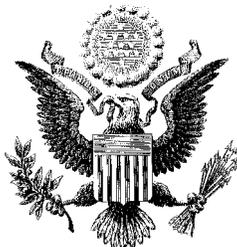


# A GUIDE TO TAX POLICY ANALYSIS: THE CENTRAL TENDENCY OF FEDERAL INCOME TAX LIABILITIES IN DISTRIBUTIONAL ANALYSIS

A JOINT ECONOMIC COMMITTEE STUDY



Vice Chairman Jim Saxton (R-NJ)

Joint Economic Committee  
United States Congress

May 2000

## Abstract

The analysis of tax data is a time intensive and complicated process. Much time and effort are spent collecting income and tax data, compiling data sets and running statistical analyses. However, it appears that relatively little time and effort are spent actually *understanding* the data and how best to present results to the public of analyses of using tax data. This is evident in the overuse of averages and the simplistic classification of taxpayers into income ranges and quintiles by tax distribution tables that are often highly publicized. This study shows that the link between income and tax liability is much more tenuous than that often presumed and that a variety of other factors can greatly affect tax liability. Specifically, this report finds that, among other things:

- Over 22 percent of all 1995 tax returns claimed zero tax liability – For calendar year 2000, the JCT estimates that 48.7 million out of 140.2 million taxpayers overall will have zero or negative federal income tax liability.
- In four out of the five income groups examined, a majority of taxpayers had tax liabilities that were either 25 percent greater than the average or 25 percent less than the average tax liability for each income group.
- In comparing federal income tax liabilities, distribution tables often misclassify and group millions of taxpayers into quintiles in which they have little tax liability in common.
  - Approximately 2.2 million taxpayers in the third quintile pay more in federal income taxes than 5.4 million taxpayers classified in the fourth quintile.
  - Over 3 million taxpayers in the fourth quintile pay more in federal income taxes than 4.1 million taxpayers classified in the fifth quintile.

Joint Economic Committee  
1537 Longworth House Office Building  
Washington, DC 20515  
Phone: 202-226-3234  
Fax: 202-226-3950  
Internet Address:  
<http://www.house.gov/jec/>



# **A GUIDE TO TAX POLICY ANALYSIS: THE CENTRAL TENDENCY OF FEDERAL INCOME TAX LIABILITIES IN DISTRIBUTIONAL ANALYSIS**

## **EXECUTIVE SUMMARY**

This study examines the misuse of averages as a sole measure of central tendency in presenting results of analyses based on income and tax data in distributional analysis. It finds that the use of averages in tax distribution tables is misleading to the public and the press and that the median is a more appropriate and representative measure to describe income and tax amounts for the taxpaying population.

Specifically, this report finds:

- Income and tax information based on tax returns filed with the IRS do not follow the pattern of a normal distribution. Hence, the use of averages is an inappropriate measure of central tendency.
- Over 22 percent of all 1995 tax returns claimed zero tax liability.
- The Joint Committee on Taxation estimates that for calendar year 2000, 48.7 million taxpayers out of 140.2 million taxpayers overall, or 34.7 percent, will have zero or negative federal income tax liability.
- For all taxpayers, the use of the average as the measure of central tendency overstates the tax liability for the “representative” taxpayer by almost 3 times the median value.
- The dispersion of taxpayers within any income group is impossible to determine from the information presented in tax distribution tables, but is shown to vary considerably.
- The grouping of taxpayers into income categories provide a false sense of precision and misleadingly suggest that taxpayers within the same groups necessarily have similar federal income tax liability.
- In four out of the five income groups examined, a majority of taxpayers had tax liabilities that were either 25 percent greater than the average or 25 percent less than the average tax liability for each income group.

- In comparing federal income tax liabilities, distribution tables often misclassify millions of taxpayers into quintiles in which they have little tax liability in common.
  - Approximately 2.2 million taxpayers in the third quintile pay more in federal income taxes than 5.4 million taxpayers classified in the fourth quintile.
  - Over 3 million taxpayers in the fourth quintile pay more in federal income taxes than 4.1 million taxpayers classified in the fifth quintile.
- The use of averages in tax distribution tables obscures the simplest facts about proposed tax policy initiatives to the public.

### **ESSENTIAL QUESTIONS TO ASSIST IN EVALUATING TAX DISTRIBUTION TABLES**

The issues raised in this paper and the following eleven questions will assist taxpayers in reviewing tax distribution tables:

1. Is the median presented as the correct measure of central tendency (or at least provided in addition to the average)?
2. What measure of income is being used (If adjusted gross income (AGI) is not presented, or some other measure that taxpayers understand, ask that it be provided)?
3. What taxes are being included in the analysis in both the before and after columns, and are they identical (i.e., comparing apples to apples)?
4. How many taxpayers reside within the displayed income categories?
5. What is the range of income and tax liability associated with each category?
6. What is the current and proposed (after full enactment of the proposed tax legislation) level of taxation (percent of total taxes paid to the government) paid by each income category?
7. What is the current and proposed (after full enactment of the proposed tax legislation) effective tax rate for each income category?
8. What are the ranges of tax cuts each income group is estimated to receive after full enactment of the tax legislation (ranges and medians should be provided instead of the often-presented average tax cut)?
9. Are the estimates presented free of imputations? If not, what imputations have been made to arrive at the estimates presented in the distributional tax tables?
10. What are the accuracy and reliability of the estimates presented in the distributional tax tables, and are data limitations disclosed or are they hidden?
11. What are some additional or hidden burdens that are not captured in the distributional tax tables (the hidden economic gains or losses resulting from a tax change, e.g., the economic increase in the stock of capital that would result from a repeal of the estate tax or the hidden burden of hiring lawyers and accountants to avoid the estate tax)?

# A GUIDE TO TAX POLICY ANALYSIS: THE CENTRAL TENDENCY OF FEDERAL INCOME TAX LIABILITIES IN DISTRIBUTIONAL ANALYSIS

He uses statistics as a drunken man uses lamp posts – for support rather than illumination.

*Andrew Lang*<sup>1</sup>

[B]efore representing the central tendency by any single number, evaluators need to look at the distribution and decide whether the indicator would be misleading.

*United States General Accounting Office*<sup>2</sup>

## I. INTRODUCTION

The analysis of tax data is a time intensive and complicated process. Much time and effort are spent collecting income and tax data, compiling data sets and running statistical analyses. However, it appears that relatively little time and effort are spent actually *understanding* the data and how best to present results to the public of analyses of tax data. This is evident in the overuse of averages and the simplistic classification of taxpayers into income ranges and quintiles by highly publicized tax distribution tables. This study shows that the link between income and tax liability is much more tenuous than that often presumed, and that a variety of other factors can greatly affect tax liability.

The taxation of individual income is a major focus of tax policy. Legislators evaluating the fundamental components of tax legislation face decisions that often affect after-tax income and wealth of taxpayers and can affect the performance of the greater economy. The presentation of tax data is necessary for the effective understanding and evaluation of tax policy by both legislators and the public. The incorrect use of descriptive statistics can have profound effects on the way tax policies are evaluated.

The official sources of tax distribution data are the Office of Tax Analysis (OTA) of the Department of Treasury, the Congressional Joint Committee on Taxation (JCT)

---

<sup>1</sup> Furman University Mathematical Quotation Server. Available online at: <http://math.furman.edu/~mwoodard/mqs/mquot.shtml>

<sup>2</sup> United States General Accounting Office. *Quantitative Data Analysis: An Introduction*. (GAO/PEMD-10.1.11), June 1992.

and, to a lesser extent, the Congressional Budget Office (CBO).<sup>3</sup> All of these organizations apply different assumptions and methodologies to the analysis of tax legislation. In addition, there are unofficial distribution tables that are publicly released by assorted advocacy groups to influence the policy process and the debate on particular aspects of tax legislation.

Many tax distribution tables released into the public domain, such as those of the Treasury Department and assorted advocacy groups, misrepresent the average as the correct measure of central tendency. Examples of these tables are provided in Appendix I. Not surprisingly, those distribution tables released to advance one point of view are the analyses most likely to misuse averages and to mislead the public. Additionally, all of the disseminators of tax distribution tables use rigid income categories to classify taxpayers that *appear* to be alike. As is commonly said, the devil is in the details.

The rest of this paper is organized as follows. Section II will briefly outline what exactly is a distribution table. Section III will then discuss the appropriate measures used to describe the central tendency of income and tax data. Sections IV and V will describe in detail why the use of averages is an inappropriate measure of central tendency for describing income and tax data, and further describe how the use of averages provides an incomplete picture in tax distribution tables. Federal income tax data from the Internal Revenue Service graphically demonstrate how the use of averages provides an illusion of precision that is false and misleading. Furthermore, these sections will explain why in order to remain impartial, distributional tax tables should *never* display averages as the sole measure of central tendency. Section VI concludes this paper. Appendix I provides examples of tax distribution tables released by the OTA and Citizens For Tax Justice and Appendix II provides a description of the data used in this paper and the limitations associated with the data.

Readers that are not familiar with distributional tax analysis, the presentation and use of distribution tables, the measures of income and methodologies used in distributional analysis are encouraged to reference “**A Guide to Tax Policy Analysis: Problems with Distributional Tax Tables,**” a previous Joint Economic Committee Study. This study also details how taxpayers can effectively evaluate the merits of different presentations used in distributional analysis and is available online at: <http://www.house.gov/jec>

---

<sup>3</sup> For a more detailed discussion of their respective rolls, see: Michael J. Graetz. “Distributional Tables, Tax Legislation, and the Illusion of Precision,” in David F. Bradford, ed. *Distributional Analysis of Tax Policy*. AEI Press. Washington, DC. 1995, page 20.

## II. The Distribution Table

A distribution table can be deceptively simple. Generally, in the left-hand column are income categories classified by either dollar cut-offs, such as, \$0 - \$10,000, \$10,000 - \$20,000, \$20,000 - \$30,000, etc., or divided into percentile groupings such as, lowest quintile, second quintile, third quintile, fourth quintile, and highest quintile. Additional columns provide information about the number of observations, income levels, taxes paid, etc., for each income category. Usually, the table provides information pertaining to the changes in taxes that are to be paid after the proposed tax legislation is enacted. The primary focus of tax analysis is the increases and decreases in taxes paid under current law in comparison to after the proposed tax legislation becomes fully effective. Table 1 provides an illustration of a simple burden table relating to a hypothetical proposal to reduce individual taxes:

Table 1.

Income Category	Change in Federal Taxes		Effective Tax Rate		Average Tax Change
	\$ (millions)	Percent	Present Law	Proposed Law	\$
			Percent	Percent	
Less than \$10,000	-20	-0.2	7.1	7.0	-300
10,000 to 20,000	-365	-1.0	8.1	8.0	-400
20,000 to 30,000	-1,300	-1.5	15.2	15.0	-500
30,000 to 40,000	-2,150	-1.9	17.6	17.3	-750
40,000 to 50,000	-2,750	-2.1	19.3	18.9	-1,100
50,000 to 75,000	-7,200	-2.3	21.2	20.7	-1,500
75,000 to 100,000	-6,600	-2.4	23.9	23.2	-2,000
100,000 to 200,000	-8,100	-2.2	26.2	25.5	-3,500
200,000 and over	-13,500	-3.1	29.2	27.6	-5,000
<b>Total, all taxpayers</b>	<b>-\$41,985</b>	<b>-2.4%</b>	<b>22.2%</b>	<b>21.5%</b>	<b>-\$650</b>

Source: Hypothetical Data. JEC Calculations

In viewing the results displayed in the second column, it is quite clear in this example that all taxpayer groups would receive a nominal reduction in tax. The lowest group receives a total reduction in their tax of \$20 million and the highest group receives a total reduction of \$13.5 billion. The third column shows the reduction in terms of percentages. The lowest group receives a 0.2 percentage reduction in tax, while the highest group receives a 3.1 percentage reduction. The fourth and fifth columns display each group's effective tax rate under present law and after the legislation becomes effective, respectively. All income groups benefit from a lower effective tax rate under the proposed legislation. The last column displays the dollar amount of the average tax cut that each member in an income category might expect to receive.

Since every income group benefits, a cursory review of the above table might lead readers to conclude that the tax proposal is beneficial for all. However, some might come to completely different conclusions. These readers may conclude that the tax legislation is not fair to the lowest income group, since the highest income group receives 32 percent of the total benefit (\$13.5 billion / \$42.0 billion) while the lowest income group receives less than ½ percent of the total benefit (\$20 million / \$42.0 billion). However, the problem with this perspective is that these numbers reflect more about the impact of the

current tax system than the tax change under consideration. In other words, in most cases such statistics primarily reflect the distribution of tax payments under the tax code before the tax change takes place. The more progressive the current tax code is, the more regressive any subsequent tax change can be made to appear. What is presented as a measure of the tax change is in reality a statistical mirage that mainly reflects the progressivity of the current tax system.

Table 1 actually provides insufficient information from which to draw an informed conclusion as to the merits of the proposed tax legislation. For example, this table does not show the current amount of taxes that each income group pays. For purposes of illustration, assume that the lowest income group currently pays no tax at all, while the highest income group pays 50% of the total tax collected. Then, based on a different measure of fairness, it could be argued that the highest income group should receive a commensurate amount of the benefits of the total tax reduction and, therefore, the proposed 32% (\$13.5 billion / \$42.0 billion) is *unfair* to the upper income group.

Additionally, Table 1 does not indicate how many taxpayers make up each income group, although this can be mathematically derived. Additional information is also necessary to effectively evaluate the proposed tax legislation, such as what items are included in income, what types of taxes are being included/excluded, and over what time horizon the effects are being measured.

### III. MEASURES OF CENTRAL TENDENCY

As Yale University law professor and former Treasury Deputy Assistant Secretary for tax policy Michael J. Graetz writes, “[t]he current practice of fashioning tax legislation to achieve a particular result in a distribution table creates the illusion of precision when such precision is impossible.”<sup>4</sup> It is statistically possible, based on averages, that some taxpayers would receive no tax cut or even face a tax increase. Furthermore, not only is precision impossible but the use of averages misrepresents the central tendency of the data.

The central tendency of the distribution of data is a point estimate or single number that corresponds to a typical, representative or middle score for a given set of data. Examples of such measures are the average, the median and the mode.

The average, or mean, is the most easily recognized and understood measure of central tendency. To calculate the average, each observation in the data is added together and then the sum is divided by the total number of observations. Some common uses of averages to describe central tendency are batting averages in baseball and student grade

---

<sup>4</sup> Michael J. Graetz. “Distributional Tables, Tax Legislation, and the Illusion of Precision.” In David F. Bradford (Editor). *Distributional Analysis of Tax Policy*. AEI Press. Washington, DC. 1995.

point averages. The use of averages is simple and easy for people to understand. However, the use of averages may not be appropriate if there are many outliers in the data or the data do not fit the pattern of a normal distribution. This is because the average as a measure of central tendency can be highly influenced by the presence of extreme values.

The median is the middle score in a set of ranked data. It represents the point in the distribution where 50 percent of the observations lie above the value and 50 percent lie below it. The median makes no assumptions about the shape of the distribution of data. Furthermore, the median is considered to be a statistically resistant measure of central tendency because the value associated with a median is not highly affected by outliers that can affect the value associated with an average.

The mode is determined by finding the value that most frequently corresponds to the data set. Simply stated, the mode is the most frequently occurring attribute or observation in a data set and is most commonly used with nominal variables.

When describing the central tendency of data, the measure that should be used is the one that best describes the data. For most income and tax data this is the median value, not the average. To see why this is the case, consider the following example displaying the seven salaries of a company in Table 2.

<b>Table 2. - Annual Income</b>	
CEO	\$1,000,000
Attorney	\$70,000
Systems Administrator	\$60,000
Economist	\$50,000
Office Administrator	\$40,000
Secretary	\$40,000
Paid Intern	\$10,500
Total	\$1,270,500
Average	\$181,500
Median	\$50,000
Mode	\$40,000

The average of these seven salaries is \$181,500. The median value is \$50,000 and the mode is \$40,000. In this instance, and in any situation where extreme outliers can skew the average, the median is a better indicator of the central tendency because the CEO's salary is an extreme outlier causing the average to lie far from the other six salaries. The median is the best single number that represents the central tendency of this data.

To further illustrate, Bill Gates, who has an estimated net worth in the billions of dollars and an unusually high income, resides in the upper most income category of any distributional tax analysis. His income alone would be enough to skew any average

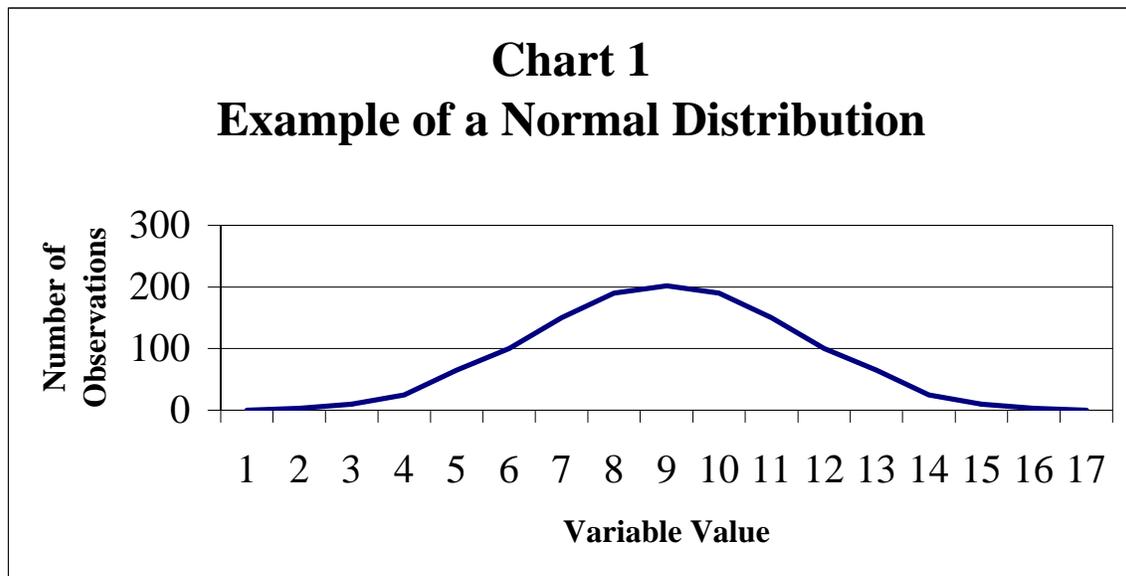
income measure in the upper percentiles. Due to the nature of income data, most official income data released by government and other statistical agencies provide the median as a measure of central tendency or at the very least provide the median along with the average.

The misuse of averages in distribution tables can hide information relating to the dispersion and the true central tendency of the data from the public, further clouding the ability to make sound decisions about tax policy. The severity of the misuse of the average as a measure of central tendency depends on how far the distribution of the data varies from a normal distribution.

#### IV. THE CENTRAL TENDENCY OF TAX DATA

The Internal Revenue Service (IRS) Public Use Tax File, prepared by the Statistics of Income Division (SOI), contains a stratified random sample of tax returns and is used to tabulate and present statistical information representative of the entire population of individual income tax returns filed with the IRS.<sup>5</sup> Using this data and a statistical software package, graphical representations of the distribution of taxpayers' tax liability by income categories becomes possible.

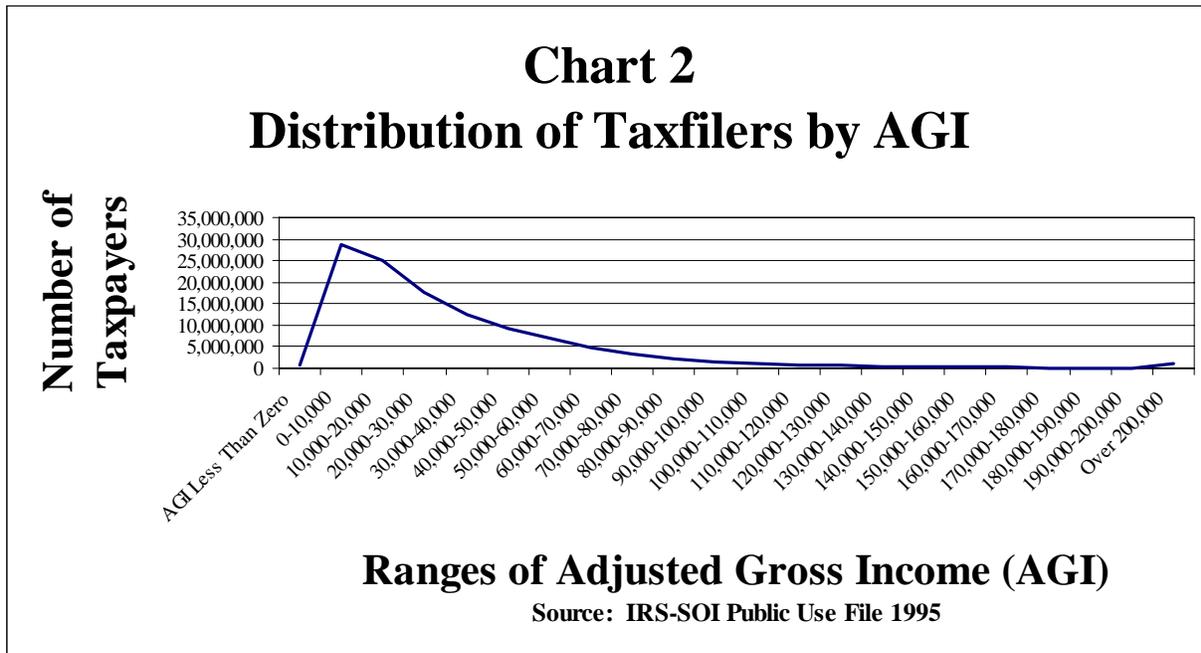
A common graphical way to present the distribution of data is by means of a simple line chart. In this fashion, a normal distribution would take on a shape similar to the following in Chart 1 below.



<sup>5</sup> For a full description of the IRS Public Use File, including sampling error and disclosure avoidance procedures, please see the Appendix II.

With normally distributed data the shape is symmetrical. Furthermore, the three measures of central tendency (average, median and mode) tend to be identical or very close to being identical. In the above example, the average, median and mode are all nine. However, data provided by the IRS show that income and tax data do not follow the pattern of a normal distribution.

For tax year 1995, the most recent public use file available, the distribution of tax returns by adjusted gross income (AGI) looks as follows in Chart 2.<sup>6</sup>

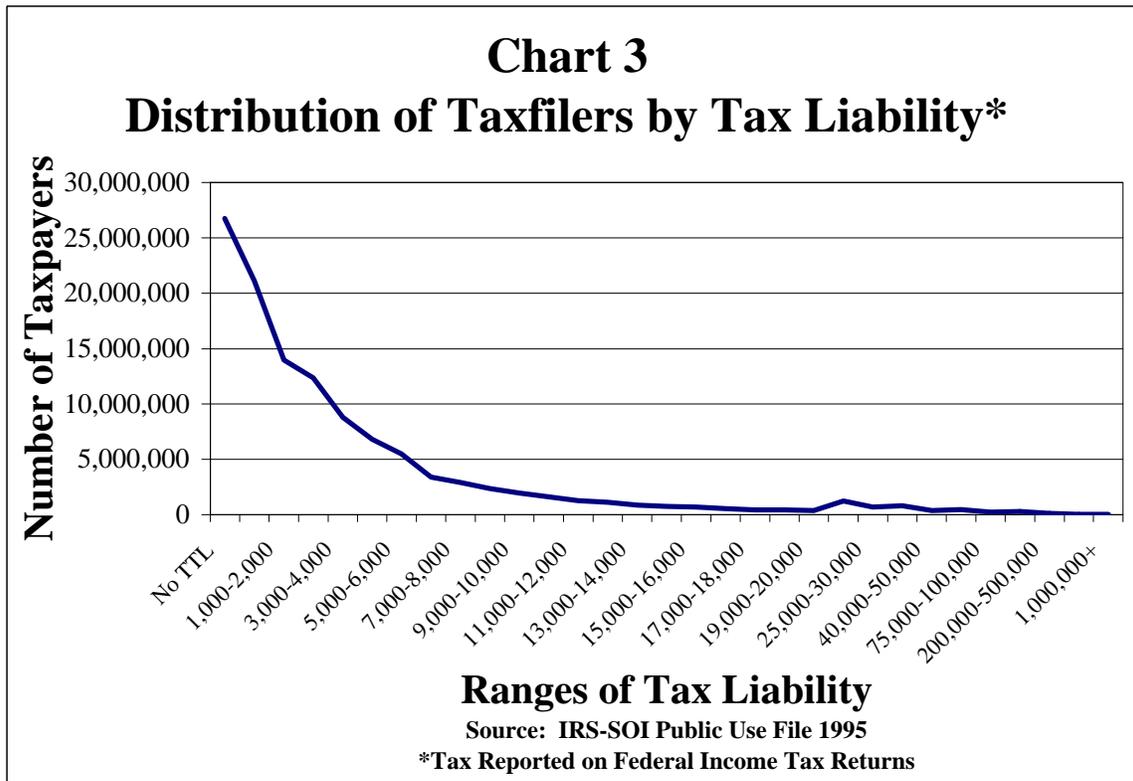


As can be seen, the distribution of tax returns based on AGI is highly asymmetrical. Furthermore, the distribution is highly skewed to the left. Due to the extreme asymmetry of the data, it would be inappropriate to use the average as an appropriate measure of central tendency when describing taxpayers based on AGI.

Chart 3 below displays how the distribution appears if the variable of analysis is federal income tax liability, or the total dollar amount that is paid to the IRS and reported straight off of a federal tax return.<sup>7</sup>

<sup>6</sup> The IRS releases aggregate statistics to the public and publishes these statistics in its “Statistics of Income Bulletin” on a lagged basis. In past years, the public use file has been published yearly on a one-year lag after the end of the filing period. The current increase in the lag has been caused by SOI’s efforts to reexamine the disclosure issues involved with the microdata. The public use files for tax years 1996 – 1998 will hopefully be released starting late this summer or early fall. Furthermore, SOI hopes to have the reexamination of its disclosure policies completed shortly so that the Tax Year 2000 Public Use File will be available in December 2002.

<sup>7</sup> Does not include payroll or excise taxes or any taxes not reported on a federal tax return.



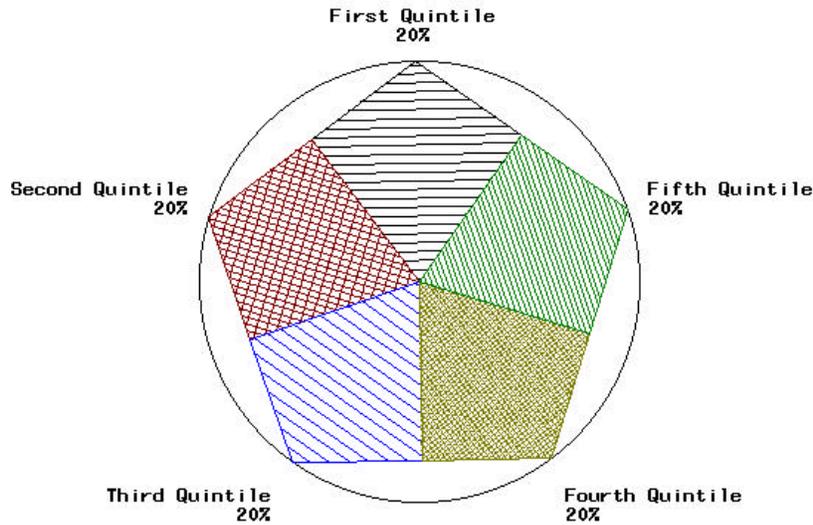
In this case, the distribution is also asymmetrical with the data highly skewed to the left. From the chart, it is observed that over 25 million tax returns have zero tax liability. Hence, any use of an average to describe taxpayers based on tax liability does not accurately represent the central tendency of the population. Furthermore, due to the skewed nature of the data, even the use of the median may not provide an accurate representation of the data.

The use of line charts is a simple way to graphically represent the distribution of data and can be created in spreadsheet software packages. A more complex chart can be used to shed light on the nuances that are often hidden in more simplistic tables. Star charts provide an interesting and novel approach to looking at the distribution of data.

Star charts are graphs created with complex statistical software packages that show statistics based on values of a variable. The center of a star chart represents the value zero. The circle enclosing the star chart represents the maximum statistic value for any one of the predefined groups. Each group value is represented by a slice. The slice with the greatest value extends out to the edge of the circle. The remaining slices are represented as proportions of the slice with the greatest value. The groups can be midpoints, quartiles, quintiles, or any programmed group that an analyst chooses to study.

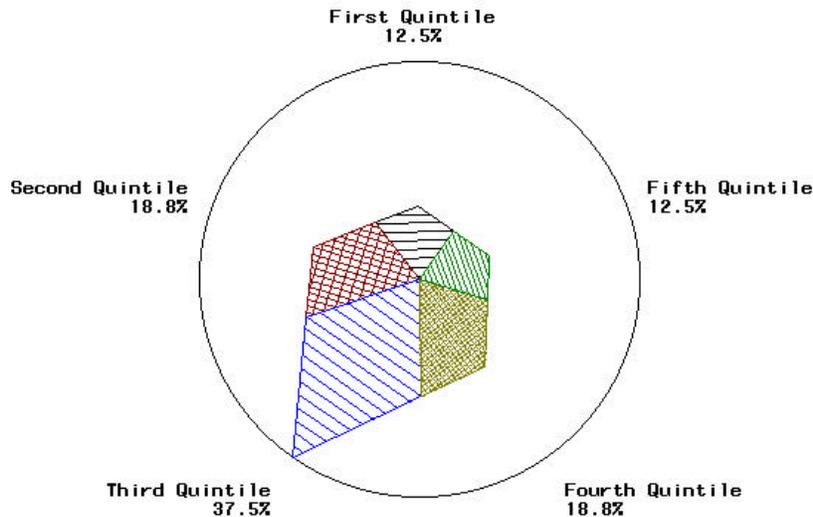
Chart 4 below provides an example of a star chart with an equal distribution. The variable of study has been grouped into quintiles. By definition, a quintile contains one-fifth of the total number of observations in a data set. If the variable under study was federal tax liability and the distribution of federal tax liability was equal for each quintile, this would imply that each quintile has the same number of total dollars as each of the other quintiles. Since each quintile group contains the same amount of total federal tax liability, each slice extends equally out to the edge of the circle.

Chart 4 – Example of An Equal Distribution



However, federal income tax liability doesn't follow an equal distribution. Chart 2 above shows that income is asymmetric and highly skewed to the right. If tax liability were normally distributed and were to follow a pattern such as that displayed in Chart 1, a star chart displaying the distribution of a variable that follows the shape of a normal distribution grouped into quintiles would look like the following example in Chart 5.

Chart 5 – Example of A Normal Distribution



This is how a variable that follows the pattern of a normal distribution displays as a star chart. The third quintile is equivalent to the middle observations that would lie underneath the height of the curve of a normal distribution displayed as a line chart, as in Chart 1 above. Since the third quintile represents the greatest value (37.5%), its slice is the longest and extends to the edge of the circle. Since both the second and fourth quintiles contain half the value as the third quintile (18.75% rounded to 18.8%), their respective slices extend halfway to the edge of the circle. Similarly, the first and fifth quintiles, or the tails of a normal distribution as displayed in Chart 1, contain only one-third the value as the third quintile (12.5%). Hence the slices representing the first and fifth quintiles extend one-third of the way to the edge of the circle. Only if a variable follows the pattern of a normal distribution similar to the pattern displayed above in Chart 5 is it appropriate to use the average as the measure of central tendency.

Tax distribution tables ultimately focus on how much more or less in taxes income groups will pay under a change in tax law. Furthermore, the majority of distribution tables that are released use the average as a measure of central tendency and group taxpayers into quintiles. Therefore, the rest of this paper will focus on federal AGI and tax liability grouped by quintiles. Using the SOI Public Use File, it is possible to calculate the average and median AGI and federal tax liability amounts for each quintile. Table 3 below displays this information for tax year 1995.

<b>Table 3. Estimated Average and Median Amounts Federal AGI and Tax Liability</b> (Rounded to Nearest \$100)		
<b>All Tax Returns</b>	Average	Median
AGI	\$35,300	\$22,100
Tax Liability	\$5,200	\$1,800
<b>First Quintile</b>		
AGI	\$1,600	\$3,700
Tax Liability	\$100	\$0
<b>Second Quintile</b>		
AGI	\$12,200	\$12,100
Tax Liability	\$500	\$400
<b>Third Quintile</b>		
AGI	\$22,400	\$22,100
Tax Liability	\$1,800	\$1,800
<b>Fourth Quintile</b>		
AGI	\$38,700	\$38,000
Tax Liability	\$4,200	\$3,900
<b>Fifth Quintile</b>		
AGI	\$101,300	\$71,600
Tax Liability	\$19,100	\$10,100

Detail May Not Add Due To Rounding.

The average and median values show some interesting contrasts in Table 3. For all tax returns, the average AGI amount is almost 60 percent more than the median. The contrast is even greater focusing on tax liability, the average of which is 189 percent greater than the median! Since the average and median are so far apart, it is obvious that the distribution of AGI and tax liability among all tax returns does not follow the pattern of a normal distribution. Hence, the average should not be used as the sole measure of central tendency.

Contradictory observations are further made focusing on the quintile levels. Focusing on tax liability, the averages and medians for the second and third quintiles are relatively close. However, the opposite is the case for the first and fifth quintiles. In the first quintile, the average tax liability is \$100 (rounded up) and the median is \$0 (this value wasn't rounded). This means that at least 50 percent of the tax returns in the bottom quintile have zero or negative tax liability. In this instance, the median is the best representative measure of central tendency.

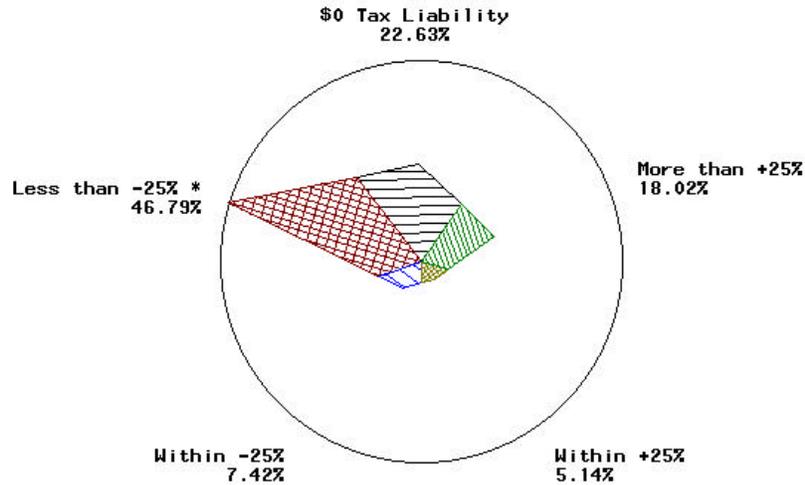
In fact, as will be demonstrated later in the paper, there are tax returns in each quintile that have zero tax liability. A study by the Congressional Joint Committee on Taxation (JCT) calculates that roughly 48.7 million taxpayers (including those taxpayers that don't file a federal income tax return) have zero or negative tax liability in calendar year 2000.<sup>8</sup> This is equivalent to 34.7 percent of the JCT's estimated number of tax units, including filing and non-filing units and *excluding* individuals who are dependents of other taxpayers and taxpayers with negative income. If these taxpayers were included in the JCT analysis, the number and percentage of taxpayers who have zero or negative tax liability would be substantially higher. This further supports using the median as the most representative measure of central tendency when describing income and tax liability amounts.

But how do the distributions of tax returns by quintile compare to that of a normal distribution? Again, Chart 5 above presented a star chart for a normally distributed variable. In order to use star charts to show the distribution of tax returns by quintile, it is necessary to define some groupings. For purposes of this analysis each quintile has been grouped further into five categories: (1) tax returns having zero tax liability; (2) returns having tax liabilities greater than zero and that are between the average amount for that quintile and the amount which is less than 25% greater than the average; (3) returns having tax liabilities that are between the average amount for that quintile and the amount which is less than 25% less than the average; (4) returns having tax liabilities greater than that amount which is 25% more than the average; and (5) returns having tax liabilities less than the amount which is 25% less than the average.

---

<sup>8</sup> United States Congress. Joint Committee on Taxation. "Distribution of Certain Federal Tax Liabilities by Income Class for Calendar Year 2000." JCX-45-00. April 11, 2000.

Chart 6 — Federal Income Tax Liability \*\*



\* For Federal Income Tax Liability Greater Than Zero  
 \*\* Compared To Average Income Tax Liability

Before turning to an analysis of quintiles, the national distribution of tax returns based on tax liability for all tax returns using the groupings defined above is displayed in Chart 6.

For tax year 1995, over 22 percent of all tax returns have no tax liability. This amounts to 26.8 million tax returns. This figure is less than the 48.7 million taxpayers identified in calendar year 2000 by the JCT.<sup>9</sup> This discrepancy is in part based on the different years under analysis and that the unit of analysis in the 1995 data is tax returns while the JCT's unit of analysis is taxpayers.

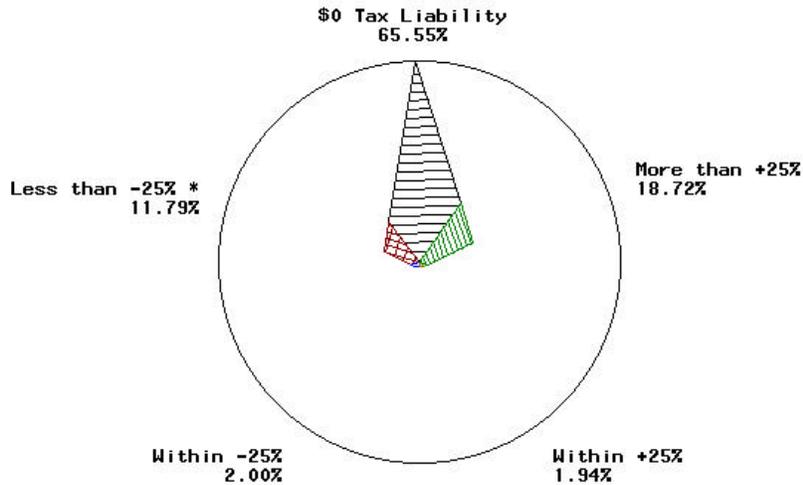
Furthermore, almost 47 percent of all returns have tax liability amounts falling between zero and 25 percent less than the average of \$5,200. If these tax returns are combined with those with zero tax liability, then over 69 percent (22.63% + 46.79%) of all returns pay less than the average tax liability. Lastly, about 12 percent of all returns have tax liabilities that are within +/- 25 percent of the average tax liability amount. In other words, and perhaps most notably, almost 88 percent of all returns have tax liabilities that are either 25 percent greater than the average or 25 percent less than the average.

Based on this information, the use of the average as the sole measure of central tendency to describe the tax liability for the entire country would be misleading. The use of the average suggests that the "representative" taxpayer has a tax liability of \$5,200, almost three times greater than the median amount.

<sup>9</sup> *Ibid.*

Chart 7 below represents the distribution of tax returns based on tax liability for the first quintile using the groupings defined above.

Chart 7 — Federal Income Tax Liability for First Quintile \*\*



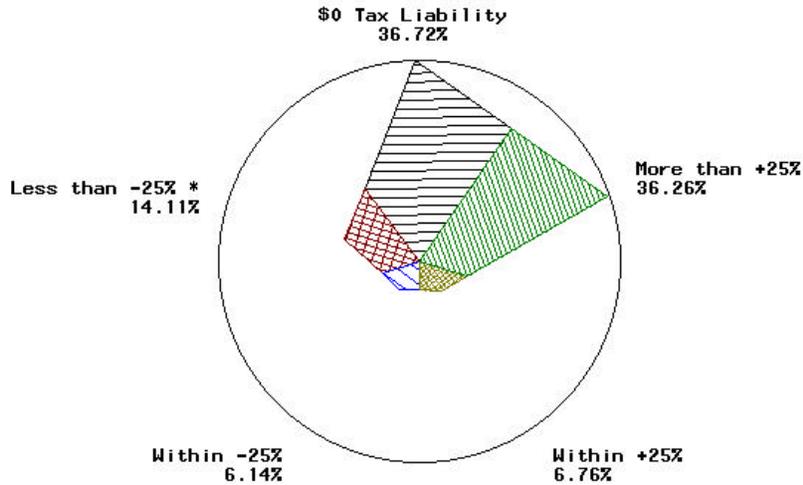
\* For Federal Income Tax Liability Greater Than Zero  
 \*\* Compared To Average Income Tax Liability

Notice that over 65 percent of the returns in the first quintile have no income tax liability. This means that over 65 percent of the returns in this quintile have more in common with the median (\$0) than with the average (\$100). Furthermore, only about 4 percent of the returns in the first quintile have tax liabilities that are within +/- 25 percent of the average tax liability amount for the first quintile of \$100. This means that over 96 percent of all returns in the first quintile have tax liabilities that are either 25 percent greater than the average or 25 percent less than the average.

It would appear that the median is definitely a more representative measure of central tendency in the first quintile than the average. The use of the average in this case misleads the reader into believing that more people in this quintile have positive tax liability than those that have zero tax liability.

A similar picture emerges for the second quintile, as Chart 8 shows. Just over 36 percent of tax returns in this quintile have zero tax liability. Also, under 13 percent of the tax returns have tax liability within +/- 25 percent of the average (\$500). In other words, over 87 percent of all returns in the second quintile have tax liabilities that are either 25 percent greater than the average or 25 percent less than the average.

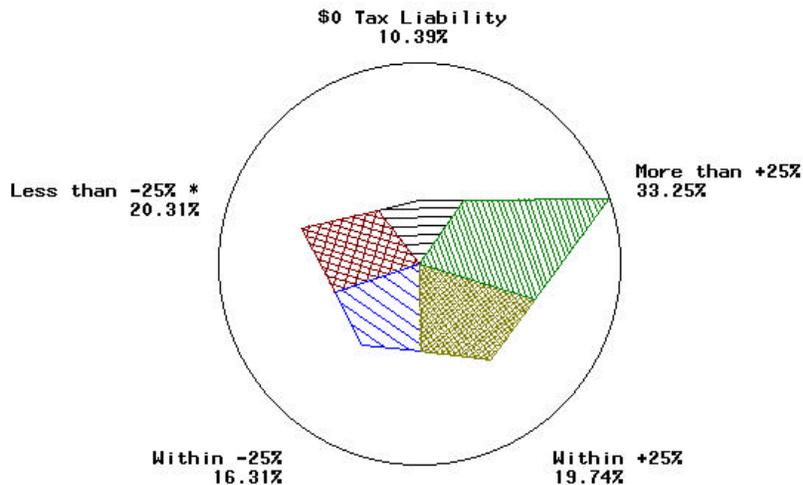
Chart 8 — Federal Income Tax Liability for Second Quintile \*\*



\* For Federal Income Tax Liability Greater Than Zero  
 \*\* Compared To Average Income Tax Liability

The third quintile, in which the average and median are similar, displays a more normal pattern as Chart 9 displays.

Chart 9 — Federal Income Tax Liability for Third Quintile \*\*

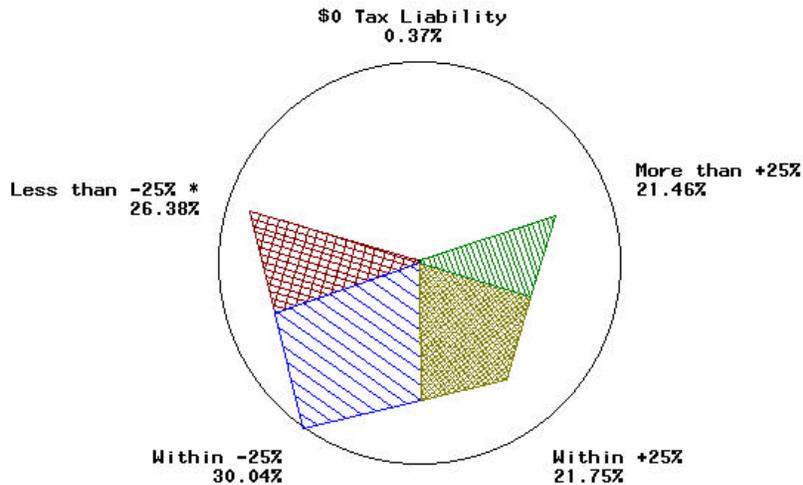


\* For Federal Income Tax Liability Greater Than Zero  
 \*\* Compared To Average Income Tax Liability

Ten percent of returns in this quintile have zero tax liability (10% of returns with AGI between \$16,700 and \$29,000). Thirty-six percent of tax returns have tax liability amounts between +/- 25 percent of the average (\$1,800). However, the overwhelming majority of tax filers in the third quintile (almost 64%) have tax liabilities that are either 25 percent greater than the average or 25 percent less than the average.

The fourth quintile is similar in distribution to the third, with less than 1 percent of returns showing zero tax liability and just over 50 percent of returns having tax liability amounts within +/- 25 percent of the average (\$4,200). The fourth quintile is the most “normal” of the quintiles, as can be seen from Chart 10 below. However, nearly half of the tax filers in the fourth quintile have tax liabilities that are either 25 percent greater than the average or 25 percent less than the average.<sup>10</sup>

Chart 10 — Federal Income Tax Liability for Fourth Quintile \*\*

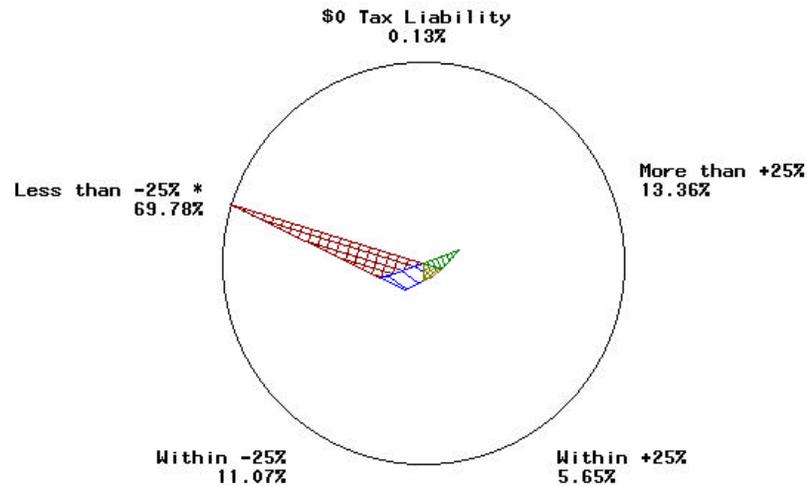


\* For Federal Income Tax Liability Greater Than Zero  
 \*\* Compared To Average Income Tax Liability

<sup>10</sup> However, almost 60 percent (57.37%) of the tax filers in the fourth quintile have tax liabilities that are either 20 percent greater than the average or 20 percent less than the average.

The fifth quintile is as non-normal as the first quintile, as Chart 11 demonstrates below. A most interesting statistic is that almost 70 percent of the returns in the fifth quintile report a tax liability amount that is *less* than 25 percent of the average. As discussed earlier, this demonstrates how a few high-income earners can have a tremendous effect on the average. Because of this, again the median is the more appropriate measure of central tendency. To report only the average would mislead the reader into believing that one-fifth of all tax returns have tax liabilities that are similar to the average amount for the fifth quintile of \$19,100 instead of the median value of \$10,100. The average tax liability amount for the fifth quintile is almost double the median value!

Chart 11 — Federal Income Tax Liability for Fifth Quintile \*\*



\* For Federal Income Tax Liability Greater Than Zero  
 \*\* Compared To Average Income Tax Liability

Therefore, using the average as the measure of central tendency when analyzing or discussing tax policy initiatives is quite misleading. The over-reliance on averages has the effect of making it appear that tax plans that aim to reduce income tax burdens overstate the benefits to the taxpayers in the upper income categories, whereas what is primarily reflected is their higher tax burden before the tax change takes effect. Additionally, even the use of the median can be misleading due to the significant dispersion of tax liability among taxpayers. However, the use of the median is less misleading than the use of the average.

The use of averages when displaying distribution data for income and tax liability misleads the public. This clouds the transparency necessary for the public to effectively evaluate the merits of any proposed tax plan. But this is only part of the story. Not only is the use of averages as a measure of central tendency misleading, but so is the use of quintiles or income categories based on AGI or any other measure of income. These

arbitrary categories imply that the taxpayers grouped into these categories are necessarily similar in economic status and pay similar taxes. This is far from the case.

## V. Misclassification of Taxpayers

It is well known to most taxpayers that tax liabilities often differ among families with the same income. This can be because of family size, filing status, whether a family itemizes their deductions or elects to take the standard deduction, whether a family pays a mortgage on their home and deducts the interest expense or rents, the nature of a family's income and many other factors. Additionally, some families are more aggressive at reducing their tax liabilities than others. For example, this can be done legally by contributing to a 401(k) plan, an individual retirement account or a medical savings account, and in many other ways as well.

The dispersion of taxpayers within any income group is impossible to determine from the information typically presented in tax distribution tables. Do most of the taxpayers within the \$20,000 to \$30,000 income range lie closer to \$20,000 or to \$30,000? All other things being equal, and from the information presented in most distribution tables, it would be expected that a taxpayer with income closer to \$30,000 would necessarily have a higher tax liability, and consequently pay a greater amount in taxes than a taxpayer with income closer to \$20,000. But this is not necessarily the case as Table 4 below begins to illuminate.

All Tax Returns	Average	Median	Minimum Amount	Maximum Amount
AGI	\$35,300	\$22,100	(\$241,700,000)	\$209,400,000
Tax Liability	\$5,200	\$1,800	\$0	\$62,560,000
<b>First Quintile</b>				
AGI	\$1,600	\$3,700	(\$241,700,000)	\$7,900
Tax Liability	\$100	\$0	\$0	\$3,764,000
<b>Second Quintile</b>				
AGI	\$12,200	\$12,100	\$7,900	\$16,700
Tax Liability	\$500	\$400	\$0	\$58,700
<b>Third Quintile</b>				
AGI	\$22,400	\$22,100	\$16,700	\$29,000
Tax Liability	\$1,800	\$1,800	\$0	\$168,300
<b>Fourth Quintile</b>				
AGI	\$38,700	\$38,000	\$29,000	\$50,700
Tax Liability	\$4,200	\$3,900	\$0	\$529,900
<b>Fifth Quintile</b>				
AGI	\$101,300	\$71,600	\$50,700	\$209,400,000
Tax Liability	\$19,100	\$10,100	\$0	\$62,560,000

Detail May Not Add Due To Rounding.

Although over 65 percent of returns in the first quintile and over 36 percent of returns in the second quintile reported zero tax liability (as shown in Charts 7 and 8 above), Table 4 shows that there are actually taxpayers in each quintile that reported zero tax liability on their federal tax returns in 1995. However, the grouping of taxpayers by income measures into quintiles suggests that there are close similarities among these taxpayers with respect to the amount of federal tax liability. The suggested correlation that higher income taxpayers always have higher tax liabilities is not necessarily the case. As Table 4 also illuminates, the maximum tax liability reported on a return classified in the second quintile was \$58,700. However, the maximum tax liability reported on a return classified in the first quintile was over 3 million dollars, \$3,764,000. It seems counterintuitive that a taxpayer ranked and classified in a lower income category can pay more in taxes than a taxpayer ranked and classified in a higher category. This is possible because millions of taxpayers have more in common with each other based on tax liability than based on income. This important fact is ignored in typical tax distribution tables.

It could be suggested that the case highlighted above is only that of an outlier and should be discarded from the sample. Not only would discarding this observation fail to highlight extreme cases in our tax system, but it would also fail to enlighten the public that taxpayer misclassification is actually a problem involving millions of taxpayers, not just a few extreme cases. Chart 12 below begins to illuminate the problem and false sense of precision of classifying taxpayers by income categories.

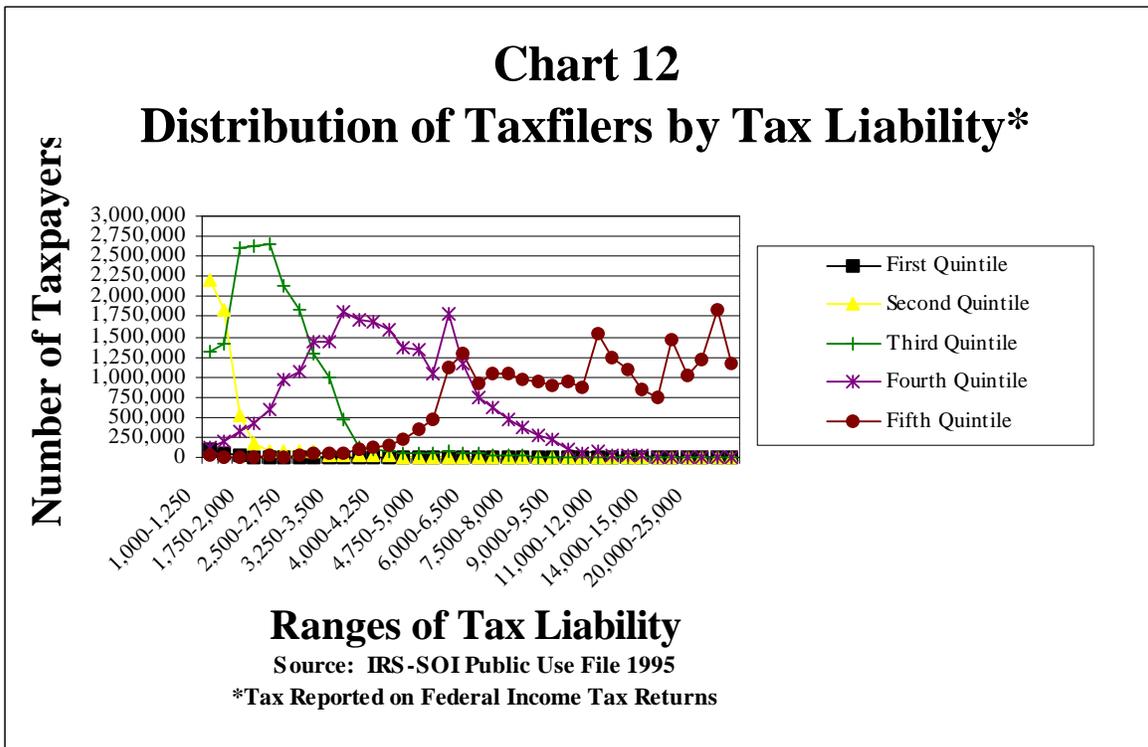


Chart 12 focuses on all tax returns that paid over \$1,000 in federal income tax in 1995, ranked by AGI and grouped into quintiles. As the chart shows, there are millions of taxpayers in the third quintile who pay more in taxes than millions of taxpayers in the fourth quintile. Similarly, there are millions of taxpayers in the fourth quintile who pay more in taxes than millions of taxpayers in the fifth quintile.

Based on Chart 12, Chart 13 below shows that there are 2.2 million tax returns in the third quintile that paid \$3,000 or more in federal income taxes, compared with 5.4 million tax returns in the fourth quintile that paid less than \$3,000, even though these taxpayers are in a higher income quintile.

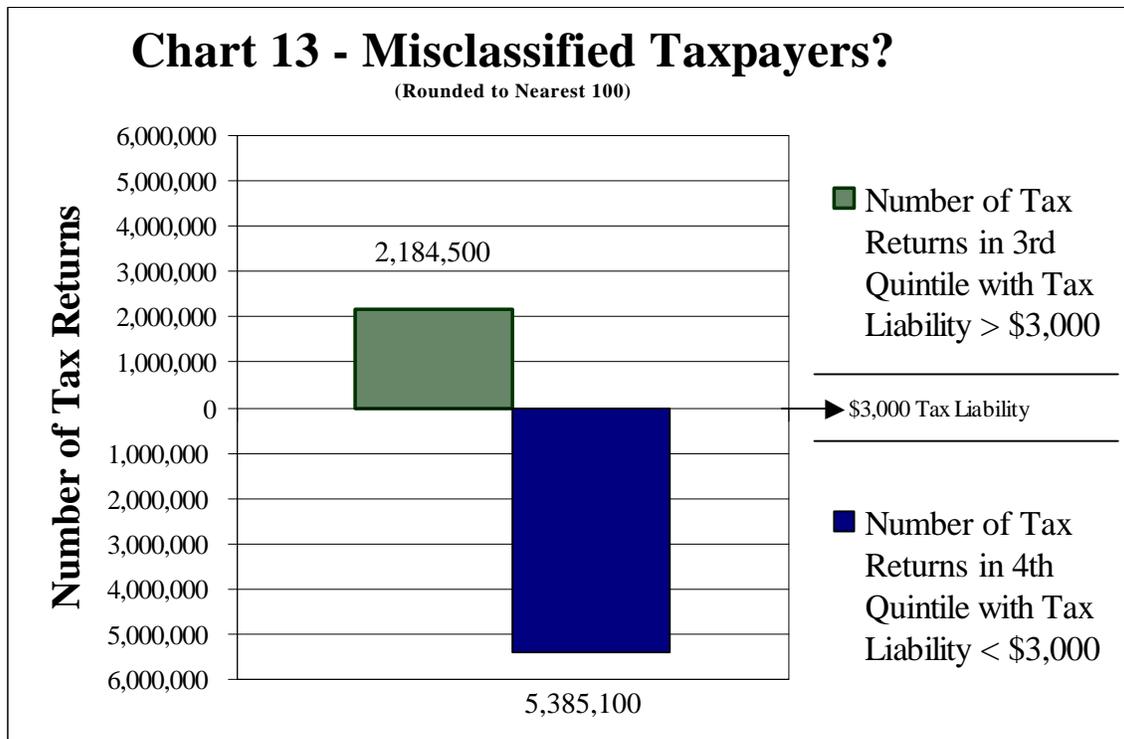
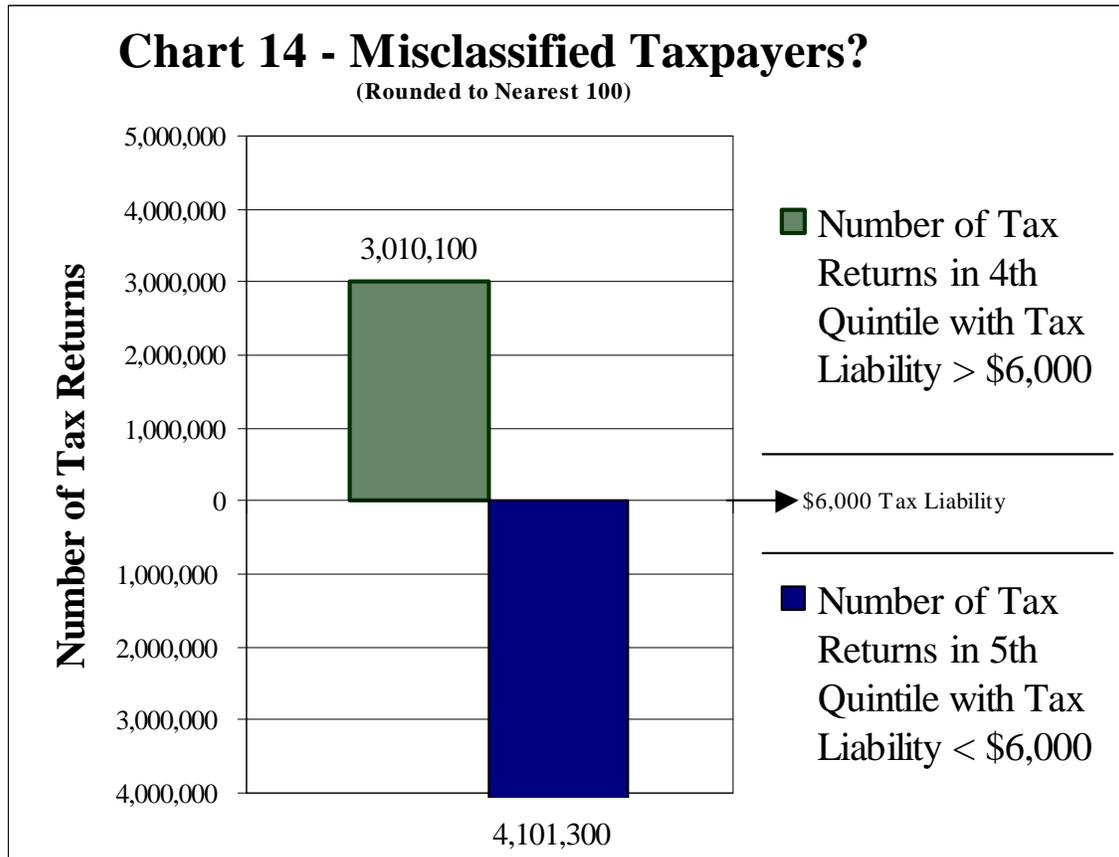


Chart 14 below sheds light on a similar story between the fourth and fifth quintiles. Even though they are in a lower income quintile, 3 million tax returns in the fourth quintile paid over \$6,000 in federal income tax in 1995, compared with 4.1 million tax returns in the fifth and “richest” quintile that paid less than \$6,000.



For tax year 1995, there were roughly 118 million federal tax returns. This amounts to about 23.6 million tax returns per quintile. Chart 13 above suggests that based on tax liability, 5.4 million taxpayers in the fourth quintile have more in common with 21.4 million taxpayers in the third quintile than they do with the other members of the fourth quintile. Similarly, Chart 14 suggests that 4.1 million taxpayers in the fifth quintile have more in common with 20.3 million taxpayers in the fourth quintile than they do with the rest of the 19 million taxpayers in their own quintile.

Ultimately, since tax distribution tables are concerned with the amount of tax currently paid and the amount of tax that is to be paid after a proposed tax legislation is enacted, it is questionable whether policy makers and the public are best served by classifying taxpayers into rigid income categories. This is especially the case when, based on income measures alone, millions of taxpayers have less in common with taxpayers of their own income categories because the amount of tax they pay is more

similar to taxpayers in other income categories. Along with the use of averages, the use of income categories without detailed descriptive language detailing their limitations misleads the public by suggesting that the numbers detailed in tax distribution tables are accurate, precise and reflect an accurate picture of the American taxpaying population.

## VI. CONCLUSION

A former Treasury Deputy Assistant Secretary for Tax Policy, Michael J. Graetz, argues that due to the current opaque nature of communicating even the simplest facts about tax policy to the American public, distributional tax tables should be abandoned as a basis for legislative decision-making.<sup>11</sup> The statistical evidence demonstrates that the process, development, presentation and release of tax distribution tables need fundamental reform.

Lastly, tax changes can alter the after-tax prices and costs of goods and services, thereby adjusting the relative mix of inputs used in production, the types of goods and services businesses offer, as well as the amount of labor and capital. Tax changes can also alter the growth path of the economy and can produce broad economic effects that are not reflected in distributional analyses. Therefore, attempts to ascertain the distributional impact of proposed tax legislation should consider the possible macroeconomic effects. Furthermore, if distributional analysis is used, it should be in a much broader context in which the effects on efficiency and the economy are fully considered.

This paper has demonstrated how the use of averages and income classifications in tax distribution tables can mislead the public. This has the effect of supporting arguments based on class conflict paradigms and fails to illuminate the public as to the nuances of the actual distribution of tax liability across the income spectrum. Unless there is greater public recognition of the improper use of averages with income and tax data and the problems associated with using broad sweeping income categories to group “like” taxpayers, the current practice of using tax distribution tables will continue to mislead the public. At the very minimum, the use of the median as a more appropriate measure of central tendency will help to illuminate the public and contribute to a more open and honest tax policy debate

Specifically, this report finds:

- Income and tax information based on tax returns filed with the IRS do not follow the pattern of a normal distribution. Hence, the use of averages is an inappropriate measure of central tendency.

---

<sup>11</sup> Michael J. Graetz. “Distributional Tables, Tax Legislation, and the Illusion of Precision.” In David F. Bradford (Editor). *Distributional Analysis of Tax Policy*, pages 75 and 76.

- Over 22 percent of all 1995 tax returns claimed zero tax liability.
- The Joint Committee on Taxation estimates that for calendar year 2000, 48.7 million taxpayers out of 140.2 million taxpayers overall, or 34.7 percent, will have zero or negative federal income tax liability.
- For all taxpayers, the use of the average as the measure of central tendency overstates the tax liability for the “representative” taxpayer by almost 3 times the median value.
- The dispersion of taxpayers within any income group is impossible to determine from the information presented in tax distribution tables, but is shown to vary considerably.
- The grouping of taxpayers into income categories provide a false sense of precision and misleadingly suggest that taxpayers within the same groups necessarily have similar federal income tax liability.
- In four out of five income groups examined, a majority of taxpayers had tax liabilities that were either 25 percent greater than the average or 25 percent less than the average tax liability for each income group.
- In comparing federal income tax liabilities, distribution tables often misclassify millions of taxpayers into quintiles in which they have little tax liability in common.
  - Approximately 2.2 million taxpayers in the third quintile pay more in federal income taxes than 5.4 million taxpayers classified in the fourth quintile.
  - Over 3 million taxpayers in the fourth quintile pay more in federal income taxes than 4.1 million taxpayers classified in the fifth quintile.
- The use of averages in tax distribution tables obscures the simplest facts about proposed tax policy initiatives to the public.

In addition to the use of averages (or the omission of the median as a measure of central tendency), tax distribution tables can mislead the public in other areas as well. The points made in this paper and the following 11 questions will assist taxpayers in reviewing distribution tables of proposed tax legislation. If citizens evaluating the merits of tax distribution tables are unable to determine the answers to the following 11 questions, more information should be requested from the authoring agency or organization. Only with the answers to all of the following questions can taxpayers make informed decisions about the merits of tax proposals.

1. Is the median presented as the correct measure of central tendency (or at least provided in addition to the average)?
2. What measure of income is being used (If adjusted gross income (AGI) is not presented, or some other measure that taxpayers understand, ask that it be provided)?

3. What taxes are being included in the analysis in both the before and after columns, and are they identical (i.e., comparing apples to apples)?
4. How many taxpayers reside within the displayed income categories?
5. What is the range of income and tax liability associated with each category?
6. What is the current and proposed (after full enactment of the proposed tax legislation) level of taxation (percent of total taxes paid to the government) paid by each income category?
7. What is the current and proposed (after full enactment of the proposed tax legislation) effective tax rate for each income category?
8. What are the ranges of tax cuts each income group is estimated to receive after full enactment of the tax legislation (ranges and medians should be provided instead of the often-presented average tax cut)?
9. Are the estimates presented free of imputations? If not, what imputations have been made to arrive at the estimates presented in the distributional tax tables?
10. What are the accuracy and reliability of the estimates presented in the distributional tax tables, and are data limitations disclosed or are they hidden?
11. What are some additional or hidden burdens that are not captured in the distributional tax tables (the hidden economic gains or losses resulting from a tax change, e.g., the economic increase in the stock of capital that would result from a repeal of the estate tax or the hidden burden of hiring lawyers and accountants to avoid the estate tax)?

Using the answers to these 11 questions, taxpayers will be able to unveil the information that is not always contained in tax distribution tables and evaluate the economic merits of proposed tax legislation. Distributional tax tables that are presented in such a manner that withhold or omit the answers to these questions, misuse the average as the sole measure of central tendency, or are based on statistically compromised data sources, should seriously be questioned on the issues of transparency, accuracy and reliability.

This is another paper in a Joint Economic Committee series on distributional tax analysis. For more information and details on how taxpayers can effectively evaluate the merits of different presentations used in distributional analysis, see the previous paper in the series, “**A Guide to Tax Policy Analysis: Problems with Distributional Tax Tables,**” is available online at: <http://www.house.gov/jec>

Jason J. Fichtner  
Senior Economist

## APPENDIX I - TABLE I

### Major Tax Cut Provisions in the Senate Finance Committee Chairman's Mark <sup>1</sup>

(1998 Income Levels)

Family Economic Income Quintile (2)	Number of Families (millions)	Average Tax Change (\$)	Total Tax Change		Tax Change as a Percent of:	
			Amount (3) (\$M)	Percent Distribution (%)	Current Federal Taxes (4) (%)	Family Economic Income (%)
Lowest (5)	21.5	-12	-264	0.4	-2.10	-0.13
Second	22.2	-64	-1428	2.3	-2.32	-0.26
Third	22.3	-274	-5095	10.0	-3.86	-0.64
Fourth	22.3	-583	-12964	21.3	-4.20	-0.81
Highest	22.3	-1789	-39837	65.5	-4.38	-0.97
Total (5)	111.3	-547	-60836	100.0	-4.19	-0.82
Top 10%	11.1	-2338	-26036	42.8	-3.93	-0.89
Top 5%	5.6	-3137	-17489	28.7	-3.58	-0.83
Top 1%	1.1	-7081	-7945	13.1	-3.06	-0.75

**Source: Department of the Treasury – Office of Tax Analysis. June 16, 1997.**

(1) This table distributes the estimated change in tax burdens due to the major tax cut proposals in the Senate Finance Committee Chairman Mark which include the following: i) a child credit; ii) a modified HOPE scholarship tax credit; iii) a deduction for student loan interest; iv) deduction for education expenses paid through State-sponsored prepaid tuition programs; v) permanent extension of Section 127; vi) education investment accounts and private prepaid tuition programs; vii) expanded front-loaded and new back-loaded IRAs; viii) Capital gains provision (lower individual rates, extension of S. 1202, and \$500,000 exclusion for gains on a principal residence; and ix) changes in the individual AMT.

(2) Family Economic Income (FEI) is a broad-based income concept. FEI is constructed by adding to AGI unreported and under-reported income; IRA and Keogh deductions; nontaxable transfer payments such as Social Security and AFDC; employer-provided fringe benefits; inside build-up on pensions, IRAs, Keoghs, and life insurance; tax-exempt interest; and imputed rent on owner-occupied housing. Capital gains are computed on an accrual basis, adjusted for inflation to the extent that reliable data allow. Inflationary losses of lenders are subtracted and gains of borrowers are added. There is also an adjustment for accelerated depreciation of noncorporate businesses. FEI is shown on a family rather than a tax-return basis. The economic incomes of all members of a family unit are added to arrive at the family's economic income used in the distributions.

(3) The change in Federal taxes is estimated at 1998 income levels but assuming fully phased in (2007) law and behavior. For the IRA provisions and education accounts, the change is measured as the present value of the tax savings from one year's contributions. The effect of the capital gains provision is based on the level of capital gains realizations under current law.

(4) The taxes included are individual and corporate income, payroll (Social Security and unemployment), and excises. Estate and gift taxes and customs duties are excluded. The individual income tax is assumed to be borne by payors, the corporate income tax by capital income generally, payroll taxes (employer and employee shares) by labor (wages and self-employment income), excises on purchases by individuals by the purchaser, and excises on purchases by business in proportion to total consumption expenditures. Federal taxes are estimated at 1998 income levels but assuming 2007 law and, therefore, exclude provision that expire prior to the end of the Budget period and are adjusted for the effects of unindexed parameters.

(5) Families with negative incomes are excluded from the lowest quintile but included in the total line.

NOTE: Quintiles begin at FEI of: Second \$16,950; Third \$32,583; Fourth \$54,758; Highest \$93,222; top 10% \$127,373; Top 5% \$170,103; top 1% \$408,551.

<b>Does the table show the answers to the following 11 essential questions?</b>	<b>Yes</b>	<b>No</b>
1. Is the median presented as the correct measure of central tendency?		X
2. What measure of income is used?	X	
3. What taxes are included?	X	
4. How many taxpayers are in each income category?	X	
5. What income range is associated with each income category?		X
6. What are the current and proposed levels of taxation for each category?		X
7. What are the current and proposed effective tax rates for each category?		X
8. What are the estimated ranges of tax cuts for each category?		X
9. Are the estimates presented free of imputations?		X
10. Are measures of error provided relating to the precision, accuracy and reliability?		X
11. Do the estimates provided account for hidden burdens?		X

The FEI concept is used in this analysis, and families with negative incomes are excluded from the lowest quintile, biasing the analysis. Furthermore, this Treasury table excludes information relating to the percentage *change* in after after-tax income, which is considered by the Treasury Department to be the most important piece of information to include in a distributional tax table. As one of the Office of Tax Analysis' own economists writes:

The only tax burden measure with some theoretical basis is the percentage change in after-tax income. It alone provides some indication of a family's change in welfare, because after-tax income represents the family's consumption possibilities in either the current or future years. In contrast, the share of the total change in tax burdens, which is often quoted in the popular press, does not convey information on a family's initial welfare position.<sup>12</sup>

The opaque nature of the exclusion of this information prevents citizens from having an informed debate regarding the "fairness" of the tax proposal under analysis.

<sup>12</sup> Julie-Anne Cronin. "U.S. Treasury Distributional Analysis Methodology." Office of Tax Analysis. Department of Tax Analysis. OTA Paper 85. September 1999. Page 34.

## APPENDIX I – TABLE II

### Effects of the House GOP Tax Plan

Income Group	Income Range	Average Income	Tax Cut (billions)	Average Tax Cut	% of Total Tax Cut
Lowest 20%	Less than \$13,300	\$8,400	\$-0.7	\$-29	0.5%
Second 20%	\$13,300 – 23,800	18,300	-3.6	-144	2.4%
Middle 20%	23,800 – 38,200	30,300	-8.9	-350	5.8%
Fourth 20%	38,200 – 62,800	49,100	-18.1	-712	11.8%
Next 15%	62,800 – 124,000	83,600	-28.8	-1,513	18.8%
Next 4%	124,000 – 301,000	173,000	-24.7	-4,866	16.1%
Top 1%	301,000 or more	837,000	-68.3	-54,027	44.6%
<b>ALL</b>		<b>\$48,700</b>	<b>\$-153.1</b>	<b>\$-1,199</b>	<b>100.0%</b>
<b>Addendum</b>					
Bottom 60%	Less than \$38,200	\$19,000	\$-13.3	\$-174	8.7%
Top 10%	\$89,000 or more	204,000	-105.8	-8,355	69.1%

Source: Citizens for Tax Justice. “House GOP Tax Plan: The Rich Get Richer.” July 27, 1999

Notes: Figures show the annual effects of (1) a 10% cut in personal income tax rates; (2) a reduction in the income tax rates on realized capital gains, from 20% to 15% (for those in all but the bottom regular tax bracket) and from 10% to 7.5% (for those in the bottom regular tax bracket); (3) elimination of the estate tax; (4) repeal of the individual Alternative Minimum Tax; (5) a \$200 interest and dividend exclusion (\$400 for couples); (6) an increase in the standard deduction for couples to double the single amount; (7) increased contribution and benefit limits for pensions and 401(k)s; (8) deductions for health insurance for people without employer plans; and (9) various corporate tax breaks. Not included are about \$3 billion a year in miscellaneous tax breaks, mostly for certain health and education expenses. All figures are at 1999 levels, showing full-year effects after phase-ins are completed.

Does the table show the answers to the following 11 essential questions?	Yes	No
1. Is the median presented as the correct measure of central tendency?		X
2. What measure of income is used?		X
3. What taxes are included?	X	
4. How many taxpayers are in each income category?		X
5. What income range is associated with each income category?	X	
6. What are the current and proposed levels of taxation for each category?		X
7. What are the current and proposed effective tax rates for each category?		X
8. What are the estimated ranges of tax cuts for each category?		X
9. Are the estimates presented free of imputations?		X
10. Are measures of error provided relating to the precision, accuracy and reliability?		X
11. Do the estimates provided account for hidden burdens?		X

The CTJ table misuses the average as the appropriate measure of central tendency, provides no detail as to the income measure used and whether taxpayers with negative incomes are excluded from the lowest income category, nor does it identify whether “taxpayers” who don’t file tax returns are included in the analysis. As the checklist above details, the lack of transparency and the exclusion of essential information from the CTJ distributional tax table, as is the case with many of the distributional tax tables released by the CTJ, only serves to bias the reader towards the preconceived notions of the CTJ.

## APPENDIX II

### 1995 STATISTICS OF INCOME PUBLIC USE TAX FILE

“The Internal Revenue Service 1995 Public Use Tax File, which contains 103,117 records, was selected as part of the Statistics of Income program that was designed to tabulate and present statistical information for the 118.2 million Form 1040, Form 1040A, and Form 1040EZ Federal Individual Income Tax Returns filed for Tax Year 1995.

The Tax Files which have been produced since 1960, consist of detailed information taken from SOI sample records. The public use versions of these sample files are sold in an unidentifiable form, with names, Social Security Numbers (SSN), and other similar information omitted. The primary uses made of these files have been to simulate the administrative and revenue impact of tax law changes, as well as to provide general statistical tabulations relating to sources of income and taxes paid by individuals.”<sup>13</sup>

Furthermore, the public use file is adjusted to comply with IRS disclosure procedures. First, taxpayers in the sample with total income or loss of \$5,000,000 or more; those with business plus farm receipts of \$50,000,000 or more; and nontaxable returns with adjusted gross incomes or expanded incomes of \$200,000 or more were subsampled at a 33 percent rate to project the identity of individual taxpayers. Second, those returns that remain in the public use file after the subsampling procedure are combined with other high income returns in a blending process to further protect the identity of individual taxpayers. Third, all lower income returns have been blurred for alimony paid and alimony received and home mortgage interest paid to financial institutions. Finally, all fields in the returns have been rounded to the four most significant digits (e.g., \$14,371 = \$14,370 and \$228,867 = \$228,900). These are the main differences between the public use file and the microdata files used by the Treasury Department’s Office of Tax Analysis and the Congress’ Joint Committee on Taxation.

However, all sample data are subject to further sampling and measurement error. To properly use the statistical data presented in distributional tax tables, the magnitude of the potential sampling error must be known; coefficients of variation (CVs) are used to measure that magnitude. Based on the microdata, the table below highlights selected coefficients of variation (CVs) for selected items, tax year 1995 at a 95-percent confidence level. The CVs and subsequent standard errors associated with the public use file will be equal to or greater than the CVs listed in the table below due to the disclosure procedures applied to the public use file by SOI as detailed above. For more information

---

<sup>13</sup> Mike Weber. United States Internal Revenue Service, Statistics of Income Division. “General Description Booklet for the 1995 Public Use Tax File.”

on SOI sampling methodology and data limitation with reference to the tax year 1995 data, please see *SOI Bulletin – Fall 1997*, page 245.

<b>Coefficients of Variation for Selected Items, Tax Year 1995</b>				
(Number of returns is in thousands – money amounts are in millions of dollars – CVs are percentages)				
<b>Item</b>	<b>Number of Returns</b>	<b>Coefficient of Variation</b>	<b>Amount</b>	<b>Coefficient of Variation</b>
Adjusted Gross Income (less deficit)	118,218	0.12	4,189,354	0.34
Salaries and Wages	101,139	0.36	3,201,457	0.56
Net capital gain	10,151	2.36	176,473	1.74
Net capital loss	5,134	3.56	9,715	3.84
Taxable social security benefits	6,598	3.12	45,715	3.78
Total statutory adjustments	18,209	1.56	41,140	2.48
Total standard deduction	83,223	0.48	413,585	0.62
Total itemized deductions after limitations	34,008	1.12	527,374	1.10
Taxable income	94,612	0.44	2,813,826	0.44
Total income tax	89,253	0.54	588,419	0.48

Source: SOI Bulletin. Fall 1997. "Individual Income Tax Returns, 1995." Page 20.

Note: SOI publishes CVs at the 68-percent confidence level. The CVs above have been changed to reflect a 95-percent confidence level.

## BIBLIOGRAPHY

- Auerbach, Alan J. “Public Finance and Tax Policy.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Barthold, A., James R. Nunns and Eric J. Toder. “A Comparison of Distribution Methodologies.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Barthold, Thomas A. “Distributional Analysis at the Joint Committee on Taxation.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Bradford, David F. (Ed.) *Blueprints for Basic Tax Reform*. Arlington, VA: Tax Analysts, 1984.
- . *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995
- Bartlett, Bruce. Brief Analysis #303: “Income Distribution.” National Center for Policy Analysis. Washington, DC. August 10, 1999.
- Browning, Edgar K. “Tax Incidence Analysis for Policy Makers.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Cronin, Julie-Anne. “U.S. Treasury Distributional Analysis Methodology.” U.S. Department of the Treasury. Office of Tax Analysis. OTA Paper 85. September 1999.
- Frenze, Christopher. “Income Mobility and the U.S. Economy: Open Society or Caste System?” Joint Economic Committee. United States Congress. January 1992.
- . “Income Mobility and Economic Opportunity.” Joint Economic Committee. United States Congress. June 1992.
- . “Treasury Department Estimates of Tax Changes: A Review and Analysis.” A Joint Economic Committee Brief. Washington, DC: Joint Economic Committee. United States Congress, July 1997.
- Graetz, Michael J. “Distributional Tables, Tax Legislation, and the Illusion of Precision.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Hubbard, R. Glenn. “Distributional Tables and Tax Policy.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Kasten, Richard A. and Eric J. Toder. “Distributional Analysis at the Congressional Budget Office.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Kennedy, Peter. *A Guide to Econometrics*. Cambridge, MA: The MIT Press, 1992.
- Maier, Mark H. *The Data Game – Controversies in Social Science Statistics*. Armonk, New York: M.E. Sharpe, Inc, 1991.
- McIntyre, Bob. “Bush Scales Back Tax Cut Plan to Trim Cost. New \$1.8 Trillion Plan Tilts Even More to Very Top.” Citizens for Tax Justice. Released May 11, 2000.

- . “House GOP Minimum Wage Plan Offers \$11 in Upper-Income Tax Brakes for Every \$1 in Wage Hikes for Low Earners.” Citizens for Tax Justice. March 7, 2000.
- . “CTJ Releases Distributional Analysis of GOP/Dems Marriage Penalty Bills.” CTJ News, released February 11, 2000.
- . “House GOP Tax Plan: The Rich Get Richer.” Citizens for Tax Justice. CTJ News, released July 27, 1999.
- Nunns, James R. “Distributional Analysis at the Office of Tax Analysis.” in Bradford, David F. (Ed.) *Distributional Analysis of Tax Policy*. Washington, DC: The AEI Press, 1995.
- Pechman, Joseph A. *Federal Tax Policy (5<sup>th</sup> Edition)*. Washington, DC: The Brookings Institution, 1987.
- Robbins, Gary and Aldona. “An Analysis of the Financial Freedom Act of 1999.” Institute for Policy Innovation. Issue Brief. July 30, 1999.
- United States Congress, Joint Committee on Taxation. *Distribution of Certain Federal Tax Liabilities by Income Class for Calendar Year 2000* (JCX-45-00), April 11, 2000.
- . *Methodology and Issues in Measuring Changes in the Distribution of Tax Burdens* (JCS-7-93), June 14, 1993
- United States General Accounting Office. *Quantitative Data Analysis: An Introduction*. (GAO/PEMD-10.1.11), June 1992.
- . *Using Statistical Sampling*. (GAO/PEMD-10.1.6), May 1992.
- . *Tax Expenditures: A Primer*. (PAD 80-26), 1979.
- United States Internal Revenue Service. Publication 17. *Your Federal Income Tax*. Washington, DC.
- . *Statistics of Income Bulletin*. Spring 1999. Washington, DC 1999.
- . *Statistics of Income Bulletin*. Winter 1998-1999. Washington, DC 1999.
- . *Statistics of Income Bulletin*. Fall 1997. Washington, DC 1997.