THE STRUCTURE OF AMERICAN INCOME INEQUALITY*

BY FRANK A. COWELL

London School of Economics

Income inequality is examined using the Panel Study of Income Dynamics and a consistent decomposition analysis. I only use inequality measures that satisfy the Principle of Transfers, have the property that a *ceteris paribus* increase in inequality within any subgroup increases overall inequality, and are independent of the scale of income and population. Decompositions are carried out by family size and by age of head for several definitions of income and income recipient. Whilst changing the time unit over which income is measured has a substantial impact on inequality, the effect of removing the between-age-group component of inequality is relatively slight.

1. INTRODUCTION

This paper is about the measurement of the inequality of income distribution amongst persons. This is hardly a new topic,[1] but recent advances in the theory of inequality measurement and the use of a particularly versatile data set make it possible to provide new light on some old questions.

Abstracting from philosophical questions of the comparability of personal welfare scales and the means by which a social welfare function is specified it might appear that in principle the mechanics of inequality measurement (and its associated topic social welfare analysis) are reasonably straightforward. You take the measured values of everyone's index of personal welfare and aggregate them in the agreed fashion. The index of personal welfare is very often some comprehensive measure of real income, and social welfare is treated as a kind of sophisticated averaging process. However in practice there are at least two major complications. Firstly, people live in family groups and pool resources. Secondly, "time and chance happeneth to them all," and we find each person's stream of earnings and other receipts subject to both systematic and random variations. Instead of a clear cut list of individuals we find a heterogeneous (and shifting) set of clumps of persons with fluctuating real incomes.

It would be helpful to know how the heterogeneity of the population and the income fluctuations affect overall inequality. Is there a substantial contribution to overall inequality that is simply a result of differences in family size? If so, can this be eliminated by appropriate respecification of measured income and of the income recipient? Is there likewise a substantial contribution to apparent

*Research funding by the SSRC under grants HR7607, HR6681/2 is gratefully acknowledged. Fatemeh Mehta, Bahram Pesaran, and Yasushi Tanaka provided valuable research assistance. A. B. Atkinson, R. Dinwiddy, J. N. Morgan and the referees of this *Review* provided very helpful comments on an earlier draft.

[1]Some aspects of the effects of different definitions of income and of income recipient on income inequality have been examined by Danziger and Taussig (1979), Lydall (1978), Morgan (1962), Morgan *et al.* (1962), Taussig (1973), van Ginneken (1981), and others. The length of the accounting period and its effect on inequality has been studied by Benus and Morgan (1975), Hanna (1948), Lillard (1977) and others.

overall inequality that is attributable purely to the general life cycle pattern of household incomes? Do transitory income variations, which may be economically unimportant, nevertheless contribute substantially to measured inequality? In each case one would like a specific idea about the relative magnitude of these effects in order to understand what the level of income inequality really is, and how changes in that underlying inequality can be separated out from other shifts in the social structure.

To make headway on all of these questions imposes special data requirements. We need details not only on the composition of families and family incomes—which is readily available on the Current Population Survey tapes—but also on family histories—which is not so readily available. Furthermore, appropriate quantitative tools are required. If one is going to assign inequality contributions to different groups in the population then it is a good idea to have this done in an unambiguous and consistent fashion. If adjustments are to be made to the data which affect the inequality in just one section of the population then it is helpful if one there is a well-defined relationship between overall inequality and inequality within that sector. Also since many are interested in the welfare implications of income distribution it is probably desirable to have a method of inequality analysis that is based on recent developments that permit interpretation in terms of fundamental welfare criteria.

Accordingly in section 3 we shall consider the theoretical foundations of a suitable technique for analysing the structure of income inequality. In section 4 we consider the practical problems of comparing incomes accruing to families with different compositions and the analysis of inequality within and between groups by family size. Section 5 re-examines these issues for groups categorized by age rather than family size. Section 6 examines the problems of extending the time unit to cover several years. However a preliminary examination of the data is called for, which is done in the next section.

2. Data Problems

For some of the issues addressed by this paper a number of data sources are readily available. For example there are in the U.S. and elsewhere a number of quite good sample surveys—such as the Current Population Survey—which would permit a decomposition analysis by, say, age and family size. However such surveys are usually limited to one-time observations and so yield rather limited information about the effect of time on income inequality. On the other hand there are several sources which yield data of time series on, say, earnings of *individuals*; but these are obviously limited in the extent to which they throw light on the inequality of disposable income of all people living in *families*.

The University of Michigan's Panel Study of Income Dynamics (PSID) goes a long way toward providing a data source that meets both the requirements of a time series on particular incomes and of the background information on family composition and the like. A detailed discussion of this source is in the Appendix, but the brief facts are these. In 1968 the Survey Research Center at Michigan set up a panel of about 5,000 families who were interviewed about their incomes and other family characteristics and were then reinterviewed year by year. The panel has evolved in three main ways: persons leaving it, persons marrying into it, and "splitoffs"—i.e. the forming of traceable subfamilies where, for example a teenager leaves home or a couple divorce. This evolution raises certain conceptual and practical problems which I discuss below. In preparing this paper I have used the tapes for the tenth wave of this study, so we have the raw materials for looking at the effects of family characteristics on income inequality over a period ranging from one to ten years.

How the "income recipients" are defined in the sample is discussed more fully in section 6 and the Appendix. The Appendix also discusses in detail the "income" variables I use, the broad specifications of which are as follows. The basic concept is *total family income* which in later sections of the paper I adjust to allow for family size. However, two other concepts are also of interest—*family factor income* and *family disposable income.* These are defined as follows:

$$\textit{family factor income} = \sum \text{labor income} + \sum \text{income from assets}$$

$$\textit{total family income} = \text{family factor income} + \sum \text{transfers}$$

$$\textit{family disposable income} = \text{total family income} - \sum \text{taxes}$$

where in each case the "$\sum$" means sum over *all who were living in the family* during the year in question.

The total panel is in fact merged from a sample that was originally randomly selected and a sample from the Survey of Economic Opportunity especially designed to overselect the poor. Accordingly, to eliminate the obvious source of bias two approaches have been used.[2] Firstly, the SEO members were dropped from the sample, leaving a total of about three and one third thousand families, and the computations were carried out on this sub-sample. Secondly, the entire sample (5,992 families) was used, each observation being weighted according to a scheme devised by the Survey Research Center of the University of Michigan to correct for possible bias. The overall pattern of the results was very similar to those obtained by the first approach, and it is the second, weighted results that are reported. A crude representation of the distribution of the three types of income over these families (for the interview year 1977 referring to incomes in 1976) is given in Figure 1. However to understand the structure of inequality we need first to introduce some purpose-made tools.

3. Principles of Inequality Analysis

Assume for the moment that we have agreed on two basic ingredients of inequality analysis—the definition of income and the definition of the income receiving unit.[3] We now need a numerical method of representing the income distribution—i.e. for summarising data such as those in section 2. Accordingly, let $y_1, \ldots, y_n$ be the "incomes" received by n "persons" $1, \ldots, n$; inequality I is

[2]See Appendix for other adjustments made to the sample.

[3]These issues are more conveniently dealt with in the next section.

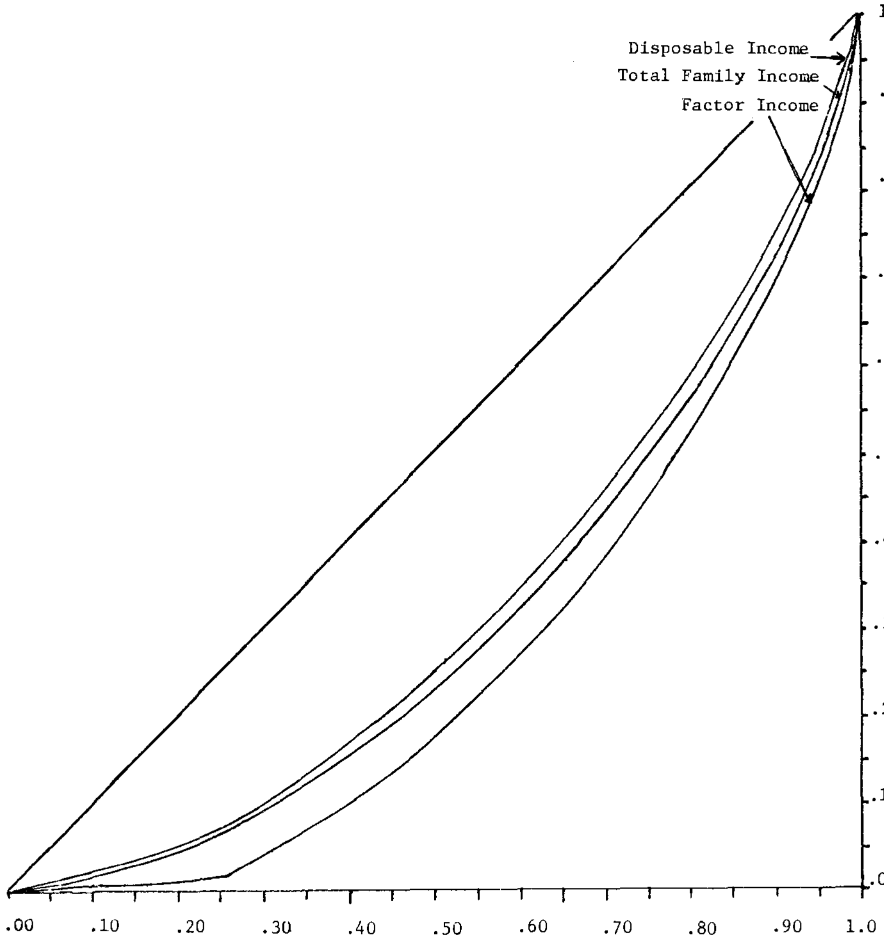

Figure 1. Lorenz Curves for Three Definitions of Income Distributed by Families

given by some function

$$I = \Phi(y_1, \ldots, y_n).$$

The first question to be answered is—what kind of function Φ should be used?

To deal with this let us introduce some convenient notation. Suppose the population $\{1, \ldots, n\}$ is divided into an arbitrary exhaustive collection of G mutually exclusive groups indexed by $g = 1, \ldots, G$, where subgroup g contains n_g persons and has mean μ_g; let inequality in group g be written I_g where

$$(1) \qquad I_g = \Phi(y_{(1)}, \ldots, y_{(n_g)}).$$

Let "between group inequality" I_B be the value of I were every member of group g to receive μ_g instead of his or her actual income, and let μ be the *population* mean. For mathematical convenience only, I shall assume that Φ is "reasonably smooth."[4]

Now consider the properties of Φ that may be desirable from the point of view of the economics of income distribution. I shall suggest just three.[5]

I. *The Weak Principle of Transfers.* Let the income distribution be modified so that \$1 is transferred from some fairly poor person to someone with more income (such that the mean μ remains constant). Then Φ must have the property that I increases.

II. *General Decomposability.* For *any* exhaustive collection of mutually exclusive subsets indexed $g = 1, \ldots, G$, Φ must be such that it is possible to write

$$I = Q(I_1, \ldots, I_G; I_B) \tag{2}$$

where the function Q is increasing in each argument and may depend on $\mu_1, \ldots, \mu_G$ and on $n_1, \ldots, n_G$.

III. *Complete Scale Invariance.* This implies that Φ has the twin properties (a) that I remains invariant under replications of the population, and (b) that I remains invariant under arbitrary changes in the scale of incomes.

Property I is assumed explicitly or implicitly in the majority of theoretical and empirical analyses. The only reason for relaxing it seems to be where the two persons considered are unlike in some relevant respect, in which case it might be reasonable to approve of a transfer from one person to another with higher income. Property III is very widely assumed, although part (b) is quite restrictive since it implies that the measure of inequality is invariant under uniform proportionate growth of all incomes. Relaxing III would introduce a number of rather complicated issues which would deflect one too far from the main theme of this article. It is property II that is so vital, and it is worth while considering this a little further. What the property means is that if I_g (inequality in group g) should increase, *with all other intra-group inequalities and inter-group inequality remaining unchanged,* then overall inequality must increase, regardless of how the population has been subdivided, and regardless of the levels of inequality elsewhere. Several well-krown inequality measures do *not* have this property,[6] and it seems to be rather mportant to eliminate such measures from our enquiries on the *structure* of income inequality. Without property II it is literally impossible to attribute

[4]Formally I need to assume Φ is everywhere continuous, and is twice differentiable in $y_i, \ldots, y_n$ for any *distinct* distribution where no two incomes are identical. In point of fact it is most unlikely that *any* proposed inequality measure will violate this extremely weak assumption. See Cowell and Shorrocks 1890.

[5]Formally these three conditions amount to

(I) Φ is S-convex (which implies that Φ is symmetric).

(II) There exists $\hat{Q}$ such that

$$\Phi(y_1, y_2, \ldots, y_n) = \hat{Q}(\Phi(y'_{(1)}, \ldots, y'_{(n_1)}), \ldots, \Phi(y^G_{(1)}, \ldots, y^G_{(n_G)}); n_1, \ldots, n_G; \mu_1, \ldots, \mu_G)$$

for any arbitrary partition of $\{1, \ldots, n\}$.

(III) $\Phi(\underbrace{y_1, \ldots, y_n}_{(1)}, \ldots, \underbrace{y_1, \ldots, y_n}_{(K)}) = \Phi(y_1, \ldots, y_n)$ and $\Phi(\lambda y_1, \ldots, \lambda y_n) = \Phi(y_1, \ldots, y_n)$

for arbitrary positive integer K and arbitrary positive scalar λ.

[6]For an example of this problem see Cowell (1983).

overall inequality to its component inter- and intra-group inequalities in an unambiguous way.

The only functions Φ that satisfy properties I to III are those that can be written[7]

$$\Phi(y_1, \ldots, y_n) = F\left(\frac{1}{\alpha^2 - \alpha}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\mu}\right]^{\alpha} - 1\right]\right) \tag{3}$$

where F is a monotonic increasing function and α is an arbitrary parameter which may be given any value from $-\infty$ to $+\infty$. We note in passing that the family contains such well-known members as the coefficient of variation ($\alpha = 2$), Theil's index ($\alpha = 1$) and the entire subfamily of Atkinson indices ($\alpha < 1$).[8] Clearly there are two further theoretical questions which must now be answered—what form should F take? What values should be assigned to α?

The form of F clearly leaves the *ordinal* properties of the measure unaltered, but will affect its *cardinal* properties and the precise nature of the decomposition and assignment of group contributions implicit in (3). As an example take the coefficient of variation

$$c \equiv \left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\mu}\right]^2 - 1\right]^{1/2}$$

and the (modified) Herfindahl index of concentration $h \equiv [c^2 + 1]/n$, each of which has the same ordering properties over Lorenz curves.[9] Write the *population share* for group g as $w_g \equiv n_g/n$ and the *income share* for group g as $v_g \equiv n_g\mu_g/n\mu$. Then the aggregation condition (2) becomes in each case:

$$h = \sum_{g=1}^{G}\left(\frac{v_g}{w_g}\right)^2 h_g + h_B$$

$$c = \left[\sum_{g=1}^{G}\frac{v_g^2}{w_g}c_g^2 + c_B^2\right]^{1/2}.$$

Clearly the marginal contribution of c_g to c depends on other subgroup inequalities and on c_B, but the marginal contribution of h_g to h only depends on v_g and w_g. Which cardinalization is more suitable?[10]

[7] For proof and discussion of this see Bourguignon (1979), Cowell (1980), Cowell and Kuga (1981a), Cowell and Shorrocks (1980), Shorrocks (1980). Note that

$$\lim_{\alpha\to 0}\frac{1}{\alpha^2-\alpha}[[y_i/\mu]^{\alpha} - 1] = -\log(y_i/\mu)$$

and that

$$\lim_{\alpha\to 1}\frac{1}{\alpha^2-\alpha}[[y_i/\mu]^{\alpha} - 1] = \log(y_i/\mu)y_i/\mu.$$

[8] For further discussion see Blackorby and Donaldson (1978), Cowell and Kuga (1981b).

[9] In fact of course the Herfindahl index violates Property III part (a), but this can easily be remedied by taking nh instead of h.

[10] For other purposes other cardinalizations are clearly appropriate. For example if one assumes that the *Social Welfare Function* is additively separable then a suitable welfare-theoretic cardinalization is that given in Atkinson (1970), Blackorby and Donaldson (1978), and the decomposition formula might then be redefined in terms of welfare levels—see Blackorby *et al.* (1981). But this step involves much stronger assumptions about the Social Welfare Function than I have insisted upon here—see Cowell (1982).

For the purposes of this paper I think the answer can be provided quite simply from the general formulae (2), (3). Let groups 1 and 2 have identical mean income but differing inequality. Suppose there are now some people who move from group 2 to group 1 (for example, migration between regions) in such a way as to leave unaltered both mean income in each group and inequality in each group. Writing (2) in the form of (3) it is easy to see that the change in overall inequality dI resulting from such a population shift dw is given by

$$\frac{\partial I}{\partial w}=[F^{-1}(I_1)-F^{-1}(I_2)]\frac{\mu_1^\alpha}{\mu^\alpha}F'\left(\sum_{g=1}^{G} v_g^\alpha w_g^{1-\alpha}F^{-1}(I_g)+F^{-1}(I_B)\right). \tag{4}$$

The last term in (4) implies that the impact of this population shift depends on the intra group inequality of groups $3, \ldots, G$ and the inter group inequality elsewhere, unless F' is constant. There appears to be no good reason why the effect of such population changes should depend on apparently irrelevant information. Accordingly I shall require $F''\equiv 0$ and adopt the simplest possible cardinalization of (3):

$$I^\alpha=\frac{1}{\alpha^2-\alpha}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\mu}\right]^\alpha-1\right] \tag{5}$$

which has the aggregation property

$$I^\alpha=\sum_{g=1}^{G} v_g^\alpha w_g^{1-\alpha}I_g^\alpha+I_B^\alpha \tag{6}$$

where the superscript α on the symbol I has been used as a convenient label for the particular family members of (5).

Finally, then, which of the continuum of measures I^α should be used? There is a good case for excluding large positive or large negative values of α since they make the measure excessively sensitive to, respectively, minor variations in very high incomes and minor variations in very low incomes. The problem is that in almost all data sets on income distribution one has to make some slightly unsatisfactory assumption about how to treat the topmost incomes and the very bottom incomes: it is therefore not a good idea to use an inequality statistic that will be dominated by the precise assumptions made. There is also an obvious appeal in examining the cases $\alpha=0$ and $\alpha=1$ since here we find that the weights assigned to the within group components $\{v_g^\alpha w_g^{1-\alpha}\}$ specialise to, respectively, the population weights $\{w_g\}$ and the income weights $\{v_g\}$. Hence I have used six members of the family (5) with α values from -1 to $+2$, and including 0 and 1. Measures with positive values of α, being particularly sensitive to income differences at the top end of the income distribution, will be known as *top-sensitive* inequality measures. Likewise, measures with negative values of α (sensitive to very low incomes) will be known as *bottom-sensitive* measures.

4. Family Size and Inequality

Let us turn immediately to one of the most obvious ways of partitioning our six thousand families—by the size of unit which they comprise. Family size is an important characteristic influencing the living standard a family can enjoy from a given total income, though, of course, it is not the only such characteristic: an adult and small child may well enjoy quite a different living standard on a given income from that enjoyed by two adults. However subclassification on a finer basis than size did not promise to provide much additional insight to the overall pattern of inequality.

We must now face squarely the issue of defining "income" and "income recipient" which are ducked in the last section. As noted in section 2 a number of methods of totalling income will be examined, but regardless of how this is done we must make allowance for the fact that for any observation in the sample many people may be sharing in the recorded income, or it may be enjoyed by just one person. So, for family i of size H_i should one be interested in family income y_i or in family income per head y_i/H_i? An argument may be advanced that neither is entirely appropriate: instead what one might be interested in is the income that an adult would have to have in order to enjoy the same standard of living as he presumably enjoys in the particular family within which he actually lives. Clearly an index of needs constructed using the information on family composition is required; the PSID tapes provide such an index based on the indices in the Orshansky poverty scales. Whilst a more extensive study would obviously require the investigation of a number of alternative ways of specifying such an index, for present purposes this criterion alone will be applied. Normalizing the index of needs q_i at unity for the average single person, we may define personal equivalent normalized needs-adjusted income (PENNI) $z_i \equiv y_i/q_i$.

Of course in view of the fact that there are (at least) three concepts of income in which we may be interested, y_i, y_i/H_i, z_i there are also (at least) three concepts of income receiver, the family itself, the number of persons in the family H_i, the number of "equivalent adults" in the family q_i. In principle this gives us a total of nine possible combinations of definitions of income and of income recipient.[11] For expositional purposes Table 1 has been constructed for the hypothetical case of a two-person family in which $y_i = \$30{,}000$, $q_i = 1.5$ to illustrate these nine possibilities. Straightaway one sees that some of the nine are unlikely to be particularly interesting—for example the first two entries in the right hand column. I have concentrated on just five of the possibilities labelled A to E. Case A represents the naive approach that takes no account of family composition at all

[11]On this point see also van Ginneken (1981). There are two further intractable difficulties. One is the lack of treatment of income differences within families. Frankly it is impossible to derive any reasonable estimates of how total family income is translated into a separate component of "command over resources" for *each* family member, and so I have ignored such possible differences and assume equal shares within each family. The second problem is what to do about children. The issue of whether "adult equivalent" weights derived from expenditure data should be used for welfare comparisons (see Pollak and Wales, 1979) has largely been met by our adoption of case E in Table 1. Yet, should new-born babies be counted along equally with adults in the weighting of persons as income receivers? Or, should children be treated as consumption goods—whose necessary maintenance costs ought to be deducted from their parents incomes? (Dinwiddy, 1980; Garfinkel and Haveman, 1978). The latter approach seems indefensible for a study of this sort where new households are formed dynamically within the sample, by teenage children "splitting off" from their parents. However, I did try a modified form of computing cases D and E by weighting each family only by the number of *adults* (18 years of age or over). The results were similar to case E, although the reductions in inequality were not quite as large. This is also the effect of weighting not by individuals (as in cases D and E) but by numbers of equivalent adults.

TABLE 1

CONCEPTS OF INCOME AND INCOME RECIPIENT

Income Concept	Income Receiving Unit: Family	Person	Equivalent Adult
Total dollar income	(A) 1 unit receiving $30,000	2 units receiving $30,000 per unit	1.5 units receiving $30,000 per unit
Dollar income *per capita*	(B) 1 unit receiving $15,000	(D) 2 units receiving $15,000 per unit	1.5 units receiving $15,000 per unit
PENNI[a]	(C) 1 unit receiving $20,000	(E) 2 units receiving $20,000 per unit	1.5 units receiving $20,000 per unit

[a]Personal equivalent normalized needs—adjusted income.

(a procedure that is still often used!). Cases B, C and D each represent partial improvements on this: B and C assign respectively the income per head or the PENNI to each family; D puts the income per head assignment on an *individual* rather than a *family* basis.

Case E seems to be the most interesting combination of assumptions, for the following reasons. Presumably (in the United States anyway) social welfare depends on the well-being of individual persons, regardless of the units in which they happen to live, the alliances they form, or whether or not they live at home: hence we focus on column 2. Presumably also we are interested in the living standards to which an income gives rise in order to compare two persons living in different types of family on a sensible basis: so z_i, row three, is the appropriate choice. Whilst one cannot claim the "refined" method E is perfect it does seem to be about the best one can do given the data limitations and an individualistic approach to economic and social questions. Some economists who are accountants under the skin may worry about selecting *any* off-diagonal entry of the array in Table 1 since, applying the analysis to the whole economy, the implicit "total income" (i.e. sum over all units of the income accruing to each unit) will not tally with total personal income in the national accounts, and indeed will differ if people regroup into different families. However this need not cause undue concern since the use of z_i rather than y_i/H_i implies that one is taking into account in some measure the economies of scale of living in families, and these economies of scale will neither be reflected in the conventional national accounts nor be invariant under changes in family structure.

We may now draw up the top half of Table 2 where we note that since each column refers to a subgroup of the population consisting of families of a given size H_g, row 2 is simply formed by writing the number of persons in group g as $m_g \equiv H_g n_g$. In row 5 ζ_g is found simply from the arithmetic mean of the z_is in group g. Note that μ_g generally increases (unsurprisingly) with H_g, whilst μ_g/H_g decreases from $g=2$ onwards. We see that ζ_g follows a pattern similar to that of μ_g/H_g, but there is a much larger gap between ζ_1 and ζ_2 due to the economies

TABLE 2

THE STRUCTURE OF FAMILY INCOME DISTRIBUTION BY SIZE OF FAMILY

		No. of Persons in Each Family H_g						
		1 Person	2 Persons	3 Persons	4 Persons	5 Persons	6 or more Persons	All Families
No. of Families	n_g	1,328	1,485	1,147	933	515	584	5,992
No. of Persons	m_g	1,328	2,970	3,441	3,732	2,575	4,105	18,151
Mean Family Income	μ_g	8,249	16,759	18,862	21,043	21,264	22,258	16,102
Mean Family Income per Person	μ_g/H_g	8,249	8,379	6,288	5,261	4,253	3,408	6,923
Mean PENNI	ζ_g	8,249	12,438	11,911	10,670	9,038	7,606	10,397
Within-Group Inequality I_g^α (case A)	$\alpha = 2$	0.322	0.375	0.416	0.287	0.217	0.226	
	1	0.262	0.261	0.240	0.207	0.175	0.202	
	$\frac{1}{2}$	0.262	0.253	0.225	0.197	0.171	0.206	
	0	0.283	0.266	0.234	0.202	0.179	0.223	
	$-\frac{1}{2}$	0.335	0.304	0.268	0.223	0.200	0.259	
	-1	0.455	0.381	0.345	0.267	0.243	0.332	

of scale of living together rather than singly. Now examine the bottom half of Table 2, which reports the results of inequality within each size class.[12] The striking thing here is that except for very large families (six or more persons) inequality I_g^α declines systematically with g; for larger sized families the sub-sample becomes too small to derive reliable results.[13]

Consider the effect on overall inequality that the features noted in the last paragraph are likely to have. Firstly it appears that there should be less inequality between groups in terms of the PENNI than in terms of total family income or family income per head. Secondly the within group component of inequality is likely to be less when *persons* are taken as the basic units than when families are taken as basic units. Thirdly, if *per capita* income is used ((μ_g/H_g) in the table) there is likely to be *more* inequality amongst persons than amongst families because of the consequent large number of very small incomes.

Let us support this intuitive reasoning with some more formal analysis. Table 3 shows the population shares w_g and incomes shares v_g used in the aggregation procedure of equation (6) for the five different cases under examination, derived

TABLE 3

POPULATION AND INCOME SHARES

	Family group						All Families
	1	2	3	4	5	6	
			Families (cases A–C)				
w_g	0.272	0.280	0.164	0.147	0.075	0.062	1.00
v_g^A (Income)	0.139	0.291	0.193	0.192	0.099	0.086	1.00
v_g^B (Income/head)	0.324	0.338	0.149	0.112	0.046	0.031	1.00
v_g^C (PENNI)	0.216	0.334	0.188	0.151	0.065	0.046	1.00
			Individuals (cases D, E)				
w_g	0.100	0.206	0.182	0.217	0.138	0.156	1.00
v_g^D (Income/head)	0.139	0.291	0.193	0.192	0.099	0.086	1.00
v_g^E (PENNI)	0.081	0.250	0.211	0.226	0.121	0.111	1.00

simply from rows 1 to 5 of Table 2. Note the substantial shifting of weights in response to different definitions (in particular contrast row 1 with row 5, or row 3 with row 7). Table 4 displays the between group inequality components I_g^α, and Table 5 displays the resulting values of total inequality for the five cases A–E and the various values of α.

Observe that in nearly every case the size ordering of the between group inequality components is A–D–B–E–C. Observe also that, for this sample at least, when PENNI is used I_B^α becomes small.[14] In other words the naive method A

[12]There are some variations in this table for cases A through to E for obvious reasons, but in view of their being so slight only case A is reported here.

[13]For all inequality estimates standard errors were computed so that overly fine partitioning of subgroups could be avoided and the significance of inequality differences checked. For example although the difference between I_2^2 and I_3^2 is of the wrong sign it is not significant at the 5 percent level.

[14]Although still statistically significant at the 5 percent level.

TABLE 4

BETWEEN GROUP INEQUALITY I_B^α

α	A	B	C	D	E
2	0.050	0.030	0.016	0.048	0.016
1	0.056	0.032	0.016	0.049	0.017
$\frac{1}{2}$	0.060	0.034	0.016	0.050	0.017
0	0.064	0.036	0.017	0.051	0.018
$-\frac{1}{2}$	0.070	0.038	0.017	0.053	0.018
-1	0.077	0.041	0.017	0.056	0.019

clearly yields a substantial and arguably spurious component of inequality between groups arranged by family size. "Correct" imputation of income and weighting by *persons* rather than by families of arbitrary size virtually eliminates this component. On reflection this is not a very amazing result, but it is reassuring to find that the adjustments (from n_g to m_g and from μ_g to ζ_g) alter inequality in the expected fashion and it is interesting to see just how small inequality between family size groups really is. However it should *not* be interpreted as saying that when concepts are correctly defined family size does not contribute to inequality. It does. The (significant) reduction of intra-group inequality with family size ensures this.

TABLE 5

OVERALL INEQUALITY I^α

α	A	B	C	D	E
2	0.408	0.398	0.372	0.425	0.356
1	0.289	0.279	0.249	0.284	0.239
$\frac{1}{2}$	0.286	0.270	0.240	0.272	0.230
0	0.311	0.284	0.251	0.285	0.240
$-\frac{1}{2}$	0.377	0.327	0.290	0.328	0.275
-1	0.528	0.424	0.379	0.424	0.352

Combining within-group and between group contributions we find (Table 5) that the inequality size orderings of the five cases is A–D–B–C–E in every case except extremely top-sensitive measures. Note that adjusting for "correct" income and income receiving unit definitions can reduce overall inequality by about one eighth to one third. But how important is such a reduction in practical terms? To see this examine Table 6 which is constructed in a manner similar to that of Table 5 except that we now examine total income *less taxes*. We expect the deduction of personal taxes to reduce measured inequality—and indeed it does, in every case. What is rather remarkable is that for $\alpha < 1$ the effect of taxes on inequality is *less* than the effect of reweighting the data to adjust for differences in family size. For example $I^{-1/2}$ is 0.377 in case A before tax and 0.313 in case A after tax; but it is reduced to 0.275 in case E before tax.

TABLE 6

OVERALL INEQUALITY AFTER TAX

α	A	B	C	D	E
2	0.288	0.290	0.252	0.306	0.239
1	0.234	0.228	0.195	0.231	0.185
$\frac{1}{2}$	0.237	0.225	0.192	0.226	0.183
0	0.260	0.238	0.204	0.238	0.193
$-\frac{1}{2}$	0.313	0.272	0.234	0.271	0.220
−1	0.430	0.346	0.302	0.342	0.277

5. AGE AND INEQUALITY

For a number of reasons considerable interest has been shown in the dispersion of incomes in different age groups and between such groups. Is there here, perhaps, a substantial contribution to the structure of income inequality which ought to be analysed?

To fix ideas let the *age of family head* determine the group to which the family is assigned, and take as age groups those designated in the first six columns of Table 7. The first four rows of Table 7 should be self-explanatory since they correspond to rows in Table 2. Evidently there is now likely to be substantial heterogeneity by family size within each group and so I have reported within-group inequality for both the "naive" case A and the "refined" case E. Clearly the choice of groupings is somewhat arbitrary, so I tried a fairly fine partition

TABLE 7

INEQUALITY BY AGE GROUP

		Age of Family Head						
		<25	25–34	35–44	45–54	55–64	≥65	All Families
No. of families	n_g	942	1,714	900	924	760	752	5,992
No. of persons	m_g	2,086	5,388	3,808	3,436	2,077	1,235	18,151
Mean family income	μ_g	9,294	15,897	21,244	23,095	16,891	9,626	16,102
Mean PENNI	ζ_g	6,890	10,423	11,084	12,564	12,516	8,456	10,397
I_g^α Within-group inequality (case A)	α = 2	0.209	0.191	0.224	0.407	0.253	1.005	
	1	0.195	0.164	0.191	0.265	0.234	0.444	
	$\frac{1}{2}$	0.207	0.166	0.194	0.255	0.246	0.382	
	0	0.242	0.178	0.212	0.272	0.277	0.369	
	$-\frac{1}{2}$	0.323	0.207	0.257	0.325	0.338	0.392	
	−1	0.536	0.267	0.365	0.441	0.458	0.454	
I_g^α Within-group inequality (case E)	α = 2	0.164	0.184	0.255	0.457	0.238	0.875	
	1	0.156	0.162	0.215	0.266	0.213	0.391	
	$\frac{1}{2}$	0.165	0.164	0.217	0.246	0.221	0.334	
	0	0.186	0.175	0.234	0.250	0.245	0.319	
	$-\frac{1}{2}$	0.232	0.201	0.271	0.280	0.299	0.333	
	−1	0.340	0.253	0.347	0.347	0.422	0.379	

into ten age groups, defined in the horizontal axis of Figure 2, and a coarser partition of six groups, designated in the column headings of Table 7. The results turned out to be so close that it seemed much simpler to present only the "coarse partition" computations in detail.

It is clear that for most values of α and for most of the age range inequality increases with age, although this effect is less marked than has been noted elsewhere in the case of the inequality of individual labour earnings. Also it appears that for given α and g the value of I_g^α is usually smaller in case E than in case A—which is what one would expect from the last section. Hence in view also of the fact that switching from weighting by families to weighting by persons concentrates the $\{w_g\}$ more in the middle-age ranges (see Figure 2) which exhibit less inequality, we expect less *within* group inequality in case E than in case A. This indeed turns out to be true and may easily be checked by subtracting the entries in Table 8 (below) from the corresponding entries in Table 5.

TABLE 8

BETWEEN AGE-GROUP INEQUALITY I_B^α

α	A	B	C	D	E
2	0.049	0.008	0.017	0.009	0.009
1	0.051	0.008	0.017	0.009	0.010
$\frac{1}{2}$	0.053	0.008	0.018	0.009	0.010
0	0.055	0.008	0.018	0.008	0.010
$-\frac{1}{2}$	0.057	0.008	0.019	0.008	0.011
-1	0.060	0.009	0.019	0.008	0.011

However it is the *between* age group components themselves that are particularly interesting. From a number of studies one has come to accept as standard the rising, concave trajectory of individual earnings plotted against age. This is to some extent reflected in the behaviour of mean family income in this sample—see μ_g in row 3 of Table 7. Does this mean then that there is a substantial component of family income inequality that is due purely to the typical path of household income over the life cycle? This is not necessarily so because, of course, we are only looking at crude *unadjusted* family income. When we examine ζ_g, the mean PENNI for group g, in row 4 of Table 7 we find that the life cycle effect on mean income is less pronounced, particularly in the middle age groups. Now recall that "correctly" counting the units of the population as people rather than families shifts the population shares $\{w_g\}$ substantially towards precisely those age groups: presumably this will reduce the inter-age group inequality component.

This supposition is reinforced by Figure 2; notice that in case E the income shares v_g are much more closely matched to the population shares w_g than they are in case A. In fact the quantitative results are even more remarkable as one can see in Table 8. Notice that the between group inequality component in the naive case A is fairly high—some twelve to twenty percent of overall inequality, depending on the precise measure you use. Now go from case A to the refined

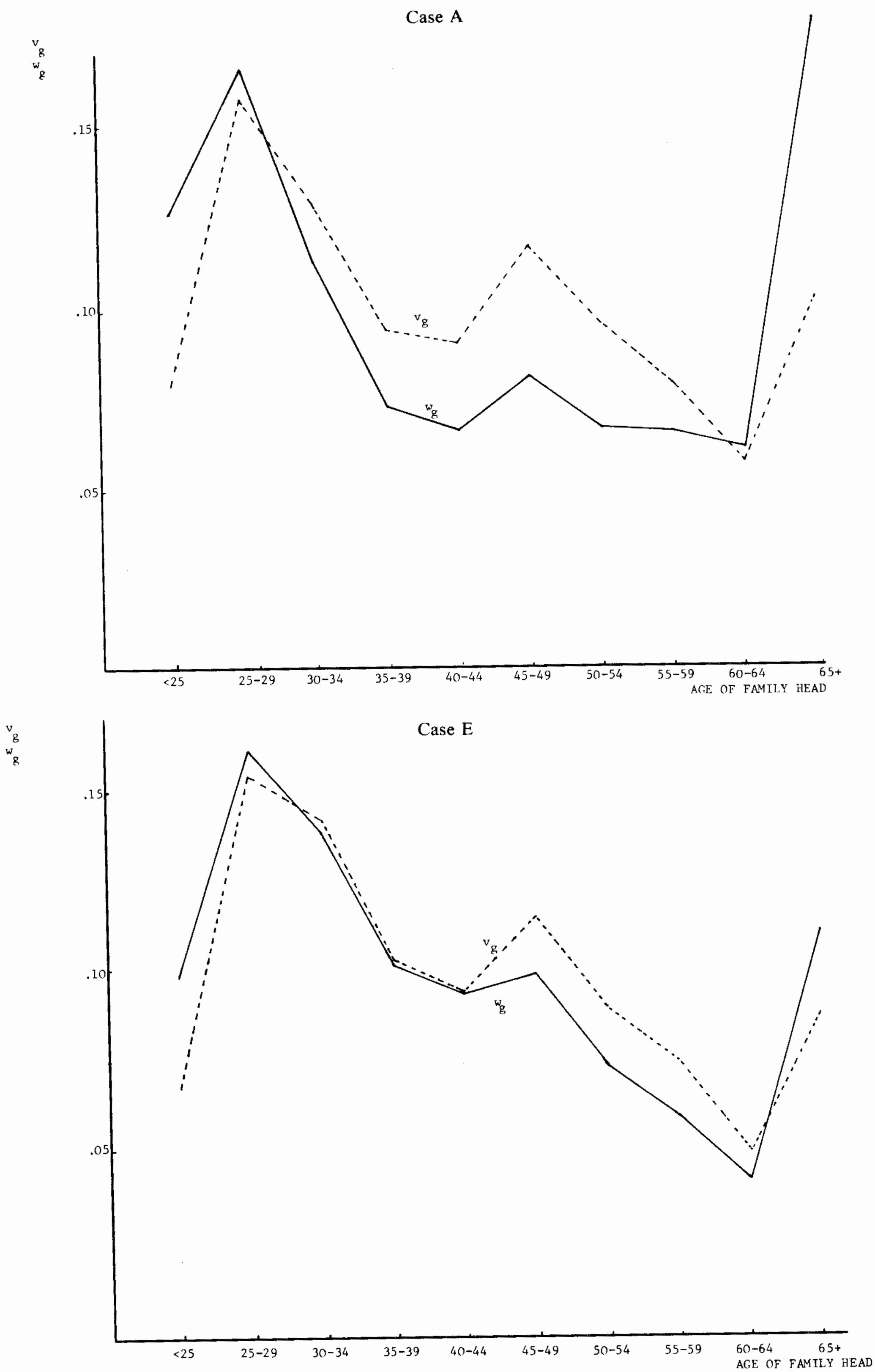

Figure 2. Income and Population Shares

case E: the between-group component is now less than one fifth of its former size and accounts for a mere three or four percent of overall inequality. As with the family size decomposition going from case A to case E greatly reduces the importance of inter-group inequality. But note by contrast that in the case of the age decomposition *every* assumption other than A yields small between group components: so the dramatic "shrinking" of the between group component is not due to a particular arbitrary choice of income and of income recipient thus defining away the problem.

It is interesting to note that this is not just a special feature of the particular definition of income used. If instead we employ "family factor income," we find a similar feature appearing for the between-group inequality component in cases A and E. Table 9 contains the between group inequality and total inequality

TABLE 9
INEQUALITY OF FACTOR INCOME
6 Age Groups

	Between Age Group Inequality I_B^α		Total Inequality I^α	
α	A	E	A	E
2	0.063	0.016	0.480	0.424
1	0.069	0.017	0.357	0.299
$\frac{1}{2}$	0.073	0.018	0.384	0.310
0	0.079	0.019	0.503	0.383
$-\frac{1}{2}$	0.086	0.020	0.942	0.645
-1	0.095	0.021	3.590	2.020

values corresponding to those in Tables 5 and 8. Notice that one still finds a dramatic reduction in between-group inequality as one switches from A to E. So the "disappearance" of between-age-group inequality does not result from an offsetting pattern of transfers amongst members of different age groups, but is due to the normalization of measured incomes and the correct weighting of families of different sizes.

However, is this result due perhaps to age of family head acting as a sort of proxy for family size? After all one expects the very young and the very old to have rather small families by contrast with those in the middle age ranges. We can examine this by doing a double decomposition analysis by the 6 family size groups and 6 age groups—36 groups in all. This enables one to calculate an "interaction term" for the two partitions, namely $J = I_{B12} - I_{B1} - I_{B2}$ where I_{B12} represents the between-group component for the 36 family-size-and-age partition and I_{B1}, I_{B2} represent the between-group components for the partition by family size and the partition by age, respectively. A large, negative J (of the order of magnitude of I_{B1} or I_{B2}) would indicate that one partition was really just acting as a proxy for the other. Now examine Table 10. If the data are treated in the "naive" fashion (case A) then there is a fairly large negative interaction—though in absolute value only one quarter to one half of I_{B1} or I_{B2}. In the "refined" presentation (case E) the interaction term is now small and positive. Notice that in each case the contribution of I_{B2} is smaller than that of I_{B1} by a statistically significant amount.

TABLE 10

INTERACTION TERM FOR 6 FAMILY SIZES AND 6 AGE GROUPS

α	Total	Between Family Size I_{B1}	Between Age Group I_{B2}	Between Family Size and Age Group I_{B12}	Interaction J
			Case A		
2	0.408	0.050	0.049	0.078	−0.021
1	0.289	0.056	0.051	0.084	−0.023
$\frac{1}{2}$	0.286	0.060	0.053	0.089	−0.024
0	0.311	0.064	0.055	0.096	−0.023
$-\frac{1}{2}$	0.377	0.070	0.057	0.106	−0.021
−1	0.528	0.077	0.060	0.120	−0.017
			Case E		
2	0.356	0.016	0.009	0.033	0.008
1	0.239	0.017	0.010	0.034	0.007
$\frac{1}{2}$	0.230	0.017	0.010	0.034	0.007
0	0.240	0.018	0.010	0.035	0.007
$-\frac{1}{2}$	0.275	0.018	0.011	0.036	0.007
−1	0.352	0.019	0.011	0.037	0.007

This throws light on an issue that has bothered several economists in recent years:[15] should conventional inequality statistics be adjusted for a between-age-group component that (arguably) is a spurious contribution to overall inequality? Whilst leaving open the issue of whether this is desirable in principle, the answer in practice seems to be that as far as the inequality of family welfare is concerned it really does not matter very much—*provided that you have used appropriate definitions of "income" and "income recipient" in the first place.* Using as an example $I^{-1/2}$ for total family income we find that the simple correction from case A to case E reduces inequality from 0.377 to 0.275; if one further deducted the inequality due to inter-age group differences one would find that $I^{-1/2}$ only fell to 0.264.

6. THE TIME PERIOD AND INEQUALITY

We turn now to an issue on which the Michigan PSID provides a superb source of information: the pattern of income distribution over periods longer than one year. It is well known—and indeed elementary reasoning suggests the

[15]The literature that was spawned by Paglin (1975) is the most obvious example. The major difficulties with interpreting that literature are (i) exclusive attention is devoted to the Gini index, which, as we have seen, is unsuitable for the kind of decomposition analysis considered here and in the Paglin literature; (ii) no criteria of significance are considered.

result—that the dispersion of individuals' annual incomes is usually greater than the dispersion of incomes averaged over longer periods of time. But can this result be confirmed for *family* income inequality in view of the fact that families are a heterogeneous, and often temporary, grouping of persons who pool income? How important is the period effect compared with the family size or the age effect?

To give a precise answer to these questions some further methodological issues must be cleared up. Firstly consider the way in which changes in family composition in the panel are handled. Let i be a family in the current "wave"—that is, the current version of the panel—and let $t-1$ denote the number of years ago that a particular data item was recorded ($t=1$ is the current year, $t=2$ is last year, $t=3$ the year before last and so on). Denote by y_{it} the total income of the family in which the current head of family i was then living. Note that it is *not* the total income $t-1$ years ago of all the current members of the family i, a concept which is difficult to measure accurately because of the separate tracing of income histories of several persons. Note also that the *same* piece of data may appear as y_{it} and y_{jt} for two different families i and j and $t \geq 2$. The reason for this is that "splitoffs" are traced and their incomes are recorded but, if suitably handled, need not lead to double counting.

The procedure can be illustrated by following the fortunes of two couples. Anne and Bill each earn a steady income: they met two years ago and were each in the panel then; last year they married; but, unfortunately, this year they have separated again. Charlie comes from a big family that was in the panel two years ago, and was then living with his folks; Charlie left home last year and met Diana, a lady of independent means but outside the panel; Charlie and Diana are now married. This story is set out schematically in Figure 3. How the data then appear in the current wave is shown in Table 11 where each entry contains the dollar income y_{it} and, in parenthesis, the number H_{it} giving the corresponding family size of the family at time t. Note two points here. One, since Diana married into the panel her \$30,000 income at $t=2$ will not be recorded, though if she subsequently walks out on Charlie her income may subsequently be traced as that of a "splitoff" family. Two, the rest of Charlie's family may still be in the panel either as a single unit or as a number of splitoffs. Each splitoff family would have the same "income history" up to the time of the split.

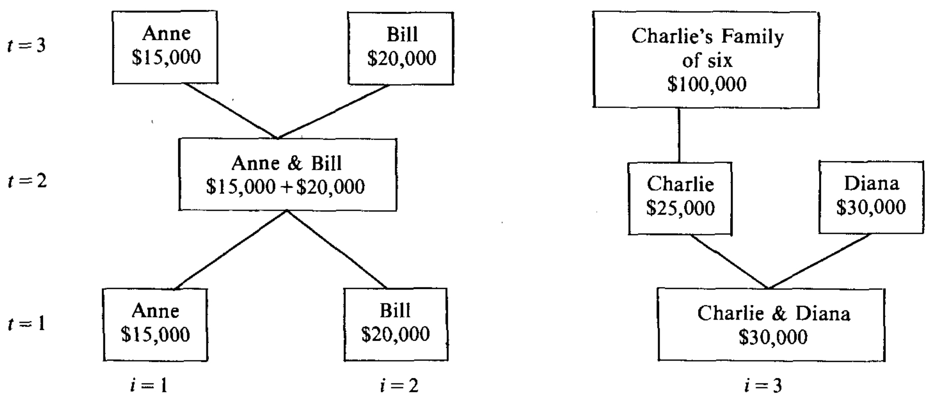

Figure 3

TABLE 11

$t \backslash i$	1	2	3
1	\$15,000 (1)	\$20,000 (1)	\$30,000 (2)
2	\$35,000 (2)	\$35,000 (2)	\$25,000 (1)
3	\$15,000 (1)	\$20,000 (2)	\$100,000 (6)

Given the characteristics of the family t years ago to which y_{it} and H_{it} refer we may obviously compute the PENNI at time t, z_{it}. This provides the required measure of income flow which we may then cumulate using the simple discounting formula

$$Z_i(r, T) = \sum_{t=1}^{T} z_{it}[1+r]^{t-1} \tag{7}$$

where r is a discount rate that has yet to be specified and the z_{it} are measured in real terms. So the interpretation of $Z_i(0.10, 7)$, say, is "the normalized income (PENNI) of the head of household i cumulated at 10 percent per annum over the last seven years." Obviously the distribution amongst persons of $Z_i(r, T)$ for some suitable values of r and T is going to be of great interest in examining real income inequality in the longer run, so presumably one wants an inequality statistic of the form[16]

$$I^{\alpha}(r, T) = \frac{1}{\alpha^2 - \alpha}\left[\sum_{i=1}^{n} \frac{H_{i1}}{m}[Z_i(r, T)/\zeta(r, T)]^{\alpha} - 1\right] \tag{8}$$

corresponding to the "ideal" inequality index for annual income (case E of the previous section). There is in fact a snag with this, to which I return below.

Table 12 reveals that the temporal averaging process implicit in taking progressively longer periods T makes a great deal of difference to measured inequality. This is despite the fact that we have taken care to cumulate the PENNI's in each period, *not* total family income or family income per head which may fluctuate more erratically because of changes in family size. In fact taking $Z_i(0.10, 10)$ as the income concept rather than z_i we find that measured inequality is reduced by between 27 and 38 percent—compare column 10 of Table 12 with column 1 (which corresponds to case E in Table 5).[17] Notice that

[16] $\zeta(r, T)$ is the mean of the $Z_i(r, T)$ and $m \equiv \sum_i H_{i1}$.

[17] In the case of incomes of a particular cohort of *individuals* it is well known that slightly different results are obtained according to whether one uses *backward* cumulation as here (comparing the inequality of $\{Z_i(r, T)\}$ with the inequality of $\{z_{i1}\}$) or *forward* cumulation (using $\{z_{iT}\}$ for comparison)—see Shorrocks (1981) for a discussion of this. However, in view of the changing family composition this becomes conceptually rather complex in the present case so the simple version of Table 12 has been presented. Nevertheless, whether one cumulates forwards or backwards, one still obtains a substantial reduction in inequality as the averaging period is lengthened—see also Benus and Morgan (1975).

TABLE 12

INEQUALITY VALUES I^{α} WHEN INCOME IS CUMULATED OVER SEVERAL YEARS

	Cumulated Period (Years) T									
	1	2	3	4	5	6	7	8	9	10
α = 2	0.356	0.423	0.508	0.384	0.325	0.284	0.259	0.241	0.229	0.219
1	0.239	0.236	0.248	0.222	0.207	0.195	0.187	0.180	0.175	0.171
$\frac{1}{2}$	0.230	0.219	0.222	0.205	0.194	0.185	0.179	0.174	0.170	0.167
0	0.240	0.222	0.221	0.205	0.196	0.188	0.183	0.178	0.175	0.172
$-\frac{1}{2}$	0.275	0.244	0.238	0.222	0.212	0.204	0.199	0.194	0.191	0.189
-1	0.352	0.289	0.277	0.258	0.246	0.236	0.230	0.225	0.222	0.220

the reduction in inequality is much more rapid for bottom-sensitive inequality measures than for the top sensitive measures.[18]

Comparing Tables 5 and 12 we find that we started out with a "naive" inequality statistic for $I^{-1/2}$ of about 0.377 and have ended up with a value of 0.189.[19] Similar reductions are found for other values of α. Does this then mean that "true" income inequality is less than half of what it appears to be? The answer obviously depends on one's interpretation of year-to-year fluctuations in people's PENNIs. If one believes that rich and poor alike have access to efficient credit markets enabling them to "smooth out" all income fluctuations foreseen and unforeseen, then clearly the asymptotic value of $I^{\alpha}(r, T)$ is what one wants, and transitory inequality is irrelevant. If one believes that in the main income fluctuations impose immediate and unavoidable hardships, then transitory inequality *is* important—at least for one's conclusions about economic welfare. Of course even in this case the long run value of $I^{\alpha}(r, T)$ would still be interesting in that one is presumably interested in that component of inequality which is in a sense "permanent" and cannot be ascribed to year-to-year fluctuations.

Finally, note a difficulty with the interpretation of $I^{\alpha}(r, T)$ as given in (8), which is virtually unavoidable at this stage. In the construction of $I^{\alpha}(r, T)$ we have to use a unique set of population weights. This means either that we simply assign each family a weight $1/n$ (case C of section 4) which we have seen is undesirable in a heterogeneous population, or we assign weights such as $\{H_{i1}/m\}$,

[18]Obviously one would like to know whether this very large effect on measured inequality is just the result of some special assumption in taking $I^{\alpha}(0.10, 10)$. However, when interest rates of $r = 0$ percent, 5 percent, 15 percent and 20 percent were used for the computation there was a barely perceptible change in the figures appearing in Table 12. Column 1 remains unaffected, of course, but even column 10 where one would expect maximum effect of the change in the discount rate changes very little. This is corroborated by Shorrocks (1981) who noted that his results were barely affected by switching from nominal to real income or to real discounted income. Moreover there would still be dramatic reduction in inequality even if families with temporarily very low incomes were eliminated from the sample. To check that the figures for Table 12 were not given a substantial bias because of the presence of a few families with very low PENNI values in some years all those with $z_i \leq \$500$ were dropped and the inequality statistics recomputed. Although $I^{\alpha}(0.10, 1)$ was noticeably reduced for $\alpha < 0$, $I^{\alpha}(0.10, 10)$ hardly changed. Inequality was still dramatically reduced.

[19]As an indication of how important this is, the overall effect of such adjustments is of the same magnitude as a combined income equalization amongst all one-person families, amongst all two-person families and amongst all three-person families.

as in equation (8), which effectively imposes the *current* population mix on an aggregate that subsumes substantial family composition changes over the decade. However, irrespective of which imperfect set of weights is chosen, the pattern of inequality reduction with T remains unaltered.

7. Concluding Remarks

If we are interested in the *structure* of inequality within the population then there is an extremely strong case for using a technique of inequality measurement that permits a satisfactory decomposition analysis. Doing so reveals some rather interesting features. Whether we look at the family size or at the age characteristic of the sample, the imputation of personal incomes (from family incomes) and the counting of *people* rather than family units of arbitrary size together play a vital role in colouring the picture of overall income inequality. Firstly, there is noticeably *less* overall inequality than would appear from a crude analysis of the raw data. Secondly inter-group inequality is very small, once such basic adjustments are made.

Nevertheless, age and family size *do* have an impact—though not principally on the inter-group component. There is less inequality amongst big families (of a given size) than amongst small families or amongst isolated individuals. There is on the whole greater inequality amongst the old, and amongst the very young than amongst the young-to-middle-aged. Both these conclusions are independent of the adjustments made to the size of incomes and numbers of income recipients.

Moreover we can throw some light on an interesting question concerning relative magnitudes—of the multitude of data refinements and adjustments concerning income, the receiving unit, the accounting period, and so on, which really matter? The answer is obviously that when one is dealing with families the particular assumption one makes in Table 1 is significant, and so also is the choice of accounting period, with the latter effect being slightly more powerful over a ten year period than the former. (By contrast "taking out" a possibly spurious age component to eliminate systematic life cycle variation has very little effect.) However the quantitative importance of these two issues raises a number of problems for further research. Clearly the role of the structure of "adult-equivalence" scales on between-group inequality needs to be more thoroughly investigated[20]. Also the impact of *family* instability on income stability and income inequality raises little-explored conceptual and empirical issues that go beyond the scope of this paper. Finally there is also further theoretical and practical work to be done on the correct interpretation of inequality measures relating to periods of different lengths when there is substantial stochastic income variability.

Appendix

The PSID has followed the economic fortunes of a nationally representative sample of American families annually since 1968, starting off with about 5,000 families (about 18,000 individuals). A valuable feature of the PSID data is that

[20]On this see also Lazear and Michael (1980) and the Appendix to this paper.

it attempts to follow all sample individuals in the original families, including those who leave home, and thus progressively builds up an extensive set of background information about the family head and spouse, including those for "new" families. The sampled families were originally visited personally and the family head interviewed if available, but since 1973 interviewing by telephone has become the usual procedure. The PSID started as an improvement on the Survey of Equal Opportunity (SEO) with a fresh cross-sectional sample from the PSID's own national sampling frame in 1968, by which time more interest had been focussed on understanding change in family economic situations than in merely counting or describing the poor. Hence, in 1968, the first year of the PSID study, about 2,000 families were those of the Census' SEO, with income less than 1.5 times the official poverty line, and a further 3,000 families were a fresh probability sample from the Survey Research Centre's national sampling frame. The SRC publish sampling weights which allow for this structure, to take into account differential response rates and to take into account the self replacing nature of the sample over time. The results reported in the paper used these weights and were checked for the sample that excluded the SEO members.

A small number of very large incomes in later years had been rounded down to conform to data handling requirements. However I am grateful to Professor G. Duncan for supplying the original observations, thus avoiding the downward bias in top-sensitive measures. In order to avoid bottom-sensitive measures assuming meaningless values sample members with impossibly low incomes (i.e. $z_i \leq 0$) were excluded. In the case of total family income this involved excluding twelve families from the original sample. In the case of factor income this involved dropping 234 further cases. So the results were reworked using top-sensitive measures only and *including* the zeros: the conclusions remained unchanged. In the analysis of $I^{\alpha}(r, T)$ we also considered excluding those with total family income less than \$500. Table 13 below shows the number of families of the "Non-SEO" sub-sample with incomes below \$500 in the interview years from 1968 to 1977.

TABLE 13

Year	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
No.	46	42	39	32	26	24	19	14	13	8

Of these about one quarter had incomes less than \$100. As can be seen the relatively small numbers throughout the years tell us that the loss in the sample size is a relatively minor one.

Three basic concepts need to be more carefully defined. They are: "Family," "Income" and "PENNI." A family or a family unit is defined as a group of persons living in a household, who are related by blood, marriage (including de facto marriage), or adoption. In occasional cases, an unrelated person has been included in the family unit if he or she shares expenses and is apparently a permanent member of the unit. The definition of a family used in this study includes single person families. Moreover, the family composition contains several dimensions, most of which are related to the family's position in the standard life cycle: marriage, birth of first child, the youngest child who has reached age six and started school, children who have left home, one spouse who died. The sex and marital status of the head, the number of children, and the age of the youngest are the main components. That the study is longitudinal means there is one record for each family extended over the years. Where there are several families derived from an original family (the case of split-offs), the early family information appears on each of their records.

Total income is the regular money income consisting of taxable income and transfers. The taxable income includes the labour part of farm income and of unincorporated business income, the incomes from wages, bonuses, overtime, commissions, and professional practice or trade, the labour part of roomers and boarders, farming or market gardening, the "asset" part of farm income, unincorporated business income and of roomers and boarders, farming or market gardening, alimony, income from rent, interest, dividends etc. Transfers include the amount of ADC/AFDC, other welfare payments, social security payments, other retirement pensions and annuities, unemployment pay (including strike benefits), worker's compensation, child support, help from relatives, supplemental security income, etc. Total family income is the total regular money income of all the members of the family, total income and family being defined above. Family factor income is defined as the total family income minus transfers (as defined above). And disposable income is the total family income minus taxes, where the taxes are the total "*estimated*" Federal income taxes.

Family money income is adjusted according to the needs of the family in question to arrive at the PENNI. This needs adjustment procedure is simply the family money income divided by the family needs. For farmers, this ratio is multiplied by 1.25 to allow for lower costs of food to them. The "family needs" is the Annual Need Standard, which is the Orshansky type poverty threshold, based on an annual food needs standard derived from the weekly food costs, which itself is based on U.S. Department of Agriculture low cost plan estimates of weekly food costs. Family needs are estimated by converting the weekly food costs to an annual amount and adjusting for economies of scale by USDA rules as follows.

Single person	add 20 percent
Two	add 10 percent
Three	add 5 percent
Four	0
Five	reduce 5 percent
Six and more	reduce 10 percent

An additional adjustment for diseconomies of small households (in rent etc.) was made as follows for the Annual Need Standard.

Single person	4.89 times the food needs
Two-person unit	3.70 times the food needs
All other units	3.0 times the food needs.

Clearly other adjustment procedures could be considered to obtain the PENNIs for each household member, though these are not reported here for reasons of space. Lazear and Michael (1980) provide such a procedure based on household production theory, and a comparative analysis using BLS consumer expenditure survey data. The principal difference between their scale and the Orshansky scale is in the relative weighting of single persons and two-person families. Taking the single person index needs q_i as unity, the Orshansky q_i index for a two-person family is 1.26 (our modified version yields 1.38); Lazear and Michael's q_i for a two-person family is 1.06—so two very nearly live as cheaply as one! However, even if we adopt this alternative scale, there is little affect on the between group inequality component in the "refined" case E. I_B^2 becomes 0.014, I_B^0 becomes 0.016, I_B^{-1} becomes 0.018—compare Table 4. The reason for this tiny difference is that there is less relative difference between the two scales for *large* families; and large families have high population shares in case E.

References

Atkinson, A. B., On the Measurement of Inequality, *Journal of Economic Theory*, 2, 244–263, 1970.

Benus, J. and Morgan, J. N., Time Period, Unit of Analysis and Income Concept in the Analysis of Income Distribution, in Smith, J. D. (ed.), *The Personal Distribution of Income and Wealth*, NBER, New York, 1975.

Blackorby, C. and Donaldson, D., Measures of Relative Equality and Their Meaning in Terms of Social Welfare, *Journal of Economic Theory*, 18, 59–80, 1978.

Blackorby, C., Donaldson, D., and Auersperg, M., A New Procedure for the Measurement of Inequality Within and Among Population Subgroups, *The Canadian Journal of Economics*, 14, 665–685, 1981.

Bourguignon, F., Decomposable Inequality Measures, *Econometrica*, 47, 901–920, 1979.

Cowell, F. A., On the Structure of Additive Inequality Measures, *Review of Economic Studies*, 47, 521–531, 1980.

———, Social Welfare Functions and Income Inequality, mimeo, London School of Economics, 1982.

———, Inequality Decomposition: Three Bad Measures, mimeo, Social Justice Project, RSSS, Australian National University and London School of Economics, 1983.

Cowell, F. A. and Kuga, K., Inequality Measurement—An Axiomatic Approach, *European Economic Review*, 15, 287–305, 1981a.

———, Additivity and the Entropy Concept: an Axiomatic Approach to Inequality Measurement, *Journal of Economic Theory*, 25, 1981b.

Cowell, F. A. and Shorrocks, A. F., Inequality Decomposition by Population Subgroups, LSE mimeo, 1980.

Danziger, S. and Taussig, M. K., The Income Unit and the Anatomy of Income Distribution, *Review of Income and Wealth*, 25, 365–375, 1979.

Dinwiddy, R., The Concept of Personal Income in the Analysis of Income Distribution, Government Economic Service Working Paper, No. 30, Departments of Environment and Transport, London, 1980.

Garfinkel, I. and Haveman, R. H., Capacity, Choice and Inequality, *Southern Economic Journal*, 45, 421–431, 1978.

Hanna, F. A., The Accounting Period and the Distribution of Income, in Hanna, F. A., Pechman, J. A. and Lerner, S. M., *Analysis of Wisconsin Income*, NBER, New York, 1948.

Lazear, E. P. and Michael, R. T., Family Size and the Distribution of Real *per capita* Income, *American Economic Review*, 70, 91–107, 1980.

Lydall, H., Some Problems in Making International Comparisons of Inequality, in Moroney, J. R. (ed.), *Income Inequality, Trends and International Comparisons*, Lexington, Mass., 1978.

Lillard, L., Inequality—Earnings vs. Human Wealth, *American Economic Review*, 67, 42–53, 1977.

Morgan, J. N., The Anatomy of Income Distribution, *Review of Economics and Statistics*, 44, 270–283, 1962.

Morgan, J. N., David, M. H., Cohen, W. and Brazer, H., *Income and Welfare in the United States*, McGraw-Hill, New York, 1962.

Paglin, M., The Measurement and Trend of Inequality—a Basic Revision, *American Economic Review*, 65, 598–609, 1975.

Pollak, R. A. and Wales, T. J., Welfare Comparisons and Equivalence Scales, *American Economic Review, Papers and Proceedings*, 69, 216–221, 1979.

Shorrocks, A. F., The Class of Additively Decomposable Inequality Measures, *Econometrica*, 48, 613–625, 1980.

———, Income Stability in the United States, in Klevmarken, A. and Lybeck, J. A. (eds.), *Statics and Dynamics of Income*, Tieto, Clevedon, 1981.

Taussig, M. K., *Alternative Measures of the Distribution of Economic Welfare*, Industrial Relations Section, Princeton University, 1973.

van Ginneken, W., Mexico's Income Distribution, paper presented to EADI conference, University of Paderborn, FRG, 1981.