Review of Income and Wealth
Series 37, Number 3, September 1991

INCOME STRATIFICATION AND INCOME INEQUALITY

BY SHLOMO YITZHAKI

Department of Economics, Hebrew University

AND

ROBERT I. LERMAN

Department of Economics, The American University

This paper develops a new index of stratification that highlights the distinction between inequality and stratification. The stratification index captures the extent to which population subgroups occupy distinct strata within an overall distribution. The indices are group specific and control for group size. A weighted sum of group stratification indices is the third term that, together with between-group inequality and a weighted sum of within-group inequality adds to the overall Gini index of inequality. The paper applies the indices and the decomposition to income inequality by ethnic groups in Israel and by race and family type groups in the U.S.

Economists and sociologists have traditionally analyzed inequality within societies from different perspectives. Economists have focused on *inequality*, usually income inequality; while sociologists have studied *stratification*, often in terms of a combined index (SES) of occupational prestige, education, and income. Although stratification remains a major organizing principle within sociology,[1] sociologists have increasingly analyzed patterns of income inequality, sometimes using standard economic methods such as the Gini coefficient and the Theil entropy measure.[2]

Articles using the term stratification continue to appear, but they typically attempt to determine the impact of the social and economic determinants of *inequality* (usually the variance) of such outcomes as earnings, incomes, occupational level or education.[3] While these analyses may contribute to our understanding of average effects of, say, education on income levels, they provide no basis for making a distinction between stratification and inequality.

Some sociologists have noted the problem of the absence of well-specified concepts and measures of stratification. As Turner put it in his 1984 book, *Societal Stratification* (p. 57):

> If one engages in only a cursory review of the literature on stratification, however, it becomes immediately evident that there is little consensus

Note: We are indebted to Joram Mayshar and Ingram Olkin for helpful discussions and to anonymous referees for helpful comments.

[1]For example, social stratification is a category for grouping articles within *Sociological Abstracts*.

[2]See, for example, Bailey (1985), Allison (1978), and Bolen and Jackman (1985).

[3]Massagli (1987) uses measures of the variance or the log of the variance to explain stratification in occupational status and earnings. Lenski (1984) does not distinguish between inequality and stratification. His Tables 1 and 4, which have income stratification in the caption, present distributions of income.

over what stratification is . . . . Typically, after a number of analytical distinctions are made—say, between inequality, class, status, and power—everything that is separated gets thrown back together and "a" theory is developed about "the" composite phenomenon.[4]

The absence of a well-accepted quantitative definition of stratification has weakened the ability of social scientists to build a convincing literature on patterns over time and differences among groups and geographic areas. This is unfortunate because stratification is a viable concept that is distinct from standard economic concepts of inequality. In this paper, we offer a concrete definition of stratification and develop an index to measure it. The new index captures the relationship between socially meaningful population groupings and socially meaningful orderings (such as income inequality, level of occupational prestige, political power, and/or SES.)

Sociological definitions of stratification rely on the concept of strata. Consider this statement from *The International Encyclopedia of Sociology*, edited by M. Mann (1984), (p. 366):

> *Social Stratification*: The division of a society into a number of strata, hierarchically arranged groupings. These groupings have assumed numerous historical and cultural variations, of which CASTES, ESTATES, and CLASSES are the most familiar. In the 1960s and 1970s, attention also turned to ETHNIC and then GENDER stratification. (*Capitals in original.*)

In Lasswell's 1965 statement, the division of society into strata involves the formation of layers:

> In its general meaning, a stratum is a horizontal layer, usually thought of as between, above or below other such layers or strata. Stratification is the process of forming observable layers, or the state of being comprised of layers. Social stratification suggests a model in which the mass of society is constructed of layer upon layer of congealed population qualities (Lasswell, p. 10, 1965).

We may think of these layers as segments of a distribution. Groups form well-defined layers, or strata, to the extent that their members differ from the rest of the population. Thus, a sound index of stratification should capture the degree of overlap between group members and others. In contrast, the concept of group inequality has to do with similarities and differences *within* the group. The lower (higher) is the index of inequality, the more (less) similar are the members of a particular group.

[4]In his attempt to put precision on the ideas of "class" and "straum", Turner proposes concepts that are similar to ours. He defines two terms:

(a) Differentiation of Homogeneous Subpopulations (DF_{HO}) = the degree and extent to which subsets of members in a society reveal common behavioral tendencies and similar attitudes so that they can be distinguished from other subsets of members in a society; and
(b) Ranking of Homogeneous Subpopulations (RA_{HO}) = the degree to which homogeneous subsets of members in a society can be linearly rank-ordered in terms of their perceived worthiness.

Turner's primary interest is in postulating the causal factors determining the levels of these variables. Unlike our derivation, Turner does not provide a specific way measuring these concepts nor does he embed the measures within a broader concept of inequality.

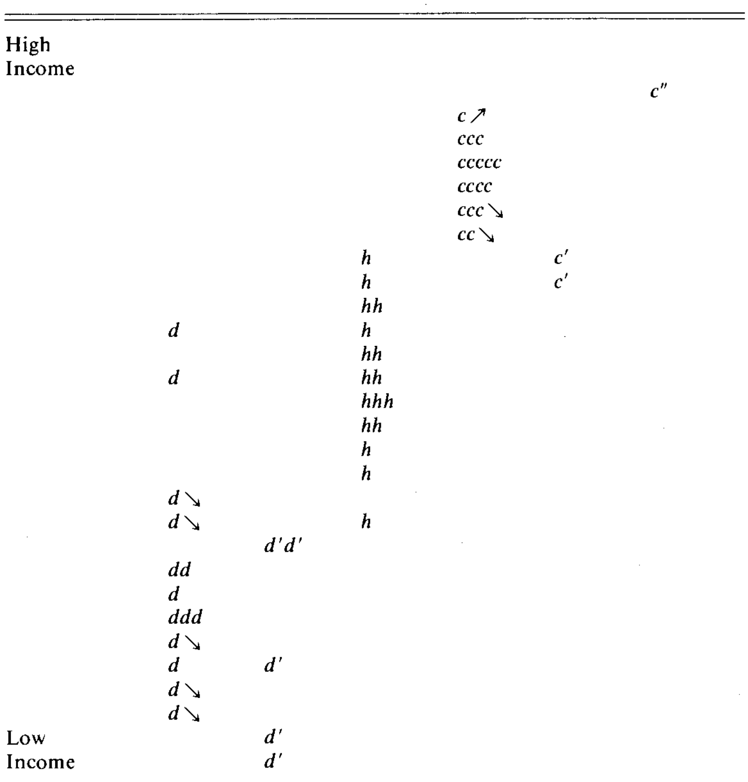

Figure 1. Illustration of Stratification and Inequality Among High School Dropouts (d), High School Graduates (h), and College Graduates (c)

Figure 1 illustrates the distinction between stratification and inequality in a hypothetical distribution by education groups, *d* (high school dropouts), *h* (high school graduates), and *c* (college graduates). In this example, college graduates show the lowest within-group inequality, followed by groups *h*, and then *d*. Stratification is at a maximum for college graduates, while high school graduates and dropouts overlap and thus are less than perfectly stratified. In Lasswell's imagery, the *h* and *d* layers intertwine. Generally, a rise in a subgroup's inequality will reduce the subgroup's stratification, as, for example, a few college graduates, *c*'s, move to the c' positions. However, there is nothing about the two concepts that assures this result. Were the shift simply from *c* to c'', the inequality of *c* would increase with no change in the overlap between group *c* and *h*. Another example is that increased inequality in the dropout distribution (the moves to d') causes a rise in the stratification of groups *h* and *d*. In spite of no change in the inequality of group *h*, the fact that fewer dropouts are in the range of *h* leaves group *h* more segmented from other groups. The increased stratification of dropouts occurs because the members of group *h* in the range of group *d* are farther away from the mean of group *d*.

As Bailey (1985) and Allison (1978) point out, decomposing the Theil index is the common approach for dividing inequality into between-group and within-group components. But such decompositions ignore the stratification dimension illustrated in Figure 1.

This paper broadens the analysis of group distributional patterns. Specifically, we develop methods for determining whether population groups (whether identified by race, occupation, education, or academic department) form distinct strata in terms of a particular hierarchy, such as income, athletic talent, or scientific prestige.[5]

The first step is to develop the stratification index and describe how its properties capture the concept of separate strata, or layers. We produce two indices, one that measures relative stratification and a second that measures the absolute degree of overlapping among groups. Next, we demonstrate how the stratification of subgroups contributes to overall inequality by decomposing the Gini index into between-group inequality, within-group inequality, and a term capturing the degree of subgroup stratification. Section 2 applies our stratification measure and decomposition procedure to income inequality in Israel and the U.S. The focus is on whether ethnic and racial groups form strata within the overall income distribution.

1. Stratification and Inequality

1.1. *The Stratification Index*

Stratification is a meaningful concept for depicting the degree to which groups overlap with respect to any hierarchical measure. The concept has meaningful applications to rankings of income, wealth, education, job status, or other non-economic variables. Consider a group of scientists who publish papers in academic journals. Assume that prestige rankings depend on the number of pages published. Scientists belong to departments and thus have rankings within their own department as well as in the overall scientific community.

Each department's mean published pages (per scientist) might determine the hierarchy of departments. One can distinguish between each department's *inequality* of published pages, which measures differences across department members, and department *stratification*, which measures the segmentation of departments from each other. In the extreme, where members of each department occupy a certain range in the distribution and no member of another department occupy that range, then departments form perfect strata, or horizontal layers, in the overall population.

Suppose the relevant population includes two departments. Department A includes the best and the worst scientists and department B includes the middle level of scientists. In this case, department B forms a stratum, but department A does not form a single group. Knowing someone belongs to department A tells us little until we learn in which group within A he belongs.

[5]In a recent study of how particular colleges determine the occupational success of Japanese males, Miyahara (1988) describes how graduates grouped by type of college are likely to occupy particular strata in the occupational distribution. Miyahara does not offer a sensitive measure of the extent to which groups do and do not overlap, but this sociological paper discusses stratification in terms of the concept of strata that underlies the measure we develop in this paper.

A stratification index that reflects layering or group hierarchy should have several attributes. It should yield measures for each subgroup, since some groups may and other groups may not form distinct strata. Members of one department may all rank above all members of all other departments, while other departments may often overlap with each other. The index for a particular group, say i, should decline as the number of other group members who overlap with group i rises. And, for a given number who overlap with the group, it should decline as the page levels of outside members are close to the midpoint of the group i distribution.

Where a group is small relative to the members of other groups, more cases of non-group members overlapping with group members are likely to occur. Differences in overlapping due to group size capture one aspect of stratification. However, measures with no controls on group size can change merely because of changes in population shares. Indeed, subgroup population shares might dominate such indices. Since we focus on the group's income patterns relative to those of the rest of the population, we develop a normalized stratification index.

Let y_{ij} be the income of member j of group i. Assume that group i has m_i members and $\sum_i m_i = k$. Let $P_i = m_i/k$, or the proportion of the population in group i. $F_0(y_{ij})$ is the value of the cumulative distribution at observation ij within the overall population, which is also the rank (normalized to be between zero and one) of observation ij divided by the number of observations in the overall population. That is, $F_0(y_{ij}) = \text{Rank}\,(y_{ij})/k$. $F_i(y_{ij})$ is the value of the cumulative distribution at ij within group i, which is also the (normalized) rank of observation j in group i. That is, $F_i(y_{ij}) = \text{Rank}\,(y_{ij})/m_i$, where the ranking is over group i only. Since the index measures stratification for each department, it is convenient to divide the population into two groups. Hence, we define $F_{ni}(y_{ij})$ where ni denotes not in group i and $F_{ni}(y_{ij}) = \text{Rank}\,(y_{ij})/(k - m_i)$, where the ranking is over all population members except group i. $F_{ni}(y_{ij})$ is the rank that observation y_{ij} would obtain were it ranked in the population of members other than group i.

If we return to the example above, $F_{ni}(y_{ij})$ is the rank of scientist ij with y published pages among scientists of all departments excluding department i. Note that $F_0(y_{ij}) = P_i F_i(y_{ij}) + (1 - P_i) F_{ni}(y_{ij})$, that is the ranking in the overall population is a weighted average of the ranking in group i and the rest of the population. $\text{Cov}_i\,(x, y)$ is the covariance between x and y among members of group i only. The index of stratification of group i is:

$$Q_i = \frac{\text{Cov}_i\,[(F_i - F_{ni}), y]}{\text{Cov}_i\,(F_i, y)}. \tag{1}$$

The numerator of Q_i is the covariance over group i between the variate and the difference between the ranking of a member of group i in his own group and the ranking he would have in the rest of the population. The denominator, which should be viewed as a normalizing factor, is the covariance between the variate and own ranking in group i. Note that Q_i is unit free.

The Q index has properties that make it sensitive to stratification or layering of groups across an overall distribution.[6] Specifically:

[6]The proof appears in an appendix available from the authors.

1. $-1 \leq Q_i \leq 1$. The index moves between -1 and 1.
2. $Q_i = 1$ if no members of other groups are in the range of the variate of group i. That is, when $Q_i = 1$, group i alone occupies a certain range in the distribution. In this sense, group i forms a perfect strata, or a horizontal layer, as is the case of department B in the example.
3. Q_i declines as more and more members of other groups are in the range of the variate of group i. Put another way, the lower is Q_i, the less group i forms a strata in the overall population.
4. Given the *number* of members of other groups who fall within the range of group i, Q_i will be lower, the closer are the members of other groups to the mean of group i. This property implies that the index is sensitive not only to the overlapping of groups, but also to the position of non-group members in the distribution of group i.
5. $Q_i = 0$ if the normalized ranks (or percentile in the cumulative distribution) of members of group i are identical to their normalized ranks in the overall population. In this case, group i does not form a strata at all. This case occurs if the rank of each person within his own group is equal to his rank in the overall population.
6. $Q_i < 0$ implies that the divergence within the rankings of members of group i in the overall population is greater than the divergence in their own group. This means that group i is not a homogeneous group in the overall population, but is composed of several different groups.
7. $Q_i = -1$ if group i is composed of two groups, the members of each group are identical, and those two groups are located at extremes of the overall distribution. That is, all members of other groups lie inside the range defined by group i. This is an extreme case, where group i is not a group at all, but rather is composed of two perfect strata.

Based on these properties, a society is stratified in terms of a characteristic, if Q_i for all i are greater than zero.[7] The closer the Q_i indices are to one, the more stratified is the society. An intuitive interpretation of Q_i becomes apparent when we divide both the numerator and denominator of Q_i by the variance of y. The numerator of Q_i is displayed as the difference between two regression coefficients and the denominator is a normalizing regression coefficient. In this case, we can write:

$$Q_i = \frac{b_i - b_{ni}}{b_i} \tag{2}$$

where b_i and b_{ni} are defined by the regressions:

$$F_{ij} = a_i + b_i y_{ij} + e_{ij}$$

$$F_{ni,j} = \hat{a}_i + b_{ni} y_{ij} + \hat{e}_{ij}$$

where both regressions are run over members of group i only.

[7] Note that our index yields group specific measures. In general, we measure a group's stratification in terms of relationship to the rest of the population. However, one could also calculate Q indices for one subgroup in terms of its overlap with any other subgroup. Thus, in a population defined into n groups, one could calculate $n*(n-1)$ indices. To see why the group against group concept makes sense, consider a society made up of white and black distributions that do not overlap and an Indian distribution spread over the entire population. Is society stratified? It will depend on the sizes of the three groups.

The regression coefficients, b_i and b_{ni}, are the predicted impact of changes in the variate on rankings. Suppose a randomly selected member of group i published one additional scientific page. Then, b_i is the expected increase in ij's rank within group i and b_{ni} is the effect on ij's rank in the rest of the population. If group i forms a perfect stratum, then the ranking in the other group is not affected ($b_{ni}=0$), and the index equals one. If group i is distributed identically like the rest of the population, then the effect on the ranking of groups will be identical ($b_i = b_{ni}$) and $Q=0$. Finally, if Q_i is negative, then the predicted change in the ranking among members of the other group is higher than the predicted change in the ranking in his own group. This would imply that there are several members of other groups between two members of group i.

We choose a normalization for the index in which $Q=1$ is the case of maximum stratification (no overlap) and $Q=0$ is the case in which the two groups have identical distributions. The third possibility, which the index designates as negative stratification, arises when an income change causes a larger change in one's ranking outside one's group than inside one's own group. In this situation, the person's group really divides into more than one grouping in terms of rankings. The fact that the index takes a negative value to reflect this phenomenon is a result of our normalization of Q to lie between -1 and 1. We could have developed an alternative normalization between 0 and 1 in which 0.5 was the case of identical distributions.

1.2. *Overlapping, Stratification, and Group Size*

Stratification means a group's isolation from members of other groups. Measuring and interpreting this concept becomes complicated when one takes account of differences in group size. Consider the case of two groups, one having 90 percent and the other with 10 percent of the total population. The potential number of members of one group that overlap with the other group differs enormously by group size. The number of small group members overlapping within the large group is at most 1/9 of the large group's population, while overlaps within the small group can be 9 times the small group's population. To avoid having the stratification index be sensitive to group size, we developed the Q index in a way that controls for each group's share of the total population. Hence, we refer to Q as the *relative* index of stratification.

However, actual perceptions of isolation may depend on the extent of *absolute* overlap. A small group may not feel isolated even if only a low proportion of other groups have incomes within its income range. Absolute overlap would still be high enough for every second person within the small group's distribution to come from another population group. Conversely, members of the large group may identify closely with each other (perceive themselves as a separate group) even if all members of other groups overlap their income range.

The absolute of overlap affects the degree of "tagging" that can link certain characteristics with a particular group. Consider the statements, "blacks are poor' and "poor people are black". The first statement is more valid the lower the average income of blacks and the lower is inequality within the black population. The second statement, or the tagging of a characteristic to a group, is more accurate the fewer members of non-blacks that are poor, that is, the lower is the absolute overlap of other groups with the black poor. Where overlap is high, we cannot identify a specific income range (say, poor) with membership in a particular group (say, black).

The O index (shown below) captures this absolute sense of overlap with an index of overlapping that is related to the Q index. O eliminates any controls on the group's share of the population. O is larger, the smaller is group size, thus reflecting the higher potential absolute overlap that can take place within the income ranges of smaller groups.

$$O_i = \frac{1-(1-P_i)Q_i}{P_i}, \tag{3}$$

where $(1-P_i)$ is the number of potential overlaps and P_i is group i's proportion of the total population. To see O_i in terms of the link between incomes of group i and group and population-wide rankings, we make use of equation (1) and write:

$$O_i = 1 + \frac{(1-P_i)\,\mathrm{Cov}_i\,(y, F_{ni})}{P_i\,\mathrm{Cov}_i\,(y, F_i)} = 1 + \frac{\mathrm{Cov}_i\,(y, R_{ni})}{\mathrm{Cov}_i\,(y, r_i)} = \frac{\mathrm{Cov}_i\,(y, R)}{\mathrm{Cov}_i\,(y, r)}, \tag{4}$$

where R_{ni} is the ranking in the rest of the population and r_i is the ranking within group i, while R is the ranking in the overall population.

The properties of O_i can be derived from the properties of Q_i. Its values for specific cases are:

- $O_i = 1$ if $Q_i = 1$
- $O_i = 1/P_i$ if $Q_i = 0$
- $O_i = (2-P_i)/P_i$ if $Q_i = -1$.

1.3. *Stratification In an Inequality Context*

Inequality and stratification are related concepts. In general, high subgroup inequality is likely to increase a subgroup's overlapping with other groups and thus reduce its stratification. However, as Figure 1 shows, increases in group inequality can have no effect or even increase stratification. Though related, inequality and stratification are different in that inequality measures how similar are members of a group to other group members, while stratification measures how different a group's members are from members of other groups. The general connection arises because the more similar subgroup members are to each other, the more we expect them to differ from others.

Group inequality and group stratification both capture interesting patterns, but a unified framework is necessary to connect the two properties and reveal the relationship between grouping patterns and overall inequality. This section develops such a framework, which shows how stratification fits within the decomposition of inequality by population subgroup.

The first step is to express the Gini coefficient in terms of the covariance between a variable and the rank of the variable. As we demonstrated in our 1984 paper, the Gini coefficient (G) of y is equal to:

$$G = \frac{2\,\mathrm{Cov}\,[y, F(y)]}{\bar{y}}, \tag{5}$$

where $F(y)$ is the cumulative distribution of y and $\bar{y}$ is the mean of y. The empirical estimate of $F(y)$ is R/k, where R is the rank of y and k is the number of observations.

Now, define n population subgroups ($i=1,\ldots,n$) and income vectors. Each group includes m_i individuals and the total population is k individuals where $k=\sum_{i=1}^{n} m_i$. The proportion of the population in group i is $P_i=m_i/k$.

The additional notation is:

$y_{i.}$ is the mean income in group i, $y_{i.}=\sum_{j=1}^{m_i} y_{ij}/m_i$;

$y_{..}$ is the mean income of the entire population, or $y_{..}=1/k\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}$;

S_i is the income share of group i;

R_{ij} is the rank of y_{ij} in the overall population, $R_{ij}=1,\ldots,k$;

$R_{i.}$ is the average rank of group i in the overall population, or $R_{i.}=1/m_i\sum_{j=1}^{m_i} R_{ij}$; note that $R_{..}=(k+1)/2$; and r_{ij} is the rank of observation i in its own group $r_{ij}=1,\ldots,m_i$.

In decomposing equation (5) into groups, first decompose the numerator, Cov $(y, F(y))$. Using the estimator of $F(y_{ij})=R_{ij}/k$ and eliminating components that are equal to zero, we can write:

$$\text{(6)}\qquad \operatorname{Cov}(y, F(y))=\frac{1}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}\left(R_{ij}-\frac{k+1}{2}\right)$$

and, by adding and subtracting $R_{i.}$, equation (7) can be written as:

$$\text{(7)}\qquad \operatorname{Cov}(y, F(y))=\frac{1}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}(R_{ij}-R_{i.})+\frac{1}{k}\sum_{i=1}^{n} P_i y_{i.}\left(R_{i.}-\frac{k+1}{2}\right)$$

the second term on right side of equation (7) is actually a covariance of a weighted population; that is,

$$\text{(8)}\qquad \frac{1}{k}\sum_{i=1}^{n} P_i y_{i.}\left(R_{i.}-\frac{k+1}{2}\right)=\frac{1}{k}\operatorname{Cov}(y_{i.}, R_{i.}).$$

The first is also a sum of covariances, since

$$\text{(9)}\qquad \frac{1}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}(R_{ij}-R_{i.})=\frac{1}{k}\sum_{i=1}^{n} P_i \operatorname{Cov}_i(y, R).$$

Hence, using equations (8) and (9), equation (7) can be written as

$$\text{(10)}\qquad \operatorname{Cov}[y, F(y)]=\frac{1}{k}\sum_{i=1}^{n} P_i \operatorname{Cov}_i(y, R)+\frac{1}{k}\operatorname{Cov}(y_{i.}, R_{i.}).$$

Substituting $F*k$ for the R's in equation (10), we obtain,

$$\text{(11)}\qquad \operatorname{Cov}[y, F(y)]=\sum_i P_i \operatorname{Cov}_i[y, F(y)]+\operatorname{Cov}(y_{i.}, F_{i.}).$$

Since

$$\text{(12)}\qquad F(y)=P_iF_i(y)+(1-P_i)F_{ni}(y),$$

we can add $(1-P_i)F_i(y)$ to the first term and subtract it to reach,

$$\text{(13)}\qquad F(y)=F_i(y)-(1-P_i)[F_i(y)-F_{ni}(y)],$$

Substituting (13) and using the properties of the covariance, we can rewrite (11) as:

$$(14)\quad \mathrm{Cov}\,(y, F(y)) = \sum_i P_i\, \mathrm{Cov}_i\,[y, F_i(y)] + \sum_i (P_i - 1)(P_i\, \mathrm{Cov}_i\,[y, F_i(y) - F_{ni}(y)] + \mathrm{Cov}\,(y_{i.}, F_{i.}).$$

Using the definition of Q_i in (1), the definition of the Gini in (5), and the income share, $S_i = P_i * (y_{i.}/y_{..})$, we obtain:

$$(15)\qquad G = \sum S_i G_i + \sum S_i G_i Q_i (P_i - 1) + \frac{2\,\mathrm{Cov}\,(y_{i.}, F_{i.})}{y_{..}},$$

where $F_{i.} = R_{i.}/k$.

The first component represents within-group inequality; the second component reflects the impact of stratification, or of intra-group variability in overall ranks, and the third term measures between group inequality.

The third term is between-group inequality, but this interpretation requires an explanation. It is twice the weighted covariance between each group's average income and the average rank divided by the overall mean income. Note in (5) that the Gini across individuals is twice the covariance of each observation's income with its rank divided by the mean income. Similarly, if we use groups as observations, then we may interpret twice the covariance between group average incomes and group average overall ranks (divided by the overall mean income) as the Gini across groups. Hence, we refer to this term as G_b, or the between-group Gini. G_b is different from the between-group Gini coefficient presented by Bhattacharia and Mahalanonis (1967), Pyatt (1976), Mookherjee and Shorrocks (1982), Das and Parikh (1982), Pyatt (1976), Mookherjee and Shorrocks (1982), Das and Parikh (1982), and Silber (1989). In their case, each group is represented by its mean income while its rank is the rankings of the group's mean incomes. That is, the group's average rank depends only on its mean income relative to the means of other groups. We take account of each observation's ranking in the overall distribution by averaging these rankings within each subgroup.[8] If there is no overlap between groups, all the methods yield the same results. However, with income overlapping between groups, it is easy to show that:

$$(16)\qquad G_b^p > G_b,$$

where G_b^p is the between-group Gini defined by Pyatt.

An example can clarify the issue. Assume that groups have different mean incomes, but the same average rank. Then, our between-group term will be zero, while G_b^p will be positive. From our perspective, the groups are equal in terms of their average positions within the income distribution.[9] In general, changes in

[8]The conventional view, as expressed by Shorrocks (1984) and by Cowell (1985), is that the Gini coefficient is decomposable into between-group and within-group components only in the case when incomes of population subgroups do not overlap. This position is correct, but it does not imply that Gini decompositions are uninteresting. Indeed, we believe that the fact that the Gini decomposes into three terms—within-group, between-group, and a third term—may enhance the results, since it captures the degree of subgroup stratification.

[9]Unlike G_b^p, the G_b term can be negative. Consider a distribution of two groups, one of which has all poor people except one extremely rich person. This group could have an average rank lower than the other group, but an average income that is higher.

individual incomes will affect all three components of (15). However, certain types of distributional changes will exert their influence on only one of the components. The first is the weighted sum of intra-group Gini coefficients. Changes in a group's Gini coefficient, with group income shares held constant, will change the first component in the same direction.

The second component is the sum of each group's stratification, weighted by the product of the group's Gini, income share, and one minus its population share. This component depends partly on the terms in the first component and partly on the distribution of ranks. The added information revealed through the second component is stratification, or the relationship between overall ranks (within each group) and incomes. The middle term is zero when all the subgroups have identical distributions. This is the result of our normalization of the Q index.

The third component depends only on the covariance between group mean incomes and group mean ranks within the overall distribution. The Gini is sensitive to both the distribution of incomes and the distribution of ranks. This is why the Gini does not decompose neatly into two between-group and within-group impacts. Thus, while the Gini produces a more complex decomposition, it provides more information than such neatly decomposable measures as the Theil index.

Still, income changes may affect only one component of the Gini. Changes in Ginis within certain groups may change component one, but have no impact on component two or three. Some changes in Q_i's may leave Ginis unchanged and influence only component two. Finally, some changes in mean ranks may leave the Ginis and Q terms unchanged and thus affect only component three. These examples may involve complex changes, but they demonstrate that the three components can play distinctive, separable roles in accounting for aspects of the income distributions.

It is interesting that increases in stratification exert a negative effect on inequality. To see the statistical explanation, consider the decomposition in equation (15).[10]

$$G = \sum_{i=1}^{n} P_i S_i O_i G_i + G_b \qquad (17)$$

Note that the subgroup inequality indices and the overlap indices have symmetrical impacts on overall inequality. At fixed levels of one set of indices (say, the subgroup Gini coefficients), increases in the other group indices (say, the overlap terms) raise overall inequality. One may think of inequality as rising with the variability of incomes (subgroup Ginis) and with the variability of ranks (subgroup overlap). But, high stratification implies a low variability of ranks. Hence, inequality and stratification are inversely related.

This relationship might seem counterintuitive but it is consistent with relative deprivation theory.[11] According to this theory, stratified societies can tolerate higher inequality than unstratified societies. As people become more engaged with each other, they have less tolerance for a given level of inequality. Consider groupings by leagues. Assume that each person confines his aspirations to his

[10]To obtain this decomposition, substitute O_i for Q_i in equation (15) and rearrange terms.

[11]See Runciman (1986) for a description of the theory and Yitzhaki (1982) for the use of the Gini coefficient to represent the theory.

assigned league. Then, as Yitzhaki (1982) shows, if the leagues are organized according to ability (thus producing high stratification by our measure), then feelings of deprivation will be lower than if all individuals are in one league.

2. Applying the Decomposition to Ethnic Group Differences in the U.S. and Israel

Our decomposition yields answers to three questions:

- how much of overall income inequality comes from within-group inequality, between-group inequality, and the combination of stratification and within-group inequality? and
- to what extent do race and ethnic groups occupy specific segments of the income distribution? What is the extent of ethnic stratification with respect to income?
- which way of grouping the population yields meaningful subgroups with respect to income rankings?

The ethnic breakdowns differ by country. In the Israeli context, the groups of families are those headed by Israelis from Europe or America, from Asia or Africa, and from Israel. The U.S. breakdown is by race and origin of the family head (black, Hispanic, and white non-Hispanic) and type of family (husband-wife, other male head, and other family head).

The data come from the March 1987 U.S. Current Population Survey (CPS) and from the 1979–80 Israel Survey of Family Expenditure. The U.S. income measure is family income including the market value of several in-kind transfers and net of taxes, while the Israeli measure is gross family income per standard adult.[12]

Tables 1 through 4 show that most inequality in both countries is due to the weighted sum of within-group income differences. Although mean incomes vary substantially by race and ethnic origin, the Gini coefficients within subgroups are nearly as high as the overall Ginis. Inequality between racial groups accounts for less than 10 percent of overall inequality. This result is especially surprising for the U.S., where income differentials by race are large. The role of between-group inequality increases when one divides the population into race-family type groupings. The stratification component—which embodies between-group and within-group elements—accounts for a very small amount of the Gini coefficients.

The stratification terms reveal which ethnic-racial-family type groups are meaningful in terms of income rankings. Recall that each stratification term (Q_i) provides an index of the extent of overlap between a particular group and the rest of the population. According to Table 1, which groups families according to the head's continent of birth, Israelis born in Europe or America do not constitute a separate group in the sense of occupying a segment of the income

[12]We used the entire sample from the Israeli survey and three CPS rotation groups from the U.S. survey. The U.S. data included all one-family (including one person) households; they make up the vast majority of households. The U.S. income definition was after-tax, after-transfer income. For a procedure that calculates Gini coefficients from weighted micro-data, see Lerman and Yitzhaki (1989).

TABLE 1

DECOMPOSITION OF ISRAELI INCOME INEQUALITY BY COUNTRY OF OWN ORIGIN: 1979-80

	Europe/ America	Asia/ Africa	Israel	Total
Mean Income per standard adult (thousands of 1980 Israeli shekels)	11.326	8.126	11.758	10.401
Population share (P)	0.450	0.320	0.230	1.000
Income share (S)	0.490	0.250	0.260	1.000
Gini coefficient (G)	0.321	0.292	0.276	0.317
Stratification index (Q)	0.000	0.090	0.160	
Within-group component	0.157	0.073	0.072	0.302
Between-group component				0.028
Stratification component	0.000	−0.004	−0.009	−0.013

Source: Tabulations by authors from 1979-80 Survey of Family Expenditure.

distribution. Both those from Asia-Africa and the Israeli-born are somewhat stratified from the other groups.

The picture changes in Table 2, where ethnic status depends on the origin of the head's father. Here, the stratification indices are highest for European-American and Asian-African groupings. Those with Israeli-born fathers are less stratified, as members of other groups overlap much more with their income distribution. These patterns differ from those in Table 1, partly because grouping by continent of birth interacts with age. In Table 1, the Israeli-born are a distinct subgroup largely because they are a younger than average group. On the other hand, those born in Europe or America do not occupy a distinct segment because they overlap substantially with the many Israeli-born who have fathers of European-American origin.[13]

In the U.S., the impact of race differences on income inequality and stratification interacts with family type. When one divides the population only into white, black, and Hispanic families, between-group inequality accounts for little of total inequality, despite large mean income differentials by subgroup. The stratification indices reveal that whites are more segmented from other groups than are blacks and Hispanics. Controlling for population size, fewer minorities overlap with the white income distribution and/or the overlaps are further from the midpoint of the white distribution than is the case for blacks relative to whites and Hispanics or Hispanics relative to blacks and whites. Since whites have much higher incomes than Hispanics or blacks, minority incomes generally do not overlap with white incomes. The median black family had an income that would have placed it in the 32nd percentile of the white distribution.

The low black stratification index shows that large proportions of white and Hispanic families have incomes that overlap with black incomes. In part, this is because blacks have incomes that are distributed more unequally than white incomes. The wider spread of black incomes opens more space within which the incomes of other groups can be situated.

[13]For an extended discussion of the Israeli results concerning income inequality as well as educational inequality, see Yitzhaki (1987).

TABLE 2

DECOMPOSITION OF ISRAELI INCOME INEQUALITY BY COUNTRY OF PARENT'S ORIGIN: 1979-80

	Europe/ America	Asia/ Africa	Israel	Total
Mean income per standard adult (thousands of 1980 Israeli shekels)	11.861	8.267	10.401	10.401
Population share (P)	0.570	0.390	0.040	1.000
Income share (S)	0.650	0.310	0.040	1.000
Stratification index (Q)	0.100	0.100	0.040	
Gini coefficient (G)	0.317	0.288	0.266	0.318
Within-group component	0.206	0.089	0.011	0.306
Between-group component				0.028
Stratification component	−0.009	−0.005	0.000	−0.015

Source: Same as Table 1.

TABLE 3

DECOMPOSITION OF U.S. INCOME INEQUALITY BY RACE AND SPANISH ORIGIN AND BY TYPE OF FAMILY: 1986

	Income After Taxes and All Transfers						
Race	Mean Income	Population Share	Income Share	Group Gini	Group Stratification	Group Overlap	Within-Group Term
Black	$20,705	0.125	0.092	0.362	−0.001	8.00	0.033
Hispanic	21,369	0.085	0.064	0.340	0.021	11.60	0.022
White	30,069	0.790	0.844	0.316	0.156	1.22	0.267
Total	$28,160	1.000	1.000	0.331			
Within-group Inequality				0.322			
Between-group Inequality				0.018			
Stratification Term				−0.009			

Source: Tabulations by authors from March, 1987 *Current Population Survey.*

Note: The income concept is family income including the earned income tax credit and in-kind transfers (the market value of food stamps, housing benefits, school lunch benefits, and noninstitutional medical benefits) less taxes (social security taxes, property taxes, and state and federal income taxes). The unit of analysis is the individual family member. The sample includes all persons in families of two or more persons.

Serious stratification of black and Hispanic families emerges when the grouping is by race and family status. Note in Table 4 that black and Hispanic families headed by women have high indices of stratification, as do white husband-wife families. By implication, these groups encounter relatively few members of other groups overlapping their distributions. Adding family status groupings within race and Spanish origin raises the proportion of inequality associated with between-group inequality from 5 percent (when groups are based on race and Spanish origin) to 15 percent (when the subgroup breakdown includes family status and race). This might be expected purely on the basis of adding new groupings. However, calculations (not shown in these tables) using family type

TABLE 4

DECOMPOSITION OF U.S. INCOME INEQUALITY BY RACE AND SPANISH ORIGIN AND BY TYPE OF FAMILY: 1986

	Income After Taxes and All Transfers						
Race by Family Type	Mean Income	Population Share	Income Share	Group Gini	Group Stratification	Group Overlap	Within-Group Term
Black							
Husband–wife	$26,943	0.063	0.061	0.291	0.075	14.7	0.018
Other male head	21,595	0.005	0.004	0.464	−0.137	100.5	0.002
Other female head	13,713	0.057	0.028	0.352	0.265	13.1	0.010
Hispanic							
Husband–wife	23,861	0.061	0.051	0.312	0.037	15.9	0.016
Other male head	18,832	0.005	0.003	0.337	0.096	105.0	0.001
Other female head	14,105	0.019	0.010	0.370	0.200	42.0	0.004
White							
Husband–wife	31,957	0.670	0.760	0.297	0.251	1.47	0.226
Other male head	25,690	0.024	0.022	0.329	−0.017	42.80	0.007
Other female head	18,020	0.096	0.062	0.363	0.080	9.60	0.022
Total	$28,159	1.000	1.000	0.331			
Within-group Inequality				0.305			
Between-group Inequality				0.050			
Stratification				−0.025			

Source: Same as Table 3.
Note: See Table 3.

groupings and no racial breakdowns confirm the dominant role of family status. Stratification and between-group inequality are nearly as high when the population is divided into family status groups as when the group definition includes both race and family status. Only the bottom tail of the income distribution of husband–wife families overlaps with the income distribution of female-headed families.

The tables all illustrate the differences between absolute and relative measures of overlap and stratification. Group size is the major determinant of the O terms, which show the extent of overlap on an absolute basis. Whites overlap least with other groups largely because whites constitute 80 percent of the population. The smallest groups—such as other male heads of Hispanic and black families—overlap most.

3. SUMMARY AND IMPLICATIONS FOR RESEARCH

This paper presents an index of stratification, integrates the idea of stratification within overall income inequality, and clarifies the distinction between subgroup inequality and subgroup stratification. Unlike the Theil index, the Gini coefficient does not decompose neatly into within-group and between-group contributions to inequality. However, we show that decomposing the Gini into three components yields revealing interpretations of the nature of inequality among population subgroups.

The index of stratification controls for group size and reveals the extent to which groups form strata, and thus are *relatively* stratified by income. Results on income inequality in Israel and in the U.S. show that these measures can identify groupings that are meaningful in the sense of being relatively similar within groups (say, in income terms) and different from those outside the group.

We see several directions for applying and extending our approach. One is to conduct analyses of other aspects of inequality (say, education or occupational prestige) and other types of groupings (say, occupational stratification with respect to income). A second application is to examine trends in stratification based on various specifications of population subgroups. Future research to derive the standard deviation of Q will be useful in judging the significance of such trends.

Another direction is to widen our approach to capture multidimensional aspects of the social hierarchy. This paper's method of integrating stratification and inequality deals with only one dimension. While sociologists write about stratification as a multidimensional concept (as, for example, a vector of power, prestige, and income), we are not aware of any measures of stratification that are multidimensional. In future work, we shall attempt to derive measures that capture population groupings in a society in multidimensional terms.

References

Allison, P., Measures of Inequality, *American Sociological Review*, 43, 865-880, 1978.

Bailey, K., Entropy Measures of Inequality, *Sociological Inquiry*, 55, 2, 200-211, 1985.

Bhattacharia, N. and Mahalanobis, B., Regional Disparities in Household Consumption in India, *Journal of the American Statistical Association*, 62, 143-161, 1967.

Bolen, K. and Jackman, R., Political Democracy and the Size Distribution of Income, *American Sociological Review*, 438-457, 1985.

Cowell, F. A., Multilevel Decomposition of Theil's Index of Inequality, *Review of Income and Wealth*, 31, 201-205, 1985.

Lasswell, T., *Class and Stratum*, Houghton Mifflin Company, Boston, Massachusetts, 1965.

Lenski, G., Income Stratification in the United States: Toward a Revised Model of the System, in Treiman, D. and Robinson, R. (eds.), *Research in Social Stratification and Mobility*, JAI Press, Greenwich, CT, 1984.

Lerman, R. and Yitzhaki, S., A Note on the Calculation and Interpretation of the Gini Index, *Economic Letters*, 15, 363-368, 1984.

———, Improving the Accuracy of Gini Estimates, *Journal of Econometrics*, 42, 43-47, 1989.

Mann, M., editor, *The International Encyclopedia of Sociology*, Macmillan Press, New York, 1984.

Massagli, M. P., Effects of Family Size on the Process of Stratification: A Structural Equation Model for White Couples in the U.S. in 1962 and 1973, in Robinson, R. (ed.), *Research in Social Stratification and Mobility*, 6, 81-112, JAI Press, 1987.

Miyahara, K., Inter-College Stratification: The Case of Male College Graduates in Japan, *Sociological Forum*, 3, 25-43, 1988.

Mookherjee, D. and Shorrocks, A. F., A Decomposition Analysis of the Trend in UK Income Inequality, *The Economic Journal*, 92, 886-902, 1982.

Pyatt, G., On the Interpretation and Disaggregation of Gini Coefficients, *The Economic Journal*, 86, 243-254, 1976.

Runciman, W. G., *Relative Deprivation and Social Justice*, Routledge and Kegan Paul, London, 1986.

Silber, J., Factor Components, Population Subgroups, and the Computation of the Gini Index of Inequality, *Review of Economics and Statistics*, 71, 107-115, 1989.

Shorrocks, A. F., Inequality Decomposition by Population Subgroups, *Econometrica*, 53, 1369-1386, 1984.

Turner, J., *Societal Stratification: A Theoretical Analysis,* Columbia University Press, 1984, New York.
Yitzhaki, S., On Stratification and Inequality By Continent of Origin In Israel, *Bank of Israel Review,* 63, December, 1988, 36-51.
———, Relative Deprivation and Economic Welfare, *European Economic Review,* 17, 99-113, 1982.