This paper represents the Statistics of Income (SOI) Division’s first release of data from its Tax Year 1999 Panel of Individual Income Tax Returns. A previous ASA paper explained the history and development of this panel so that only a brief review of the panel’s history and design will be provided in this paper. SOI’s mission is to produce and publish data on the operation of the Federal tax system. Policy analysis and the development of recommendations on the operation of the tax system are not part of SOI’s mission. SOI microdata files, tabulations, and articles are accepted as the nonbiased starting point for policy discussions by individuals of all ideological backgrounds. The fact that virtually all of SOI’s published tabulations are based on cross-sectional samples where the sampling frames and sampling techniques are established and well-known certainly helps SOI fulfill this mission. The publication of tabulations based on panel samples, however, presents a more complicated situation as will be discussed later. The purpose of this paper is to work through some of those complications and to arrive at a series of panel tabulations that can be viewed in the same unbiased light as the more standard SOI tabulations. Already today, income tax return panels provide policy organizations such as the Treasury Department’s Office of Tax Analysis (OTA) and Congress’s Joint Committee on Taxation (JCT) with powerful policy analysis tools that are not available to researchers outside of those organizations. But it is not OTA or JCT’s responsibility to provide voluminous amounts of tabular panel data to the public; it is SOI’s responsibility, and this paper is hopefully a first step in meeting that responsibility.

Background
Each year, the Statistics of Income Division produces a sample of individual income tax returns. The Tax Year 1999 sample included 176,966 returns sampled in 92 stratifications. The sampling rates ranged from 100 percent to .05 percent based on classifications of income and the type of forms and attachments included on each return. The 1999 Edited Panel is an 83,434-return subsample of the 1999 cross-sectional sample. The 1999 Edited Panel contains only 21 stratifications with sampling ranging from 100 percent to .05 percent.

The base year of this panel represents a sample of tax returns. Subsequent years represent a sample of the returns filed by individuals listed as taxpayers on the 1999 base year return. This is a significant difference because it means that the base year sample unit can break apart into two returns through divorce or double the number of individuals in the unit through marriage. Even worse, a unit can divide into two returns through divorce and then, through a second marriage for each original taxpayer, end up representing four individuals. It is these changes that present problems in tabulating, presenting, and interpreting income tax return panel data.

Potential Solutions
One solution to the changing marital status problem is to follow only the primary taxpayer listed on the tax return. The main problem with this approach is that approximately 95 percent of primary taxpayers listed on jointly filed returns are male, thus, a significant gender bias would be introduced into any analysis.

Another possible solution to the changing filing status problem would be to follow both the primary and secondary taxpayers separately. The main problem with this approach is the complexity involved in trying to divide up income between the primary and secondary taxpayers on jointly filed returns. Even if the

---


2 For additional information on the sample design of the annual Complete Report sample, see Internal Revenue Service, Statistics of Income Individual Income Tax Returns, Publication 1304, 1999, “Section 2: Description of Sample.”
income could be divided correctly, the act of doing so has implications. For example, do married individuals make independent or joint economic decisions? If their incomes are divided, how is the joint decision-making aspect retained in the data?

Finally, another possible solution is to simply examine only those panel units where the marital status has not changed. The main problem with this approach is that it excludes all taxpayers who, during the course of the study, either get married, divorced, or had a spouse die. If changes in a taxpayer’s marital status or the death of a spouse affect his or her economic well-being and decision-making process, then that information is lost under this approach.

Obviously, none of these solutions is really adequate, and perhaps the best solution is to utilize all three and compare the results. Unfortunately, such an exercise is beyond the scope of this paper. But given time and resource constraints, and the basic structure of the panel, the easiest and quickest solution to implement is the third solution: examine only those panel units where the filing status has not changed.

An Analysis of Panel Units That Did Not Change Marital Status from 1999 to 2003
The first step is to subset the file to only those panel units where there are returns present for all 5 years of the study. This is not a required step in analyzing panel data. For example, one might want to examine only two points in time, 1999 and 2003, in which case the file would only need to be subset to returns where both of those years were present. But for this paper, the 5-year average Adjusted Gross Income (AGI) is computed and used in subsequent tables, and in order to keep the basis for all tables consistent, only panel units with returns present for all 5 years will be used. (Another solution would be to impute missing returns, but that is beyond the scope of this paper.)

As Figure 1 shows, in 1999, the panel contained an estimated 127 million returns or panel units. But, as of 2003, only 106 million panel units had filed returns for all 5 years. Where did the 21 million panel units go? First, any single taxpayer who died during this time period obviously is part of the 21 million missing units, as are any 1999 filers who no longer met the filing threshold for any or all of the subsequent years. Another portion represents taxpayers who should have filed a return but did not. Often, these taxpayers file, but do so in a subsequent calendar year. Roughly 3 percent of the returns filed each year are for a previous tax year. In other words, the returns are eventually filed with the IRS, and generally within 2 years of the due date. Because of the way returns are selected for this panel, these returns will eventually be sampled and included in the panel file. But this presents SOI with an interesting publication issue. Should the tabulation of panel data be held up for 2 years while we await the addition of 3 percent of 1 year’s data? For example, the file used for this paper is only complete for the period 1999 to 2001. This is a topic for further research.

The second step is to subset the file to those panel units where a return is filed in every year and only one return is filed each year. As is shown in Figure 1, by 2003, this step removes another 3.4 million returns from the panel. These 3.4 million returns generally represent joint filers who divorced and where each taxpayer now files independently of his or her former spouse and couples who on at least one occasion during this 5 year period filed using a marital status of married filing separately. Note that it is possible to add items from a married couple’s two married filing separately returns to generate a combined return, but this process was not undertaken for this paper.

The final step is to subset the file to those panel units where a return is filed in every year and only one return is filed each year and where the marital status does not change. As Figure 1 shows, 14.9 million panel units were removed in this step. Only 87.6 million panel units remain. They generally consist of taxpayers who married during the 1999-2003 period or married couples where one of the spouses died during this period.

As Table 1 shows, in order to create the database that will be used for the subsequent tabulations in this paper, 31 percent of the panel units or base year returns, accounting for 19.4 percent of base year AGI, have been removed. Further research must be conducted to understand the impact of removing these panel units, including answering an important fundamental question: is it even legitimate to produce tabulations where 31 percent of the units have been removed. And if so, what data about the 31 percent should also be presented?
1999-2003 Edited Panel Tables
Table 2 is probably the most basic and straightforward panel tabulation that it is possible to produce. It is produced using the 87.6 million weighted panel units where each panel unit filed one and only one return for each year of the 5 year period under study and where each panel unit maintained the same marital status for the entire 5 year period. The panel units are classified by the AGI shown on the 1999 return and by the AGI shown on the 2003 return. The 2003 AGI amounts, as well as all other amounts shown in this paper, have been deflated to 1999 levels using the price deflator applied in other SOI Individual taxation data.

It should be noted that returns filed by dependents are included in Table 2. If an individual can be claimed as a dependent by another taxpayer, yet has income sufficient to require the filing of a return, the individual is required to file a tax return that is separate from the return on which he or she was claimed as a dependent. In the sample design of this panel, as in the standard SOI individual cross-sectional samples, no attempt was made to create a separate sample stratum for dependent returns. Thus, if sampled, a dependent return represents a unique panel unit as does the return, if sampled, on which that individual was listed as a dependent. Dependents, however, may exhibit significant income changes when they move from dependent status to independent tax filer. For example, a college student earning $4,000 a year at McDonald’s may graduate and earn $40,000 in his or her first professional job. In Table 2, this situation cannot be separated from the case of an adult who is 35 years old and supporting a family who moves from an income of $4,000 in 1999 to $40,000 to 2003. Consequently, Table 3 excludes returns filed by base year dependents. This eliminates another 7.2 million panel units. But as can be seen from comparing both tables, the reduction in panel units is almost exclusively in the $1 under $10,000 AGI class.

A possible concern with Table 3 is that it only presents two points in time. A taxpayer may have earned $50,000 in 1999 and $50,000 in 2003 indicating no real change in income. But what if the taxpayer earned only $10,000 in 2000, 2001, and 2002? The 5-year average income is significantly different than the income at the beginning and the end points of the study period. Consequently, Table 4 is classified by the 1999 AGI and by the 5-year average AGI (in 1999 dollars). As mentioned earlier in the paper, Table 4 is the reason why, in constructing the database of panel units to be used in this study, only panel units where a return was filed for the entire 5-year period were used. As noted earlier, another alternative would be to ease this restriction and develop an imputation method for the missing data. Such an approach was beyond the scope of this paper but should be explored in future research. Imputations of this nature may become essential as the panel ages and more panel units are found to be missing at least one return over the course of the study and thus reducing the number of panel units available for tabulations such as Table 4. Finally, another way to present the 5 year average AGI is in terms of the percentage change from the 1999 AGI. This has been done in Table 5.

---

Figure 1 -- Derivation of 1999-2003 Edited Panel Sample Used in Subsequent Tabulations

<table>
<thead>
<tr>
<th>TaxYear</th>
<th>At least one return present in all years</th>
<th>Column (1) &amp; only one return present in each year</th>
<th>Column (2) &amp; the same marital status in all years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999 thru 2000</td>
<td>120,887,311</td>
<td>119,794,388</td>
<td>114,807,823</td>
</tr>
<tr>
<td>1999 thru 2001</td>
<td>115,810,399</td>
<td>113,770,493</td>
<td>104,860,374</td>
</tr>
<tr>
<td>1999 thru 2003</td>
<td>105,938,164</td>
<td>102,549,251</td>
<td>87,617,774</td>
</tr>
</tbody>
</table>

Notes:  
* 2002 and 2003 data are for returns received by IRS through Calendar Year 2004.  
  Additional returns for 2002 and 2003 were filed in Calendar Years 2005 and 2006.  
* Married filing separately returns have been removed in columns 2 and 3 to simplify processing  
* Base year prior-year returns (approximately 9,000 weighted returns) have been removed.  
* Base year single panel members who married another panel member in a subsequent year (approximately 4,000 weighted returns) have been removed.
Table 1 - 1999-2003 Full Edited Panel and Limited Edited Panel Differences

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No adjusted gross income</td>
<td>1,016,365</td>
<td>547,216</td>
<td>469,149</td>
<td>(13,874,990)</td>
</tr>
<tr>
<td>$1 under $10,000</td>
<td>26,210,180</td>
<td>13,381,189</td>
<td>12,828,991</td>
<td>61,349,284</td>
</tr>
<tr>
<td>$10,000 under $20,000</td>
<td>23,966,960</td>
<td>14,953,415</td>
<td>9,013,545</td>
<td>132,599,506</td>
</tr>
<tr>
<td>$20,000 under $30,000</td>
<td>18,359,111</td>
<td>12,513,685</td>
<td>5,845,426</td>
<td>144,237,142</td>
</tr>
<tr>
<td>$30,000 under $40,000</td>
<td>13,368,846</td>
<td>9,700,429</td>
<td>3,668,417</td>
<td>127,144,988</td>
</tr>
<tr>
<td>$40,000 under $50,000</td>
<td>9,812,207</td>
<td>7,584,758</td>
<td>2,227,449</td>
<td>99,027,042</td>
</tr>
<tr>
<td>$50,000 under $75,000</td>
<td>16,897,458</td>
<td>13,882,868</td>
<td>3,014,590</td>
<td>182,512,574</td>
</tr>
<tr>
<td>$75,000 under $100,000</td>
<td>7,755,507</td>
<td>6,653,302</td>
<td>1,102,205</td>
<td>94,321,971</td>
</tr>
<tr>
<td>$100,000 under $200,000</td>
<td>7,188,685</td>
<td>6,271,959</td>
<td>916,726</td>
<td>118,481,487</td>
</tr>
<tr>
<td>$200,000 under $500,000</td>
<td>1,891,017</td>
<td>1,640,006</td>
<td>251,011</td>
<td>71,761,851</td>
</tr>
<tr>
<td>$500,000 under $1,000,000</td>
<td>355,710</td>
<td>309,944</td>
<td>45,766</td>
<td>30,922,895</td>
</tr>
<tr>
<td>$1,000,000 under $1,500,000</td>
<td>88,847</td>
<td>76,779</td>
<td>12,068</td>
<td>14,611,433</td>
</tr>
<tr>
<td>$1,500,000 under $2,000,000</td>
<td>38,160</td>
<td>33,102</td>
<td>5,058</td>
<td>8,705,708</td>
</tr>
<tr>
<td>$2,000,000 under $5,000,000</td>
<td>57,547</td>
<td>49,710</td>
<td>7,837</td>
<td>23,435,069</td>
</tr>
<tr>
<td>$5,000,000 under $10,000,000</td>
<td>14,176</td>
<td>12,123</td>
<td>2,053</td>
<td>14,064,831</td>
</tr>
<tr>
<td>$10,000,000 or more</td>
<td>8,711</td>
<td>7,289</td>
<td>1,422</td>
<td>33,815,615</td>
</tr>
<tr>
<td>Total</td>
<td>127,029,487</td>
<td>87,617,774</td>
<td>39,411,713</td>
<td>1,143,116,446</td>
</tr>
</tbody>
</table>
Table 2 - Tax Year 1999 filers present in 2000, 2001, 2002, and 2003 with no change in marital status by 1999 AGI class and 2003 AGI class in 1999 dollars

<table>
<thead>
<tr>
<th>2003 AGI Class</th>
<th>Number of Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>No adjusted gross income</td>
<td>1,501,084</td>
</tr>
<tr>
<td>$1 under $10,000</td>
<td>325,234</td>
</tr>
<tr>
<td>$10,000 under $20,000</td>
<td>4,078,377</td>
</tr>
<tr>
<td>$20,000 under $30,000</td>
<td>1,337,586</td>
</tr>
<tr>
<td>$30,000 under $40,000</td>
<td>526,959</td>
</tr>
<tr>
<td>$40,000 under $50,000</td>
<td>141,323</td>
</tr>
<tr>
<td>$50,000 under $75,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$75,000 under $100,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$100,000 under $200,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$200,000 under $500,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$500,000 under $1,000,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$1,000,000 under $5,000,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$5,000,000 under $10,000,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$10,000,000 or more</td>
<td>1,432</td>
</tr>
</tbody>
</table>

Table 3 -- Nondependent Tax Year 1999 filers present in 2000, 2001, 2002, and 2003 with no change in marital status by 1999 AGI class and 2003 AGI class in 1999 dollars

<table>
<thead>
<tr>
<th>2003 AGI Class</th>
<th>Number of Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>No adjusted gross income</td>
<td>1,501,084</td>
</tr>
<tr>
<td>$1 under $10,000</td>
<td>325,234</td>
</tr>
<tr>
<td>$10,000 under $20,000</td>
<td>4,078,377</td>
</tr>
<tr>
<td>$20,000 under $30,000</td>
<td>1,337,586</td>
</tr>
<tr>
<td>$30,000 under $40,000</td>
<td>526,959</td>
</tr>
<tr>
<td>$40,000 under $50,000</td>
<td>141,323</td>
</tr>
<tr>
<td>$50,000 under $75,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$75,000 under $100,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$100,000 under $200,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$200,000 under $500,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$500,000 under $1,000,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$1,000,000 under $5,000,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$5,000,000 under $10,000,000</td>
<td>1,432</td>
</tr>
<tr>
<td>$10,000,000 or more</td>
<td>1,432</td>
</tr>
</tbody>
</table>

Total | 8,046,899 |
Table 4 - Non-dependent Tax Year 1999 filers present in 2000, 2001, 2002, and 2003 with no change in marital status by 1999 AGI class and average 1999-2003 AGI class in 1999 dollars

<table>
<thead>
<tr>
<th>1999 AGI Class</th>
<th>No under under under under under under under under under under under under under or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10,000 under $10,000</td>
<td>38,612</td>
</tr>
<tr>
<td>$10,000 under $20,000</td>
<td>52,256</td>
</tr>
<tr>
<td>$20,000 under $30,000</td>
<td>52,256</td>
</tr>
<tr>
<td>$30,000 under $40,000</td>
<td>52,256</td>
</tr>
<tr>
<td>$40,000 under $50,000</td>
<td>52,256</td>
</tr>
<tr>
<td>$50,000 under $75,000</td>
<td>2,143</td>
</tr>
<tr>
<td>$75,000 under $100,000</td>
<td>177</td>
</tr>
<tr>
<td>$100,000 or more</td>
<td>12</td>
</tr>
</tbody>
</table>

Total | 80,462,860 | 392,619 | 5,060,111 | ... | 281,535 | 68,010 | 29,238 | 43,214 | 9,867 | 4,658 |


<table>
<thead>
<tr>
<th>1999 AGI Class</th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>-100%</td>
<td>75% - 75%</td>
<td>50% - 75%</td>
</tr>
<tr>
<td>$1 under $10,000</td>
<td>318,914</td>
<td>64,306</td>
</tr>
<tr>
<td>$10,000 under $20,000</td>
<td>1,116,609</td>
<td>4,184,732</td>
</tr>
<tr>
<td>$20,000 under $30,000</td>
<td>1,051,765</td>
<td>3,960,772</td>
</tr>
<tr>
<td>$30,000 under $40,000</td>
<td>827,310</td>
<td>3,466,821</td>
</tr>
<tr>
<td>$40,000 under $50,000</td>
<td>659,362</td>
<td>2,753,416</td>
</tr>
<tr>
<td>$50,000 under $75,000</td>
<td>619,138</td>
<td>2,576,348</td>
</tr>
<tr>
<td>$75,000 under $100,000</td>
<td>571,149</td>
<td>2,782,426</td>
</tr>
<tr>
<td>$100,000 under $200,000</td>
<td>827,999</td>
<td>2,457,699</td>
</tr>
<tr>
<td>$200,000 under $500,000</td>
<td>636,488</td>
<td>1,638,637</td>
</tr>
<tr>
<td>$500,000 under $1,000,000</td>
<td>72,676</td>
<td>87,984</td>
</tr>
<tr>
<td>$1,000,000 under $1,500,000</td>
<td>18,383</td>
<td>16,747</td>
</tr>
<tr>
<td>$1,500,000 under $2,000,000</td>
<td>8,411</td>
<td>6,958</td>
</tr>
<tr>
<td>$2,000,000 under $5,000,000</td>
<td>10,533</td>
<td>9,042</td>
</tr>
<tr>
<td>$5,000,000 under $10,000,000</td>
<td>2,263</td>
<td>1,572</td>
</tr>
<tr>
<td>$10,000,000 or more</td>
<td>1,234</td>
<td>85</td>
</tr>
</tbody>
</table>

Total | 6,181,886 | 6,895,918 | 26,804,431 | 79,966,259 | 29,229,611 | 7,191,660 | 2,757,735 | 1,509,432 | 3,729,540 |