓ-B	**130.92**	(0.20)	**112.83**	(**0.39**)	**115.92**	(**0.33**)	**112.10**	(0.43)	**103.96**	(**0.40**)	**117.75**	(0.19	**117.01**	(**0.44**)	**129.91**	(0.18)
I	KS																
	Beta	6.509	(0.31)	6.001	(0.29)	6.891	(0.29)	7.853	(0.31)	7.161	(0.26)	7.362	(0.29)	7.435	(0.26)	7.567	(0.26)
	LN-B	**6.423**	(**0.33**)	**5.950**	(**0.32**)	**6.729**	(**0.31**)	**7.645**	(0.32)	**6.937**	(**0.29**)	**7.197**	(**0.32**)	**7.296**	(**0.30**)	7.452	(0.28)
	InvΓ-B	6.464	(0.31)	5.952	(0.29)	6.822	(0.27)	7.702	(**0.34**)	7.107	(0.23)	7.273	(0.25)	7.312	(0.24)	**7.063**	(**0.48**)
	AD2																
	Beta	212.45	(0.31)	155.52	(0.29)	267.57	(0.27)	560.43	(0.30)	351.57	(0.25)	415.20	(0.28)	436.94	(0.25)	514.84	(0.25)
	LN-B	**139.00**	(**0.32**)	**107.05**	(**0.31**)	**158.93**	(**0.30**)	274.53	(0.32)	**194.88**	(**0.28**)	**234.79**	(**0.29**)	**248.61**	(**0.28**)	286.99	(0.26)
	InvΓ-B	185.45	(0.30)	144.63	(0.29)	204.32	(0.28)	**257.90**	(**0.42**)	250.95	(0.24)	269.22	(0.25)	266.89	(**0.28**)	**215.16**	(**0.57**)
F	KS																
	Beta	3.354	(**0.40**)	3.577	(**0.31**)	3.296	(**0.36**)	3.154	(**0.42**)	4.030	(**0.47**)	4.271	(0.27)	3.618	(**0.33**)	3.400	(0.38)
	LN-B	**3.292**	(0.39)	**3.512**	(**0.31**)	**3.219**	(0.35)	**3.088**	(**0.42**)	**3.962**	(**0.47**)	**4.188**	(**0.28**)	**3.548**	(**0.33**)	**3.337**	(**0.39**)
	InvΓ-B	3.375	(0.38)	3.548	(0.22)	3.329	(0.29)	3.171	(0.41)	3.996	(0.45)	4.288	(0.22)	3.618	(0.31)	3.384	(**0.39**)
	AD2																
	Beta	52.79	(0.33)	69.40	(0.29)	56.04	(0.31)	46.09	(**0.34**)	89.92	(0.42)	108.23	(0.28)	62.48	(**0.31**)	0.31	(**0.38**)
	LN-B	**49.32**	(0.32)	63.35	(**0.43**)	52.59	(0.35)	**43.68**	(0.32)	80.06	(**0.53**)	**90.94**	(**0.31**)	**57.00**	(0.29)	0.29	(0.37)
	InvΓ-B	69.34	(**0.36**)	**51.47**	(0.19)	**45.94**	(**0.52**)	51.93	(0.24)	**74.33**	(0.34)	126.02	(0.20)	66.38	(0.25)	**0.25**	(0.35)

Table 6: Kolmogorov-Smirnov (KS) and Quadratic Anderson Darling (AD2) statistics with p-values (in brackets). N = Nondurables; D = Durables; I = Insurances; F = Food. In boldface the figures that in any given wave and commodity category minimize statistics or maximize p-values.

		Waves															
	pdf	1989		1991		1993		1995		1998		2000		2002		2004	
N	**Beta**	0.004	(0.44)	0.005	(0.86)	0.008	(0.70)	0.010	(**0.26**)	0.010	(**0.46**)	0.015	(0.19)	0.013	(**0.59**)	0.014	(**0.64**)
	LN-B	0.003	(0.43)	0.004	(0.86)	0.007	(0.68)	0.009	(0.25)	0.010	(**0.46**)	0.015	(**0.20**)	0.013	(0.58)	0.014	(**0.64**)
	InvΓ-B	**0.00**	(**0.66**)	**0.003**	(**0.91**)	**0.006**	(**0.73**)	**0.008**	(0.25)	**0.008**	(0.45)	**0.013**	(**0.20**)	**0.011**	(0.58)	**0.012**	(0.62)
D	**Beta**	0.004	(0.88)	0.003	(0.94)	0.005	(0.68)	0.004	(0.77)	0.004	(0.85)	0.004	(0.79)	0.005	(0.66)	0.005	(0.66)
	LN-B	**0.002**	(**0.93**)	**0.002**	(**0.99**)	**0.004**	(0.57)	**0.003**	(**0.87**)	**0.002**	(**0.97**)	**0.003**	(**0.86**)	**0.004**	(**0.70**)	**0.003**	(**0.74**)
	InvΓ-B	0.004	(0.90)	0.003	(0.96)	0.005	(**0.71**)	0.004	(0.82)	0.004	(0.87)	0.004	(0.78)	0.006	(0.65)	0.006	(0.62)
I	**Beta**	0.003	(0.77)	0.004	(0.63)	**0.002**	(**0.95**)	0.003	(0.71)	0.002	(0.94)	**0.003**	(0.76)	0.003	(0.62)	**0.002**	(0.76)
	LN-B	**0.002**	(**0.97**)	**0.003**	(**0.89**)	0.003	(0.87)	**0.002**	(**0.83**)	0.003	(0.81)	**0.003**	(**0.77**)	**0.002**	(**0.92**)	**0.002**	(**0.95**)
	InvΓ-B	0.003	(0.76)	0.004	(0.64)	**0.002**	(0.94)	0.003	(0.69)	**0.001**	(**0.97**)	**0.003**	(**0.77**)	**0.002**	(0.68)	0.003	(0.54)
F	**Beta**	**0.001**	(**1.00**)	**0.002**	(**0.99**)	**0.004**	(**0.78**)	**0.006**	(0.61)	**0.004**	(**0.80**)	**0.005**	(0.74)	**0.005**	(**0.80**)	**0.007**	(**0.53**)
	LN-B	**0.001**	(**1.00**)	**0.002**	(0.98)	**0.004**	(**0.78**)	**0.006**	(**0.62**)	**0.004**	(0.79)	**0.005**	(**0.75**)	**0.005**	(0.79)	**0.007**	(0.52)
	InvΓ-B	0.002	(**1.00**)	0.003	(0.98)	0.005	(0.77)	0.007	(0.60)	0.005	(0.77)	0.006	(0.73)	0.006	(**0.80**)	0.008	(0.52)

Table 7: Average absolute deviation statistics with p-values (in brackets). Average absolute deviations computed by comparing empirical and theoretical pdfs. N = Nondurables; D = Durables; I = Insurances; F = Food. In boldface the figures that in any given wave and commodity category minimize average absolute deviations or maximize p-values.

LN-B	$m > 0$	$m < 0$
$s < 1$	F	N
$s > 1$	D, I	

InvΓ-B	$\theta > 1$	$\theta < 1$
$p > 1$	F	N
$p < 1$	D, I	

Table 8: A taxonomy of household budget share distributions according to the estimated parameters (m, s) and (p, θ). N = Nondurables; D = Durables; I = Insurances; F = Food.

	$\xi > 0$	$\xi < 0$	$\kappa >> 3$ low σ	$\kappa \geq 3$ high σ
$\mu/med \simeq 1$	F	N		N, F
$\mu/med < 1$	D, I		D, I	

Table 9: A taxonomy of household budget share distributions according to estimated sample moments. N = Nondurables; D = Durables; I = Insurances; F = Food. μ = mean; med = median; σ = standard deviation; ξ = skewness; κ = kurtosis.

European Central Bank Working Paper Series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.europa.eu).

1016 "When does lumpy factor adjustment matter for aggregate dynamics?" by S. Fahr and F. Yao, March 2009.

1017 "Optimal prediction pools" by J. Geweke and G. Amisano, March 2009.

1018 "Cross-border mergers and acquisitions: financial and institutional forces" by N. Coeurdacier, R. A. De Santis and A. Aviat, March 2009.

1019 "What drives returns to euro area housing? Evidence from a dynamic dividend-discount model" by P. Hiebert and M. Sydow, March 2009.

1020 "Opting out of the Great Inflation: German monetary policy after the break down of Bretton Woods" by A. Beyer, V. Gaspar, C. Gerberding and O. Issing, March 2009.

1021 "Rigid labour compensation and flexible employment? Firm-level evidence with regard to productivity for Belgium" by C. Fuss and L. Wintr, March 2009.

1022 "Understanding inter-industry wage structures in the euro area" by V. Genre, K. Kohn and D. Momferatou, March 2009.

1023 "Bank loan announcements and borrower stock returns: does bank origin matter?" by S. Ongena and V. Roscovan, March 2009.

1024 "Funding liquidity risk: definition and measurement" by M. Drehmann and K. Nikolaou, March 2009.

1025 "Liquidity risk premia in unsecured interbank money markets" by J. Eisenschmidt and J. Tapking, March 2009.

1026 "Do house price developments spill over across euro area countries? Evidence from a global VAR" by I. Vansteenkiste and P. Hiebert, March 2009.

1027 "Long run evidence on money growth and inflation" by L. Benati, March 2009.

1028 "Large debt financing: syndicated loans versus corporate bonds" by Y. Altunbaș, A. Kara and D. Marqués-Ibáñez, March 2009.

1029 "The role of fiscal transfers for regional economic convergence in Europe" by C. Checherita, C. Nickel and P. Rother, March 2009.

1030 "Forecast evaluation of small nested model sets" by K. Hubrich and K. D. West, March 2009.

1031 "Global roles of currencies" by C. Thimann, March 2009.

1032 "Assessing long-term fiscal developments: a new approach" by A. Afonso, L. Agnello, D. Furceri and R. Sousa, March 2009.

1033 "Fiscal competition over taxes and public inputs: theory and evidence" by S. Hauptmeier, F. Mittermaier and J. Rincke, March 2009.

1034 "The role of the United States in the global economy and its evolution over time" by S. Dées and A. Saint-Guilhem, March 2009.

1035 "The role of labor markets for euro area monetary policy" by K. Christoffel, K. Kuester and T. Linzert, March 2009.

1036 "Search in the product market and the real business cycle" by T. Y. Mathä and O. Pierrard, March 2009.

1037 "What do asset prices have to say about risk appetite and uncertainty?" by G. Bekaert, M. Hoerova and M. Scheicher, March 2009.

1038 "Are 'intrinsic inflation persistence' models structural in the sense of Lucas (1976)?" by L. Benati, March 2009.

1039 "'Real Time' early warning indicators for costly asset price boom/bust cycles: a role for global liquidity" by L. Alessi and C. Detken, March 2009.

1040 "The external and domestic side of macroeconomic adjustment in China" by R. Straub and C. Thimann, March 2009.

1041 "An economic capital integrating credit and interest rate risk in the banking book" by P. Alessandri and M. Drehmann, April 2009.

1042 "The determinants of public deficit volatility" by L. Agnello and R. M. Sousa, April 2009.

1043 "Optimal monetary policy in a model of the credit channel" by F. De Fiore and O. Tristani, April 2009.

1044 "The forecasting power of international yield curve linkages" by M. Modugno and K. Nikolaou, April 2009.

1045 "The term structure of equity premia in an affine arbitrage-free model of bond and stock market dynamics" by W. Lemke and T. Werner, April 2009.

1046 "Productivity shocks and real exchange rates: a reappraisal" by T. A. Peltonen and M. Sager, April 2009.

1047 "The impact of reference norms on inflation persistence when wages are staggered" by M. Knell and A. Stiglbauer, April 2009.

1048 "Downward wage rigidity and optimal steady-state inflation" by G. Fagan and J. Messina, April 2009.

1049 "Labour force participation in the euro area: a cohort based analysis" by A. Balleer, R. Gómez-Salvador and J. Turunen, May 2009.

1050 "Wealth effects on consumption: evidence from the euro area" by R. M. Sousa, May 2009.

1051 "Are more data always better for factor analysis? Results for the euro area, the six largest euro area countries and the UK" by G. Caggiano, G. Kapetanios and V. Labhard, May 2009.

1052 "Bidding behaviour in the ECB's main refinancing operations during the financial crisis" by J. Eisenschmidt, A. Hirsch and T. Linzert, May 2009.

1053 "Inflation dynamics with labour market matching: assessing alternative specifications" by K. Christoffel, J. Costain, G. de Walque, K. Kuester, T. Linzert, S. Millard and O. Pierrard, May 2009.

1054 "Fiscal behaviour in the European Union: rules, fiscal decentralization and government indebtedness" by A. Afonso and S. Hauptmeier, May 2009.

1055 "The impact of extreme weather events on budget balances and implications for fiscal policy" by E. M. Lis and C. Nickel, May 2009.

1056 "The pricing of subprime mortgage risk in good times and bad: evidence from the ABX.HE indices" by I. Fender and M. Scheicher, May 2009.

1057 "Euro area private consumption: Is there a role for housing wealth effects?" by F. Skudelny, May 2009.

1058 "National prices and wage setting in a currency union" by M. Sánchez, May 2009.

1059 "Forecasting the world economy in the short-term" by A. Jakaitiene and S. Dées, June 2009.

1060 "What explains global exchange rate movements during the financial crisis?" by M. Fratzscher, June 2009.

1061 "The distribution of households consumption-expenditure budget shares" by M. Barigozzi, L. Alessi, M. Capasso and G. Fagiolo, June 2009.