Incoine Distribution, Income Inequality and Life Cycle Effects – A Nonparametric Analysis for Switzerland

MICHAEL GERFIN

I. INTRODUCTION

The analysis of income distribution and income inequality is of major importance for public economics. It yields important information on both the need for new and the success of existing social policy measures. Therefore, it is not surprising that there is a huge literature on these topics. It is surprising, however, that “the main problem in the statistical description of an income distribution, ... the specification of the [income] density function” (KAKWANI 1980) has received relatively little attention in the literature. The usual approach is to impose a functional form a priori and to estimate the unknown parameters of the assumed density function. To my knowledge the first study that estimated the density function using nonparametric methods was by HILDENBRAND and HILDENBRAND (1986). They found a bimodal income density function in the UK, which does not correspond to any of the functional forms usually employed. This example illustrates that it seems important to start the analysis of income distribution and income inequality with estimating the income density function.

At any given time the income distribution is influenced by the distribution of demographic variables. One of the most important demographic variable is age reflecting that people are at different stages in their life cycle. Part of the observed inequality in the income distribution is probably caused by this life cycle effect. There seems to be a consensus in the literature that this is an important issue. There is, however, no consensus on how to account for this problem in the measurement of income inequality. The prime example is the controversy caused by the paper by PAGLIN (1975). The usual approach is to construct discrete age groups and decompose the inequality into intra-group inequality and inter-group inequality. The inter-group inequality is interpreted as a measure for inequality caused by life cycle effects. The main problem with this approach is the arbitrariness of the age group boundaries, and the results can be sensitive to the choice of the age group width (see e.g. FORMBY et al. 1989).

In order to control for life cycle effects PUDNEY (1993) suggested to analyze the bivariate structure of income and age using nonparametric estimation methods. The income density conditional on age and the regression of income on age give some

* Economics Institute, University of Bern, Gesellschaftsstrasse 49, CH–3012 Bern. This research is financed by the Swiss National Science Foundation (Schweizerischer Nationalfonds) under grant 1215-037478.93. I want to thank BERWIN TURLACH and STEVE MARRON for some GAUSS codes and clarifications regarding binning methods in nonparametric estimation. This paper is an abridged version of GERFIN (1994).
indication of the importance of these life cycle effects. These bivariate structures can be very informative at the exploratory stage of the analysis.

PUDNEY (1993) also suggested to estimate age-conditional inequality indices. These conditional indices should be largely purged of the life cycle effects and should hence measure inequality unrelated to the life cycle. Since there is no theoretical guideline on how to specify these relationships, PUDNEY suggests to use nonparametric estimation methods. Furthermore, this approach avoids the problems of constructing arbitrary age groups.

This paper has two purposes: first, it reproduces PUDNEY’s analysis with other data, and second, it tries to evaluate the nonparametric estimation approach to the measurement of income inequality. The paper is organized as follows: section II presents the estimation of the univariate income density function. In section III, I estimate the bivariate income-age density function, the conditional income density function, and the regression of income on age. Section IV presents the estimation of conditional inequality measures and compares them with the conventional inequality indices. The data are described in Appendix A. Appendix B provides an overview of the employed nonparametric estimation methods.

II. THE INCOME DENSITY FUNCTION

A natural starting point for the analysis of income distributions is to estimate the underlying density function. The conventional approach to do this is to assume a particular parametric form or to simply construct a histogram of the income distribution. But it is well known that histograms are very sensitive to the choice of both the origin and the binwidth (see e.g. SCOTT 1992). Therefore, it is preferable to estimate the density function, e.g. by using nonparametric kernel methods. This will yield an accurate picture of the shape of the density (e.g. skewness or multiple modes). There is a rapidly growing literature on nonparametric density estimation, including some very good textbooks (e.g. SILVERMAN 1986, SCOTT 1992). Therefore, I do not present the details of nonparametric kernel estimation in the main body of this paper. Appendix B contains the relevant information in order to make this paper reasonably self-contained. It should be noted that the estimation employs recently developed techniques to speed up the kernel estimation. By this, the main drawback of kernel estimation, its computational slowness with large data sets, is removed. From this viewpoint the nonparametric methods become an attractive tool for econometric analysis.

Figure 1 shows the estimated income density functions for the year 1982. The data are described in Appendix A. These data have already been used in a number of studies on income distribution and poverty in Switzerland (e.g. BUHMANN 1988). Unfortunately, at the present time there are no cross-section data available for a later date. Income is defined as disposable household income. No attempts have been made to adjust income with respect to family size. To estimate the density, I delete the bottom and top percentile
of the income data because the outlying observations are not very informative for the density and tend to bias the result with the employed estimation technique.

Figure 1: Income density, 1982, confidence bounds, lognormal distr.

Figure 1 shows that the income distribution is unimodal and resembles conventionally used distributions. Figure 1 also shows confidence intervals for the estimated density function (see e.g. Härdle et al. 1993). Also plotted is the log-normal distribution. The coefficients of this distribution are estimated by maximum likelihood. It is evident from Figure 1 that the lognormal distribution is rejected by the data. This implies that e.g. an analysis of the effects of income redistribution based on the lognormal distribution would yield biased results.

III. BIVARIATE INCOME-AGE RELATIONSHIP

In this section I present an explorative analysis of the income-age relationship in order to gain some information of the influence of the age distribution on the income distribution. In a first step I estimate the bivariate density of income and age by using nonparametric kernel methods. Age is defined as age of household head.

Figure 2 shows the estimated bivariate density for 1982. The density conforms with a priori expectations: the mode is concave with respect to age. This impression is confirmed by the contour plot displayed in Figure 3. This is in contrast to the findings.
in PUDNEY (1993) where no systematic correlation between age and income was detected.

The density of income conditional on age is plotted in Figure 4. It has to be interpreted as follows: take any vertical slice through the surface parallel to the income axis. The resulting curve is the pdf of income conditional on that particular age. The conditional
densities first shift to the right with increasing age and then shift back. This again reflects the expected concave relationship between income and age. This is further illustrated by Figure 5 that shows the conditional densities for selected values of age. The shift of the conditional densities is obvious in this graph.

Figure 4: Conditional density (household income)

Figure 5: Some examples of conditional densities
These perspective 3-d and contour plots are very useful for giving a general impres­sion of the income-age relationship. However, they are often not easy to interpret and can be misleading because the impression is largely dominated by the modal relationship. The mode is a rather unconventional measure and hence it is neccessary to have information on the relationship at the mean or the median. Therefore, I estimate the nonparametric mean regression of income on age. The result is displayed in Figure 6. According to this plot there is a strong concave relationship between income and age. This corresponds to the results in numerous studies for other countries, in general based on parametric methods.

Figure 6: Nonparametric Regression $E(y|a)$

IV. AGE-CONDITIONAL INCOME INEQUALITY

The results of section III indicate a strong relationship between income and age. This implies that it is likely that the observed income inequality is influenced by life cycle effects. In order to analyze this hypothesis I estimate several inequality indices condition­ional on age. The indices I consider are the Atkinson index and indices belonging to the class of general entropy measures. The Atkinson index, proposed by ATKINSON (1970) is defined as
where $E(\cdot)$ is the expectation operator, $y$ is income, $\mu = E(y)$ is mean income, $n$ is the sample size, and $\varepsilon$ is the inequality aversion parameter ranging from 0 (complete insensitivity to inequality) to $\infty$ (concern only for the poorest member of the society). The second line of (1) is convenient for specifying the age conditional index. To obtain this we replace the expectations in (1) by the corresponding expectations conditional on age, i.e. the corresponding regression. Hence the age conditional Atkinson index is given by

$$A_\varepsilon(a) = 1 - \left[ E(y^{1-\varepsilon} | a) \right]^{1/(1-\varepsilon)} E(y^{-1})^{-1}$$

(2)

The regressions can easily be obtained by employing the nonparametric kernel regression described in Appendix B.

The general entropy (GE) measure is defined as

$$E_\alpha = \left[ 1 / (\alpha^2 - \alpha) \right] \left[ \left( 1 / n \sum_{i=1}^{n} (y_i / \mu)^\alpha \right) - 1 \right]$$

$$= \left[ 1 / (\alpha^2 - \alpha) \right] \left[ E(y^\alpha) E(y)^{-\alpha} - 1 \right]$$

(3)

for $\alpha \neq 0$ and $\alpha \neq 1$. For $\alpha = 1$ the entropy measure is the Theil index defined as

$$E_1 = (1 / n) \sum_{i=1}^{n} (y_i / \mu) \log(y_i / \mu)$$

$$= E(y)^{-1} E[y \log(y)] - \log[E(y)]$$

(4)

In order to obtain the conditional index all the expectation terms in (3) and (4) have to be replaced by the respective conditional expectation terms.

I estimate four conditional inequality indices: the Atkinson with $\varepsilon = 0.5$ and $\varepsilon = 2$, the Theil index, and the GE index with $\alpha = -0.5$. The Theil index and the Atkinson index ($\varepsilon = 0.5$) are sensitive with respect to high income, whereas the Atkinson index ($\varepsilon = 2$) and the GE index ($\alpha = -0.5$) are sensitive with respect to low income. The complete sample is used for the estimation. Figure 7 displays the estimated conditional inequality
indices. The main difference, apart from scaling, between the results for the two groups is that the low income sensitive indices indicate a considerable inequality at the beginning of the life cycle, whereas the high income sensitive indices are roughly constant until the age of 40. From 40 to 70 all indices show a continuous increase in inequality. The behavior of the indices at the end of the age range is more erratic which is due the the relative sparsity of data.

The entropy indices are the only ones that are additively decomposable by population subgroups. Additively decomposable means that the index can be written as a weighted sum of the group inequality indices (the intra-group inequalities) plus a inter-group inequality term based on mean income and group size. The Atkinson index is only decomposable, but not additively decomposable. Recent studies on demographic effects on the income distribution are usually based on additively decomposable indices. The intra-group elements of general entropy indices computed in the usual way hence serve as a good comparison for the nonparametrically estimated conditional indices. A very good description of how to compute the intra-group and inter-group indices is given in Jenkins (1991).

Figure 8 shows the age-conditional Theil and GE indices and the intra-group Theil and GE indices calculated in the usual way, using age groups with width 10 years. Interestingly, the intra-group indices are very close to the estimated conditional index curve, except for high age groups. In the present case, the nonparametric regression approach does not yield information that could not be obtained with the traditional method. This result can also interpreted as a test of the sensitivity of the traditionally computed indices with respect to the arbitrary age grouping (see Gerfin 1994 for an example where the traditional approach can be misleading).
Table 1 shows the results of the decomposition of inequality into intra-group and inter-group effects using the Theil index and the GE index with $\alpha = -0.5$. Standard errors of the inequality indices have been obtained using the bootstrap method (500 replications). The high Theil index in the 80-90 age group is caused by two outlying observations with very high income. This is reflected by the high standard error in this case. The low-income sensitive GE index indicates a high inequality in the youngest age group which the Theil index does not pick up. This emphasizes the importance of using several inequality indices. The inter-group effect which can be interpreted as inequality due to life cycle effects is relatively large. For the Theil index the inter-group effect accounts for about 15% of total inequality. For the GE index this proportion is lower (about 12%). Hence there seems to be a considerable life cycle effect in the observed income inequality in Switzerland in 1982.
Table 1: Decomposition of income inequality by age groups (1982, Household Income)

<table>
<thead>
<tr>
<th>Age group</th>
<th>Theil Index</th>
<th>GE index (α = -0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-30</td>
<td>0.052 (0.003)</td>
<td>0.398 (0.088)</td>
</tr>
<tr>
<td>30-40</td>
<td>0.044 (0.003)</td>
<td>0.133 (0.012)</td>
</tr>
<tr>
<td>40-50</td>
<td>0.045 (0.003)</td>
<td>0.144 (0.014)</td>
</tr>
<tr>
<td>50-60</td>
<td>0.068 (0.007)</td>
<td>0.165 (0.013)</td>
</tr>
<tr>
<td>60-70</td>
<td>0.165 (0.038)</td>
<td>0.286 (0.038)</td>
</tr>
<tr>
<td>70-80</td>
<td>0.080 (0.010)</td>
<td>0.161 (0.014)</td>
</tr>
<tr>
<td>80-90</td>
<td>0.265 (0.143)</td>
<td>0.274 (0.119)</td>
</tr>
<tr>
<td>90+</td>
<td>0.054 (0.016)</td>
<td>0.127 (0.044)</td>
</tr>
</tbody>
</table>

Intra-group index | 0.074 (0.008) | 0.245 (0.025) |
Inter-group index  | 0.013 (0.005) | 0.033 (0.013) |
Total Inequality   | 0.088 (0.011) | 0.277 (0.044) |

Notes: Intra-group index is weighted sum of age group indices (see e.g. JENKINS 1991). Bootstrap standard errors in parentheses.

V. CONCLUSIONS

The aim of this paper was to illustrate the use of nonparametric estimation techniques in the analysis of income distribution and income inequality with focus on identifying life cycle effects. The general conclusion is that these techniques provide a useful complementary tool for this analysis. The nonparametric methods are especially useful at the exploratory and descriptive stage of the analysis. They provide important information on the shape of the income distribution and on the relationship between income and demographics such as age.

However, concerning the measurement of income inequality the nonparametric methods did not yield different results than the traditional methods. The reason to employ the nonparametric methods was to control for life cycle effects without having to divide the data in arbitrary age groups. Because this arbitrary grouping can be problematic (see GERFIN 1994 for examples) the nonparametric methods can be used as a test for the traditional methods. According to the traditional decomposition analysis about 12 to 15% of the total inequality can be attributed to life cycle effects.

On the computational level the new techniques used in this paper to speed up estimation seem very promising. It took about the same time to compute the nonparametric estimate of the conditional inequality indices as to compute these indices with the traditional method. Hence computation time is no longer an argument against nonparametric kernel methods.

Further work on these topics is necessary in several directions: From the econometric point of view data driven bandwidth selection is an important issue because the bandwidth chosen by rule of thumb tend to oversmooth the result (see Appendix B). The main work in my view is to extend the analysis to income dynamics, where these nonparametric estimation methods seem to be very promising, given the fact that studies
on income dynamics find that the main cause for income changes is "luck" (see e.g. VARIAN 1980). These random changes of income over time can be described by their density function that can be estimated nonparametrically. Finally, more work on the identification of demographic effects on income inequality is necessary. This includes the adjustment of income with respect to family size.

REFERENCES


HÄRDLE, W. and D.W. SCOTT (1992), "Smoothing in low and high dimensions by weighted averaging using rounded points", Computational Statistics 1, 97-128.


Appendix A: Data

The data used in this paper come from the Swiss Income and Wealth Survey (SIWS) 1982. The 1982 sample is stratified and can be made a random sample of the population of permanent residents in Switzerland by using sampling weights (see e.g. Bühmann 1988). The income data were collected from the tax offices and contain very detailed information on all income sources. For the same individuals, except the permanent foreign residents, the income data were also collected for the years 1978 and 1980. Hence the SIWS data can be used as panel data for a subsample. However, only the data for 1982 can be seen as random cross-section data. Therefore, the analysis in the present paper is based on the data for 1982. Table A.1 presents the descriptive statistics for the SIWS data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
<th>1% percentile</th>
<th>99% percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>42.355</td>
<td>38.707</td>
<td>0</td>
<td>1301.440</td>
<td>6.036</td>
<td>132.340</td>
</tr>
<tr>
<td>Age</td>
<td>48.6</td>
<td>17.0</td>
<td>21</td>
<td>97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Income is disposable household income per year in 1000 SFr. Age id age of household head. All figures are unweighted.

Appendix B: Nonparametric kernel estimation

a) Univariate densities

Denote the observations of the variable X by \( X_i, i = 1, \ldots, n \), where n is the sample size. The kernel density estimate \( \hat{f}_h \) of the density \( f \) is constructed by averaging over scaled kernel functions centered in the points \( x_i \), i.e.

\[
\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)
\]

(B.1)

where \( K(\cdot) \) is the kernel function. The two main conditions to be fulfilled by the kernel function are symmetry and that it integrates to unity. Hence a natural choice for kernel functions are symmetric probability density functions, e.g. the normal density. In this paper I use the quartic or biweight kernel which has the advantage of a compact support on \([-1,1]\). The quartic kernel is given by

\[
K(u) = \frac{15}{16} (1 - u^2)^2 I(|u| \leq 1),
\]
where $I(\cdot)$ is the indicator function taking the value one if the expressions in parentheses is true. The bandwidth parameter $h$ determines the range in which the local averaging takes place. The wider the bandwidth the more averaging takes place, and the smoother the resulting picture. There are ways to determine data-driven bandwidths that are optimal in some sense, e.g. least squares cross validation. In this paper I will rely on rules of thumb in determining the bandwidth and postpone systematic bandwidth selection to further research. The rule of thumb in univariate density estimation is to set $h$ equal to $1.06\sigma n^{-1/5}$, where $\sigma$ is the standard deviation of $X$. The bandwidth chosen by rule of thumb generally leads to smoother estimates than an optimally chosen bandwidth.

b) Bivariate density and regression

The kernel estimator of the bivariate density of a $n \times 2$ data set, $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$, is defined as follows

$$
\hat{f}(x) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{2} K \left( \frac{X_i - X_\hat{j}}{h_j} \right) \right\}
$$

(B.2)

where $x = (x_1, \ldots, x_n)'$. The same univariate kernel is used in each dimension but with a different smoothing parameter in each direction. Kernels in this multiplicative form are called product kernels. SCOTT (1992) recommends using product kernels for estimating multivariate densities. Using product kernel it is easy to take account of the different scaling of the variables by choosing the bandwidth accordingly. SCOTT (1992) gives the following rule of thumb for bandwidth choice in the 2-dimensional case: $h_i = \sigma_i n^{-1/6}$.

Given bivariate data $(X_1, Y_1), \ldots, (X_n, Y_n)$ the nonparametric kernel regression estimator (the so-called Nadaraya-Watson estimator) $E(y \mid x)$ is defined as

$$
\hat{m}_h(x) = \frac{(nh)^{-1} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right) Y_i}{(nh)^{-1} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)}
$$

(B.3)

Kernel estimators are notoriously slow to compute when they are applied directly to the data. Therefore, several techniques to speed up kernel estimation have been proposed. These techniques are based on the idea of discretizing the data. An intuitively easy approach to the idea of discretization of the data is the so called Weighted Averaging of Rounded Points (WARPing) method proposed by HÅRDLE and SCOTT (1992) to make
calculations in nonparametric density estimation and regression faster. These methods require to bin the data before estimation. Space limitations prevent a detailed description of the binning procedure which can be found in GERFIN 1994. FAN and MARRON (1993) report speed gains by the factor 100 over direct implementations employing matrix algebra and by the factor of 10,000 over direct implementations employing do-loops (for the programming language GAUSS which was used throughout this paper). Using these methods it took only several seconds to compute the estimates with data sets of about 7000 observations. Incorporating sampling weights is straightforward by simply weighting the bincounts accordingly.

ZUSAMMENFASSUNG


ABSTRACT

This paper illustrates the use of nonparametric estimation techniques in the analysis of income distribution and income inequality with focus on identifying life cycle effects. I consider kernel estimates of the univariate income density and of the bivariate income-age density. The income-age relationship is also explored by using kernel regression. It turns out that there is a strong concave relation. To analyze the degree of inequality I estimate age-conditional inequality measures nonparametrically. Employing inequality indices belonging to the class of General Entropy measures, which are additively decomposable by population subgroups, allows to compare the results with the traditional approach. It turns out that there is hardly any difference. About 12-15% of total inequality can be attributed to life cycle effects.