



Laboratory of Economics and Management

Sant'Anna School of Advanced Studies

Piazza Martiri della Libertà, 33 - 56127 PISA (Italy)

Tel. +39-050-883-343 Fax +39-050-883-344

Email: lem@sss sup.it Web Page: <http://www.lem.sssup.it/>

# LEM

## Working Paper Series

### **The Distribution of Consumption- Expenditure Budget Shares. Evidence from Italian Households**

Matteo Barigozzi<sup>‡</sup>  
Lucia Alessi<sup>§</sup>  
Marco Capasso<sup>°</sup>  
Giorgio Fagiolo<sup>\*</sup>

<sup>‡</sup> Max Planck Institute of Economics, Jena, Germany

<sup>§</sup> European Central Bank, Frankfurt, Germany

<sup>°</sup> Utrecht University, The Netherlands

<sup>\*</sup> Scuola Superiore Sant'Anna, Italy

**2008/18**

**September 2008**

ISSN (online) 2284-0400

# The Distribution of Consumption-Expenditure Budget Shares. Evidence from Italian Households

Matteo Barigozzi\*    Lucia Alessi†    Marco Capasso‡

Giorgio Fagiolo§

September 2008

## Abstract

This paper explores the statistical properties of household consumption-expenditure budget shares distributions (HBSDs) —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 HBSDs 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 HBSDs 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 HBSDs. 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.

---

\*Sant’Anna School of Advanced Studies, Laboratory of Economics and Management, Pisa, Italy, and Max Planck Institute of Economics, Jena, Germany.

†European Central Bank, Frankfurt, Germany and Sant’Anna School of Advanced Studies, Laboratory of Economics and Management, Pisa, Italy.

‡Urban & Regional research centre Utrecht (URU), Faculty of Geosciences, Utrecht University, and Tjalling C. Koopmans Institute (TKI), Utrecht School of Economics, Utrecht University.

§Corresponding Author. Address: Sant’Anna School of Advanced Studies, Laboratory of Economics and Management, Piazza Martiri della Libertà 33, I-56127 Pisa, Italy. Email: giorgio.fagiolo@sssup.it

# 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 (HBSs henceforth) contain useful information to shed light on this issue. Indeed, the HBS for a given commodity category (CC)  $g$  is defined as the ratio between the expenditure for CC  $g$  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 CC to prices and individual-specific variables as total expenditure or income, household size, head-of-household age, and so on.<sup>1</sup>

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.<sup>2</sup> 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 CC under study.<sup>3</sup>

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.

---

<sup>1</sup>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.

<sup>2</sup>See Aitchison and Brown (1954); Prais (1952); Banks et al. (1997); Blundell et al. (2007) among others.

<sup>3</sup>An Engel curve describes how the expenditure for a given CC 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 (henceforth, HCE) and HBS distributions change over time and between different CCs. 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).<sup>4</sup> 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 CCs. 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.<sup>5</sup> 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 HCEs and HBSs, for which reliable and detailed cross-section data are also available. This is somewhat surprising because —as Attanasio (1999) notices— understanding consumption is crucial to

---

<sup>4</sup>For example, Caselli and Ventura (2000) show that models based on the representative-agent assumption impose almost no restrictions on HCE and HBS 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).

<sup>5</sup>See among others Chatterjee et al. (2005), Clementi and Gallegati (2005), Axtell (2001), Bottazzi and Secchi (2006) and Fagiolo et al. (2008).

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 HCE distributions are well-approximated by log-normal densities (or, as they put it, are “more log-normal than income”).<sup>6</sup> In a complementary paper (Fagiolo et al., 2007), we argue that log-normality is valid only as a first approximation for Italian total HCE 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 HCE aggregate distributions only and nothing is said on the statistical properties of HCE or HBS distributions disaggregated among CCs. 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 HCE and HBS distributions (HCEDs and HBSDs henceforth) for a sequence of 8 waves between 1989 and 2004. We focus on four CCs: nondurable goods, food, durable goods, and insurance premia (which are rarely studied in the literature).<sup>7</sup>

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

We begin with a descriptive analysis aimed at empirically exploring the stability of HBSDs over time. Estimated sample moments show that the shape of the HBSD of each given CC

---

<sup>6</sup>Log-normality of HCE 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.

<sup>7</sup>Food is actually a subcategory of nondurable goods, but for its intrinsic importance we consider it as a separate CC throughout the paper.

<sup>8</sup>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.

did not dramatically change over the time interval considered. However, for any given wave, there emerges a lot of across-CC heterogeneity in the observed shapes of HBSDs. We also show that the underlying HCEDs —for any given wave and CC— 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 HCEDs. 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 HBSDs 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 HBSs stem from.

We find that in Italy, for all the waves under study and for all the CCs, the proposed density family —using either approximation— outperforms the Beta in fitting observed HBSDs 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 HBSDs across different CCs. Furthermore, the estimated parameters of the proposed density allow to reproduce an economically-meaningful taxonomy of CCs, 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 HCEDs and HBSDs. 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 CC 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..<sup>9</sup>

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 CCs: 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 food and non-food items, excluding what falls in the other categories described below and in the following ones: maintenance payments, extraordinary maintenance of household's dwelling, rent for the dwelling, and mortgage payments. The expenditures for food include spending on food products in shops and supermarkets, and spending on meals eaten regularly outside home.<sup>10</sup> Household expenditures for durable goods correspond instead to items belonging to the following categories: precious objects, means of transport, furniture, furnishings, household appliances, and sundry articles.

---

<sup>9</sup>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.

<sup>10</sup>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).

Finally, the CC 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 (excluding compulsory automobile liability insurance), and health insurance policies (accidents and sickness).

The sum of HCE for nondurable goods, durable goods and insurances makes up on average 80% of total expenditures. Therefore, the variable recording household aggregate expenditure available in the database does not correspond to the sum of expenditures on the four CCs considered here, as is obtained by aggregating more CCs. Data for other CCs other than the four that we study are also available, but the sample size is so small that prevents any reliable distribution analysis.

HBSs are computed as ratios of nominal yearly quantities. More formally, our data structure consists of the distribution of yearly HBSs defined as

$$B_i^{h,t} = \frac{C_i^{h,t}}{C^{h,t}}, \quad (1)$$

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 CCs,  $C_i^{h,t}$  is the (nominal) HCE of household  $h = 1, \dots, H_t$  for the CC  $i$ , and  $C^{h,t}$  is the (nominal) total HCE of household  $h$ . All HCE 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,  $B_i^{h,t} \in (0, 1)$ . 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 HBSs. 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 HCEs 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 HBSs in terms of household-income ratios without any appreciable differences in the results as far as descriptive analyses were concerned.<sup>11</sup>

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 HBSs as agnostic variables that have an economic meaning ‘per se’. Moreover, note that Engel curves describe the relationship between conditional averages of HCEs (or HBSs) for a particular CC and levels of income or total HCEs, where averages are computed conditional to levels of income or total HCEs, and possibly other explanatory variables. In this paper, we begin instead to study the statistical properties of unconditional HBSDs, that is —for any CC and wave— we pool together households irrespective of their income or total HCE, 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 CC and wave, as done in Engel-curve studies.<sup>12</sup> This is because the goal of the paper is simply to characterize the distributional shape of *unconditional* HBSDs 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 HBSDs, for each CC and wave, to total household resources and investigate how conditional HBSDs change as income or total HCE increase.<sup>13</sup> Furthermore, one might extend the univariate approach employed here to a multivariate perspective. Indeed, the present study of HBSDs can be considered as a first approximation to the more thorough analysis of the statistical properties of the multivariate distribution  $(C_1^{h,t}, \dots, C_K^{h,t})$ , where  $K$  is the overall number of CCs which total HCEs are disag-

---

<sup>11</sup>Data, scripts, and additional results are available from the Authors upon request. All statistical exercises were performed using MATLAB<sup>®</sup>, version 7.4.0.287 (R2007a).

<sup>12</sup>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 HBS 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  $E(C_i^{h,t} | C^{h,t}, \dots)$  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  $(i, t)$ .

<sup>13</sup>As discussed above, this task is fairly more general than studying Engel curves only.

gregated 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 CCs can be recovered from the study of HBSDs, as the denominator of HBSs contains information on the correlations between  $C_i$  and all the unobserved components of the multivariate variate  $(C_1^{h,t}, \dots, C_K^{h,t})$ .

### 3 Statistical Properties of HBSDs: Descriptive Analysis

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

Let us start with HCEDs. Figure 1 shows kernel density estimates of the logs of HCEDs for waves 1989, 1993, 1998, and 2002.<sup>14</sup> A visual inspection of the four panels indicates that, with the exception of insurance premia, the shape of any given HCED is fairly constant over time. This evidence seems to be confirmed by Table 1, where we report estimated sample moments for logged HCEDs, and by Figure 2, which shows their evolution over time. Notice that sample means show a positive trend in time because we are considering nominal quantities. However, insurance HCEDs 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 CCs, meaning that the across-CC 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 HCEDs for different commodities in 2004.<sup>15</sup> As expected, the correlations among HCEDs are all strongly positive and significant.

We turn now to a descriptive analysis of HBSDs. Figure 3 shows the plots of kernel-density estimates for 1989, 1993, 1998, and 2002.<sup>16</sup> We immediately see that they are fairly stable over time. Conversely, as expected, their shapes differ significantly across CCs. HBSDs

---

<sup>14</sup>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.

<sup>15</sup>Correlations are computed only for households with non-zero expenditure for all CCs. This means using a sample size of about 1000 households for each wave.

<sup>16</sup>Kernel density estimation is performed so as to avoid possible biases due to the fact that BSs are defined over a bounded set (i.e., the unit interval).

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-CC heterogeneity that clearly emerges in the shape of HBSDs 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 HBSDs 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 HCEs.<sup>17</sup>

Figure 4 plots the time evolution of the first four sample moments of HBSDs. 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 for all other HBSDs (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-CC and over-time. In fact, the main message coming from the foregoing descriptive analysis is that HBSDs 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 HBSDs' correlation structure. Table 4 reports the correlation matrix for 2004, together with the p-values for the null hypothesis of no correlation. Figure 5 plots instead the time evolution of the correlations between the distributions of nondurable goods, durable goods and insurance premia.<sup>18</sup> Note that all correlations are fairly stable over time and exhibit signs consistent with the economic intuition. Indeed, nondurables are neg-

---

<sup>17</sup>No appreciable differences are found by replacing the denominator of HBSs with the sum of nondurables, durables and insurance HCE (i.e., by replacing the variable total expenditure provided by the Bank of Italy with the sum of the four CCs of HCEs employed in this study).

<sup>18</sup>As in the HCED case, the correlation between HBSDs are computed only for households with non-zero expenditure for all the commodities (about 1000 households for each wave).

atively 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 nondurable goods, including food. Notice also that the correlation between insurance and all other HBSDs 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 HCEDs and HBSDs (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) HCED and for HBSD are stable over time, with the same exceptions found before.

## 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 HBSDs (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 HBS should be consistent with the statistical properties of the underlying HCEDs employed to compute HBSs. Second, it should be flexible enough to accommodate —for each wave— the observed across-CC heterogeneity in the shape of HBSDs. Third, the parameters of the density should embody some economic meaning and allow one to taxonomize CCs according to their (high, low) level.

Let us begin with the first point. The existing literature shows that aggregate HCEDs are typically log-normally distributed.<sup>19</sup> Figure 6 indicates that a log-normal density provides reasonable fits also for our HCEDs disaggregated across our four CCs. Table 1 confirms this

---

<sup>19</sup>Battistin et al. (2007) find that result for U.K and U.S. total HCEDs. Furthermore, in Fagiolo et al. (2007) we show that, as a first approximation, similar evidence is true also for the Italian total HCE. 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.

finding, as the logs of disaggregated HCEDs 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).<sup>20</sup> Robust estimators for the third and fourth moments (see previous section) also support log-normality of HCEDs. In fact, according to standard bootstrap tests, robust skewness and kurtosis of logged HCEDs are often close to their expected values in normal samples (0 and 1.233, respectively).

Let  $C_1, \dots, C_K$  be the expenditure levels of a given household in a representative time period, where  $K$  is the number of CCs considered.<sup>21</sup> The HBS of CC  $i$  is defined as

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

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  $j = 1, \dots, i - 1, i + 1, \dots, K$ . Obviously,  $B_i \in (0, 1)$  as required. From equation (2), it follows that the cumulative distribution function (cdf) of  $B_i$  reads:

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

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:

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

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:

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

<sup>20</sup>As already mentioned, we do not have data for all CCs. Therefore, we can not exactly check the assumption that *all* HCEDs disaggregated into CCs 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 CC in order to keep the identity  $C^{h,t} \equiv \sum_{i=1}^4 C_i^{h,t} + C_5^{h,t}$ .

<sup>21</sup>In our case  $K = 5$ , as we consider four CC plus the composite commodity  $C_5^{h,t} = C^{h,t} - \sum_{i=1}^4 C_i^{h,t}$ .

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.<sup>22</sup>

As a result, the shape of HBSD  $B_i$  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  $C_i$ s 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 HCEDs (see Table 2).<sup>23</sup> The significant correlation between HCEDs thus implies that  $Z_j(i)$ —as well as HBSDs— 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.<sup>24</sup> 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.<sup>25</sup>

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  $K \rightarrow \infty$  then  $S_i$  converges in distribution to an an inverse-Gamma ( $\text{Inv}\Gamma$ ) density, which performs well in approximating the sum also for very small  $K$ .<sup>26</sup> Therefore there may be some gains in considering an  $\text{Inv}\Gamma$  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:

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

---

<sup>22</sup>More generally, if  $X$  and  $Y$  are log-normally distributed with parameters  $(\mu_X, \sigma_X)$  and  $(\mu_Y, \sigma_Y)$ , and covariance  $\sigma_{XY}$ , 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  $D$ .

<sup>23</sup>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.

<sup>24</sup>See Beaulieu et al. (1995) for the case of independent summands and Mehta et al. (2006) for the case of correlated summands.

<sup>25</sup>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.

<sup>26</sup>The  $\text{Inv}\Gamma$  random variable is simply defined as the inverse of a  $\Gamma$  random variable, i.e. if  $X \sim \Gamma(\eta, \theta^{-1})$  then  $X^{-1} \sim \text{Inv}\Gamma(\eta, \theta)$ .

Using (4), we get:

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

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 7 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 despite its parsimony, the density (7) is sufficiently flexible to accommodate different shapes for HBSDs.

On the other hand, if we assume an  $\text{Inv}\Gamma$  approximation for the distribution of a sum of log-Normals, then the distribution of  $S_i$  depends on two parameters  $(\theta, p)$  and its pdf reads:

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

Once again, using (4) we obtain the pdf of  $B_i$  (henceforth,  $\text{Inv}\Gamma$ -B), which reads:

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

Figure 8 shows the shape of the density (9) for selected values of  $\theta$  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 HCEDs.

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. We shall comment parameter estimates in Section 6, where they will be employed to classify the CCs 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 HBSDs, we choose as a benchmark the Beta density (Evans et al., 2000), whose pdf reads:

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

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.<sup>27</sup> However it lacks any consistency requirements with respect to the underlying shape of HCEDs, because in general it cannot be derived as the density of HBSs stemming from log-normally distributed expenditure levels.

We employ the Average Absolute Deviation (AAD) as a simple measure of goodness-of-fit (see, e.g., Bottazzi et al., 2008). The AAD represents a measure of agreement between the empirical and the theoretical frequencies. For any given CC  $i$  and wave  $t$  (labels are suppressed for the sake of simplicity), the AAD is defined as:

$$AAD = \frac{1}{N} \sum_{m=1}^N |\phi_{B_i}(x_m) - f_{B_i}(x_m; \bullet, \bullet)|, \quad (11)$$

where  $N$  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  $\phi_{B_i}$  and the theoretical frequency  $f_{B_i}$ . The latter is obtained using equations (7), (9), or (10),

---

<sup>27</sup>An alternative, less parsimonious, benchmark to the Beta is the Generalized Beta, see e.g. Mauldon (1957) and Sepanski and Kong (2007).



when parameters are replaced by their maximum-likelihood estimates.

According to the values obtained for the AAD (see Table 6), in 62% of the cases the Beta distribution is outperformed by either the LN-B or the Inv $\Gamma$ -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 Inv $\Gamma$ -B approximation fits better the data. In the remaining 38% of the cases the results are ambiguous and it is not possible to rank the three alternative distributions accordingly to their goodness-of-fit.

So far, the performance of the three densities have been compared one against the other using the empirical AAD levels that they obtain. However, nothing is said on the accuracy with which each single distribution actually fits the data. In other words, 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 the distributions of the AADs via simulation and we have computed the relevant p-values of the empirical values obtained before, i.e., the probability mass to the left of the observed AAD values.<sup>28</sup>

According to p-values in Table 6, the Beta distribution fits the data better than the other two densities only in 25% of the cases. Conversely, the LN-B density provides a better fit in 50% of the cases, whereas in 19% of the cases the Inv $\Gamma$ -B approximation wins the competition. Only in a couple of cases the results are ambiguous.

It is interesting to note that, according to these results, the Inv $\Gamma$ -B provides good fits for nondurable HBSs, while the LN-B works better for durable goods and insurance premia. Food HBSs seem to be well described by either the Beta or the LN-B. Notice also that both AAD 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 HCEDs.

A graphical analysis of the goodness-of-fit for two waves (2000 and 2004) is provided in

---

<sup>28</sup>To proxy the distribution of AAD for a given  $f_B$  and a CC of HBSs 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; (ii) on the randomly-extracted sample, re-estimate the parameters of  $f_B$  by maximum likelihood and compute the AAD; (iii) repeat this procedure a large number of times  $m$  to get the proxy for the distribution of AAD. Of course the foregoing steps should be repeated for any given empirical sample, i.e. for any wave and CC considered, and for any of the three densities studied. In what follows, we have set  $m = 1000$  and we have considered  $N=100$  bins.

Figures 9 and 10. The LN-B density provides better fits for the left tail of nondurables HBSs 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 Inv $\Gamma$ -B performs well only on the right tail of nondurable HBSDs.

## 6 Towards a Taxonomy of Commodity Categories

The foregoing analysis suggests that the LN-B and Inv $\Gamma$ -B densities are a statistically-satisfactory parametric model for Italian HBSDs, 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 HCEDs. 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 CCs.

To begin with, notice that it is very hard to taxonomize our four CCs on the basis of HCED estimated parameters. Indeed, the sample moments reported in Table 1 are similar for all CCs. However, inspection of Table 5 reveals that estimated parameters for LN-B and Inv $\Gamma$ -B—as happened also for sample moments—feature a much higher heterogeneity. This difference between HBSDs and HCEDs is not surprising, as the HBSDs contain more information than the HCED ones, namely the information about household-budget allocation behavior, which is itself the factor that can allow one to discriminate among CCs.

Therefore, it is tempting to employ the information coming from estimated parameters of both LN-B and Inv $\Gamma$ -B densities in order to build a taxonomy of the four CCs. More precisely, we shall employ the study of the shape of the LN-B and Inv $\Gamma$ -B densities performed in Section 4 to classify our CCs with respect to the high/low values of their estimated parameters  $(m, s)$  and  $(p, \theta)$ . 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 CCs exhibiting (average) estimates for  $m \lesssim 0$  and  $s \lesssim 1$ , whereas for the Inv $\Gamma$ -B density we will differentiate between CCs with (average) estimates for  $p \lesssim 1$  and  $\theta \lesssim 1$ . The two resulting taxonomies are shown in Table 7. Note that durable goods and insurance premia have similar characteristics, i.e. they have low dispersion and are right-skewed, while the HBSDs of nondurable goods are more disperse and left-skewed. Food HBSDs are similar

to the latter in that are quite disperse, but are right-skewed.

The taxonomy in Table 7, apart from its statistical soundness, has also a rather interesting economic meaning, related with Engel’s classification of commodities. Indeed, it discriminates between CCs that are more likely to be related to necessary goods (as nondurables) and CCs that are more likely to be related to luxury goods (as durables).

Notice also that, although the parameters of both the LN-B and the Inv $\Gamma$ -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 7 and estimated sample moments of HBSDs. Indeed, suppose to classify now CCs on the base of estimated sample moments only (i.e., without fitting the HBSDs with any parametric model). In particular, suppose to focus on estimates of the mean ( $\mu$ ), the median ( $med$ ), standard deviation ( $\sigma$ ), skewness ( $\xi$ ) and kurtosis ( $\kappa$ ). Let us take the number of observations outside the estimated interval  $[\mu - \sigma, \mu + \sigma]$  as a measure of dispersion of HBSDs: the larger this number the higher the dispersion around the mean. Let us also say that a HBSD 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  $\kappa \gg 3$ .

Given this setup, one gets the two taxonomies of Table 8. 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 HBSDs have mean lower than the median ( $\mu/med < 1$ ), low dispersion, they are highly right-skewed ( $\xi > 0$ ) and fat-tailed ( $\kappa \gg 3$ ). Nondurable HBSDs 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 ( $\kappa \geq 3$ ). These results make economic sense: indeed, we do not expect many exceptional events when dealing with the consumption of nondurable goods, while extreme observations are more common when dealing with durable goods. Finally, food HBSDs display high dispersion, a mean similar to the median, are right-skewed, but have tails which are thinner than durables HBSDs.

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 economically-meaningful classifications of CCs, 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 Inv $\Gamma$ -B densities (Table 7) should be preferred to the one based on sample moments (Table 8) 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 HBSD, 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 HCEDs and HBSDs 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-CC heterogeneity. We have then derived a family of parsimonious parametric models (densities) for HBSDs that are consistent with the statistical properties of observed HCEDs (which HBSDs 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 CCs, 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 CC 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 BS distributions, one might think to study the shape (and the moments) of HBSDs 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

HBSDs is affected by increasing income levels (and across different CCs).

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 HBSD is studied, either parametrically via e.g. multivariate extensions of our LN-B and Inv $\Gamma$ -B approximations, or non parametrically via multivariate kernel analyses. This might allow one to fully incorporate into the study the underlying across-BS 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

- Aitchison, J. and J. A. C. Brown, “A synthesis of Engel curve theory”, *The Review of Economic Studies* 22 (1954), 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.
- 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 Verhältnisse des Königreichs Sachsen”, *Zeitschrift des Statistischen Bureaus des Königlich 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”, LEM Working Paper, 2007-24, Laboratory of Economics and Management, Sant’Anna School of Advanced Studies, Pisa, Italy (2007).
- 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-tailedness 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).

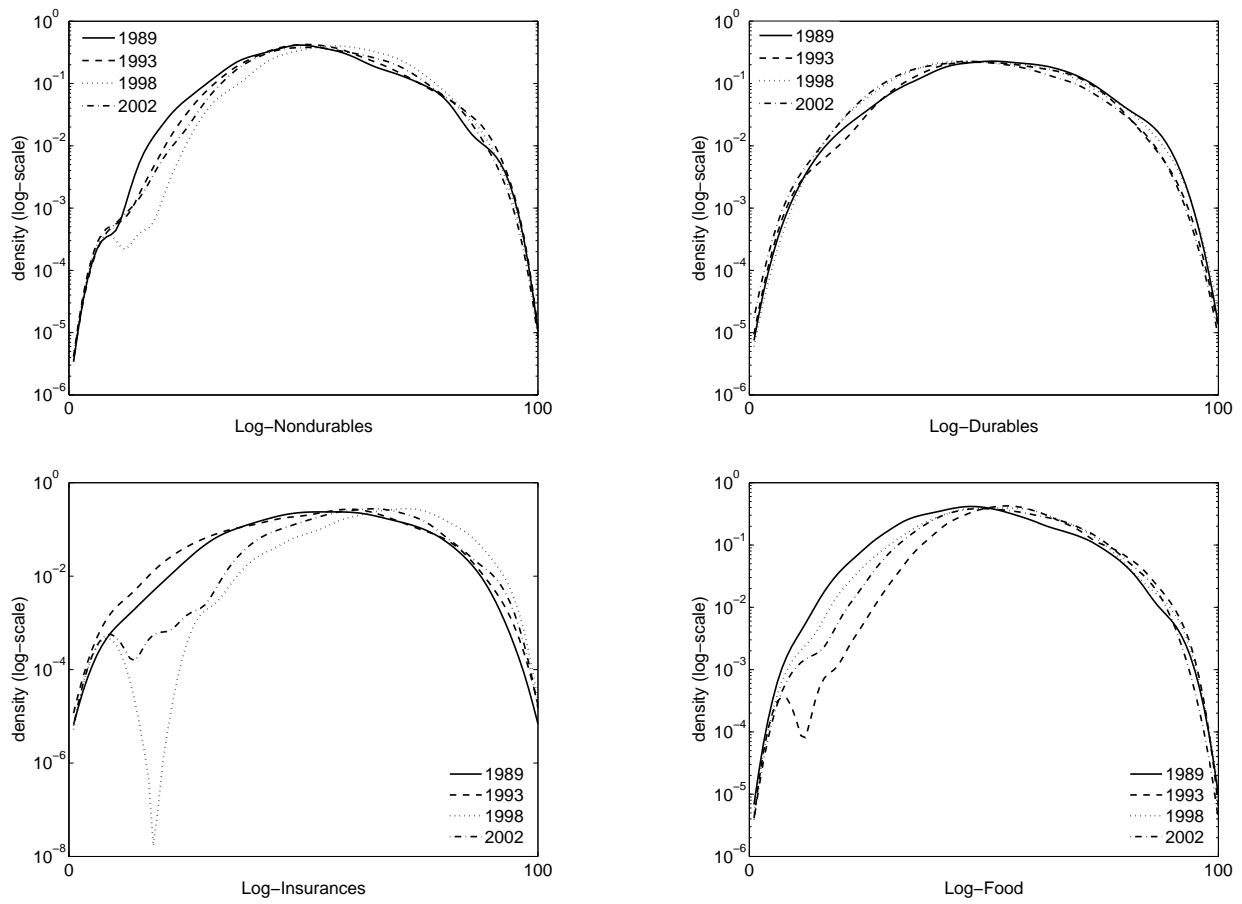


Figure 1: Kernel-density estimates of logged HCE distributions and their evolution over time.



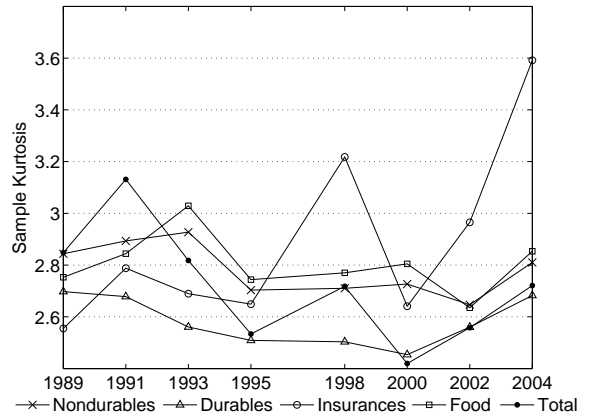
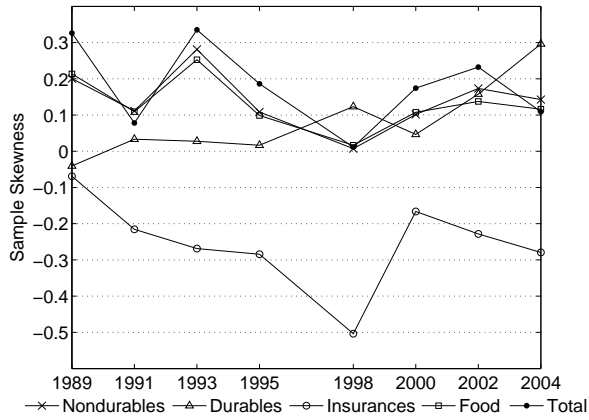
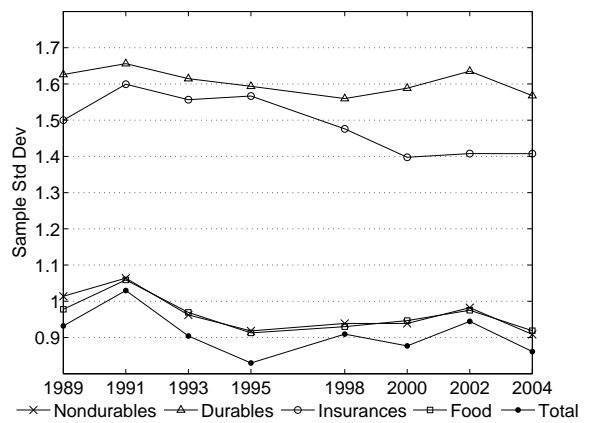
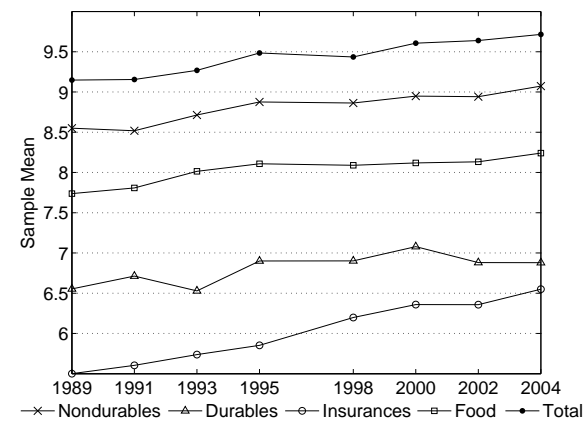


Figure 2: Sample moments of logged HCE distributions and their evolution over time.

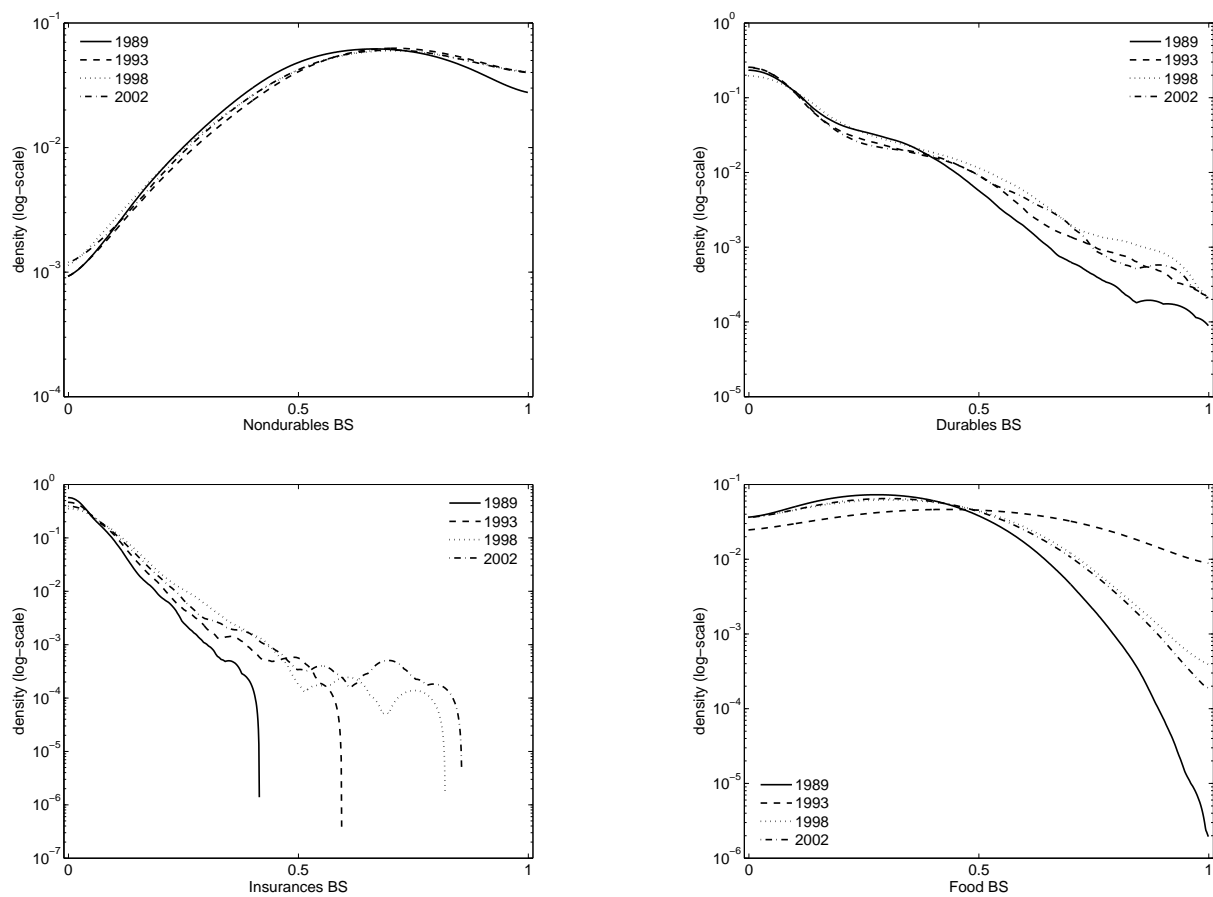


Figure 3: Kernel-density estimates of HBS distributions and their evolution over time.

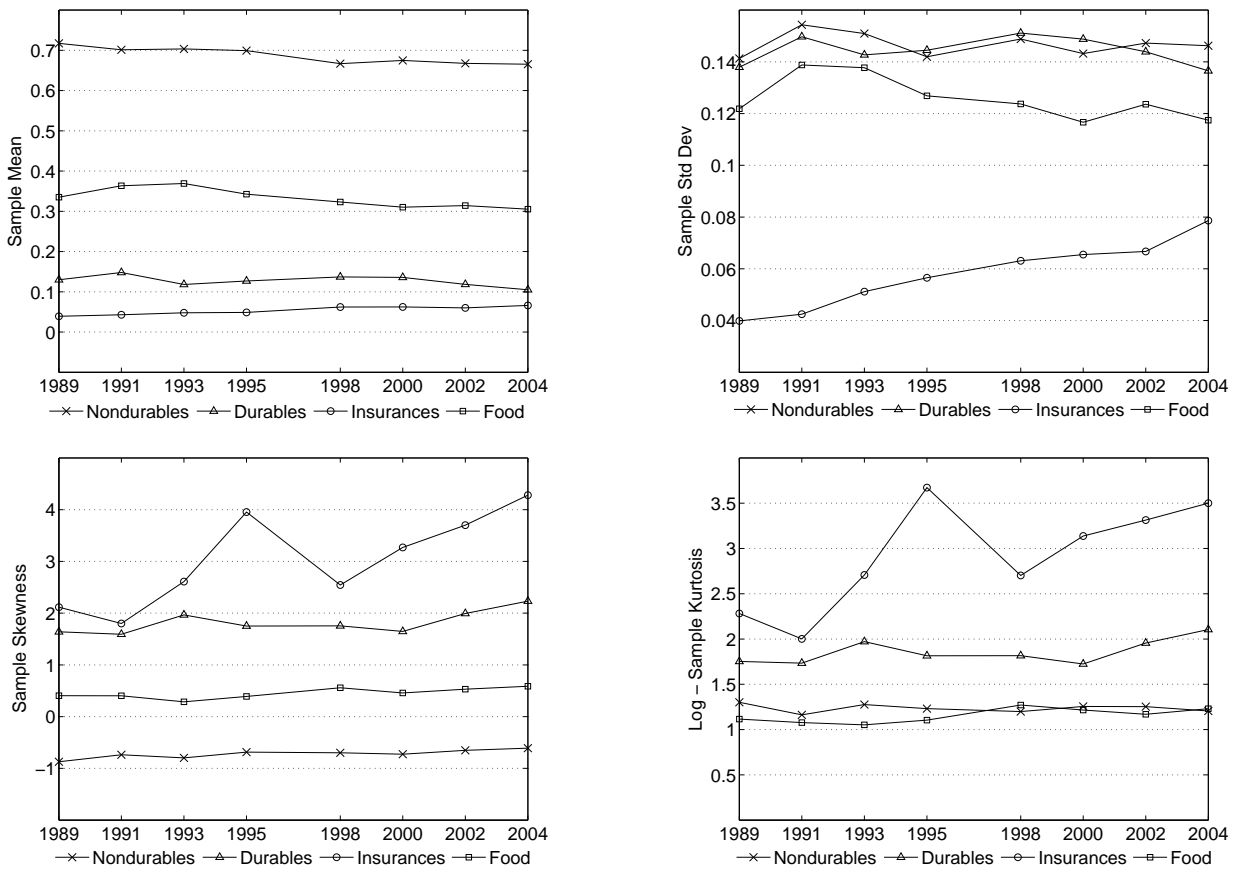


Figure 4: Sample moments of HBS distributions and their evolution over time.

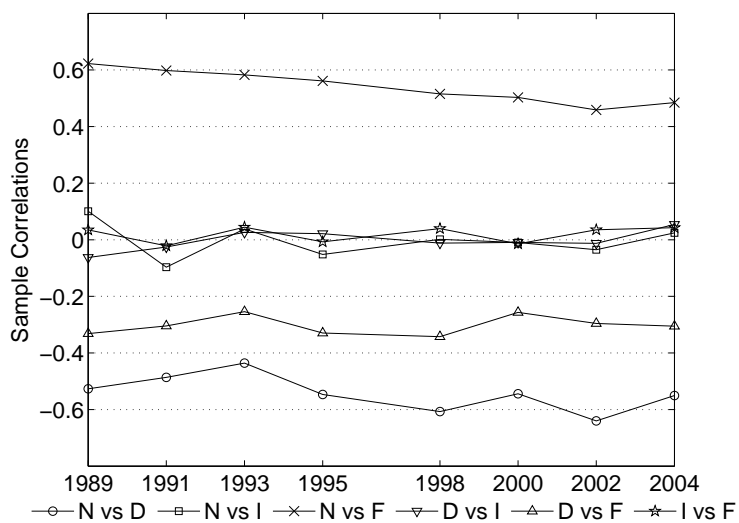


Figure 5: Correlations between HBS distributions and their evolution over time. N = Non-durables; D = Durables; I = Insurances; F = Food.

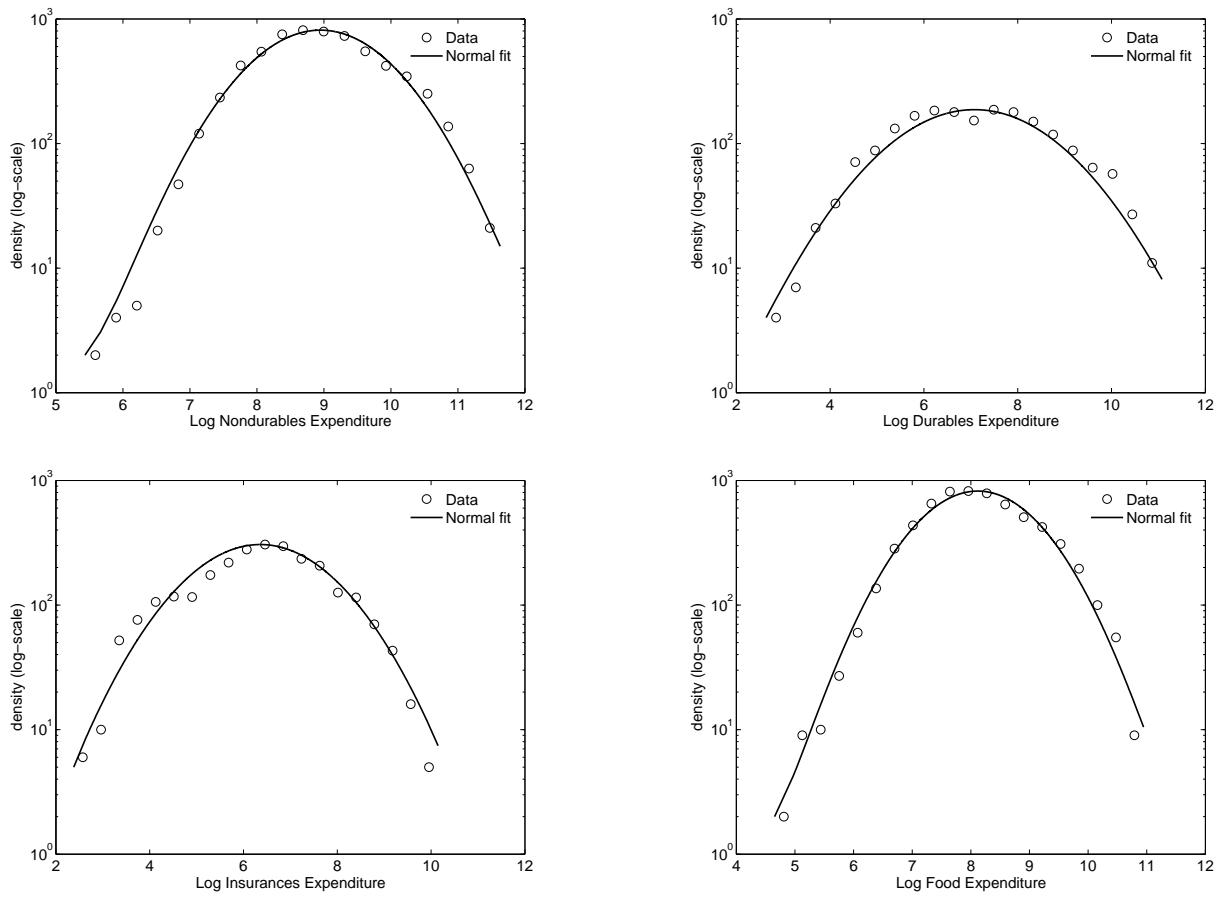


Figure 6: An example of normal fits to logged HCE distributions. Wave: 2000.

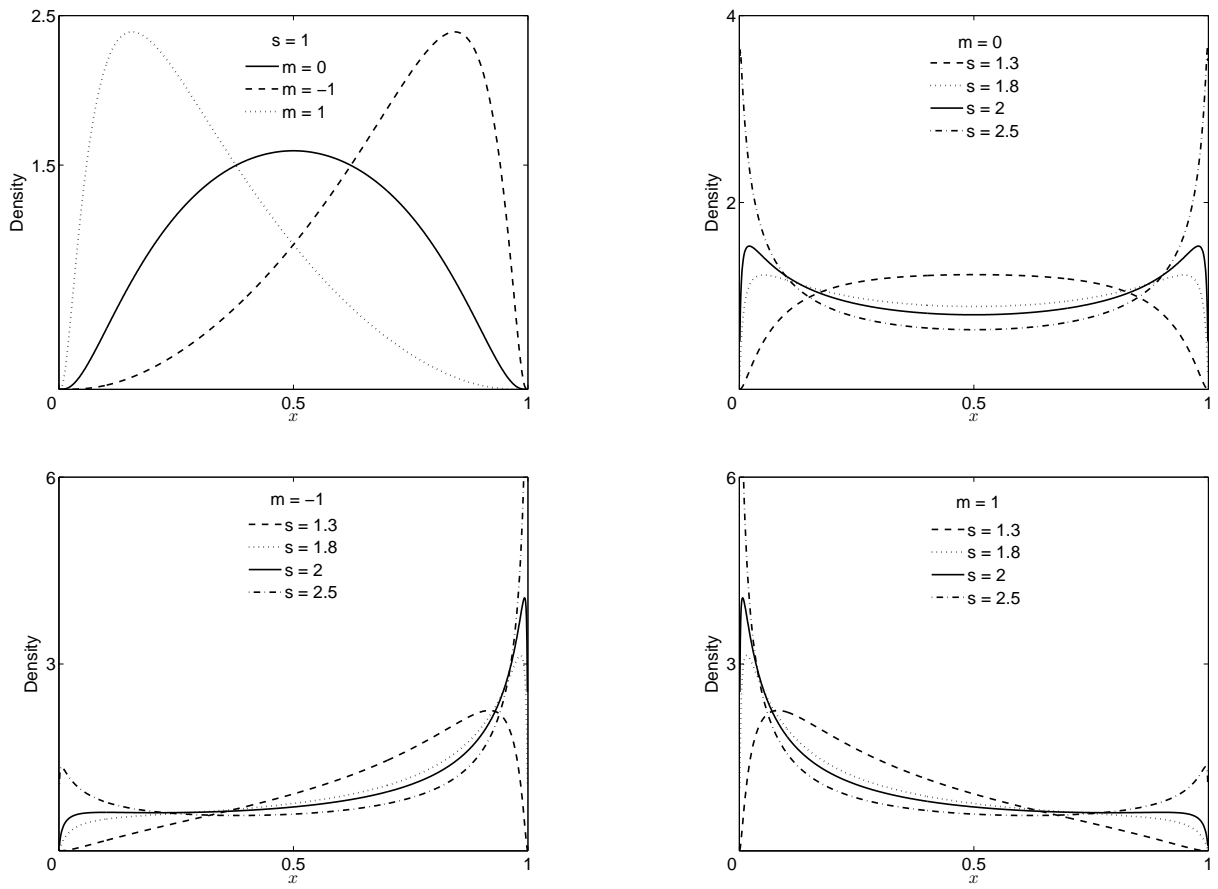


Figure 7: The LN-B approximation to HBSDs. Different shapes of  $f_{B_i}$  as parameters  $m$  and  $s$  change.

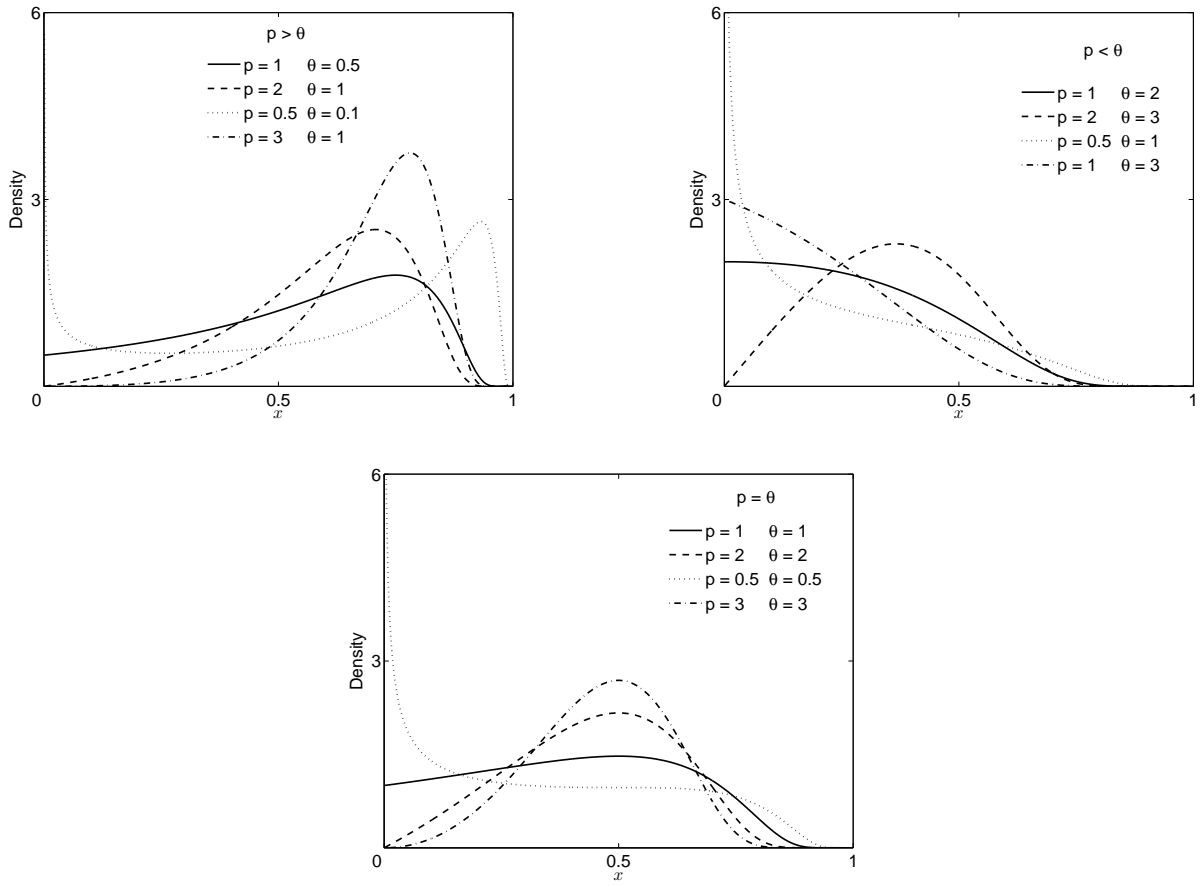


Figure 8: The  $\text{Inv}\Gamma$ -B approximation to HBSDs. Different shapes of  $f_{B_i}$  as parameters  $p$  and  $\theta$  change.

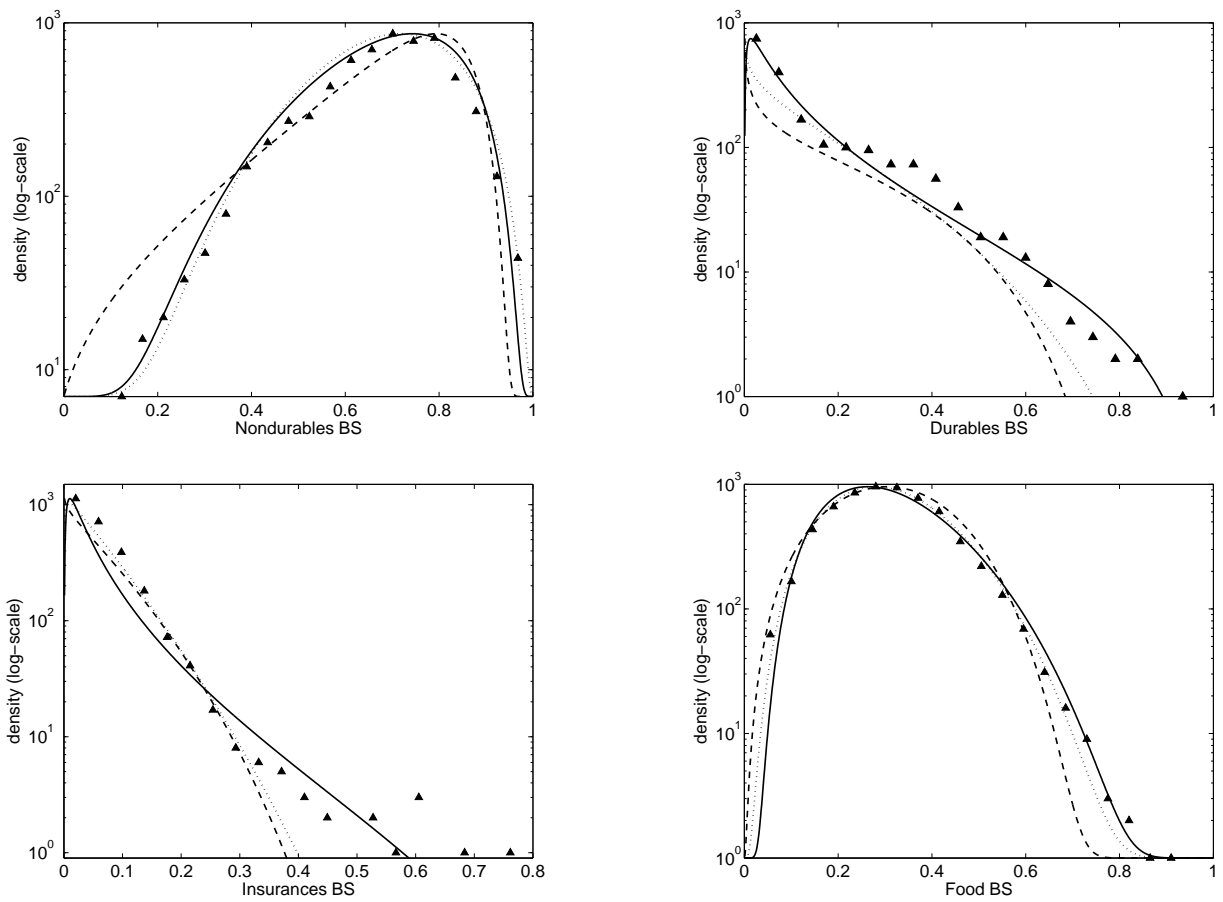


Figure 9: Fitting alternative densities to HBSDs. Wave: 2000. Beta: dotted line. LN-B : solid line. Inv $\Gamma$ -B: dashed line.

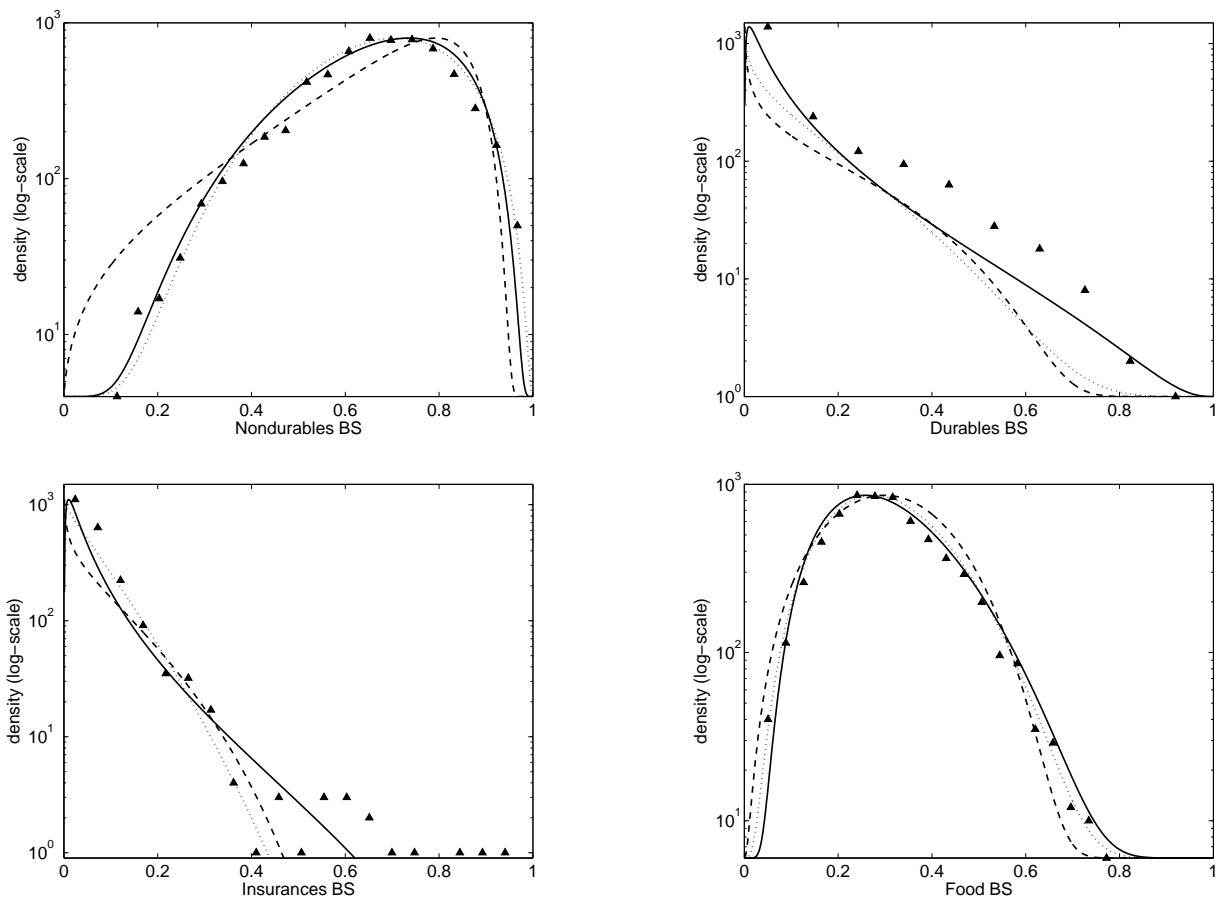


Figure 10: Fitting alternative densities to HBSDs. Wave: 2004. Beta: dotted line. LN-B : solid line. Inv $\Gamma$ -B: dashed line.



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

Table 1: Moments of logged HCE 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 CC.

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

Table 2: Correlations among HCEDs 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.

	Stats	Waves								
		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 HBS 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 CC.

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

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

LN-B		Waves								
	Parameters	1989	1991	1993	1995	1998	2000	2002	2004	Avg
N	m	-1.04	-0.98	-0.98	-0.95	-0.77	-0.81	-0.78	-0.77	-0.88
	s	0.76	0.83	0.83	0.77	0.74	0.72	0.75	0.74	0.77
D	m	2.47	2.28	2.69	2.57	2.43	2.45	2.64	2.80	2.54
	s	1.33	1.33	1.42	1.40	1.37	1.37	1.35	1.34	1.36
I	m	3.74	3.66	3.61	3.61	3.25	3.22	3.26	3.17	3.44
	s	1.18	1.24	1.31	1.31	1.25	1.18	1.17	1.20	1.23
F	m	0.74	0.61	0.59	0.71	0.80	0.86	0.85	0.89	0.76
	s	0.59	0.66	0.66	0.61	0.62	0.60	0.62	0.60	0.62
Inv $\Gamma$ -B		Waves								
	Parameters	1989	1991	1993	1995	1998	2000	2002	2004	Avg
N	p	1.79	1.45	1.38	1.52	1.93	1.96	1.74	1.81	1.70
	$\theta$	0.47	0.37	0.34	0.41	0.67	0.66	0.58	0.62	0.52
D	p	0.66	0.68	0.57	0.63	0.64	0.64	0.61	0.58	0.62
	$\theta$	3.09	2.70	2.78	3.08	2.81	2.83	3.12	3.26	2.96
I	p	0.99	0.95	0.86	0.78	0.92	0.93	0.91	0.71	0.88
	$\theta$	23.14	20.14	15.95	13.37	12.69	12.50	12.46	7.30	14.69
F	p	3.07	2.38	2.52	2.86	2.73	3.02	2.77	3.00	2.79
	$\theta$	5.41	3.51	3.68	4.82	5.00	6.02	5.34	6.09	4.98

Table 5: Estimated parameters of LN-B and Inv $\Gamma$ -B vs. waves. Avg: average values over the whole period. N = Nondurables; D = Durables; I = Insurances; F = Food.

		Waves															
	pdf	1989		1991		1993		1995		1998		2000		2002		2004	
<b>N</b>	<b>Beta</b>	0.004	(0.435)	0.005	(0.863)	0.008	(0.696)	0.010	<b>(0.259)</b>	0.010	(0.456)	0.015	(0.194)	0.013	<b>(0.588)</b>	0.014	(0.635)
	<b>LN-B</b>	0.003	(0.426)	0.004	(0.862)	0.007	(0.683)	0.009	(0.252)	0.010	<b>(0.461)</b>	0.015	<b>(0.197)</b>	0.013	(0.578)	0.014	<b>(0.642)</b>
	<b>Inv<math>\Gamma</math>-B</b>	<b>0.002</b>	<b>(0.664)</b>	<b>0.003</b>	<b>(0.913)</b>	<b>0.006</b>	<b>(0.731)</b>	<b>0.008</b>	(0.248)	<b>0.008</b>	(0.446)	<b>0.013</b>	<b>(0.197)</b>	<b>0.011</b>	(0.582)	<b>0.012</b>	(0.623)
<b>D</b>	<b>Beta</b>	0.004	(0.881)	0.003	(0.939)	0.005	(0.680)	0.004	(0.770)	0.004	(0.848)	0.004	(0.793)	0.005	(0.661)	0.005	(0.661)
	<b>LN-B</b>	<b>0.002</b>	<b>(0.925)</b>	<b>0.002</b>	<b>(0.983)</b>	<b>0.004</b>	(0.573)	<b>0.003</b>	<b>(0.867)</b>	<b>0.002</b>	<b>(0.972)</b>	<b>0.003</b>	<b>(0.862)</b>	<b>0.004</b>	<b>(0.695)</b>	<b>0.003</b>	<b>(0.742)</b>
	<b>Inv<math>\Gamma</math>-B</b>	0.004	(0.902)	0.003	(0.959)	0.005	<b>(0.710)</b>	0.004	(0.820)	0.004	(0.873)	0.004	(0.782)	0.006	(0.654)	0.006	(0.616)
<b>I</b>	<b>Beta</b>	0.003	(0.765)	0.004	(0.627)	<b>0.002</b>	<b>(0.951)</b>	0.003	(0.710)	0.002	(0.935)	<b>0.003</b>	(0.760)	0.003	(0.621)	<b>0.002</b>	(0.758)
	<b>LN-B</b>	<b>0.002</b>	<b>(0.966)</b>	<b>0.003</b>	<b>(0.885)</b>	0.003	(0.867)	<b>0.002</b>	<b>(0.828)</b>	0.003	(0.814)	<b>0.003</b>	(0.767)	<b>0.002</b>	<b>(0.915)</b>	<b>0.002</b>	<b>(0.946)</b>
	<b>Inv<math>\Gamma</math>-B</b>	0.003	(0.764)	0.004	(0.642)	<b>0.002</b>	(0.942)	0.003	(0.685)	<b>0.001</b>	<b>(0.974)</b>	<b>0.003</b>	<b>(0.771)</b>	<b>0.002</b>	(0.684)	0.003	(0.536)
<b>F</b>	<b>Beta</b>	<b>0.001</b>	<b>(1.000)</b>	<b>0.002</b>	<b>(0.986)</b>	<b>0.004</b>	<b>(0.783)</b>	<b>0.006</b>	(0.611)	<b>0.004</b>	<b>(0.799)</b>	<b>0.005</b>	(0.743)	<b>0.005</b>	<b>(0.795)</b>	<b>0.007</b>	<b>(0.526)</b>
	<b>LN LN-B</b>	<b>0.001</b>	(0.999)	<b>0.002</b>	(0.980)	<b>0.004</b>	(0.781)	<b>0.006</b>	<b>(0.618)</b>	<b>0.004</b>	(0.791)	<b>0.005</b>	<b>(0.745)</b>	<b>0.005</b>	(0.794)	<b>0.007</b>	(0.520)
	<b>Inv<math>\Gamma</math>-B</b>	0.002	<b>(1.000)</b>	0.003	(0.981)	0.005	(0.774)	0.007	(0.601)	0.005	(0.786)	0.006	(0.729)	0.006	(0.801)	0.008	(0.523)

Table 6: AAD statistics with p-values (in brackets). AADs 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 CC minimize AAD or maximize p-values.

<b>LN-B</b>	$m > 0$	$m < 0$	<b>Inv<math>\Gamma</math>-B</b>	$\theta > 1$	$\theta < 1$
$s < 1$	F	N	$p > 1$	F	N
$s > 1$	D, I		$p < 1$	D, I	

Table 7: A taxonomy of HBS distributions according to the estimated parameters  $(m, s)$  and  $(p, \theta)$ . N = Nondurables; D = Durables; I = Insurances; F = Food.

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

Table 8: A taxonomy of HBS distributions according to estimated sample moments. N = Nondurables; D = Durables; I = Insurances; F = Food.  $\mu$  = mean;  $med$  = median;  $\sigma$  = standard deviation;  $\xi$  = skewness;  $\kappa$  = kurtosis.