Statistics of Income Partnership Studies: Evaluation of Preliminary Estimates

Paul B. McMahon, Internal Revenue Service

This is the fourth paper in a continuing series on efforts to improve the U.S. Internal Revenue Service's Statistics of Income Partnership Studies. The previous papers in this series dealt with updating the sample design in the face of changing tax legislation and regulations (McMahon, 1993; McMahon, 1991; McMahon, Collins, and O'Conor, 1990). This paper shifts the focus to the estimation problems encountered with the introduction of a new product, a preliminary data file and tabulations.

First some background on the project, the sample design and its goals are briefly described, then we inspect the current product, review the procedures used, and discuss proposed improvements. We close with comments on future research.

■ Background

We are now in the fifth decade of producing annual data on partnership operations in the United States. The firms in which we are interested are unincorporated businesses with two or more owners and are required to report income, deductions, distributions, tax credits and preferences, and other information to the U.S. Internal Revenue Service annually.

The advantages of using administrative records as data sources have been well explored, but this particular case, the use of tax records, deserves a few additional remarks. Unlike many types of administrative records, such as birth certificates, driver's licenses or auto registrations, these source documents are rich in detail. For example, the income and deduction schedule contains 22 items, including Gross Receipts, Cost of Goods Sold and Depreciation Deduction. Many of these items must be supported by subsidiary schedules reporting how the figures were determined. Cost of Goods Sold, for example, is supported by a schedule that includes data on purchases, inventories, personal withdrawals and labor, while Depreciation requires details on each property, including cost and purchase date. With all these additional data the physical size of these reports can be weighty indeed.

Often the supporting schedules are computer generated listings, handwritten accounts or in other non-standardized formats, with only a very small fraction sending the reports in an electronic medium. While much of the report may be customized, the definitions and derivations of the data are strictly regulated, but the variety of reporting characteristics raises the cost of abstracting the data.

This partially explains why only selected items are abstracted from the paper documents onto computer records. These records -- transactions, in accounting terms -- are then posted to the Business Master File accounts. "Posted" means that there is an exact match between a transaction and an account on the master list and that both the account and the transaction record are updated. These updated transactions are then passed on for further uses.

One of these uses is as the sampling frame for the selection of the Statistics of Income samples. The Partnership Study uses many items on these records to compute values for receipts and income that are consistent with economic definitions (rather than the more complex tax law versions). These computed values, with assets and an industry code, are used in the sampling operation to divide the population of about 1.5 million companies into about 70 strata (for the Tax Year 1993 study -- about 45 strata in recent years).

Once a record is classified into a stratum, it is then selected (or not) based on the prescribed sampling rate for that class and a random number generated from the Employer Identification Number. The

sampling rates range from less than 0.2 percent to 100 percent, yielding a sample size of about 30,000 records. These records are selected as soon as they are processed, usually soon after the return is received.

A partnership's report is due three and a half months after the close of its business year (often June 30th). The strict deadlines and very real penalties for late filing are offset by a policy of granting virtually automatic extensions to those filing deadlines, provided only that there will not be an increase in tax liabilities. As a result, a significant number of firms send in their returns in mid-October. This policy stretches out the sample selection period to a full year and, as we shall see, has a profound effect on the preliminary estimates.

There are two main uses for our complete report's data: for analysis of historical trends using the data from many years and as a basis for predicting the effects of various economic and, especially, taxation policies. The U.S. Department of the Treasury's Office of Tax Analysis and Congress' Joint Committee on Taxation are the clients for this second use. The microdata simulations they conduct as part of their research require the actual sample records. (While the privacy provisions of the tax code permit these two organizations access to the microdata, it constrains production of public-use data files to such an extent that we have not found a way to construct one for this series.)

In recent years, especially, the Office of Tax Analysis has had ever greater need for the data early in the year. After the 1992 election, for example, there were a large number of tax proposals in the legislature that had to be evaluated. The various Statistics of Income studies of Individuals, Corporations and Partnerships form a basis for such analysis, though only in the past few years has there been a need for preliminary Partnership data. This need is the major impetus behind both the preliminary estimates and this paper. Since the clients desire preliminary estimates before Congress reconvenes in January, the preparation of the data, including the calculation of weighting factors, must be complete in early December.

This raises another timing factor that directly affects the estimation process. The tax processing systems are revised on an annual basis, with the new versions becoming operational each January. The computer programs that form this system are highly complicated and must be tested, for numerous changes are always required.

Since the sample selection process uses this system for the sampling frame, the selection programs must also be amended at this time. The statisticians and analysts preparing the preliminary files and estimates must also ensure that the sample selection system for the next year is ready. The system generates a considerable number of reports, files and controls, supporting several diverse studies, each with complex designs (including the SOI Corporations projects). Testing the annual modifications is critical to the success of future projects -- a task demanding significant attention between the end of October and mid-January. Thus, the preliminary estimation procedures must be relatively straight forward and expedient, because critical resources are fully committed elsewhere.

■ Initial Results

The Tax Year 1991 preliminary data were provided to the Treasury's Office of Tax Analysis in January 1993, more than three months before the first data from the completed study became available. We accelerated the procedures for the following year, delivering the preliminary data in early December 1993.

The preliminary files contained records processed through the data abstraction and editing operations as of some date. Other records that were incomplete or had identifiable errors were excluded from the early sample file, as were records that contained only administrative data from the sample frame. (The items on the transaction records are too few to be useful in the economic analysis, and some are of uncertain quality.)

The outlier review was similarly abbreviated. Those records that had to be researched were simply excluded from the preliminary file and tabulations,

while those less obviously outliers were noted but included. (These records were adjusted, if necessary, then included in the final file.)

Cutting off a sample in this way has a price, if only in the increased imprecision that a smaller sample size implies. To assess this price, we look at a selection of variables that are used in the sample's stratification or are used by the subject matter analysts as flags for potential problem areas. In addition, we consider the estimated number of active entities (the target population) and the number of partners.

If the returns included in the early processing were much the same as those processed later, then we would expect the results to be somewhat different, but within bounds set using the sample error from the final sample as a guide. Thus, for comparison's sake, we first examine the coefficients of variation for the two years in question. Table 1, below, shows that there has been little change between the years, as would be expected. The decrease in the coefficient for Total Assets doubtless arises from the ten percent growth in the population of the certainty strata reserved for those firms with assets over $100 million, while the improvement in Net Income is due

Table 1.--Error in Partnership Preliminary Estimates

Items	Preliminary Estimates' Relative Error		Final Estimates' Coefficients of Variation	
	1991	1992	1991	1992
Partnerships	0.8	2.0	0.5	0.5
No. of Partners	0.7	7.7	1.9	2.4
Total Assets	-4.4	2.2	0.7	0.4
Receipts	-4.7	1.0	0.4	0.3
Net Income	-0.6	4.8	1.8	0.7
Net Loss	-12.3	-10.1	1.0	1.0
Portfolio Inc.	5.4	1.6	1.4	1.4
Depreciation	-11.8	-1.8	0.9	0.7
Taxes Ded.	-6.5	-0.7	0.8	0.7
Sal. & Wages	1.5	6.3	0.8	0.7

to an increase in the proportion of companies reporting a profit.

The relative errors for the preliminary estimates are, in most cases, of a larger magnitude than the sampling error, and do not share the same stability as the coefficients of variation. Several preliminary estimates show improvement from the 1991 to the 1992 study, especially taxes and depreciation. Indeed, the average of a number of preliminary estimates shows a marginal improvement.

The 2 percent overstatement of the number of firms for Tax Year 1992 could have affected some of the economic models. For the past seven years, the population has declined by between 2.5 and 3.5 percent each year. The early data overstatement suggested that the effects of the 1986 Tax Reform Act had run their course (Petska, 1991; Petska and Nelson, 1990). The final estimate, however, showed a 2.7 percent decline.

We separated the estimates of Net Income and Net Loss to place them on a similar basis to the other estimates. It serves another purpose, highlighting as it does the severe underestimate of the Losses. This problem is clearly a consequence of the filing extension policies, for losses do not cause owners to owe more taxes, thus meeting the main criterion for permission to delay.

The preliminary estimate for Net Income went from being pretty much on target for 1991 to a nearly 5 percent overstatement in the early 1992 estimates. At the same time, there was a marginal improvement in the early estimate of Net Loss. A common thread is clearly suggested.

In developing the final estimates, we use a simple expansion weight. The robustness of this procedure is important to the final sample file, in part because the evaluation of the tax legislation depends on factors determined only after the sample has been made available for the analysis. If the sample and population counting were completed, then the weights would be the ratio of population to sample for each stratum. The estimator, then, for a population total

for some item would be:

$$\hat{X} = \sum w_i x_{ij} = \sum (N_i/n_i) x_{ij} \quad .$$

The preliminary estimator, relying as it does on the cut off sample and the available processes, replaces the N_i in the weight with a predicted value. This, along with the smaller sample size (n_i) can cause some weights used in the preliminary estimates to be 40 percent higher than the final weight.

That common thread in the overestimate of Net Income for Tax Year 1992 is evident in the overestimate of the number of active partnerships. This arose because the projected frame populations were also about two percent over the actual count (which wasn't available for another six weeks). These over-projections were not evenly distributed over the strata, as shown in Table 2, below.

Table 2.--Error in Projected Frame Population Relative Error (Percent)		
Strata Group	1991	1992
Certainty (largest)	-0.3	5.4
Large Firms	-0.3	4.6
Medium	-0.1	2.4
Small	0.4	1.4

This caused higher preliminary weights, especially in the certainty classes, that tended to improve the statement of loss firms' activities. But because firms with profits tend to file earlier in the year, it also caused overstatement for profit firms.

Why the degradation in the population projections? In part this arose because the population filings were projected for twice as long a period. This was not factored into the decision of when to cut off, largely because this was only the second year for these "advanced" estimates, so our experience did not suggest what the trade off would be.

Another cause, which had a more significant impact, was poor modelling. Figure 1 shows the pattern of weekly partnership return volumes for the sample selection system late in the year. The overall population has been declining for several years, but this pattern is reasonably stable. Although we show only three years in this graph, the pattern has been repeated for over a decade. This stability is due to both the regulations and the processing deadlines, with minor shifts due to patterns of the holidays. A small shift in the deadlines caused the volume for the 1992 Study to spike on week 46 (encircled), the one we used for the projection.

Figure 1: Filing Pattern
(End of Year)
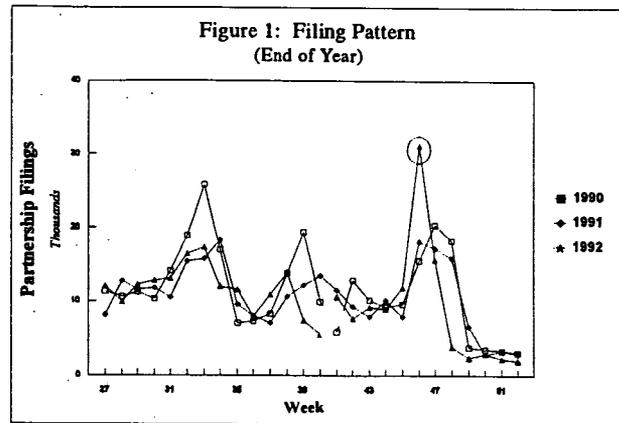
The projections, in essence, are the strata population counts through the cut-off plus estimates for the records that have not yet been filed and administratively processed. The early spike in volume caused the adjustments for the unprocessed population to be too large. In turn, the overestimate caused the over-inflation of the expansion weights, and, as Table 2 shows, this particularly affects the certainty classes.

Only a single year's strata populations were available for the projections because the sample design had undergone a major adjustment. Data from earlier years were available, but the sampling plan in use for the 1988 through 1990 studies hadn't been designed with either preliminary estimates or distributional features as goals. Thus, estimates from these years tended to rely on few, if any, observations for one of the revised strata. Also, as we have already noted, the preliminary estimates had to employ existing programs and frameworks, with minimum staff

support, due to the high demand for the key individuals at the time of the cut off.

Still, the overall estimated population could have been more accurate, for the frame definition hadn't changed and there were a few more years worth of data (counts from years before 1988 must be excluded because the 1986 Tax Reform Act reversed the growth trend of two decades). Indeed, an earlier estimate of the population was well within half a percent of the actual count, but, because the tax law was being changed, we felt that this earlier estimate was, perhaps, too low.

■ Planned Improvements

In November 1994 we will again be producing a preliminary data file and estimates, this time for the 1993 Tax Year. Drawing on the experiences of the past two years, we will make modifications to our previous practices, both in developing the strata population estimates and adjusting for the late filers. As we noted earlier though, tax reform dramatically changed the partnership population. The "phased-in" features of the law delayed part of the effect and leaves us with only five years worth of data as a base for the predictions.

Predicting the populations has two components: estimating the grand total and predicting the distribution among the strata. We will use the total to scale the individual strata estimates, so we first address that issue.

As we saw in Figure 1, there are patterns in the processing that repeat each year. Does this extend to the proportion of the population processed through some week? If so, the predictor could be quite simple, say:

$$\hat{N} = N_{yw} \, (N_{(y-1)} / N_{(y-1)w}) \, ,$$

relying on a reference week (w) and an expansion ratio from the previous year's counts.

Figure 2 shows that about week 42 there is a period where the predicted values have been on target.

We will also use week 31 as a confirmation prediction, for it appears that a half percent adjustment should give nearly the same result.

Figure 2: Prediction Error
Ratio of Predicted to Actual Population
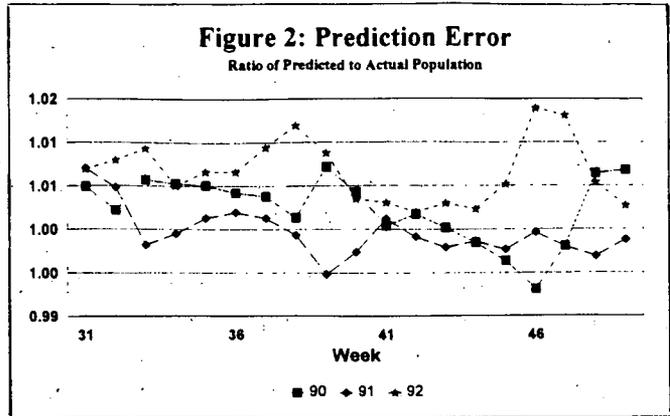
Unfortunately, the second part of predicting the strata populations isn't quite as straight forward. The sample design for the 1993 study has about 20 additional strata, with most of the rest redefined to some extent. A tally of the sampling frame for recent prior years might solve some of these problems, but resource constraints and timing issues leave these data unavailable. Moreover, since the strata used before the 1991 study were not designed with either preliminary data or distributional characteristics in mind, we are left with only two years' samples to project a distribution across the strata.

All of the predicted strata populations must begin with the actual counts of those classes through the cut-off week. Otherwise, some strata would have predictions smaller than the actual cut-off counts (as happened when we explored this option to 6 of the 45 classes for the 1991 study). Thus, we are trying to estimate only the last two months or so of filing and processing.

The number of sample selections late in Tax Years 1991 and 1992 in the strata for the certainty and large strata will easily support estimates for our predictions, because the sample sizes are reasonable and the weights small.

The small and medium classes will, on the other hand, have few observations and much larger weights. Estimating the end-of-year populations for the new

strata definitions is a simple weighted tally from the prior studies' data files. We plan to reweight the records with selection weeks before the reference date, using population and sample data as of that reference period, and tally only the early records.

Adjustments to the current study's counts, whether the stratum size is small or large, will then take the form of the overall ratio estimate above. A second approach will be substitution:

$$\hat{N} = N_{yw} + (N_{(y-1)} - N_{(y-1)w}) \ .$$

This uses the difference in the previous year's estimates in place of those not yet counted. We also group the strata into asset classes, repeating the methods above, then prorate the results according to the current distribution within the asset strata group.

This yields four predictions which are then scaled and reviewed for reasonableness. A few extreme predictions will be removed, and each stratum's prediction is the average of the remaining estimates.

These procedures are similar to those used for the 1991 study, and, as Table 2 demonstrates, these procedures worked well enough.
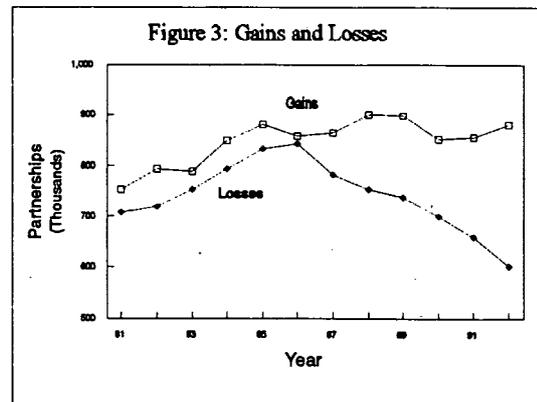
■ **Adjustments**

The action of cutting off the sampling process to create the preliminary data files also creates a nonresponse issue. We described the impact of this self-inflicted nonresponse on the initial naive estimates in the second part of this paper. The next step is to moderate these effects through appropriate strategies.

There are two main classes of strategies for dealing with nonresponse: imputation and weighting adjustments. In both cases adjustment classes are created, sometimes crossing strata boundaries, to create homogeneous groups "like" the nonrespondents. After our review above, it can be small surprise that we first consider separating the profitable firms from the others.

Most imputation methods begin with at least some information about the sample units that haven't "responded." In our case, we don't even know the number. Substitution can be considered a form of unit imputation, which in this case would involve using records from the end of the previous year, but poses problems in both analysis and execution.

On the analysis side, the use of records from the previous study would interfere with the economic models used for the tax proposal evaluations. Equally serious, the changes in the regulations means that each year's study uses records with somewhat different items and record layouts. So substitution requires programming resources which, as we've already noted, aren't there.

This leaves weighting adjustments. Unfortunately, we cannot have post-strata counts or marginal totals (for raking) any more than we can have the final population counts. If the proportion of firms with profits were relatively stable over the years, though, we could adjust the sample accordingly. But, as Figure 3 shows, since the 1986 Tax Reform Act, that proportion has shifted dramatically.


Figure 3: Gains and Losses

Just as the population predictions must rely on the current counts, then, so will the adjustments for the proportion of firms with gains (or the proportion of losses). We propose comparing the change in these proportions from the reference week in the prior year to that year's final estimate. We will use

the ratio of these proportions to adjust the current study's, as shown below:

$$\hat{R}_{py} = R_{pyw} \, (R_{p(y-1)} / R_{p(y-1)w})\ .$$

Yet, as Figures 4 and 5 show, we have very different coverage problems for different sized companies.

Figure 4: Largest Firms Cumulative Filing
(Percent of Final Count)
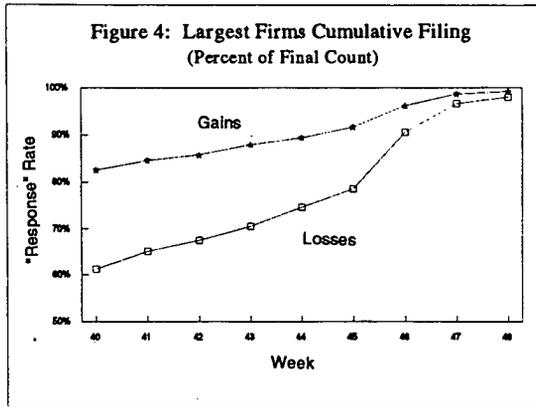
These charts show the cumulative weekly results estimated from the 1992 sample. We start with week 40 (about the first week of October) because data abstraction doesn't begin until the middle of September, and it takes about six weeks to process the 80 percent of the sample we feel is needed for the preliminary estimates. Realistically, a cut-off later than mid-November would entail producing the preliminary estimates after the new year.

These graphs illustrate the filing pattern differences between the largest and smallest firms. Overall, more than 90 percent of the population has been processed by the beginning of October (week 40); yet, as Figure 4 shows, only about 70 percent of the largest (certainty strata) firms have filed. Contrast this to the greater than 90 percent "response" rate for the small companies in Figure 5.

A comparison based on whether a company has earned a profit or not is even more revealing. While the smallest firms show little difference in these categories' patterns, a gulf exists for the largest firms.

Yet, as with most establishment studies, the proportion of the sample or population responding is

Figure 5: Smallest Firms Cumulative Filing
(Percent of Final Count)
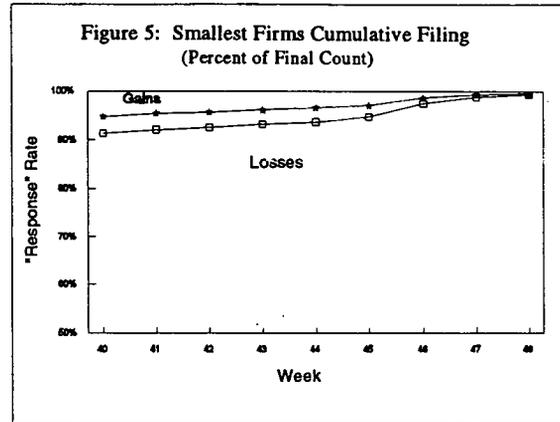
not, by itself, a reliable indicator of quality. By the beginning of November (week 44, a realistic cut-off point), over 95 percent of the filings have been tallied, but only 88 percent of the total profits and 78 percent of the losses were accounted for. This is clearly arising from the later filing behavior of the larger companies, which number only 20,000 -- only 65 percent of their contribution to the total amount of losses has been tallied by that week.

These results suggest that a simple division of the smaller strata into those firms with net income and those without would largely correct the problem for the 98 plus percent of the population they represent. The distance between the "response" rate and "coverage" rate for the largest 1.4 percent, however, signals the difficulty that would be encountered with this approach for the large concerns that dominate the estimates.

■ A Simulation

We simulated a cut-off date using week 42 (the middle week of October) and proceeded with estimating the variables listed in Table 1, using the weighting adjustments described above. We also estimated the final population, while ignoring the known result, as we would in practice.

However, we departed from what we would use in practice in that we used only eight adjustment classes. In practice for the 1993 Study, we would expect to use far more, but this will illustrate our intent. Further, we used the existing strata and cut

off at an earlier date, to simulate a "bad case" scenario.

The roughly 20 percent improvement in the estimated amount of net losses gained through the adjustment process would still leave at least an 8 percent error in an actual case. The increased error on Total Assets (from near zero to about 4 percent), the eight-fold increased error on the number of partners and on other items is a source of concern (Table 3).

Table 3.--Adjusting for Noncoverage (Relative Error of Preliminary to Final Estimates) (Percent)		
Item	Basic Estimates	Profit or Loss Adjustment
Partnerships	0.07	0.07
No. of Partners	1.1	8.6
Total Assets	0.1	3.7
Receipts	-2.1	-2.0
Net Income	2.4	-1.0
Net Loss	-27.3	-18.3
Portfolio Inc.	-0.3	9.2
Depreciation	5.4	5.4
Taxes Ded.	-6.5	-2.5
Sal. & Wages	-3.2	-1.7

Perhaps expanding the number of adjustment classes will reduce the error still further, but the effect on the other estimates urges caution. Moreover, our review of the stratum by stratum results suggests that this adjustment is unstable. Thus, we concluded that this procedure requires more study before implementation.

■ **Further Research**

As we've noted elsewhere, the sample design used for the 1993 Study has significant changes from recent prior years. These changes are the result of the tax law, regulatory and administrative systems changes so common in recent days. The main changes were increasing the use of industry codes in the stratification and making the Receipts/Income classes dependent on the Asset class. We plan to examine the impact of these changes, both on the final estimates and on the process of developing preliminary estimates.

Though not covered here, the preliminary data processes effect on industry estimates is under investigation. It will be the subject of another paper in this series. We will also be considering various improvements to the estimation process and the effect of delayed response on the final data.

And, of course, we expect that the tax law, regulations and administrative processes will continue to be revised, bringing more opportunities for change, and resulting in the continuation of papers in this series.

■ **References**

McMahon, Paul (1993), "Statistics of Income Partnership Studies: Evaluation of the Revised Sampling Plan," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

McMahon, Paul (1991), "Statistics of Income Partnership Studies: Sampling Plan Redesign II," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

McMahon, Paul; Collins, Richard; and O'Conor, Karen (1990), "Revising the Statistics of Income Partnership Sampling Plan," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Petska, Tom (1991), "Further Examination of Tax Shelters in the Post-Reform World," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Petska, Tom and Nelson, Susan (1990), "Partnerships and Tax Shelters: An Analysis of the Impact of the 1986 Tax Reform," *Proceedings of the Section on Survey Research Methods, American Statistical Association.* ■