

Chapter 7: Studies Linking Income & Wealth



THE MISMEASURE OF MAN' S WELL-BEING: Refining Realized Income Measures with Wealth, Portfolio, and Mortality Information

Barry Johnson, Statistics of Income Division, Internal Revenue Service
Jenny Wahl, Economics Department, Carleton College

Economists and policymakers often rely on realized income to gauge individual well-being. Attractive for its ease of calculation, this measure is nonetheless seriously flawed, in part because people have some ability to choose how much income to realize at a given time. Income from capital is particularly susceptible to manipulation. In this paper, we build upon the path-breaking work of Projector and Weiss (1969) to ascertain the degree of mismatch between realized income and wealth and to suggest ways to construct better indices of wellbeing.

Empirical evidence from a recently compiled Internal Revenue Service data set shows just how imperfect realized income can be as a measure of economic well-being. Linked federal estate and income tax returns reveal that wealthy individuals – particularly those in their prime working years -- realize very low returns on capital. What is more, less-wealthy retirees tend to realize larger returns than more-wealthy retirees.

Our data also allow us to impute wealth on the basis of realized income, portfolio allocation, and other important factors. We offer here some initial results that suggest how one might use income data to predict an individual's wealth.

EXISTING RESEARCH

The research most closely associated with ours is a set of studies conducted by C. Eugene Steuerle (1983, 1985). Steuerle used a database containing wealth and income

data from federal tax returns to examine the relationship between realized income and the underlying wealth that generates at least a portion of that income.

Steuerle's work serves as a partial blueprint for ours, although our data are much more extensive. Not only are Steuerle's samples smaller and more restricted than ours, they fail to contain weights that reflect the probability of a match between estate and income tax records. Nor did Steuerle's data have weights to conform the decedent population more closely to the living population.

Other research has informed ours as well. The years since Steuerle's work have witnessed the advent of data sets such as the Federal Reserve Board of Governor's triennial Survey of Consumer Finance (SCF), the University of Michigan's Panel Study of Income Dynamics, and the Census Bureau's Survey of Income and Program Participation.

One result of this inquiry is the creation of measures of well-being that blend components of income and annuitized values of certain assets (see for example Ringen 1988, Radner 1990, Wolff et al. 2004). Much of this work — particularly in the federal government — centers upon measures of poverty (see for example Bauman 1999, Mishra et al. 2002). The appropriate treatment of wealth in poverty indices remains a significant source of debate, however (see Short and Ruggles 2004).

Recent work (especially Kennickell 1999, 2001) explores the possibility of modeling the relationship between wealth and income for the very wealthy as well. The work we describe here suggests that estate tax data may prove particularly useful in this effort.

DATA

U.S. federal tax records offer a bountiful source of intergenerational data on income and wealth. For our ongoing research, the Statistics of Income Division of the Internal Revenue Service has selected a set of federal estate tax returns and matched it to income tax returns filed by the decedent in the year before death, to gift tax returns filed during the decedent's lifetime, and to income tax returns filed by beneficiaries. These data are referred to collectively as the Estate Collation Study. The core data for this paper come from a stratified sample of federal estate tax returns (form 706) filed in 1992 and 1993 for people who died in 1992 and left estates of at least \$600,000. Matched to the estate tax returns are income tax returns (form 1040) filed by the decedent in 1991. The final Estate Collation sample consists of returns for 3,767 decedents.

Before analyzing the data, we needed to account for certain factors. Estate tax returns did not have equal probabilities of being matched to income tax returns due to primarily to late filing and errors in the secondary Social Security numbers on returns of joint filers. We therefore generated weights for the sample that reflect the unequal probabilities of a successful match. The first step was to create an adjustment factor to balance to the original population totals, essentially treating unmatched records as non-respondents. We then used auxiliary data, post-stratification, and raking to adjust the sampling weights and compare them to estimates from other sources of administrative data. Johnson and McMahon (2002) describe this process in greater detail.

A second feature of the data that requires adjustment has to do with differences in traits between the living population and the decedents. The 1992 estate tax population consisted of some 60,000 individuals with gross estates of at least \$600,000, the estate-

tax filing threshold in effect at that time. These decedents represented less than 1 percent of the U.S. population in 1992, according to the U.S. Census Bureau, and accounted for 2.8 percent of all 1992 deaths. While female decedents comprised 51.2 percent of the total U.S. resident population in 1992, female decedents made up only 43.5 percent of the 1992 estate tax population. The majority of male decedents -- 65.8 percent -- were married, as compared to 56.8 percent in the general population. Most female estate-tax decedents -- 61.5 percent -- were widowed, much higher than the 11.2 percent observed for their living counterparts in 1992. More than 87 percent of male decedents were 60 years old or older, while 14.4 percent of the living male population was in that age group in 1992. Likewise, 94.5 percent of female estate tax decedents were age 60 or older, while just 18.9 percent of living women were in that age group in 1992. Eller et al. (1992) contains a more complete description of the 1992-estate-tax population.

These statistics highlight one of the potential deficiencies of using data from estate tax returns to study the living population. As Smith (1985) points out, estate-tax data provide an excellent means of making statements about the deceased, but do not of themselves allow inferences about the living population. To compensate for the age bias and produce estimates more representative of the living population, we re-weight the file using reciprocals of mortality rates (by age and sex), adjusted by a differential that reflects the lower mortality rates experienced by the wealthy. Richer people tend to live longer because they enjoy access to better healthcare, safer occupations, and superior nutrition. Johnson and Woodburn (1994) provide a full discussion of weight adjustments.

Another potential limitation of the Estate Collation file concerns married decedents. While the estate tax return should contain complete information on the

decedent's portfolio, many, perhaps most, married decedents filed income tax returns jointly with a surviving spouse. Yet we do not directly capture the assets of the surviving spouse for the purpose of calculating returns to capital. We make a partial adjustment for this by including the full value of any property owned jointly by the decedent and surviving spouse in our asset base, including all community property and property owned as tenants-in-common. But we still miss the value of assets owned solely by the surviving spouse. While we have experimented with imputing values for these assets, we make no adjustment for them here. In some of our analysis, we do try to account for possible differences between married and other decedents – for example, by including dummy variables in various regressions.

One further data concern: the reporting of certain assets on federal estate tax returns is idiosyncratic. For example, the full face value of life insurance is included in the decedent's total gross estate for tax purposes. In addition, the tax code allows certain adjustments in asset value, such as the special valuation of real estate used for farming or certain business purposes. Where possible, we modify the data to compensate for these reporting anomalies. In the case of life insurance, for instance, we impute an equity value using data from the 1992 SCF.

Table 1 shows income and assets by source for our matched data, using weights appropriate for the estate-tax population. It also reports estimates generated using weights and asset values adjusted to represent the living population. As might be expected, the share of net income subject to tax attributable to salaries and wages is substantially higher for the living population than for those in the decedent group. Likewise, income from businesses is much higher for the living group. The portfolio

estimates for these two groups reveal differences as well. The share of the portfolio pertaining to business assets is higher in estimates for the living population than that for decedents, as is the share made up of real estate. The proportion of investments in tax-exempt bonds is significantly higher in the estimates for the decedent group than for those in the living population. A comparison of the adjusted data set with estimates from the 1992 SCF indicates that the adjusted estate values are consistent with patterns seen in the 1992 living population.

But other data issues remain. Certainly, many decedents must have been aware that they were close to death, so their portfolios could differ from those of the general population. Decedents (and their executors) naturally had incentives to report the lowest possible legal values for tax purposes on both income and estate tax returns. We believe that the high audit rate for estate-tax returns ensures that evasion is relatively rare, although informal transfers of small items such as jewelry surely take place. In addition, the truncation of the distribution due to the estate-tax filing threshold means that we must be cautious in generalizing from these data to populations other than the relatively wealthy. Finally, limitations due to the timing and retention of IRS masterfile data (the source of income-tax data for this study) mean that the only income data available are for income earned during calendar year 1991. Some income-producing assets could have been sold in 1992 prior to a decedent's death, but we cannot track that transaction. What is more, Kennickell (2001), among others, has suggested that pooling several years of data smoothes out year-to-year fluctuations in income caused by events such as change in employment status, receipt of inheritances, or realization of capital gains. Our one year of income data could therefore contain substantial transitory components. Despite these

flaws, we think our data are more promising than other micro-level data sources for realized property income because they do not suffer from the amount of underreporting and item non-response present in most survey data.

EMPIRICAL FINDINGS

Figure 1 shows the distributions of income and wealth across net-worth deciles for individuals in the Estate Collation sample. The graph shows that, as wealth increases, income also increases. However, the rate of growth for income is significantly less than that of wealth, as evidenced by the flatter slope of the income line. This means that the realized rate of return on assets actually tends to decrease as individuals accumulate more wealth.

Calculating Realized Rates of Return on Capital

Realized rates of return differ from actual economic rates of return by the amount of unrealized income or other income from capital not reported on a tax return. For our sample, the mean return on capital for all ages is 4.6 percent, with the mean return for those aged 50 to 59 at 2.9 percent. These rates are lower than those associated with a reasonably risk-free, low-paying savings account. For instance, six-month CDs generated an average 5.91 percent return during 1991. What is more, our measure for realized return to capital is likely biased upward because some included income items overstate the actual return.

Economic theory suggests that higher-risk, less-liquid assets generate higher economic rates of return. We think it implausible that these wealthy and presumably

investment-savvy decedents would have been satisfied with the relatively low rates they realized on complex portfolios of stock, bonds, real estate, partnerships, and the like. Instead, we believe that the pattern of realized returns offers evidence of careful tax planning, which became more important the more the individual had at stake. Indeed, the fact that people appear in high-wealth categories suggests that these individuals were successful in generating both a high economic return and a low realized return (and thus low taxes).

Table 2 shows estimated average realized rates of return on capital across different classes of wealth for decedents of various ages as well as for the living population. For those aged 70 or older, the table reveals that those with estates of \$10 million or more realized lower returns than those with estates less than \$1 million. Also notable is that individuals in their prime work-years tend to realize lower returns on capital than retirees. This finding reinforces our planning argument: if people earn taxable labor income, they may wish to realize relatively less capital income than those who are not working.

Rate-of Return Regression Analysis

Regression analysis might allow us to say more about the influence of one's portfolio upon realized rates of return on capital and on stock. Consistent with Steuerle, we find that realized rates of return varied inversely with the value of the particular asset in question, holding other relevant variables constant.

Yet modeling rates of return from estate and income tax data is fraught with problems. Income generated from various assets that could appear on an estate tax return

can be categorized in many different ways, for instance. Consequently, we do not draw conclusions from this rate-of-return analysis.

Estimating Wealth from Components of Gross Income

Rather than refine the rate-of-return analysis, we construct a model that predicts wealth from components of realized income and adjustments to income. A very simple model poses total assets as a function of various types of income reported on the 1040, along with the value of interest deductions and an index for the importance of deductions and adjustments to income. Table 3 reveals these results. For the overall sample, total assets are an increasing function of age in the relevant range. Nearly all income components have a positive relationship to total assets, with the largest coefficient associated with dividend income. The regression weighted to the living population suggests that an extra dollar of dividend income implies an increment to total assets of \$83. In turn, this result gives us a point estimate of only 1.2 percent for the rate of return on assets that yield dividends. The coefficient on taxable interest implies a higher estimated yield of 10.7 percent. The negative coefficient on tax-exempt interest seems odd; as we shall see, however, all but the highest-income people generate a positive relationship between tax-exempt interest and asset value. A similar result holds for capital gains and other income.

Deductions from income as well as income components plausibly might relate to the amount of assets held. The regression results indicate, for instance, that higher interest deductions are associated with more total assets. This result makes sense, given that the deductions probably act as a proxy for the value of real estate. One other variable

of note is “propadj.” This variable indicates the proportion of gross income subject to tax that is made up of adjustments and deductions such as those pertaining to Keogh plans, IRAs, and SEPs. Higher deductions can thus be associated with the building up of assets. Consequently, we might expect that, for a given level of gross income, people with a higher “propadj” would have higher total assets; indeed, the regression coefficient on “propadj” is positive.

Segmenting our data helps us craft even better predictions of total assets. Total assets are closely related to the amount of reported dividends. The relationship is even stronger for dividend amounts above a threshold of about \$2000. Total assets are also correlated positively – though less strongly -- with other income components such as taxable and tax-exempt interest.

These relationships suggest categorizing decedents on the basis of dividends received, with special treatment for those who realized very small amounts. Table 4 reports the results of regressions by dividend class. These results indicate that non-corporate and tax-exempt interest income are more important in predicting total assets for people reporting relatively small amounts of dividend income, whereas dividend income matters more in predicting total assets for those receiving larger amounts of dividend income.

The predictions from the set of regressions reported in Table 4 appear quite promising, because they yield estimates of well-being that are much more closely related to total assets than are income measures. Significantly, the Pearson coefficient relating the predicted value to actual total assets is .79, whereas the coefficient for taxable income is .22 and for gross income is .48. Predicted values from regressions pertaining to

decendents indicate that these values are also better than income measures at ranking observations. The Spearman coefficient for the predicted value is .70, as compared to a coefficient of .56 for taxable income and .67 for gross income. The Spearman rank results are not as clear for the regressions pertaining to the living population. Because the way this index is constructed can yield a lower value when rank shifts are more frequent but relatively more minor, however, we give more weight to the Pearson results. What they suggest is that we may have found a useful technique to gross up income components to yield a predicted value of well-being for wealthy people.

CONCLUSION

The evidence shown here indicates that income from capital is in many ways a voluntary event. Realized property income can vary dramatically across wealth and age classes, most likely reflecting tax considerations rather than differences in true economic returns. Indeed, wage income may be doing the same, particularly for executives who earn substantial amounts of non-wage compensation that receive preferential tax treatment. Income alone is no longer a reasonable way to assess individual well-being.

Our research suggests that merged estate and income tax records offer an effective way to gauge individual well-being among the wealthiest portion of the population. In short, we show how various components of income and deductions associated with capital assets can be combined to yield a predicted value for total assets that is highly correlated with actual assets. Applying our methods may therefore help reduce the degree of mismeasurement in man's well-being.

SOURCES

- Bauman, Kurt. Extended Measures of Well-being: Living Conditions in the United States:1998. *Current Population Reports*. Washington, DC, March 2003.
- Eller, Martha Britton; Erard, Brian; and Chih-Chin Ho. Federal Taxation of Wealth Transfers, 1992-1995. *Statistics of Income Bulletin* (Winter 1996-1997): 8-63.
- Johnson, Barry W. and Paul B. McMahon. Using Auxiliary Information to Adjust for Nonresponse in Weighting a Linked Sample of Administrative Records. *Statistics of Income Bulletin* (Winter 2002).
- Johnson, Barry W. and Louise Woodburn. The Estate Multiplier Technique, Recent Improvements for 1989. In Barry Johnson. ed. *Compendium of Federal Estate Tax and Personal Wealth Studies*. Washington, DC: Department of the Treasury, Internal Revenue Service, Pub. 1773, 1994.
- Kennickell, Arthur. Using Income Data to Predict Wealth. Paper presented at the American Economics Association meetings (January 1999). Also found at <http://www.federalreserve.gov/pubs/oss/oss2/papers/wealth.income.7.web.pdf>.
- Kennickell, Arthur. Modeling Wealth with Multiple Observations of Income: Redesign of the Sample for the 2001 Survey of Consumer Finances. (October 2001). <http://www.federalreserve.gov/pubs/oss/oss2/papers/scf2001.list.sample.redesign.9.pdf>.
- Mishra, Ashok K., Hisham S. El-Osta, Mitchell J. Morehart, James D. Johnson, and Jeffrey W. Hopkins. Income, Wealth, and the Economic Well-being of Farm Households. *Agricultural Economics Report No. 812*. Washington, DC: U.S. Department of Agriculture Farm Sector Performance and Well-being Branch,

Resource Economics Division, Economic Research Service, July 2002.

Projector, Dorothy S. and Weiss, Gertrude S. Income-Net Worth Measures of Economic Welfare. *Social Security Bulletin* 32 (November 1969).

Radner, Daniel B. Alternate Estimates of Economic Well-being by Age Using Data on Wealth and Income. *Social Security Administration Office of Policy Office of Research and Statistics Working Paper Number 42* (March 1990).

Ringen, Stein. Direct and Indirect Measures of Poverty. *Journal of Social Policy* 17 (1988): 351-65.

Short, Kathleen and Patricia Ruggles. Accounting for Net Worth in Measuring Poverty. *Proceedings of the 28th General Conference of The International Association for Research in Income and Wealth* (2004). www.iriw.org.

Smith, James. Comment. In Martin David and Timothy Smeeding, eds. *Horizontal Equity, Uncertainty, and Economic Well-being*. Chicago: University of Chicago Press, 1985.

Steuerle, C. Eugene. Relationship Between Realized Income and Wealth, a Report from a Selected Sample of Estates containing Farms or Business. *Statistics of Income Bulletin* (Spring 1983).

_____. Wealth, Realized Income, and the Measure of Well-being. In Martin David and Timothy Smeeding, eds; *Horizontal Equity, Uncertainty, and Economic Well-being*. Chicago, University of Chicago Press, 1985.

Wolff, Edward N., Ajit Zacharias, and Asena Caner. Levy Institute Measure of Economic Well-being: United States, 1989, 1995, 2000 and 2001. Annandale-on-Hudson, N.W.: The Levy Economics Institute, 2004.

FIGURE 1

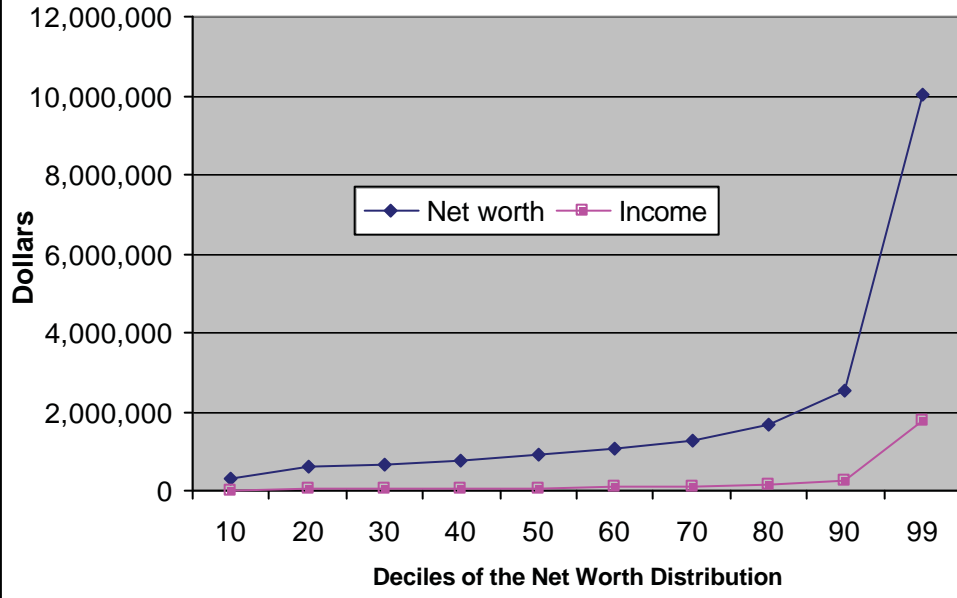


Table 1: Income and Wealth by Source

	Estate Tax Decedent Population Estimate			Estimate for Living Population		
	Average amount of item	Item as a percentage of Net income subject to tax	Net Worth	Average amount of item	Item as a percentage of Net income subject to tax	Net Worth
Income by source						
Wages and salaries	16,702	15.74	0.99	62,781	45.61	4.40
Dividends	20,209	19.04	1.20	12,195	8.86	0.86
Taxable interest	28,910	27.24	1.72	22,003	15.99	1.54
Tax-exempt interest	16,591	15.63	0.99	10,069	7.32	0.71
State tax refunds	831	0.78	0.05	1,046	0.76	0.07
Alimony	73	0.07	0.00	324	0.24	0.02
Schedule C income or loss	2,278	2.15	0.14	5,777	4.20	0.41
Net capital gain or loss	13,239	12.47	0.79	15,233	11.07	1.07
Capital gain distribution	65	0.06	0.00	42	0.03	0.00
Supplemental gain or loss	97	0.09	0.01	(1,035)	-0.75	-0.07
Pension & annuity income	7,095	6.68	0.42	8,138	5.91	0.57
Partnership and S-corp income	14,565	13.72	0.87	29,729	21.60	2.08
Rents, royalties, REMIC	6,567	6.19	0.39	8,474	6.16	0.59
Estate and trust income	1,110	1.05	0.07	770	0.56	0.05
Farm income	162	0.15	0.01	(403)	-0.29	-0.03
Reported other income, Social Security, unemployment comp.	4,779	4.50	0.28	248	0.18	0.02
Gross income subject to tax	109,093	102.79	6.49	146,351	106.33	10.26
Total adjustments	657	0.62	0.04	1,757	1.28	0.12
Net income subject to tax	106,135	100.00	6.31	137,633	100.00	9.65
Exemptions	2,946	2.78	0.18	3,813	2.77	0.27
Interest deduction	3,153	2.97	0.19	9,334	6.78	0.65
Other deductions (Standard Ded. or Itemized less mortgage int.)	22,692	21.38	1.35	20,194	14.67	1.42
1040 taxable income	96,045	90.49	5.71	127,247	92.45	8.92
	Average amount of item	Item as a percentage of total wealth		Average amount of item	Item as a percentage of total wealth	
Wealth by source						
Stock	421,610	23.86		254,487	15.81	
Closely held stock	150,026	8.49		199,561	12.40	
Personal residence	141,503	8.01		180,781	11.23	
Real estate	228,478	12.93		285,483	17.73	
Tax-exempt bonds	217,058	12.28		128,061	7.95	
Cash, bonds, notes and mortgages	362,225	20.50		250,740	15.57	
Noncorporate assets	46,305	2.62		69,086	4.29	
Other assets	202,546	11.46		241,424	15.00	
Total wealth	1,766,938	100.00		1,609,940	100.00	15
Debts	86,234	4.88		183,727	11.41	

Rates of Return to Capital, Estimates for the Living Population, 1992					
	\$600,000 under \$1 million	\$1million under \$5 million	\$5 million under \$10 million	\$10 million or more	All
Age					
Under 50	4.1	4.1	3.1	4.6	4.0
50 under 60	2.6	2.9	3.1	5.5	2.9
60 under 70	4.6	5.0	4.3	4.8	4.8
70 and older	8.4	5.1	6.0	5.7	6.9
All ages	5.0	4.2	4.2	5.2	4.6

	\$600,000 under \$1 million	\$1million under \$5 million	\$5 million under \$10 million	\$10 million or more	All
Age					
Under 50	2.6	3.3	2.8	7.3	3.0
50 under 60	3.2	2.6	3.7	5.0	2.8
60 under 70	4.4	5.8	5.1	4.8	5.2
70 and older	7.4	5.4	5.5	5.5	6.5
All ages	6.7	5.2	5.2	5.4	5.9

	Coefficients living populations	Coefficients decedent population	Means	S.D.
total assets			7542341	29912054
married	123755	-266952	0.52	0.5
age	**71893	90987	70.96	16.42
agesqrd	** -699	-795		
propadj	**728681	**559004	0.93	0.21
ncorpinc	**1.77	**2.01	46916	570688
divinc	**83.35	**67.51	93373	379337
farminc	-0.35	-0.16	-3042	216106
intinc	**9.37	**13.22	90982	354940
teintinc	** -5.51	** -2.39	77143	372086
cginc	-0.61	-1.32	71858	419944
penaninc	2.06	1.52	12620	87503
othinc	** -.70	** -1.57	2019	367147
nonkinc	**1.20	**5.81	72378	215246
intded	**7.63	**8.94	16541	174756
1/weight	** -1733084	-2386298		
Adj Rsqr	0.73	0.68		
N obs	3767	3767		

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

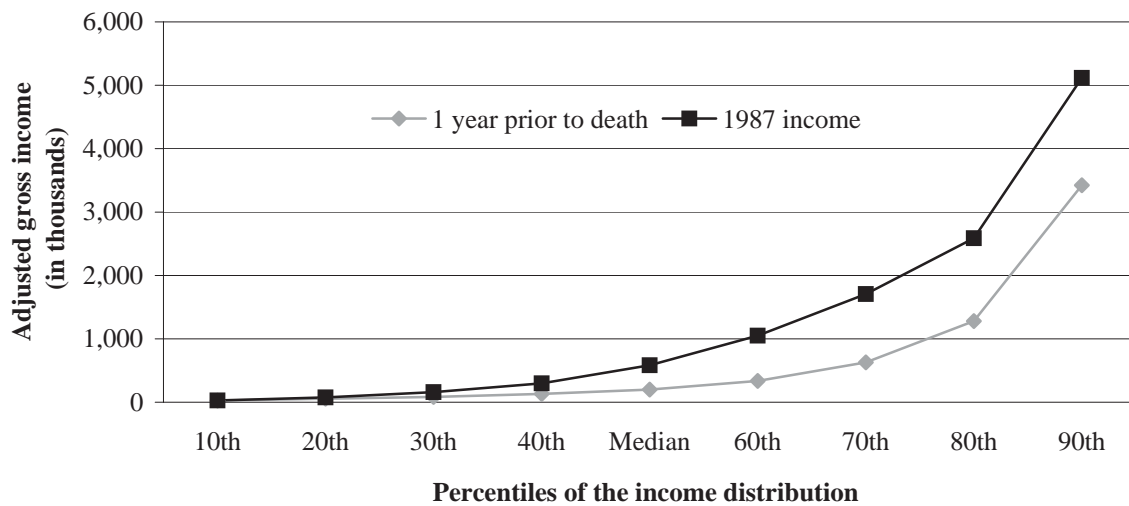
Despite the limitations and challenges discussed in the previous section, the FPDD gives a unique opportunity to learn more about the way that incomes change as people age and contemplate the end of their lives and also provides a snapshot of the wealth that was the source of a portion of that income. 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

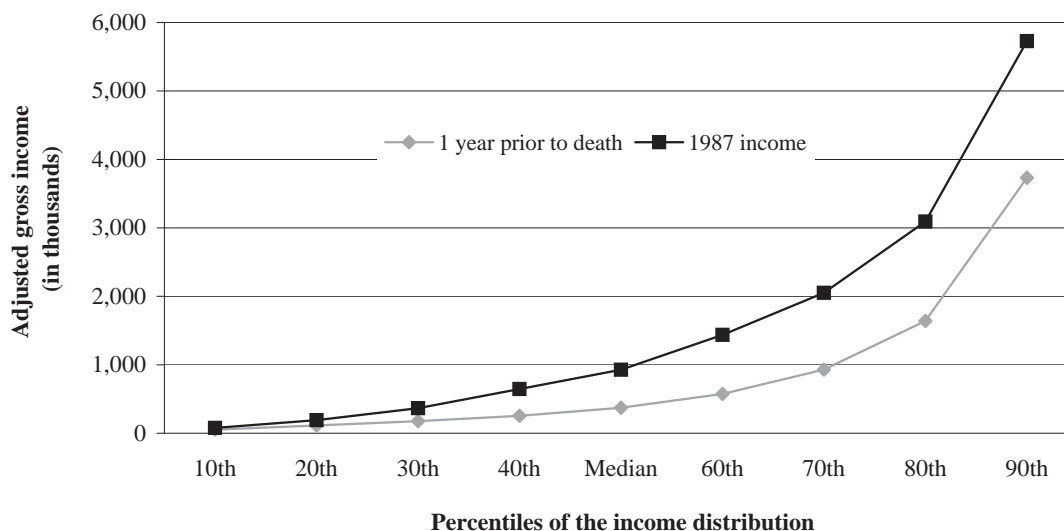
percent, of the 4,778 panel decedents who were joint filers in 1987 were male. The mean and median ages of females in the FPDD were 65 and 66, respectively, in 1987 and 76 and 78 at death. The mean and median age for males in 1987 were 63 and 64, respectively, and 75 and 76 at death. These statistics 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.

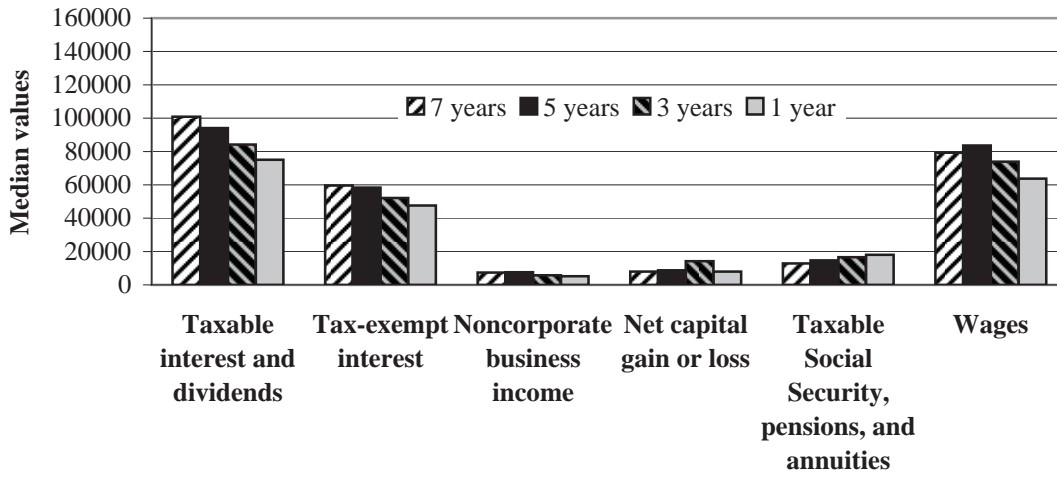
Figures 5 and 6 decompose AGI into major components for selected years over the 7-year period preceding a panel decedent's year of death [10]. For single filers, overall, median values for wages, taxable interest and dividends, and income from noncorporate businesses decreased as individuals aged. Median 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.

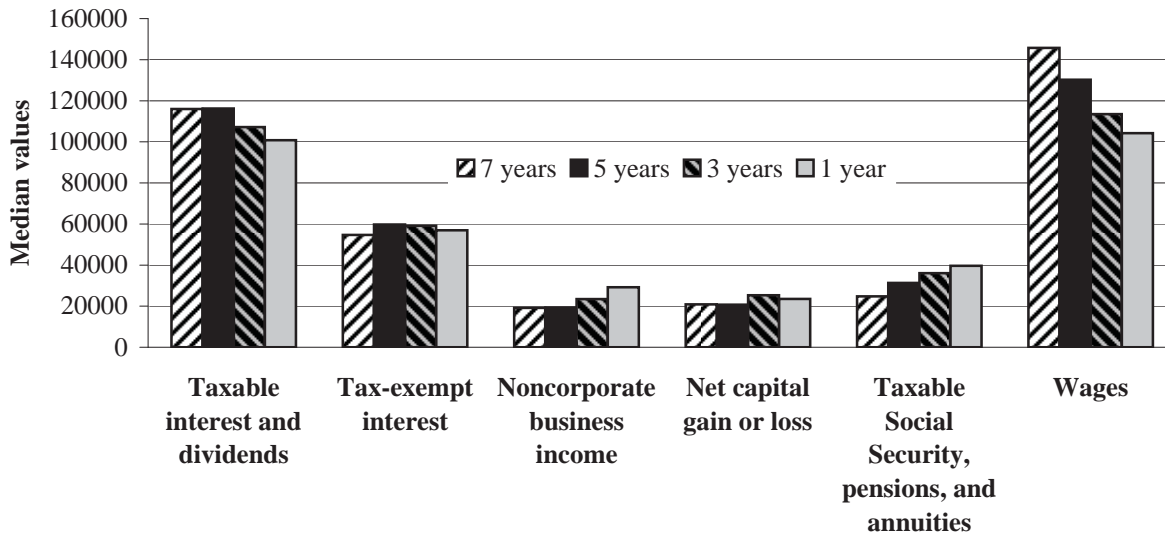
Figures 5-6 showed that, as panel members aged, the share that wage income contributed to AGI decreased, while the patterns of change in income from other sources varied somewhat, depending on filing status and wealth class. It has been noted that the realization of income derived from assets is a more 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].

► 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

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

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.

The Income-Wealth Paradox: Connections Between Realized Income and Wealth Among America’s Aging Top Wealth-Holders

*Barry Johnson, Internal Revenue Service;
Kevin Moore, Board of Governors of the Federal Reserve System;
and Lisa Schreiber, Internal Revenue Service*

Meaningful measures of individual economic well-being are essential for the equitable administration of Government social and economic policies. Realized income, which includes both wage and property income, is a frequently cited measure of both economic well-being and inequality, chiefly because wage income, the largest component for most households, is relatively easy to observe and measure (Steuerle, 1985). Some researchers, however, have argued that the “stock dimension” of asset ownership provides economic advantages, such as economic security, political privilege, and power that should also be considered in any study of well-being (Wolfe and Zacharias, 2006). Policymakers, the media, and the general public often incorrectly conflate income and wealth, using them interchangeably when trying to make inferences about the well-being of various segments of the population. This is particularly problematic because these two are not as closely correlated as is generally assumed, particularly among the very wealthy.

For the very wealthy, the discordant relationship between income and wealth is the result of the dynamic nature of the income reported by this segment of the population. Two studies using panel data from U.S. Federal income tax returns have shown that the composition of the group of individuals whose incomes place them near the top of the income distribution changes dramatically over time (Frenze, 1992; U.S. Treasury, 2007). The U.S. Treasury Department study found, for example, that fewer than half of those in the top 1 percent of the income distribution in 1996 were still in the top 1 percent in 2005. This volatility increased at the very top of the distribution, so that only about 25 percent of the individuals in the top 1/100th percent in 1996 remained in the top 1/100th percent in 2005. The Treasury report concluded that the income of many of the highest-income taxpayers is transitory and generally declines over time (U.S. Treasury, 2007).

The transitory composition of income quintiles over time can be partially attributed to decreases in wage income for individuals above retirement age. Also, for wealthier individuals, return on capital becomes an increasingly important source of income. For the very wealthy, however, income from capital can be particularly susceptible to manipulation to minimize tax liability. For example, it has been shown that rates of return on investments decline as wealth increases among the very wealthy (Steuerle, 1985; Wahl and Johnson, 2004). If this is the case, then, for these very wealthy individuals, measures of well-being that focus solely on realized income will understate their true economic status.

This paper is intended to add to the understanding of the ways in which income from various sources changes with age for the very wealthy. It makes use of a special longitudinal panel of U.S. income tax data linked to wealth data reported on U.S. estate tax returns filed for wealthy decedents. The relatively high estate tax filing threshold places these individuals at the top of the U.S. wealth distribution. Combined income and wealth data in the Statistics of Income Family Panel Decedent Dataset (FPDD) allow investigation of changes in the composition of realized income over time and also provide insights into asset management strategies employed by this elite group. In addition, this paper investigates the relationship between income and end-of-life wealth through the use of the portfolio data reported on the estate tax returns. Due to the limitations of the tax data, it incorporates data from the U.S. Survey of Consumer Finances to estimate these panel members' place in the overall U.S. distributions of income and wealth.

Tax Return Data

The Statistics of Income Division (SOI) of the United States Internal Revenue Service collects statistical data from most major Federal tax and information returns. These data are used by both the U.S. Congress and the Executive Branch of the Government to evaluate and develop tax and economic policy. Among these are annual studies of the *United States Estate (and Generation-Skipping Transfer) Tax Return* (Form 706) and the *U.S. Individual Income Tax Return* (Form 1040).

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.¹ The return must be filed within

¹ The estate tax filing thresholds for 1994–2003 are listed in Table 1.

9 months of a decedent's death, although a 6-month extension is frequently granted. All of a decedent's assets, as well as the decedent's share of jointly owned and community property assets, are 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.

Form 1040 is filed by individuals or jointly by couples to report annual income, including wages, interest, dividends, capital gains, and some types of business income. The Statistics of Income Division of the Internal Revenue Service conducts annual studies of these filings, extracting detailed information from a statistical sample of returns as they are filed and producing microdata sets and tabulations that are widely used to evaluate and manage the U.S. tax system and the economy. The SOI stratified sample design oversamples high-income taxpayers to ensure accurate estimates of the often unique financial characteristics of this elite group. In 1987, SOI incorporated a panel component, the Family Panel, into its annual cross-sectional samples in order to include all members of a tax family (primary and secondary filers and their dependents) in a panel that represented the cohort of tax families filing returns in 1988 for Tax Year 1987 (Schirm and Czajka, 1991). For the initial year, the Family Panel included 89,755 returns, not counting returns filed by dependents.

The Tax Family Concept

The unit of observation for the SOI 1987 Family Panel was defined as a tax family, which included an income taxpayer, spouse, and all dependents (not limited to children) claimed by either. Thus, a tax family could represent single income tax filers, as well as joint filers and their dependents.² An interesting complication of the tax family concept is the treatment of married couples who, for various reasons, elected to file income taxes separately. For the purposes of the followup in the later years of the panel, only a partner whose separately filed return was selected into the 1987 panel sample was permanently 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

² 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. No dependents are included in the analysis presented in this paper.

independently selected. Thus, the tax family differs significantly from the more common “household” measure used by many national surveys (Czajka and Schirm, 1993).

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 family-owned 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.

Survey of Consumer Finances

The Survey of Consumer Finances (SCF) is a survey of household balance sheets conducted by the Board of Governors of the Federal Reserve System in cooperation with the SOI division of the IRS. Besides collecting information on assets and liabilities, the SCF collects information on household demographics, income, relationships with financial institutions, attitudes toward risk and credit, current and past employment, and pensions (Bucks; Kennickell; Mach; and Moore, 2009).

The SCF uses a dual frame sample design to provide adequate representation of the financial behavior of all households in the United States. One part of the sample is a standard multistage national area probability sample (Tourangeau et al., 1993), while the list sample uses the SOI individual income tax data file to oversample wealthy households (Kennickell, 2001). Wealth data from the SCF are widely regarded as the most comprehensive household-level data available for the United States. Sample weights constructed for the SCF allow aggregation of estimates to the U.S. household population level in a given survey year (Kennickell and Woodburn, 1999; Kennickell, 1999).

The Data

Starting in 1994, the sample for SOI’s annual estate tax studies included any Form 706 filed for a deceased 1987 Family Panel member. The Family Panel Decedent Dataset (FPDD) was begun in 1994 as a combination of these estate tax returns and their corresponding individual income tax return

data. Between 1994 and 2003, there were 5,557 estate tax returns identified as having been filed for 1987 Family Panel members who died.³

The FPDD includes income data spanning 1987 to 2003 and estate tax data ranging from 1994 to 2003.⁴ A total of 72,373 income tax returns were available for the members of FPDD. Table 1 presents the distribution of decedents by year of death, along with the applicable estate tax filing threshold. The rightmost column shows only those 5,162 decedents whose gross estates at the time of death were at least \$1 million in constant 2003 dollars and for whom a Form 1040 was filed in the last year prior to death.

For 98.2 percent of decedents captured in the FPDD, income tax data were available for each tax year between 1987 and the last full year prior

Table 1. Filing Threshold and Number of Decedents, by Year of Death

Year of Death	Number of decedents	Filing threshold in nominal dollars	Number of decedents with assets of \$1M or more in constant 2003 dollars
1994	417	600,000	385
1995	480	600,000	440
1996	521	600,000	478
1997	574	600,000	520
1998	538	625,000	487
1999	635	650,000	586
2000	609	675,000	559
2001	667	675,000	605
2002	636	1,000,000	630
2003	480	1,000,000	472
Total	5,557	N/A	5,162

³ An additional 755 Estate tax returns were filed for decedents who died prior to 1994, the date that SOI began collecting these data for panel members, so that these decedents are excluded from this analysis. Estate returns of visitors to the panel (individuals who were married to existing panel members for periods after 1987) were not included in the final dataset since income data were only available for those years that they were associated with an original panel member. Estate returns of dependents were also excluded.

⁴ Up until 1996, individual income tax data were collected and edited by SOI. Starting in 1996, a reduced set of data collected by IRS for administrative purposes was available. These data were not subject to the edit review that is routinely part of SOI data collection and may be subject to additional nonsampling error and subtle differences in data definitions (see Johnson and Schreiber, 2006).

to death. For an additional 1.3 percent of all decedents, only one return was missing from this time series, leaving only a handful of decedents for whom more than one return was missing from the panel.⁵

The design of the FPDD poses several analytical challenges. Longitudinality introduces problems with the tax family concept because, over time, a filing unit may change composition, and this change is usually accompanied by changes in filing status (Czajka and Schirm, 1993). In addition, the selection criteria for inclusion in the FPDD changed during the sample period due to changes in the estate tax filing threshold. Another important consideration is that an estate tax return includes only a decedent's share of a married couple's assets, while income tax returns for married couples who file jointly report income attributable to both partners. Finally, with a few exceptions, such as tax-exempt interest income, 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.

Although the income tax filing status reported for members of the FPDD was much more stable over time than that of the general population, changes are inevitable. In particular, married persons may divorce, single persons may marry, couples who customarily file jointly may elect to file separately or vice versa, or one or both spouses of a married couple may die. The longer the time series, the greater the possibility for one of these events to occur. Table 2 shows panel members for whom a tax return was filed in the last year prior to death and compares each panel member's filing status in the year prior to death with that reported for earlier

Table 2. Filing Status Stability of Panel Members for Whom a Form 1040 was Filed 1 Year Prior to Death

Includes only those panel members who died between 1994 and 2003 with gross assets valued at \$1 million or more in constant 2003 dollars

Filing Status	Number	Number of years prior to death filing status unchanged			
		3	5	7	9
Single	1,688	1,421	1,230	1,062	766
Joint	3,474	3,399	3,343	3,305	2,693
Total	5,162	4,820	4,573	4,367	3,459

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 Social Security number (SSN)—a unique personal identifier used for tax administration. 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 greatly reduced the chances of SSN errors in the data.

tax periods. Filers are grouped into two broad categories, single filers and joint filers.⁶ Using this classification, filing status was constant for 67 percent of all panel members over the 9 years preceding death. Individuals who were single filers at death were much more likely to have changed filing status in the years preceding death than those who were joint filers. Only 45 percent of all individuals who were single filers in the year prior to death had been single over at least the 9 years examined. This result is influenced by couples for whom one spouse died and those who divorced or separated during the period. Of individuals who were joint filers at death, 78 percent had been married for at least the previous 9 years. Filing status was significantly more static over the 7 years preceding death for both groups, with no change for 85 percent of all filers, 63 percent of single filers, and 95 percent of joint filers. This paper focuses on filers with constant filing status for the 7 years prior to death and at least \$1 million (in constant 2003 dollars) in gross wealth as reported in estate tax filings.

Income Components

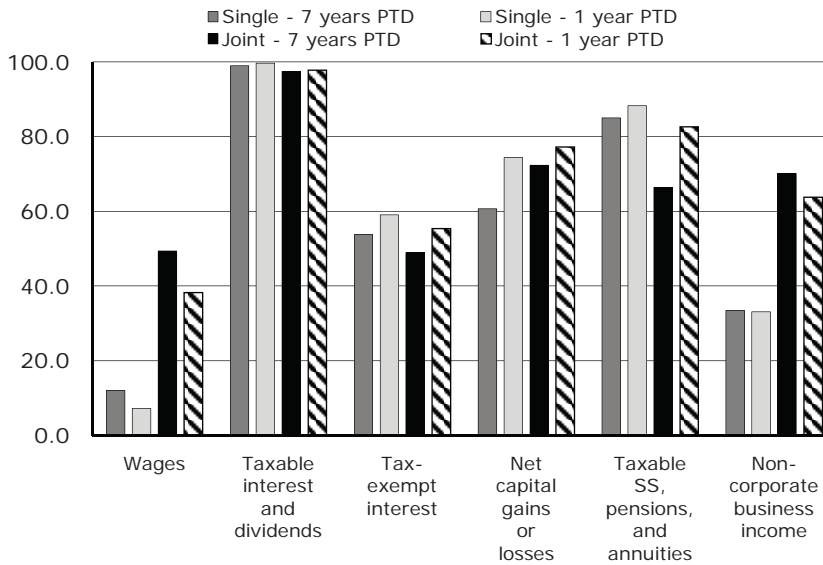
The filers in the sample used in this analysis are a very selective slice of all taxpayers in any given year. Many members of the sample have a high level of total income, but, owing to the nature of the sample selection, it is difficult to gauge where these filers fall in the overall distribution of income. One possibility is to compare weighted mean total income by year in the FPDD to the distribution of a comparable total income measure constructed from SCF data.⁷ The comparison reveals that weighted mean total income by year from the FPDD is above the 95th percentile of the SCF income distribution in each year in which the two data sources overlap (Tax Years 1988, 1991, 1994, 1997, 2000, and 2003).⁸

Figure 1 provides some basic information on the fraction of filers with different types of income, by the number years prior to death. The most striking point to note from this Figure, but hardly surprising, is the extremely high incidence of income derived from various assets, regardless of filing status or the number of years prior to death. Over 96 percent of both types of filers have taxable interest and dividend income, and about one-half have

⁶ The category "single" includes individual income tax return filers who were unmarried, widowed, or married but filing separately.

⁷ All estimates are weighted using weights that reflect the original family panel selection probabilities of the primary and, if present, secondary filer. All dollar values are reported in constant 2003 dollars.

⁸ In comparable years, weighted median total income in the FPDD falls between the 70th and 90th percentiles of the SCF income distribution.

Figure 1. Percentage of Filers with Various Types of Income

tax-exempt interest income. For single filers, about 65 percent have net capital gains or losses. Over 70 percent of joint filers report this type of income. About 35 percent of single filers and 65 percent of joint filers also receive income from noncorporate businesses. Given that the average age at death in the sample is 77, it is not surprising that taxable Social Security, pension, and annuity income is common among both groups of filers, while wage income is the least common type of income received.

Figures 2a–c present the (unconditional) mean values of various types of income by filing status, years prior to death, and end of life wealth category.⁹ The most striking feature of the Figures is the difference in mean total income across wealth groups. Depending on filing status and number of years prior to death, mean total income is 5 to 10 times larger for the \$10 and \$20 million wealth group (Figure 2b) than for the less than \$10 million wealth group (Figure 2a). Somewhat smaller differences exist between the middle and the top wealth groups. Mean total income for the \$20 million or more wealth group (Figure 2c) is only 2 to 6 times larger.¹⁰

The Figures also reveal that income derived from taxable interest and dividends, tax-exempt interest, and capital gains is an important source

⁹ Gross estate valued on the date of a decedent's death is used as the measure of wealth throughout this analysis.

¹⁰ Similar results are found when comparing the median and the 75th and 95th percentile values of total income across wealth groups.

Crossing the Bar: Predicting Wealth from Income and Estate Tax Records

Lisa Schreiber Rosenmerkel and Jenny Wahl

Draft 6-14-2011

Clarifying the connections between income and wealth is essential for ascertaining individual economic status and establishing informed policy. For the very rich, realized income may reveal little about true well-being because tax liability can drive decisions about form and timing. Fortunately, estate tax records provide a superb additional source of information.

We have constructed a unique data set that links together several years of U.S. Individual Income Tax Returns (Forms 1040) for persons who died between 1996 and 2002, as well as the U.S. Estate and Generation–Skipping Transfer Tax Return (Form 706) when the decedent’s estate size exceeded the filing threshold.¹ The included individuals were members of a panel representing the cohort of tax families (primary and secondary filers and their dependents) that filed Form 1040 in Tax Year 1987.² We use the data to do three things: (1) predict the probability of filing a Form 706 from Form 1040 information,³ (2) estimate individual wealth from Form-1040 data via a Heckman two-step approach that corrects for selection bias, and (3) outline an approach for predicting the amount of total gross estate that will ultimately appear on Forms 706.

¹ Jacobson et al. (2007) describes the tax treatment of estates. Total gross estate includes the decedent’s assets plus the relevant share of jointly owned and community property. Property over which the decedent had a general power of appointment, most life insurance proceeds, and certain inter vivos transfers are also included.

² A tax family is defined as the taxpayer, spouse, and all dependents as claimed on the Form 1040. Married couples who elect to file separately are treated as two distinct tax families. Only the partner whose return was selected into the sample was included in the panel. As a result, the tax family differs significantly from the more common “household” measure used by many national surveys (Czajka and Schirm 1993).

³ An alternative is predicting the probability of surpassing a given wealth threshold. The threshold for filing estate tax returns changed over time – \$600,000 for 1996-97, \$625,000 for 1998, \$650,000 for 1999, \$675,000 for 2000-01, and \$1 million for 2002. We report results for both the probability of filing and of exceeding the largest threshold of \$1 million (in real dollars).

I. A Behavioral Model: Linking Theory to Estimation

A simple model of individual choice might look something like this

$$\text{Max } U = f(C, L, G) \quad (1)$$

subject to

$$C + G = wH + N \quad (2)$$

and

$$T = H + L, \quad (3)$$

where U is lifetime utility, C lifetime value of consumption, L lifetime leisure hours, G lifetime value of gifts bestowed (including bequests), H lifetime work hours, w hourly wage rate, N lifetime nonwage income (including gifts received and inheritances), and T length of life. We can express the change in wealth at time t (ΔW_t) and wealth at time t (W_t) as follows:⁴

$$\Delta W_t = w_t H_t + N_t - (C_t + G_t) \quad (4)$$

$$W_t = \Delta W_t + \Delta W_{t-1} + \Delta W_{t-2} + \dots \quad (5)$$

In other words, the change in wealth in period t equals income in period t less consumption plus gifts bestowed at time t . Wealth at time t is the sum of net additions to wealth up to that point (including any unspent inheritances and gifts received).

Equations (4) and (5) are basically accounting identities; the amounts people choose to add to wealth and realize as income in any period are of course functions of tax treatment.

Modeling what occurs the last period of life (T) brings home the messiness of uncertainty: a person may not exhaust all resources because he or she mispredicts the date of death. The decedent-to-be may also wish to influence the behavior of potential

⁴ ΔW_t can be negative and, in theory, so could W_t . Real-world policy regarding bankruptcy and debt, particularly at the end of an individual's life, could add more constraints.

heirs.⁵ This process is complicated by any desire to minimize overall tax liability, which can depend on policy regarding capital gains, gifts, and estates as well as income.

Consequently, W_T is likely the outcome of a complex interaction of constraints and preferences about own consumption, own leisure, gift-giving, bequests, tax avoidance, and mistakes. Nonetheless, equations (2), (4), and (5) make clear that wealth at any point in time, including the time of death, is related to prior income:

$$W_T = f(w_T H_T, w_{T-1} H_{T-1}, \dots, N_T, N_{T-1}, \dots). \quad (6)$$

Transforming equation (6) to something that could be estimated using tax data requires us to face several issues: (1) complete lifetime income information for a reasonably sized sample of individuals simply is not available, in tax data or anywhere else, (2) tax considerations could imply different allocation and timing of income for people at different points in the life cycle and in different wealth categories, (3) end-of-life wealth amounts are available from tax data only for decedents whose estates surpass a certain threshold size – that is, people had to cross two bars before we could observe their wealth, and (4) the income-wealth relationship for the living population could vary from that associated with a group of decedents. These are significant, but not insurmountable, issues.

To address the first two, we include demographic and portfolio data from a given year's income tax return for each individual.⁶ Variables indicating age, gender, and filing

⁵ For a discussion of the strategic bequest motive, see for example Bernheim et al. (1985).

⁶ We have experimented with using income information from a fixed number of years before death, as well as with income information from several years. See Johnson et al. (2009) for information about income patterns for several years prior to death. Adjusted gross income (AGI) is highly correlated across years for decedents whose estates did not file a 706 return, but less so for estate-tax decedents. What is more, the relationship of gross estate to AGI of whatever year also depends upon macroeconomic conditions. As a consequence, the choice of year for AGI observation could indeed matter for estimation.

Yet our research is motivated in part by the desire to predict the probability of filing an estate tax return (or exceeding a given wealth threshold) from given income tax information but unknown date of

status act as proxies for life-cycle differences.⁷ Breaking income into its components – salary, dividends, tax-exempt interest, and the like – allows us to evaluate coefficients and assess implied rates of return on different asset types. This may help us ascertain whether variations in tax treatment affect portfolio composition across different wealth categories, at least indirectly. Equation (6) thus becomes something like this:

$$W_T = \alpha + \beta_i Y_i + \gamma_j D_j + \varepsilon, \quad (7)$$

where Y_i refers to the i th component of income and D_j refers to the j th demographic trait.

James Heckman's seminal research provides a method for us to overcome the third issue: selection bias associated with truncated data.⁸ This procedure first estimates a probit model of the form

$$X = a + b_i y_i + c_j d_j + e, \quad (8)$$

where $X=1$ when an estate tax return is filed (or a given wealth threshold is exceeded) and 0 otherwise, and vectors y and d include relevant income-tax and demographic

death. We decided that this objective is best served by using a cross-section of income-tax returns rather than one that requires knowledge of death dates.

The second experiment – using multiple years of income-tax data – is part of our continuing work. Here, we have to grapple with the issue of using different numbers of years of information for decedents (depending on the year of death) versus using a given number of years for everyone, the latter of which implies again that we know the year of death. Because many components of income are highly correlated across years, we hope that inferences from a single year of income information will hold if we instead use multiple years.

⁷ The work presented here distinguishes people married and filing jointly in 1993 from those who were single, married filing separately, widowed, or separated. We have done some preliminary work that uses filing status information from multiple years but have not yet settled on the best way to incorporate this into the analysis. Estate wealth pertains to the decedent, but income reported on a 1040 could include spousal wages, non-labor income from jointly owned assets, and the like. We have experimented with different ways to cope with this – assigning half the income in the case of joint filers, for example, and analyzing long-married and long-single persons separately. The clearest way to present our current work is simply to include gender and filing status as of 1993 in our regression analysis.

⁸ Heckman (1979). An alternative is the Tobit model, discussed in Tobin (1958). Because one of our tasks is to use a probit model to determine the characteristics of filers, using the Heckman technique to estimate wealth grows naturally out of those results. We plan to use the Tobit model in future research as a check on our findings here.

information.⁹ The predicted values from the probit regression are then used to construct an inverse Mills' ratio λ to correct for selection bias in equation (6):

$$W_T = \bar{\alpha} + \bar{\beta}_1 Y_i + \bar{\gamma}_j D_j + \phi \lambda + \varepsilon. \quad (9)$$

The resulting coefficients on Y_i and D_j are unbiased, after correcting standard errors for heteroskedasticity.

What remains to be addressed is the fit of the model for the general population. Our data require us to include only people known to have died, because only they report wealth information that we can observe via Form 706. One way to cope with this is simply to assume that our data are representative as to mortality rates. Then we could estimate wealth for the decedent population using the inverse probabilities of death for particular age groups to obtain wealth for the living population. Because unweighted sample sizes are fairly small, however, this may not be appropriate for age groups in which mortality rates are quite low.¹⁰ What is more, anticipation of death could encourage some individuals – the very old, for instance – to adjust spending patterns to reflect the decreased uncertainty about the end of life.

Besides estimating equation (9) for all decedents in our sample, we therefore also estimate it for two subsamples: decedents who very likely would have died between 1996 and 2002 simply because humans have limited lifespans, and decedents who had a relatively low *ex ante* probability of death. That is, we focus on groups of very old and

⁹ We use lower-case letters for the explanatory vectors because this step must naturally include at least one identifying variable not included in the second step.

¹⁰ The distribution of AGI for the living population (obtained at <http://www.irs.gov/pub/irs-soi/06in05tr.xls>) is fairly similar to that for our decedent population, although the age distribution is undoubtedly different. This is not terribly surprising, as the sample is intended to represent the income-tax-filing population, which hopefully includes representativeness in terms of mortality rates. Even so, we cannot be sure that “representativeness” extends to the income-wealth relationship for the entire population, in part because relatively few young people die and even fewer leave large estates.

fairly young decedents to ascertain whether results using these data differ appreciably from those generated by the full sample.

II. Data

The data used here consist of 8,942 individuals who filed a Form 1040 in 1993 and died between 1996 and 2002; 4,226 of these decedents also had wealth that exceeded the Form 706 filing threshold.¹¹ By weighting the sample to reflect the population,¹² we find that the number with a Form 706 constitutes about 8.8 percent of decedents.¹³

We obtained the original income-tax data from panel data collected by the Statistics of Income (SOI) Division of the Internal Revenue Service. Each year, SOI obtains a stratified sample of income-tax returns. Several years ago, SOI incorporated into its annual cross-sectional sample a panel component that represents the cohort of tax families filing Form 1040 for the 1987 Tax Year.¹⁴ These data are called the 1987 Family Panel and consisted initially of 89,755 returns.

¹¹ The correlation between AGI and estate wealth is largest for AGI reported 4 years before death for the full sample, 5-6 years for the old subsample, and 3 years for the young subsample. Consequently, we did not want to use income information from the earliest filing years – 1992 is 10 (or possibly 11) years before the date of death for our latest-dying individuals, for example. And Form 1040 filed in the year of death typically reports activity for only part of a year. We therefore did not want to use Forms 1040 filed for Tax Year 1996 or later. What is more, some decedents did not have to file a Form 1040 in the year before death because much of their income went toward deductible medical expenses. This left us with two good possibilities for income-tax filing years: 1993 and 1994. We report results using Form-1040 information from Tax Year 1993 here; results using Forms 1040 from Tax Year 1994 are not substantially different and are available from the authors.

¹² Choudry (2001a, 2001b, 2001c), Czajka and Schirm (1991) and Schirm and Czajka (1992) offer additional information about the weights used in the sample. The weights are the inverse of the probability of initial selection into the 1987 Family Panel. They therefore do not easily translate into something with precise meaning for our sample. Although not optimal for our purposes, the weights still indicate something reasonable about the number of individual filers that a given observation represents. We therefore report all results using these weights, as the unweighted sample is far from representative.

¹³ Jacobson et al. (2007) report that fewer than 2 percent of the estates of adult decedents filed a Form 706 from 1916 to 2004, although the figure grew considerably in the 1990s, which may help explain the agitation for reform and ratcheting up of the estate tax filing threshold after 1997.

¹⁴ For additional description, see Schirm and Czajka (1992), Nunns et al. (2008), and Johnson and Schreiber (2006).

Starting in 1994, SOI began to include in its annual sample of estate tax returns all Forms 706 for 1987 Family Panel members who died in that year. Between 1994 and 2003, SOI gathered 5,557 estate tax returns for 1987 Family Panel members.¹⁵ Over the last two years, we have also identified 1987 Family Panel members who died between 1996 and 2002 but whose estates were not required to file a Form 706.

We have constructed two datasets, one based on income tax returns and the other on individual decedents. Each observation in the return-based data set consists of information collected from a given Form 1040 (plus additional data from the corresponding Form 706 where available) filed between 1988 and the year of death for a Family Panel member who died between 1996 and 2002. Each observation in the individual-based data set encompasses information on all the Forms 1040 (and Form 706 where present) for a given Family Panel decedent. Table 1 summarizes the number of observations (unweighted) in each data set.

TABLE 1: Number of observations in two relevant datasets

	Estate ≥ Form 706 Filing Threshold	Estate < Form 706 Filing Threshold	Total
1040-return-based data (1987-year of death)	55,160	60,061	115,221
Individual-based data			
death year 1996	522	557	1,079
death year 1997	575	564	1,139
death year 1998	541	618	1,159
death year 1999	642	647	1,289
death year 2000	636	700	1,336
death year 2001	668	773	1,441
death year 2002	645	873	1,518
Total 1996-2002	4,229	4,732	8,961

¹⁵ See Johnson and Schreiber (2006) and Johnson et al. (2009) for additional discussion of the Family Panel Decedent Dataset.

The decedents in the individual-based data set ranged in age from 14 to 99.4 years in 1993.¹⁶ By inspecting the distribution of death ages, we find that 10 percent of decedents died by age 52.2, 25 percent by age 65.1, 90 percent by age 89.5, 95 percent by age 92.5, and 99 percent by age 97.4.

A rough method of obtaining a subsample of Family Panel members who very likely would have died between 1996 and 2002 is to put a lower bound on the age of a person filing a Form 1040 in 1993. For our old “likely-to-have-died” sample, we chose a cutoff age of 82.8 for Filing Year 1993. If these individuals filed the 1993 Form 1040 in timely fashion, they would have reached age 92.5 (or the 95th percentile) by the end of Filing Year 2002. This unweighted subsample consists of 956 individuals. For our young “unlikely-to-have-died” sample, we include individuals reporting an age of 55.4 or younger for Filing Year 1993. This unweighted subsample includes 1,509 individuals.

III. Variable Choice and Regression Results for Individuals

The selection issue revisited

Two of our analytical tasks are determining how best to predict from detailed Form-1040 information the likelihood of a later Form-706 filing and to model the relationship of income to wealth for decedents.¹⁷ The first is likely a prerequisite for the second, because unobservable factors affecting the probability of a decedent’s surpassing the Form-706 filing threshold could reasonably affect the size of the estate as well. Call

¹⁶ Because income can fall short of the Form 1040 filing threshold, not all decedents filed an income tax return in every year. Of the 8,961 individuals who died between 1996 and 2002, for example, only 8,942 filed a Form 1040 in 1993.

¹⁷ We cannot include in our analysis persons who died in the period 1996-2002 but did not file a Form 1040 in Tax Year 1993 – those omitted may include elderly or retired persons whose 1993 income falls below the zero-bracket amount but who still may have assets. We speculate that omitted persons are unlikely to fall in the upper part of the wealth distribution, however, which will probably be the focus of any analysis using these data.

the individuals whose estates file an estate tax return “F706 decedents” and those whose estates fall short of the estate-tax filing threshold “non-F706 decedents.” Once we have a method of determining how an eventual decedent “selects into” F706 status, we can use this to correct for selection bias in a regression that has wealth information only for F706 decedents, as outlined in equations (8) and (9) in section I. The resulting unbiased coefficients can in turn help us predict wealth for non-F706 decedents.¹⁸

Variation in tax treatment for income earned from different sources means that certain types of assets may appear more attractive to wealthier taxpayers. Assets that generate tax-exempt income or unrealized capital gains (or realizable capital losses) might particularly appeal to richer individuals. Thus, both the presence of a Form 1040 item and its size may help predict the probability of filing a Form 706 and the size of total gross estate.

Descriptive information

Charts 1 and 2 display information about the presence and average size (in \$2001) of various Form-1040 items (for Filing Year 1993) for F706 and non-F706 decedents.¹⁹ Virtually all members of both groups report adjusted gross income (AGI). But whereas nearly half of F706 decedents report tax-exempt interest, for example, fewer than 10 percent of non-F706 decedents do.²⁰ Mean real AGI is over \$128,000 for F706

¹⁸ The techniques we currently have available may not be as refined as we would like. Quantile regression, for example, would allow us to estimate the median (or other quantiles); it is more robust than OLS regression when outliers are present. See for example Hao and Naiman (2007). We hope to extend our analysis using quantile regression analysis once we have the requisite computing software.

¹⁹ Because our Form 706 information comes from returns filed in different tax years, we converted all dollar amounts to dollars of a given year (2001).

²⁰ For the old subsample, just over 60 percent of F706 decedents report tax-exempt interest income in 1993, whereas only about 13 percent of non-F706 decedents report this item of income. For the young subsample, the figures are 22 and 1.5 percent, respectively.

decedents, about 5 times the mean real AGI for non-F706 decedents.²¹ The numbers next to the y-axis labels indicate the ratio of the relevant figures for F706 and non-F706 decedents. For example, 6.64 times as many F706 decedents report tax-exempt income as non-F706 decedents. But the average amount of tax-exempt income by F706 decedents is over 33 times that reported by non-F706 decedents.

CHART 1: Presence of Form-1040 items, by Form-706 filing status

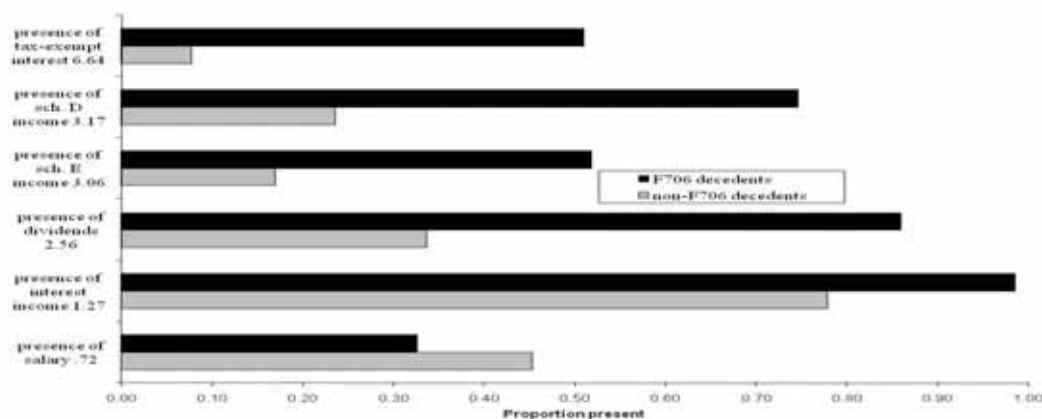
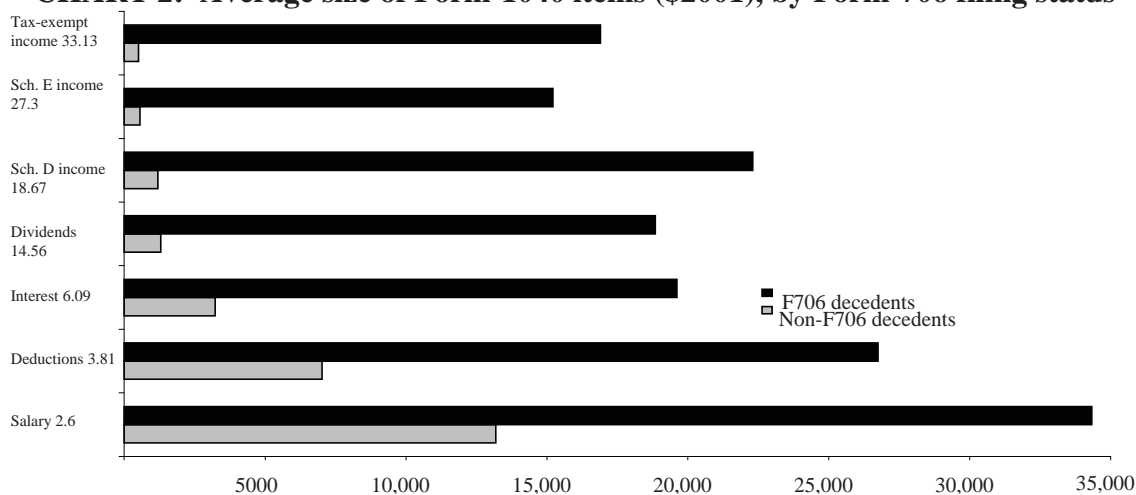


CHART 2: Average size of Form-1040 items (\$2001), by Form-706 filing status



²¹ Mean real AGI in 1993 for elderly F706 decedents is 4.24 times that for non-F706 decedents; the figure is 5.82 for the young subsample. We do not show AGI on chart 3 because it is so large relative to the other items, making the chart difficult to read.

We have also examined other moments of the distribution, particularly variance and skewness, for several variables. In sum, F706 decedents exhibit higher mean values for many variables (particularly tax-exempt interest and dividend income) as well as greater variability.

Chart 3 indicates the averages of ratios of particular line items to AGI for F706 and non-F706 decedents for the full sample. Note particularly the differences for deductions and for tax-exempt income. The average deductions-to-AGI ratio for non-F706 decedents is 0.795; it is only 0.242 for F706 decedents. The respective figures for the tax-exempt-income/AGI ratios are 0.016 and 0.204.

CHART 3: Averages, ratio of Form-1040 items to AGI, by Form-706 filing status

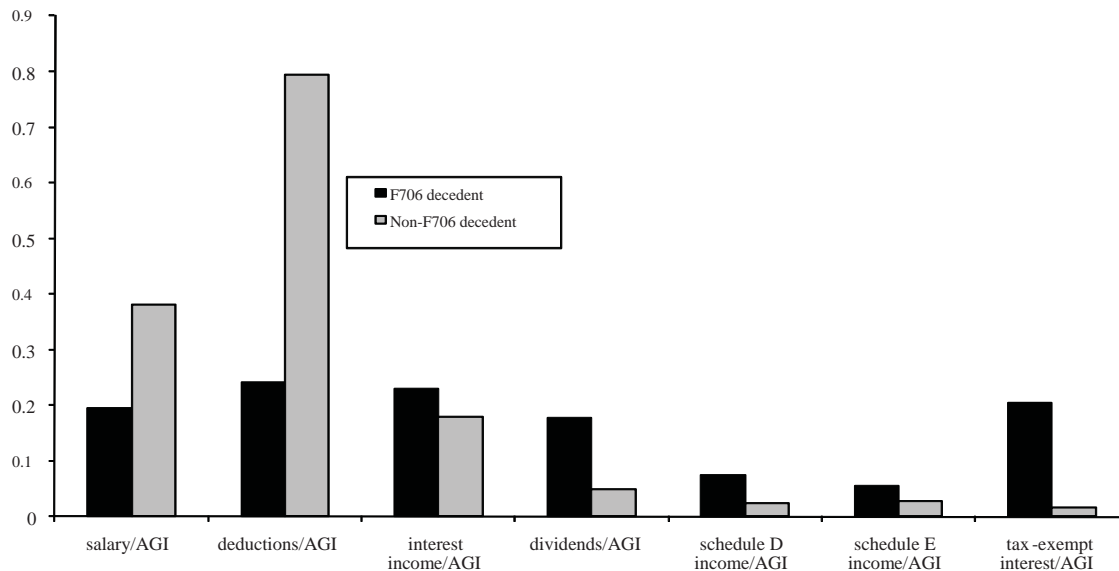


Chart 4 shows the F706 decedent/non-F706 decedent ratios of the proportions shown in Chart 3 for the full sample and the two subsamples for Forms 1040 that reported positive AGI. The F706 decedent/non-F706 decedent ratio for deductions/AGI for the full sample equals 0.30 (0.253/0.856), for instance, and the full-sample ratio for tax-exempt income/AGI equals 13.1 (0.21/0.016). The difference between F706 and non-F706 decedents is especially notable for the young subsample with respect to dividends, capital gains and losses as reported on Schedule D, and tax-exempt interest income. Young F706 decedents with positive AGI in Filing Year 1993 exhibit nearly 16

times the average dividend/AGI ratio as their non-F706 counterparts, for instance, whereas the same figure is only 1.87 for the old subsample and 3.45 for the full sample.

CHART 4: Ratios, F706 decedents to non-F706 decedents (AGI >0)

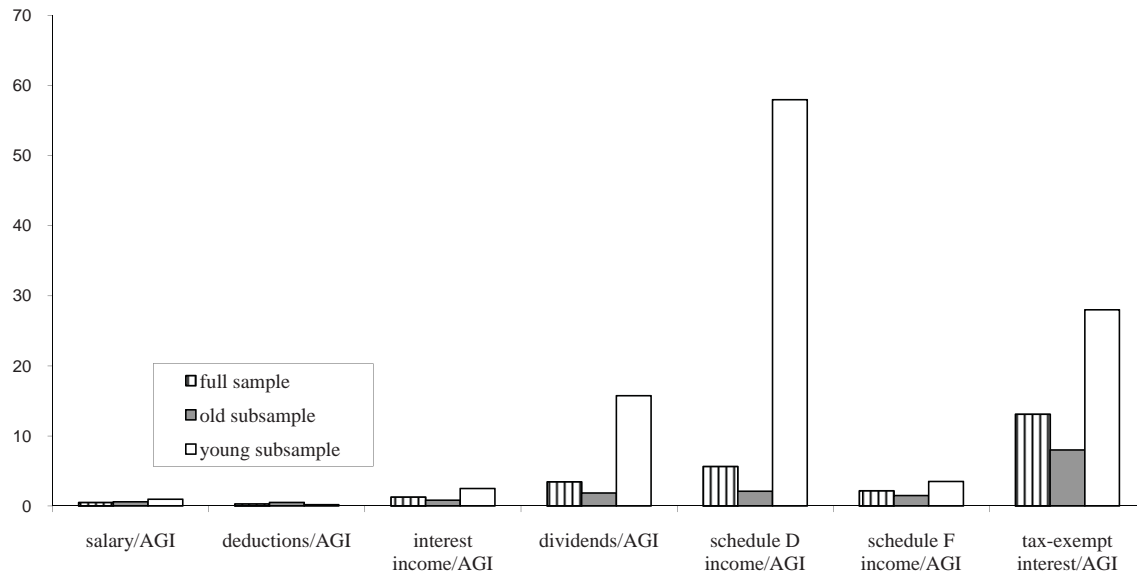


Table 2 gives the percent married filing jointly in 1993 for various demographic groups. The smallest numbers are for females in the old subsample. This reflects the relatively longer lifespan for females and the small average gap in ages for married couples in the U.S.²² Interestingly, the percentages are all fairly close for men in the two subsamples, although older unmarried men are probably more often widowed whereas younger unmarried men may be more likely to be divorced or never married, relative to the full sample. Another intriguing pattern is the gender difference between F706 and non-F706 decedents in the young subsample. The percent for male F706 decedents is only 8 percentage points larger than for non-F706 decedents, but the corresponding figure is 27 percentage points for females. Although we decline to speculate, we find it

²² For data on U.S. lifespans, see Shrestha (2006). The average spousal age gap in the U.S. has fallen from about 5 years in 1900 to about 2 years in 2000. Rolf and Ferrie (2008).

fascinating that such a large proportion of relatively young wealthy female decedents were married.²³

TABLE 2: Percent married filing jointly (Form 1040) in 1993, by gender and Form-706 filing status

	Non-F706 decedents	F706 decedents
Full Sample		
Male	73	77
Female	51	38
Old Subsample		
Male	59	60
Female	21	10
Young Subsample		
Male	59	67
Female	61	88

Step 1: Predicting the probability of filing an estate tax return

The first step of the Heckman method calls for a probit model designed to predict the probability of filing an estate tax return. Given the information presented above, we include the following as independent variables: age of the filer, size of AGI, presence of tax-exempt income, presence of dividend income, gender (“male”=1 for males, 0 otherwise), filing status (“married”=1 for married persons living with spouses in 1993, 0 otherwise), and an interactive variable (“male”*“married”) to account for potentially different effects of marital status upon men and women. Interpreting the coefficients on a

²³ Henry James’s novel *The Wings of the Dove* springs to mind.

probit regression can be challenging, so we instead present the odds ratios from a similar logit regression in Table 3.²⁴

TABLE 3: Odds ratios, logit regressions for probability of estate exceeding the Form-706 filing threshold

	Males		Females		Overall		
	Single	Married	Single	Married	Full	Old	Young
Filing age	0.98	1.00	1.04	1.00	1.01	0.96	1.06
AGI	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Presence of tax-exempt income	2.87	4.33	3.41	4.15	3.96	3.45	3.72
Presence of dividend income	23.76	3.39	3.91	7.19	5.05	2.76	8.94
Male					0.71	0.51	6.88
Married					0.35	0.18	1.26
Male*married					1.87	3.40	0.39
Percent concordant	85.5	86.7	89.8	86.4	86.6	88.1	85.7
N	1120	4280	1348	2194	8942	956	1509

Table 3 shows two things: we can use a parsimonious set of regressors to predict accurately the probability of filing an estate tax return,²⁵ and we observe different odds ratios depending upon gender and filing status (and, to some extent, age group), particularly for the variable indicating the presence of dividend income. For example, single males with dividend income are more than 20 times as likely to generate an estate tax return as their counterparts with no dividend income, *ceteris paribus*. The corresponding figure is far lower for all other groups. In the wealth estimation analysis, we aggregate the data for purposes of estimating the selection bias variable (λ). To

²⁴ All coefficients are significant. We report here the weighted results using actual AGI as a regressor. The odds ratio indicates that a person with \$x of AGI is just as likely to file a Form 706 as a person with \$(x+1), all else constant – an extra dollar of AGI simply is not that influential. But, if we use ln AGI as a regressor, the odds ratio ranges from 2.08 to 2.87 – that is, an extra 1 percent of AGI does matter in predicting the likelihood of Form-706 filing.

²⁵ We experimented with different regressors but found that several Form 1040 items are correlated with other Form 1040 items. We also had to determine the placement of variables into the two steps of the regression. Ultimately, we decided that the presence of certain items on the Form 1040 (dividends and tax-exempt interest) seemed more important for predicting the likelihood of filing but the size of various items more plausibly belonged in the wealth regression.

evaluate policies regarding the likelihood of exceeding the Form-706 filing threshold, however, these results suggest that analysts might want to conduct separate studies by gender and filing status.²⁶

Step 2: Predicting wealth

Table 4 shows coefficients from the wealth regression for the full, old, and young samples.²⁷ The first two columns both pertain to the full sample; the first column includes a λ constructed from the probability of exceeding the Form-706 filing threshold and the second a λ created from the probability of exceeding the maximum filing

TABLE 4: Wealth Regression

	Full sample (x706)*	Full sample (xthresh)**	Old subsample*	Young subsample*
Intercept	ns	3,657,754	ns	11,945,626
Dividends	69.52	69.06	35.31	18.45
Interest	ns	ns	22.62	ns
Tax-exempt interest	7.09	6.00	19.46	ns
Deductions	24.70	24.78	12.08	24.13
Age at filing	ns	-51,129	ns	ns
Sch. D income	ns	ns	-4.19	-1.40
Sch. E income	4.01	3.84	2.86	4.80
Male	ns	ns	1,719,209	-7,859,775
Married	-1,525,068	-1,375,439	ns	-11,527,675
Male*Married	ns	ns	ns	8,792,592
λ	1,192,116	510,842	ns	ns
Adjusted R-squared	0.482	0.481	0.735	0.276

Note: "ns" means not statistically significant.

* λ is constructed from a probit model that uses probability of exceeding the current-year filing threshold as the dependent variable.

** λ is constructed from a probit model that uses probability of exceeding the maximum filing threshold (\$1 million in 2002) as the dependent variable.

²⁶ Recall, however, that “married” indicates filing status (married filing jointly) in Tax Year 1993. “Single” filers include not only long-single persons but also recently widowed individuals, whose income and wealth patterns might more closely resemble those for “married” persons. In future research we hope to refine our marital-status indicator.

²⁷ We use weights in the regression, but we report the underlying unweighted number of observations.

threshold for gross estate (\$1 million in 2002). Table 5 offers means and standard deviations for the variables in the wealth regression.

TABLE 5: Means and Standard Deviations for Wealth Regression Variables

	full sample		old subsample		young subsample	
	mean	S.D	mean	S.D.	mean	S.D.
gross estate	2,332,377	211,061,905	2,381,685	149,374,390	3,525,805	103,103,314
dividends	18,841	1,607,562	24,497	2,126,588	7,689	1,131,482
interest	19,611	1,292,951	25,621	1,927,411	8,724	940,353
tax-ex.int.	16,896	1,124,990	22,396	1,282,347	4,081	398,406
deductions	26,742	1,777,739	25,360	2,283,583	39,247	1,319,098
age at filing	71.47	151.18	86.46	41.5	48.81	70.51
sch. D inc.	22,301	4,216,457	16,328	2,696,479	32,567	6,938,998
sch. E inc.	15,212	3,681,648	10,142	5,368,819	29,036	4,206,000
male	0.53	6.19	0.34	7.54	0.76	5.06
married	0.59	6.11	0.27	7.09	0.72	5.36
male*married	0.41	6.1	0.2	6.41	0.51	5.95
λ	1.4	7.27	1.14	8.52	1.41	8.41
N (unweighted)	4,226		542		533	

Note: All dollar amounts are reported in constant (2001) dollars.

In the full-sample regressions, note particularly the large coefficients on dividend income. They indicate that \$1 in additional dividend income yields nearly \$70 of estate wealth, or an estimated rate of return on dividend-bearing assets of only about 1.4 percent.²⁸ Compare this to an estimated rate of return (column 1) on tax-exempt assets of 14.1 percent and on assets yielding Schedule-E income of 24.9 percent. Of course, these assets do not truly generate these rates of return; the coefficients correspond to realized returns.²⁹ What we find for dividends is in line with previous research; the comparatively

²⁸ The value of a consol equals its coupon divided by its rate of return. A crude way of estimating the rate of return on the underlying asset generating dividends, then, is to act as if an asset worth \$69.52 generates \$1 of dividends in perpetuity, thus implying a rate of return equal to $1/69.52 = 1.4$ percent.

²⁹ Note that the implicit yield on tax-exempt interest income for the old subsample – about 5.1 percent – is reasonably close to posted yields during the time period in question. The yields on state and local bonds

low realized return suggests that people may choose investments in part to time income for tax purposes.³⁰ The next section explores this possibility in greater detail.

Note as well the difference in coefficients on dividend income for the full, old, and young samples. The implicit rate of return is greatest for the young subsample, followed by that for the old subsample. Although we cannot definitely say why, these results suggest that these decedents chose investments that yielded relatively more immediate cash, perhaps to pay off mortgages and child-rearing expenses (younger decedents) and medical bills (older decedents).³¹

The coefficients on Schedule-D income, Schedule-E income, and deductions are also worth discussing. Schedule D reports capital gains and losses. What the negative coefficients for the old and young subsamples may imply is that wealthier people, especially those close to death and those in prime working years, may take more advantage of the timing of capital losses.³² What is more, the step-up in basis at death for assets with accrued capital gains means that the elderly may tend to avoid realizing gains, thus saving their heirs future capital gains taxes.³³ The large implicit yield on Schedule-E income (income from rental real estate, royalties, and partnerships) is consistent with our knowledge of the increasing importance of limited partnerships over the period 1989-2004 (Jacobson et al. 2007). That the yield is especially large for the old subsample may

from 1993 to 2002 ranged from a low of 4.85 in 2002 to a high of 5.95 in 1999.

<http://www.federalreserve.gov/releases/h15/data.htm>.

³⁰ Johnson et al. (2009).

³¹ The latter is also suggested in Johnson et al. (2009).

³² The step-up in basis at death for assets with accrued capital gains does not work in reverse – accrued capital losses have no benefit for heirs. Consequently, wealthier people – especially those who anticipate leaving a large estate fairly soon– could find it especially advantageous to realize accumulated capital losses during their lives.

³³ This may especially be true for decedents whose spouses inherit the bulk of the estate. Because spousal bequests are fully deductible, accrued capital gains do not generate a tax burden via the estate tax for these heirs.

indicate that some of these decedents were holding onto assets that enabled them to retain control over a family business. The positive coefficient on deductions indicates that wealthier people take more deductions – not too surprising, as richer individuals are more aware of deduction possibilities, deductions include items likely to be larger for wealthier taxpayers (including other taxes, mortgage interest, and charitable contributions), and deductions in this regression may act partly as a proxy for AGI.³⁴ The relatively smaller coefficient for the old subsample could reflect a weaker connection between housing and terminal wealth, perhaps due to downsizing or mortgage payoffs among the elderly.

The significance of the coefficient on λ in the full-sample regressions tells us that selection bias is indeed an issue. Obtaining unbiased coefficients on the independent variables in equation (7) requires us to use the Heckman two-step method, as outlined in equations (8) and (9). That the coefficient on λ is positive indicates, not surprisingly, that unobserved factors positively associated with estate wealth are also positively associated with the probability of filing a Form 706.

Interestingly, the coefficient on λ is not significant for the old and the young subsamples. The argument we put forth to explain the coefficients on dividends could apply here as well: realized income and underlying wealth more closely match for people who have significant current out-of-pocket expenses – those with children at home, mortgages to pay, or large medical bills. Observable factors thus do a good job at predicting both the probability of exceeding the Form-706 filing threshold and the size of wealth for the relatively old and the relatively young. For the elderly, anticipation of death may also alter income realization patterns. Knowing that you can't take it with

³⁴ The original data included an amount for itemized deductions for itemizers. We assigned the standard-deduction amount to non-itemizers.

you, and knowing that you don't have much more time to enjoy your wealth, could mean that realized AGI (and thus current spending) more closely mirrors underlying wealth for older persons. For the young, we note that the coefficients on "male" are large (in absolute value) in both stages as compared to the same coefficients for the full sample. This observed trait may perform especially well in partitioning the data so as to reduce selection bias.

The differing results for the full, old, and young samples suggest that constructing the wealth distribution for the living population from that for the dead could be complicated. Accounting for the differences in the degree of uncertainty about impending death could be part of this process. Our future research will grapple with the modeling of the income-wealth relationship for living adults at the extremes of the age distribution.

An aside on income timing and wealth

Previous research suggests that wealthier people time the receipt of income so as to minimize tax liability, pointing to the lower realized rates of return on various income categories associated with higher-wealth individuals.³⁵ Our work reinforces those findings.

Table 6 offers regression results for F706 decedents from different wealth percentiles. Column 1 includes only F706 decedents whose gross estate fell in the top 90 percent, column 2 includes only those with gross estate in the top 50 percent, and so

³⁵ See Johnson et al. (2009) and Steuerle (1985).

forth.³⁶ The generally increasing size of the coefficients from left to right across the table for dividends, tax-exempt interest, and Schedule-E income suggests that, the wealthier the individual, the lower the realized rate of return on the assets generating these sorts of income.

TABLE 6: Wealth Regressions by Percentile of Wealth

	Top 90%	Top 50%	Top 10%	Top 5%	Top 1%
Intercept	4,740,924	7,092,792	19,082,851	30,882,135	83,423,691
Dividends	69.33	69.62	71.49	72.63	77.80
Interest	5.44	5.45	ns	ns	ns
Tax-ex. interest	6.24	6.53	7.80	7.96	ns
Deductions	24.48	24.53	24.23	23.76	22.99
Age at filing	-59,763	-90,182	-239,437	-375,011	-1,074,325
Sch. D income	ns	ns	ns	ns	ns
Sch. E income	3.52	3.57	3.64	3.54	3.98
Male	ns	ns	5,365,229	ns	ns
Married	-1,405,757	-2,176,530	-5,808,024	-10,279,937	-22,606,035
Male*married	ns	ns	ns	ns	ns
Adj. R squared	0.481	0.480	0.480	0.480	0.490
N (unweighted)	4,126	3,664	2,350	1,737	1,043

Note: These regressions include only observations for which a Form 706 was present, so no λ variable appears.

IV. Practical Considerations: Estimating Wealth Reported on Forms 706

Thus far, the analysis has focused on wealth estimation across the entire spectrum. But policy makers might be interested in a different question: can our models help predict the amount of wealth that will be reported on Forms 706? This estimate in turn will indicate something about the amount of wealth that will be transferred – often across

³⁶ Because these regressions include only those known to have filed an estate tax return, they do not have a λ variable. Note that these regressions, like the wealth regressions above, include weights.

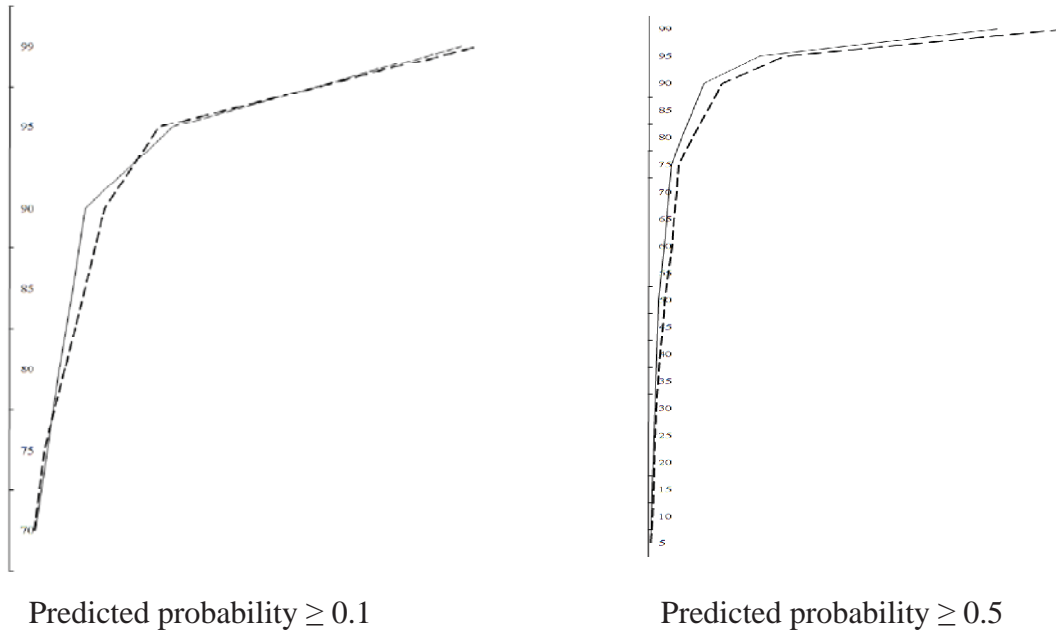
generations -- during a certain period of time, as well as the size of the expected estate tax base.

Because we use estate tax returns from several different years, and because the filing threshold increased over the period, we conduct this part of the analysis using information about whether a particular decedent left at least a certain amount of gross estate rather than whether his or her estate filed a Form 706. We have only a finite period for which we obtained estate tax returns; thus, we do not capture wealth information for those who died after 2002. Effectively, what we estimate is wealth associated with those who filed a Form 1040 in 1993 and who died between 3 and 9 years later with gross estate of at least \$1 million in real dollars.

We proceed the same way in this section as we did in the previous one – first predicting the probability of exceeding a given wealth threshold and then estimating the amount of wealth for each individual. To see how well our model predicts the total amount of gross estate reported in the period 1996-2002, we construct cumulative distribution functions (cdf's) of actual and predicted wealth. Because of the truncation of the actual wealth distribution, however, a straight comparison of values may not be particularly informative. Recall that the first step of the model generates a predicted probability. People either leave an estate exceeding a given wealth threshold or they don't; one way to cope with the truncation issue is to look only at actual and predicted values of wealth for people whose predicted probability exceeds a certain size. A larger cutoff means excluding more people whose estates actually do surpass the threshold; a smaller cutoff means including more people whose estates actually will fall short of the threshold – in other words, a classic type I-type II error tradeoff.

In each graph in Chart 5, the predicted gross estate is denoted by the dotted line and the actual gross estate is the solid line. The left graph indicates the cdfs for predicted and actual wealth of all persons whose predicted probability of filing exceeded 0.1. The right graph includes only cdfs of those for whom the predicted probability

CHART 5: Cumulative density functions, actual and predicted wealth



exceeds 0.5. Because we have no information on actual wealth for persons whose estates did not file a Form 706, we only report the percentiles for which we can make a meaningful comparison.³⁷

The choice made for the threshold predicted probability depends on how much of the distribution one wishes to model. Suppose only the very top of the wealth

³⁷ For example, only the top 30 percent of persons whose predicted probability of filing is at least 0.1 were actually required to file a Form 706. The higher the threshold for predicted probability, the greater percentage of persons who were required to file – which explains why the vertical axis extends farther down for the right-most graph. Note that we only map up to the 99th percentile – actual wealth for the very top wealth-holders far exceeds predicted wealth, so the cdfs intersect again, between the 99th and 100th percentiles. We do not show this, however, because doing so would compress the graphs so much at the lower percentiles that no distinction between the cdfs would be visible.

distribution – say the top 5 percent -- is of interest. Choosing a low threshold makes sense – this will capture virtually all of the actual F706 decedents; the captured non-F706 decedents are likely to generate predicted wealth that falls below the top of the distribution. As the left-most graph shows, imposing a threshold of 0.1 for the predicted filing probability will yield wealth that is slightly overpredicted at the 90th percentile, slightly underpredicted at the 95th percentile, and slightly overpredicted at the 99th percentile.³⁸

V. Conclusions

The research presented here suggests that linked estate and income tax records offer a promising data source for investigating a variety of important economic and policy issues. These include predicting whether an individual's terminal wealth will exceed a given threshold, imputing wealth from income and demographic information (particularly for high-wealth taxpayers), determining the degree of income realization across different wealth classes and age groups, modeling non-compliance, understanding unintended consequences of the estate tax, and estimating the potential tax base associated with an estate tax.

Our work reveals that we can accurately predict the probability of a decedent's estate filing a Form 706 from a relatively small set of Form-1040 information. We find that adjusting for this selection issue is important if we wish to estimate wealth from data on income. We also show that portfolios differ significantly across wealth classes, and

³⁸ Technical considerations made including the horizontal axis on the graph difficult. Here are the numbers for actual (predicted) gross estate for the various percentiles for the left-most graph: 90th percentile \$2,993,263 (\$2,372,494), 95th percentile \$4,695,133 (\$5,127,255), 99th percentile \$14,645,974 (\$14,164,627).

that people with greater wealth tend to have smaller realized yields on their assets. This strongly suggests that income data underestimate the differences in true well-being across individuals, and that wealth (or imputed wealth) – particularly for high-wealth people – presents a particularly useful alternative. The information provided on linked Tax Forms 706 and 1040 undeniably provide a singular source of data for mapping the connections between income and wealth.

SOURCES

- Bernheim, Andre Shleifer, and Lawrence Summers. (1985) The Strategic Bequest Motive. *Journal of Political Economy* 93(6), 1045-76.
- Choudhry, G. Hussain. (11 June 2001) Development of Family Panel Weights for Tax Years 1994-96. Unpublished internal document prepared by Westat for Internal Revenue Service, Statistics of Income Division.
- _____. (13 June 2001) Development of Family Panel Weights for Tax Years 1987-1993. Unpublished internal document prepared by Westat for Internal Revenue Service, Statistics of Income Division.
- _____. (13 June 2001) Evaluation of Family Panel Weights for the Tax Year 1987. Unpublished internal document prepared by Westat for Internal Revenue Service, Statistics of Income Division.
- Czajka, John, and Alan Schirm. (1991) Cross-Sectional Weighting of Combined Panel and Cross-Section Observations, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC.
- Czajka, John L and Bonnye Walker (1989) Combining Panel and Cross-Sectional Selection in an Annual Sample of Tax Returns. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Hao, Lingxin, and Daniel Naiman. (2007) *Quantile Regression*. Sage.
- Heckman, James. (1979) Sample Selection Bias as a Specification Error. *Econometrica* (47), 153–61.
- Jacobson, Darien, Brian Raub, and Barry Johnson. (Summer 2007) The Estate Tax: Ninety Years and Counting. *Statistics of Income Bulletin*, Internal Revenue Service.
- Johnson, Barry, Kevin Moore, and Lisa Schreiber. (2009) The Income-Wealth Paradox: Connections between Realized Income and Wealth Among America’s Aging Top Wealth-Holders. 2009 IRS Research Conference.
- Johnson, Barry, and Lisa Schreiber. (2006) Creativity and Compromise. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Nunns, James, Deena Ackerman, James Cilke, Julie-Anne Cronin, Janet Holtzblatt, Gillian Hunter, Emily Lin, and Janet McCubbin. (2008) Treasury’s Panel Model

for Tax Analysis. *OTA Technical Working Paper 3*. U.S. Department of Treasury, Office of Tax Analysis. <http://www.treas.gov/offices/tax-policy/library/OTAtch03.pdf>.

Rolf, Karen, and Joseph Ferrie. (2008) The May-December Relationship Since 1850: Age Homogamy in the U.S. Working Paper. <http://paa2008.princeton.edu/download.aspx?submissionId=80695>.

Schirm, Alan, and John Czajka. (1992) Weighting a Panel of Individual Tax Returns for Cross-Sectional Estimation. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Shrestha, Laura. (2006) Life Expectancy in the United States. CRS Report for Congress, <http://aging.senate.gov/crs/aging1.pdf>.

Steuerle, Eugene. (1985) Wealth, Realized Income, and the Measure of Well-Being. In Martin David and Timothy Smeeding, eds., *Horizontal Equity, Uncertainty, and Economic Well-Being*. National Bureau of Economic Research.

Tobin, James. (1958) Estimation of Relationships for Limited Dependent Variables. *Econometrica* 26(1), 24-36

A Comparison of Wealth Estimates for America's Wealthiest Decedents Using Tax Data and Data from the Forbes 400

Brian Raub, Barry Johnson, and Joseph Newcomb, Statistics of Income, IRS¹

Presented in "New Research on Wealth and Estate Taxation"

Introduction

Measuring the wealth of the Nation's citizens has long been a topic of interest among researchers and policy planners. Unfortunately, such measurements are difficult to make because there are few sources of data on the wealth holdings of the general population, and most especially of the very rich. Two of the better-known sources of wealth statistics are the household estimates derived from the Federal Reserve Board of Governor's Survey of Consumer Finances (SCF) (Kenneckell, 2009) and estimates of personal wealth derived from estate tax returns, produced by the Statistics of Income Division (SOI) of the Internal Revenue Service (Raub, 2007). In addition, Forbes magazine annually has produced a list, known as the Forbes 400, that includes wealth estimates for the 400 wealthiest individuals in the U.S. While those included on the annual Forbes 400 list represent less than .0002 percent of the U.S. population, this group holds a relatively large share of the Nation's wealth. For example, in 2007, the almost \$1.6 trillion in estimated collective net worth owned by the Forbes 400 accounted for about 2.3 percent of total U.S. household net worth (Kopczuk and Saez, 2004).

This article focuses on the estimates of wealth produced for the Forbes 400 and their relationship to data collected by SOI (see McCubbin, 1994, for results of an earlier pilot of this study). In all years examined, the threshold for inclusion in the Forbes estimates was well above the effective estate tax filing threshold; therefore, there should be complete overlap between these two sources of wealth information. Examining the Forbes and SOI data together should provide insights into strengths and weakness of both data sources for economic analysis. To the

extent that data from these sources prove different, this analysis should also provide insight into issues that influence measures of wealth derived from estate tax data and contribute to a fuller understanding of economic well-being. Finally, by examining, to the extent possible, the form of wealth held by these top wealth holders, this research may inform the ongoing debate surrounding the effects of the Federal estate tax on the general economy.

Background

Estate tax returns provide a rich data source from which to study the nation's wealthiest individuals. The return contains a complete listing of a decedent's assets and debts, a demographic profile of the decedent, and information on the costs of administering the estate. All assets are valued as of the date of the decedent's death, although special valuation rules apply for certain uses of real estate and for assets that decline in value within 6 months of a decedent's death.² The decedent's share of jointly held assets, as well as any assets over which the decedent held power of appointment are also reported, as is the face, but not cash, value of all life insurance policies. Valuations of business assets are frequently subject to significant discounts when there is evidence that a decedent's fractional ownership share affects marketability.

Forbes magazine has published a list of the 400 richest Americans annually since 1982. The wealth estimates for the individuals on the list are produced through investigative research and take into account interviews with employees, competitors, customers, attorneys, ex-spouses and securities analysts, as well as reviews of Securities and Exchange Commission (SEC) filings and legal documents. Privately held companies are valued by coupling estimates of revenues or profits with prevailing price-to-revenues or price-to-earnings ratios for similar public companies. Asset values are rounded to the nearest 100 million dollars and stocks are valued on August 31. Forbes acknowledges that its analysts do not have knowledge of all assets belonging to the

individuals on the list, and describes the published estimates as “deliberatively conservative,” (Miller, 2009).

For its estimates, Forbes magazine uses net worth as the measure of wealth. For the SOI data, net worth is calculated as total assets minus debts, where total assets is calculated using the value of each asset on the day that the owner died. An estimate of the cash value of life insurance replaces the reported face value and any reduction in value for real estate allowed under IRC§ 2032A is disregarded in this calculation.

The Data

An exact match, based on name, between estate tax returns filed for 1982 – 2008 and the 1,378 individuals who have appeared in the publicly available listings of the Forbes 400 covering the same period yielded 376 individuals who had been listed in the Forbes 400 and for whom an estate tax return was available in SOI’s database. This group also included a small number of individuals for whom estate tax returns were filed in 2009 or 2010 but for whom only limited data were available because the returns had not yet been processed by SOI. Estate tax returns could not be located for 20 deceased individuals who had appeared in the Forbes 400. With additional research, we were able to verify that the individuals’ estates were not required to file for various reasons.³

By comparing SOI data to the annual thresholds for inclusion on the Forbes 400 list, we were also able to identify 26 individuals whose net worth at death, as reported on the estate tax return, was sufficient for inclusion in the contemporaneous Forbes 400, but who never appeared on the list. For about half of these individuals, the failure to appear on the list appears to be due to the method Forbes uses to assign wealth that is dispersed across families. For some of the others, the primary source of wealth was from creative endeavors, which may be particularly

difficult to value prior to an auction or other sale; in some cases these assets may increase substantially in value because the creator has died.

Results

The final Forbes-SOI matched dataset contained 376 records, consisting of 304 male and 72 female decedents. Male decedents had a mean age of 80, with more than half the sample aged 80 or older. Most, 74 percent, were married, with only 26 percent widowed, single, divorced or separated. For this group, the mean estate tax net worth was \$405.9 million, while the median value was \$174.1 million. Only 18 percent of female decedents were married, while 82 percent were widowed or otherwise single. The mean estate net worth for women was slightly higher than that for men, \$495.9 million, but the median was lower, \$110.4 million.

As seen in Figure A, the asset and debt holdings of decedents in the dataset differed by gender and marital status. For each class of decedent, stock, which includes shares of both publicly traded and closely held corporations, accounted for the largest share of any asset type, between 41.1 and 50.8 percent of total assets. Bonds, cash, real estate, including real estate partnerships and real estate investment trusts (REITS), and business assets, defined here as unincorporated businesses, typically combined to make up the bulk of the non-equity portion of the estate, although the percentage held in each asset type varied markedly by gender and marital status. Males held, on average, more of their portfolios in business assets and less in bonds than their female counterparts. This may reflect both the higher incidence of business ownership among men and the much higher percentage of women who were widowed, as widowed decedents may have divested active ownership interest in a late spouse's businesses following his or her death. For both males and females, unmarried decedents held, on average, a greater

share of their portfolios in bonds and cash than married decedents, and smaller shares in stock, business assets, and real estate. Males, had, on average, significantly more debt than females.

Figure A – Mean Percentage of Total Assets, Calculated for Asset Components, and Debt-to-Asset Ratio, by Sex and Marital Status

Asset Type	Males			Females		
	All	Married	Unmarried	All	Married	Unmarried
Stock	45.6	47.1	41.1	43.2	50.8	41.5
Bonds	12.2	10.8	16.0	17.8	4.8	20.5
Business assets	12.4	13.1	10.6	6.4	7.9	6.1
Real estate	9.8	10.0	9.1	11.5	23.2	8.9
Cash	5.4	5.2	6.1	6.9	4.5	7.4
Other assets	5.2	5.4	4.6	8.0	6.5	8.3
Mortgages/notes	4.3	3.0	8.1	2.6	0.6	3.1
Retirement assets	1.3	1.6	0.3	0.8	1.6	0.6
Insurance	0.5	0.6	0.4	< 0.1	0.1	0.1
Debt-to-asset ratio	8.3	8.0	9.3	4.0	4.1	3.9

One of the challenges posed by this analysis is the fluid nature of the Forbes 400 listing. Over the period examined, the threshold for inclusion on the Forbes list increased in most years, starting from an inflation-adjusted low of \$225 million in 1982 and rising to \$1.3 billion by 2007. Because the threshold for inclusion might increase as others' fortunes surpass those listed in any given year, individuals included on the list in one year may be absent in the next even if their wealth remained more or less constant. Members can drop off the list for other reasons as well; for example a member's wealth may diminish due to economic conditions or because of a significant charitable donation. Individuals may also die, making room for others.

Figure B provides information on the status of a decedent's membership on the Forbes 400 list at the time of death for the full linked sample in three categories: those who were listed either at the time of death or within 1 year of death; those whose net worth declined for some observable reason between the time that they were listed and death; and those who dropped off

the list prior to death seemingly because the growth of their individual net worth did not keep pace with increases in the inclusion threshold.⁴

Figure B – Financial and Demographic Data, by Forbes Membership Status

Asset Type	On the List at Death		Dropped from List Prior to Death			
	Male	Female	Net Worth Declined		Threshold Increased	
			Male	Female	Male	Female
Number of observations	148	33	67	12	89	27
Mean net worth (SOI)	\$621.1 million	\$845.6 million	\$213.8 million	\$419.8 million	\$192.6 million	\$102.4 million
Mean Last Forbes estimate	\$1,530.1 million	\$1,347.9 million	\$689.6 million	\$838.3 million	\$404.7 million	\$417.3 million
Mean age	79.1	80.5	79.9	87.4	80.5	88.2
Mean Forbes Ratio	0.48	0.60	0.32	0.62	0.46	0.30
Mean % of Total Assets:						
Stock	53.1	48.7	38.1	50.1	38.6	33.5
Bonds	11.2	16.9	13.7	12.2	13.0	21.1
Business assets	10.2	4.4	13.4	8.6	15.4	8.0
Real estate	8.0	10.4	11.3	14.8	11.5	11.2
Cash	4.2	4.6	7.4	3.1	5.9	11.4
Other assets	5.5	9.0	5.2	5.5	4.7	7.9
Mortgages/notes	3.8	2.2	5.6	5.6	4.2	1.9
Retirement assets	0.9	0.7	2.1	< 0.1	1.4	1.2
Insurance	0.4	< 0.1	0.7	< 0.1	0.8	0.2
Debt-to-asset ratio	6.8	2.4	12.4	1.9	7.8	6.8

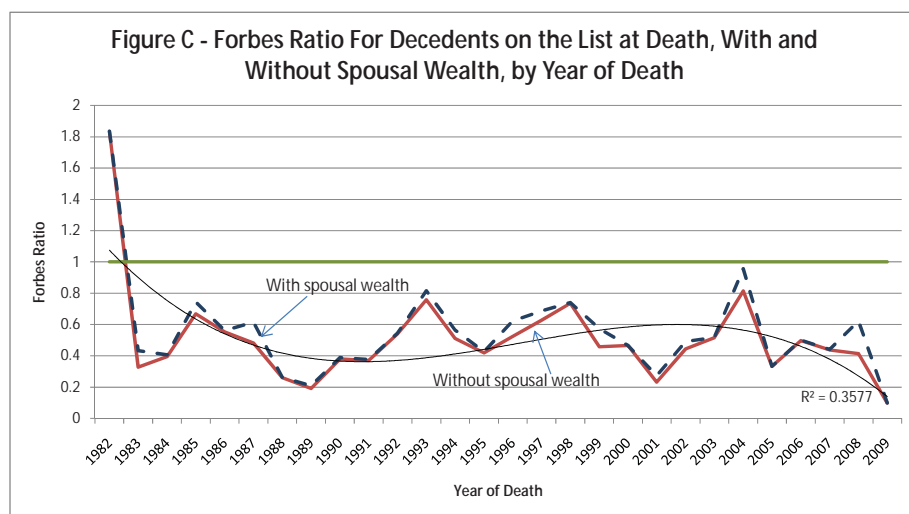
Note: All money amounts have been converted to constant 2008 dollars.

As expected, the average reported estate tax net worth in Figure B is highest for decedents in the Listed at Death category; the lowest average was reported for those in the Threshold Increased category. This pattern is also apparent in the average Forbes estimates of wealth. Interestingly, the Forbes Ratio, defined as the value of net worth reported on estate tax returns divided by the Forbes estimate of net worth, also varies between these groups, but the pattern is somewhat different. For the decedents in the Threshold Increased group, the mean Ratio is 0.42. For decedents in the Net Worth Declined group, the mean ratio is 0.37. For decedents in the Listed at Death group, the mean ratio is 0.50. ANOVA analysis rejects the null hypothesis that these three population means are equal at the 10% level ($F = 2.51, p=0.824$).

Comparing Forbes and Estate Valuations

Figure C displays the average Forbes Ratio by year of death, limited to the 181 decedents who were on the list at death, whose estate and Forbes estimates we would expect to be most

closely aligned. If the estimates were identical, we would expect a straight horizontal line at 1.0. For these data, however, the average ratio is less than 1.0 for all but one year. While there is considerable year to year variation in the mean value, ranging from a low of 0.10 to a high of 1.83, the overall trend shows a slight decrease between 1982 and 2010. In aggregate, the average Forbes Ratio was 0.50, meaning that the values reported for tax purposes were about half of those estimated by Forbes. However, for 15.5 percent of decedents, the estate net worth was actually larger than the Forbes value.



Although some of the year-to-year variation shown in Figure C is due to a relatively small number of observations in most years, some of it is certainly structural. A significant example is the differential treatment of assets held in trust between the Forbes estimates and the tax data. For valuation purposes, Forbes treats assets held in trust as belonging to the income beneficiary of the trust, even if that beneficiary has no control over the disposition of the *corpus*, or assets, of the trust. This is not the case for tax purposes where only assets that a decedent owned outright, or assets held in trust for which the decedent retained control or the power of

appointment are included in the estate for tax purposes.⁵ The lowest Forbes ratios can be attributed to this difference.

Another structural difference between the Forbes and tax estimates is in the treatment of family wealth. The Forbes estimates for married persons treat joint wealth, including wealth owned by extended family members in some cases, as if it belongs to the listed individual, while for estate tax purposes only assets owned solely by a decedent are included. Thus, for married individuals in our data set, some of the difference can be attributed to the missing spousal/family wealth from the tax data. We can partially adjust for this by adding the full value of jointly held assets, both those held as joint tenants and as community property, to the reported tax values. While still omitting assets owned solely by a spouse or other family members, the dashed line in Figure C shows that, for some years, this additional wealth increases the Ratio closer to 1. In fact, for all married individuals in the file, the average overall Forbes Ratio increases from .46 to .53 when including these jointly held assets.

The explicit recognition of valuation discounts when valuing assets for tax purposes is another significant difference between the Forbes and tax values.⁶ These discounts can be as large as 50 percent. In 2004, SOI began collecting data on the incidence and magnitude of discounts reported for estate tax valuation purposes. Almost two-thirds, or 64.4 percent, of estate tax returns in our data set filed on or after 2004 included some assets for which a valuation discount had been reported. Adding the value of these discounts to the estate tax values increased the average Forbes Ratio for this sub-group from .47 to .54. Adjusting for both the value of discounts and the value of spousal wealth, the average Forbes Ratio for this sub-group increased to .58. It is widely believed that the use and size of valuation discounts has increased over time. These may account for a portion of the relatively small decline in the Ratio over the period.

While these structural issues explain a portion of the differences between the Forbes estimates and tax values, there also seem to be some financial profiles that are easier for Forbes to estimate than others. Figure D shows that portfolio differences, in aggregate, between individuals with high and low Forbes Ratios, grouped in quartiles based on the ratio value, can be quite striking.⁷ For those whose ratio was at least 0.78, the highest quartile group, the portfolio is dominated by stocks and bonds. Non-corporate business assets make up a relatively low proportion, as do real estate and cash. In contrast, for those whose Forbes Ratio was less than 0.10, the lowest quartile group, the data reveal an aggregate portfolio that is quite diversified. Equity investments account for less than one-third of the total, with real estate assets making up more than 15 percent, three times larger than for the higher ratio group. Other notable differences for the lowest-ratio group include a much higher concentration of business assets and mortgages and notes. Business assets are notable because of the subjective nature of valuing them, especially relatively small businesses for which a portion of the value may be derived from the personality or skills of a founder. The death of such a key person may have a strong adverse affect, at least initially, on the valuation of the firm and could account for a significant share of the difference between the tax and Forbes values in some cases. Overall, the portfolio of the high-ratio group is dominated by assets for which there is a ready market, making valuation for both estate and Forbes purposes relatively straightforward, while that of the low-ratio group contains higher concentrations of assets for which valuation is much more subjective, and more frequently subject to valuation discounts.

In addition to the differences in portfolio composition among the three groups, the disparate holding of debt is equally striking. While debts, on average, represented less than 2 percent of the total assets for the high-ratio group, they accounted for an average of nearly 10

unable to observe the incidence or magnitude of debts, and explain some of the overall gap between the tax data and the Forbes estimates.

Figure D – Mean Percentage of Total Assets, Calculated for Asset Components, and Debt-to-Asset Ratio, by Forbes Ratio Quartile Group

Asset Type	Group		
	Low quartile (ratio < 0.10)	Middle quartiles (0.10 < ratio < 0.78)	High quartile (ratio > 0.78)
Stock	31.3	61.5	61.2
Publicly traded	14.9	26.9	31.2
Closely held	16.3	34.6	30.1
Bonds	13.6	11.3	15.2
Business assets	12.8	7.1	8.5
Real estate	18.4	5.3	5.3
Cash	6.8	5.4	2.3
Other assets	8.8	5.6	6.1
Mortgages/notes	5.7	2.3	0.4
Retirement assets	1.1	1.2	1.0
Insurance	1.3	0.1	<1.0
Debt-to-asset ratio	9.5	6.3	1.8

To further assess the relationship between the Forbes estimates and wealth reported on estate tax returns, we used OLS regression to model the Forbes Ratio using demographic and financial data from all of the tax records in the data set.⁸ The dependent variable is a natural log transformation of the Forbes Ratio. The results of this regression are shown in Figure E. The R-squared value of this model is 0.3579, and the model has a joint F value of 11.18. Using GLS type heteroskedastic robust standard errors did not substantially change the results of this model. Binary variables were included to account for marital status and gender. Additionally, a binary variable was included to indicate the group that had died while on the Forbes list.

Figure E - An Estimate of the Forbes Ratio

Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	0.44481	0.11872	3.75	0.0002
nw08	<0.00000	<0.00000	8.82	<.0001
married	-0.16204	0.07368	-2.20	0.0285
male	-0.02348	0.04258	-0.55	0.5818
MarriedMale	0.14234	0.07887	1.80	0.0720
AGE	-0.00186	0.00128	-1.46	0.1457
nyod ¹	-0.00614	0.00210	-2.92	0.0038
DiedOnList	-0.05848	0.02795	-2.09	0.0372
lnStock ²	0.00882	0.00384	2.30	0.0223
lnBonds ²	0.00729	0.00186	3.91	0.0001
lnCash ²	-0.02837	0.00740	-3.83	0.0002
lnMortnote ²	0.00016	0.00214	0.07	0.9412
lnRetirement ²	-0.00191	0.00201	-0.95	0.3426
lnBusiness ²	0.00105	0.00215	0.49	0.6254
lnRealest ²	0.00466	0.00388	1.20	0.2308
lnInsure ²	-0.00364	0.00227	-1.60	0.1099
lnOther ²	-0.00701	0.00516	-1.36	0.1749
lnDebts ²	-0.00300	0.00248	-1.21	0.2271

¹Year of death value is re-centered at 1982=0. ²All assets are expressed as portfolio shares.

The group of decedents whose Forbes Ratio was greater than 1, meaning that the estate tax value of net worth exceeded the Forbes estimate, was relatively small, consisting of only 49 individuals, or about 13 percent of all decedents. To investigate whether this group could better be modeled as distinct from those whose tax values were less than the Forbes estimate, we constructed an alternative model that included interaction variables based on this grouping, but there was not a significant enough reduction in the sum of squared residuals to justify its use over our original model.

Of the factors that contribute to modeling the Forbes Ratio, several seem particularly important. The decedent's marital status and the interaction term of marital status and gender are statistically significant in this model. We interpret this result to mean that differences between the way that the Forbes and tax estimates assign ownership of assets within families contribute significantly to the differences we observe in estimates from each. Year of death is also

statistically significant and negative. Although the parameter estimate for this variable is very small, it does suggest a slight decline in the average Forbes Ratio over time of 0.6 percent per year that is consistent with Figure C. The decedent's net worth is also statistically significant and positive, though the coefficient is extremely small. All else equal, wealthier decedents seem to have a higher Forbes Ratio. This observation is consistent with Figure C.

The coefficients for stock, bonds, and cash are all statistically significant at the 90 percent level. An increase in the share of the portfolio held in stocks and bonds has a predicted effect of increasing the SOI value relative to the Forbes estimate, and moving the Ratio closer to 1 for most cases. An increase in the share of the decedent's portfolio held in cash will have the opposite predicted effect. We speculate that in some cases, large amounts of cash in an estate represent the sale of an asset for less than the Forbes estimated value.

Summary

Using a unique data set that combines estimates of wealth from Forbes magazine's annual listings of the wealthiest 400 Americans for 1982-2009, with data reported on federal estate tax returns filed for deceased current and former listees, we have shown that, on average, the values reported for tax purposes are approximately half those estimated by Forbes. Although the ratio of these two estimates, dubbed the Forbes Ratio, varies a great deal due to the relatively small sample size and the unique nature of each decedent in this data set, we have shown that the trend of this average has been relatively stable over time, with a slight decline overall. Using detailed information reported on the tax returns, we have also shown that a portion of the difference measured by the Ratio is due to structural differences between the two sets of estimates. Specifically, Forbes includes the joint assets of married persons in their estimates, while for tax purposes only assets owned by the deceased partner are reported. Adding even a

small portion of the surviving spouse's assets to the tax-based net worth values decreases the difference between these estimates significantly.

The estate tax data also allowed us to examine portfolio allocation differences between individuals whose Forbes Ratio was relatively close to 1 and those for whom the Forbes Ratio was extremely low. Based on tabular data and regression results, we conclude that tax values and Forbes values were in closest agreement when valuation issues were relatively objective and when individuals had a relatively small amount of debt, but much further apart when the portfolio was dominated by assets for which valuation required a greater degree of subjectivity or were difficult to observe, and when individuals held relatively more debt.

This research highlights the inherent difficulties of valuing assets which are not highly liquid. The portfolios of very wealthy individuals are made up of highly unique assets and often the value of assets, such as businesses, are very closely tied to the personality and skills of the owner. Determining a precise value for these assets can involve more art than science. Previous researchers have suggested that differences between tax and Forbes estimates are due to evasion; however, these are high dollar-value returns filed for very well-known decedents and so they tend to be carefully prepared by licensed professionals. This suggests that, while values reported for tax purposes may be conservative, they will fall within legally defensible parameters. Estimates of value for other purposes may be much more optimistic, but perhaps no more precise than those provided to the IRS, and so, contrary to the Forbes' assertion that its methodology is conservative, these results suggest that it is actually generous in some cases. Without actually selling a difficult to value asset, it may be impossible to determine a precise "market value," especially in times of economic volatility. These data provide an important reminder that data reported for administrative purposes can be legally acceptable yet fundamentally different from

those collected for other purposes. This should be an important consideration when attempting to use data to answer specific research questions or when comparing results from several sources and even suggests that, in some cases, multiple data sources should be examined in order to obtain truly robust results.

Notes

¹The views expressed in this paper represent the opinions and conclusions of the authors alone and do not necessarily represent the opinions of the Internal Revenue Service or the Treasury Department.

²Under Internal Revenue Code (IRC) Section 2032A, real estate used in an ongoing business or farm can be valued based on its present use, rather than its full market value. The reduction in value is limited, and the heirs must agree to continue the business or farm for 10 years after the decedent's death in order to qualify.

³For example, for some Forbes decedents, it was clear that most of the Forbes-identified wealth was owned through a continuing trust and not outright by the decedent.

⁴Decedents were assigned to "Net Worth Declined" category by Forbes based on information from Forbes that indicated either a large financial setback or because Forbes revised a prior year estimate when there was reason to believe that estimate was much too high. Decedents were assigned to the "Threshold Increased" category based on the Forbes Magazine estimates for 'near misses' in the year they were dropped from the list.

⁵The power of appointment is the power to determine who will be given the authority to dispose of certain property under the will.

⁶Common reasons for discounting an asset value below fair market include: problems with the physical condition of the asset; recognition that a business may have derived significant value

from the active participation of the decedent; or, most commonly, shared ownership arrangements which make the asset unattractive to a willing buyer, as in the case of assets held in family limited partnerships (FLP). For a more complete description of FLPs, see Raub, 2007 p. 121.

⁷The analysis in this section is limited to data collected by SOI after 1988 to take advantage of additional data which distinguishes between publicly traded and closely held corporate stock in the decedent's portfolio. While publicly traded stock can be accurately valued using exchange data, closely held stock values are more subjective in nature. It is important to note that the portfolio allocations for the 181-decedents used in the preceding analysis were very similar to those of this smaller, 136-decedent group.

⁸This analysis includes 362 observations from the full 376 observation data set; the omitted records were incomplete because SOI processing had not been completed at the time of the analysis. No additional detailed demographic or portfolio data were available from Forbes.

References

Kennickell, Arthur. (2009) Ponds and Streams: Wealth and Income in the U.S., 1989 to 2007. <http://www.federalreserve.gov/pubs/feds/2009/200913/200913pap.pdf>

Kopczuk, Wojciech and Saez, Emmanuel (2004) Top Wealth Shares in the United States 1916-2000: Evidence from Estate Tax Returns. <http://www.nber.org/papers/w10399>

McCubbin, Janet. (1994) Improving Wealth Estimates Derived from Estate Tax Data. In *Compendium of Federal Estate Tax and Personal Wealth Studies*, Barry W. Johnson, ed. Internal Revenue Service Publication 1773.

Miller, Matthew. (2009) America's Richest People: Complete Methodology, How Forbes compiles its list of the 400 richest Americans. <http://www.forbes.com/2009/09/29/forbes-400-gates-buffett-ellison-rich-list-09-methodology.html>

Raub, Brian. (2007) Federal Estate Tax Returns Filed for 2004 Decedents. <http://www.irs.gov/pub/irs-soi/04esreturnbul.pdf>

Raub, Brian. (2008) "Personal Wealth, 2004," <http://www.irs.gov/pub/irs-soi/08fallbulpw.pdf>