

The Tax Year 1999-2003 Individual Income Tax Return Panel: A First Look at the Data

Michael E. Weber, Statistics of Income Division, Internal Revenue Service
1111 Constitution Avenue, NW, Washington, DC 20000

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 paper¹. 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 return². 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

¹ Weber, Michael (2005), "The 1999 Individual Income Tax Return Edited Panel," 2005 *Proceedings of the American Statistical Association, Social Statistics Section, Government Statistics Section*, Alexandria, VA: American Statistical Association.

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 data³.

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 in 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.

³ AGI is shown in constant dollars, calculated using the U.S. Bureau of Labor Statistics consumer price index for urban consumers. U.S. Department of Labor, Bureau of Labor Statistics, *Monthly Labor Review*.

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

	At least one return present in all years	Column (1) & only one return present in each year	Column (2) & the same marital status in all years
Tax Year	(1)	(2)	(3)
1999	127,029,487	127,029,487	127,029,487
1999 thru 2000	120,887,311	119,794,388	114,807,823
1999 thru 2001	115,810,399	113,770,493	104,860,374
1999 thru 2002	111,048,409	108,251,388	96,043,680
1999 thru 2003	105,938,164	102,549,251	87,617,774

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

Size of AGI	Full 1999-2003 Edited Panel		Limited 1999-2003 Edited Panel		Difference		Percentage Difference	
	Number of Returns	Amount of AGI	Number of Returns	Amount of AGI	Number of Returns	Amount of AGI	Number of Returns	Amount of AGI
No adjusted gross income.....	1,016,365	-49,057,319	547,216	-35,182,329	469,149	(13,874,990)	46.2%	28.3%
\$1 under \$10,000.....	26,210,180	132,336,387	13,381,189	70,987,103	12,828,991	61,349,284	48.9%	46.4%
\$10,000 under \$20,000.....	23,966,960	357,434,358	14,953,415	224,834,852	9,013,545	132,599,506	37.6%	37.1%
\$20,000 under \$30,000.....	18,359,111	453,687,690	12,513,685	309,450,548	5,845,426	144,237,142	31.8%	31.8%
\$30,000 under \$40,000.....	13,368,846	464,230,987	9,700,429	337,085,999	3,668,417	127,144,988	27.4%	27.4%
\$40,000 under \$50,000.....	9,812,207	438,993,580	7,584,758	339,966,538	2,227,449	99,027,042	22.7%	22.6%
\$50,000 under \$75,000.....	16,897,458	1,031,747,639	13,882,868	849,235,065	3,014,590	182,512,574	17.8%	17.7%
\$75,000 under \$100,000.....	7,755,507	666,429,881	6,653,302	572,107,910	1,102,205	94,321,971	14.2%	14.2%
\$100,000 under \$200,000.....	7,188,685	944,083,593	6,271,959	825,602,106	916,726	118,481,487	12.8%	12.5%
\$200,000 under \$500,000.....	1,891,017	546,818,812	1,640,006	475,056,961	251,011	71,761,851	13.3%	13.1%
\$500,000 under \$1,000,000.....	355,710	241,057,746	309,944	210,134,851	45,766	30,922,895	12.9%	12.8%
\$1,000,000 under \$1,500,000.....	88,847	107,343,480	76,779	92,732,047	12,068	14,611,433	13.6%	13.6%
\$1,500,000 under \$2,000,000.....	38,160	65,801,348	33,102	57,095,640	5,058	8,705,708	13.3%	13.2%
\$2,000,000 under \$5,000,000.....	57,547	172,372,870	49,710	148,937,801	7,837	23,435,069	13.6%	13.6%
\$5,000,000 under \$10,000,000.....	14,176	97,281,129	12,123	83,216,258	2,053	14,064,871	14.5%	14.5%
\$10,000,000 or more.....	8,711	215,765,177	7,289	181,949,562	1,422	33,815,615	16.3%	15.7%
Total.....	127,029,487	5,886,327,358	87,617,774	4,743,210,912	39,411,713	1,143,116,446	31.0%	19.4%

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

1999 AGI Class	1999-2003 Average AGI Class																		
	Total	No AGI	Number of Returns																
			\$1 under \$10,000	\$10,000 under \$20,000	\$20,000 under \$30,000	\$30,000 under \$40,000	\$40,000 under \$50,000	\$50,000 under \$75,000	\$75,000 under \$100,000	\$100,000 under \$200,000	\$200,000 under \$500,000	\$500,000 under \$1,000,000	\$1,000,000 under \$1,500,000	\$1,500,000 under \$2,000,000	\$2,000,000 under \$5,000,000	\$5,000,000 under \$10,000,000	\$10,000,000 or more		
No adjusted gross income.....	496,602	244,713	103,300	53,236	38,361	16,612	14,740	11,493	3,829	4,628	4,894	260	262	39	107	116	12		
\$1 under \$10,000.....	7,291,321	77,293	3,733,190	2,783,108	502,121	122,257	40,331	22,531	4,288	6,200	-	-	-	-	-	-	-		
\$10,000 under \$20,000.....	14,138,652	32,600	1,103,155	9,341,917	2,968,718	472,937	125,178	65,591	18,796	7,486	2,098	170	-	5	-	-	-		
\$20,000 under \$30,000.....	12,401,452	6,453	92,175	2,054,290	7,156,297	2,465,181	425,060	158,102	26,034	17,683	170	5	-	-	-	-	-		
\$30,000 under \$40,000.....	9,658,256	6,394	16,461	318,777	1,905,998	4,903,657	1,930,748	516,779	46,365	12,503	556	17	-	-	-	-	-		
\$40,000 under \$50,000.....	7,572,662	4,796	6,173	77,148	448,226	1,434,585	3,551,698	1,926,940	98,253	20,056	4,062	-	556	170	-	-	-		
\$50,000 under \$75,000.....	13,866,782	11,631	2,140	37,306	219,033	597,781	1,822,815	9,240,646	1,716,485	201,309	17,629	-	-	5	-	-	-		
\$75,000 under \$100,000.....	6,647,392	177	556	8,136	25,391	96,145	147,910	1,577,060	3,755,049	1,013,742	21,156	2,072	-	-	-	-	-		
\$100,000 under \$200,000.....	6,263,968	2,802	2,081	556	2,140	31,397	48,824	342,598	1,077,802	4,405,489	336,803	11,365	81	2,030	-	-	-		
\$200,000 under \$500,000.....	1,638,337	4,541	766	619	1,113	2,098	3,194	15,341	35,019	455,894	1,030,824	77,690	6,204	1,533	3,476	24	-		
\$500,000 under \$1,000,000.....	308,924	821	-	-	56	170	-	783	56	11,977	123,159	139,607	22,584	4,134	4,335	1,216	25		
\$1,000,000 under \$1,500,000.....	76,553	12	56	-	-	-	-	-	-	356	14,340	30,609	20,003	6,481	4,389	308	-		
\$1,500,000 under \$2,000,000.....	32,989	100	-	5	-	-	-	-	-	79	4,174	9,257	8,023	5,487	5,117	505	242		
\$2,000,000 under \$5,000,000.....	49,572	215	56	5	-	-	-	17	-	17	537	10,437	9,177	7,508	18,720	2,337	544		
\$5,000,000 under \$10,000,000.....	12,113	61	-	-	-	-	-	-	10	-	31	46	1,109	1,835	4,949	3,306	765		
\$10,000,000 or more.....	7,286	12	-	-	-	-	-	-	-	-	5	-	11	15	2,116	2,054	3,069		
Total.....	80,462,860	392,619	5,060,111	14,675,103	13,267,454	10,142,820	8,110,499	13,877,882	6,781,986	6,157,419	1,560,438	281,535	68,010	29,238	43,214	9,867	4,658		

Table 5 --Tax Year 1999 nondependent 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

1999 AGI Class	1999-2003 Average Indexed AGI Percentage Change from 1999 AGI					Total	1999-2003 Average Indexed AGI Percentage Change from 1999 AGI				
	Negative						Positive				
	-100%	75% - 100%	50% - 75%	25% - 50%	0.1 - 25%		0.1 - 25%	25% - 50%	50% - 75%	75% - 100%	100%
\$1 under \$10,000.....	77,293	14,570	88,039	318,916	1,034,061	7,291,321	1,178,392	952,202	637,414	542,640	2,447,793
\$10,000 under \$20,000.....	32,600	12,008	171,947	1,116,009	4,184,732	14,138,652	4,650,586	1,884,650	929,415	486,283	670,422
\$20,000 under \$30,000.....	6,453	13,994	172,635	1,051,765	3,960,772	12,401,452	4,973,418	1,385,268	438,329	175,801	223,016
\$30,000 under \$40,000.....	6,394	8,433	163,984	827,310	3,446,821	9,658,256	3,894,720	880,707	255,018	81,405	93,464
\$40,000 under \$50,000.....	4,796	10,187	117,338	659,362	2,753,416	7,572,662	3,289,105	523,171	116,643	57,437	41,206
\$50,000 under \$75,000.....	11,631	6,914	263,230	1,052,621	5,507,597	13,866,782	5,958,052	781,723	152,324	62,679	70,011
\$75,000 under \$100,000.....	177	8,692	152,127	571,149	2,782,426	6,647,392	2,684,583	303,025	74,243	24,924	46,047
\$100,000 under \$200,000.....	2,802	10,159	231,935	822,798	2,457,690	6,263,969	2,169,316	343,000	88,947	50,879	86,443
\$200,000 under \$500,000.....	4,541	13,940	125,620	363,488	554,625	1,638,337	361,686	110,951	52,451	18,622	32,412
\$500,000 under \$1,000,000.....	821	5,717	47,946	72,676	87,984	308,924	49,727	17,816	7,575	6,455	12,208
\$1,000,000 under \$1,500,000.....	12	2,841	18,023	18,383	16,747	76,553	9,823	3,943	2,999	1,008	2,775
\$1,500,000 under \$2,000,000.....	100	2,312	8,540	7,411	6,094	32,989	3,828	1,848	829	509	1,519
\$2,000,000 under \$5,000,000.....	215	4,484	14,488	10,533	9,042	49,572	4,661	2,541	1,235	615	1,758
\$5,000,000 under \$10,000,000.....	61	2,058	3,835	2,263	1,572	12,113	1,150	558	199	116	300
\$10,000,000 or more.....	12	1,832	2,199	1,234	853	7,286	564	256	113	57	166
Total.....	147,906	118,139	1,581,886	6,895,918	26,804,431	79,966,259	29,229,611	7,191,660	2,757,735	1,509,432	3,729,540