Creativity and Compromise: Constructing a Panel of Income and Estate Tax Data for Wealthy Individuals*

Barry W. Johnson and Lisa M. Schreiber, Internal Revenue Service

The Statistics of Income Division (SOI) of the IRS collects statistical data from all major Federal tax and information returns that are used by both the Congressional and Executive branches of the Government to evaluate and develop tax and economic policy. Among these are annual studies of Form 1040, *U.S. Individual Income Tax Return,* and Form 706, *United States Estate (and Generation–Skipping Transfer) Tax Return*.

Form 1040 is filed annually by individuals or married couples to report income, including wages, interest, dividends, capital gains, and some types of business income. In 1987, SOI undertook a major revision of the sample of Forms 1040 included in its annual studies in order to include a panel component, along with the usual cross-sectional sample. Cross-sectional samples provide reliable coverage of population totals and support annual budget projections as well as a wide range of other research; panels are more useful for estimating behavioral responses to hypothetical tax law changes. The new sample design was created to include all members of a tax family (primary and secondary filers and their dependents) in the panel, and represented the cohort of tax families filing returns in 1988 for Tax Year 1987. It included 39 strata based on income, filing status, and total receipts from businesses and farms (see Czajka and Schirm, 1991; Schirm and Czajka, 1991). For the base year, the initial SOI Form 1040 sample included 114,700 returns, 88,000 of which were panel members, not counting returns filed by dependents, which were added at a later time.

In 1994, the sample for SOI's annual estate tax studies was changed so that data from any Form 706 filed for a deceased 1987 Family Panel member would be collected. A Federal estate tax return, Form 706, must be filed for every U.S. decedent whose gross estate, valued on the date of death, combined with certain lifetime gifts made by the decedent, equals or exceeds the filing threshold applicable for the decedent's year of death. The return must be filed within 9 months of a decedent's death, although a 6-month extension is often requested and granted. All of a decedent's assets, as well as the decedent's share of jointly owned and community property assets, are included in the gross estate for tax purposes and reported on Form 706. Also reported are most life insurance proceeds, property over which the decedent possessed a general power of appointment, and certain transfers made during life. Assets are valued on the day of the decedent's death, although an estate is also allowed to value assets on a date up to 6 months after a decedent's death if market values decline. Special valuation rules and a tax deferral plan are available to an estate that is primarily composed of a small business or farm. Expenses and losses incurred in the administration of the estate, funeral costs, the decedent's debts, bequests to a surviving spouse, and bequests to qualified charities are all allowed as deductions against the estate for the purpose of calculating the tax liability.

▶ The Tax Family Concept

The initial unit of observation for the SOI 1987 family panel was defined as a tax family, which included a taxpayer, spouse, and all dependents (not limited to children) claimed by either. Thus, a tax family could represent single filers (widowed, divorced or separated, or those who were never married), as well as married filers and their dependents. Dependents did not need to live in the same household as the parent to be included in the tax family; however, information on dependents whose incomes fell below the filing threshold was generally not available unless reported on the parent's return. Coresident family members who were not claimed as dependents were not included in the tax family. An interesting complication of the tax family concept is the treatment of married couples who, for various reasons,

*Johnson, Barry W. and Schreiber, Lisa M. (2006), "Creativity and Compromise: Constructing a Panel of Income and Estate Tax Data for Wealthy Individuals," American Statistical Association, Proceedings, Section on Survey Research Methods, (forthcoming).

elected to file separately. For the purposes of the SOI panel, only the partner whose separately filed return was selected into the sample in 1988 was included in the panel; the only way for both spouses of a married couple filing separately in 1988 to have been permanently included in the family panel was for returns filed by each spouse to have been independently selected. Thus, the tax family differs significantly from the more common "household" measure used by many national surveys (Czajka and Schirm, 1993) [1].

▶ The Data

Between 1987 and 2004, there were 6,614 Federal estate tax returns filed for 1987 Family Panel members or visitors [2]. Of these, 5,659 estate tax returns were identified as having been filed for permanent 1987 Individual Family Panel members who died between 1994 and 2004 [3]. These 5,659 decedents form the core of the SOI Family Panel Decedent Data Set (FPDD) [4].

Individual income tax data were collected by SOI for the 1987 Family Panel from Tax Year 1987 through Tax Year 1996. SOI data consist of both the set of data items that are collected for administrative processing of Form 1040 and all attachments, as well as many more detailed data items required for complex statistical and economic analysis of taxpayer behavior. In addition, data collected by SOI are extensively tested and adjusted to minimize nonsampling error related to taxpayer mistakes and errors introduced during the data transcription process. For tax years after 1996, SOI continued to collect administrative data related to the Family Panel members, but due to problems of panel drift decided to discontinue SOI processing of panel member returns, electing instead to develop new panels based on lessons learned from this initial exercise. The most convenient source of the administrative data for 1997 to 2004 is the Compliance Data Warehouse (CDW) maintained by the IRS Office of Research. The CDW houses, among other things, a complete archive of administrative data for Form 1040 and selected attachments in a normalized relational database. Its primary purpose is generalized statistical research on taxpayer behavior, so that very little information which can be used to identify individual taxpayers is available. In fact, only a four-digit name control and a masked Social Security number (SSN) for the primary filer of a return are available to most users of this dataset. Special permission was required to gain access to tables that link the actual SSN with the masked version. Combining data from SOI and the CDW, a total of 72,373 income tax returns filed for Tax Years 1987-2003 were available for the FPDD.

Ideally, an income tax return would be available for every tax period between 1987 and a decedent's year of death. For 98.2 percent of decedents, this was the case. For 1.3 percent of all decedents, only 1 return was missing from the time series 1987 through the last full year prior to death, leaving only a handful of decedents for whom more than 1 return was missing from the panel [5].

A panel sample of income tax filers, the elements of which have at their core two common factors, that of being sampled based on 1987 reported income and that of having an estate tax return filed sometime after that, poses interesting analytical challenges. Two of these relate to selecting appropriate reference periods and determining how to treat changes in tax family composition over time. In addition, the selection criteria for inclusion in the FPDD changed during the sample period due to changes in the estate tax filing threshold, which ranged from $600,000 in gross assets in 1994 to $1.5 million in 2004. Another important consideration is that only a decedent's share of a married couple's assets is reported on an estate tax return, while income tax returns for married couples who file jointly report income attributable to both partners. Because income tax data were obtained from two different sources, there are also variations in the available data items from different tax years, subtle differences in data definitions, and differences in data quality. Finally, with a few exceptions, only income subject to taxation is reported on a tax return, and that reported income may be subject to both accidental and intentional misreporting by the taxpayer.

The FPDD includes individual income tax data for Tax Period 1987 for all sampled tax families by definition. It also includes an estate tax return for at least one member of each tax family. This suggests two relevant reference periods for research purposes, either 1987 or the year of death reported on the estate tax return. Selecting 1987 as the reference period is advantageous for

some research because the probability of being selected into the file is known, making it theoretically possible to produce population estimates from the file. However, since wealth valuation data in the file are for deaths between 1994 and 2004, the time series of income data vary from about 7 years to 17 years, which might be limiting for certain types of analysis.

Because one of the prime features of the FPDD is the connection of income to wealth, the date of death—that is, the date for which wealth data are available—is also an attractive reference period. The income stream that would be most relevant in this case would be income reported in the years immediately prior to death. Focusing on income in this way would be appropriate for studying changes in income sources and savings habits as individuals approach the end of their lives, and analyzing the relationship between wealth and realized income. Given that years of death in the FPDD range from 1994-2004, a disadvantage of this approach is the difficulty of controlling for intertemporal differences in economic conditions that affect rates of return and therefore influence portfolio allocation decisions. This dynamic nature of portfolio allocation decisions, often indicated by the realization of capital gains, also makes it difficult to align income earned in one period with assets observed in another, even when the two periods are relatively close.

Longitudinality introduces problems with the tax family concept because, over time, a filing unit may change composition, which is usually accompanied by changes in filing status (Czajka and Radbill, 1995). For example, married persons divorce, single persons marry, couples who customarily file jointly may elect to file separately and vice versa, dependent filers may file independently, or one spouse of a married couple may die. Tax families for married persons can be particularly complex. As a result, an individual might appear in the panel as: a primary filer on a joint return married to an original panel member or visitor (spouse who entered the panel after 1988); a married primary filer on a separate return whose spouse may or may not be in the panel; a secondary filer on a joint return (married to an original panel member or to a visitor); and as a single filer. The longer the time series is carried forward, the greater the possibility for combinations of these events to occur. There are a number of strategies for handling these changes in tax family composition. The most straightforward is to limit analysis to only those filing units that do not change over time. However, this approach tends to introduce a bias since the more stable filing units will tend to have more stable incomes. A second approach is to focus analysis on person level data, imputing income for each individual in the tax family.

Figures 1 and 2 show panel members grouped into two broad categories, single filers and joint filers, in order to examine changes in filing status over time [6]. Looking first at each panel member's filing status in 1987, Figure 1 shows that, overall, filing status changed for 24.6 percent of all filers between 1987 and the year prior to death [7]. There was slightly more stability for single filers, only 15.2 percent of whom filed a joint return at some point during the period; 26.4 percent of joint filers became single filers sometime between 1987 and death. Figure 2 shows each panel member's filing status in the year prior to death and compares it to income tax returns filed for earlier tax periods. Only filers for whom a Form 1040 was available for at least 7 years prior to death were included in the figure [8]. Using this criterion, filing status was constant for 85.1 percent of all panel members over the 7 years preceding death. Individuals who were single filers at death were much more likely to have changed filing status in the

Figure 1—Filing Status Stability, Using 1987 as Reference Year

Filing status	Return present 1987	Filing status unchanged 1987 to 1 year prior to death	
		Number	Percentage
Single	881	747	84.8
Joint	4,778	3,518	73.6
Total	5,659	4,265	75.4

Figure 2—Filing Status Stability Using Year of Death as Reference Year

Filing status	Return filed year prior to death	Number of years prior to death filing status unchanged			Percentage unchanged for 7 years
		3	5	7	
Single	1,865	1,586	1,370	1,186	63.6
Joint	3,744	3,681	3,630	3,588	95.8
Total	5,609	5,267	5,000	4,774	85.1

years preceding death than those who were joint filers. Only 63.6 percent of all individuals who were single filers in the year prior to death had been single over the 7 years examined, reflecting both couples for whom one spouse died and those who divorced or separated during the period. Almost 95.8 percent of individuals who were joint filers at death had been married for at least the previous 7 years.

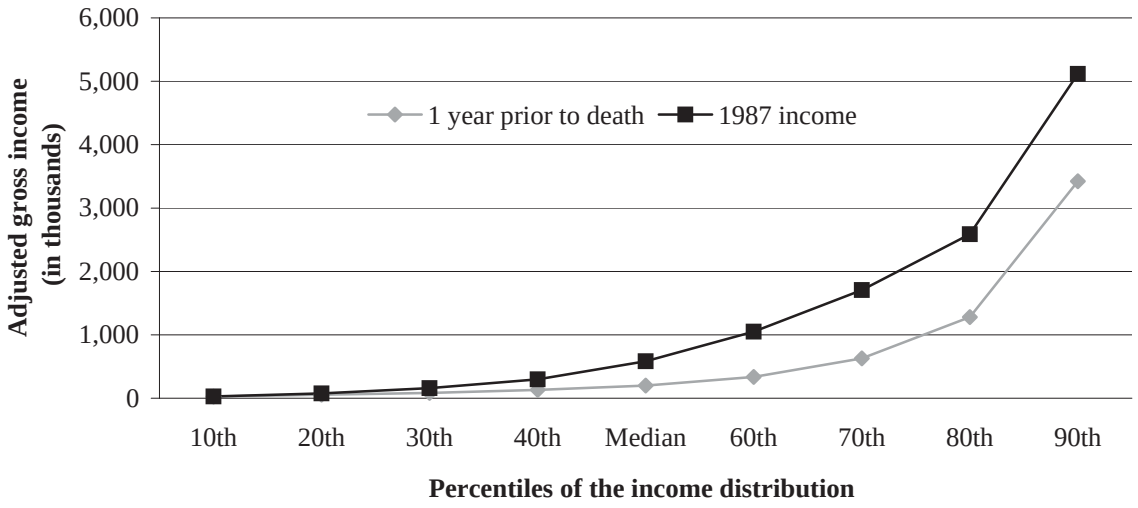
▶ Descriptive Statistics

Despite the limitations and challenges discussed in the previous section, the FPDD gives a unique opportunity to learn more about the way that incomes change as people age and contemplate the end of their lives and also provides a snapshot of the wealth that was the source of a portion of that income. This section briefly describes individuals in the FPDD. For this analysis, filing units are again examined in two broad groups, single filers and joint filers, all estimates are unweighted, and all money amounts have been converted to constant 2001 dollars [9].

There are 5,659 decedents in the FPDD. In 1987, the base year of the panel, 881 were single filers, 48.2 percent of whom were female. The majority, 64.3 percent, of the 4,778 panel decedents who were joint filers in 1987 were male. The mean and median ages of females in the FPDD were 65 and 66, respectively, in 1987 and 76 and 78 at death. The mean and median age for males in 1987 were 63 and 64, respectively, and 75 and 76 at death. These statistics indicate that many of the decedents in the FPDD were at or nearing retirement in 1987, the inception of the panel.

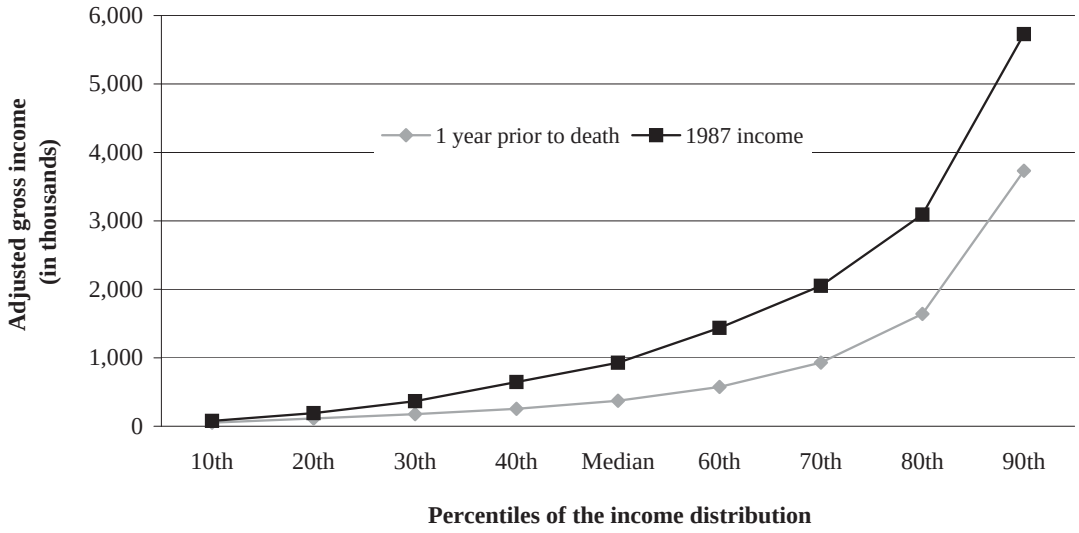
For all filing units whose filing status did not change between 1987 and the year prior to death, reported adjusted gross income (AGI) declined over this period, which is not surprising given that most individuals in the panel were transitioning from work into retirement over the period covered by the panel. For single filers, mean AGI declined from almost $2.0 million in 1987 to $980,000 at death. Figure 3 shows that this decline was an overall flattening and downward shift of the AGI distribution for these filers, with relatively little change for those in the lower percentiles and with the largest differences in the middle of the distribution. Median AGI, for example, declined from about $580,000 in 1987 to almost $200,000 in the year prior to death, a decrease of 65.6 percent. A similar pattern is shown in Figure 4 for joint filers, for whom mean AGI declined from $2.2 million to $1.7 million between 1987 and the death of

Figure 3—Income Distribution in 1987 and Year Prior to Death, Single Filers*



* Dollar amounts are unweighted and in constant dollars.

Figure 4—Income Distribution in 1987 and Year Prior to Death, Joint Filers*



* Dollar amounts are unweighted and in constant dollars.

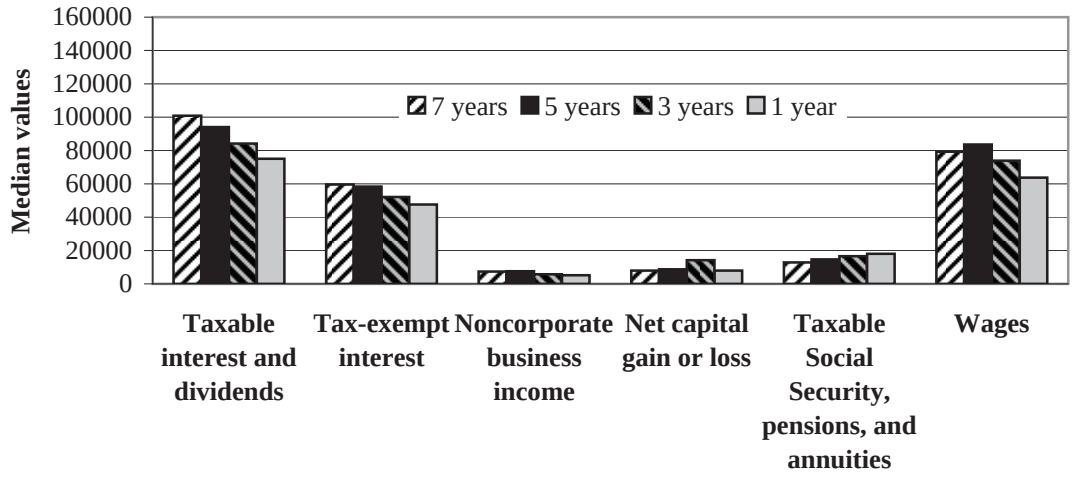
one partner. Median AGI for joint filers declined nearly 60.0 percent, from almost $930,000 to about $370,000, while AGI for those in the 90th percentile declined less over the period, about 35.0 percent.

Figures 5 and 6 decompose AGI into major components for selected years over the 7-year period preceding a panel decedent's year of death [10]. For single filers, overall, median values for wages, taxable interest and dividends, and income from noncorporate businesses decreased as individuals aged. Median values for tax-exempt interest, derived from investments in bonds issued by State or local governments, also declined, overall, for the 7-year period shown in Figure 5. However, for wealthier decedents, those with $5 million or more in gross assets at death, income from tax-exempt bonds increased over this period. For all single decedents, taxable Social Security, combined with pension and annuity income, increased over time, while gains from sales of capital assets were relatively stable.

Figure 6 shows that, while the income distributions for single and joint filers exhibit similar downward shifts over time, the sources of these declines differ between the two groups. For joint filers, income from wages, as well as interest and dividends from taxable investment assets, declined over the 7 years preceding the death of one spouse, but income from most other sources was either stable or increased over this period. Most notable was the relative stability in tax-exempt income for joint filers, overall. For the wealthiest joint filers, however, those where one spouse owned $10 million or more in gross assets at death, tax-exempt income increased by 40 percent over the period examined. For these wealthy filers, income from noncorporate businesses increased by almost 27.0 percent over time.

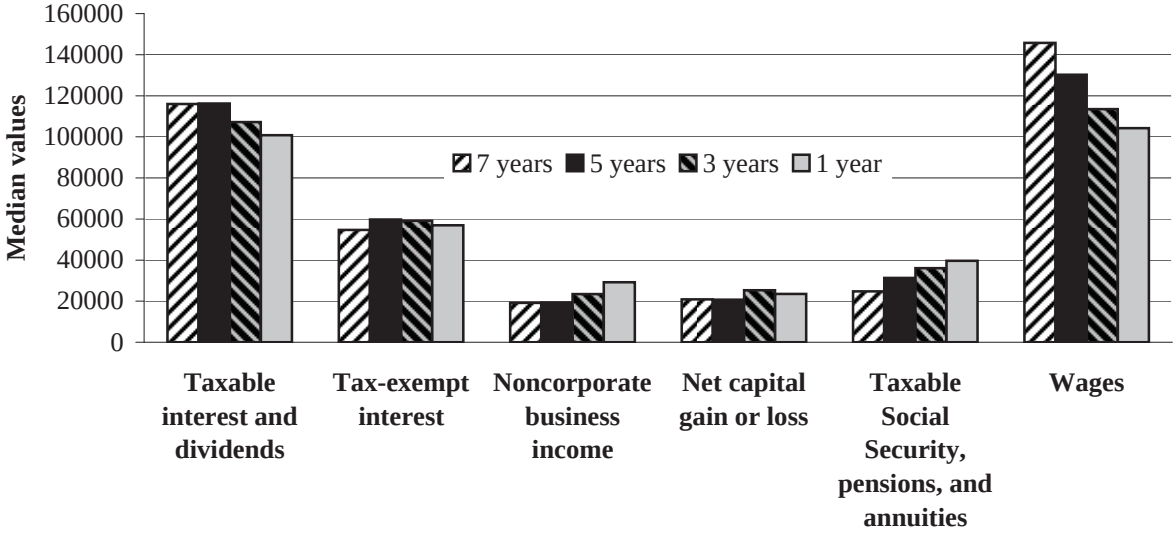
Figures 5-6 showed that, as panel members aged, the share that wage income contributed to AGI decreased, while the patterns of change in income from other sources varied somewhat, depending on filing status and wealth class. It has been noted that the realization of income derived from assets is a more or less voluntary event. Wealthy individuals, those for whom return on investments makes up a relatively large source of income, have the ability to allocate their portfolios in order to take maximum advantage of preferences built into the tax code, to reduce risk, and to vary income significantly according to their own consumption needs. According to Steuerle (1985), the voluntary nature of capital income recognition implies that "taxes paid and benefits received will vary tremendously among persons in fairly identical circumstances." He goes on to state that, because of the voluntary nature of income recognition, using income

Figure 5—Changes in Income Composition, Selected Years Prior to Death, Single Filers*



* Dollar amounts are unweighted and in constant dollars.

Figure 6—Changes in Income Composition, Selected Years Prior to Death, Joint Filers*



* Dollar amounts are unweighted and in constant dollars.

as a classifier in statistical analyses will be inaccurate or misleading for many purposes.

For many decedents, income reported on a tax return in the year prior to death will be closely correlated with the assets reported on an estate tax return filed at death [11]. It is, therefore, possible to estimate rates of return on various asset classes. Rates of return are estimated as income attributable to each class of assets as reported on Form 1040 and its attachments in the last year prior to death, divided by the value of those assets reported on Form 706. Figure 7 shows median values for estimated rates of return for all capital assets, for investment assets that produce taxable income, and for tax-exempt bonds. For single filers with gross assets under $1 million, the rate of return on capital was 4.27 percent. This rate declined for individuals in higher wealth classes, and was just 2.13 percent for single filers with $10 million in gross assets at death. Likewise, rates of return on investments that produced taxable interest or dividends declined with gross asset size. It is interesting to note, however, that the rate of return on tax-exempt investments was fairly stable for single filers, regardless of their wealth. These trends, when combined with those seen previously in Figures 5 and 6, suggest a systematic reordering of the portfolio, over time, favoring tax-exempt income sources over those that produce taxable income. For joint filers, rates of return show a similar pattern across wealth classes, although there was more variation across wealth categories for rates of return on tax-exempt bonds than was seen for single filers [12].

Figure 7—Selected Rates of Return One Year Prior to Death, by Size of Gross Assets

Asset	Size of gross assets	Single	Joint
Return on capital assets	All	2.74	2.84
	Under $1 million	4.27	4.31
	$1 million, under $5 million	3.27	3.52
	$5 million, under 10 million	2.40	2.48
	$10 million or more	2.13	1.85
Return on taxable bonds and stocks	All	2.92	2.15
	Under $1 million	3.83	3.01
	$1 million, under $5 million	3.08	2.37
	$5 million, under 10 million	2.58	2.20
	$10 million or more	2.65	1.77
Return on tax-exempt bonds	All	5.72	5.12
	Under $1 million	5.77	5.72
	$1 million, under $5 million	5.84	5.49
	$5 million, under 10 million	5.72	5.17
	$10 million or more	5.65	4.40

▶ Conclusion

Panel data consisting of income reported by wealthy taxpayers provide important opportunities to study the ways in which income changes over time. When paired with wealth data from Federal estate tax returns, the resulting data set provides a rare opportunity to learn more about the relationship of wealth to realized income, which is an important consideration in many public policy debates, and about changes in income that occur as people near the ends of their lives. These data, however, present many challenges to researchers, a number of which have been explored in this paper. Techniques for dealing with problems that arise due to the longitudinality of the data set, differences in reporting units on income and estate tax returns for joint filers, the dynamic nature of investment portfolios, and many other challenges must be explored before the full potential of the FPDD can be realized. However, the preliminary statistics presented in this paper suggest that there is much that can be learned by addressing these issues using even the most basic assumptions.

▶ Endnotes

[1] Dependents are not included in the analysis presented in this paper.

[2] Estate tax returns filed prior to 1994 were identified by matching panel member SSNs to the IRS Master File. Due to the limited amount of estate tax data available from the Master File for these pre-1994 decedents, they are not included in the FPDD.

[3] Estate tax returns were filed for an additional 57 panel members, but they were missing key documentation or schedules at the time of SOI processing and had to be rejected.

[4] Visitors to the panel were not included in the final dataset since income data were only available for

the period of time that they were associated with an original panel member.

[5] Missing returns can occur either because a taxpayer was not required to file in a given year, or because of an error in reporting a taxpayer's SSN. The latter occurred mainly in the case of secondary SSNs in the 1987 panel. After the period covered by this study, the IRS implemented processing improvements that have reduced these types of errors.

[6] The category "single" includes filers who were unmarried, widowed, and married individuals who elected to file separately since the data on these returns should reflect income attributable to one individual.

[7] The year prior to death is used because a return filed for the year of death would usually reflect income earned during only that portion of the year during which a decedent was alive.

[8] "Seven years" is used since that is the maximum number of full-year income tax returns that would be available for 1987 panel members who died in 1994.

[9] Values were converted to constant dollars using the GDP chain-type price index. Source: Bureau of Economic Analysis.

[10] Only those panel members whose filing statuses did not change over the 7 years preceding their years of death are included in Figures 5 and 6.

[11] In some cases, assets that generated income reported in the year prior to death may have been sold and the proceeds either consumed or invested differently prior to reporting on Form 706; however, no attempt to adjust the data was made for this analysis.

[12] For joint filers, asset values reported for the decedent spouse were doubled in an attempt to approximate the full value of a married couple's asset holdings. This approach will likely overstate the combined asset holdings, in aggregate, causing rates of return to be understated somewhat.

▶ References

Czajka, John L. and Radbill, Larry M. (1995), "Weighting Panel Data for Longitudinal Analysis," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC.

Czajka, John L. and Schirm, Allen L. (1991), "Cross-Sectional Weighting of Combined Panel and Cross-Sectional Observations," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC.

Czajka, John L. and Schirm, Allen L. (1993), "The Family That Pays Together: Introducing the Tax Family Concept, with Preliminary Findings," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC.

Schirm, Allen L. and Czajka, John L. (1991), "Alternative Designs for a Cross-Sectional Sample of Individual Tax Returns: The Old and the New," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC.

Steuerle, Eugene (1985), "Wealth, Realized Income, and the Measure of Well-Being," in David, Martin and Smeeding, Timothy, editors, *Horizontal Equity, Uncertainty, and Economic Well-Being,* University of Chicago Press, Chicago, IL.