A Decomposition Analysis of the Trend in UK Income Inequality
Author(s): Dilip Mookherjee and Anthony Shorrocks
Source: *The Economic Journal*, Vol. 92, No. 368 (Dec., 1982), pp. 886-902
Published by: Wiley on behalf of the Royal Economic Society
Stable URL: https://www.jstor.org/stable/2232673

The Economic Journal, **92** (*December* 1982), 886–902
Printed in Great Britain

A DECOMPOSITION ANALYSIS OF THE TREND IN UK INCOME INEQUALITY*

Dilip Mookherjee and Anthony Shorrocks

Post war Britain has experienced a number of dramatic demographic and social changes that have undoubtedly influenced the distribution of income. There has been a rise in the number of pensioners, whose incomes are relatively low, a considerable increase in the number of young and old households living on their own, and a remarkable expansion in the proportion of working wives. The potential importance of these structural changes suggests that their impact on the income distribution needs to be examined systematically before any attempt can be made to interpret the observed trend in overall income inequality adequately.

The method usually adopted for this purpose attempts to create a comparable time series of distributions based on the structural characteristics pertaining in one particular year. For example, if we wished to investigate the effect of shifts in the age structure of the population, this approach would disaggregate the distribution in one year, t_1, and estimate what the distribution would have been like if the age structure had been the same as that observed in some other year t_0, on the assumption that other distributional characteristics (such as age–income differentials and inequality within age groups) remain invariant to this demographic change. The comparison between the actual distribution at t_1 and the constructed distribution at t_1 with t_0's age structure is taken to indicate the impact of changes in the age structure over the period concerned. This comparison can be presented in terms of quantile shares, summary measures of inequality or the entire Lorenz curves.

This approach, known as 'shift-share analysis' or 'standardising' the inequality series (see Semple (1975), Love and Wolfson (1976), Dinwiddy and Reed (1977), Royal Commission (1977), Rowley and Henderson (1978), MacLeod and Horner (1980)), is perfectly adequate for simple and straightforward structural changes. When several changes are occurring simultaneously, however, it becomes difficult to identify the contribution of the individual factors, and to assess their relative importance to the overall trend in inequality. Furthermore, the combined impact of several structural influences will not in general sum up to the inequality change requiring explanation.[1]

The purpose of this paper is to propose an alternative method of analysing

* We are grateful to Juan Dolado and Frank Cowell for assistance during the early stages of this study and to Richard Glendinning and Fatimeh Mehta for help with the computations. Valuable comments were made by a referee, and by participants at a Public Economic Study Group meeting at University College London. Financial support was provided by the SSRC.

[1] In the words of MacLeod and Horner (1980, p. 7) '...this method of analysis suffers from an index number problem. This is suggested by the fact that the...contributions...together explain a Gini increase of 0·017 rather than the observed increase of 0·012'. In point of fact, this problem is not due to the shift-share approach itself, but rather its application to Gini coefficient values: see the discussion of the Gini 'decomposition' in Section I.

the impact of structural changes on the trend in income inequality that has a number of advantages over the 'standardisation' approach outlined above. It is based on the use of additively decomposable inequality measures, and involves disaggregating total inequality in each year into various components and examining the time paths of these individual contributions. The conventional type of cross section decomposition in a single year associates two main contributions (the 'within group' and 'between group' terms) with each disaggregation characteristic. These contributions are in turn completely specified by information about the subgroup mean incomes, population shares and inequality values. In the decomposition of the *trend* in inequality, we identify three main components corresponding to the influence of changes in subgroup inequality values, shifts in the subgroup population shares and relative variations in the subgroup mean incomes. The decomposition is consistent, in the sense that the sum of the contributions of these three factors is equal to the change in overall inequality, and links up in an obvious way with the annual cross-section decompositions. Furthermore this approach can be easily extended to analyse the effect of many structural factors operating simultaneously.

For our empirical results we have made use of published material from the *Family Expenditure Survey*. Restricting attention to published information provides us with a longer series than available with the FES micro data tapes, but the data are inferior in other respects. In particular, we are able to disaggregate the distribution by age, or sex, or household size, but the published tables do not allow us to perform the simultaneous multiple characteristic decomposition that is, in principle, feasible. In this study we have chosen to disaggregate by age, on the grounds that age acts as a proxy for most of the important structural changes of recent years. Changes in the age structure and in the propensity of persons to form separate households are reflected in the proportion of households in different age groups. The increase in female participation rates is captured, along with other factors, in the shape of the age–income profile for households.

Disaggregating by age has the additional advantage of enabling us to comment on some of the issues raised by Paglin (1975). Paglin argues that age differences contribute spuriously to inequality and suggests an adjustment to the observed degree of inequality that subtracts out the contribution due to differences in mean income across age groups. This adjustment has dramatic consequences, since it reduces the level of inequality by approximately 25% and converts a mildly declining inequality time trend (for the United States) into one that falls significantly. Paglin's paper inevitably attracted a number of responses (see Johnson (1977), Danziger *et al.* (1977), Minarik (1977), Kurien (1977)) and continues to be the subject of controversy. It is not our intention to provide a comprehensive re-examination of Paglin's thesis, but we do attempt to clarify a number of points.

In the next section we begin by examining the age decomposition of the Gini coefficient and demonstrate why it is preferable to conduct the analysis in terms of decomposable inequality measures. In section II we report cross-section age decomposition results for the FES household income distribution in each of the years 1965–80 using a variety of inequality measures. Section III examines the trend in inequality in more detail. We show how changes in inquality within the age groups, in the age structure and in the age–income profile have contributed to the overall trend, and provide an assessment of their quantitative impact. Our results suggest that the trend in inequality over the period 1965–80 is mostly due to the changing age–income relationship. In section IV we discuss the implications of these findings for the adjustments that might be made to the time series of aggregate inequality values, and, in particular, for the trend revision proposed by Paglin. A short summary of the paper and its conclusions is given in section V.

I. THE AGE DECOMPOSITION OF INEQUALITY MEASURES

Consider a population of n individuals (or households) with mean income μ, and let y_i denote the income of individual (or household) i. If N_k represents the subset of individuals in age group k, and this group numbers n_k with mean μ_k, then the aggregate value of the Gini coefficient can be written

$$\begin{aligned} G &= \frac{1}{2n^2\mu}\sum_i \sum_j |y_i - y_j| \\ &= \frac{1}{2n^2\mu}\sum_k \Big(\sum_{i\in N_k}\sum_{j\in N_k} |y_i-y_j| + \sum_{i\in N_k}\sum_{j\notin N_k}|y_i-y_j|\Big) \\ &= \sum_k \left(\frac{n_k}{n}\right)^2 \frac{\mu_k}{\mu} G^k + \frac{1}{2n^2\mu}\sum_k \sum_{i\in N_k}\sum_{j\notin N_k}|y_i-y_j|, \end{aligned} \tag{1}$$

where G^k is the Gini value for the k-th age group alone. If the range of incomes in any group k does not overlap with that of any other group h, then

$$\sum_{i\in N_k}\sum_{j\in N_h} |y_i-y_j| = n_k n_h |\mu_k - \mu_h| \qquad h \neq k, \tag{2}$$

and (1) becomes

$$G = \sum_k \nu_k^2 \lambda_k G^k + \tfrac{1}{2}\sum_k \sum_h \nu_k \nu_h |\lambda_k - \lambda_h|, \tag{3}$$

where $\nu_k = n_k/n$ is the proportion of the population in age group k and $\lambda_k = \mu_k/\mu$ is its mean income relative to that of the whole population.

In other circumstances, when the subgroup income ranges overlap (as will normally be the case), equation (3) is modified to

$$G = \sum_k \nu_k^2 \lambda_k G^k + \tfrac{1}{2}\sum_k \sum_h \nu_k \nu_h |\lambda_k - \lambda_h| + R, \tag{4}$$

where the term R depends upon the frequency and magnitude of overlaps between the incomes in different subgroups and is usually called the 'interaction-effect'. This is the form of the decomposition of the Gini coefficient due to Bhattacharya and Mahalanobis (1967), Rao (1969) and Pyatt (1976).

The first term in equation (4) represents some kind of average of the inequality values within each age group (although the weights will not normally sum to unity), whilst the second term can be expressed as

$$\hat{G} = \tfrac{1}{2}\sum_k \sum_h \nu_k \nu_h |\lambda_k - \lambda_h| = \frac{1}{2n^2\mu}\sum_k \sum_h n_k n_h |\mu_k - \mu_h|, \tag{5}$$

and corresponds to the value of the Gini coefficient in (1) if the incomes of all individuals are replaced by the mean income of the subgroup to which they belong. Such a procedure eliminates all within age group income variations, leaving the inequality due solely to the difference in mean incomes by age. For this reason we refer to it as the pure 'age effect'. However, there still remains in equation (4) the 'interaction effect' R, which is impossible to interpret with any precision, except to say that it is the residual necessary to maintain the identity. Furthermore, the way in which it reacts to changes in the subgroup characteristics is so obscure that it can cause the overall Gini value to respond perversely to such changes. For instance, an increase in subgroup inequality, keeping subgroup means and population sizes constant, might be expected (on the basis of the first term on the right-hand side of (4)) to increase G. But the value of R can decline sufficiently to offset this and cause G to fall.[1]

Alternative inequality measures can be disaggregated in a fashion similar to that of equation (3), regardless of whether or not the subgroup income ranges overlap. Thus the awkward interaction effect is not required. The entire class of such measures is derived in Shorrocks (1980) under relatively weak restrictions. The subset of indices that satisfy population replication and mean independence is given by the single parameter 'Generalised Entropy' family

$$\left.\begin{aligned} I_c &= \frac{1}{n}\frac{1}{c(c-1)}\sum_i\left[\left(\frac{y_i}{\mu}\right)^c - 1\right] \quad c \neq 0, 1, \\ I_0 &= \frac{1}{n}\sum_i \log\left(\frac{\mu}{y_i}\right), \\ I_1 &= \frac{1}{n}\sum_i \frac{y_i}{\mu}\log\frac{y_i}{\mu}. \end{aligned}\right\} \tag{6}$$

This class includes the Theil coefficient (I_1), the mean logarithmic deviation (I_0) and monotonic transformations of the coefficient of variation and all the Atkinson (1970) family of indices. The parameter c reflects different 'perceptions of inequality' with lower values indicating a higher degree of 'inequality aversion'.

The decomposition equations for the Generalised Entropy measures are

$$I_c = \sum_k \nu_k(\lambda_k)^c I_c^k + \frac{1}{c(c-1)}\sum_k \nu_k[(\lambda_k)^c - 1] \quad c \neq 0, 1, \tag{7}$$

$$I_0 = \sum_k \nu_k I_0^k + \sum_k \nu_k \log(1/\lambda_k), \tag{8}$$

$$I_1 = \sum_k \nu_k \lambda_k I_1^k + \sum_k \nu_k \lambda_k \log \lambda_k. \tag{9}$$

The first term in these equations (the 'within group component') is a simple weighted sum of the subgroup inequality values. The second term is the 'between-group component', reflecting the inequality contribution due solely

[1] Suppose that subgroup 1 contains three individuals with incomes {3, 4, 14} and that subgroup 2 has two persons with incomes {8, 11}. Then $G^1 = 0{\cdot}349$, $G^2 = 0{\cdot}079$ and the overall Gini value is 0·29. If a redistribution of subgroup 1 incomes results in the new set of incomes {1, 7, 13}, G^1 increases to 0·381 (and G^2 is unchanged). But the overall Gini *declines* to 0·27.

to differences in the subgroup means. In a decomposition by age, this term corresponds to the pure 'age effect' in exactly the same way that $\hat{G}$ in equation (5) does so for the Gini coefficient.

When dealing with summary inequality statistics it is always possible that the pattern of results obtained with one index differs from that obtained with another. As a check on this eventuality we present results derived from the use of three measures selected from the Generalised Entropy family (6) and compare these with the Gini decomposition. It turns out that the picture is broadly consistent across these indices, so no ambiguity arises in the interpretation of the evidence. Apart from the Gini coefficient, the measures selected are I_0, I_1 and

$$I_2 = \frac{1}{2n\mu^2}\sum_i (y_i - \mu)^2 = \tfrac{1}{2}CV^2, \tag{10}$$

where CV is the coefficient of variation.

II. DECOMPOSITION RESULTS FOR SEPARATE YEARS

The data used for this study have been extracted from published statistics of the *Family Expenditure Survey*, which give the distribution of household income by age of head of household from 1965 onwards.[1] The data exist only in grouped form and inequality values therefore have to be estimated. This was done by numerical interpolation within each income range using a program developed by Frank Cowell. The program provides alternative estimates under a variety of interpolation procedures, including upper and lower bounds. Since these different estimates of the inequality value are generally bunched together, we report only those based on the assumption of an exact Pareto distribution within each income interval.[2]

The data requirements for the estimation program comprise the mean income and frequency in each income range, and the interval limits. This creates a problem since the FES reports the mean of each income interval for the aggregate distribution, and the mean income of each age group as a whole, but not

[1] The reliability of our results and conclusions is naturally influenced by the quality of FES data, although some of the obvious deficiencies, such as the possibility of response bias, may not be particularly serious given that we are primarily concerned with annual changes in inequality over a fairly long time horizon. Dinwiddy and Reed (1977, Appendix B) document some of the potential drawbacks to the FES figures and provide several references on this issue. Our own feeling is that our results are likely to be more sensitive to the concepts chosen for the purposes of publication, rather than to the quality of the FES data themselves. We would, for example, have preferred to work with household incomes net of tax and adjusted for family size, but such information is not published for the separate age groups.

[2] Because the interpolation procedure can be applied independently to each age group and to the aggregate distribution, we will normally obtain different estimates for the right and left hand sides of equations (7), (8) and (9). The correspondence between the direct estimate of aggregate inequality and the indirect estimate derived from the right hand side of these equations could be used as a check on the reliability of estimation from grouped data. This is an interesting consequence of the use of decomposable inequality measures. However, in our particular case, the Pareto interpolation procedure combined with the assumption of identical interval means for each age group guarantees that the direct and indirect estimates of aggregate inequality will correspond exactly. We were therefore unable to use this check on the reliability of the grouped data inequality estimates.

Table 1

Age Decomposition of Aggregate Inequality by Year: Various Inequality Measures

Index ...	I_0			I_1			I_2			Gini	
Year	Aggregate inequality	Within age group component	Age effect	Aggregate inequality	Within age group component	Age effect	Aggregate incquality	Within age group component	Age effect	Aggregate inequality	Age effect
1965	0·196	0·160	0·036	0·182	0·149	0·033	0·215	0·185	0·031	0·326	0·136
1966	0·194	0·162	0·032	0·181	0·152	0·030	0·214	0·186	0·028	0·326	0·128
1967	0·190	0·159	0·031	0·178	0·150	0·028	0·212	0·186	0·026	0·323	0·124
1968	0·196	0·154	0·042	0·183	0·146	0·037	0·215	0·181	0·034	0·328	0·141
1969	0·211	0·167	0·044	0·197	0·158	0·039	0·237	0·201	0·036	0·338	0·146
1970	0·215	0·170	0·046	0·195	0·154	0·041	0·229	0·191	0·038	0·339	0·151
1971	0·223	0·171	0·052	0·205	0·159	0·046	0·242	0·201	0·041	0·347	0·156
1972	0·211	0·162	0·049	0·190	0·147	0·043	0·214	0·175	0·039	0·337	0·151
1973	0·228	0·169	0·059	0·208	0·157	0·051	0·246	0·199	0·047	0·350	0·167
1974	0·225	0·170	0·055	0·206	0·156	0·049	0·244	0·198	0·046	0·348	0·167
1975	0·224	0·164	0·060	0·199	0·147	0·052	0·221	0·174	0·047	0·347	0·168
1976	0·219	0·159	0·060	0·199	0·146	0·053	0·224	0·176	0·048	0·347	0·171
1977	0·216	0·162	0·054	0·194	0·146	0·048	0·214	0·170	0·044	0·343	0·163
1978	0·217	0·160	0·057	0·196	0·145	0·051	0·217	0·170	0·047	0·346	0·168
1979	0·226	0·161	0·065	0·199	0·142	0·057	0·214	0·162	0·052	0·350	0·176
1980	0·237	0·167	0·070	0·212	0·151	0·061	0·240	0·184	0·056	0·358	0·184

Note. The age effect for the Gini coefficient corresponds to $\hat{G}$ given in equation (5). Within age group components and age effects for the other indices correspond to the two terms in equations (7)–(9).

the interval means for each age group.[1] These were assumed equal to the mean income of the corresponding interval in the aggregate distribution.[2]

Table 1 presents aggregate inequality values computed for each of the years 1965–80. All four of the indices selected show a modest upward trend in household income inequality over the entire period with the largest year-on-year changes recorded in the periods 1968–9, 1971–3 and 1979–80. Fluctuations in the annual inequality values are more pronounced for the index I_2, and this may reflect greater sensitivity to the annual change of sample in the Survey.

In the first part of the table aggregate inequality, as measured by I_0, is decomposed into the within and between age group (or pure 'age effect') components as indicated by the two terms on the RHS of equation (8). The age effect constitutes 16–30% of total inequality and rises markedly over the period both in absolute value and as a proportion of aggregate inequality. The within age group component, on the other hand, displays little trend over the entire period. It is therefore clear that the upward trend in aggregate inequality observed for household incomes during 1965–80 is due almost entirely to the rise in the age effect. These features re-emerge when inequality is measured by either I_1 or I_2. For the Theil coefficient, I_1, the pattern is very similar to that for I_0, and the increase in the age effect over the period again accounts almost entirely for the rise in the aggregate value. In the case of both I_2 and the Gini coefficient, the upward trend in the age effect actually exceeds the increase recorded for aggregate inequality. This is evident from the decline in the within age group component for I_2, which would be more pronounced were it not for the large increase between 1979 and 1980.

Although the contribution of the within age group component shows little net change over the entire 15 year period, there are variations year by year, possibly due in part to sample changes in the FES and to errors in estimating the inequality values. Since the within age group component is substantially larger than the age effect, these annual fluctuations are by no means negligible and it would be wrong to dismiss their impact as being insignificant: the importance of the contribution depends very much on the precise period under consideration.

III. ANALYSING THE TREND IN HOUSEHOLD INCOME INEQUALITY 1965–80

In Table 2 the changes in aggregate inequality during successive five-year sub-periods are decomposed into the within and between age group components for the three indices I_0, I_1 and I_2. Between 1965 and 1970 both terms increase

[1] A further problem arises from the absence of an upper limit for the top income range. This upper limit was set arbitrarily at 5 times the lower limit of the top interval, but estimates under alternative assumptions revealed that the computations are affected very little. For example, the lower bound of the highest income range in 1975 was £150 (per week). Assuming an upper bound of £750 generated an I_0 estimate of 0·2235 for the whole population, compared to 0·2233 when the upper bound was set at £500 and 0·2236 when the upper bound was £2,000. The corresponding range for the 30–40 age group was 0·1044–0·1047 and that for those over 65 was 0·2291–0·2292.

[2] The reliability of this imputation can be checked by applying the published frequencies to the imputed interval means to produce estimates of the mean income for each age group that can be compared with the true age means. Performing such an exercise revealed no significant or systematic divergence between the estimated and true means for any age group.

Table 2

The Contribution of Changes in Within and Between Age Group Components to the Trend in Aggregate Inequality

Index ...	I_0			I_1			I_2		
Change in	Aggregate inequality	Within age group component	Age effect	Aggregate inequality	Within age group component	Age effect	Aggregate inequality	Within age group component	Age effect
Period:									
1965–70	0·020	0·010	0·010	0·014	0·006	0·008	0·013	0·006	0·007
1970–5	0·008	−0·006	0·014	0·004	−0·008	0·012	−0·007	−0·017	0·010
1975–80	0·013	0·003	0·010	0·013	0·004	0·009	0·019	0·011	0·008
1965–80	0·041	0·007	0·034	0·031	0·002	0·028	0·025	0	0·025

by similar amounts, causing a significant rise in the aggregate inequality value. Over the next 5 years the age effect continues to rise markedly, in contrast to the within age group term which falls. In the final five-year period both within and between age group components are again positive and cause another significant rise in aggregate inequality.

This pattern of contributions to the trend in aggregate inequality can be understood by considering the time paths of the variables from which the within and between group components are constructed. Table 3 provides such information for selected years. Data on the age structure confirm the rise in the proportion of young (under 40) and old (over 65) households during 1965–80 and the corresponding fall in the proportion of those aged 40–65. The main feature of the early years is the increasing number of young households. After 1970 their proportion continues to increase, but the rising percentage of households over 65 takes over as the most prominent characteristic. The details given of inequality within each age group show that the tendency for household inequality to increase with age and other general features are common to all three of the inequality measures. The evidence points to a rise in inequality within almost all age groups during the first 5 years. For most age groups this is followed by a decline in inequality over the next five years and another upward movement in the last sub-period. Thus the initial increase in the within group component reported in Table 2 can be explained as a genuine increase in within age group inequality, partially offset by the shift towards younger households, which increases the weighting attached to groups with lower inequality values. The position almost exactly reverses after 1970, and the within group component tends to fall because the reduction in inequality within the majority of age groups dominates the rise due to the age structure shift towards older groups of households with high inequality values. In the final years, the rise in inequality within the majority of age groups appears again to be the main influence, with the changes in age structure having an ambiguous impact because the rise in the proportion of households over 65 occurs simultaneously with a fall in the proportion over 60.

As regards the time series of mean incomes, Table 3 shows little trend in the relative incomes of the youngest households or those aged 50–65. There is, however, an upward movement in the relative household incomes of those aged 30–50, possibly as a result of increased female participation rates; and a decline in the relative position of households over 65, possibly due to an increase in the number of single pensioners living alone. Since the groups aged 30–50 have above average incomes and those over 65 are below the mean, the overall effect is to produce a greater arching of the age–income profile, which would, *ceteris paribus*, augment the between age component. The shift of population towards groups with relatively low incomes reinforces this trend. Hence the progressive rise in the age effect that characterises Tables 1 and 2.

It is clear from the above comments that the trend in aggregate inequality is the net result of numerous individual contributions arising from a variety of causes and often counteracting one another. The necessity of providing the type of disaggregated information given in Table 3 is, however, unsatisfactory on several counts. To begin with, the volume of data reported rapidly becomes excessive. We have only provided data for selected years but could, and perhaps should, have included data for the intervening years. Furthermore, had a two way classification by, say, age and household size been available, the space required to record (and the patience required to read) this information would be enormous. A second unsatisfactory aspect is the fact that, whilst we may be able to identify the various contributory effects, there is no way of knowing the quantitative importance of their impact on aggregate inequality. This is particularly relevant when the effects are working in opposite directions.

Table 3

Time Paths of Population Shares, Relative Incomes and Inequality Values by Age Group

	Age group					
	0–30	30–40	40–50	50–60	60–65	65+
Population shares ($\nu_k = n_k/n$)						
1965	0·105	0·166	0·200	0·209	0·092	0·228
1970	0·140	0·173	0·190	0·176	0·098	0·224
1975	0·147	0·180	0·163	0·169	0·094	0·246
1980	0·134	0·204	0·160	0·168	0·075	0·259
Relative mean incomes ($\lambda_k = \mu_k/\mu$)						
1965	0·95	1·03	1·27	1·22	0·89	0·60
1970	0·96	1·10	1·32	1·22	0·90	0·55
1975	1·01	1·17	1·36	1·23	0·90	0·51
1980	0·97	1·17	1·42	1·25	0·87	0·50
Within age group inequality: I_0						
1965	0·088	0·083	0·111	0·163	0·201	0·272
1970	0·110	0·088	0·125	0·177	0·247	0·268
1975	0·117	0·105	0·121	0·177	0·226	0·229
1980	0·141	0·122	0·131	0·179	0·225	0·213
Within age group inequality: I_1						
1965	0·088	0·085	0·114	0·159	0·194	0·291
1970	0·107	0·090	0·124	0·168	0·234	0·289
1975	0·110	0·103	0·115	0·160	0·208	0·258
1980	0·125	0·117	0·124	0·166	0·212	0·236
Within age group inequality: I_2						
1965	0·103	0·100	0·141	0·193	0·237	0·411
1970	0·125	0·107	0·149	0·200	0·294	0·412
1975	0·123	0·118	0·128	0·178	0·239	0·369
1980	0·131	0·136	0·141	0·191	0·248	0·321

These problems can be overcome if decomposable inequality measures are used, since the aggregate inequality value depends only on the subgroup mean incomes, population sizes and inequality values, and this enables the inequality trend to be assigned to changes in these three factors. Thus it is possible to keep the reported information to a minimum and to indicate the relative importance of the contributory influences, without sacrificing the ability to capture the salient features. To see how this can be done we look in some detail at the index I_0, for which the cross-section decomposition was provided earlier in equation (8), and for which the trend decomposition is relatively straightforward. We have already demonstrated that the age decomposition pattern for the three indices I_0, I_1 and I_2 are broadly similar, so restricting attention to one of the indices should not unduly affect the results.

Applying the difference operator to both sides of equation (8) gives

$$\begin{aligned}\Delta I_0 &= I_0(t+1) - I_0(t) = \Delta\left(\sum_k \nu_k I_0^k\right) - \Delta\left(\sum_k \nu_k \log \lambda_k\right)\\ &= \sum_k \nu_k(t)\,\Delta I_0^k + \sum_k I_0^k(t+1)\,\Delta\nu_k - \sum_k \log\lambda_k(t+1)\,\Delta\nu_k - \sum_k \nu_k(t)\,\Delta\log\lambda_k,\end{aligned} \tag{11}$$

where Δ represents the change in the variables from year t to $t+1$. Equation (11) is an exact decomposition of the change in I_0 into four terms that can be interpreted, respectively, as the impact of intertemporal changes in within age group inequality, the effect of changes in the population shares of the age groups on the 'within group' and 'between group' components, and the influence of changes in the relative incomes of the age groups. Computing these individual components provides a way of identifying the contributions of different factors to the trend in overall inequality, as measured by the index I_0.

The aggregation weights in equation (11) are base period values for ν_k and final period values for I_0^k and λ_k. An alternative decomposition is available that switches around these base and final period values. Although the particular choice is unlikely to make much difference to the results, it is perhaps appropriate to adopt a compromise between the base and final period weights, and to replace (11) with

$$\Delta I_0 = \sum_k \bar{\nu}_k \Delta I_0^k + \sum_k \bar{I}_0^k \Delta\nu_k - \sum_k \overline{\log\lambda_k}\,\Delta\nu_k - \sum_k \bar{\nu}_k \Delta\log\lambda_k, \tag{12}$$

where $\bar{\nu}_k = \frac{1}{2}[\nu_k(t) + \nu_k(t+1)]$, and $\bar{I}_0^k$, $\overline{\log\lambda_k}$ are similarly defined. Note that equation (12) is still an exact decomposition.

Applying (12) to the FES data over the period 1965–80 produces contributions that are generally consistent with the pattern anticipated on the basis of the data in Table 3, except in one important respect. Changes in the relative incomes of the age groups tend to have a negative impact on the inequality trend, although we had anticipated a positive contribution, due to the progressive arching of the age–income profile. This happens because the last term in equation (12) reflects changes in the relative mean incomes $\lambda_k = \mu_k/\mu$, which in turn depend on both the subgroup means μ_k *and* the population shares ν_k, since $\mu = \Sigma_k \nu_k \mu_k$. Thus changes in the population shares influence not only the second and third terms on the RHS of equation (12), but also the last term.

It seems more reasonable to identify the impact of changes in I_0^k, ν_k and μ_k (rather than I_0^k, ν_k and λ_k), and for this reason the decomposition given in (12) is not entirely satisfactory. To separate out the effect of changes in μ_k, we can rewrite the last term in (12) as

$$\begin{aligned}-\sum_k \bar{\nu}_k \Delta\log\lambda_k &= \sum_k \bar{\nu}_k \Delta\log\left(\sum_l \nu_l\mu_l/\mu_k\right) \qquad (13)\\ &= -\log\left[1 - \sum_k \lambda_k(t+1)\,\Delta\nu_k\right] + \log\left[1 + \sum_k \theta_k(t)\,\Delta\mu_k/\mu_k(t)\right] - \sum_k \bar{\nu}_k\Delta\log\mu_k,\end{aligned}$$

where $\theta_k = \nu_k \lambda_k$, the income share of subgroup k. Substituting (13) into (11) or (12) produces another exact decomposition, but for computational purposes it appears sufficient to use the approximation

$$-\sum_k \bar{\nu}_k \Delta \log \lambda_k \simeq \sum_k \bar{\lambda}_k \Delta \nu_k + \sum_k (\bar{\theta}_k - \bar{\nu}_k) \Delta \log \mu_k. \tag{13a}$$

Combining (12) and (13a) provides us with the form of decomposition which we finally employed, and which distinguishes the following four components of the change in I_0:

$$\sum_k \bar{\nu}_k \Delta I_0^k, \tag{14a}$$

$$\sum_k \bar{I}_0^k \Delta \nu_k, \tag{14b}$$

$$\sum_k (\bar{\lambda}_k - \overline{\log \lambda_k}) \Delta \nu_k, \tag{14c}$$

$$\sum_k (\bar{\theta}_k - \bar{\nu}_k) \Delta \log \mu_k. \tag{14d}$$

Expression (14a) represents the impact of changes in within subgroup inequality, whilst (14b), (14c) indicate the effect of changes in the population shares on the 'within group' and 'between group' components, respectively. (14d) is the contribution to ΔI_0 attributable to relative changes in the subgroup means.[1]

Table 4

Decomposition of the Trend in Aggregate Inequality: Index $I_0(\times 10^3)$

	Change in aggregate inequality ΔI	Contribution to ΔI due to changes in*: Within age group inequality $\Sigma \nu_k \Delta I_0^k$	Population shares: $\Sigma I_0^k \Delta \nu_k$	Population shares: $\Sigma(\lambda_k - \log \lambda_k)\Delta \nu_k$	Mean age group incomes $\Sigma(\theta_k - \nu_k)\Delta \log \mu_k$
1965–6	−1	6	−4	−2	−2
1966–7	−5	−2	−1	−1	−1
1967–8	7	−5	1	1	10
1968–9	15	12	1	1	2
1969–70	5	3	0	0	2
1970–1	8	0	1	1	6
1971–2	−12	−7	−2	−2	−2
1972–3	17	3	4	4	6
1973–4	−3	2	−1	0	−3
1974–5	−1	−6	0	−1	5
1975–6	−5	−5	1	1	−2
1976–7	−4	3	−1	−1	−5
1977–8	2	−1	0	0	3
1978–9	9	2	−1	2	6
1979–80	10	6	0	−1	5
1965–70†	20	13	−3	−1	11
1970–5	8	−8	2	2	12
1975–80	13	5	−2	3	7
1965–80	41	10	−3	4	30

* These contributions correspond to the expressions given in (14a)–(14d), respectively.

† Figures for longer subperiods are the sum of the annual values over the relevant period.

[1] It reflects *relative*, rather than *absolute*, changes in μ_k since expression (14d) is zero if all age means change in the same proportion.

Changes over time in the aggregate value of I_0, derived from the FES data, are reported in Table 4 together with the decomposition into the four contributions corresponding to (14*a*)–(14*d*). The annual values are generally (and not surprisingly) small, and for the purposes of presentation the true figures have been raised by a factor of 1,000. Due to the sampling fluctuations inherent in survey data, and to errors arising from the estimation process, it would be unwise to place any great significance on the year-on-year values, particularly where these are small. However, at the bottom of the table we provide a breakdown over longer periods, constructed by summing the annual changes over the relevant years, and these may be regarded as more reliable.

The second column of Table 4 represents the *ceteris paribus* impact of changes in inequality within each age group. The values assigned to the three five year sub-periods show positive contributions in the first and last sub-periods, with a negative effect in between. This is consistent with the tendency, evident from inspection of Table 3, for inequality within age groups to rise, then fall and subsequently to move upwards again. The third and fourth columns indicate the impact on the within age group and between group components, respectively, of changes in the population age structure. These two terms taken together reflect the shift towards young and old households. The last column gives the contribution of relative changes in the mean incomes of the age groups and shows a sequence of predominantly positive contributions between successive years that combine to generate a significant inequality-augmenting effect in each of the five year sub-periods. For the whole period 1965–80, the figures provided in the last row of Table 4 suggest that the upward trend in aggregate inequality is due primarily to the greater arching of the age–income profile, with a small additional contribution from changes in inequality within each age group. However, shifts in the age structure have no significant impact over the whole period.

IV. STANDARDISING THE INEQUALITY SERIES

In the light of the preceding analysis, we now consider the various ways in which the inequality values might be adjusted to compensate for the influence of structural changes. In addressing this question in the context of factors related to age, it is natural to begin by examining the adjustment advocated by Paglin (1975). Paglin argues that age differences exaggerate the 'true' degree of inequality and that the appropriate reference distribution should not require complete equality of incomes, but only equality of incomes within each age group. He suggests that the aggregate inequality value, as normally computed, should be reduced by the amount of inequality that would remain if income differences within each age group were entirely eliminated: in other words, by the value of the 'age effect'. Paglin applies this correction to values of the Gini coefficient, and as a consequence encounters the decomposition problems mentioned in section I above.[1] But the procedure can also be

[1] Use of the Gini coefficient introduces an unnecessary source of confusion, since the interaction effect in equation (4) enables Paglin's basic arguments to be converted into a variety of specific adjustment procedures. Some of Paglin's critics in 1977 touched upon this question, and it was recognised as a central issue in Paglin's (1977) reply.

applied to the Generalised Entropy measures which are not subject to the same disadvantages.

Paglin's proposal leads to changes in both the level and trend in the sequence of inequality values. As regards the annual inequality levels, the impact of the revision on our FES data is readily seen in Table 1. For the indices I_0, I_1 and I_2, subtracting out the age effect leaves the within group component value and reduces measured inequality by around 20–25 %. The reduction in the measured Gini value would be somewhat higher – about 45 % – and is considerably above Paglin's figures for the United States (possibly reflecting the higher proportion of elderly households in the United Kingdom). This evidence might be thought to indicate that the 'true' degree of inequality is substantially less than that previously imagined, but there is absolutely no justification for such an interpretation, since the basis of the computations has been altered. One could obtain a similar result simply by replacing the inequality index I with the index $\frac{1}{2}I$. Thus it is difficult to see how any significance can be attached to the degree to which Paglin's adjustment reduces the annual inequality levels.

The same cannot be said of the revision of the inequality *trend*, since there is now a basis for judging whether any particular inequality value is high or low, namely the figures in previous or subsequent years. Applying Paglin's adjustment to the data in Table 1, by subtracting the age effects from aggregate inequality, converts an apparent upward trend in overall inequality into a time profile that is more or less horizontal (in the case of I_0, I_1 and I_2) or even declining (in the case of the Gini coefficient). This may prompt a re-evaluation of the evidence regarding the movement towards or away from an egalitarian income distribution. However, the more detailed analysis of the factors contributing to inequality, carried out in section III, casts serious doubts on whether Paglin's method of standardising the inequality series can be justified.

In eliminating the age effect, Paglin subtracts the last two columns of Table 4 from the trend in aggregate inequality. This procedure implicitly assumes that changes in within group inequality and part of the age structure shift are 'true' determinants of the inequality trend, whilst the movement in the age–income profile and the remaining part of the shift in the age structure contribute spuriously. We see no reason to treat the two population share terms asymmetrically, nor to discard arbitrarily the impact of the changing age–income pattern, which we have shown to be the major influence on the observed inequality trend. As regards the latter, a case might be made if it was discovered that the improvement in the incomes of middle aged households relative to those younger and older was due largely to changes in household composition; but it seems likely that it also reflects the increase in female participation rates and other factors that cannot be dismissed as not 'really' affecting inequality among households.

Controlling for the impact of changes in age structure and household composition provides an alternative adjustment to the inequality series that may meet with more widespread approval. Such an adjustment recognises that demographic movements over time can cause the overall inequality value to vary without any corresponding change in the relative positions of representative households of different types. Data constraints have prevented us from examining shifts in both the age structure and household composition simultaneously, but Table 4 indicates that controlling for variations in the age structure alone (by subtracting out the figures in the two population share columns) would have no impact on the inequality trend over the whole period 1965–80, nor any appreciable effect during each of the five-year sub-periods.

Performing this adjustment would correspond closely to standardising the series on the 1965 age structure by applying the 'shift share' method to the index I_0. The results would not be identical, however, because our procedure takes into account the entire sequence of values for the variables I^k, ν_k and μ_k, whereas shift share analysis typically considers only the values in the base and final years. Thus the distinction closely parallels that between a Laspeyre index with constant weights and a chain index with weights revised annually. It is this correspondence between our decomposition approach and shift-share analysis that enables us to undertake conceptual experiments of the type which ask 'what would 1980 income inequality have been if the age structure had not changed in the period since 1965?'

V. SUMMARY AND CONCLUSION

The purpose of this paper has been to consider the impact of structural influences on income distribution, and in particular on the time trend of inequality in the distribution of UK household incomes for the period 1965–80. We have shown how decomposable inequality measures can be used both to analyse cross-section data in individual years in the conventional way (as in Table 1) and to disaggregate the inequality trend into various contributory influences.

Because of data limitations we have been forced to confine attention to a decomposition by one household characteristic, and chose a breakdown by age. But the population could have been disaggregated by household size, region, occupation or a number of other attributes. Of these, household size seems the most interesting since the social changes underlying the trends in the household age structure and the age–income relationship have also been responsible for systematic movements in the composition of households. In fact a simultaneous decomposition into age–household size groups might well add further insights into the determinants of the trend in observed inequality, and is a worthy topic for further investigation once micro data are available for a long enough period.[1]

The FES data used in this study show a modest upward movement in the overall inequality value observed for household incomes during 1965–80. Our

[1] Cowell (1980, Section 5) has examined how the issue of family size affects the cross-section decomposition by age. As expected, equivalence scale adjustments to household income flatten out the age–income profile, particularly in the 30–65 age range. This will tend to reduce the between age component in the decomposition. Moving from a economic unit based on households to one based on individuals leads to a further reduction in this term, since increased weight is attached to the middle age groups whose adjusted mean incomes are broadly similar. Taken together these changes reduce the between age group components by 75 %, at which point they are only just statistically significant. This result emphasises the drawbacks of working with household income distribution data that provide insufficient information on the size of families.

analysis by age group has shown that this increase is attributable to the rise in the 'age effect' which incorporates the influence of the age structure and the mean incomes of the age groups, but does not depend on inequality within each age group. In fact if the 'age effect' is subtracted from aggregate inequality, as Paglin (1975) advocates, the observed upward trend in household inequality either falls to almost zero (when any of the indices I_0, I_1 or I_2 are used) or is reversed (if the Gini coefficient is employed as the measure of inequality). A more detailed allocation of contributions was given in Table 4 for the index I_0. The figures suggest that the major disequalising influence over the whole period has been the changing age-income profile, which has become progressively more arched. Changes in within age group inequalities made a further positive contribution to this upward trend, but shifts in the age structure had no significant effect in either direction.

Our results assist in understanding the factors affecting the trend in inequality, and also provide a more secure foundation for discussions concerning the way that the observed trend might be revised to eliminate spurious influences. We accept the general principle that, because of structural changes, the distribution observed in different years may not be strictly comparable and that some adjustment to the observed trend in inequality may be necessary. In particular, there may be a case for standardising for shifts in the age structure. However, we see no justification for the adjustment recommended by Paglin, since this arbitrarily eliminates intertemporal variations in the age–income profile, which in turn incorporate factors that represent true movements in the relative living standards of households of different ages.

London School of Economics

Date of receipt of final typescript: May 1982

REFERENCES

Atkinson, A. B. (1970). 'On the measurement of inequality.' *Journal of Economic Theory*, vol. 2, pp. 244–63.

Bhattacharya, N. and Mahalanobis, B. (1967). 'Regional disparities in household consumption in India.' *Journal of the American Statistical Association*, vol. 62, pp. 143–61.

Cowell, F. A. (1980). 'The structure of American income inequality.' Mimeograph, London School of Economics.

Danziger, S., Haveman, R. and Smolensky, E. (1977). 'The measurement and trend of inequality: comment.' *American Economic Review*, vol. 67, pp. 505–12.

Dinwiddy, R. and Reed, D. (1977). *The Effects of Certain Social and Demographic Changes on Income Distribution.* Background Paper No. 3 to the Royal Commission on the Distribution of Income and Wealth. London: HMSO.

Johnson, W. (1977). 'Measurement and trend of inequality: comment.' *American Economic Review*, vol. 67, pp. 502–4.

Kurien, C. J. (1977). 'Measurement and trend of inequality: comment.' *American Economic Review*, vol. 67, pp. 517–9.

Love, R. and Wolfson, M. C. (1976). *Income Inequality: Statistical Methodology and Canadian Illustrations.* Ottawa: Statistics Canada.

MacLeod, N. and Horner, K. (1980). 'Analysing postwar changes in Canadian income distribution.' In *Reflections on Canadian Incomes* (Economic Council of Canada). Ottawa: Canadian Government Publishing Centre.

Minarik, J. J. (1977). 'Measurement and trend of inequality: comment.' *American Economic Review*, vol. 67, pp. 513–6.

Paglin, M. (1975). 'The measurement and trend of inequality: a basic revision.' *American Economic Review*, vol. 65, pp. 598–609.

—— (1977). 'The measurement and trend of inequality: reply.' *American Economic Review*, vol. 67, pp. 520–31.

Pyatt, G. (1976). 'On the interpretation and disaggregation of Gini coefficients.' ECONOMIC JOURNAL, vol. 86, pp. 243–55.

Rao, V. M. (1969). 'Two decompositions of the concentration ratio.' *Journal of the Royal Statistical Society*, Series A, vol. 132, pp. 418–25.

Rowley, J. C. R. and Henderson, D. W. (1978). 'Partial standardisation of Canadian family incomes for changes in composition and other characteristics.' Economic Council of Canada.

Royal Commission on the Distribution of Income and Wealth (1977). *Report No. 5: Third Report on the Standing Reference*. London: HMSO.

Semple, M. (1975). 'The effect of changes in household composition on the distribution of income 1961–73.' *Economic Trends*, no. 266, pp. 99–105.

Shorrocks, A. F. (1980). 'The class of additively decomposable inequality measures.' *Econometrica*, vol. 48, pp. 613–25.