The Effect of Late-Filed Returns on Population Estimates: A Comparative Analysis

Brian Raub, Cynthia Belmonte, Paul Arnsberger,
Melissa Ludlum[1]
[1]IRS, Statistics of Income, 1111 Constitution Ave NW, Washington DC 20024

Abstract
This paper will examine the effect of late-filed returns on population estimates for three Statistics of Income (SOI) programs. Estimates for populations of interest for each study are produced by drawing stratified, random Bernoulli samples of tax and information returns as they are filed, over periods that span a predetermined number of years. While this methodology results in the inclusion of the majority of targeted returns, a small number of returns for each study are filed beyond the data collection period. These late-filed returns may introduce non-response bias into the population estimates, which might be mitigated by post-stratification or weighting adjustments. This paper will be a case study on the effects of truncated sampling periods on population estimates, using the three SOI programs as case studies, and will provide a comparative analysis.

Key Words: non-response bias, post-stratification, Bernoulli, population estimates

1. Data Sources and Background

The Statistics of Income (SOI) division of the Internal Revenue Service (IRS) collects and disseminates detailed data based on samples of administrative records, including tax and information returns. Estimates for populations of interest for SOI studies are produced by drawing stratified, random Bernoulli samples of tax and information returns as they are filed, over periods that span a predetermined timeframe. While this methodology results in the inclusion of the majority of targeted returns, a small number of returns for each study are received beyond the data collection period. These "late-filed" returns may introduce non-response bias into the population estimates, which might be mitigated by post-stratification or weighting adjustments. (The term “late-filed return” as used in this paper does not address the compliance, or lack thereof, of return filings with statutory requirements.) Using three SOI studies with varying sampling frames, this paper will function as a case study on the effects of truncated sampling periods on population estimates.

The data presented in this paper are derived from two sources--sample data produced by SOI and administrative data obtained from the IRS Masterfile for the population of returns filed. SOI sample data typically include detailed, error-perfected financial and other information about the tax filing entity. SOI sample data are used to produce population estimates that are used in statistical studies and for analysis of tax policy. Data obtained from the IRS Masterfile include limited information for the population of filers. This information is generally used for a variety of purposes related to tax administration.

SOI conducts annual studies of a wide range of filers, including individuals, corporations, partnerships, estates, trusts, tax-exempt charitable organizations, and many other filers. This paper focuses on three SOI studies – the Estate Tax study, the Private Foundation study, and the Exempt Organization study.

1.1 The Estate Tax Study

With its annual Estate Tax study, SOI extracts demographic, financial, and asset data from Federal estate tax returns. The annual study allows production of a data file for each filing, or calendar, year. By focusing on a single year of death for a period of 3 filing years, the study allows production of periodic year-of-death estimates. A single year of death is examined for 3 years, as over 98 percent of all returns for decedents who die in a given year are filed by the end of the second calendar year following the year of death. Data included in this paper are for Year of Death 2004 and were obtained from returns filed in Calendar Years 2004-2006.

The estate of a decedent who, at death, owns assets valued in excess of the estate tax applicable exclusion amount, or filing threshold, must file a Federal estate tax return, *Form 706, U.S. Estate (and Generation-Skipping Transfer) Tax Return*. For decedents who died in 2004, the exclusion amount was $1.5 million. Alternate valuation may be elected only if the value of the estate, as well as the estate tax, is reduced between the date of death and the alternate date. The estate tax return is due 9 months from the date of the decedent's death, although a 6-month filing extension is allowed. In some cases, longer filing extensions may be permitted.

For the Year of Death 2004 Estate Tax study, there were 11,817 Form 706 returns in the sample selected from a population of 42,424. The SOI Estate Tax study is classified into strata based on year of death, the size of total gross estate, and age of the decedent. For the Year of Death 2004 study, there were a total of 57 sampling strata, with sampling rates ranging from 4 percent to 100 percent.

1.2 The Private Foundation and Exempt Organization Studies

The annual SOI studies of private foundations and exempt organizations collect detailed financial data, as well as information on charitable and grant-making activities and compliance with IRS regulations from information returns filed by exempt organizations. Studies are conducted for a single tax year and include samples of returns filed and processed during the 2 calendar years immediately following the target tax year. Data discussed in this paper for the Private Foundation and Exempt Organization studies were obtained for Tax Year 2004 returns filed in Calendar Years 2005 and 2006. The Tax Year 2004 samples include organizations with accounting periods beginning in Calendar Year 2004 (and ending between December 2004 and November 2005), for which returns were filed and processed to the IRS Business Masterfile during Calendar Years 2005 and 2006. While this 2-year sampling period ensures almost complete coverage of the target population, there are still a number of returns processed after the close of the second year (i.e., December 31, 2006 for the Tax Year 2004 study), which are generally excluded from the samples.

Private foundations and nonexempt charitable trusts are required to file Form 990-PF (*Return of Private Foundation or Section 4947(a)(1) Nonexempt Charitable Trust Treated as Private Foundation)* annually. Similarly, certain exempt organizations are required to file Forms 990 *(Return of Organization Exempt from Income Tax)* or Form 990-EZ *(Short Form Return of Organization Exempt from Income Tax)*. SOI conducts

annual studies based on samples of Forms 990-PF, 990, and 990-EZ filed for a given tax year. These information returns are due 5 months after the close of the organization's accounting period, although a 3-month filing extension is allowed. In some cases, additional filing extensions may be granted.

For the Tax Year 2004 Private Foundation study, there were 7,805 Form 990-PF returns in the sample, selected from a population of 80,570. The SOI Private Foundation study is classified into strata based on the size of end-of-year fair market value of assets, with each stratum sampled at a different rate. Sampling rates ranged from 1 percent for private foundations with total assets less than $125,000 to 100 percent for private foundations with total assets of $10 million or more.

The Tax Year 2004 exempt organization sample of section 501(c)(3) filers comprised 15,070 Forms 990 and 990-EZ, selected from a population of 279,415. End-of-year book value of assets was the stratifying variable for the exempt organization study. Sampling rates ranged from 1 percent for exempt organizations with total assets less than $500,000, to 100 percent for those with total assets of $50 million or more.

2. Late-Filed Returns

To examine the effect of late-filed returns on each of the studies, an augmented sampling frame, which incorporates 2 years of additional return filings, was constructed from IRS Masterfile data. The following tables show the number of late-filed returns received within the current and augmented sampling frames, as well as the percentage of selected financial variables represented by returns received inside and outside of the sampling period.

Table 1, below, shows the percentage of Year of Death 2004 Forms 706 filed, total gross estate, and net estate tax reported for returns filed over a 5-year collection period (2004-2008), by size of gross estate and by age of the decedent More than 98 percent of all Year of Death 2004 Forms 706 filed over the 5-year period were received within the 3 years, 2004 through 2006, from which returns were sampled. However, the estates of younger decedents filed returns outside of the 3-year sampling frame proportionately more often than the estates of their older counterparts. For example, nearly 4 percent of returns filed for decedents under 40 were received in 2007 and 2008. The percentage of total gross estate represented by late-filed returns was 1.1 percent, with the corresponding figure for net estate tax only 0.5 percent. These smaller percentages are attributable to the fact that late-filed returns were smaller on average than other returns and were proportionately more often nontaxable, as shown in the following tables.

Table 1: Estate Tax Returns Filed for 2004 Decedents, IRS Masterfile Data by Age of Decedent, 2004-2008

Calendar Year	*Returns*	*Total gross estate*	*Net estate tax*
2004-2006	**98.4%**	**98.9%**	**99.5%**
Under 40	96.4%	97.0%	100.0%
40 under 50	97.0%	97.5%	98.9%
50 under 65	97.2%	97.6%	98.7%
65 and over	98.6%	99.0%	99.6%
2007-2008	**1.6%**	**1.1%**	**0.5%**
Under 40	3.6%	3.0%	0.0%
40 under 50	3.0%	2.5%	1.1%
50 under 65	2.8%	2.4%	1.3%
65 and over	1.4%	1.0%	0.4%

Table 2 examines the same population as the previous table, classified by size of total gross estate. The table shows that returns for the smallest estates, those with between $1.5 and $2 million in gross estate, were filed in the 2 years immediately succeeded the sampling period twice as frequently as were returns for the largest estates.

Table 2: Estate Tax Returns Filed for 2004 Decedents, IRS Masterfile Data by Size of Total Gross Estate, 2004-2008

Calendar Year	*Returns*	*Total gross estate*	*Net estate tax*
2004-2006	**98.4%**	**98.9%**	**99.5%**
$1.5 million < $2.0 million	98.2%	98.2%	98.9%
$2.0 million < $3.0 million	98.3%	98.3%	99.1%
$3.0 million < $5.0 million	98.4%	98.4%	99.1%
$5.0 million < $10.0 million	98.8%	98.8%	99.5%
$10.0 million and over	99.1%	99.6%	99.7%
2007-2008	**1.6%**	**1.1%**	**0.5%**
$1.5 million < $2.0 million	1.8%	1.8%	1.1%
$2.0 million < $3.0 million	1.7%	1.7%	0.9%
$3.0 million < $5.0 million	1.6%	1.6%	0.9%
$5.0 million < $10.0 million	1.2%	1.2%	0.5%
$10.0 million and over	0.9%	0.4%	0.3%

Table 3 examines the same population as Tables 1 and 2, classified by tax status of the return. It shows that nontaxable returns were filed outside of the sampling period more than twice as often as taxable returns.

Table 3: Estate Tax Returns Filed for 2004 Decedents, IRS Masterfile Data by Tax Status, 2004-2008

Calendar Year	*Returns*	*Total gross estate*	*Net estate tax*
2004-2006	**98.4%**	**98.9%**	**99.5%**
Taxable	99.1%	99.4%	99.5%
Nontaxable	97.9%	98.2%	N/A
2007-2008	**1.6%**	**1.1%**	**0.5%**
Taxable	0.9%	0.6%	0.5%
Nontaxable	2.1%	1.8%	N/A

Table 4 illustrates the extent to which estimates based on Form 990-PF data collected from the current 2-year sampling period might be enhanced by using the augmented sampling frame. More than 98 percent of the Tax Year 2004 private foundation returns included in the augmented sampling frame were processed in the 2 years immediately following the close of the tax year. A closer examination reveals that the percentage of returns received and processed during the first 2 years increases with asset size. For example, 97.9 percent of returns filed by small organizations (those with assets less than $1,000,000) were processed during the 2005-2006 period, compared to 99.2 percent of the returns of medium-sized foundations (those with assets between $1 million and $50 million), and 99.7 percent of the returns of the largest foundations (those with assets of $50 million or more).

Table 4: Tax Year 2004 Private Foundation Information Returns: IRS Masterfile Data by Calendar Year and Size of Organization, 2005-2008

Calendar Year	*Returns*	*Assets*	*Revenue*	*Charitable Disbursements*	*Excise Tax on Net Investment Income*
2005-2006	**98.3%**	**99.5%**	**99.4%**	**99.5%**	**99.5%**
Small	97.9%	98.8%	98.7%	98.9%	98.8%
Medium	99.2%	99.3%	99.3%	99.4%	99.5%
Large	99.7%	99.6%	99.6%	99.7%	99.6%
2007-2008	**1.7%**	**0.5%**	**0.6%**	**0.5%**	**0.5%**
Small	2.1%	1.2%	1.3%	1.1%	1.2%
Medium	0.8%	0.7%	0.7%	0.6%	0.5%
Large	0.3%	0.4%	0.4%	0.3%	0.4%

Table 5 shows the breakdown of data from Forms 990 and 990-EZ returns by filing period and size of assets. As with private foundations, the vast majority of Tax Year 2004 returns were filed in the first two years after the end the tax year Again, a large portion of the returns filed in the final 2 years of the augmented sampling frame are from small organizations – those with total assets less than $100,000. Consequently, late filers of Forms 990 add little to the aggregate totals for most of the financial variables collected: less than 1 percent of total assets, revenue, and net worth.

Table 5: Tax Year 2004 Exempt Organization Information Returns, IRS Masterfile Data by Processing Year and Size of Organization, 2005-2008

Calendar Year	*Returns*	*Assets*	*Revenue*	*Net Worth*
2005-2006	**97.3%**	**99.3%**	**99.3%**	**99.2%**
Small	95.7%	96.4%	96.2%	96.1%
Medium	98.3%	98.8%	98.8%	97.8%
Large	99.2%	99.3%	99.3%	99.3%
2007-2008	**2.7%**	**0.7%**	**0.7%**	**0.8%**
Small	4.3%	3.6%	3.8%	3.9%
Medium	1.7%	1.2%	1.2%	2.2%
Large	0.8%	0.7%	0.7%	0.7%

3. Current Treatment of Late Filers

Although the Estate Tax, Private Foundation, and Exempt Organization studies share a common challenge in addressing the effect of late-filed returns on population estimates, each of the three studies currently uses a different approach in dealing with this challenge.

Year of Death population estimates for the Estate Tax study include weight adjustments for late-filed returns. Weight adjustment factors are calculated based on past late filing patterns using historical data from the IRS Masterfile and are updated periodically. The aim of using these weight adjustments is to improve the overall population estimates as well as estimates for the subpopulations of returns that have historically filed late with greater frequency. As shown in Table 6, weight adjustment factors varied by size of estate, tax status of return, and age of decedent. For each size of estate and age combination, non-taxable returns received a higher adjustment factor than taxable returns. Estates with $10 million or more in gross estate received weight adjustment factors based on tax status regardless of age, as illustrated in the top portion of the table. For estates with less than $10 million in gross estate, weight adjustment factors were assigned based on tax status and age.

Table 6: Weight Adjustment Factors for Year of Death 2004
Estate Tax Population Estimates

	Total gross estate ≥ $10 million	
	Taxable	*Non-taxable*
All ages	1.004	1.013
	Total gross estate < $10.0 million	
Age	*Taxable*	*Non-taxable*
Under 40	1.036	1.052
40 under 50	1.019	1.035
50 under 65	1.018	1.028
65 and older	1.009	1.02

Table 7, shows the aggregate effect of weight adjustment factors on the Year of Death 2004 estate tax estimates. The number of returns increased about 1.5 percent compared to a 1.2 percent increase in total gross estate and less than a 1 percent increase in net estate tax. The differences in the impact of weight adjustments on these three variables is consistent with the fact that late-filed returns comprised proportionately more small returns and non-taxable returns than the population as a whole.

Table 7: Effect of Weight Adjustments on Estimates of Year of Death 2004
Estate Tax Population
[Money amounts are in millions of dollars]

	Returns	*Total Gross Estate*	*Net Estate Tax*
Unadjusted estimate	41,599	183,657	22,075
Estimate with weight adjustment	42,239	185,921	22,220
Percentage increase	1.54	1.23	0.66

In contrast, population estimates for the Private Foundation study do not include standard adjustment factors to account for returns filed after the close of the two-year sampling period. Instead, during file closeout, efforts are made to identify and include late-filed returns of private foundations that would have been sampled at the 100-percent rate (i.e., organizations with fair market value of assets of $10 million or more). These include returns of organizations sampled in previous study years, as well as returns of organizations posting to the IRS Masterfile for the first time. Potentially, this can extend the 2-year sampling frame by four to five months, the typical length of time between the end of the sampling period and the creation of the final study file. Table 8, shows population estimates for selected variables from SOI's Tax Year 2004 Private Foundation study. The table includes population estimates from returns processed during the regular 2-year sampling period, as well as enhanced population estimates including adjustments for late-filed returns. Only 11 large-case, late-filed returns were added to the Tax Year 2004 sample. These returns represented 100^{th} of 1 percent of the population estimate, and about a one-fifth of 1 percent addition to total revenue, charitable disbursements, and net investment income excise tax.

Table 8: Tax Year 2004 Private Foundation Data from SOI Estimates Including Added Late-Filed Returns

[Money amounts are in millions of dollars. Detail may not add to totals because of rounding.]

	Returns	*Assets*	*Revenue*	*Charitable disbursements*	*Excise tax on net investment income*
Additional data from late-filed returns	11	453	129	54	1
Enhanced SOI estimate	76,897	509,924	58,668	32,125	469
Additional data as percentage of total	0.01	0.09	0.22	0.17	0.21

The Exempt Organization study includes no weight adjustments and no attempt is made to add returns to the sample that are filed outside of the two-year sampling frame. Adjustments to the sample are made for certain organizations that file returns within the 2-year sampling period. Examples of these adjustments include rejecting short-year returns and those that file with an incorrect subsection code; and adding returns that have posted incorrectly to the Masterfile as duplicate, below the filing threshold, or with incorrect total assets.

Using IRS Masterfile data as a proxy, we can mimic the Private Foundation study's technique of processing returns from the certainty strata that are filed within five months after the close of the normal sampling period. Based on the Masterfile data, 21 large-case returns would have been added to the Tax Year 2004 sample. These returns would have accounted for a one-third of 1 percent addition to the aggregate totals for assets, revenue and net worth.

4. Strengths/Weaknesses

These analyses reveal a number of strengths and weaknesses for each of the three approaches to the late-filer problem. The weight adjustment approach, as employed for the Estate Tax study, potentially improves the overall population estimates. It also may improve estimates for subpopulations for which returns have historically been filed late with the greatest frequency. The adjustments seem to be an effective means of counteracting any bias that may result from the existing sampling period. To the extent that late filers create bias in the Estate Tax study estimates, the weight adjustment approach may mitigate the bias.

On the other hand, the weight adjustment approach may not always be an effective method of predicting filing habits. The weight adjustments are developed from observed trends in historical data; this information may not always reliably predict future filing patterns. Although the characteristics of late filers have been relatively stable over time, significant changes to the estate tax law could alter these patterns.

The inclusion of large, late-filed returns in the Private Foundation study provides for more complete coverage of the target population by including returns that would have been selected with certainty within the defined sampling period. Additionally, this approach ensures that files are suitable for time-series analysis of a specific organization or panel of organizations. This strength may be unique to data for tax-exempt organizations, whose information returns, in most cases, are not subject to the same disclosure and confidentiality rules as data obtained from tax returns filed by other types of organizations and individuals.

The primary weakness of including large, late-filed returns only in the enhanced estimate is the inconsistency that it introduces. Slight variances in tax return processing, sample file creation or review, or sample file delivery date can affect the sampling period from which the enhanced estimate is drawn from year to year. Further, the method fails to address late-filed returns of smaller organizations, which account for the largest share of the late-filing population.

The "no-adjustment" approach that is used in the Exempt Organization study ensures a consistent sampling frame with a well-defined sampling period. This approach employs the Bernoulli sample over a 2-year period and does not include arbitrary additions or discontinuations. Because the population is framed as the estimate of filers of the 2-year period and not as the "universe" of filers, the bias does not exist.

Because, unlike the weight adjustment method used in the Estate Tax study, the "no-adjustment" approach does not attempt to account for late filers, it could consistently underestimate the number of returns filed by smaller organizations. By omitting some large case returns that are received outside of the defined sampling period, this approach also provides a somewhat less complete dataset for time-series panel analysis than does the Private Foundation study.

5. Conclusions / Future Research

Late-filed returns present a common challenge for studies of data obtained from tax returns, such as the Estate Tax, Private Foundation, and Exempt Organization studies. Although, for each of the studies, the number of late-filed returns is modest in comparison to the number of returns filed within the defined sampling period, the absence of these returns may introduce bias into the population estimates.

Currently, each of the three studies discussed in this paper uses a unique approach to mitigate the potential bias introduced by late filers. The weight adjustment method, employed for the Estate Tax study, improves some aspects of the study’s estimates, but could become distorted if filing patterns observed in historical data do not continue into the future. The enhanced Private Foundation estimate, which is obtained by including targeted returns received after the end of the sampling period, benefits time-series analysis. However, it creates inconsistencies in the year-to-year sampling period. The "no adjustment" method used for the Exempt Organization study provides a distinct sampling period, but does not address the exclusion of relatively small filers from the estimates.

The unique characteristics of late filers in each of the three studies discussed in this paper, as well as the benefits and shortfalls of using each of the three approaches to address the later-filer problem, provide a number of opportunities for further research.

This analysis will be expanded to research additional tax years and years of death in order to explore historical filing patterns. This effort will attempt to isolate an optimal sampling period that balances population coverage with timeliness of completion of the estimates. Additionally, weighting adjustments, similar to those in use for the Estate Tax study, will be developed for the Private Foundation and Exempt Organization studies. The adjustments will be examined for accuracy, as well as their effect on organization-level data from year-to-year.