STATISTICS OF INCOME PARTNERSHIP STUDIES: EVALUATION OF THE REVISED SAMPLING PLAN

Paul B. McMahon, Internal Revenue Service
Statistics of Income, R:S:P, P.O. Box 2608, Wash., D.C. 20013

KEY WORDS: Administrative Records, Missing Stratifying Data

This paper is the third in a series on efforts to improve the Internal Revenue Service's Statistics of Income Partnership studies. The first report, presented at the 1990 Joint Statistical Meetings, dealt with an assessment of the then current design and gave an outline for a revision (McMahon, Collins and O'Conor, 1990). The second paper described the problems of population projections, planning assumptions and sample allocation (McMahon, 1991).

In this effort, before describing the effects of the revisions, we will discuss the background of the series of studies and look at the previous design and the nature of the revisions. We will then describe some ideas for future efforts.

Background

What is a Partnership?

When a business is formed, there are a number of choices that must be made about its legal structure. If there are multiple owners, the benefits that the various forms of partnerships have are frequently attractive. Family businesses may prefer the relative informality possible with this form, while larger firms may see advantages for ad hoc cooperative ventures.

One key benefit of the partnership structure is that the profit of the firm is not subject to an income tax, although the activities of the firm must still be reported to the IRS. Each owner receives from the firm his own portion of the income, deductions, allowances and tax credits, which are combined with any income and so forth that the owner has from other sources and are reported, instead, on his own Individual tax return.

A partnership may apportion the income, deductions and tax advantages among its owners in many ways, and not necessarily proportional to the investments. One variation on the partnership theme is the

Limited Partnership, some of which have "Certificates of Partnership Interest" openly traded on the stock exchanges. In this context, "limited" refers to the limiting of some owners' responsibility for a partnership's debts to the amount invested (although at least one of the owners must be an "unlimited" or general partner).

This structure of "passed-through" tax advantages and constrained liability made the limited partnership a very attractive vehicle for tax shelters, especially in the real estate area. Given the high marginal tax rates of the late seventies, these shelters became very popular and by 1986 comprised a major proportion of the Partnership population (Petska and Nelson, 1990, and Petska, 1991).

The popularity of the tax shelters, not only in the Real Estate area, but also in such diverse industries as Beef Cattle Feed Lots and Equipment Leasing, caused the population to expand rapidly between 1976 and 1986, as can be seen in Figure A. The growth rate for that decade averaged over five percent, double the rate for the previous decade.

Figure A
Partnership Population
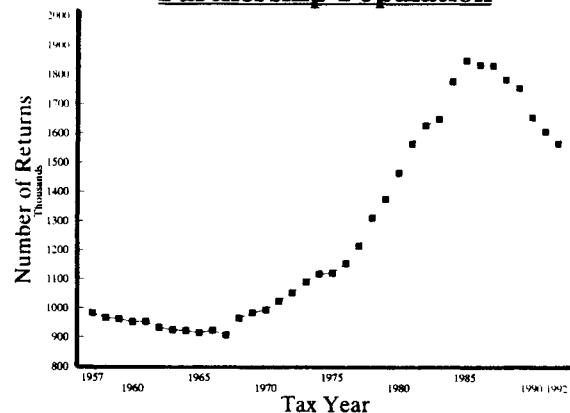
Tax Reform

The 1986 Tax Reform Act restricted the tax benefits and lowered the marginal tax rates, thereby reducing the incentives for forming firms with tax shelter goals. As a result, there was a dramatic

change in the pattern of Partnership growth, with the population decreasing by 13 percent over the next five years.

The decrease in the population was not uniform across all domains. The number of largest firms, for example, continued to increase by about --28 percent since 1986. When the sample design used during 1990 was developed (in 1979), this category dominated a large number of the estimates. Thus, the returns in that class were selected with a probability of 1. By 1989, large firms dominated the sample in another way -- over half the sample resources were allocated to that class alone. Given the limited budget, the continued growth in the class represented a reduction in the sample resources available for other strata.

Sample Redesign

During the 1980's, Partnership returns were sampled from the weekly transactions of the IRS' Business Master File system, using strata based on the amount of receipts, assets, and whether or not it was a real estate business. (See Figure A in the 1990 paper -- McMahon, Collins and O'Conor -- for a summary of the design.) The data from these transactions are largely included for tax enforcement reasons, but some are also quite useful for stratification.

Still, as with most administrative record systems, use of Master File items for stratification also had its limitations, in that we were dependent on tax reporting regulations. There were eight strata, for example, for records with assets "Zero or not reported." These strata arose in response to a change in the administrative environment. Certain companies, such as family farms, were exempted from the requirement to report asset data, including Total Assets. Since there were very different types of firms claiming the exemption, and other reasons that companies might not report their asset value, these strata contributed a significant amount of the overall variance for many estimates.

For the revised design, outlined in Appendix Table I, rather than rely on separate strata, we addressed this issue another way. Since much of the data reported by a company are interrelated, and we could readily identify records with the exemption, we decided to estimate the asset value of exempted firms from other reported information. Since vari-

ous industries hold their assets in different proportions, seven predictor formulae were needed. The goal was not to predict actual amounts, but to use this inferred asset value for stratification.

With the removal of the "Zero" asset classes, the strata boundaries for both assets and the receipts (or income) classifier were significantly realigned, with the smallest categories much more constrained. The boundaries on the smallest asset classes, for example, shrink from "Under 100,000" to "Under 35,000." For the companies with the largest economic size, the strata boundaries were raised, so that the amount of sample resources allocated to the small and medium strata could be increased. The total number of strata remained the same, at 45.

There are other constant features to the designs, such as the retention of the selection mechanism. This device uses the Employer Identification Number and large prime numbers to generate a pseudorandom number. A range of values for this randomized number that is proportional to the sampling rate is then used to decide what returns are in the sample. These ranges are nested across the strata. In practice, this means that a company would be included in succeeding years once it has been selected, if it at least retains its size (Harte, 1984). This retention of firms, which was effective even with the sample design revisions, serves to reduce the variance on the year-to-year change estimates.

Another constant is the use of separate strata for Real Estate Operators. This single minor industry accounts for nearly a third of the Partnership population and, if not set aside, would claim a like proportion of the sample's resources. Yet those sample units would contribute to only a relatively small amount of assets and liabilities. Thus, we target the sample size for these strata at half the proportional allocation, about a sixth of the total.

Impact of Design Revisions

The goals of the redesign effort, then, focused on restructuring the strata to counter the effects of inflation and population growth, while maintaining the quality of the estimates. This restructuring was also to reduce the size of the certainty classes, allowing resources to be redirected to strata for small and medium firms. The secondary goal was to improve the estimates at the Industry Division level.

By 1989 the resource demands by the certainty classes had reduced the smaller firms' representation in the sample to a minimum and pressed the total sample size beyond the budget. As an interim response, the 1989 and 1990 studies addressed this problem by creating a new stratum for returns reporting either Total Assets over $75 million or Receipts or size of Net Income over $10 million; then, most of the records remaining in the certainty strata were subjected to a 50 percent sampling rate. An unpublished assessment of this procedure showed that this approach maintained the reliability of key estimates. It did not, however, raise the allocation of the sample to the strata for small and medium companies nor address the quality of the Industry Division estimates.

Had the strata definitions not been changed, we estimate that the certainty class size for 1991, even with the reduction procedure outlined above, would have exceeded 13,000 returns, almost half the sample. Using this method, the actual certainty classes' sample size was held to about 8,500, meeting the target of reducing the size of the certainty classes.

The other half of the main objective was to retain the same level of reliability. As Figure B shows, there is only a small difference between the 1990 and 1991 coefficients of variation on the stratifying variables, as might be expected. Indeed, the 0.2 percent difference for Total Assets, for example, can be attributed to tax and accounting rule changes and normal year-to-year variation. Salaries and Wages, on the other hand, is not a stratifier, yet shows a

---

Figure B
All Industries: Coefficients of Variation
(in percent)

	SOI 1990	SOI 1991
Total Assets	0.5	0.7
Receipts	0.4	0.4
Net Income	1.4	1.8
Net Deficit	1.8	1.0
Salaries & Wages	2.4	0.8
Depreciation	1.2	0.9
Taxes Deduction	1.4	0.8

---

reasonable improvement. Similar effects are seen in other items, including Depreciation and Taxes.

While the use of IRS business codes was helpful

in identifying highly visible industries, like Real Estate Operators, less common codes are of uncertain quality. Thus, it was not possible to improve the accuracy of the less populous industries' estimates directly through allocating the sample. Instead, we examined the adequacy of the sample for those less populous industry divisions and adjusted the sampling rates in the allocation process, to ensure sufficient coverage within the strata. The result was that the maximum weight dropped from over 2,400 to below 1,300, and the variability of the weights was similarly reduced. The test of this effort, though, lies in the comparison of the Coefficients of Variation between the 1990 and 1991 studies. (See Figure C, below.)

Figure C
Coefficients of Variation by Industry Division
(in percent)

Industry Division	Total Assets		Salaries and Wages	
	1990	1991	1990	1991
All	0.5	0.7	2.4	0.8
Agriculture	8.1	5.2	23.1	9.2
Mining	9.9	2.2	11.1	7.4
Construction	9.1	5.4	11.2	7.4
Manufacturing	10.0	1.0	2.3	1.9
Transportation	10.9	1.5	4.3	2.6
Trade	5.7	1.7	3.2	2.3
Finance*	1.2	1.7	23.9	2.8
Real Estate	0.9	0.9	6.9	8.2
Services	1.4	1.1	1.3	1.1

(*Excludes Real Estate Operators)

The improvements in the accuracy of the estimates of Total Assets, for the Industry Divisions Mining, Construction, Manufacturing and Transportation demonstrate just what we hoped to accomplish. The estimates for Real Estate and Services were unaffected, while the effect in the Finance area was small. The improvement in the Agriculture estimate was not a target, but a welcome side effect.

The pattern continues when we examine the cv estimates for Salaries and Wages (Figure C) and other variables. The gains are not as dramatic as for Total Assets, but this may be due to the reporting variations the Partnership Return form permits. Of particular interest is the labor payments on Form

8825, for reporting real estate rental income and deductions (where the cv went from 6.9 percent to 8.2 percent) and labor on the Schedule F for Agriculture (23.1 to 9.2 percent).

Another factor that might be affecting the comparison was that the 1990 study employed a restratification procedure to correct for an initial data entry problem in the sampling frame, involving real estate rental income. This procedure would cause the variances for the 1990 study to be somewhat understated for variables highly correlated to receipts (total assets and its components were not affected).

The design revisions will, however, complicate some year-to-year comparisons, especially in small minor industries. Most of these cases involve estimates of less than 10,000 partnerships in the 1990 study -- as in the Vegetable and Melon Farm industry, which declined from 3,400 to 1,700 estimated firms. In these cases, a single sample return's weight could account for the largest proportion of the estimated frequency. The small number of sample returns these estimates are based on also implies a generally lower level of reliability for estimates from such industries.

Given the total sample resources available and the limitations of the sampling frame, it was not possible to design a selection procedure that would consistently support industry analysis below the division level.

In conclusion, while the modifications to the Partnership design have had the intended effects, there are some situations where changes are largely or entirely due to the design revisions and the consequent affect on the weights. Overall, however, we feel that the higher quality of the 1991 design and processes should lead to better comparisons in later years.

Refinements

The first year's operation of any substantial revision also reveals areas where there might be improvements. Considering the increase in the marginal tax rates at the upper end of the income scale for individuals, the Partnership population's trend, now in decline, may reverse itself. In such a case, to maintain the coverage of the less populous industries, we would again have to turn to raising the

minimum amounts for the certainty classes. At the same time, we would prefer to leave the majority of the strata boundaries unchanged, to minimize the impact on the estimates for less populous industries.

The highly reactive nature of the Partnership population to changes in the economic realm are not the only source of challenges for this design. Since we use the IRS' administrative records system as our sampling frame, changes to that system can have profound effects, as well. The addition of a key item could permit a modification of one of the stratifiers to improve some variables of interest to the main customers. Similarly, a decision by IRS not to pickup a particular line item from the partnership returns could present major problems. A change needn't affect all records to have serious consequences, as was shown in the Eighties with the asset reporting exemption.

In fact, such a change has already occurred. In an effort to reduce taxpayer reporting burden, the Internal Revenue Service expanded the asset reporting exemption to include all firms with both total assets and receipts under $250,000. This change affects one third of the strata and over two thirds of the population. Neither the current strata boundaries nor the asset predictor formulae were developed with this scenario in mind.

So, of course, this sets the stage for another revision to the design. Appendix Table II shows an incomplete outline of the changes due to be implemented during 1994 (for the tax year 1993 study). First, we addressed the possibility of a renewed growth in the certainty classes by realigning the income/receipts strata. In particular, the boundary for the records with the largest sizes (in absolute value) of income/receipts is raised from $10 million to $25 million. An intermediate set of nine strata were constructed to allow more selective reductions in the sample at the higher end of the design, with an expanded role for the industry codes.

The industry code available on the sampling frame has, as we mentioned above, limitations in its reliability, which is why it has been sparingly used in the past. However, recent initiatives to improve this code for various tax administration goals is expected to address many of the problems. Further, our use of these variables as stratifiers is quite broad for the selection process.

Although it is not clear from the figure, the

number of strata has been increased. The Tax Year 1991 Study had 45 classes, 20 of which were reserved for smaller Real Estate Operators. The plan for Tax Year 1993 maintains 20 classes for small and medium real estate partnerships, but adds 16 strata for the less populous industries. We expect that, with the initiative to improve the industry coding and using the new strata to ensure sufficient coverage in areas such as transportation, expanded use of post-stratification will further improve the results for these studies.

Research

Since a large proportion of the population might no longer need to report their asset holdings, we need to examine the effect this has on the estimates and our major users. If the main interest is in overall asset data, and if the dominance of the largest firms completely masks the effect of nonreporting, then no action may be required. This has not been established, since the interest in microdata files can mean that few data items are ignorable for long.

Another area of long-term interest has been the effect of Partnership activities on the direct taxpayers, especially on individuals. Several studies have been attempted in the past (e.g., Petska, 1993) with limited success, due to chains of ownership, where an individual may own an interest in more than one partnership and have that interest either directly or through other partnerships or fiduciaries or corporations (including all of the above). The weighting issues in such cases are quite complex, especially given the stratification and disproportionate sampling rates of both the Statistics of Income Individual Studies and the Partnership Studies.

References

Harte, J. M. (1984), Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS, *Proc. of the Sect. on Survey Res. Meth.,* Amer. Stat. Assn., 603-608

McMahon, P. (1991), Statistics of Income Partnership Studies: Sampling Plan Redesign II, *Proc. of the Sect. on Survey Res. Meth.,* Amer. Stat. Assn.

McMahon, P., Collins, R. and O'Conor, K. (1990), Revising the Statistics of Income Partnership Sampling Plan, *Proc. of the Sect. on Survey Res. Meth.,* Amer. Stat. Assn.

Petska, T. (1993), The Effects of Tax Reform on the Structure of U.S., *Proc. of the Sect. on Survey Res. Meth.,* Amer. Stat. Assn.

Petska, T. (1991), Further Examination of Tax Shelters in the Post-Reform World, *Proc. of the Sect. on Survey Res. Meth.,* Amer. Stat. Assn.

Petska, T. and Nelson, S. (1990), Partnerships and Tax Shelters: An Analysis of the Impact of the 1986 Tax Reform, *Proc. of the Sect. on Survey Res. Meth.,* Amer. Stat. Assn.

Appendix Table I:
SOI 1991 Partnerships
Strata Descriptions and Populations

	Population
Assets 100,000,000 or more	2,110
Assets less than 100,000.000 and the larger of receipts or income 10,000,000 or more	6,624
Assets 25,000,000 under 100,000,000 and the larger of receipts or income 3,500,000 under 10,000,000 ...	2,955
Assets 25,000,000 under 100,000,000 and the larger of receipts or income less than 3,500,000	3,252
Assets less 25,000,000 and the larger of receipts or income 3,500,000 under 10,000,000	13,016

Income measures less than 3,500,000:

Absolute Value of ASSETS ($)	Absolute Value of Receipts or Income*			
	Under 40,000	40,000 under 150,000	150,000 under 350,000	350,000 under 3,500,000
		PIA CODE 6511		
Under 100,000	147,367	22,013	3,408	2,448
100,000 under 350,000	60,093	60,428	9,156	1,939
350,000 under 1,000,000	15,008	50,939	28,666	9,809
1,000,000 under 5,000,000	6,127	11,770	28,232	44,539
5,000,000 under 25,000,000	888	719	1,218	19,054
		PIA CODE other than 6511		
Under 35,000	391,110	99,949	36,021	20,836
35,000 under 150,000	97,841	74,029	40,735	32,504
150,000 under 600,000	58,440	37,285	24,883	36,661
600,000 under 3,500,000	21,315	20,967	15,730	27,289
3,500,000 under 25,000,000	3,366	2,337	2,963	14,191

* Note: This is the larger of a Receipts or a Net Income absolute value, with both items a composite of several reported amounts in the transaction record.

Appendix Table II:
SOI 1993 Partnerships
Selected Strata Definitions with Populations

Assets $100 million or more	2,400
Assets less than $100 million and Receipts/Income $25 million or more	2,200

	Real Estate	Farms, Trade, Finance & Services	Mining and all others
Assets $25M under $100M			
Receipts/Income $5M under $25M	1,500	6,300	1,900
Receipts/Income less than $5M	1,600	2,100	200
Assets less than $25M			
Receipts/Income $5M under $25M	1,800	1,200	400

Absolute Value of Receipts/Income ($)

Assets ($)	Under 50,000	50,000 under 100,000	100,000 under 250,000	250,000 under 500,000	500,000 under 1,000,000	1,000,000 under 5,000,000
			Real Estate Operators			
Under 250,000	197,000	37,000	15,000		7,700	
250,000 under 750,000 ⋮	24,000	34,000	33,000		11,000	
5,000,000 under 25,000,000		1,600		1,400	4,500	14,000

Farms, Trades, Finance and Services

Assets ($)	Under 40,000	40,000 under 100,000	100,000 under 250,000	250,000 under 1,000,000	1,000,000 under 2,500,000	2,500,000 under 5,000,000
Under 250,000 ⋮	360,000	102,000	83,000	71,000	14,000	
2,500,000 under 5,000,000 ⋮		2,200	2,000	5,600	3,400	
10,000,000 under 25,000,000		1,200		3,600		1,000

Mining, Construction, Manufacturing and Transportation

Assets ($)	Under 40,000	40,000 under 100,000	100,000 under 250,000	250,000 under 500,000	500,000 under 1,000,000	1,000,000 under 5,000,000
Under 250,000 ⋮	56,000	26,000	25,000	9,800	2,800	
5,000,000 under 25,000,000		300		250		700