Recent Changes in Earnings Distributions in the United States: Age and Cohort Effects

*by Kelvin R. Utendorf**

In this article, the author uses large, Social Security administrative data sets to examine changes in earnings distributions in the United States over the 1980s through the mid-1990s. Because the earnings information contained in these data sets comes directly from the W-2 forms filed by employers, self-reporting errors and top-coding problems, common in other data used for this type of analysis, are minimized. Previous research has documented an increase in overall earnings inequality during the 1970s and the 1980s. The author finds that this upward trend in overall earnings inequality continues into the mid-1990s, despite a period of nearly constant or slightly decreasing earnings inequality from 1988 through 1992. The data also suggest that between-group earnings inequality, whether dividing the sample into groups by age group or by birth cohort, is increasing. Despite the increase in between-group earnings inequality over the period examined, however, within-group earnings inequality remains by far the largest contributor to overall earnings inequality.

*Division of Economic Research, Office of Research, Evaluation and Statistics, Office of Policy, Social Security Administration.

Acknowledgments: Special thanks go to Ben Bridges and Mike Leonesio for providing helpful comments on earlier drafts of this article. I am also grateful to Julie-Anne Cronin, a discussant at the January 1999 Allied Social Sciences Association session in honor of Daniel Radner, for her insights.

I. Introduction

Earnings have traditionally served as one measure of a person's well-being. Other things equal, an increase in an individual's earnings is generally thought to signify an improvement in that individual's lot. Likewise, at a group or national level, increases in average earnings are often viewed as an indication that the group or nation is, in some sense, better off than before the increase. Increasing earnings disparity among groups is commonly viewed as being bad, however. An increase in the earnings of one group relative to those of another group could mean that society as a whole is worse off, depending on one's point of view. Examining changes in earnings distributions provides us with insights into the welfare of individuals and groups in society. In addition, the ability to forecast changes in earnings distributions plays a central role in accurately projecting the future status of Social Security's Old-Age, Survivors, and Disability Insurance (OASDI) Trust Funds. One part of developing an earnings distribution forecast involves understanding how past earnings distributions have changed over time, with the hope that this understanding will provide insights as to what to expect in the future. This article, an extension of Utendorf (1998), will describe changes in earnings distributions in the United States for the overall population, as well as for age and birth cohort subgroups of the population, for the period 1981-95, using an unusually large sample derived from Social Security Administration (SSA) earnings records.

A rather extensive economics literature has developed that examines changes in earnings distributions, particularly earnings distributions in the United States, over the 1970s and the 1980s, and, to a lesser extent, the 1990s. Many of these studies found that earnings inequality increased markedly from the mid-1970s through the 1980s and beyond. Levy and Murname (1992), in their survey article that deals with earnings trends and

earnings inequality, state, "Nineteen-hundred-seventy-nine marked the beginning of a sharp acceleration in the growth of earnings inequality, particularly among men."[1] Freeman (1995), in a paper that examines the impact of increased international trade on the wages of less-skilled workers, remarked, "Researchers using several data sources—including household survey data from the Current Population Survey, other household surveys, and establishment surveys—have documented that wage inequality and skill differentials in earnings and employment increased sharply in the United States from the mid-1970s through the 1980s and into the 1990s."[2] Many others, such as Blackburn, Bloom, and Freeman (1990/1991), Karoly (1992), Bernstein and Mishel (1997), Gottschalk (1997), Johnson (1997), and Topel (1997), have documented increasing earnings inequality over various parts of the period spanning the mid-1970s to the mid-1990s.

Others, notably Robert Lerman, question whether the overall wage distribution has changed to any significant degree from the mid-1980s. Lerman (1997) argues that studies limited to examining specific segments of the population, such as men only or full-time workers only, are not appropriate for assessing overall changes in earnings distributions caused by market and institutional forces. He used a sample of all wage earners constructed from the 1984, 1987, 1988, 1991, 1992, and 1993 panels of the Survey of Income and Program Participation (SIPP) to find the following:

> Earnings inequality did increase for some groups of workers, and some forces, such as trade and technology, may have affected overall inequality. However, according to several indicators, the combined effects of changes in demand, supply, and institutional forces over the last decade did not generate higher wage inequality in the U.S. labor market as a whole.[3]

There are also studies that have gone beyond looking at changes in the overall earnings distribution by focusing on between- and within-group differences for various age and/or cohort subgroups of the population. Dooley and Gottschalk (1984), using CPS data from the 1968 through 1979 surveys, argue that, even after controlling for education, experience, and the unemployment rate, earnings inequality among male labor force cohorts increased over the period of the study, meaning that changes in the relative earnings of young to older workers or growth in the proportion of younger workers did not account for the increasing inequality among males.[4] Katz and Murphy (1992) conclude that earnings inequality within groups defined by education, age/experience level, and gender was 30 percent higher in 1987 than in 1970. Burtless (1990) argues that the increase in male earnings inequality was not brought about by changes in demographic variables, but rather that the pay of males with similar education and experience levels had become more unequal over time. In their survey article, Levy and Murnane (1992) state that between-group earnings inequality among age groups increased over both the 1970s and the1980s.[5] Katz and Murphy (1992) contend that these changes in between-group earnings inequality observed over the 1970s and 1980s can be explained by labor supply shifts (changes in the rates of growth of different labor force groups) together with relatively stable growth in the relative demand for college educated workers. Gottschalk (1997) shows that between-group inequality among age/experience groups increased significantly for men over the 1970s and 1980s, but that it has not changed a great deal since the late 1980s. He also provides evidence of increasing between-group inequality among different age/experience groups of women over the 1980s and the early 1990s.

The impact of the baby boom on the labor market has generated a great deal of research regarding birth cohort effects on earnings. Welch (1979) examines the relationship between birth cohort size and wages. He finds that large sized birth cohorts tend to initially depress not only the wages of new entrants into the labor market, but the hours and weeks worked as well. He estimates earnings decreases as high as 13 percent for college graduates due solely to the large cohort size of the baby boom. While Welch found that the earnings-depressant effects of a large cohort size gradually decline over time, Berger (1985) contends that the negative cohort size effect on earnings worsens with experience because of the fact that workers proceed through their careers more slowly.[6] Murphy and Welch (1992) argue that the entry of the baby boomers into college, and later into the labor force, had large effects on earnings distributions. They maintain that the relative wage of college graduates from the baby boom generation fell because of enormous expansion of their relative numbers and age/experience differentials grew because of the influx of inexperienced college and high school graduates into the labor force. Johnson (1997) also concludes that the large increase in the relative supply of college educated workers in the 1970s depressed relative wages.[7]

This article describes changes in earnings distributions over the period 1981-95 for the overall population, as well as by age and cohort subgroups, using data extracted from SSA's earnings records. Since these data have not been extensively used for this type of research in recent years, section II of the article describes them thoroughly. These data offer several advantages over the typical public-use survey dataset, as indicated in the next part of the article. Section III contains a description of the methods used to examine changes in the patterns of earnings distributions over the period. Gini indices have long been used to examine earnings inequality issues. Two relatively recent innovations, used in this article, are introduced in this third section. The results presented in section IV show, as do most of the papers in the economics literature in this area, that earnings inequality increased over the early to mid-1980s. Beyond the mid-1980s, however, the pattern is less clear. From 1988 onward, earnings inequality appears to fall, then rise, then fall again, with the end result being that earnings inequality in 1995 is higher by a statistically significant amount than earnings inequality in 1988, but with the magnitude of this increase being very small. Section V concludes the article.

II. Data Description

I used a subset of files from the Social Security Administration's Continuous Work History Sample (CWHS) family of files for this project. In particular, I used the 1-percent sample 1957-90 Longitudinal Employee-Employer Data (LEED) file and the 1991, 1992, 1993, 1994, and 1995 files from the 1-percent sample annual Employee-Employer (EE-ER) file series in order to examine earnings distributions over the period 1981-95. When used in combination with the CWHS 1-percent sample Active file,[8] I have information on the year of birth, annual Social Security taxable wages, and total wage compensation for a 1-percent sample of Social Security numbers (SSNs) for which wage and salary employment was reported over the period 1981-95.[9, 10]

I chose to explore the interval 1981-95 for two reasons. First, increasing earnings inequality over the 1970s and the early to mid-1980s has been well documented in numerous studies,[11] while there is less work covering the late-1980s and early-1990s. By examining distributions over 1981-95, I was able to corroborate the results of other studies for the period of the early to mid-1980s, while using a consistent method to study possible changes in earnings distributions for more recent years.

A second reason for choosing the period 1981-95 arises because of limitations in the Social Security Administration's administrative data themselves. In 1978, a change from quarterly wage reporting to annual wage reporting took place. As part of this change, total wage and salary compensation (taken directly from the W-2 information reported by employers to SSA) became available within the CWHS file system.[12] However, as with any major modification, there were difficulties in the years immediately following the change to annual wage reporting with late posting, duplicate reports, and other processing problems. In order to avoid possible data quality problems with this transition period, I elected to use 1981 as the earliest year in the study.[13]

Data from the CWHS files provide several advantages over the data typically used for this type of research. First and foremost, because they come directly from the W-2 form, the CWHS earnings data do not exhibit any of the "self-reporting" problems that have the potential for being present in most, if not all, public-use surveys. In particular, it is believed that individuals toward the upper end of the earnings distribution have a higher tendency to underreport their earnings in surveys, which is troublesome when the point is to examine earnings distributions. Data from the CWHS do not have this problem since individuals generally do not have a choice regarding what is reported on their W-2 forms.[14]

A second advantage of using data from the CWHS is that the earnings data are not top-coded. Public-use datasets top-code earnings in an effort to help mask the identity of individuals with high earnings who otherwise might be identifiable with a combination of their actual earnings and other characteristics in the file. The Annual Demographic Survey (March CPS Supplement) of the Current Population Survey (CPS), for example, presently top-codes wage and salary earnings so that an individual does not show earnings in any one job of more than $100,000 per year.[15] The percentage of individuals in a survey affected by this can vary from year to year, which means top-coding alone, other things equal, could cause measures of earnings inequality to vary from year to year. Given that one of my objectives is to make accurate observations regarding the fluctuations in earnings distributions from year to year, using data that are not top-coded is important.

A third advantage of the CWHS is the large number of observations available. For the LEED and EE-ER files, a 1- percent sample, based on specified digits from the last four digits of the SSN, of those with wage and salary earnings in the specified year(s) is drawn from Social Security's records. Sample sizes for the years studied range from about 970,000 to about 1.36 million observations, depending on the year in question.[16] Such large sample sizes eliminate concerns about having too few individuals in any particular group under study.

There are certain disadvantages to using these particular data when analyzing changes in earnings distributions. Probably the biggest drawback is the lack of certain types of socioeconomic information for the individuals in the dataset, particularly the lack of information regarding educational attainment. Many studies have pointed to differences in education and increased returns to education as possible reasons behind changes in earnings distributions over the 1970s and the 1980s. Without any way to identify schooling differences among individuals in the sample, the ability to explain changes in earnings distributions using these data is limited.[17]

A second disadvantage to using these data is the work required to make them suitable for research purposes. The primary reason that the Social Security Administration collects this information is to assist in effectively administering the program so that the monthly benefit payments to recipients are delivered on a timely basis and in the correct amounts. Researchers within SSA, in effect, have access to these data as an afterthought and thus necessarily spend a great deal of time making them useful for research purposes.

In addition to the limited demographic information, each observation in the sample contains two earnings variables, Old-Age and Survivors Insurance and Disability Insurance (OASDI) taxable earnings and total wage and salary earnings.[18] OASDI taxable earnings are earnings, up to the annual maximum taxable earnings amount,[19] by individuals covered by the Social Security program. The total wage and salary earnings information comes directly from an individual's W-2 form, as indicated earlier, *regardless of whether that individual is covered by the Social Security program.* There are observations for which the amount in the total wage and salary earnings field in the dataset is less than the amount in the OASDI taxable earnings field. This could occur if the individual in question contributed to a tax-deferred saving plan, since the earnings amount reported in the total wage and salary field in the dataset does not account for contributions to such plans.[20] It is also possible for the amount in the total wage and

salary earnings field to be less than the amount given in the OASDI taxable earnings field, due to the way that Social Security processes these data. For example, if a correction is made to the OASDI taxable earnings amount, it is generally the case that the corresponding total wage and salary earnings amount is not updated to reflect the change since, from a programmatic standpoint, the total wage and salary earnings amount is not important in the determination of benefits. Therefore, for observations covering the years 1981-93, where the total wage and salary earnings amount was less than the OASDI taxable earnings amount, the former was increased to the level of the latter in order to give a better accounting of the individuals' true total earnings, reflecting both the view that true total earnings should include the deferred earnings, as well as the belief that the OASDI taxable earnings amounts on record are the more accurate of the two.[21, 22]

For the years 1994 and 1995, additional earnings information is available in the EE-ER files. In particular, Medicare (HI) taxable earnings are available for each observation, which is important since, beginning in 1994, the ceiling on maximum HI taxable earnings was eliminated. Therefore, the Medicare taxable earnings variable provides (potentially) an excellent measure of true total wage and salary earnings, since even deferred earnings are taxed for Medicare purposes. A measure of deferred compensation is also included with the information for the years 1994 and 1995.[23]

This additional information provides somewhat of a dilemma, though. One can filter the files so that the largest of either (a) the total wage and salary earnings variable plus the deferred compensation variable, (b) the OASDI taxable earnings variable, or (c) the HI taxable earnings variable is used as the variable of analysis, true total wage and salary earnings. However, making use of the deferred compensation and the HI taxable earnings information would decrease my ability to make comparisons across years, since the yearly series would no longer be consistently calculated. Accordingly, I have performed all of the analyses using both the "new" method (looking for the largest value among HI taxable earnings, OASDI taxable earnings, and total wage and salary earnings plus deferred compensation) and the "old" method (using only the variables available to me for the years 1981-93) for the years 1994 and 1995. As the results show, using the additional information in the 1994 and 1995 files does have an impact.[24]

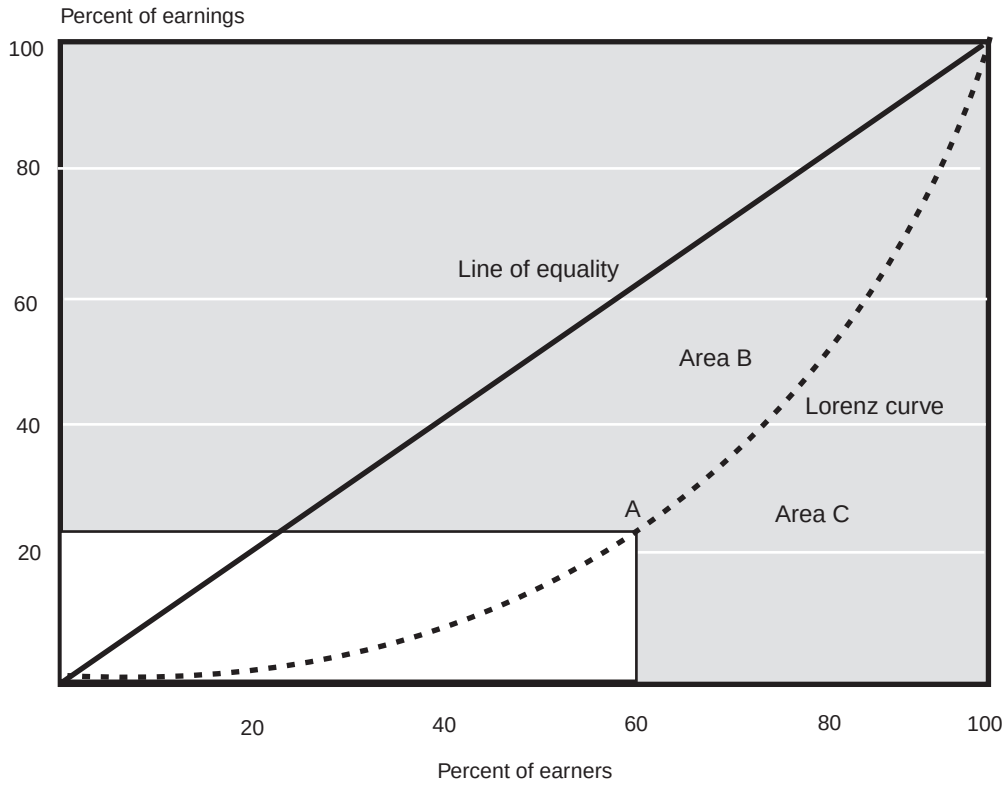
I applied one other significant filter to the data by eliminating all observations for individuals younger than age 13 or older than age 86.[25] These arbitrary age cutoffs were chosen to eliminate from the sample those observations for which either very young individuals or very old individuals had large wage and salary earnings.[26] While it is conceivable that the very young or the very old might have significant levels of income, it is much less likely that individuals in either of those groups would have large wage and salary earnings.

III. Measuring Earnings Inequality

Many different measures of earnings inequality have been developed over the years.[27] Perhaps the most commonly used measure, and the measure I employ in this article, is the Gini coefficient. The traditional Gini coefficient is defined as being half of the absolute mean difference in earnings between each pair of individuals in the sample, relative to mean earnings for the sample.[28] In other words, it is a measure of the spread between the earnings of all pairs of individuals in the sample.

The Gini coefficient can be represented graphically with the use of a Lorenz curve, as in chart 1. The Lorenz curve in the example is a plot of the cumulative percentage of total earnings vs. the cumulative percentage of earners, where the observations are ranked from lowest earnings to high-

Chart 1.—Lorenz curve for the entire sample, based on total wage earnings, 1981



est earnings. Point A on the Lorenz curve in chart 1, for example, shows that the bottom 60 percent of the earners in the sample (bottom with respect to their position in the earnings distribution) earned approximately 25 percent of the total wage and salary earnings in the United States in 1981. The "Line of Equality" shows where the Lorenz curve would be positioned were it the case that everyone in the sample had equal earnings. Therefore, the greater the area between the Lorenz curve and the Line of Equality, the greater the inequality present in the sample. The traditional Gini coefficient is equal to the ratio of the area between the Line of Equality and the Lorenz curve and the area beneath the Line of Equality— in other words, Area B divided by Areas B+C. As Area B gets smaller (meaning the Lorenz curve gets nearer to the Line of Equality and inequality decreases), the Gini coefficient gets smaller.

Two recent innovations regarding the Gini coefficient enhanced its usefulness for this project: work by Barrett and Pendakur (1995) regarding the asymptotic distribution of generalized Gini indices and work by Yitzhaki and Lerman (1991) on Gini decomposition. What follows is a brief overview of each of these new developments as well as an explanation of their importance.

S-Gini Indices

The traditional Gini index, though widely used, has been criticized because it does not allow inequality to be measured under different value judgements regarding the importance of one part of the earnings distribution relative to another. Partly in response to this criticism, Donaldson and Weymark (1980) and Yitzhaki (1983) independently developed what are known as the S-Gini class of inequality indices, and in particular, the S-Gini relative indices of inequality that I use in this article. The S-Gini indices depend on a parameter, $\delta \geq 1$, that can be adjusted to reflect the sensitivity of the index to different parts of the earnings distribution.[29] For $\delta > 2$, for example, the index places more weight on the earnings of those at the lower end of the earnings distribution. When $\delta = 2$, the index places equal weight on all of the observations and corresponds to the traditional Gini coefficient.

The real value of the S-Gini indices for this article, though, is that they are calculated by using Lorenz curve ordinates[30] and therefore use information from every part of the range of earnings. Beach and Davidson (1983), Bishop, Chakraborti, and Thistle (1989), and Bishop, Formby, and Smith (1991) developed statistical inference techniques to study income and earnings inequality by examining Lorenz curve dominance among different distributions. Barrett and Pendakur extend the previous work on S-Gini indices by deriving their large sample properties, using methods similar to those used by Bishop et al. for Lorenz curves, thereby making it possible for S-Ginis to be used for statistical inference.[31] With the traditional Gini coefficient, one is unable to assess whether there is a statistically significant difference between a Gini of .530 and one of .540, for example. By deriving the large sample properties of the S-Gini indices, however, Barrett and Pendakur make it possible to determine whether there is a statistically significant difference between S-Gini estimates. This is important in this article because it allows inferences regarding the likelihood that the distribution of earnings, as measured by the S-Gini coefficient, has changed over time.

Gini Decomposition

Typically, the Theil entropy inequality measure has been used in studies such as this because it decomposes nicely into two terms that can be thought of as measures of between- and within-group inequality. However, recent work by Yitzhaki and Lerman (1991) on decomposing the Gini coefficient has breathed new life into the measure and has allowed me to use a consistent measure of inequality across all parts of this article.

Yitzhaki and Lerman showed that the Gini index of inequality can be decomposed into three terms, one term representing between-group inequality, a second term representing the weighted sum of within-group inequality indices, and a third term representing the weighted sum of group stratification indices.[32, 33] Stratification is a concept borrowed from sociology and refers to the division of a society into hierarchically arranged groupings where the members of a group have similar qualities. Yitzhaki and Lerman develop what they refer to as "indices of stratification" that capture the degree of overlap between group members and nongroup members with respect to some characteristic, namely earnings. I used what they have defined as a relative index of stratification to examine the extent to which the earnings of certain age and cohort groups overlap with the earnings of other age and cohort groups.

The relative stratification index of Yitzhaki and Lerman, Q_i, ranges from -1 to 1.[34] If it is the case that $Q_i=1$, then no members of groups other than *i* have earnings within the range of earnings spanned by the members in group *i*, meaning group *i* forms a perfect stratum. As Q_i decreases from 1, group *i* forms less and less a stratum in the overall population as the earnings of more and more nongroup *i* members fall within the range spanned by group *i* earnings. At $Q_i=0$, group *i* does not form a stratum at all since the relative rank of each person within his or her own group is identical to his or her rank in the overall population. Negative values for Q_i mean that "group" *i* really is not a single group, but is instead composed of several different groups. Finally, were $Q_i=-1$ to be true, "group" *i* would actually consist of two distinct groups, with those groups located at opposite ends of the earnings distribution. In this case the earnings of everyone in the sample *other* than those in group *i* would lie between the ranges of the two segments of group *i* earnings, meaning that group *i* would form two perfect, distinct strata.

IV. Results

In this section I present the results[35] for the entire sample, as well as for subsamples distinguished by age and by birth cohort. The S-Gini coefficients discussed throughout this section, as well as their corresponding asymptotic standard errors, were calculated on the basis of 100 sample quantiles.

The decision to use 100 quantiles for calculating the S-Ginis was somewhat arbitrary, though the S-Ginis presented for the case of $\delta=2$ (the "traditional" Gini coefficient, where all observations are equally weighted) are identical to those calculated via the covariance method of Lerman and Yitzhaki (1984) to at least three decimal places.[36, 37]

Entire Sample

Table 1 shows the real sample means, real medians, and the S-Gini coefficients for total wage and salary earnings for the entire sample.[38] For 1981-95, using the (a) rows for 1994 and 1995, both the mean and median real earnings of those in the sample grew slowly over the period of this study, with mean earnings increasing by about 13 percent and median earnings increasing by an even smaller 4 percent over the 15-year period. The years 1994(b) and 1995(b), using measured earnings at least as great as those for 1994(a) and 1995(a), show slightly faster growth in mean real earnings over the period, as expected. Under either the (a) or (b) calculation method, the table shows that median real earnings peak in 1987 and generally decline thereafter. The stagnant earnings growth might be partly attributed to an increase in part-time/part-year employment over the period of the study, since the sample contains individuals with both types of employment. Also, slow growth in the per hour wage rate over substantial portions of the period under study no doubt contributed to the slow growth in mean real earnings.

Table 1.—Mean and median earnings for the entire sample, based on total wage earnings, 1981-95

Year	Number of observations	Mean earnings (1992 dollars)	Median earnings (1992 dollars)	S-Gini ($\delta=2$)[1]
1981	1,025,211	$19,492	$15,060	0.484
				(0.00031)
1982	980,636	20,130	15,055	.503
				(0.00051)
1983	1,100,081	20,520	15,389	.502
				(0.00054)
1984	1,132,264	20,911	15,484	.507
				(0.00071)
1985	1,181,248	21,337	15,882	.503
				(0.00049)
1986	1,171,792	21,485	15,892	.508
				(0.00053)
1987	1,212,791	22,083	16,051	.518
				(0.00084)
1988	1,252,347	22,069	15,925	.520
				(0.00078)
1989	1,280,141	21,858	15,879	.516
				(0.00065)
1990	1,311,110	21,811	15,872	.513
				(0.00061)
1991	1,301,301	21,681	15,725	.513
				(0.00068)
1992	1,308,211	22,045	15,741	.521
				(0.00075)
1993	1,318,221	21,930	15,580	.523
				(0.00071)
1994(a)`[2]	1,326,205	21,879	15,477	.526
				(0.00085)
1995(a)	1,359,143	21,984	15,654	.522
				(0.00065)
1994(b)	1,326,205	22,238	15,502	.532
				(0.00090)
1995(b)	1,359,143	22,415	15,679	.529
				(0.00072)

[1]The asymptotic standard errors are in parentheses beneath the S-Gini coefficients. One can perform one-tailed or two-tailed standard normal z-tests of the statistical significance of the difference between any two of the S-Gini coefficients given above by calculating the following statistic, $z = \frac{G_1 - G_2}{\sqrt{SE_1^2 + SE_2^2}}$, where the G's are the S-Gini coefficients and the SE's are the standard errors.

[2]The statistics calculated for 1994(b) and 1995(b) are generated using additional earnings information (tax-deferred and Medicare-taxable earnings) available for those two years only. The entries for 1994(a) and 1995(b) do not make use of this additional information and are, therefore, calculated in a way that is consistent with the calculations in years 1981-93.

The last column in table 1 contains S-Gini coefficients (with asymptotic standard errors beneath the coefficients) for each year of the study. Clearly, by this measure, earnings inequality generally increased (by statistically significant amounts) over the period of the study. In an earlier article covering the period 1981-93,[39] I observed that earnings inequality had generally decreased from 1988 on and speculated that perhaps this signaled a turnaround in the trend of increasing earnings inequality. After updating those data and adding two additional years of observations, it appears that, while earnings inequality decreased slightly over the period 1988-91, the overall trend is still upwards. In fact, the S-Gini coefficients for the years 1993-95(a) are statistically significantly greater (higher inequality) than those in any other year in the study. These results contrast with the SIPP-based findings of Lerman (1997). The Gini coefficients he calculated declined by 1.4 percent over the period 1986-95, and, in particular, decreased by 3.2 percent over the years 1992-95.[40]

Comparing the statistics for years 1994 and 1995 in the (a) rows with those in the (b) rows clearly shows the effect of being able to use the HI taxable earnings and deferred compensation variables in calculating total wage and salary earnings. Mean earnings for 1994 and 1995 are higher for (b) than for (a), reflecting the fact that there are many cases in which HI taxable earnings for an observation are greater than the combination of wage and salary earnings and deferred compensation. The S-Gini coefficients, being higher for (b) than for (a), also reflect the fact that earnings picked up when the HI taxable earnings variable is utilized generally increase the earnings of those in the upper portion of the earnings distribution. As it is generally believed that the OASDI and HI taxable earnings variables are of higher quality than the total wage and salary earnings and deferred compensation variables in SSA's administrative records, it is likely that the statistics for 1994 and 1995 presented in the (b) rows more accurately reflect reality for those two years.

Table 2 shows earnings share by decile, as well as the decile dollar breaks, for the entire sample over 1981-95. The most striking thing about table 2 is the large increase in the share of earnings garnered by the decile at the top of the earnings distribution. This increase, from 33.14 percent in 1982 to 35.83 percent in 1995(a), comes at the expense of all other parts of the earnings distribution.[41] The earnings shares also generally reaffirm the patterns found in the S-Gini coefficients in table 1. The earnings share increased for the upper decile over 1981-88, while the shares generally fell for

Table 2.—Earnings share by decile, 1981-95

Decile	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994(a)[1]	1995(a)	1994(b)	1995(b)
	Percent of earnings																
1.....	0.35	0.31	0.30	0.30	0.31	0.30	0.29	0.29	0.30	0.31	0.31	0.29	0.27	0.27	0.28	0.26	0.28
2.....	1.38	1.28	1.26	1.23	1.29	1.23	1.20	1.22	1.25	1.29	1.30	1.24	1.21	1.18	1.23	1.16	1.21
3.....	2.83	2.68	2.67	2.62	2.71	2.63	2.56	2.58	2.63	2.70	2.71	2.64	2.60	2.55	2.63	2.51	2.58
4.....	4.68	4.49	4.49	4.41	4.49	4.40	4.29	4.30	4.35	4.42	4.41	4.32	4.28	4.24	4.32	4.18	4.25
5.....	6.70	6.46	6.47	6.38	6.43	6.37	6.25	6.21	6.26	6.29	6.27	6.17	6.13	6.10	6.15	6.01	6.05
6.....	8.82	8.54	8.58	8.50	8.53	8.48	8.32	8.27	8.32	8.32	8.29	8.18	8.15	8.12	8.15	8.01	8.01
7.....	11.27	10.91	10.98	10.91	10.92	10.88	10.66	10.58	10.63	10.61	10.60	10.47	10.45	10.44	10.43	10.30	10.26
8.....	14.39	13.93	14.02	13.97	13.95	13.89	13.60	13.50	13.56	13.54	13.53	13.40	13.38	13.37	13.34	13.21	13.14
9.....	18.80	18.25	18.26	18.26	18.17	18.12	17.73	17.63	17.74	17.70	17.74	17.63	17.68	17.67	17.64	17.48	17.40
10.....	30.79	33.14	32.96	33.42	33.20	33.69	35.09	35.42	34.97	34.84	34.85	35.67	35.84	36.08	35.83	36.89	36.84
	Decile dollar breaks in 1992 dollars																
1.....	$1,533	$1,458	$1,429	$1,430	$1,541	$1,469	$1,491	$1,498	$1,516	$1,567	$1,560	$1,500	$1,446	$1,400	$1,484	$1,402	$1,485
2.....	3,959	3,858	3,867	3,869	4,103	3,974	4,008	4,035	4,072	4,214	4,202	4,122	4,039	3,921	4,088	3,926	4,094
3.....	7,199	7,092	7,220	7,235	7,587	7,455	7,463	7,481	7,543	7,689	7,659	7,611	7,484	7,361	7,566	7,374	7,578
4.....	11,114	11,039	11,229	11,256	11,624	11,531	11,611	11,536	11,539	11,619	11,512	11,503	11,358	11,244	11,463	11,264	11,483
5.....	15,060	15,055	15,389	15,484	15,882	15,892	16,051	15,925	15,879	15,872	15,725	15,741	15,580	15,477	15,654	15,502	15,679
6.....	19,434	19,420	19,928	20,155	20,587	20,649	20,820	20,659	20,562	20,495	20,329	20,410	20,254	20,149	20,275	20,200	20,328
7.....	24,707	24,714	25,369	25,704	26,214	26,292	26,492	26,257	26,135	26,012	25,831	25,992	25,808	25,714	25,799	25,805	25,898
8.....	31,731	31,736	32,599	33,098	33,714	33,814	33,977	33,729	33,593	33,429	33,274	33,549	33,357	33,246	33,312	33,386	33,472
9.....	42,511	42,761	43,310	44,327	45,055	45,291	45,546	45,377	45,232	45,034	44,907	45,434	45,536	45,527	45,753	45,779	46,024

[1]The statistics calculated for 1994(b) and 1995(b) are generated using additional earnings information (tax-deferred and Medicare-taxable earnings) available for those two years only. The entries for 1994(a) and 1995(a) do not make use of this additional information and are, therefore, calculated in a way that is consistent with the calculations in years 1981-93.

the other deciles, corresponding to the increasing S-Gini coefficients over that period found in table 1. From 1988 to 1991, the earnings share for those in the upper decile declined slightly, while shares for those in the other deciles increased, explaining the decreasing S-Gini coefficients over that period. From 1992 to 1994(a), earnings shares for the lower eight deciles showed generally large (in percentage terms) decreases, while shares for the upper two deciles showed increasingly large gains as one moves upwards along the earnings distribution, a fact again reflected in the large increase in the S-Gini coefficient from 1992 to 1994(a) in table 1. Finally, note that the numbers for 1994(b) and 1995(b) indicate that the share of earnings received by the upper decile is even larger, if one believes in the higher accuracy of the HI taxable earnings variable in SSA's administrative earnings files, than indicated in (a) columns. The disparity in earnings shown in the (b) columns of table 2 are larger than those in the (a) columns, just as one would expect given the S-Gini coefficients in table 1.

Sample Subdivided into Age Groups

Table 3 presents various earnings distribution statistics by age group. I arbitrarily chose to set the age ranges for the groups at 14-24, 25-34, 35-54, 55-64, and 65-85.[42] The first age group encompasses individuals who are still in school or who are relatively early in their careers. Those in the second age group are likely out of school and working, but members of this group are more likely to change jobs several times while searching for the "right" job. The third group consists of individuals in the core of their working lives, persons who are likely to have fewer dropout years than those in the younger two age groups. Group four, the 55-64 year olds, are probably nearing the end of their working careers and are preparing to retire. Finally, age group five is composed of those aged 65 or older, who are at or beyond what is considered to be the traditional retirement age, but who are still working either full or part time.

The "Mean earnings" numbers in table 3 show that those in both the 14-24 and 25-34 age groups had lower real mean earnings in 1995(a) than they did in 1981, while those in the other age groups all enjoyed increases in mean earnings over the period. Partly this might be a function of individuals staying in school longer, or of relatively more individuals pursuing a college education in 1995 than in 1981, particularly for the youngest age group. The patterns of earnings increases and decreases for each of the groups is also interesting. Earnings for the youngest age group generally declined over time. For all of the other age groups, mean earnings generally increased until 1987 or 1988, then mostly declined slightly from that point.

Also interesting are the patterns present in the "Proportion of sample" figures in table 3. The aging of the baby boom generation stands out clearly, as demonstrated by the fact that the number of 14-24 year olds in the sample decreased dramatically, while at the same time the relative number of 35-54 year olds increased substantially.[43] The only other age group to increase in relative size over the period 1981-95 is the 65-85 year olds, with all of the increase occurring since 1988. In part, this might be a sign that individuals are working longer, perhaps part time, or that they are reentering the labor force after retirement. The finding for 65-85 year olds is consistent with recent work on the labor force participation of older workers.[44]

As could be predicted from the mean earnings and proportion of sample numbers, the share of earnings garnered by the youngest two age groups fell sharply from 1981 to 1995(a). Also, as expected, those in the 35-54 age group enjoyed a large increase in their share of earnings, with an increase from about 43 percent of the earnings in 1981 to nearly 58 percent of the earnings in 1995(a). Those in the top age group also enjoyed an increase of nearly 15 percent in their share of earnings over the period.

As found by Utendorf (1998) and others, the largest contributor to overall earnings inequality is within-group inequality. The "Within-group Ginis" and "Within-group inequality term" numbers in table 3 bear this out. The youngest and the oldest groups had the most unequal distribution of earnings within their groups, with the Gini coefficients for both groups being greater than the overall Gini coefficient for every year in the sample. Especially for those in the 65-85 year old group, there is a great deal of earnings disparity, with the Gini coefficients reaching as high as 0.702 in 1994(a). Interestingly, those in the 25-34 age group had the lowest within-group inequality for every year in the sample. It is unclear why earnings inequality within that group would be substantially lower than within any other group. Within-group earnings inequality generally increased over the period for every age group in the study, which agrees with what Katz and Murphy (1992) found for the 1970-87 period.[46]

The "Stratification index" numbers in table 3 show that the youngest three age groups became increasingly stratified over the period 1981-95(a) in the sense that they increasingly occupied distinct segments of earnings distributions during those years. For example, in 1995(a), over 89 percent of the 14-24 year olds in the sample had earnings below the median earnings of the entire sample, with this higher concentration of 14-24 year old earners at the lower end of the earnings distribution leading to the high relative stratification number. The fact that the stratification index for the oldest age group became increasingly negative implies that that "group" increasingly became more than one group. It is likely that there were relative increases in the numbers of those in the oldest age group who continued to work full time and enjoy relatively higher earnings, separating themselves, at least along the earnings dimension, from the low earners in their age group.

The between-group inequality term in table 3 generally increases over the period, from 0.089 in 1981 to 0.104 in 1995, meaning that, by this measure, the age groups identified in the article became less equal with regard to earnings over the period.[47] This result is to be expected given the decrease in the relative share of earnings by all but the 35-54 and the 65-85 age groups. In addition, the increasing stratification of the three youngest age groups would imply that the earnings of those three groups are growing relatively less equal. This

finding of increased earnings inequality between age groups corresponds to the conclusions of several authors that earnings inequality has increased across age/experience groups.[48]

Comparing the 1994 and 1995(a) and (b) columns once again demonstrates the consequences of considering HI taxable earnings (the (b) columns) when formulating the total wage and salary earnings variable. The mean earnings for the various age groups are higher in 1994(b) and 1995(b) than in their (a) counterparts, as is to be expected. Also, given that using the HI taxable earnings information likely leads to increases in the reported earnings of high earners more often than in the reported earnings of low earners, it isn't surprising that the overall and within-group Gini coefficients are higher for the (b) columns than the (a) columns. It is somewhat surprising that the stratification index for 14-24 year olds is lower in the (b) columns than in the (a), meaning that using HI taxable earnings decreases the distinctiveness of the youngest age group along the earnings dimension. The reason(s) behind this are unclear and merit further work.

Sample Subdivided into Birth Cohorts

Table 4 shows the decomposition of the overall annual Ginis by birth cohort for total wage and salary earnings. As with choosing the age categories to use, I made somewhat arbitrary choices with regard to the years spanned by each birth cohort.

Table 3.—Earnings distributions by age group, based on total wage earnings, 1981-95

Variable	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994(a)[1]	1995(a)	1994(b)	1995(b)
Mean earnings (1992 dollars):																	
14-24	$8,820	$8,442	$8,162	$8,161	$8,200	$8,070	$8,095	$7,845	$7,684	$7,518	$7,238	$7,028	$6,851	$6,948	$6,654	$6,965	$6,751
25-34	20,371	20,132	20,304	20,577	20,941	20,961	21,265	21,003	20,767	20,555	20,117	20,163	19,941	19,783	19,744	19,955	20,026
35-54	26,686	27,779	28,254	28,910	29,285	29,495	30,398	30,346	29,986	29,664	29,280	29,836	29,635	29,428	29,510	30,019	30,108
55-64	25,169	26,551	26,624	27,250	27,560	27,652	28,235	28,789	27,857	27,489	27,198	27,635	27,379	27,295	27,652	27,897	28,478
65-85	12,767	14,171	14,090	14,531	14,950	15,084	15,377	16,252	15,197	15,477	15,021	14,874	14,237	14,106	14,477	14,290	14,743
Proportion of sample:																	
14-24	0.272	0.257	0.244	0.244	0.236	0.232	0.227	0.225	0.219	0.211	0.201	0.197	0.194	0.193	0.191	0.193	0.191
25-34	.281	.283	.283	.284	.287	.287	.286	.285	.283	.280	.277	.271	.265	.259	.255	.259	.255
35-54	.317	.330	.341	.344	.350	.356	.364	.369	.377	.388	.401	.410	.418	.424	.431	.424	.431
55-64	.100	.102	.103	.100	.099	.096	.094	.092	.090	.090	.090	.090	.090	.090	.090	.090	.090
65-85	.029	.028	.029	.028	.028	.029	.029	.029	.030	.031	.031	.032	.034	.034	.033	.034	.033
Earnings share:																	
14-24	.123	.108	.097	.095	.091	.087	.083	.080	.077	.073	.067	.063	.061	.061	.058	.060	.058
25-34	.294	.283	.280	.280	.281	.280	.276	.272	.269	.264	.257	.248	.241	.234	.229	.233	.228
35-54	.434	.455	.469	.475	.480	.489	.500	.507	.518	.527	.541	.555	.564	.571	.578	.573	.579
55-64	.130	.134	.134	.131	.128	.124	.120	.120	.115	.114	.113	.113	.113	.112	.113	.112	.114
65-85	.019	.020	.020	.020	.020	.020	.020	.022	.021	.022	.022	.022	.022	.022	.022	.022	.022
Within-group Ginis:																	
14-24	.510	.517	.521	.523	.517	.525	.538	.531	.529	.525	.527	.528	.531	.553	.530	.554	.536
25-34	.406	.416	.416	.416	.413	.418	.427	.424	.423	.419	.421	.426	.429	.431	.425	.435	.431
35-54	.423	.450	.444	.450	.447	.452	.462	.464	.460	.458	.458	.466	.466	.465	.464	.473	.472
55-64	.438	.470	.470	.478	.480	.488	.500	.517	.507	.506	.509	.520	.523	.521	.523	.529	.534
65-85	.593	.636	.641	.653	.650	.651	.662	.684	.673	.682	.675	.685	.686	.702	.689	.705	.693
Stratification index:																	
14-24	.257	.268	.291	.294	.310	.304	.271	.299	.307	.321	.325	.343	.347	.282	.364	.280	.340
25-34	.132	.131	.140	.144	.148	.147	.145	.147	.147	.150	.147	.148	.149	.151	.154	.151	.152
35-54	.130	.150	.155	.158	.159	.162	.181	.188	.191	.191	.193	.203	.209	.213	.214	.218	.218
55-64	.086	.096	.086	.082	.073	.070	.071	.070	.062	.057	.056	.053	.047	.048	.050	.049	.052
65-85	-.088	.126	-.134	-.143	-.148	-.143	-.155	-.176	-.171	-.180	-.174	-.183	-.184	-.198	-.188	-.20	-.192
Overall Gini	.484	.503	.502	.507	.503	.508	.518	.520	.516	.513	.513	.521	.523	.526	.522	.532	.529
Between-group inequality term	.089	.090	.094	.095	.094	.094	.096	.096	.097	.096	.098	.100	.102	.103	.104	.104	.104
Within-group inequality term	.434	.454	.451	.455	.452	.457	.467	.469	.465	.463	.463	.471	.472	.474	.470	.480	.478
Stratification term	-.039	-.041	-.042	-.043	-.042	-.043	-.045	-.045	-.046	-.046	-.048	-.050	-.052	-.051	-.053	-.052	-.053

[1]The statistics calculated for 1994(b) and 1995(b) are generated using additional earnings information (tax-deferred and Medicare-taxable earnings) available for those two years only. The entries for 1994(a) and 1995(a) do not make use of this additional information and are, therefore, calculated in a way that is consistent with the calculations in years 1981-93.

I elected to use birth cohorts that covered 10-year periods in order to simplify the analysis and in the belief that the birth cohorts chosen provide insights into the overall effects the cohort one is born into have on one's place in the earnings distribution. In addition, the birth cohorts were chosen with the idea of keeping the individuals in the sample between the ages of 13 and 86 at the beginning and ending years of the study period.[49] Each of the birth cohorts chosen experienced major, possibly life-shaping events as a group. Most of the 1909-18 birth cohort grew up during the Roaring Twenties and became adults during the Great Depression; many of those born between 1919 and 1928 grew up during the Great Depression and experienced World War II as young adults; a large number of those in the 1929-38 birth cohort had the post-

Table 4.—Earnings distributions by birth cohort, based on total wage earnings, 1981-95

Variable	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994(a)[1]	1995(a)	1994(b)	1995(b)
Mean earnings (1992 dollars):																	
1909-18	$16,079	$16,218	$14,625	$13,716	$13,230	$12,563	$12,074	$12,210	$11,104	$10,892	$10,116	$9,690	$8,969	$8,332	$8,730	$8,394	$8,776
1919-28	26,163	27,028	26,624	26,549	26,014	25,198	24,599	24,313	21,749	20,050	18,510	17,015	15,223	14,208	13,575	14,386	13,815
1929-38	27,235	28,656	29,274	29,922	30,323	30,427	31,225	31,267	30,271	29,499	28,592	28,530	27,379	26,265	25,439	26,843	26,183
1939-48	25,541	26,729	27,558	28,708	29,495	30,097	31,470	31,861	31,774	31,740	31,482	32,284	32,280	32,114	32,264	32,913	33,038
1949-58	18,414	19,153	20,304	21,510	22,732	23,620	24,860	25,300	25,810	26,168	26,449	27,486	27,834	28,197	28,808	28,687	29,364
1959-68	7,013	7,551	8,204	9,230	10,606	11,950	13,602	14,869	16,065	17,233	17,988	19,114	19,941	20,782	21,759	20,975	22,080
Proportion of sample:																	
1909-18	0.036	0.030	0.025	0.020	0.017	0.015	0.013	0.011	0.010	0.009	0.008	0.007	0.006	0.006	0.005	0.006	0.005
1919-28	.114	.109	.104	.094	.086	.078	.070	.063	.056	.050	.044	.039	.035	.031	.027	.031	.027
1929-38	.141	.140	.139	.134	.132	.130	.129	.126	.125	.123	.120	.117	.112	.106	.100	.106	.100
1939-48	.204	.204	.204	.200	.200	.200	.202	.202	.204	.206	.208	.209	.210	.211	.211	.211	.211
1949-58	.297	.291	.285	.281	.280	.282	.285	.288	.292	.296	.300	.304	.308	.313	.317	.313	.317
1959-68	.208	.226	.244	.271	.285	.296	.302	.309	.313	.316	.320	.324	.328	.334	.340	.334	.340
Earnings share:																	
1909-18	.029	.024	.018	.013	.010	.008	.007	.006	.005	.004	.003	.003	.002	.002	.002	.002	.002
1919-28	.153	.147	.134	.118	.103	.088	.074	.064	.051	.041	.033	.026	.021	.017	.014	.017	.014
1929-38	.196	.198	.198	.191	.185	.177	.171	.165	.156	.148	.140	.131	.120	.108	.096	.109	.097
1939-48	.267	.270	.273	.273	.271	.270	.271	.269	.268	.267	.266	.266	.265	.263	.259	.264	.260
1949-58	.280	.277	.281	.286	.293	.298	.302	.305	.312	.317	.323	.330	.336	.342	.348	.342	.347
1959-68	.075	.085	.097	.119	.139	.159	.175	.192	.208	.223	.234	.244	.256	.269	.282	.267	.280
Within-group Ginis:																	
1909-18	.564	.617	.632	.657	.659	.667	.672	.693	.690	.693	.690	.711	.726	.744	.726	.745	.727
1919-28	.426	.464	.470	.487	.500	.521	.550	.591	.600	.622	.641	.662	.676	.703	.693	.706	.698
1929-38	.423	.455	.450	.455	.455	.462	.475	.485	.480	.483	.490	.512	.523	.535	.547	.542	.557
1939-48	.420	.443	.440	.449	.446	.454	.466	.469	.466	.466	.465	.474	.474	.474	.476	.483	.484
1949-58	.408	.415	.416	.418	.417	.424	.437	.438	.441	.442	.445	.455	.457	.459	.459	.466	.466
1959-68	.517	.520	.519	.512	.488	.474	.463	.442	.431	.421	.421	.425	.429	.433	.433	.437	.439
Stratification index:																	
1909-18	-.076	-.110	-.127	-.147	-.154	-.149	-.147	-.168	-.159	-.163	-.153	-.173	-.188	-.203	-.189	-.205	-.190
1919-28	.110	.109	.086	.066	.036	.014	-.014	-.046	-.082	-.114	-.136	-.159	-.171	-.195	-.188	-.198	-.192
1929-38	.113	.130	.132	.132	.129	.127	.133	.134	.123	.110	.095	.076	.048	.022	-.002	.023	-.001
1939-48	.097	.112	.118	.125	.128	.132	.153	.162	.167	.169	.169	.176	.180	.181	.178	.185	.182
1949-58	.155	.141	.140	.136	.132	.128	.130	.133	.136	.137	.141	.151	.159	.168	.174	.171	.178
1959-68	.340	.310	.293	.264	.250	.233	.208	.206	.195	.188	.176	.169	.163	.160	.155	.160	.155
Overall Gini	.484	.503	.502	.507	.503	.508	.518	.520	.516	.513	.513	.521	.523	.526	.522	.532	.529
Between-group inequality term	.093	.091	.093	.092	.086	.081	.079	.075	.078	.078	.082	.091	.101	.110	.108	.111	.109
Within-group inequality term	.430	.451	.450	.456	.453	.457	.466	.466	.461	.458	.458	.466	.468	.469	.468	.476	.476
Stratification term	-.038	-.040	-.041	-.040	-.035	-.030	-.026	-.021	-.022	-.023	-.027	-.035	-.045	-.047	-.047	-.054	-.056

[1]The statistics calculated for 1994(b) and 1995(b) are generated using additional earnings information (tax-deferred and Medicare-taxable earnings) available for those two years only. The entries for 1994(a) and 1995(a) do not make use of this additional information and are, therefore, calculated in a way that is consistent with the calculations in years 1981-93.

WWII expansion in the United States shape their view of the world; many in the sample born between 1939 and 1948 came of age in the 1960s; most of the individuals born between 1949 and 1958 grew up during the 1960s and had their early adult lives influenced by the economic chaos surrounding the OPEC oil shocks of the 1970s; and many of those in the 1959-68 birth cohort enjoyed the long economic expansion of the 1980s as they moved into adulthood.

The "Mean earnings" numbers in table 4 show the expected pattern. Those in the 1909-18 and the 1919-28 birth cohorts have generally declining real mean earnings over the period as those who remain in the labor force move to part-time/part-year employment. Those in the most recent three birth cohorts, 1939-48, 1949-58, and 1959-68, show generally increasing mean earnings over time as they either move into their prime earnings years (those in the 1939-48 and the early 1949-58 birth cohorts) or they move from part-time jobs while in school to full-time, post-education jobs (the most recent birth cohort). Interestingly, only the 1929-38 birth cohort experiences increasing, then decreasing mean earnings over the period of the study. Evidently, enough of the older members of the birth cohort move into part-time/part-year employment after "retirement" to cause mean earnings for the group to begin to decline after 1988.

The figures in the "Proportion of sample" section of table 4 show the effects of attrition, either due to death or to exiting the labor force, by those in the oldest birth cohorts. The proportion of those in the sample from the oldest three birth cohorts declines steadily over time, while the proportion of those in the two most recent birth cohorts increases over the period in question. On the surface, it appears odd that the proportion of those born between 1939 and 1948 initially declines, but then increases, over the years of the study since those individuals should be at or near their prime working ages throughout the years studied. However, when one examines the fluctuation in the actual number of observations over the period, one does see small declines in the numbers of individuals in that birth cohort towards the terminal years of the study. The sample size is shrinking more quickly, however, because of the attrition from the three oldest birth cohorts, thus leading to the slight increases in the proportion of the sample attributed to those born between 1938 and 1948.

Given the Mean earnings and the Proportion of sample numbers, the figures in the "Earnings share" category of table 4 are to be expected, at least for those in the oldest three and the two most recent birth cohorts. The earnings share garnered by those in the oldest three birth cohorts declined rather steadily, for the most part, over the years 1981-95(a). Those in the two most recent birth cohorts enjoyed relatively large increases in earnings share, particularly those in the 1959-68 birth cohort. The share of earnings gained by those in the 1939-48 birth cohort actually declined slightly over the 15 years of the study, despite the fact that the group's mean earnings increased and that they formed a slightly larger part of the sample in 1995 than they did in 1981. The explosive growth of the mean earnings of those in the 1959-68 birth cohort, as well as their relative increase in size, account for most of the decline in the earnings share experienced by the 1939-48 birth cohort.

Several interesting patterns emerge from the "Within-group Ginis" numbers in table 4. For every birth cohort but the most recent one, the within-group Gini coefficients increase, for the most part, throughout the 1981-95(a) period.[50] For the 1959-68 cohort, though, the within-group Gini coefficients decrease substantially from 1981 to 1991 before reversing course and increasingly slightly from 1992 through 1995(a). This seems to indicate that the increasing mean earnings of the group come about because those who were low earners in 1981 "caught up" somewhat, over the years of the study, with those who were higher earners in 1981. Another pattern present in the numbers is that the smaller birth cohorts (smaller in the sense of being a smaller proportion of the sample) generally have the higher Gini coefficients. It is likely that a relatively small number of individuals had high earnings for the 1909-18 birth cohort in 1994(a), while the rest of the group had relatively low earnings, thus leading to the high 0.744 Gini coefficient. Finally, it is interesting that the 1959-68 birth cohort had the second highest Gini coefficient each year through 1984. From that point on, the decline in Gini coefficient of that birth cohort, combined with the increases in the Gini coefficients of the other birth cohorts, results in the 1959-68 birth cohort having the lowest within-group Gini coefficient by 1995(a).

The results presented in the "Stratification index" portion of table 4 are mixed. The increasingly negative index numbers of the oldest birth cohort indicate that the "group" became less and less one group over the period, at least along the earnings dimension. This corroborates the story told by the within-group Gini coefficients for the 1909-18 group of there being a group of relatively high earners and a group of relatively low earners in that birth cohort, especially during the latter years of this study. The 1919-28 birth cohort went from being a slightly stratified group to being more than one distinct group over the period 1981-95(a), while the 1929-38 birth cohort moved from being slightly stratified in 1981 to being nonstratified in 1995(a). The most recent birth cohort was moderately stratified in 1981, but, like the 1929-38 birth cohort, occupied less and less of a distinct stratum as time went by. The 1939-48 cohort, on the other hand, became increasingly stratified over the period, moving from being the second least stratified group in 1981 to the most stratified group by 1995(a).

Within-group inequality is again the most important contributor to overall earnings inequality for the birth cohorts chosen, as shown by the "Within-group inequality term" numbers in table 4. While between-group inequality exists (as shown by the "Between-group inequality term" numbers), it clearly does not influence overall earnings inequality as much as within-group inequality. It is interesting, however, that within-group inequality is relatively constant over the period 1987 to 1995(a), and that much of the growth in overall earnings inequality comes from growth in between-group inequality, at least by this measure of earnings inequality.

The differences between the (a) and the (b) columns in table 4 are similar to the differences between the (a) and (b)

sections of tables discussed earlier in the article. Using HI taxable earnings in determining total wage and salary earnings in the (b) columns results in higher mean earnings for every birth in 1994 and 1995 compared with the (a) columns. These differences are not distributed evenly across the earnings distribution within age groups, however, resulting in higher within-group Gini coefficients across the board for 1994 and 1995 in column (b). To the extent that using HI taxable earnings captures "true" total wage and salary earnings, the numbers in the 1994 and 1995 (b) columns are likely to represent reality more closely than those presented in the (a) columns.

V. Conclusion

A thorough understanding of earnings provides valuable insights into the economic well-being of individuals and groups within society. A detailed knowledge of earnings and changes in patterns of earnings is also necessary in order to accurately forecast the financial future of the Social Security program, either under current law or under various plans to reform the program.

This article uses Social Security Administration data to examine changing earnings distributions in the United States over the 1980s and early to mid-1990s. These unique data provide several advantages over data typically used in studies of this sort. Because the earnings information comes directly from the W-2 forms filed by employers, these data minimize the problem of self-reporting errors that are often present in survey data. Also, we are on the verge of having access to even better data for this type of analysis as more and more years of HI taxable earnings, with no taxable maximum after 1993, become available. Finally, because of the large number of observations contained within the dataset, I am able to provide better tests of the statistical significance of year-to-year fluctuations in earnings inequality, even when the data are segmented into age and cohort groups.

First and foremost, I find that earnings inequality continues to trend upwards for the overall United States population, though the increase in earnings inequality from 1988 to 1995 is small in magnitude overall. In an earlier article that examined the years 1981-93, I speculated that the upward trend in earnings inequality might have leveled off, or even perhaps reversed, because of a decrease in the S-Gini coefficients over the period 1988-92. After updating the data and adding two additional years of observations, it seems clear that the dip observed for the years 1988-92 was merely a pause and that earnings inequality is still trending upwards. The S-Gini coefficients for the years 1993-95 are higher than those for any other year in the study, and the results are statistically significant at very high levels of confidence.

The second important point to be made is that the share of earnings going to the upper decile of the earnings distribution continues to increase at the expense of all other deciles of the distribution. In 1995, nearly 36 percent [51] of all earnings in the United States accrued to the 10 percent of the population at the upper end of the earnings distribution. The upper two deciles garnered over 53 percent of the earnings in that year. More work is needed to pinpoint why the earnings share of the upper decile continues to increase and to consider the long-run effects.

The real mean earnings of those in the 14-24 age group fell dramatically (by nearly 25 percent) over the period 1981-95. Whether this represents a "worsening" of their condition is not clear. The decrease in mean earnings might simply be an indication that more individuals were staying in school longer (and therefore working part-time/part-year jobs) in order to better prepare themselves for future careers. However, it could also be an indication that for many young people, part-year/part-time jobs are the only types of employment available.

Another idea to come from this work is that between-group inequality, when dividing the sample into either various age groups or into birth cohort groups, is increasing. This is in contrast to a division of groups along race and/or gender dimensions as in Utendorf (1998), but agrees with much of the literature that examines changes in earnings distributions by age/experience. Although the increases in between-group inequality presented in this article are relatively small, they are, nevertheless, real. Still, the contribution to overall earnings inequality by between-group inequality is small when compared to that of within-group inequality.

Future work will examine more thoroughly the changes that have taken place in the upper part of the earnings distribution. In addition, SSA anticipates being able to match administrative information on total wage and salary earnings to public-use survey files, such as the Survey of Income and Program Participation and the Current Population Survey. Such matches will provide information on educational attainment and household demographic characteristics, thereby improving the explanatory power of the future analyses.

Notes

[1] Levy and Murnane (1992), p. 1333.

[2] Freeman (1995), pp. 17-18.

[3] Lerman (1997), p. 24.

[4] Age is often used as a proxy for experience in these types of studies. Generally, the age of the respondent at the time of the survey, less an estimated labor force entry age, is used as that respondent's labor force experience level.

[5] Levy and Murnane (1992), p. 1357, indicate that the median earnings in 1979 of 45-54 year old male college graduates were 47 percent higher than the 1979 median earnings of 25-34 year old male college graduates. By 1987, this percentage had fallen to 45 percent, thus slightly negating the increasing between-group inequality along the age dimension found in the rest of the population.

[6] Berger argues that workers from large cohorts experience a form of "congestion" in which they experience fewer opportunities for advancement and smaller pay increases.

[7] Johnson indicates that relative wages might have been depressed by as much as 5 percent because of the large influx of

college graduates into the labor force in the 1970s. See Johnson (1997), p. 45.

[8] The CWHS Active file is a 1-percent sample of all individuals with Social Security numbers who have a record of earnings posted to SSA's Master Earnings File.

[9] These files also contain other information, such as indicators of race and gender, and information for additional years not directly relevant to this study. For a more comprehensive introduction to the CWHS family of files, see Smith (1989).

[10] I do not include self-employment income in my analysis. The Social Security Administration receives information on self-employment income only to the extent that it is taxable for OASDI purposes.

[11] See Levy and Murnane (1992).

[12] Prior to 1978 there is an estimate of total earnings based on taxable earnings up to the taxable maximum. For those individuals at or above the taxable maximum, the estimate of total earnings was derived from the value of the taxable maximum combined with information regarding the quarter in which the individual's taxable earnings reached the taxable maximum.

[13] I spent a great deal of time running consistency checks and testing the data in general to determine their fitness for use in this type of exercise. Many of my questions about or problems with the data were cleared up by Creston Smith and his colleagues in SSA's Office of Research, Evaluation and Statistics' Division of Earnings Statistics and Analysis.

[14] Individuals generally have a difficult time legally preventing their actual wage and salary earnings from appearing on their W-2 forms, except to the extent that they can contribute to tax-deferred saving plans. It is also possible to misrepresent actual wage and salary earnings on W-2 forms to the extent that individuals participate in the "underground" economy.

[15] In fact, for several years covered by this article, the top-code limit for wage and salary earnings in the March CPS Supplements was $75,000.

[16] The number of observations varies from year to year, depending on the size of the workforce with wage and salary earnings. Also note that the individuals in the observation pool differ to some extent from year to year because of permanent or temporary changes in employment status.

[17] The early articles in this series on earnings distributions will be limited to descriptive analyses. Future work utilizing a public use dataset linked to SSA administrative data will provide a better basis for explaining the reasons behind the changes seen in patterns of earnings inequality.

[18] Only individuals with positive earnings in one of the earnings variables were included in the sample.

[19] Throughout the article I use the phrase "taxable maximum" to refer to the OASDI taxable maximum.The OASDI taxable maximum is automatically updated each year in proportion to the increase in the United States average wage level. See any recent *Annual Statistical Supplement* to the *Social Security Bulletin* for more information about the OASDI taxable maximum.

[20] The law limits the amount of wage and salary earnings that one can defer in any given year. In 1995, individuals could defer no more than $9,240 of their pretax earnings into 401(k)-type plans, for example.

[21] Obviously, for individuals with true total earnings above the annual taxable maximum earnings amount, this sort of adjustment will capture only part of the missing true total earnings, since the OASDI taxable earnings variable does not (generally) exceed the taxable maximum in the dataset. This means that true earnings are likely somewhat understated for these high earners and that the earnings inequality measures calculated, therefore, likely understate the true degree of inequality. In fewer than 0.7 percent of the observations in any given year OASDI taxable earnings are at the taxable maximum, while, at the same time, total wage and salary earnings are lower than the taxable maximum. Therefore, the degree to which the earnings measures presented later in the article are affected should be rather small.

[22] In addition, it is likely that I do not capture any of the deferred compensation for certain other individuals in the sample. For example, there are many observations where reported total wage and salary earnings amounts are greater than the taxable maximum (and therefore greater than the taxable earnings amount). There is not enough information in the dataset to determine whether these individuals had any deferred compensation. Consequently, true total wage and salary earnings amounts, particularly for high earners, are probably somewhat understated.

[23] The deferred compensation variable contains information on earnings that are not subject to the income tax but are subject to OASDI and HI taxation. For example, contributions to tax-deferred 401(k) plans would not show up in the file(s) in the total wage and salary earnings field, but rather in the deferred compensation field.

[24] There were fewer than 10 observations for either 1994 or 1995 out of nearly 1.33 and 1.36 million observations, respectively, for which the OASDI taxable earnings variable exceeded the HI taxable earnings variable. For nearly 12 percent of the observations in both 1994 and 1995, the HI taxable earnings variable was greater than the combination of the total wage and salary earnings and the deferred compensation variables. In almost 7 percent of the cases, the combination of the total wage and salary earnings and the deferred compensation variables exceeded the HI taxable earnings variable. For all other cases, the different earnings variables were equal.

[25] In 1995, for example, this eliminated about 0.2 percent of the sample.

[26] Thanks go to my colleague, David Weaver, for running a check on some of the more unusual cases (several individuals well into triple digit ages with large wage and salary earnings amounts reported) against SSA's Master Beneficiary Record. In the majority of the unusual cases, it was clear that the earnings files had incorrect years of birth.

[27] See Braun (1988) and Slottje (1989) for a detailed comparison of the various measures of income or earnings inequality.

[28] In other words, $G = \frac{1}{2 n^2 \bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| y_i - y_j \right|$ where y represents earnings. See Deaton and Muellbauer (1986, pp. 232-237) for a thorough discussion of the traditional Gini coefficient.

[29] Like the traditional Gini coefficient, the S-Gini has an intuitive geometric interpretation. Referring back to chart 1, the S-Gini is twice the weighted area between the Line of Equality and the Lorenz curve, where the weights depend on the observation's rank in the earnings distribution. The S-Gini indices are constructed so that

the coefficients must lie between zero and one, just as with the traditional Gini coefficient.

[30] Lorenz curve ordinates can be thought of as "points" along the curve.

[31] Their techniques for deriving the asymptotic variance of the S-Gini estimators do not require knowledge of the underlying distribution from which the data are drawn. This is important in that the 1-percent sample from the CWHS family of files is a stratified cluster probability sample, which would typically affect the sampling errors from estimation. The distribution-free property of the S-Gini indices minimizes the importance of this complication.

[32] Please see Yitzhaki and Lerman (1991) for a very thorough discussion of stratification and its relationship to measures of inequality, particularly its relationship to the Gini index. Their paper also contains a complete description of the derivation of the stratification indices and their properties.

[33] The decomposition works in the following way. A Gini coefficient, referred to as the within-group Gini, is calculated for each of the individual groups being studied by restricting the sample to members of that group only. Then an overall within-group inequality term is calculated by multiplying the within-group Gini coefficient for a group by the share of total earnings attributable to that group and summing these products across all groups. Next, a stratification index is calculated for each of the groups in question using the methods set forth by Yitzhaki and Lerman in their paper. An overall stratification term is computed by summing the products of the stratification index for each group, the share of total earnings attributable to that group, the within-group Gini for that group, and one minus the proportion of the sample in the group. Finally, a between-group inequality term is derived for the sample using techniques found in the Yitzhaki and Lerman paper. The overall Gini coefficient is given by the sum of the overall within-group inequality term, the overall stratification term, and the between-group inequality term. Since there is some overlap between the stratification term and the overall within-group inequality and the between-group inequality terms in the text, I point out where this overlap matters.

[34] Yitzhaki and Lerman (1991), p. 318.

[35] Note that, as indicated in Section II of the article, duplicate calculations were carried out on two slightly different data sets for both 1994 and 1995. The tables in this article reflect that by having both (a) and (b) columns (or rows, depending on the table) for both years. The statistics calculated for (b) are generated using the additional earnings information (variables for deferred earnings and HI taxable earnings) available in the files for those two years. Those for (a) do not make use of this additional information, meaning the method used to calculate the total wage and salary earnings variables in (a) is consistent across all of the years of the study.

[36] Barrett and Pendakur (1995) use 20 quantiles in their paper. The sensitivity tests they performed indicated that increasing the number of quantiles to 100 did not significantly improve the accuracy of their S-Ginis. Tests I conducted on my sample showed significant improvement in the accuracy of the estimated S-Gini with an increase of quantiles used from 20 to 100, but little or no gain from increasing the number of quantiles beyond 100. In order to reduce the computational burden, I chose to use 100 quantiles for this article.

[37] As I indicated in the previous section, the S-Gini indices are "ethically tunable" in that one can adjust the δ parameter to place more weight on the part of the earnings distribution with which one is most concerned. Since the techniques developed by Yitzhaki and Lerman are designed to decompose the traditional Gini coefficient, I chose to use $\delta=2$, which corresponds to the traditional Gini coefficient, in my S-Gini calculations for consistency across the measures presented in the article.

[38] I used the Total Personal Consumption Expenditures deflator to adjust earnings for changes in the price level over time. All earnings are given in terms of 1992 dollars.

[39] See Utendorf (1998).

[40] My results are more in line with Lerman's CPS-based work. The Gini coefficients that he calculated using the CPS show growth in earnings inequality to be basically flat over the period 1986-92 but increasing sharply over the years 1992-95.

[41] I am still investigating the tremendous one-year jump in the earnings share garnered by the upper decile from 1981 to 1982. I suspect it is some artifact of the data rather than a representation of reality.

[42] While I have enough observations to divide the sample into even more tightly focused age groups, it was not clear to me that there was anything to be gained by doing so. Because certain of the calculations are fairly computer intensive, I decided to focus on the broader groups.

[43] Also note that, since 1986, the relative number of 25-34 year olds in the sample fell as the baby boom generation moved out of that group and into the 35-54 age group.

[44] A colleague of mine, Michael Leonesio, used SSA administrative data to show that the labor force participation rates of both older men and older women have increased since the mid- to late 1980s. See Leonesio (1998).

[45] One possible explanation, put forth by Julie-Anne Cronin, the paper's discussant at a Society of Government Economists session at the Allied Social Science Associations (ASSA) meetings in New York in 1999, is as follows: For those aged 25-34, the earnings of individuals with lower educational attainment but longer work experience might be similar to the earnings of those with higher educational attainment but shorter work experience. It is also possible that some of the results shown in the tables are driven by gender differences, as pointed out to me by Michael Leonesio. I intend to examine the interaction of gender, age, and birth cohort in future work.

[46] S-Gini coefficients ($\delta= 2$) calculated for each of the five age groups used in table 3 show that for differences in the year-to-year pairings within an age group to be statistically significant at the .05 level, it generally is the case that the differences must be at least: .009 for the 14-24 age group; .003 for the 25-34 age group; .003 for the 35-54 age group; .006 for the 55-64 age group; and .013 for the 65-85 age group.

[47] It is not correct to simply divide a coefficient in the "Between-group inequality term" row for a particular year by the coefficient in the "Overall Gini" row for the corresponding year to arrive at a percentage of inequality attributable to between-group inequality. There are components of between-group (as well as within-group) inequality present in the coefficients of the stratification term that would not be properly accounted for by doing this.

[48] See Levy and Murnane (1992), Katz and Murphy (1992), and Gottschalk (1997).

[49] A different possible configuration of birth cohorts that I might examine in future work would be to use two 10-year birth cohorts prior to the baby boom years, a birth cohort spanning the baby boom

years (or perhaps one birth cohort spanning the early baby boom and another spanning the late baby boom years), and a final birth cohort covering the "baby bust" years.

[50] S-Gini coefficients ($\delta = 2$) calculated for each of the six birth cohort groups show that for differences in the year-to-year pairings within a group to be statistically significant at the .05 level, it generally is true that the differences must be at least: .017 for those in the 1909-18 birth cohort; .017 for those born in the 1919-28 cohort; .007 for those born between 1929 and 1938; .004 for those in the 1939-48 birth cohort; .004 for those in the 1949-58 cohort; and .003 for those born between 1959 and 1968.

[51] This number increases to nearly 37 percent if one uses HI taxable earnings to build a total wage and salary earnings variable.

References

Barrett, G. and K. Pendakur. 1995. "The Asymptotic Distribution of the Generalized Gini Indices of Inequality." *Canadian Journal of Economics*, Vol. 28, No. 4b (November), pp. 1042-1055.

Beach, C. and R. Davidson. 1983. "Distribution-free Statistical Inference with Lorenz Curves and Income Shares." *The Review of Economic Studies*, Vol. 50, No. 4 (October), pp. 723-735.

Berger, M. 1985. "The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence." *Journal of Political Economy*, Vol. 93, No. 3 (June), pp. 561-73.

Bernstein, J. and L. Mishel. 1997. "Has Wage Inequality Stopped Growing?" *Monthly Labor Review*, Vol. 120, No. 12 (December), pp. 3-16.

Bishop, J., Chakraborti, S., and P. Thistle. 1989. "Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves." *The Review of Economics and Statistics*,Vol. 71, No. 4 (November), pp. 725-727.

Bishop, J., J. Formby, and W. Smith. 1991. "International Comparison of Income Inequality: Tests for Lorenz Dominance Across Nine Countries." *Economica*, Vol. 58 (November), pp. 461-477.

Blackburn, M., D. Bloom, and R. Freeman. 1990/91. "An Era of Falling Earnings and Rising Inequality?" *Brookings Review*, Vol. 9, No. 1 (Winter), pp. 38-43.

Braun, D. 1988. "Multiple Measurements of U.S. Income Inequality." *The Review of Economics and Statistics*, Vol. 70, No. 3 (August), pp. 398-405.

Burtless, G. 1990. "Introduction" in *A Future of Lousy Jobs?* G. Burtless (ed.). Washington, DC: The Brookings Institution.

Deaton, A. and J. Muellbauer. 1980. *Economics and Consumer Behavior*. New York, NY: Cambridge University Press, pp. 214-239.

Donaldson, D. and J. Weymark. 1980. "A Single-Parameter Generalization of the Gini Indices of Inequality." *Journal of Economic Theory*, Vol. 22, No. 1 (February), pp. 67-86.

Dooley, M. and P. Gottschalk. 1984. "Earnings Inequality Among Males in the U.S.: Trends and the Effect of Labor Force Growth." *Journal of Political Economy*, Vol. 91, No. 1 (February), pp. 59-89.

Freeman, R. 1995. "Are Your Wages Set in Beijing?" *The Journal of Economic Perspectives*, Vol. 9, No. 3 (Summer), pp. 15-32.

Gastwirth, J. 1972. "The Estimation of the Lorenz Curve and Gini Index." *Review of Economics and Statistics*, Vol. 54, No. 3 (August), pp. 306-316.

Gastwirth, J. and M. Gail. 1985. "Simple Asymptotically Distribution-Free Methods for Comparing Lorenz Curves and Gini Indices Obtained from Complete Data." *Advances in Econometrics*, Vol. 4, pp. 229-243, R. Basmann and G. Rhodes, Jr. (eds.). Greenwich, CT: JAI Press.

Gottschalk, P. 1997. "Inequality, Income Growth, and Mobility: The Basic Facts." *The Journal of Economic Perspectives*, Vol. 11, No. 2 (Spring), pp. 21-40.

Johnson, G. 1997. "Changes in Earnings Inequality: The Role of Demand Shifts." *The Journal of Economic Perspectives*, Vol. 11, No. 2 (Spring), pp. 41-54.

Karoly, L. 1992. "Changes in the Distribution of Individual Earnings in the U.S.: 1967-1986." *The Review of Economics and Statistics*, Vol. 74, No. 1 (February), pp. 107-115.

Katz, L. and K. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, Vol. 107, No. 1 (February), pp. 35-78.

Leonesio, M. 1998. "Notes on Retirement and the Effects of Raising Social Security's Normal Retirement Age (NRA) and Early Retirement Age (ERA)." Unpublished paper, February.

Lerman, R. 1997. "Reassessing Trends in U.S. Earnings Inequality." *Monthly Labor Review*, Vol. 120, No. 12 (December), pp. 17-25.

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

_____. 1985. "Income Inequality Effects by Income Source: A New Approach and Applications to the United States." *The Review of Economics and Statistics*, Vol. 67, No. 1 (February), pp. 151-156.

_____. 1989. "Improving the Accuracy of Estimates of Gini Coefficients." *Journal of Econometrics*, Vol. 42, pp. 43-47.

Levy, F. and R. Murnane. 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature*, Vol. 30, No. 3 (September), pp. 1333-1381.

Murphy, K. and F. Welch. 1992. "The Structure of Wages." *The Quarterly Journal of Economics*, Vol. 107, No. 1 (February), pp. 285-326.

Shorrocks, A. 1983. "Ranking Income Distributions." *Economica*, Vol. 50, No. 197 (February), pp. 3-17.

Slottje, D. 1989. *The Structure of Earnings and the Measurement of Income Inequality in the U.S.* New York, NY: North-Holland.

Smith, C. 1989. "The Social Security Administration's Continuous Work History Sample." *Social Security Bulletin*, Vol. 52, No. 10 (October), pp. 20-28.

Social Security Administration. 1998. *Annual Statistical Supplement* to the *Social Security Bulletin*. Washington, DC: U.S. Government Printing Office.

Topel, R. 1997. "Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality." *The Journal of Economic Perspectives*, Vol. 11, No. 2 (Spring), pp. 55-74.

Utendorf, K. 1998. "Recent Changes in Earnings Distributions in the United States." *Social Security Bulletin*, Vol. 61, No. 2 (Spring), pp. 12-28.

Welch, F. 1979. "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust." *Journal of Political Economy*, Vol. 87, No. 5 (October), pp. S61-73.

Yitzhaki, S. 1983. "On an Extension of the Gini Inequality Index." *International Economic Review*, Vol. 24, No. 3 (October), pp. 617-628.

Yitzhaki, S. and R. Lerman. 1991. "Income Stratification and Income Inequality." *The Review of Income and Wealth*, Vol. 37, No. 3 (September), pp. 313-329.