Projektbereich A
Discussion Paper No. A–575

A non parametric analysis of distributions of household income and attributes

by

Werner Hildenbrand *)
Alois Kneip *)
Klaus Utikal *)

March 1998

*) Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged.

SONDERFORSCHUNGSBEREICH 303
Rheinische Friedrich-Wilhelms-Universität Bonn
Lennéstraße 37
D–53113 Bonn
Germany
Telefon (0228) 73 92 42
Telefax (0228) 73 79 40
E–Mail: with2 @ track.or.uni-bonn.de

Typeset in LaTeX by Klaus Utikal

Abstract

We study time changes of the distribution of income, age, occupation, household size via their marginal and conditional densities. The data consists of cross sectional samples from the population of Brithish households drawn over the years of 1968 - 1995. Estimation is carried out nonparametrically and no assumptions on the form of the underlying densities is made.

Key words: kernel density estimation, cross sectional data, income distributions

JEL C13, C14, C21

1 Introduction

An analysis of the distribution of income and related socioeconomic variables is of major importance in economics and other fields. The term "related variables" refers to attributes like age, family size, occupational status etc, which are highly correlated with income and play an important role in many economic models.

In this paper we will concentrate on the study of household incomes and attributes in Great Britain. The data comes from the Family Expenditure Survey (FES) for the years 1968-1995. Similar surveys exist for other countries (e.g. the Enquête Budget Famille in France (see Section 2)); they could be analyzed in an anologous way.

There is a considerable literature dealing with income distributions. A lot of theoretical and applied work is done from the point of view of welfare analysis, see ,e.g., Gottschalk and Smeeding (1997). The focus there lies on measuring "inequality" of the income distribution in given populations. A common approach is to construct specific parameters which quantify the degree of inequality such as quantiles of the income distribution, Gini coefficient, Lorentz curve; for definitions see ,e.g., Atkinson (1983). These parameters are compared across countries and their time evolution is studied. [1]

Variables like income, age, family size also play an important role in demand analysis. They are seen as basic determinants influencing household consumption expenditure on different commodities like food, fuel, transportation, services, etc. So-called "demand systems", see, e.g., Deaton and Muellbauer (1981) or Blundell et al., (1993) try to model the consumption of an individual household as a function of such variables and prices. An important problem in macroeconomics is to model the time evolution of mean consumption of the whole population. Typical approaches to model mean consumption use vector autoregressive models based on *mean* income, see, e.g., Deaton (1992). On the other hand, it is obvious that changes in mean consumption generally depend on the changes of the whole income distribution, and it is a simplification to assume that this distribution only enters through its mean. In fact, in the economic literature there has been a controversial discussion concerning the possible additional influence of distributional effects like "increasing inequality" of income. Hildenbrand and Kneip (1997) propose a general approach to model the influence of changing income and attribute distributions on consumption. A basic idea is to look for time invariances of such distributions. This is illustrated by the following simple example. Let $f_1, f_2, \ldots$ be the income densities arising in different

[1] In this literature one is usually interested in analyzing "individual" income. Generally, such individual income is determined from household income by using so-called equivalence scales. In this paper we do not follow this approach.

years $t = 1, 2, \ldots$. In general, approaches where changing income distributions only enter through their mean would be completely satisfactory only if we had invariance of the relative income distributions:

$$\bar{x}_t \, f_t(\bar{x}_t x) = \bar{x}_s \, f_s(\bar{x}_s x) \qquad \text{for all years } t, s$$

where $\bar{x}_t$ denotes mean income in year t. Then changes in income distributions are completely parametrized by their means $\bar{x}_t$ since the density of relative income $x_t/\bar{x}_t$ does not change over time.

The above invariance introduces a specific condition on the time evolution of income densities. It is shown below that this simple transformation already may provide a reasonable first approximation but that other more general transformations on income as well as other concommitant variables can improve invariance. Generally the search for invariance of suitably transformed densities can be summarized in the following problem: *given yearly samples of observations, find a family of simple transformations that lead to a family of densities which change very little with time.* A simple transformation is parsimonious if it depends only on a few parameters which may be changing with time. Ideally these parameters can be interpreted in economic terms (like mean or variance etc. of the income distribution). Moreover they will be suitable for predictions of future transformations (hence future income densities) from the past.

Before we can search for transformations of income densities we are faced with the problem of density estimation. It turns out that these cannot be considered to be of simple parametric form except for special subgroups, hence need to be estimated nonparametrically. The literature on nonparametric density estimation is vast and several different methods have been established, see e.g, Silverman (1986) . The method of kernel estimation described in Section 3 is conceptually very simple and most widely accepted. Moreover, for samples of large size, as available to us and as described in Section 2, it is our experience (and can also be shown mathematically) that estimates are close to each other even if they are obtained by very different nonparametric methods.

The transformations discussed in Section 4 and 5 are based on simple methods, e.g., standardization, logarithmic transformation etc. The goodness of the transformations is usually improved upon by partitioning the population into subgroups. In several cases it can be argued that invariance is already satisfactorily achieved this way, whereas for other variables, such as income, transformation remains indispensable.

The problem of predicting future densities using the above method remains open to research. Also the question of how to devise a general method of finding an appropriate transformation remains unanswered in this paper. We refer to a forthcoming paper of Kneip and Utikal (1998) for a mathematical solution to

this problem, based on the semiparametric analysis of a general family of scale transformations. The economic implications of these findings remain up to date unexplored.

2 Data

The data to be used in the analyses of the following sections come from the U.K. Family Expenditure Survey (FES). This survey is carried out yearly since 1957 on the initiative of the British government. Each year a total of approximately 7000 households, i.e., about 0.5 % of all British households, record their expenditures on a large variety of consumption items such as bread, different types of meat etc. A "household" is defined loosely as a group of persons living under the same roof who share at least one common meal. The information is obtained through interviews of the household members as well as "diaries" in which participants are asked to keep records of all expenses during a fourteen day period. The response rate lies at around 68 % of all selected households and is varying yearly. Also included in the survey are different forms of income and other household characteristics. For a precise definition of the variables, sampling units, sampling designs, interviewing and field work, confidentiality, reliability etc. we refer to the respective yearly FES manuals as well as the Family Survey Handbook of Kemsley et al (1980).

In the present study we use information on household income as well as on age, and occupational status of the household head included in the survey. A "household head" is the husband of a married couple, in all other cases it is essentially the person who owns or rents the dwelling. The income variable provided is net, i.e., disposable weekly household income, which essentially equals gross income minus taxes and social security deductions, however retirement payments have not been deducted.

Similar surveys are carried out in different countries, e.g., the CEX (USA), EPF (Spain), EBF (France). All these surveys are cross sectional (i.e. different household samples in different years). They are considerably different from existing panel surveys in which a cohort of households is followed up over time as is done, e.g., in the PSID (USA) and GSEP (Germany).

The neophyte reader should be warned however that no conclusions may be drawn from any data without carefully scrutinizing the precise definition of the variables observed. This is particularly important to keep in mind when, as is in this case, the definitions change even slightly over time. This point is further illustrated in the next section.

3 Statistical Methods

For the estimation of densities the time honored histogram (i.e. frequency diagram) has proven to be a useful tool. It is as easy to compute as it is simple to understand and to explain. However, there are obvious disadvantages to it as well. One such is that a continuous density is estimated by a discontinuous function. Also, it would be awkward to study time changing densities by overlaying several histograms in one picture. Several methods have been divised to overcome the shortcomings of histograms and produce higher precision estimators, see Silverman (1986). Among these we will restrict our attention to kernel density estimators. A well developed theory exists and software is available from many sources. Also it will be seen that they include histograms as a very special case.

Given n observations $X_1, \ldots, X_n$ of a random variable X a nonparametric estimator of its density $f(x)$ is the well-known kernel density estimator

$$\hat{f}(x) = \frac{1}{h} \sum_{i}^{n} K\left(\frac{x - X_i}{h}\right) \tag{3.1}$$

where K is a chosen kernel function satisfying the conditions that it be nonnegative and integrate to one. For example, for the estimates of this paper we used the Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \tag{3.2}$$

On the choice of kernel and the *bandwidth parameter* h we further comment below. Note that for $x = x_1, x_2, \ldots$ where $x_{i+1} - x_i = h$ and $K(x) = 1/h$ for $-h/2 < x \leq h/2$ and zero otherwise, we get back the histogram estimator when we compute $\hat{f}(x_i)$ as in (3.1) and define $\hat{f}(x) = \hat{f}(x_i)$ for all other $x_i - h/2 < x \leq x_i + h/2$.

It is well known that for large samples this estimator is asymptotically unbiased and normally distributed with variance

$$\mathrm{var}\{\hat{f}(x)\} = \frac{c_k}{nh} f(x)$$

where

$$c_k = \int K^2(u)\, du\,.$$

The approximate mean squared error (mse) of $\hat{f}(x)$ is given as the sum of the variance plus the square of its bias by

$$\mathrm{mse}\{\hat{f}(x)\} = \mathrm{var}\{\hat{f}(x)\} + b^2(x) \tag{3.3}$$

$$b(x) = \frac{1}{2} h^2 f(x)'' \, k_2 \tag{3.4}$$

where

$$k_2 = \int u^2 \, K(u) \, du$$

(Note that for the Gaussian kernel $c_k = 1/(2\sqrt{\pi})$ and $k_2 = 1$).

As a general rule it can be stated that, except for their smoothness properties, the kernel estimates do not depend much on the kernel chosen, however the bandwidth parameter used is of crucial importance. Several band width selection methods have been devised, see Simonoff (1996). In the analyses of Section 4 we have used as criterion to minimize the integrated mean squared error $\int \mathrm{mse}(\hat{f}(x)) \, dx$. It can be seen that this optimal bandwidth b_{opt} has to satisfy

$$h_{opt} = k_2^{-2/5} c_k^{1/5} \{f''(x)^2 \, dx\}^{-1/5} n^{-1/5} \tag{3.5}$$

To solve this equation we need to estimate the derivative of the unknown density f''. This involves the choice of a second bandwidth parameter. The resulting bandwidth $\hat{h}_{opt}$ therefore is no more than an estimate for h_{opt}. One simple approximation described in Silvermann (1986) is to replace in (3.5) the unknown function f'' by the derivative of a parametric approximation to the unknown density f. This rule we find usefull to get a first idea of the size of the bandwidth when symmetric, mound shaped densities are to be estimated. A more general approach is to specify a dependence on $\hat{h}_{opt}$ of the smoothing paramter in the estimation of f'' and then solve the resulting equation iteratively. It was shown by Engel et al (1994) that their proposed algorithm (which is used also by us) is convergent. Moreover, $\hat{h}_{opt}$ is asymptotically convergent for increasing sample sizes to the solution h_{opt} of (3.5). This convergence is of the reasonably fast stochastic order of $n^{-1/2}$. The estimated mean squared error obtained when using the estimator with this estimated optimal bandwidth is of order $O_P(n^{-4/5})$. This yields a an estimator which can be expected to be of high precision for moderately large sample sizes.

For the study of densities over a range of serveral years we first compute the optimal bandwidths and then estimate the densities with a bandwidth averaged over the optimal bandwidths of the years considered. The average bandwidths of all estimates displayed are cited in the captions. They can be used to estimate the variability of the estimates, applying formula (3.3).

We note that the estimates of relative income densities (see Secition 4) were all carried out on logarithmic incomes and thereafter the log income density estimates transformed back, using the fomula

$$f_X(x) = f_Y(y(x)) \left| \frac{dy(x)}{dx} \right|$$

where $y(x) = log(x)$. The reported optimal bandwidths refer to the estimates of logged income densities. It is not hard to see that this procedure is equivalent

to estimation of the (unlogged) income densities using a kernel with variable bandwidth $x\,b$ instead of b. The proposed method of transforming the data has been discussed in Wand et al (1991). Without applying this method the structure of the densities in the very low income range would not be captured appropriately by the estimators.

To illustrate the method we consider the relative household income data of 1984. A common histogram of these data is shown in Fig 3.1 [left]. A kernel density estimate [right] of bandwidth 0.3 displays a curve estimate of the same basic shape.

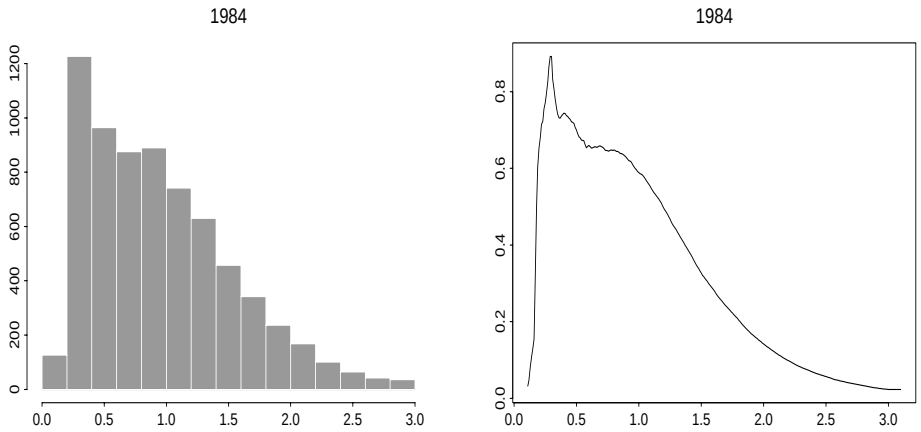


Fig 3.1: histogram(15 classes) and kernel density estimator (bandwidth = 0.3)

A kernel density estimator using the optimal bandwith (0.08) displays a different shape, see Fig. 3.2 [left]. A large singularity emerges at very low relative income levels. Running over the same data a histogram estimator with extremely fine granulation reveals a group of several hundred households concentrating at very low incomes (see Figure 3.2 [right]).

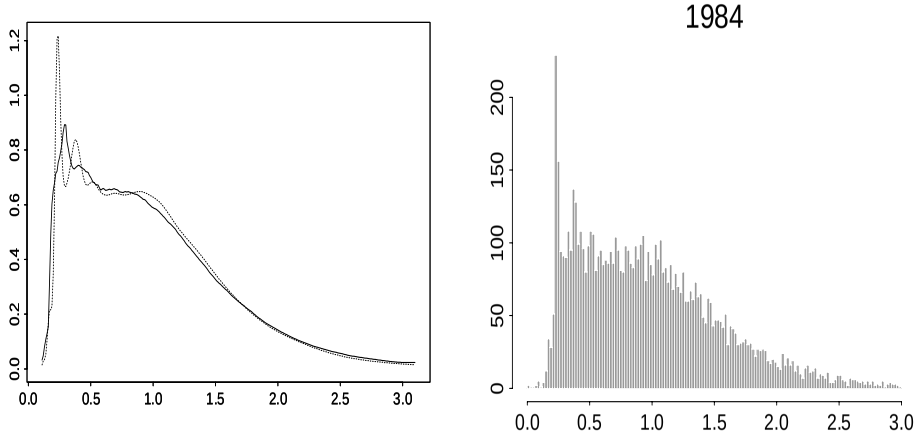


Fig 3.2: kernel density estimator [left] bandwidth=0.3 (solid line) and optimal bandwidth=0.08 (broken line). Finely granulated histogram estimator [right].

We also note this same feature to persist over all years following 1984 while it is absent before, see Fig. 3.3. This may be explained partly by the fact that in 1984 the definition of net income was changed not to include housing benefits any longer. We will therefore generally restrict our analysis to income data prior to 1984 which will illustrate the methods just as well. Also can it be expected that certain subgroups of the population such as the full-time employed are little effected by this change at the very low end of the income range. This is the reason why for this group our analysis is extended to the whole time range.

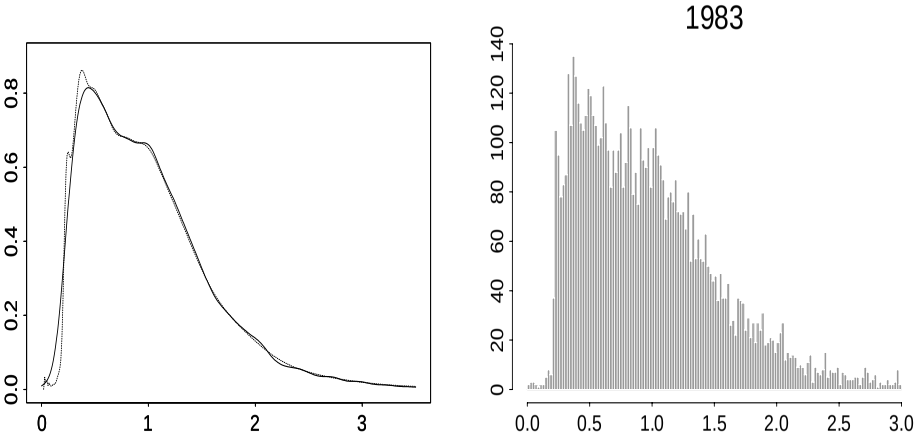


Fig 3.3: kernel density estimator [left] bandwidth=0.3 (solid line) and optimal bandwidth=0.09 (broken line). Finely granulated histogram estimator [right].

We emphasize that this important feature of the data might well have remained hidden to the ignorant analyst when using an "intuitive" choice of band-

width which produces "plausible" estimates and that it is the merit of the optimal bandwidth choice that makes the estimator sensitive to it.

We conclude this section by briefly mentioning the problem of multivariate density estimation. For a good introduction to this subject we refer to the monography of Scott (1992). Similarly to (3.1) one can define for a bivariate sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ the kernel smoother

$$\hat{f}(x,y) = \frac{1}{h_1}\frac{1}{h_2}\sum_{i=1}^{n} K_{12}\left(\frac{x - X_i}{h_1}, \frac{y - Y_i}{h_2}\right)$$

where K_{12} is a bivariate nonnegative function with integral 1. For convenience we consider in the present work only products of univariate normal kernels, i.e.

$$K_{12}(x,y) = K(x)\,K(y)$$

where K is defined in (3.2). Similarly to the univariate case the bias, mean squared error, bandwidth selection matrix etc. have been studied for the multivariate case. Also several new problems have arisen in this interesting field which is still in plain development, for a current list of references see e.g. Simonoff (1996).

4 The distribution of household income

The data described in Section 2 provides cross sectional samples of household incomes over 28 years. The sample sizes are very large, consisting of approximately 7000 households each year. Given the large samples the kernel estimates of the income densities can be expected to have a high degree of precision.

The estimated densities of net weekly income (in pounds) are plotted in Fig. 4.1 for the years 1968, 1973, 1978, 1983. Since (nominal) income is growing steadily over the years it is not surprising that the estimated densities change very much.

It seems natural to consider *relative* income instead of *nominal*; relative income is obtained by dividing the nominal income of every household by the population mean. The densities of relative income are now comparable on a common scale of mulitples of mean income, see Fig. 4.2.

It can be seen that there are only very few households with extremely low income and that there is no clear upper bound for high income, where the distribution tails off rather slowly. Note that for reasons of presentation have the densities been cut off at incomes of three times the mean.

The multimodal structure of these densities may be explained as the result of a superposition of unimodal densities characterizing some influential subpopulations. This is suggested by Figure 4.3 and 4.4 which show estimates of relative income densities for the subgroups of full-time employed and unemployed household heads. Note that for each subgroup relative income refers to income normalized by the mean income of this subpopulation. The resulting densities are roughly unimodal with modes at very different locations[2]. Corresponding modes in the relative income density of the total population (Fig. 4.2) are clearly visible.

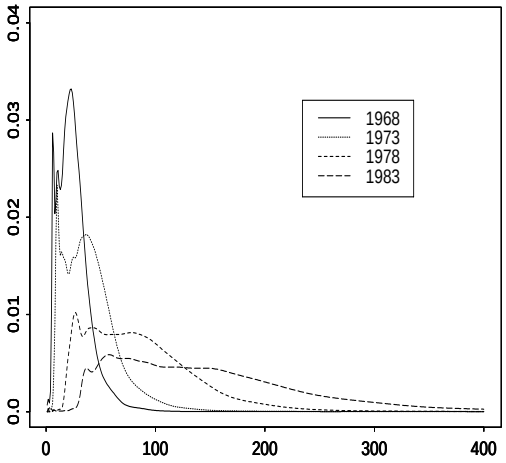


Fig. 4.1: Total population: nominal income densities

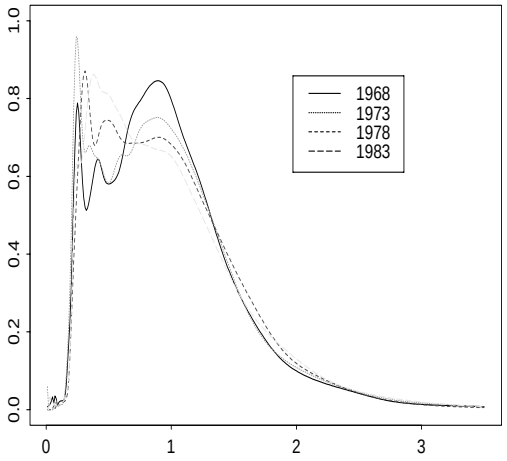


Fig. 4.2: Total population: relative income densities $(0.079)^3$

When considering the time evolution of income densities, we have already noted above that the distribution of nominal income changes rapidly over time. The transformation from nominal to relative income leads to much more "invariant" distributions. Nevertheless, this invariance is far from perfect. A close inspection of Figure 4.2 still reveals a trend in the time evolution of these relative income densities: there is an increasing number of poor households, while the hight of the middle class peak around mean income decreases over time. This subject is further pursued in Kneip and Utikal (1998). Stratification with respect to suitable subpopulations can lead to densities which are more stable over time as can be seen from Fig. 4.3 and 4.4. In particular, the densities of

[2]In Section 2 we already mentioned a change in the definition of income in 1984 by which the population of full-time employed remains unaffected. For this reason only incomes of this latter group have been studied beyond 1983.

[3]optimal average bandwidth parameters used in the smoothing are shown in parenthesis ()

full-time employed household heads are more invariant over time than those of the total population. In fact, part of the time trend characterizing the latter can simply be explained by the growth of such subgroups; it is well known that the percentage of full-time employed persons is consistently decreasing in the U.K. while unemployed have been constantly on the rise until recently.

One might try to eliminate some the remaining differences between the densities in Fig. 4.2 -4.4 by applying more sophisticated transformations incorporating higher order moments as is done next. Let X_{it} denote nominal income of an individual household i and define the *standardized log income* Z_{it} by

$$Z_{it} = \frac{log(X_{it}) - \mu_t}{\sigma_t}$$

where μ_t and σ_t^2 denote mean and variance of $log(X_{it})$ within the population. If the underlying distributions were *exactly* normal then for any year t the resulting density f_t^* of Z_{it} generated by this transformation would be standard normal. Hence, under this hypothesis the densities f_t^*, $f_{t+1}^*, \ldots$ would be completely time invariant. Note however that time invariance does not require log normality. In fact, quite generally it seems reasonable to expect an improved time invariance after applying this transformation, because it eliminates differences in location *and in dispersion* between the distributions over time. For example, it is well known that the variances of the logarithmized relative incomes increase with time, hence standardization eliminates this effect.

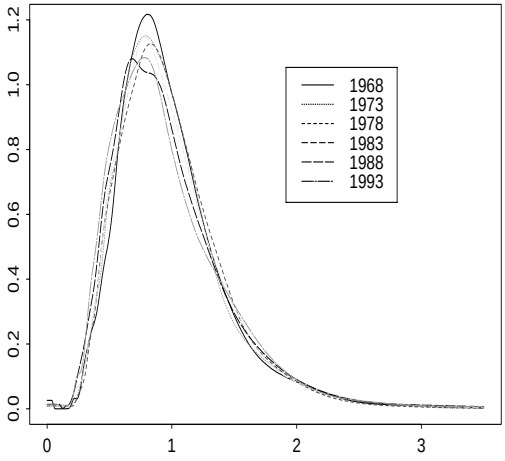


Fig 4.3: Full-time employed: relative income densities (0.084)

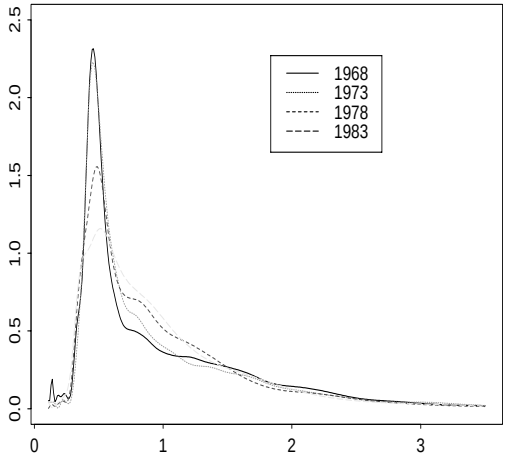


Fig. 4.4: Unoccupied: relative income densities (0.089)

Figure 4.5 and 4.6 show the estimated densities f_t^* of standardized log income for the total population and for the subgroup of full-time employed. There seems to result quite a satisfactory time invariance of the densities of full-time employed. In principle this can be tested though this has not yet been done by the authors. The benefit of this hypothesis is obvious since the study of the income density evolution would completely reduce to a study of the evolution of the parameters μ_t and σ_t.

The log standarized income densities of the total populations are obviously not exactly invariant, but there seems to be a further gain in stability compared to relative income densities.

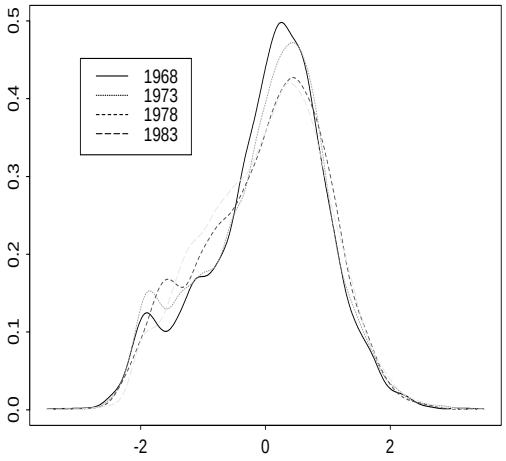


Fig. 4.5: Total population: standardized log income densities (0.127)

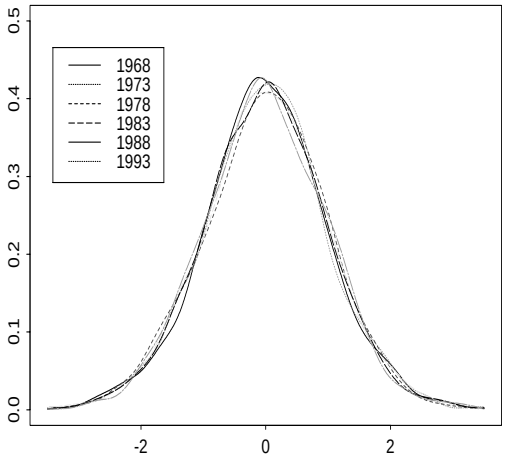


Fig. 4.6: Full-time employed: standardized log income densities (0.188)

5 The joint distributions of household income and household attributes

In the previous section we have seen that a stratification with respect to occupational status leads to much simpler structured income densities. In general, from an economic point of view there is considerable interest in a joint analysis of income and other household attributes (like age, household size, etc.). This is of particular importance in demand analysis. Consumption and savings depend

on income as well as on attributes. In economic literature age is often considered as the most important covariate, see Deaton (1992).

In the following we will thus concentrate on the joint distribution of age and income. In the FES data the variable "age" refers to "age of the household head". The corresponding age distribution therefore does not represent the distribution of ages of all individuals in Great Britain. Obviously there are very few household heads which are below 20 years of age.

In the preceeding section we have analyzed in detail the marginal distribution of income. Kernel density estimators can also be employed to study the marginal distribution of age. Fig. 5.1 shows estimated age densities for the years 1968-1971 and 1980-1983. It can be seen that indeed there are few young and few very old households. The densities are high in the range between 25 and 70 years. The structural details are quite irregular and not easy to interpret. In fact, estimated optimal bandwidths are very small. This must be seen as a consequence of the fact that only a small bandwidth can provide an adequate modeling of the rapid increase of the density after age 20 as well as of the rapid decrease around age 70. More stable estimates could be achieved by a locally adaptive choice of bandwidth, selecting larger bandwidths in the region between 30 and 70 years of age.

Our major interest lies in comparing the age densities over time. There is no visible development *within* each of the four years ranges. Up to some minor fluctuations the age densities between 1968-1971 appear to be very close to each other and the same holds true for those obtained for 1980-1983 or any other range of 4 consecutive years. However, there exists a striking difference between the two age density families of Fig. 5.1 corresponding to the early seventies and early eighties. In contrast to the seventies one recognizes a "young household" peak in the eighties, indicating thus a long-run socioeconomic change.

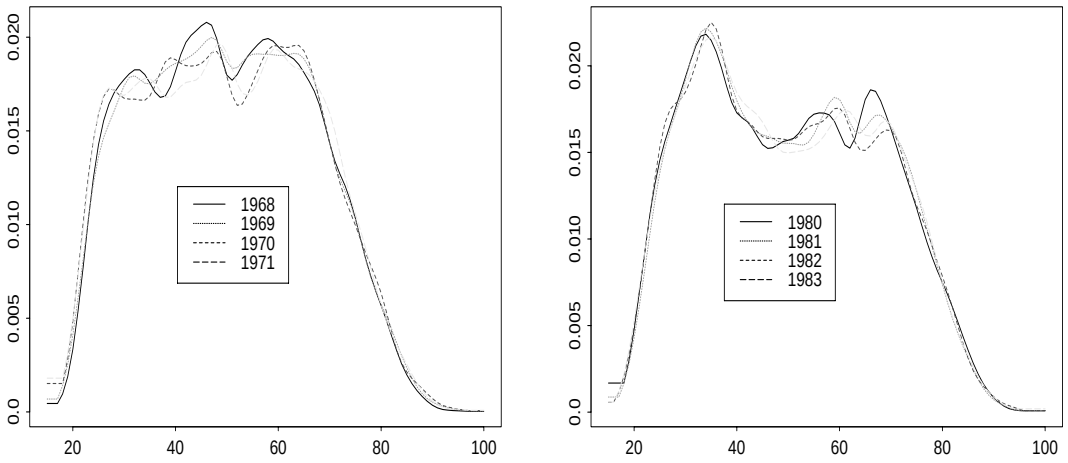
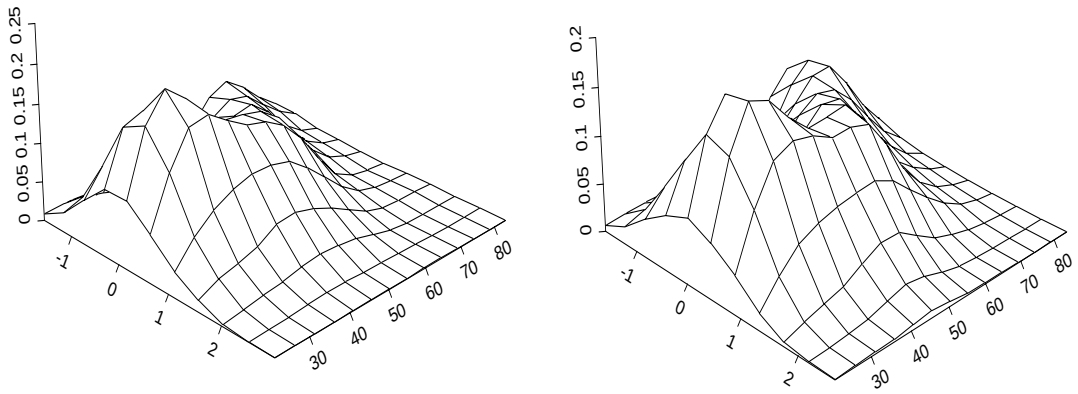


Fig 5.1: Total population: age densities (1.84) [left], (1.91) [right].

Recall the discussion of time invariances in Section 4. We saw that nominal income densities change rapidly from year to year, and only after applying suitable transformations we could speak of an approximate invariance of the transformed income densities. The situation is quite different for the marginal distribution of age. It is neither necessary nor feasible to look for such parametric transformations. Age densities change very slowly from year to year, they are *approximately time invariant in the short-run*. Only if one is mainly interested in long-run analysis, then changes in the age densities have to be taken into account.

Let us now turn to the joint distribution of age and income. It does not make much sense to consider the joint distribution of age and *nominal* income. It is clear from the marginal distribution of nominal income that this distribution will change rapidly over time. However, following our discussion of marginals there is some hope to find regularities in the joint density of age and *standardized log income*. Fig. 5.2 shows two-dimensional kernel estimates of this joint density for two different years.



1968 1978

Fig 5.2: Total population: Joint density of standardized log income and age

The two densities look similar to a certain extent. As an overall tendency one recognizes that in both years very young and very old households are poorer than those in their mid-life. The vast majority of very poor is over 60 years of age. However, there is also an important group of poor young households. Households belonging to middle class and upper middle class (according to their income) are most frequent around 50 years of age. A perhaps surprisingly low number of rich households are over 70.

A deeper understanding of the time development of these features can be achieved by analyzing stratified age densities based on a stratification into income classes. We used four such classes defined by their position in the density of standardized log income (denoted by x): $-2 \leq x < -1$ (poor); $-1 \leq x < 0$ (lower middle class); $0 \leq x < 1$ (upper middle class); $1 \leq x < 2$ (rich).

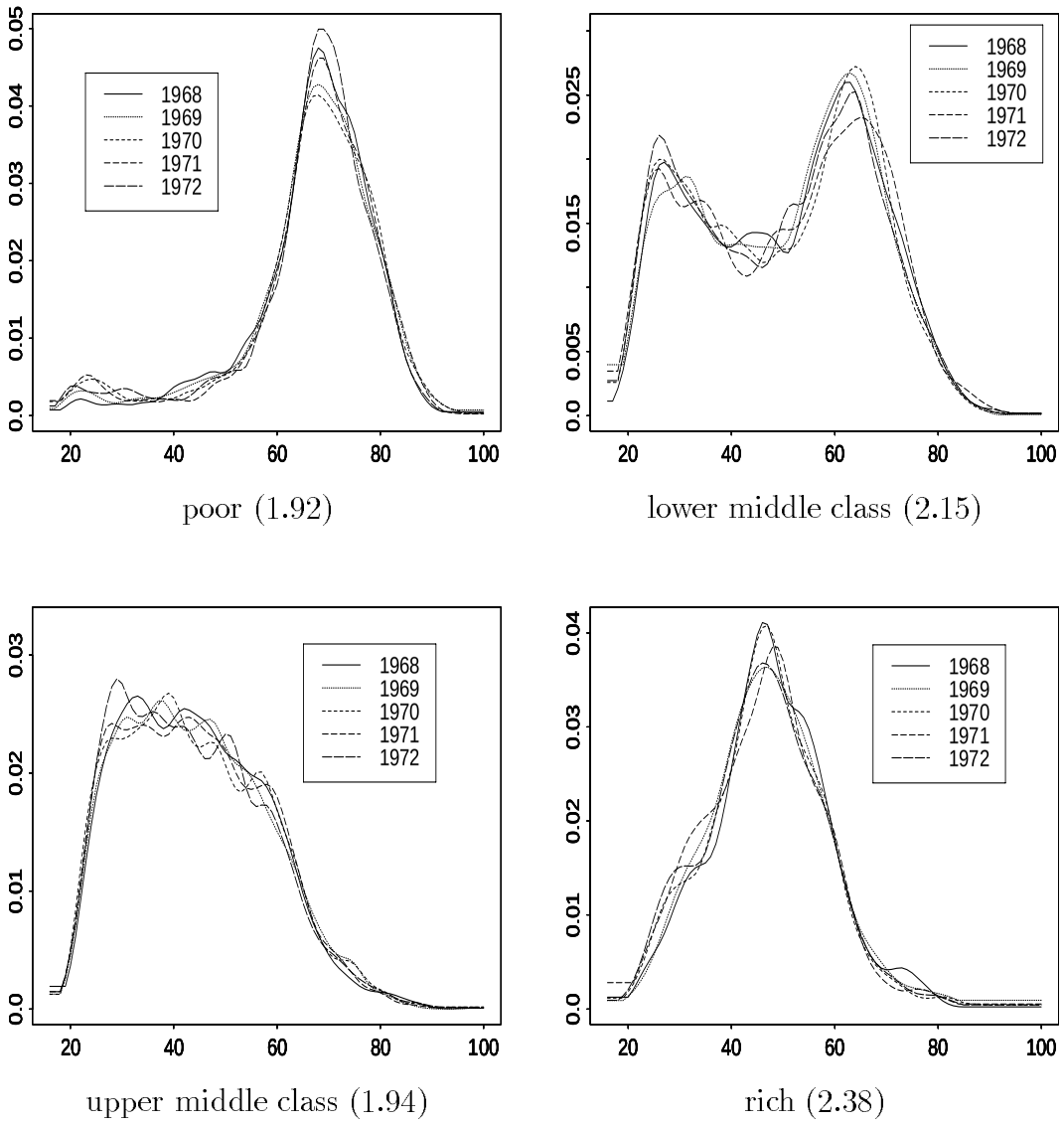


Fig 5.3: Total population: age densities for six consecutive years stratified by income classes, average optimal bandwidths in paranthesis

Fig. 5.3 represents the resulting age distributions for the four income classes in the years 1968-1972. One notices immediately that the age distributions for different income classes are very different in location and shape. More precisely, as could already be seen from the bivariate density above, we find a concentration of the poor around 70 years of age. The bimodal distribution of the lower middle class is concentrated around 30 and 70 years. The group of rich households shows a unimodal distribution concentrated around age 50.

Fig. 5.3 also shows that *within* groups the stratified densities change very slowly in between 1968 and 1972. The same result is obtained for other ranges of five consecutive years. This can be seen, for example, in Fig. 5.5 which shows the age densities of poor households in the years 1991-1995. On the other hand, there are considerable lon-run changes in the stratified densities which are most pronounced for the group of poor households. The latter is demonstrated by Fig. 5.4 which shows the resulting development in between 1968 and 1995. One clearly recognizes the appearance of a subgroup of poor young households. It should be noted that this socioeconomic phenomenon also explains part of the increase of the "poor household peak" in the income distribution. In summary, in a way similar to our discussion of the marginal distribution of age we can draw the following qualitative conclusions about the time development of the joint distribution of age and standardized log income.

a) Stratified age distributions change very slowly from year to year, they are *approximately time invariant in the short-run.*

b) In the long-run there is a clearly visible trend for the group of poor households. This trend has to be taken into account in a long-run analysis.

c) There are *drastic* differences of the age distributions *between* income classes. These difference are much more important than the time changes within classes.

The joint distributions of income and other household attributes such as family size and occupational status can be studied in a similar way. The discrete nature of these variables even simplifies the analysis. It can be shown that the qualitative conclusions obtained by a) - c) remain valid when replacing age by family size or occupational status. A detailed analysis is beyond the scope of this paper.

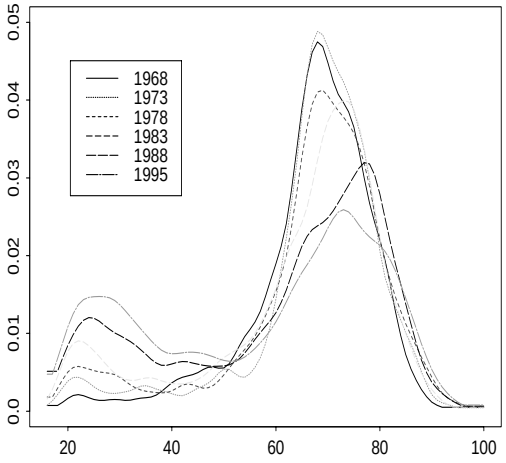


Fig 5.4: Poor households: long-run changes of densities (2.31)

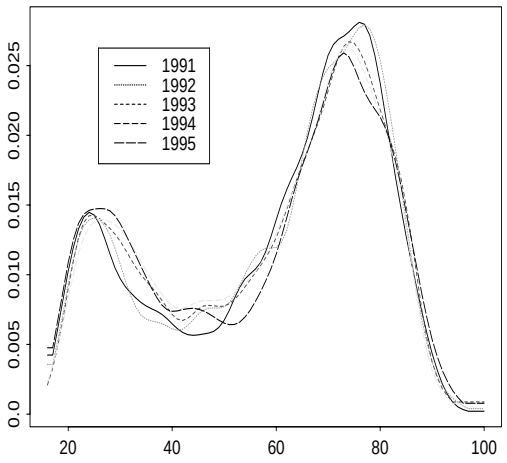


Fig 5.5: Poor households: short-run changes of densities (2.56)

6 Summary

To study the distributions of household income and certain household attributes we have used nonparametric methods of density estimation. Our major objective was to find invariances in the time evolution of these densities after suitable transformations. The use of nonparametric methods is natural in this context since large samples are available but there is no a priori acceptable functional form of the population distribution. Estimation by the method of kernel functions is seen to be superior to density estimation by histograms. The problem of bandwidth choice is crucial. The method of optimal bandwidth determination is shown to be preferable to any "intuitive" choice.

Bivariate density estimates are plotted as surfaces in three dimensional spaces only to give an overall impression while a detailed study of multivariate densities proceeds via marginal and conditional densities.

The search for invariance is the basic theme in the study of time evolution of densities. One looks for simple transformations of variables such that the corresponding densities become "approximately" equal. The problems of description, modelling, and prediction of changing densities are reduced this way to a study of these transformations.

In the particular case of income densities one obtains a rough approximation to invariance by simply standardizing incomes by their means. When applied

to subpopulations this invariance is improved. A better invariance of income densities is obtained by standardizing the logarithm of nominal income. This transformation also leads to satisfactory results for the joint distribution of income and age.

7 References

Atkinson, A.B. (1983). *The Economics of Inequality.* Clarendon Press, Oxford.

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

CEX. Institute for Social Research, University of Michigan, Consumer Expenditure Survey, United States Department of Labor, Bureau of Labor Statistics.

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

Deaton, A., and Muellbauer, J. (1980). An almost ideal demand system, *American Economic Review 70*, 312-326.

EBF. Enquête Budget de Famille (1979,1984-85,1989). Division "Condition de Vie des Ménages", Institut National de la Statistique et des Études Économiques, Paris.

Engel, J., Herrmann E., Gasser, T. (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives, *Nonparametric Statistics*, Vol.4,pp.21-34

EPF. Instituto Nacional de Estadistica (INE), Encuesta de Presupuestos Familiares, Madrid, Spain.

ESCR Data Archive at the University of Essex, Family Expenditure Survey, Annual Tapes 1968-1986, Department of Employment, Statistics Division, Her Majesty's Stationary Office, London.

Gottschalk, P. and Smeeding, T. (1997). Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature 35*, 633-687.

GSOEP Deutsches Institut für Wirtschaftsforschung, Berlin, German Socio-Economic Panel.

Hildenbrand, W. and Kneip, A. (1997). Demand aggregation under structural stability. Discussion Paper A-560, Sonderforschungsbereich 303, University of Bonn.

Kemsley, W.F., Redpath, R.D., and Holmes, M. (1980). Family Expenditure Survey Handbook, Her Majesty's Stationery Office, London.

Kneip, A. and Utikal, K.J. (1998). Some methods in the study of distributional evolution, *in preparation.*

PSID. Institute for Social Research, University of Michigan, The Panel Study of Income Dynamics.

Scott, D.W. (1992). *Multivariate Density Estimation.* John Wiley.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Simonhoff, J.S. (1996). *Smoothing Methods in Statistics.* Springer Verlag, New York.

Utikal, K.J. (1996).*Invariant points of low dimensional curve families* Departmen t of Economics, University of Bonn, SFB 303, Projektbereich A, Discussion Paper No. A-516

Wand, M.P., Marron, J.S., and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*