

Sampling and the IRS

By Howard M. Weinman

Howard M. Weinman examines areas where sampling is used, ways in which taxpayers may use sampling to their advantage and the pitfalls to avoid.

Introduction

Sampling has been used by taxpayers in connection with preparation of federal tax returns for decades. Both taxpayers and the IRS use sampling on audit. This article explores areas in which sampling is commonly used, ways in which taxpayers can try to use sampling to their advantage and pitfalls to avoid when either the taxpayer or the IRS uses a sample. Because of a recent change in the IRS position regarding samples by taxpayers, in many situations the use of statistical sampling has become more advantageous.

This article focuses on the procedural aspects of sampling under IRS guidance and on the legal consequences of different types of sampling. It does not delve into the mathematics of sampling—taxpayers routinely engage professional statisticians for that¹—but does introduce some very basic statistical concepts, particularly point estimates and confidence intervals, only to the extent that they are necessary to understand the procedural and legal issues.

Statistical and Judgment Samples

It is first important to understand that there are two kinds of samples: *statistical* and *judgment*. Each is subject to different legal treatment in connection with tax matters.

Statistical samples are based on mathematical principles.² They involve (usually) a large volume of data to be sampled, called the population. The sample is selected *at random*. In this context, “random” does not mean pointing at stack of boxes in a warehouse and using it as the sample. Detailed rules create a

regime in which every member of the population has a known chance of being selected. Using these procedures, the warehouse sample could be expected to come from boxes in many different stacks. If this key first step is performed correctly, then, based on principles of probability and statistics, it is possible to estimate characteristics of the population being sampled. These may be reflected in a point estimate and a confidence interval, as explained below.

A simple example could involve a determination of the basis of the stock of a corporation that is acquired in a stock-for-stock exchange. In a tax-free “B” reorganization of this type, the basis of the acquiring corporation for the stock of the acquired corporation carries over from the basis in the hands of the former shareholders. If the acquired corporation was a subsidiary of another corporation, it should be possible (if sometimes complex) to determine what the former parent corporation’s basis was for the stock. However, if there is a tender offer for stock of a publicly held corporation with thousands of shareholders, it would be next to impossible to obtain the necessary information from each shareholder. In this situation, the acquiring corporation might obtain data from a statistical sample of former shareholders. The sample may show that the average price paid by the sampled former shareholders for their stock was \$50 per share. This would be the point estimate. From this, it could be inferred that all shareholders paid, on average, \$50 per share, so the acquiring corporation’s basis would be \$50 multiplied by the number of acquired shares.

Importantly, it is possible to specify a *range* around the point estimate, called a confidence interval, that reflects the likely accuracy of the sample. Looking at the corporate acquisition again without delving into the mathematics, it is intuitive that if the acquiring corporation took a second sample of former share-

Howard M. Weinman is a Partner in the Washington D.C. office of Crowell & Moring LLP.

holders, the average stock basis in that sample might be \$52, and in a third sample, it might be \$49. Each sample can only be expected to approximate the exact value, which could be obtained only if data were provided by every former shareholder. The confidence interval reflects the variability inherent in the sampling. For example, the 95-percent confidence interval of the first sample might be from \$45 to \$55. For the second sample, it might be from \$47 to \$57. The 95-percent confidence interval of a sample means that, if the taxpayer sampled the population 100 times, it could expect the exact answer would be within the 95-percent confidence intervals of 95 of the samples. Thus, the 95-percent confidence interval reflects a high likelihood that the “right” answer is somewhere within the interval.

In real life, depending on whether the acquired company was a staid public utility with a consistent stock price, or a “high beta” technology company with a volatile stock price, the variation among the different sample results might be mere pennies or several dollars. The underlying variability of the sampled population affects the size of the confidence intervals (likely narrower for the utility and broader for the dot-com).

The sample design also affects the confidence intervals. It is intuitive that, if the sample is made larger, the breadth of the confidence intervals should shrink, because there is less of a chance that the sample will miss the true characteristics of the sampled population. Professional statisticians are adept at designing samples with usefully small confidence intervals but with minimal sampling expense.³

Sample design can also *stratify* the sample. Suppose in the stock purchase example, there were five large shareholders each holding 10–15 percent of the stock and that the rest was held by hundreds of public shareholders. The statistician could create one stratum of large shareholders, all of whom would be included in the sample, and another stratum of small shareholders, only some of whom would be sampled. Intuitively, if the sample includes exact data on the large shareholders (which have 50 percent to 75 percent of the total shareholdings), the degree to which the remaining sample of smaller shareholders can vary is significantly reduced.

Similarly, the confidence intervals can be specified in different sizes. Instead of a 95-percent confidence interval, the statistician could determine a 90-percent confidence interval. The 90-percent confidence interval would be smaller because it is less likely that

the exact answer would fall within the confidence interval. As will be seen in the following discussion, the IRS has focused on 95-percent confidence intervals as the standard.

In contrast to statistical samples, *judgment* samples, as the name implies, depend on judgment as to the reliability of the result. Judgment samples are not selected purely at random and thus do not have the statistical foundation necessary for the computation of confidence intervals. For example, if the acquiring corporation obtained basis data only from the ten largest former shareholders of its new subsidiary, that would be a judgment sample. A point estimate could be derived from these shareholders’ average share basis, but no confidence intervals could be constructed around this point estimate. The IRS and the courts treat judgment samples very differently than statistical samples.

The IRS Position on Sampling

General Guidelines

The IRS has issued many items of guidance on sampling, some published and some internal, covering a variety of specific areas of tax law. These will be discussed in detail below. However, a look first at the basic principles the IRS follows today may be useful. The most recent iteration of that guidance is the *Field Directive on the Use of Estimates from Probability Samples*, issued November 2009.⁴ Except as noted, the discussion below is based on the field directive.

As a general rule, the IRS does not accept taxpayer sampling in every situation. The threshold issue is whether the taxpayer’s books and records would provide a more accurate answer or sampling would not conform to GAAP. Even if records are voluminous, if they are computerized and can be analyzed by machine, sampling may not be appropriate.⁵

Assuming that sampling is permitted, the IRS may accept taxpayer statistical samples that meet certain criteria. However, judgment samples are rejected unless conducted on the basis of agreement to such a sample by the IRS examination team.⁶

As discussed above, a statistical sample provides not only a point estimate of the characteristic being measured but also confidence intervals. The existence of confidence intervals is also critical to IRS acceptance of a taxpayer sampling study (unless the IRS has agreed to a judgment sample). The reason is that the IRS ordinarily allows the taxpayer the amount shown not by the point estimate, but

by the least favorable (to the taxpayer) 95-percent confidence interval. Suppose, for example, that the taxpayer uses statistical sampling to establish the portion of its meals and entertainment expense that is subject to the 50-percent disallowance under Code Sec. 274(n). The sample shows that 35 percent of the expense is subject to disallowance, with 95-percent confidence intervals ranging from 28 percent to 42 percent. In this case, the IRS would subject 42 percent (the least favorable 95 percent confidence limit, rather than the point estimate of 35 percent) to the Code Sec. 274(n) disallowance rule.

Why is this? Possibly, it is because the IRS will not give the taxpayer the benefit of the doubt where, because of a very small sample, poor sampling design, and/or high variability of the sampled population, there is a wide range of possible answers. The use of the least favorable confidence limit gives the taxpayer an incentive to provide the IRS with a well designed study that confines the range of likely answers represented by the confidence intervals. In fact, under its more recent policy, if the confidence interval is sufficiently small, the IRS will reward the taxpayer by accepting the point estimate. If the precision of the estimate (*i.e.*, the size of one side of the confidence interval), divided by the estimate itself, does not exceed 10 percent, the point estimate will be accepted. Returning to our example, suppose the confidence intervals were from 32 percent to 38 percent. In that case, the requirement would be met,⁷ and the IRS would accept the point estimate of 35 percent (rather than 38 percent, the least favorable confidence limit) as the amount subject to Code Sec. 274(n) disallowance. If the precision is more than 10 percent, but is less than 15 percent, of the point estimate, the IRS will accept an estimated value that moves away from the least-favorable confidence limit and toward the point estimate as the relative precision drops from 15 percent to 10 percent. The more precise statistical analysis may be more expensive and time-consuming, but it could be well worth the effort.

If a statistical sample design falls short of IRS guidelines, the taxpayer is allowed an opportunity to correct the sample to bring it into compliance if possible. However, if this is not possible, the IRS will

Proper sample design, whether by the IRS or the taxpayer, is critical to the result.

reject the sample as applied to the entire population that was sampled and will only accept the results as they apply to the items actually sampled. (That is, if five percent of the population was sampled in a fatally flawed study, the IRS will accept the results only as to the five percent sampled and will not project the study result to the other 95 percent.)

On the IRS side, as part of the *Joint Audit Planning Process*⁸ in connection with LMSB audits,

“[t]he audit team should consider the use of statistical sampling or alternative testing methods (*e.g.*, mutually agreed upon judgment sampling) if records are voluminous.”⁹ Interestingly, the IRS, at least as far back as 1992,¹⁰ took the position that the IRS should also be held to the least favorable (to the IRS) confidence limit when the IRS conducted a statistical sample on audit. The IRS was apparently concerned that the sample result could otherwise be attacked by the taxpayer as arbitrary. This IRS position is now embodied in INTERNAL REVENUE MANUAL §4.47.3.3 (August 31, 2002). However, the manual allows the examination team to use the point estimate if “the sampling error at the 95 percent confidence level is immaterial in relation to the point estimate.”¹¹ Possibly, the IRS will apply the same 10 percent precision standard to itself as it applies to taxpayers.

Finally, as noted in the *Joint Audit Planning Process*, the IRS will consider judgment samples, if agreed to by the taxpayer.¹²

Published Guidance¹³

Transfer Pricing

Reg. §1.482-1 permits use of statistical analyses, among other factors, in establishing arm’s-length pricing between related parties.

Interest Expense Allocation

Under Reg. §1.861-9T(h)(1)(ii), taxpayers who allocate interest expense according to fair market value of assets may use statistical methods of valuation for fungible tangible personal property.

Qualified Appraisals

Certain charitable contributions of property must be supported by a qualified appraisal. Reg. §1.170A-13(c)(3)(ii)(K) indicates that statistical sampling may

be used but that “a justification for using sampling and an explanation of the sampling procedure employed” must be provided. A similar rule under Reg. §1.468B-3(b)(4)(v) applies to appraisals that support deductions or losses claimed for transfers of certain types of property to qualified settlement funds.

Revolving Credit Plans

Rev. Proc. 64-4¹⁴ allowed statistical sampling to determine the percentage of charges under revolving credit plans that could be treated as installment sales. In addition to a detailed prescribed safe-harbor sampling procedure, taxpayers could use other methods conforming to the *Report of Committee on Standards of Probability Sampling for Legal Evidence—Admissibility of Data from Probability Samples*, attached to the revenue procedure.

Trading Stamps

The trading stamp method¹⁵ permits a taxpayer issuing such stamps to reduce its gross receipts by the estimated cost of redeeming the stamps. Rev. Proc. 72-36¹⁶ permits the taxpayer to use statistical samples to determine expected future redemptions. A detailed safe-harbor method is prescribed, but the taxpayer may also use other statistical methods acceptable to the IRS that meet the requirements of the *Report of Committee on Standards of Probability Sampling for Legal Evidence—Admissibility of Data from Probability Samples*.¹⁷

Stock Basis in “B” Reorganizations

Rev. Proc. 81-70¹⁸ permits a corporation that has acquired stock of a subsidiary from the public in a “B” reorganization to estimate its basis for the stock by statistical sampling, on the ground that the taxpayer does not have adequate records to determine the basis independently. Apparently for the first time in published guidance, the IRS specified that the taxpayer must construct a sample in which the 95-percent confidence limits do not exceed 10 percent of the point estimate, a standard that the IRS used in subsequent guidance on other issues.

Employee Scholarships

Under a safe-harbor procedure, an employer-related private foundation may make grants to children of employees if, among other things, grants are awarded to no more than 10 percent of eligible employee children. In determining how many children are eligible, Rev. Proc. 85-51¹⁹ specifically provides that “statistical or sampling techniques” may *not* be used.

Qualified Intermediaries

The IRS maintains a “qualified intermediary” program under which certain foreign financial institutions may act as withholding agents for taxes on U.S.-source income received by foreign persons without revealing the identities of the foreign recipients of the income. As part of the program, an independent external auditor must determine if the correct amounts were withheld. Under Rev. Proc. 2002-55,²⁰ the external auditor may use statistical sampling to determine whether there was underwithholding. The point estimate of the underwithholding may be projected to all accounts in the sampled population to determine the qualified intermediary’s liability.

Medical Expense Reimbursement Plans

In Rev. Rul. 2003-43,²¹ the IRS considered health flexible spending arrangements and health reimbursement arrangements under which employees used credit or debit cards provided by the employer subject to certain safeguards to assure that the cards were only used for eligible medical expenses. The IRS concluded that where, among other requirements, every use of the card was substantiated as an allowable medical expense, the amounts involved were excludable from employee income. Where, however, the employer did not substantiate every charge, but requested substantiation only on a sampling basis, the amounts involved could not be excluded.

Fifty-Percent Disallowance of Meal and Entertainment Expenses

Under Rev. Proc. 2004-29,²² taxpayers may use statistical sampling to determine the extent to which meal and entertainment expenses are subject to the 50-percent disallowance of Code Sec. 274(n). The disallowed amount is based on the least favorable 95-percent confidence interval, unless the precision of the change in the estimated deductible amount, divided by the change in the estimated deductible amount, does not exceed 10 percent. In that case, the point estimate may be used. Judgment samples are prohibited.

Advance Payments

Rev. Proc. 2004-34²³ permits a taxpayer to defer a portion of advance payments for certain goods and services to the next succeeding tax year. Generally, the deferrable amount is determined on the basis of the taxpayer’s “applicable financial statements,” but if the taxpayer cannot make that determination, a

statistical analysis is acceptable. No specifics as to the statistical sampling are provided, but since the taxpayer must obtain IRS approval for a change in method of accounting, the parameters would presumably be worked out in that process.

Deduction for Domestic Production Activities

Code Sec. 199 allows a deduction, within limits, equal to a percentage of qualified production activity income. Rev. Proc. 2007-35²⁴ permits statistical sampling to be used to determine the allowable deduction “if the taxpayer can demonstrate a compelling reason for its use.” The revenue procedure gives several examples of appropriate and inappropriate use of sampling. For a taxpayer with over 5,000 products, and which would need one staff day to determine the portion of gross receipts from each product that qualifies as domestic production gross receipts, sampling would be appropriate. For a taxpayer with a computerized sales journal enabling it to determine the place of manufacture for each sale, sampling is inappropriate because exact information can be extracted from the taxpayer’s computerized records.

Rev. Proc. 2007-35 includes several other important rules. Judgment sampling is prohibited. Each unit in the sampled population must have a known, nonzero chance of selection. No unit selected for the sample can be replaced; if the unit lacks adequate documentation, it must be treated as nonqualifying. The allowable amount will be computed to the least-favorable confidence interval, unless the precision of the estimated difference divided by the estimated difference is 10 percent or less, in which case the point estimate may be used.

Employer Provided Cell Phones

In Notice 2009-46,²⁵ the IRS suggested allowing employers to use a method similar to Rev. Proc. 2004-29 to determine the taxable fringe benefit resulting from employee cell phone use.

Unpublished Guidance

LIFO Inventories

IN APPEALS SETTLEMENT GUIDELINES, DOLLAR-VALUE LIFO SEGMENTS OF INVENTORY (September 1, 1998),²⁶ the IRS recognized that statistical sampling could be used, but criticized taxpayers for trying to use nonstatistical methods, particularly those that excluded portions

of the inventory from sampling entirely. The guidelines state that “[t]he deliberate exclusion of whole segments of the inventory from a sample does not usually satisfy the requirements for a representative or statistical sample.”²⁷

Timber Casualty Losses

THE TIMBER CASUALTY LOSS AUDIT TECHNIQUES GUIDE (September 2005), Chapter 5,²⁸ permits statistical sampling to be used to estimate the volume of timber lost by casualty. No further details are provided.

Cost Segregation Studies

THE COST SEGREGATION AUDIT TECHNIQUES GUIDE (March 2008), particularly Chapter 5,²⁹ discusses statistical sampling of properties of the same type (e.g., a retail chain having “cookie cutter” stores in numerous malls). The guide generally approves statistical sampling, though holding any sampling error (*i.e.*, the least favorable 95-percent confidence interval) against the taxpayer. Judgment sampling is permitted, but “warrants greater scrutiny by the examiner.”³⁰

Research and Experimentation

The allowability of the research credit under Code Sec. 41 has become a contentious audit issue. Research credit claims are now a Tier I compliance issue.³¹ Part of the problem results from the fact that the basis for claiming the credit depends on facts that are not necessarily reflected in the taxpayer’s books and records. Taxpayers often attempt to fill the gaps by conducting studies of, e.g., the proportion of various employees’ time that is spent on qualifying research. Because of the prohibitive cost of developing this information for all projects or all employees, taxpayers often take a representative sample and extrapolate the results. Similarly, because of the large volume of data, the IRS will often utilize sampling on audit.

Insofar as sampling by taxpayers is concerned, if the taxpayer has conducted a valid statistical sample, the principles discussed above in the FIELD DIRECTIVE ON THE USE OF ESTIMATES FROM PROBABILITY SAMPLES, applicable to samples by taxpayers generally, should apply. However, judgment samples by the taxpayer are categorically rejected. THE RESEARCH CREDIT CLAIMS AUDIT TECHNIQUES GUIDE (May 2008)³² states:

Non-statistical sampling, commonly referred to as “judgment sampling”, possesses none of the scientific safeguards inherent in statistical sampling. The only assurance of accuracy stems from

the judgment of the sampler. The projection of results from non-statistical sampling would only be correct by pure chance. There is no authority that allows taxpayers to use a judgment sample to compute the research credit. Taxpayers would be entitled to the credit only on the activities or expenses that they substantiated from the sample. Any activities or expenses not substantiated should be disallowed.

The IRS's guidelines as to its own sampling on audit of the taxpayer's claimed research credit are contained in the FIELD DIRECTIVE, USE OF SAMPLING METHODOLOGIES IN RESEARCH CREDIT CASES (March 2002).³³ The directive takes the position that sampling of research credit claims by the IRS on audit is permissible, despite taxpayer arguments that research projects are too varied in nature to permit sampling. However, the examination team may only use sampling where there are numerous projects. Sampling is particularly called for where there are a few large projects and many small ones, since if the IRS audited only the large projects, it would be difficult to sustain adjustments to the small ones. Even if the taxpayer agreed to project the adjustments from the large projects to the small ones, that could overstate the allowable credit, since the large projects could be more likely to involve allowable research and to have supporting documents than the small ones. As in the case of taxpayer samples, the IRS may be held to the least favorable confidence interval that results from its sample.

If there are too many projects to be sampled, the directive suggests alternative strategies. The IRS might audit only a few representative projects and give the taxpayer the results. If the taxpayer is dissatisfied, it might be unwilling to provide documentation on the remaining projects, in which case the credit might be denied for lack of substantiation.

As another alternative, the IRS and the taxpayer could mutually agree on a judgment sample. However, the IRS cautions its examination teams that a judgment sample should only be undertaken pursuant to a closing agreement binding the taxpayer to the result.³⁴ Otherwise, the taxpayer might walk away from an unfavorable result. Note that a taxpayer that agrees to a judgment sample may also be well served by a closing agreement, lest the IRS walk away from an unfavorable result.

Pre-Filing Agreement

One procedural alternative to potentially contentious analysis of sampling methodology on audit is to enter into a pre-filing agreement with regard to the issue

with the IRS under Rev. Proc. 2009-14.³⁵ In this way, the taxpayer and the IRS may be able to work more cooperatively on the sample design.³⁶

Sampling in the Courts

Beyond IRS guidance, of course, a taxpayer would be interested in its chances of prevailing in litigation involving sampling. This would be relevant not just to the possibility of actual litigation but to the effect it would have on IRS Appeals' determination of the hazards of litigation in settling an issue.

In this connection, there is a distinction to be drawn between judicial treatment of statistical samples and judgment samples. As might be imagined, statistical samples are much more likely to be accepted but may be rejected where the sample, however carefully conducted, measured the wrong thing. Judgment samples put the taxpayer in a weaker position unless the government has agreed to the sample, but courts have sometimes accepted judgment samples, though holding weaknesses against the taxpayer.

Statistical Samples

There appears to have been surprisingly little litigation as to the validity of statistical sampling in federal tax cases.³⁷ One frequent sampling area involves employee tips. As an example, *L.D. Ross*³⁸ and *J.A. Cohen*³⁹ each described different aspects of an IRS sampling operation at Atlantic City casinos in which time periods and locations were selected at random and tips were analyzed. Adjustments were based on the least favorable (to the IRS) 95-percent confidence interval. The Tax Court accepted the statistical sampling methodology.

Statistical sampling was used by the taxpayer in *Southgate Master Fund, LLC*.⁴⁰ In that case, a stratified statistical sample of a large group of nonperforming loans was sufficient to establish, along with other evidence, that the loans were not worthless when they were acquired. (However, claimed losses were disallowed on other grounds.) Similarly, in *Tele-Communications, Inc.*,⁴¹ the taxpayer used a statistical sample of 10 percent of the subscribers to a cable TV franchise over a 47-month period to determine the likely remaining life of subscriptions in determining future cash flows from the franchise so as to be able to assign a value to it. The court accepted this methodology.

In *National Westminster Bank, PLC*,⁴² the taxpayer used a random statistical sample of three trading days for each of seven tax years at issue to support testi-

mony of its executives that its U.S. branch provided arm's-length rates to related parties in money market transactions. Although the government's expert witness criticized the statistical analysis, he did not present an opinion of his own. The court held that the government had failed to show that there was an issue of material fact as to its proposed transfer-pricing adjustments. As a consequence, the taxpayer was entitled to summary judgment.

On the other hand, a statistical sample, no matter how rigidly adhering to procedural requirements, will be worthless if it samples the wrong population or the wrong characteristics. Thus, in *Prudential Overall Supply*,⁴³ the taxpayer provided millions of industrial garments and related equipment to customers through rental and service agreements. Because of rapid wear and tear on the garments as well as functional obsolescence, the taxpayer deducted their cost when the garments were placed in service. On audit, the IRS asserted that the cost should have been capitalized and depreciated. The IRS supported its adjustment by taking a random statistical sample, utilizing the least favorable (to the IRS) 95-percent confidence interval, and establishing lives of more than a year based on dates on tags on the sampled garments. However, as noted by the court, the tags referenced by the IRS bore the date of manufacture, rather than the later placed-in-service date, and accordingly the statistical information developed was meaningless and did not support the adjustment.

Similarly, in *Banc One Corp.*,⁴⁴ the taxpayer acquired two banks and attempted to establish the facts necessary to amortize the value of the acquired banks' core deposits. The taxpayer engaged a consulting firm, which performed a statistical sample of acquired accounts over a four-year period to determine account closure rates. However, the years measured were those following the acquisitions. The court held that "determination of the useful life of an asset ... must be based on facts existing as of the close of the tax year in issue."⁴⁵ Accordingly, the results of the statistical sample were rejected, and since no other evidence of useful life had been presented, amortization deductions were denied.⁴⁶

Judgment Samples

In a way, the leading case (because it was decided by the Supreme Court) supporting a judgment sample is *Fior D'Italia, Inc.*,⁴⁷ though it did not involve sampling in the strict sense. The taxpayer was a restaurant that the IRS believed was underpaying FICA taxes on employee tips. Using data from the restaurant's Forms 8027, the IRS was able to determine the full amount of charged tips and credit card charges for the tax year. The IRS divided the charged tips by the credit card charges to determine a tip ratio. The IRS then applied this same ratio to cash gross receipts of the restaurant to determine cash tips

and then aggregated the charged and computed cash tips to compute FICA taxes. There was really no sample here (because the IRS had complete credit card data), except in the sense that the IRS used credit card charges as its sample and then made the judgment that cash pay-

ments bore the same tip ratio. The Supreme Court held that the IRS was not "unreasonable" in making this estimation. Since the taxpayer had not challenged the specific aspects of the computation (such as whether cash tips might properly be in a lower ratio than charge tips), just the overall reasonability, the IRS adjustment was sustained. Thus, the case does not tell us too much about sustainability of judgment samples. However, it can be read to say that a judgment sample will be more acceptable if the standard to be met is only whether the result is reasonable, rather than whether the result is exactly correct.

Judgment samples are frequently used in cases involving research and experimentation tax incentives. Thus, judgment samples were used in *Lockheed Martin Corp.*⁴⁸ (four of the taxpayer's largest projects, amounting to about 65-percent coverage of the amounts at issue); *Norwest Corp.*⁴⁹ (eight out of 67 projects); *Yellow Freight System, Inc.*⁵⁰ (nine out of an unspecified number of projects); and *Union Carbide Corp.*⁵¹ (five out of over 100 projects). It is important to keep in mind that these sample projects were agreed to by both the taxpayer and the government, with the idea of extrapolating the results of the litigation to the taxpayer's other projects. Neither side was allowed to choose them unilaterally.

Where the taxpayer alone chooses a judgment sample to establish a tax attribute and then must litigate

It is ... important to understand that there are two kinds of samples: *statistical* and *judgment*. Each is subject to different legal treatment in connection with tax matters.

the issue, the taxpayer is on weaker (but not necessarily unsustainable) ground than if a statistical sample had been used. The fundamental decision on which these cases are based is *G.M. Cohan*.⁵² In that case, although the taxpayer lacked adequate records to support a deduction in a situation where it was clear that he had incurred at least some deductible expenses, the Second Circuit rejected a Board of Tax Appeals holding that no deduction was allowed, and instead held that “the Board should make as close an approximation as it can, bearing heavily if it chooses upon the taxpayer, whose inexactitude is of his own making.”⁵³

This point (and the risks assumed by the taxpayer by conducting a judgment sample instead of a valid statistical sample) is illustrated by two virtually identical cases, *Lutheran Mutual Life Ins. Co.*⁵⁴ and *Phoenix Mutual Life Ins. Co.*⁵⁵ Both cases involved insurance companies that conducted a judgment sample of their independent agents to find out how much time the agents spent on policy loans. Commissions paid to the agents that could be allocated to that activity could be deducted from the taxpayers’ investment income. There were no actual records of time the agents spent on the loans.

In *Lutheran Mutual*, the survey was designed and conducted by the assistant vice president of marketing, who was not a professional statistician. The recipients of the survey questionnaire were personally selected by him, not at random, making it a judgment sample. Less than 10 percent of the agents were sampled. The sample was biased toward full-time agents, who were more likely to have policy loan activity. About 25 percent of the agents sampled did not respond; these nonresponses were not followed up. Finally, both the questioner and responder knew why the questions were being asked.

The district court rejected the survey results and denied the deduction. On appeal, the Eighth Circuit held the taxpayer to the following standard:⁵⁶

To establish the trustworthiness of a survey, it must be shown (1) that a proper “universe” was examined and a representative sample was chosen; (2) that the persons conducting the survey were experts; (3) that the data were properly gathered and accurately reported; (4) that the sample design, the questionnaires, and the manner of interviewing met the standards of objective surveying statistical techniques; and (5) that the interviewers, as well as the respondents, were unaware of the purpose of the survey.

Based on the facts as discussed above, the survey obviously fell well short of these criteria. However, the Eighth Circuit, while affirming, framed its holding in these terms:⁵⁷

Although we might have been inclined to accept the results of the survey, applying a little Kentucky windage to compensate for the empirical inadequacies inherent in the somewhat amateurish manner in which the survey was conducted, we cannot say that the district court erred in holding that the taxpayer had failed to prove by a preponderance of the evidence that it was entitled to deduct [the amounts at issue].

In other words, notwithstanding that the Eighth Circuit upheld the district court, it found that it also could have upheld a trier of fact who concluded that the taxpayer was entitled to some amount. However, the inadequacies of the survey would be held against the taxpayer, echoing the rule in *Cohan*.

The Eighth Circuit’s invitation was, in effect, accepted in *Phoenix Mutual*, in a factual scenario that the Tax Court described as “virtually identical” to *Lutheran Mutual*.⁵⁸ As in *Lutheran Mutual*, the survey was designed by a taxpayer employee and was made of the group most likely to have policy loan activity (top producers), who may have been aware of the reasons for the survey. Here, the court accepted the general validity of the survey. However, since the survey was biased and may have had other flaws, the court allowed the taxpayer only about 70 percent of the benefit indicated by the survey results. The court specifically invoked *Cohan* in making its estimation.

Cohan has also figured prominently in several research credit cases, with similar mixed results.⁵⁹ Although these cases do not involve judgment sampling *per se*, their references to *Cohan* hold open the possibility that taxpayer judgment sampling might be accepted as a basis for claiming the research credit, though as in *Phoenix Mutual*, the taxpayer may bear the burden of its inexactitude.

One interesting question is whether judgment samples can be excluded from evidence under the principles of *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,⁶⁰ which bans “junk science” presented by expert witnesses. The very nature of the *Cohan* doctrine, however, is that the evidence presented is not well substantiated. The evidence is admitted, but negative inferences resulting from its defects are held against the taxpayer. Ac-

cordingly, *Daubert* should not prevent taxpayers from continuing to attempt to introduce judgment surveys, though as noted above, the very nature of such surveys puts the taxpayer at risk that, although some may be accepted, some (as in *Lutheran Mutual*) will be viewed as so bad as to be excluded outright.

Of interest in this regard is *Exxon Corp.*,⁶¹ a post-*Daubert* case in which the court conducted a judgment sample. *Exxon* involved percentage depletion deductions. For an integrated gas producer like Exxon, which added value to the gas through transportation and processing after it emerged from the well, the deduction was to be based on a “representative market or field price” (RMFP) for unprocessed gas sold at the wellhead. Importantly, the determined RMFP did not have to be precise, but could be an approximation.⁶² To establish the RMFP, Exxon produced a study of over 2,000 gas contracts in the relevant geographical area.⁶³ However, under many of these contracts, the producer provided substantial transportation of the gas or processed the gas, thus potentially adding to its value, which would improperly inflate the price on which Exxon’s depletion would be based.

Faced with the daunting task of reviewing 300,000 pages of contracts,⁶⁴ the court was able to eliminate most without specific examination on the basis that Exxon could not show that the contracts did not involve transportation or processing of the gas. The parties agreed that some of the remaining contracts could be used to determine the RMFP, leaving 433 contracts in dispute.⁶⁵ Of these, 130 involved contracts where the gas producer might have processed the gas. The court examined “roughly 50” of the 130 contracts and “emphasized transactions that involved large volumes of gas.”⁶⁶ In other words, the court took a judgment sample focusing on the largest contracts. Among these 50 contracts, the court found 18 that did not involve processing and thus could be included in determining the RMFP. The court excluded the remaining reviewed contracts, as well as those that the court had not reviewed, from the RMFP computation.⁶⁷

Similarly, to determine whether transportation from the well had impermissibly added to the price of the gas, the court examined “a substantial number” (exact number unspecified) of additional contracts and found that 23 did not involve such

transportation.⁶⁸ The court further reasoned that in the case of single- (as opposed to multiple-) well contracts, it made economic sense for the delivery point to be close to the well. After evaluating “a reasonable number” (exact number unspecified) of contracts, the court apparently accepted all 115 contracts in this category as eligible to be included in the RMFP computation.⁶⁹ Again, the court appears to have taken a judgment (or at least unexplained) sample of contracts to reach its conclusion as to the characteristics of a greater number.

Standing in stark contrast is the court’s approach to Exxon’s analysis of a sub-issue in the case. Exxon attempted to establish that the area covered by gas leases was typically small, so that it would be impossible for the producer to provide significant transportation of the gas so long as it was delivered to the purchaser somewhere on the lease plot. To support this, one of Exxon’s experts sampled 300 out of the more than 2,000 contract files and determined that the median lease plot was only 233 acres. The court, specifically referring to *Daubert*, rejected the expert’s analysis as valueless, on the basis that the expert did not (i) explain why he took only a sample; (ii) explain which contracts he had sampled; (iii) take account of multiple-well leases; (iv) compute a mean (instead of just a median); or (v) measure variability.⁷⁰

Besides the problem of a potentially bad sample, there are times when, because of the underlying fact to be proved and the threshold of proof, *no* sample by the taxpayer will be sufficient. This is illustrated by *CRST, Inc.*⁷¹ In 1980, the trucking industry was deregulated. ICC operating authorities, formerly quite valuable, became almost worthless, as it became easy to obtain new ones. The taxpayer trucking company claimed to have abandoned its permanent ICC operating authorities in connection with the deregulation and claimed a 1980 abandonment loss for their cost. The taxpayer was to replace the supposedly abandoned authorities with a new, national operating authority. However, it did not obtain its new authority until mid-1981. Meanwhile, the taxpayer purported to be operating under various temporary authorities rather than those it had purportedly abandoned. The taxpayer performed a study of 10 percent of its shipments following the purported abandonment to show it was not using the “abandoned” permanent authorities, only other, temporary authorities.

The court in effect held that, since legal standard was abandonment, *no* survey could meet the nec-

essary burden of proof because the taxpayer had to show an affirmative act of abandonment of the authorities, which it had not done. Although it could have been operating under the temporary authorities as identified in its study, the taxpayer could just as well have been using the permanent authorities that it had claimed to have abandoned. Thus, the study was ineffective in establishing the abandonment.

Conclusion

The conclusions here are straightforward. Statistical sampling is an accepted part of IRS audits and, in many instances, return preparation, at least where extraction of data from the taxpayer's books and records is impracticable. If the relevant amounts are reflected in the taxpayer's books and records, the IRS may insist that no survey will suffice. Proper sample design, whether by the IRS or the taxpayer, is critical to the result.

If the taxpayer must sample to establish a benefit, it will have to balance the cost of sampling against bearing the detriment of an unfavorable confidence interval.

Judgment samples have substantial litigation hazards, at a minimum. If some weaknesses are revealed, the taxpayer may be able to apply *Cohan*, though risking a considerable "haircut" by the court as in *Phoenix Mutual*. Also, some studies are judged to be so bad, as in *Lutheran Mutual*, that the court will refuse to apply *Cohan*. As indicated by *Fior D'Italia* and *Exxon*, however, a court may be more hospitable if the object of the sampling is simply to establish that a position is reasonable or an approximation, rather than to determine an exact amount.

Finally, there may be limited situations, like the *CRST* abandonment case, where the legal standard to obtain the benefit is so difficult that no survey, no matter how well designed, will be accepted.

ENDNOTES

¹ For a more detailed look at the statistical aspects, see Wendy Rotz & David Hudson, *IRS Addresses Use of Statistical Sampling in Research Credit Cases*, 2006 TNT 104-50 (May 31, 2006); Mary Batchner, *Statistical Sampling: A Potential Win for Business Taxpayers*, 53 TAX EXECUTIVE 464 (Nov. 2001).

² For general background, see, e.g., Rudolf J. Freund & William J. Wilson, *STATISTICAL METHODS* (2d ed. 2003).

³ For an article focusing on how sample design can affect the reliability of the result, see Jeri Mulrow & Eric Falk, *What's the Risk Due to Sampling in Tax Audits and Planning?* 90 TAX NOTES 939 (Feb. 12, 2001).

⁴ LMSB-4-0809-032 (Nov. 3, 2009), available at www.irs.gov/pub/irs-utl/statsamplingidd.pdf. The field directive modifies a prior document of the same name issued March 14, 2002.

⁵ See Rev. Proc. 2007-35, 2007-1 CB 1349, discussed below.

⁶ *Research Credit Claims Audit Techniques Guide (RCCATG): Credit for Increasing Research Activities §41*, LMSB-04-0508-030 (May 2008), available at www.irs.gov/businesses/article/0,,id=183208,00.html; *Field Directive, Use of Sampling Methodologies in Research Credit Cases* (March 2002), available at www.irs.gov/pub/irs-utl/field_directive_samp_method_research_credit_cases.pdf.

⁷ The precision of the estimate is three percent (one side of the six percent confidence interval). The estimate itself is 35 percent. $3\% / 35\% = 8.6\%$, which is within the 10-percent precision required to be allowed to use the point estimate of 35 percent.

⁸ Available at www.irs.gov/pub/irs-utl/09-17-03_joint_audit_planning_process_with_cover.pdf.

⁹ *Id.*, at 8.

¹⁰ Litigation Guideline Memorandum TL-97 (Sept. 9, 1992).

¹¹ *Id.* §4.47.3.3.1(3)(C).

¹² Judgment samples can lead to very unpredictable results; the taxpayer should agree to one only with great caution as to how the sample is selected. The taxpayer can complain if an examiner uses a judgment sample without its agreement, but the taxpayer may then be burdened with a more intrusive audit to develop a statistical sample. Depending on how poor the judgment sample is, the taxpayer may be better off arguing in Appeals that the examination team's work was so inadequate that it should be disregarded or materially adjusted in the taxpayer's favor. This may require the taxpayer to produce a statistical sample of its own to support the protest.

¹³ The IRS has issued published guidance as to acceptable sampling in a number of specific areas. These are discussed below. Some of these documents are modifications of earlier positions.

¹⁴ Rev. Proc. 64-4, 1964-1 CB 644, *modified and amplified*, Rev. Proc. 65-5, 1965-1 CB 720.

¹⁵ Reg. §1.451-4.

¹⁶ Rev. Proc. 72-36, 1972-2 CB 771.

¹⁷ Although trading stamps themselves are seldom issued today, many taxpayers use this method to account for their customer loyalty programs. See INTERNAL REVENUE MANUAL §4.43.1.12.6.4 (July 23, 2009),

available at www.irs.ustreas.gov/irm/part4/irm_04-043-001-cont05.html.

¹⁸ Rev. Proc. 81-70, 1981-2 CB 729. See also Notice 2004-44, 2004-4 CB 32, soliciting public comments on whether the procedures prescribed by Rev. Proc. 81-70 are still appropriate. Where such a statistical sampling is conducted that involves less than 50 percent of the acquired corporation's shareholders, knowledge that the acquiring company gains as to purchase dates by former shareholders of their stock will not taint a subsequent tax-free distribution under Code Sec. 355 (Reg. §1.355-6(f)(5), Example 3).

¹⁹ Rev. Proc. 85-51, 1985-2 CB 717.

²⁰ Rev. Proc. 2002-55, 2002-2 CB 435.

²¹ Rev. Rul. 2003-43, 2003-1 CB 935.

²² Rev. Proc. 2004-29, 2004-1 CB 918.

²³ Rev. Proc. 2004-34, 2004-1 CB 991.

²⁴ Rev. Proc. 2007-35, 2007-1 CB 1349.

²⁵ Notice 2009-46, 2009-23 IRB 1068.

²⁶ Available at www.irs.gov/pub/irs-utl/dollar_value_lifo_-_segments_of_inventory.pdf.

²⁷ See also Reg. §1.472-8(e)(1) (allowing computation of an index "by double-extending a representative portion of the inventory in a pool or by the use of other sound and consistent statistical methods").

²⁸ Available at www.irs.gov/businesses/corporations/article/0,,id=148213,00.html.

²⁹ Available at www.irs.gov/businesses/article/0,,id=134152,00.html.

³⁰ In contrast, see FSA 200151004 (Sept. 7, 2001), in which the taxpayer attempted to reclassify certain assets to shorter lives under ACRS. The taxpayer supported the

ENDNOTES

- reclassification with a judgment sample of “a very small percentage of the population,” which included only five years although 10 years were involved, and which excluded any sampled item if there was inadequate documentation (rather than sampling an additional item or treating the item as if no reclassification was warranted). The FSA held that the result could not be extrapolated to the rest of the asset population.
- ³¹ See generally www.irs.gov/businesses/corporations/article/0,,id=200599,00.html.
- ³² Available at www.irs.gov/businesses/article/0,,id=183208,00.html#15.
- ³³ Available at www.irs.gov/pub/irs-utl/field_directive_samp_method_research_credit_cases.pdf.
- ³⁴ *Audit Techniques Guide: Credit for Increasing Research Activities (i.e. Research Tax Credit)—Sampling Methodologies* (June 2005), available at www.irs.gov/businesses/article/0,,id=156863,00.html.
- ³⁵ Rev. Proc. 2009-14, IRB 2009-3, 324.
- ³⁶ The author participated in one case in which, although the taxpayer was rejected from the pre-filing agreement program, the IRS team manager arranged for early assignment of an IRS statistician to the audit to work with the taxpayer in designing the sample, essentially producing the same result.
- ³⁷ For an extensive discussion of judicial treatment of statistical sampling *not* involving federal tax cases, see PHILLIP I. GOOD, *APPLYING STATISTICS IN THE COURTROOM* (2001).
- ³⁸ *L.D. Ross*, 58 TCM 1069, Dec. 46,236(M), TC Memo. 1989-682.
- ³⁹ *J.A. Cohen*, 60 TCM 1509, Dec. 47,054(M), TC Memo. 1990-650.
- ⁴⁰ *Southgate Master Fund, LLC*, DC-TX, 2009-2 USTC ¶ 50,593.
- ⁴¹ *Tele-Communications, Inc.*, 95 TC 495, Dec. 46,970 (1990).
- ⁴² *National Westminster Bank, PLC*, FedCl, 2006-1 USTC ¶ 50,107, 69 FedCl 128.
- ⁴³ *Prudential Overall Supply*, 83 TCM 1545, Dec. 54,723(M), TC Memo. 2002-103.
- ⁴⁴ *Banc One Corp.*, 84 TC 476 (1985), *aff'd per curiam*, CA-6, 815 F2d 75.
- ⁴⁵ *Id.*, at 499.
- ⁴⁶ See also *Southern Bancorporation, Inc.*, CA-4, 88-1 USTC ¶ 9,344, 847 F2d 131.
- ⁴⁷ *Fior D'Italia, Inc.*, 2002-2 USTC ¶ 50,459, 536 US 238.
- ⁴⁸ *Lockheed Martin Corp.*, CA-FC, 2000-1 USTC ¶ 50,401, 210 F3d 1366.
- ⁴⁹ *Norwest Corp.*, 110 TC 454, Dec. 52,758 (1998).
- ⁵⁰ *Yellow Freight System, Inc.*, ClsCt, 92-1 USTC ¶ 50,029, 24 ClsCt 804.
- ⁵¹ *Union Carbide Corp.*, 97 TCM 1207, Dec. 57,753(M), TC Memo. 2009-50. The author's firm was counsel to the taxpayer.
- ⁵² *G.M. Cohan*, CA-2, 2 USTC ¶ 489, 39 F2d 540 (1930).
- ⁵³ *Id.*, at 543–44.
- ⁵⁴ *Lutheran Mutual Life Ins. Co.*, CA-8, 87-1 USTC ¶ 9257, 816 F2d 376.
- ⁵⁵ *Phoenix Mutual Life Ins. Co.*, 96 TC 497, Dec. 47,252 (1991).
- ⁵⁶ *Lutheran Mutual Life Ins. Co.*, *supra* note 54, 816 F2d at 378, *citing Pittsburgh Press Club*, CA-3, 78-1 USTC ¶ 9,457, 579 F2d 751 and Judicial Conference of the United States, *Handbook of Recommended Procedures for the Trial of Protracted Cases*, 25 F.R.D. 351, 429 (1960).
- ⁵⁷ *Lutheran Mutual Life Ins. Co.*, *supra* note 54, 816 F2d at 379.
- ⁵⁸ *Phoenix Mutual Life Ins. Co.*, *supra* note 55, 96 TC, at 545.
- ⁵⁹ *E.V. Fudim*, 67 TCM 3011 Dec. 49, 867(M), TC Memo. 1994-253 (1994); *N.E. Eustace*, 81 TCM 1370, Dec. 54,280(M), TC Memo. 2001-66; *Union Carbide Corp.*, 97 TCM 1207, Dec. 57,753(M), TC Memo. 2009-50 (2009); *A.R. McFerrin*, CA-5, 2009-1 USTC ¶ 50,430, 570 F3d 672; *Trinity Industries, Inc.*, DC-TX, 2010-1 USTC ¶ 50,219.
- ⁶⁰ *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 US 579 (1993). See also Fed. R. Evid. 702. For an example of the application of *Daubert* to statistical analyses, see *Allapattah Services, Inc. v. Exxon Corp.*, 61 FSupp2d 1335 (S.D. Fla. 1999).
- ⁶¹ *Exxon Corp.*, FedCl, 2000-1 USTC ¶ 50,116, 45 FedCl 581, *aff'd in part and rev'd in part on other issues*, CA-FC, 2001-1 USTC ¶ 50,348, 244 F3d 1341.
- ⁶² *Id.*, at 700.
- ⁶³ It is not clear from the opinion how this sample was developed, other than that it was based on public filings and other information available to Exxon's experts.
- ⁶⁴ *Exxon Corp.*, *supra* note 61, 45 FedCl, at 603.
- ⁶⁵ *Id.*, at 699.
- ⁶⁶ *Id.*, at 700 (emphasis added).
- ⁶⁷ *Id.*, at 701.
- ⁶⁸ *Id.*, at 705–06.
- ⁶⁹ *Id.*, at 707.
- ⁷⁰ *Id.*, at 679–83.
- ⁷¹ *CRST, Inc.*, 92 TC 1249, Dec. 45,761 (1989), *aff'd*, CA-8, 90-2 USTC ¶ 50,426, 909 F2d 1146.