
Nonsampling Errors in Data Abstraction From Administrative Records

Paul B. McMahon, Internal Revenue Service

The various Statistics of Income studies use verification samples to assess the data abstraction process. For the most part, this consists of comparing an already completed record against an independently abstracted copy. Differences between the two are then resolved, with a record of the source of the difference maintained. The purpose of this system is to detect patterns of errors so that the editing clerks can have immediate feedback on problems they encounter and so that future training can be linked to real situations.

Our purpose differs from the purpose of the verification process because we seek to evaluate the effect of errors on estimates. We chose the Partnership Study for this review because it is neither the largest nor smallest sample, nor the most heavily stratified nor least, nor are the forms the most complex nor simplest. And, of course, we also chose it because the author is familiar with the sample design, data, and history.

We will begin with a brief description of the population, the sampling frame, and sampling procedures, then discuss some peculiarities of the returns filed. We will then turn our attention to the editing process and especially how the quality review sample was chosen. Finally, we will present some results based on the quality review data, with comments on what changes we expect to incorporate in future quality review sampling procedures.

■ Background

The Statistics of Income series of economic reports date back to 1916, always using the tax forms filed by companies and individuals as the source. Today, we continue this series, but the main purpose has changed from publishing standard tables. Rather, the focus now, and for most of the past two decades, is to provide microdata to Treasury's Office of Tax Analysis and Congress's Joint Committee on Taxation for their assessments of tax laws, both current and proposed.

Assessing the direct effects of the laws leads to a strong reliance on our annual Corporation and Individual Studies. There are strong indirect effects as well, where the advantages of certain investments are passed through to the owners via trusts or partnerships. The growth of tax shelters, frequently organized as limited partnerships, in the Seventies and early Eighties illustrates how these secondary applications can attract significant attention. They were, after all, a main target of the 1986 Tax Reform Act.

Partnership filings were dramatically affected by this change in the law. The population ceased its growth, which had averaged 5 percent per year, shrinking by almost a quarter between 1986 and 1993. Moreover, the long trend of net losses (in excess of profits) by this population has ended. Yet, though an important part of the business population, the Internal Revenue Service abstracts only a handful of data items into its computer records from the filed returns, since there is no direct taxation of partnership income. The assessments by Treasury and Congress depend on far more items being available on each record than are on the IRS's Business Master File (or that would be needed for the standard published tabulations, for that matter).

Still, the Master File is a useful sampling frame for our purposes, containing a number of items for stratification. The design we use for the Partnership Studies has 73 strata based on industry, total assets and receipts, or net income (or loss). Within the strata, a bernoulli sample is drawn, using a permanent random number (see Harte, 1984), at rates ranging from under 0.2 percent to certainty (100 percent). The use of permanent random numbers increases the number of firms retained in the sample from one year to the next, thereby reducing the variance for estimates of change.

Over the course of the sampling period, Calendar Year 1995 in this case, a population of 1,579,505 partnership returns was subjected to sampling. Of these, 31,458 were selected for the sample and additional data

abstraction. (For a discussion of the design, see McMahon, 1990, 1991, 1993, and 1995.)

Working in the administrative environment imposes certain constraints that might affect the final estimates. One is that although there is a deadline for filing, the law allows delays for a variety of reasons, including missing key documents, fire, floods, and litigation. These delayed filers are, as a group, quite different from the normal filers, reporting disproportionately large amounts of losses (see McMahon, 1994). We minimize the effect of deadline extensions by keeping the sampling period open from January through December, but some firms are delayed in filing for years. We offset the few very late returns by including the handful received during the sampling period for the three most recent prior years.

Moreover, sometimes taxpayers find that there was an error of some sort in their original returns. They may file amended reports almost immediately or many years later, but usually well after the closing date for the project. These are excluded from all Statistics of Income Studies, including Partnerships.

Another constraint for Statistics of Income projects is that since this is not an enforcement study, contact with taxpayers is discouraged and exceedingly rare. The data we include in our files are as reported in the original filing, not based on any review (or tax audit) of the firms or on later corrections by either the taxpayer or IRS actions. Indeed, because contact could cause behavior changes and make a sample firm's selection in following years likely, we guard against audits based only on selection for our studies.

These limitations imply the presence of nonsampling errors, mainly of omission, beyond the scope of this article. There are also significant advantages in using tax records for economic studies, arising from mandatory response (with real penalties) and standardized accounting terms. True, different forms of businesses, such as a partnership or corporation, are subject to quite distinct filing rules, including tax forms, but the attachments to the returns are often the same, as in the case of the Real Estate Rental Income schedule (Form 8825) or the Depreciation schedule (Form 4562).

The data from these standard schedules are often of great interest to our sponsors, and we must abstract this information while ensuring that strata identifiers are maintained. We do this by creating control records as each return is selected for the sample. These control records also retain many of the Business Master File Transaction Record items.

These control records form the basis for further data abstraction from paper returns. Our concern here lies with the quality of this further data abstraction for the Tax Year 1994 Partnership Study, although, since some data from the original administrative abstraction is used, those data are also considered.

This abstraction process differed from the earlier administrative operation in that the process involved computerized testing of relationships among edited items, ensuring consistency. This consistency testing relies heavily on the accounting structure of the various forms and schedules: details adding to totals, or that one figure is some set percentage of another (for relationships are part of the law). There are, however, some tests that use the expected weighted values to confirm that extreme values are real and not simply the result of a tad too much pressure on the keyboard. These large value tests are especially useful for those cases where the item cannot be compared to any combination of others.

Whenever the tests highlighted a problem, editing clerks would make necessary revisions, relying on the paper source documents still in their hands. Occasionally, they updated amounts used in the stratification process, but such changes did not affect the strata identifiers. A screening program searched for possible misstratified records and potential outliers at the completion of editing. Also, the subject matter analysts and the project's statistician (the author) reviewed those records.

Nine records were selected as possibly misstratified, based on changed stratification amounts and other factors. Only five (of the more than 30,000 records in the sample) were reassigned to self-representing classes. Two of these five were identified on nonstratification characteristics, with the other three in the same Finance industry (holding out the possibility that most of these

problem cases might be removed through some small design modification). This result from the misstratification review confirms that, at least for stratification purposes, the original "revenue processing" data are of reasonable quality.

Our main focus here, though, is on the quality of the SOI editing and abstraction process. In reviewing this aspect, we will rely on two independent studies: a "Supporting Schedules Review" and a standard quality assurance study.

■ Quality Assurance Sample

As in many quality review situations, the products under study are considered as "within tolerance" or not. It is assumed that the tolerance set is sufficient to ensure the end product's usefulness. When a product is not within tolerance, it is taken to be a problem, without regard to the amount by which it exceeds the margin.

In the normal course of events, the differences and problems are resolved (with the erroneous data corrected) between the original editor and the Q.A. editor on the final sample's record (but not the Q.A. sample's). The data on the causes of the difficulties are maintained and tabulated so that problem areas can be identified and addressed in future studies' editing classes. According to economists on the study, two areas are frequently on the problem list: the Real Estate Rental and Depreciation schedules.

The quality review process starts as a systematic sample of completed records, selecting every 14th record for each of the 15 to 20 editing clerks, or roughly 7 percent of the full sample. The selected Quality Assurance records were then submitted to a second editor for an "independent" reprocessing (that is, the original edit and the Q.A. sample's edit were done by different people). From the start, we were aware that this sampling procedure was supplemented in the early processing, but it seemed to affect only a small number of records, and usually in the most constrained strata.

We planned to use the implicit stratification to estimate the magnitude of differences between the two edits, as below.

$$\hat{X}_o - \hat{X}_q = \sum(N_i / n_i)(n_i / n_{iq})(x_{ijo} - x_{ijq})$$

$$= \sum(N_i / n_{iq})(x_{ijo} - x_{ijq})$$

Here, the x_{ijo} refers to the original, first clerk's edit of the j th return in stratum i , and x_{ijq} to the second quality assurance edit. The sample size n_{iq} is the count of quality review records in the i th stratum, N_i the population count, and n_i the full sample count for that stratum.

Since we need to compare the results in a number of different fields, we computed the relative error from the edits, as below.

$$\text{Relative Error} = \frac{(\text{Clerk 1}) - (\text{Clerk 2})}{(\text{Clerk 1})}$$

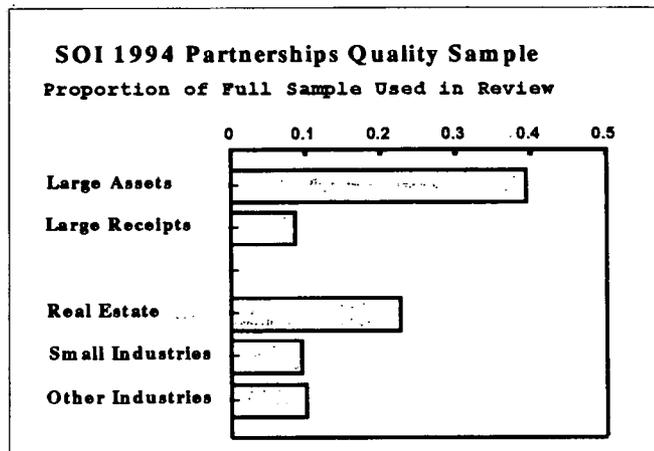
Using the first clerk's data as the denominator suggests that we consider this the more reliable information, as in a testing framework, but we choose it instead because that is the same basis as data presented to the public. We do not mean to imply that either abstraction is giving the "true" values.

This formulation requires that the Q.A. sample records be matched to and merged with the final, full sample file records. The exact match was straightforward, since each record was assigned a unique identifier at the time of selection from the filing population. Using the final sample file, though, introduces the possibility that post-editing adjustments, made by the project's economists, could exist.

In this situation, we are trying to use a specific purpose sample (the Q.A. selections) for a quite different goal than originally intended. It is to be expected, then, that problems would arise. Our first was almost immediate: instead of the 2,100 expected Q.A. sample records, we received a file with over 4,300. Where did these additional records come from?

When a lull in processing records for different studies or a delay in shipping sample returns across the country occurred, rather than temporarily reassign clerks to regular IRS operations, more returns were included in the Q.A. review. These additional records were selected by deterministic means rather than changing the skip

interval or using a second randomizing technique. As the table below demonstrates, they were not uniformly distributed among the sampling classes either.



Unfortunately, there was no indication in the records as to which is an ordinary, random quality sample return and which is an "additional" Q.A. Sample selection. Some of the additional quality selections were based on the desire to ensure that the very largest companies' data were accurate. Given the dominance these giants have over the estimates, this is a sound precaution. The definition of what constituted a "large" firm was based on a changeable amount of Total Assets. One week, it might be companies with more that \$500 million; the next, \$350 million.

It is clear that the economist's worries about Real Estate data were also reflected in the choice of those records added to the systematic selection. What effect do these records have on the analysis? We compared a couple of weighted estimates, using the same fields and assuming that the additional records were selected at random.

	Quality Sample (Estimates in billions of dollars)	Full Sample
Total Assets	\$4,560	\$2,300
Receipts	960	660

Applying a weight of about 2.5 to the records with the highest amounts of Total Assets clearly distorts the estimates. Virtually all overestimation of assets from the Q.A. sample is due to the largest assets stratum, with

no discernable effect from the oversupply of Real Estate firms. At the same time, not all estimates are affected equally, as shown by the receipts figures. This leads us to present estimates of the error arising from the data abstraction with and without the top assets class.

Our initial set of error estimates contained a few numbers that did not fit a reasonable profile. After all, we are looking at a fairly constrained circumstance, with all of those consistency tests as well, so errors above about a half a percent deserve special inquiry.

Item	With Top Assets	Without Top Assets
Cost of Goods Sold	185	0.17
Cash	50	50
Other Assets	77	77

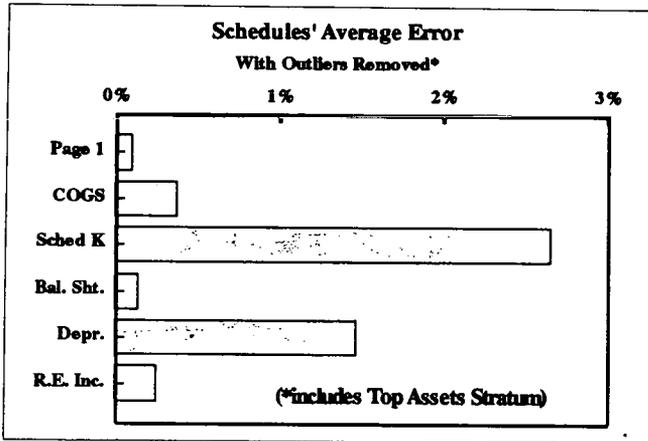
Clearly something other than a simple clerical error was at play here, and it seemed likely to be a general process applied to a Finance Division industry, so we approached the economists once again. Were there any special rules or automated edits?

There are, but that was not what caused the differences we see in the above table. There are a few very large partnerships that are little more than consolidations of other operations (also in our population and sample), which leads to a sizable overestimate of the size of the population's holdings. This situation had been known for quite a while, but the nature of the issue for these outliers was not made public. These three outliers, already in self-representing strata, are treated to a special "reduction" (dividing the amounts reported by a large number) by the economists as a final step prior to delivery of the products to our main sponsors.

We removed these records from further consideration in the study at hand.

As we described above, our goal is to identify areas and schedules and areas where the errors are largest. We selected the first page of the tax form; Cost of Goods Schedule; Partners' Shares of Income, Credits, Deduc-

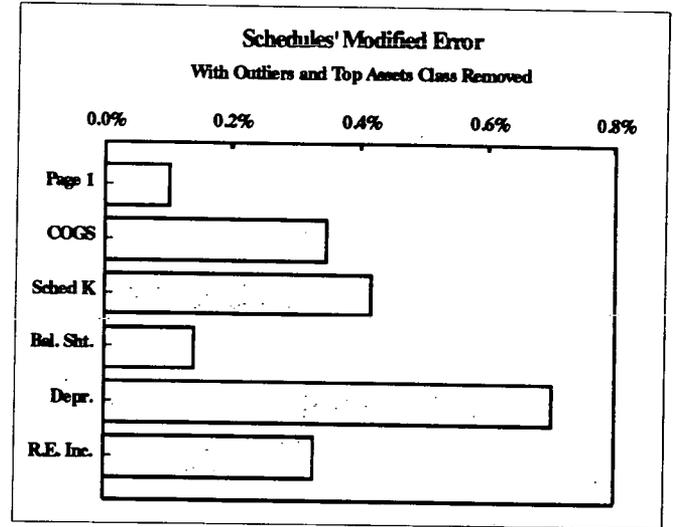
tions, etc. page (Schedule K); Balance Sheet; and the two sections that concerned the subject matter people, the Real Estate Rental and Depreciation schedules. From these sets of data items, we then selected several from each group as representative, calculating the average error as before.



The Page 1 and Real Estate data are not affected by the removal of the outliers, but the Cost of Goods Sold schedule and the Balance Sheet both show considerable improvement. The Balance Sheet average error, for example, is a reasonable 0.1 percent, where before it had been a hundred times larger.

From this vantage, it appears that the problems are on the Partners' Share and Depreciation pages. Depreciation is an expense related to certain types of asset holdings, so we might expect that this area's size be a reflection of the balance sheet's, but the asset holdings' average error is just a tenth of the relative error we see in the depreciation area. This arises because many types of assets, especially financial holdings such as stocks and bonds, are not subject to depreciation. Thus, the extreme size of the depreciation errors could be a reflection of the weighting problems in the top stratum.

The Partners' Share, Schedule K, is a summary of the information given to each partner, who in turn uses it on his or her tax return. A few items merely reiterate data reported on, say, page 1 or the Real Estate schedule, but most are only available here, such as charitable donations and income from tax-exempt municipal bonds. Unfortunately, consistency-testing this part of the return is quite limited, since there are few totals to bal-



ance against. Ratio and range tests are also of little use here because several of the items are known to fluctuate wildly from year to year (such as sales of capital assets). For these reasons, we are not surprised to find larger errors on this page of the return.

With both the outliers and the top assets class removed, we see a different picture (Schedules' Modified Error, above). The upper limit on this graph, 0.8 percent, is only a fourth the size of the previous chart's.

The schedules have a less than 0.4-percent average error arising from the data abstraction and editing processes. The exception is in Depreciation, as the economists predicted. However, the Real Estate form that worried them does not have a materially different error effect than other schedules, which suggests that the resources spent there could be reassigned to other areas. We note that, contrary to our expectations, the size of the Schedule K (Partners' Shares) error is largely attributable to the top assets category.

We have focused on the schedules, rather than specific items, but improvements in the abstraction process will come from addressing specific problems. In reviewing the depreciation form, for example, we saw that the situations with the larger error effects have something in common: they are isolated, in the midst of verbiage and on a back page.

These error effects apply to the estimates of the cur-

rent year but may be taken to be roughly constant in size over the years. There are other aspects to using administrative data that can cause biases in longitudinal estimates for the unwary. Such a case lies in the "Other" schedules.

■ Supporting Schedules

A business Income Statement lists the various revenues and costs associated with running the firm. For a single operation, this list can be exhaustive, but on a generalized form, only the most frequent and significant are usually included. So it is with the Partnership Return. On the form's first page is a generalized income statement (it does not include passive income), which has two lines for reporting miscellaneous, infrequent and small items: "Other Income" and "Other Deductions."

Such catchall categories give rise to a concern that information is being reported here that really ought to have been reported elsewhere. Perhaps the company has all the information on a spreadsheet, and rather than copying all that information to the tax form only fills in the "Other" lines. The amounts reported on those lines must be substantiated in some detail on supporting schedules, though, so any company reporting in that way would append its own version of the income statement. Thus, a review of those supporting schedules could reveal any significant misstatements.

We were asked to explore the reporting of these items by the Department of the Treasury's Office of Tax Analysis, a major client, but here as part of a feasibility study, we will only consider the income amount and its major components. This study applied an additional edit to the returns with the largest reported "Other Income" amounts, separate from the usual processing.

Top "Other Income" Returns' Characteristics

[Money amounts are in billions of dollars]

	Largest Firms	National Estimates
Number of returns	314	1,495,000
Other Income	\$20.1	\$31.3

This small group of records, less than 0.03 percent

of the population, clearly dominates any analysis, accounting for about 64 percent of the estimated population total. Because the amount of Other Income is only about 3.7 percent of the \$844 billion "income from all sources," scant attention is usually given to the details.

These results are reasonably comparable to a similar review after the 1986 Tax Reform Act went into effect. We consider the breakdown of the "Other Income" schedule for only a few significant items covering about a third of the full amount. About \$12 billion of the Other Income amount was only classifiable as miscellaneous. As the chart below illustrates, the amount of Interest dominates these data, accounting for half.

One might expect that perhaps these data should have been reported elsewhere, such as on the Partners' Shares section, Schedule K, under portfolio interest income. But, as usual in legal documents, there are caveats. In this case, the instructions call for interest received "... in the ordinary course of a trade or business ..." to be reported as Other Income. This exception applies to finance operations and certain types of installment sales. From the viewpoint of tax regulations, then, we cannot conclude that there is anything awry with the reporting.

The other side of the coin is that from an economic analysis view, an estimate of Interest would be seriously understated. However, this situation does not extend to all types of income reported as "Other." Some items are not separately addressed elsewhere, and some do not have a significant reportage in Other Income.

Relationship of Other Income Details

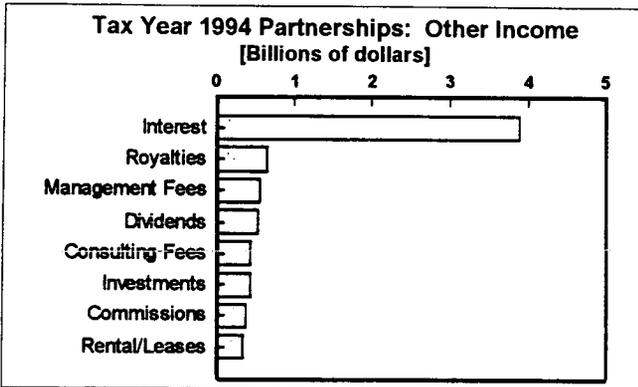
[Money amounts are in billions of dollars]

Item	"Other Schedule"	Adjusted Total*	Percent Understated
Interest	3.90	25.5	15.3
Royalties	0.66	3.1	21.3
Dividends	0.55	4.9	11.2
Rentals	0.35	172.5	0.2

*Includes portfolio and "other" amounts.

Of these, rental income from the "Other" schedule is an ignorable part of the overall rents collected. The Royalties and Dividends items included in the Other Income Schedule, on the other hand, are a major part of

the total partnership income from these sources but are small relative to the \$844 billion all-source income.



The bottom line for the “Other” schedules is that although there seems to be significant amounts present on these reports, they are properly classified under the tax regulations. Of course, the fact that only the companies with the largest amounts of Other Income were included suggests that these firms’ returns were produced by an accountant or lawyer. In turn, we can only conclude that most of the population value is properly reported. Nevertheless, this reporting characteristic does generate a bias in the measure of current income from these sources.

As we noted above, these results are consistent with an earlier study. Thus, for recent years, the time series involving these amounts would provide reasonable regression estimates. However, if that time line included data from before the 1986 Tax Reform Act, the change in the reporting requirements could inflict a serious bias.

Although not discussed in detail above, we also inspected the “Other Deductions” schedule. There, the amounts were what one might expect to see in a miscellaneous category, with the only exception, labor, being too small (less than 1 percent) to materially affect any longitudinal estimates of wages.

■ Conclusions

With the ad hoc “Other Schedules” data, we see that the published data can paint an incomplete picture of the economic activities of a sector of the business population. We would like to report that data will be extracted from these filings to enlarge upon the published

information, but unless a major sponsor asks for the details on such items, there will be no change in the current processing.

Was there misallocation of sources of revenue and expenses to the “Other” lines of the form? We concluded that there was not, so the sponsors can now focus on other problems. Moreover, given current budget constraints, even small changes require strong justification.

With regard to the Quality Sample, as we noted, too few resources were expended on the review of returns with large incomes or receipts. This does lie in our domain, and we have instituted modifications for the 1995 study. Moreover, additional codes are now in place to allow us to review those selected under the various criteria separately. Later analysis should show whether the increased attention to the large Income classes was useful. The budget reductions, however, may result in a quality sample too small for analysis of less common situations.

■ Further Research

In the above discussion, we limited ourselves to a few reporting characteristics and data abstraction quality. We have not covered unreported financial information or adjustment strategies. These data may be “missing” from some records because the tax regulations permit some companies to forego filling in a schedule such as the balance sheet. If no adjustment were made, then, the economic estimates would be understated. Yet, for tax assessment and analysis purposes, the absence of these data is of little consequence.

This means that within the segment of the population that qualifies for the exception on reporting assets, for example, we have reports from a self-selecting group, possibly intermixed with required reporters. What is the extent of nonreportage, and what is its impact on national estimates?

We are also exploring refinements to the weighting process, including considerations of sample selection changes that might promote better industry division-level figures. Another area under consideration is longitudi-

nal estimation. While the data that are published focus on current reports, the natural comparisons to previous years are omnipresent. As yet, however, we have not provided any estimates of the variance one might expect for such data.

■ Acknowledgments

The author appreciates the assistance of Sandra Neale and Nancy Ahn in acquiring the Quality Sample File and Timothy Wheeler in deciphering the "Quality" results and the Other Schedule data.

■ References

- 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.
- McMahon, P. (1991), "Statistics of Income Partnership Studies: Sampling Plan Redesign II," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- McMahon, P. (1993), "Statistics of Income Partnership Studies: Evaluation of the Revised Sampling Plan," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- McMahon, P. (1994), "Statistics of Income Partnership Studies: Evaluation of Preliminary Estimates," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- McMahon, P. (1995), "Statistics of Income Partnership Studies: Evaluation of the Expanded Sampling Plan," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- McMahon, P.; Collins, R.; and O'Connor, K. (1990), "Revising the Statistics of Income Partnership Sampling Plan," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Petska, T. (1991), "Further Examination of Tax Shelters in the Post-Reform World," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Petska, T. and Nelson, S. (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.
- Wheeler, T. (1994), "Partnership Returns, 1992," *Statistics of Income Bulletin*, Fall 1994, Volume 14, Number 2, Internal Revenue Service.