Origins of the Estate and Personal Wealth Sample Design

Paul B. McMahon, Jr., Internal Revenue Service

In Estates and Personal Wealth, we have two studies with different populations under consideration. The Estates Study is concerned with the assets, debts, and taxes left by a decedent who had more than a certain amount of wealth. The Personal Wealth Study, on the other hand, is focused on the wealth holdings of the living. For Estates, essentially all the population appears on a sampling frame, but, to study the living, we must rely on proxies that can be observed for only a portion of the distribution, the portion in the tail.

One set of samples is the source for the data in both series of studies.

We will first briefly describe the interest in these populations. The "questionnaire" in this set of surveys is an administrative record, the Form 706, *Estate Tax Return*, and the sampling frame is a system of electronic records derived from the initial filing. We will provide a bit of background on these as well.

We focus on the studies initiated since 1982, with strata designs that changed somewhat over that time. While some previous papers have addressed certain estimation issues, such as with the Personal Wealth Estimation (Johnson and Woodburn, 1994), there have been only the briefest descriptions of the strata design or concepts.

Our goal, then, is to show how the different requirements for studies of the two populations affect this one sample design, and how that design has evolved in the light of tax law changes.

Finally, we will discuss some future directions for the series, in light of pending legislation.

▶ Analysts and Uses

The two main sponsors of these studies are the Office of Tax Analysis in the Department of the Treasury and Congress's Joint Committee on Taxation. Their objective is to gather data for oversight on the operation of the tax laws and, in this case, on Estate Taxes, and projecting the effects of proposed changes to those laws. This is not limited to the revenue aspects of the tax laws.

That is, this study has to meet two uses. First, the measurement of current law, and second, determining the effect on the living population who have estates large enough for the eventual filings. In order to look at trends in the analysis, we need to be concerned about the effect of economic conditions at the time of the observations (the date of death), the time of life considerations (youthful spenders versus middle-age savers, for example), and what the sociologists call age cohorts, where history affects economic decisions (the Depression generation's thrift).

There is also an underlying philosophical question: Does the operation of the Estate Tax, in concert with a graduated income tax, prevent the concentration of wealth into few hands? At the beginning of the twentieth century, some politicians, like Theodore Roosevelt argued in favor of the Estate Tax on just this issue. More recently, there have been numerous articles this past spring in the *New York Times* and the *Wall Street Journal*, for example, on the concentrations of incomes. Income is often taken as a proxy for wealth; so, this question is clearly of continued interest.

Indeed, using data from Estate Tax Returns dating back to 1916, the National Bureau for Economic Research (NBER) published a working paper that considers this very concentration issue (Kupczuk and Saez, 2004). Although the data used in that study are from many years in the past, the sample designs for most of those years actually originated in the mid-1980's and reflect the plans developed for sampling more recent tax filings.

▶ The Administrative Records

The basic data for these studies use the records that arise from what some have called the "Death Tax." It is more accurate, though, to call it a transfer tax, as the

change of an asset's title to some beneficiary or heir is the proximate cause for this tax or its complement, the gift tax. The tax return, which acts as the questionnaire for our studies, is Form 706, *Estate Tax Return*.

The assets that are considered for this tax are everything owned by the decedent: art, bonds, cars, personal effects, through to zoom lenses and beyond. That is, the filing is based on a complete inventory of an individual's possessions. In this, it is similar to the information that the Federal Reserve attempts to obtain in its Survey of Consumer Finance.

There are major differences between the data collected for the Federal Reserve surveys and the IRS studies, however. First, the tax form also includes insurance payments to the estate and gifts made before the decedent's death, which would not be included in the Finance Survey. Then, the law permits deductions for the costs of such items as estate administration, the funeral, and legal counsel, as well as exempting the contributions to charities and the spouse of the decedent.

Another difference is that the value of the assets is usually assessed at the time of death, not as of some common reference date for all respondents.

The main difference, though, arises from the populations these two sets of studies targets. The Survey of Consumer Finance seeks to estimate the holdings of all households, while the Estates and Personal Wealth studies are limited to individuals who exceed a certain threshold set by the tax code.

Figure 1.--Estate Tax Return Filing Thresholds for Selected Years

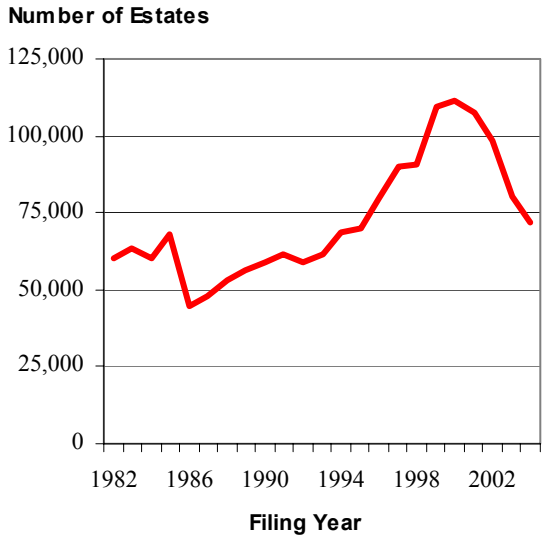
Year of Death	Gross Estate Threshold
1997	$600,000
1998	$625,000
1999	$650,000
2000 & 2001	$675,000
2002 & 2003	$1,000,000
2004 & 2005	$1,500,000
2006 – 2008	$2,000,000

If the value of those possessions at the time of the decedent's death is below the threshold amount shown in Figure 1, then there is no estate tax. That threshold varies depending on the year of the decedent's death. It is currently $1.5 million, rising to $2 million in January 2006. These values have been updated in the tax code periodically; in 1977, for example, the threshold was $60,000.

Filing is not required for smaller estates, though some do if the value is near the boundary. This may be due to the difficulty in itemizing all of an estate's assets. In those cases, amended returns will be filed, and perhaps a tax assessed, but such cases are outside the scope of this set of studies; we are only concerned with initial filings.

One can see the effect of raising the threshold quite clearly in Figure 2. In 1986, the exclusion was doubled, to $120,000, with a resultant sharp drop in filings and again, after the 2001 tax bill passed, which raised the limit several times in succession.

Figure 2.--Annual Filings of Estate Tax Returns



While the law and regulations provide one source of limitations on the studies, and thereby the design, another is in the physical properties of the documents and the processing regimen.

The Estate Tax Return is filed on paper as a large package with sections that are partly structured and partly

respondent-created. While Form 706 is, on the surface, highly standardized, the space allowed for some schedules (such as a list of heirs) is sometimes insufficient. This leads the attorney or executor to create substitute schedules of their own design.

The filing regulations also mandate the inclusion of the will, unless the decedent died intestate, appraisals of real property, and the death certificate. While the last may be relatively standardized, the will and appraisal(s) are not.

Moreover, all of these filings are subjected to an audit review, unlike the small proportion of Individual Tax Returns. Such audits keep the return unavailable for considerable lengths of time. Thus, the Statistics of Income studies must capture the return first and cannot wait for the entire population to become available; the sample must be selected as the returns are processed through the administrative pipeline.

The filing deadline for these documents is 9 months after the decedent's death. Extensions to this deadline are often required, because it takes time to locate some financial records, and for some assets to come to light. Since evaluating the effect of changes to the law is an objective, focus on a particular year of death means we must continue the selection over more than 2 years: the focus year and at least the following 15 months.

In practice, given the administrative environment, the minimum effective sampling period is 3 years. The additional months arise from the cycle of updating the computer programs, where the latest versions are introduced each January.

We want to use an electronic record in the sampling of these estates because, while selecting the returns as paper records ensures their retention for statistical purposes, this direct approach is costly and difficult and limits stratification options. The 1977 Study's manually-selected sample was limited to three strata, for example, and required considerable daily coordination with the ten national Service Centers where the returns were filed.

Yet the use of the computer records also gives rise to limitations. Ignoring audit trail codes, tracking data, and name and address information, there were only 16 amounts available in 1982, less than we can use today, but not by much. Most of those, 13, were involved in the calculation of the tax liability. This left a bare handful as possibly useful for sampling purposes, including some of the "code" fields.

Decedent's Year of Death was available. This was, and is, a tax-related field due to changes in the filing threshold; so, it was an administrative requirement.

For 1982, though, the Statistics of Income Division managed to convince the other interested parties within the Service that the age of the decedent could be useful. Rather than have a clerk calculate the age, though, the Service decided to include the Date of Birth. Gender, which could have been an important stratifier, is not available.

► The Stratifiers

Longitudinal studies in the sociology field have long noted that there are three effects to the group under observation: current events, time of life, and age cohort. We cannot easily address this last effect, that of the age cohort, at least not in the near future, because the observations on this group trickle in over such a long time.

We could address the aspect of current events' effect by focusing on all the decedents in a single year. "Current events," in this context, means not only the operation of economic conditions, but also the tax provisions then in force. Years ending in 2, 6, or 9 were selected; so, the first focus year included in this review is 1982.

Likewise, we could address the "time of life" through the age of the decedent (since we have the dates for both birth and death). This sociological concern has an economic component in the nature of financial holdings. For example, middle-aged people are often counseled to focus their investment strategy on growth, while retirees frequently look to revenue- producing equities. One tax consideration that arises is the unrealized capital gains included in the estate. By considering the age of the decedent, then, we can improve the measures in the composition of estates.

Age can also improve the reliability of the personal wealth estimates, which depend on this factor in the construction of the weighting classes.

Age and a focus year, though, would not aid in reducing the sampling error of the monetary estimates all that much, though. For that, we needed a variable that was reasonably correlated with the key amounts of interest. Given that this is a general sample to support ambiguous analysis (at the time of the design, anyway), that left Total Gross Estate as the monetary stratifier.

▶ Selection Method

Since the selection process was computerized, we took advantage of a Bernoulli mechanism, the "Transformed Taxpayer Identification Number," used in selecting other IRS Business Master File samples, such as for the Corporations and Partnership Studies (Harte, 1986). This permanent random number procedure was meant to improve the year-to-year estimates of change by increasing the likelihood of an entity being included in the sample in succeeding years. Clearly, this is not an issue for Estates, but it did reduce the programming burden.

The selection probabilities were set within strata, with those records with a Transformed Taxpayer Identification Number below the designated probability selected for the sample.

In addition to that selection process, a 1-percent Continuous Work History Sample (CWHS) set of ending digits for the Social Security numbers was employed. We felt that, since some of the CWHS digits were in use for the Statistics of Income Individual Study, this might allow a greater overlap between the two studies.

▶ Strata Boundaries

There are two sets of boundaries that need to be determined: age, and size of Gross Estate. Fortunately, in the later case, our task was simplified by the administrative systems. Each return was assigned a Gross Estate Code, manually, based on the size of the Estate. At the time this design was first implemented, the value itself was not available.

Gross Estate Codes, shown in Figure 3 below, with a value of less than 6 were for returns below the filing threshold in 1982, and thus were not subjected to the Bernoulli sampling. These smaller estates were filing for the record only, though we did sample them using the CWHS digits.

Figure 3.--Defining the Gross Estate Code

Size of Gross Estate	Code
Under $300,000	1 - 5
$300,000 under $500,000	6
$500,000 under $1,000,000	7
$1,000,000 under $5,000,000	8
$5,000,000 or More	9

Determining the age groups was a more difficult problem. The sample has to address two populations: the estates affected by the tax law and the living population for the Personal Wealth Estimates. In addition, we made the assumption that the age distributions within the Gross Estate categories would have a significant impact; so, we planned separate age classes for the various Gross Estate Codes. The reasoning was that, as age increases, the opportunity to accumulate wealth also increases. Thus, the median age for the smaller estates' decedents would be less than that for larger estates.

The data we had available at that time were from the 1977 Estates Study, which as we noted above had but three strata based on the size of Gross Estate. The estimates were tallied into 5-year bands. As one might expect, given the nature of the population under consideration, most of the low age-groups were empty of observations.

Over the years from 1977 to 1982, though, the number of estates in each category grew, even as the total number declined due to a rise in the filing threshold. This growth resulted from both inflation effects and the normal growth of the economy.

That growth adjustment only addresses the expected filing volume, not the population of interest. To address this, we need a further adjustment to predict the population of the living wealthy. That adjustment was the inverse of the mortality rate developed by the National

Center for Health Statistics, NCHS (then, in 1980, the data were in a pamphlet; now, they are available on their Web site).

The main reason for using the estimated wealthy population instead of the expected filings of estate Tax Returns is that we wished to focus on the scarcity of "youthful decedents." This mortality- weighted set of estimates allowed us to determine, in effect, what age a "youthful decedent" might be.

We used the Dalenius-Hodges' cumulative square root of the frequency method to find reasonable strata boundaries, with a goal of choosing five groups (Dalenius and Hodges, 1959). In the end, a sixth was added because there were a fair number of cases where there was no age reported. In later years, this "Age Unknown" group was folded into the highest- age category because research showed that these decedents actually were members of that group, and the numbers became quite small.

While the strategy outlined above was applied to the estates within the focus year, some felt that, with appropriate "aging" of assets for decedents from other years, we might be able to create better Personal Wealth estimates. Hence, as is seen in Table 1, some strata are reserved for "young," nonfocus-year decedents.

The later sample design tables show this strategy was revisited after the first focus year, and the strata for nonfocus-year filings expanded, duplicating the strata outline of the focus year. This revision reflected an increase in funding for this series of projects, as well as better meeting the need for data on the annual processing operations.

▶ Sample Allocation

Weighted strata variances for the value of Gross Estate (the value of all of an estate's assets) were available from the prior 1977 study. Since the data collection is from administrative records, without any costs related to contacting a taxpayer, we simply assumed that the costs were essentially the same regardless of the stratum. The sample size was set at about 13,000 records per year.

Neyman Allocation (with a set sample size or otherwise) also requires a population estimate. Since we are primarily interested in the effect of the tax law as it is applied in a given year, and that law has effects on the living as well as the estates, the appropriate population was the same as the one used to find the age-strata breaks.

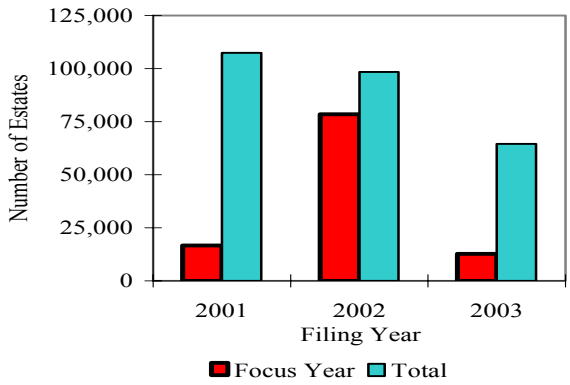
For the initial 1982 study, we allocated sample to strata under the plan for sampling the returns over 3 years, concentrating only on the year of death of the decedent, and ignoring the year of filing the administrative record.

Since the "Personal Wealth" population is more numerous than the Estates population, there were a lot of cases where the allocation prescribed more sample than there were expected estate filings. Thus, the allocation was reiterated several times, removing the certainty strata each time, before the final design's sample sizes were derived.

These sample sizes, when divided by the expected filing volumes, became the sampling probabilities used in the Bernoulli selection. These are the sampling rates shown in Tables 1 through 5, below, exclusive of the CWHS sample selections.

As a result of the filing pattern, as in the example shown in Figure 4, only about 15 percent of the sample, or about 2,000 records, were to be designated in the first year of the study, and a similar amount in the final year of the set.

Figure 4.--Estates For Decedents Who Died During 2001



Starting with the 1986 Estates Study, while the allocation of the sample to the focus year was set at the target 10,000 to 15,000 records, the difference between the expected sample size in any given filing year and the target was allocated to the nonfocus- year records within a filing year. Thus, using 2005 as an example (Focus Year 2004), while the overall sample size is about 10,000 records, about 3,000 were allocated to estates of decedents who died before 2004 or in 2005.

The allocation for nonfocus-year returns used the expected filing volume of returns, instead of the population of the wealthy used in the allocation for the focus-year strata.

► Changes--1986 to 2004

The initial design, in Table 1, shows the result of having age stratification dependent on the Gross Estate class. Although we show a zero probability of selection for the "Under $300,000" Gross Estate classes and other strata, those records were subjected to the 1-percent CWHS selections.

For the 1986 version of the design, shown in Table 2, the age groups were made independent of Gross Estate and were replicated for the nonfocus- year decedents. This also resulted in new age boundaries.

(Note, in this table and in subsequent ones, we will not show the classes that fall below the filing threshold due to space constraints. We used red to highlight the changes as well.)

The 1989 edition of the design, Table 3, also shows only a minor change: the introduction of an age group "65 under 75."

The next significant change arose for the 1992 study (Table 4). Here, we were finally able to replace the Gross Estate Code with the actual amount and thus expand the stratification. This design outline stood for about a decade.

The anticipated changes to the Estate Tax Law in 2001 left the design, Table 5, in some question. As a result, instead of planning to select the earliest filings for the Focus Year (2001 decedents) at the same rates as filings in later years, we planned on the initial year's sample to support estimation by itself. The focus-year pattern was also amended; so, the Statistics of Income studies will coincide with the Federal Reserve Board's Survey of Consumer Finance.

As of this writing, the tax law is still subject to change, but at least one update, having the strata boundaries match the filing thresholds, is planned for 2007.

► Future Research

The current trend for the tax law suggests that, in a few years, we will be canvassing the entire population, and, under some legislation, this part of the tax code would expire. However, at some future time, there may again be reason to sample a successor tax return, for one lesson from history is certainly that the Estate Tax may someday be revived. We hope that, should that arise, this paper might be of some help to that future statistician.

One more immediate issue that the Estates and Personal Wealth studies have is that the original filings on which they are based may be prone to errors in the reporting, and especially underreporting of financial assets. When such problems are discovered, the executor or lawyer will file amended returns. While such amendments are possible with other types of tax filings, because the sole person knowledgeable about the various holdings for an estate has passed away, it may be that the effect would be more serious. At this time, we simply do not have the data to examine this issue.

However, we are starting to accumulate a database that might permit such research in a few years.

► References

Dalenius, T. and Hodges, J.L. Jr. (1959),"Minimum Variance Stratification," Journal of the American Statistical Association.

Johnson, Barry W. and Woodburn, R. Louise (1994), "The Estate Multiplier Technique, Recent Improvements for 1989," Compendium of Federal

Estate Tax and Personal Wealth Studies, Publication 1773 (4/94), Department of the Treasury, Internal Revenue Service.

Harte, J. M. (1986), "Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Kopczuk, Wojiech and Saez, Emmanuel (2004), "Top Wealth Shares in the United States, 1916-2000: Evidence From Estate Tax Returns," Working Paper 10399, National Bureau of Economic Research (http://www.nber.org/papers/w10399).

Table 1.--Strata and Selection Probabilities, 1982

Age of Decedent	*Size of Gross Estate (Based on Gross Estate Code)*				
	Under $300,000	***$300,000 under $500,000***	***$500,000 under $1,000,000***	***$1,000,000 under $5,000,000***	***$5,000,000 or More***
Decedent Died in 1982					
Under 45	0	1.00	1.00	1.00	1.00
45 under 55		0.50	1.00		
55 under 60		0.35			
60 under 70			0.50		
70 or Older		0.10	0.25		
Unknown		0.10	0.25		
Decedent Died in a Year Other Than 1982					
Under 45	0	1.00	1.00	1.00	1.00
45 or Older, or Unknown	0	0	0	0	1.00

Table 2.--Strata and Selection Probabilities, 1986

Age of Decedent	***Size of Gross Estate (Based on Gross Estate Code)***		
	$500,000 under $1,000,000	**$1,000,000 under $5,000,000**	**$5,000,000 or More**
Decedent Died in 1986			
Under 40	1.00	1.00	1.00
40 under 50	1.00	1.00	1.00
50 under 65	0.35	1.00	1.00
65 or Older, or Unknown	0.07	0.50	1.00
Decedent Died in a Year Other Than 1986			
Under 40	1.00	1.00	1.00
40 under 50	0.25	0.35	1.00
50 under 65	0.04	0.50	1.00
65 or Older, or Unknown	0.01	0.01	1.00

Table 3.--Strata and Selection Probabilities, 1989

Age of Decedent	***Size of Gross Estate (Based on Gross Estate Code)***		
	$500,000 under $1,000,000	**$1,000,000 under $5,000,000**	**$5,000,000 or More**
Decedent Died in 1989			
Under 40	1.00	1.00	1.00
40 under 50	1.00	1.00	1.00
50 under 65	0.50	1.00	1.00
65 under 75	0.12	0.50	1.00
75 or Older, or Unknown	0.12	0.50	1.00
Decedent Died in a Year Other Than 1989			
Under 40	1.00	1.00	1.00
40 under 50	0.25	0.35	1.00
50 under 65	0.05	0.06	1.00
65 under 75	0.03	0.05	1.00
75 or Older, or Unknown	0.03	0.05	1.00

Table 4.--Strata and Selection Probabilities, 1992

Age of Decedent	Size of Gross Estate: $600,000 under $1,000,000	$1,000,000 under $2,000,000	$2,000,000 under $5,000,000	$5,000,000 under $10,000,000	$10,000,000 or More
Decedent Died in 1992					
Under 40	1.00	1.00	1.00	1.00	1.00
40 under 50	1.00	1.00	1.00	1.00	1.00
50 under 65	0.22	0.44	1.00	1.00	1.00
65 under 75	0.10	0.20	0.40	1.00	1.00
75 or Older, or Unknown	0.03	0.06	0.18	1.00	1.00
Decedent Died in a Year Other Than 1992					
Under 40	1.00	1.00	1.00	1.00	1.00
40 under 50	0.15	0.20	1.00	1.00	1.00
50 under 65	0.06	0.11	0.33	1.00	1.00
65 under 75	0.06	0.11	0.33	0.45	1.00
75 or Older, or Unknown	0.03	0.05	0.16	0.22	1.00

Table 5.--Strata and Selection Probabilities, 2001

Size of Gross Estate	Age of Decedent: Under 40	40 under 50	50 under 65	65 or Older
Decedent Died in 2001				
$675,000 Under $1,000,000	1.00	1.00	1.00	0.13
$1,000,000 under $1,500,000	1.00	1.00	1.00	0.20
$1,500,000 under $2,000,000	1.00	1.00	1.00	0.20
$2,000,000 under $3,000,000	1.00	1.00	1.00	0.40
$3,000,000 under $5,000,000	1.00	1.00	1.00	0.80
$5,000,000 under $10,000,000	1.00	1.00	1.00	1.00
$10,000,000 or More	1.00	1.00	1.00	1.00
Decedent Died in a Year Other Than 2001				
Under $1,000,000	1.00	0.01	0.01	0.01
$1,000,000 under $1,500,000	1.00	0.01	0.01	0.01
$1,500,000 under $2,000,000	1.00	0.01	0.01	0.01
$2,000,000 under $3,000,000	1.00	0.02	0.02	0.02
$3,000,000 under $5,000,000	1.00	0.04	0.04	0.04
$5,000,000 under $10,000,000	1.00	0.11	0.11	0.11
$10,000,000 or More	1.00	1.00	1.00	1.00