

2007

The IRS Research Bulletin

Publication 1500

The IRS
Research
Bulletin

Proceedings of the 2007 IRS Research Conference

Research
Analysis
Statistics

Comments on Tax Practitioners—Perspectives and Impact

John Scholz, Florida State University

The three articles on tax practitioners presented in this volume provide creative examples of how research focusing on practitioners rather than taxpayers can help improve our understanding of tax compliance and of enforcement policies. All three underscore the critical role played by practitioners in the compliance process and hence the importance of including strategies toward practitioners in any well-considered enforcement policy. The Bloomquist, Albert, and Edgerton article (henceforth BAE) notes, for example, that practitioners prepared 64 percent all of returns filed in 2005, accounting for 74 percent of reported tax. Tax practitioners are a relatively small professional group with considerable knowledge about reporting and considerable potential influence over the accuracy of reporting for the majority of income tax. An efficient enforcement program needs to consider carefully what can best be achieved by dealing with practitioners as an alternative to dealing directly with the overwhelming number of individual taxpayers.

The three articles also emphasize different aspects of the compliance impacts of practitioners, which is important for developing the broad perspective necessary to understand their role in the compliance process. The Taxpayer Advocate Services article (henceforth TASA) by Wilson et al. studies tax practitioners in order to diagnose potential errors involved in IRS audits. This approach makes use of the expertise of tax practitioners to develop an alternative means of providing a standard of compliance by which to judge the relative accuracy of audits. It relies on the assumption that the presence of practitioners will tend to reduce possible errors made by auditors, at least in the case of audits involving earned income credits. The implication of this research approach is that we can compare the results of represented and unrepresented audits in order to identify issues and audit categories where IRS performance may need to be modified. In their case, they find that the difference for earned income credits is significant and sufficiently large to suggest the need for improved techniques and training.

BAE makes use of the IRS's Automated Underreporter (AUR) program to evaluate the performance of tax practitioners. It reflects the assumption that the difference in reporting discrepancies between represented and unrepresented taxpayers can indicate to what extent different types of practitioners

increase or decrease the compliance level of their clients. The implication here is that techniques can be developed for identifying particular practitioners and practitioner types whose clients are consistently found to have higher rates of noncompliance than their appropriate comparison group. To the extent that such identification techniques prove valid, and that effective safeguards can be created to protect misidentified practitioners, these techniques can be incorporated into general enforcement programs. Targeting individuals and groups of practitioners who can influence tens, hundreds, or even thousands of tax returns may prove to be a more efficient use of scarce enforcement resources than targeting individual taxpayers.

Finally, the National Society of Tax Professionals study (henceforth NSTPS) by Whitlock uses the rapport between tax professionals and their clients to gather information on compliance problems facing taxpayers. The anecdotal data provided through 100 practitioners was used to analyze the relative frequency of various obstacles to compliance. The implication here is that the practitioner's wealth of experience about compliance problems can be tapped systematically to identify the most frequent problems facing taxpayers and possibly to help devise more efficient means of mitigating these problems through changes in laws, administrative procedures, or tax forms. Again, the advantage of surveying practitioners rather than taxpayers stems both from the broader perspective enjoyed by these practitioners and by their greater knowledge of the tax system.

Using Practitioners for Audit Diagnostics: The Case of the Earned Income Credits

The imaginative TASA focuses on a taxpayer population that is particularly likely to suffer from a lack of resources to confront auditors if mistakes are made. Taxpayers eligible for earned income credits are generally poor and lacking in skills necessary for complying with documentation and reporting problems. As TASA notes: "Recent focus groups and interviews with taxpayer representatives have noted various barriers imposed by the correspondence audit process. Some of these obstacles include lost paperwork, inconsistent requests for documentation, and poor communication."

The study very carefully documents that taxpayers represented by tax professionals in audits end up with more favorable outcomes than those who are not represented. Whether one looks at the proportion found to be eligible, to emerge with full EIC claims intact, to not owe more tax, or to have no change in taxes due, represented taxpayers do better than unrepresented taxpayers. The impact of representation in reducing the likelihood of a

negative tax change holds even when the significant differences in the types of taxpayers who use representation are taken into consideration, controlling for filing status, dependents (the unexplained DDb score in the article), adjusted gross income, and type of audit issue.

The article utilizes a large and well-considered design for gathering data, and an advanced matching approach to control for the selection bias problem—the problem that differences between the taxpayers who do and do not have representation may fully explain the observed differences in audit results. If true, then the observed impacts of representation on audit outcomes will disappear once these differences are accounted for. The matching approach used in TASA to control for these differences provides a convincing result.

This excellent analysis provides very solid evidence for the authors' primary contention, with some but not all details normally reported in academic journals made available in the appendix. The analysis is based only on changes in taxes, which is arguably the most important of the observed differences between the groups. But the reader is left wondering whether the selection bias correction affects the other variables as well. The uncorrected results are impressive because they included a number of measures to assess the difference between the groups—shouldn't each variable be assessed for bias in the same manner? More importantly, after introducing the means to correct for selection bias, Tables 11 through 15 appear to be based on the uncorrected data rather than the selection-bias controlled data. So, we unfortunately do not know how much of these differences can be attributed to representation, and how much to the documented differences between the groups.

The final recommendation in TASA is that the study should be replicated on more recent data. This is far too mild, given the high quality of this research, the limitations on timeliness of data affecting all tax-related research, and our limitations as a society to devote resources to research. The research makes a very clear case that representation causes a major difference in outcomes, a case that could be strengthened by the suggested further applications of the same selection bias analysis. The remaining question is what if anything to do about these findings. Earlier in the report, the authors make the relevant suggestion: "An affirmative answer to this question [of differences in outcomes] would highlight the need to reformulate IRS compliance programs that verify the EIC in such a way as to minimize the use of audits and/or modify the way the audits are conducted." That is a fitting conclusion for the article, and one that IRS research should be encouraged to make as a recommendation!

There is the unaddressed question about interpreting the differences between audits with and without representation. In the EIC case, the problems of compliance for the taxpayers noted in the article, combined with institutional pressures to be tough in auditing these low-income returns (“EIC audits represent approximately 43 percent of all IRS individual taxpayer audits,” as reported in the article), provide a strong basis for concluding that auditors are indeed most likely to be overly aggressive and to err on the side of denying credits that should have been granted. Extending this comparison approach to uncover audit problems in areas beyond EIC might require more serious attention to the possibility that representation actually enhances error by inducing auditors to underassess taxes rather than spend the necessary time to deal with internal and court appeals that are likely when practitioners are present. In the EIC case, however, it is most plausible to conclude that there is a major problem of justice here that the IRS and Congress need to address.

Using Record Matching To Detect Problem Practitioners: Finding the Bad Apples

BAE makes use of the impressive amount of data now available in the IRS Compliance Data Warehouse, first to investigate the impact of practitioners on tax compliance, and second to provide a basis for a new technique to identify the bad apples among practitioners associated with noncompliance. The article is in part a response to the GAO study they cite that found a considerable range of errors in hypothetical tax returns prepared by a sampling of tax preparers. BAE attempts to use extensive IRS data to provide a better assessment of the math errors and potential misreporting discrepancies found in returns prepared by practitioners, and also to do something about reducing the “discrepancies” by focusing on practitioners associated with the highest levels of these discrepancies.

BAE finds first that returns filed by practitioners involve fewer math errors but have more misreporting discrepancies than those by self-preparers. They also note that the percentage of practitioner returns containing misreporting discrepancies falls as firm size increases, and they investigate several other indicators that affect the relative number of discrepancies. Given these discrepancies, they move to their main task of developing a method for identifying practitioners associated with higher levels of discrepancies. The main thrust of the paper focuses on devising improved enforcement targeting methods, rather than on understanding the role of practitioners in the compliance process, and should be evaluated accordingly.

The major question in evaluating this paper is about the quality of the primary measure of reporting accuracy. Two indicators of accuracy are used in the first section—the presence or absence of a math error, and the presence of a nonzero dollar discrepancy in the IRS's Automated Underreporter (AUR) program. The second measure is also the basis for the second section of the paper. While the math error measure is well-defined and uncontroversial, the AUR measure is trickier to assess. The AUR Program attempts to match taxpayer income and deduction information submitted by third parties to amounts reported on individual income tax returns, and to use mismatches to identify potential audit cases. One report found that 55 percent of originally identified cases (78 percent of the 70 percent actually pursued) resulted in increased tax liability, suggesting that the measure has some ability to identify potential tax discrepancies.¹ A 55-percent success rate may be reasonable for a screening device for targeting some form of audit procedure, although it does not provide much confidence as a measure related to accuracy of reporting. In addition, the AUR measure focuses only on one kind of detectable discrepancy, and does not take advantage of other audit-based measures like those based on the Taxpayer Compliance Measurement Program or its recent replacement. It, therefore, provides a very specialized and limited measure of discrepancies in reporting associated with noncompliance.

Tables 1-5 report these measures of discrepancy to make various comparisons relating to preparers. For example, Table 1 compares discrepancies between self-preparers and two categories of preparers. Ideally, the authors would like to conclude that the differences were due to the preparers. However, most of these comparisons are subject to the same selection problem as those discussed for TASA—the 2-percent higher discrepancy rate for paid preparers in 2005 might be due to the influence of preparers, but might also be due to differences between taxpayers who do and do not use paid preparers. Table 3 at least controls for some possible variance across taxpayers by focusing on line items, but this still does not provide much assurance that the observed differences could not best be explained in terms of the difference between taxpayers who do and do not use preparers. While these tables may be useful from the perspective of identifying areas with the greatest discrepancy, they tell us little about whether preparers make any difference at all in levels of discrepancy or compliance after controlling for differences in taxpayers who go to preparers. This is a precaution that is not clearly

¹ United States Treasury, *More Information Is Needed To Determine the Effect of the Automated Underreporter Program on Improving Voluntary Compliance*, August 2003, Audit Reports (Audit # 200340024), Reference Number 2003-40-180 (last viewed October 26, 2007, at <http://treasury.gov/tigta/auditreports/2003reports/200340180f.pdf>)

noted in the text. For example, the statement that “paid preparers account for a higher number and a larger percentage of tax returns with a potential AUR discrepancy” is accurate in terms of targeting enforcement at areas with higher discrepancies, but is very misleading if interpreted as meaning that preparers are somehow responsible for these discrepancies. The data as presented in the tables are not capable of supporting the latter argument, and do not without further analysis provide a justification for focusing on individual practitioners.

The main business of the article, however, is in developing a method of utilizing available data already used to detect individual returns in order to detect individual practitioners whose returns consistently have higher rates of discrepancy than would be expected. The problem here involves how to determine appropriate levels of expectation. Ideally, one would compare all tax returns prepared by each preparer with all “similar” tax returns to determine the deviation associated with this preparer. Then, enforcement efforts could target those preparers associated with the highest rates of deviation. The comparison with similar tax returns helps reduce the selection problem by using only similar returns as a basis for evaluation.

But how can one identify the set of similar returns to provide a targeting mechanism within the constraints of data and computer availability? This is the critical question, and the answer in the article is not completely clear. BAE appears to determine similarity in terms of geography, comparing the actual discrepancies observed for a preparer with the expected discrepancy for taxpayers drawn randomly from the same Zip Codes in which the preparer’s clients live. Figure 1 suggests that only taxpayers using practitioners were used for the comparison, which would provide a much better comparison group than would a comparison based on all taxpayers. Additional matching on other characteristics like income level, taxpayer status, and particular line items could further improve the potential accuracy of the comparison group.

The true test of the suggested procedure or an improved version featuring better matches, however, will depend on how well the system works in targeting practitioners for audits. Just as with other features of the AUR system, it is the success rates of the resultant enforcement actions (compared with an appropriate baseline) that will ultimately determine whether this procedure is useful at all. Do audits of the returns of practitioners receiving high scores with this procedure in fact turn out to produce greater corrections of underreporting than average?

This is where the selection effect problem takes on very practical significance. If the selection effect is the only reason that practitioners differ in deviation rates—that is, if one practitioner’s higher deviation rate is only

due to the characteristics of the individuals who use that practitioner—then, the system of targeting practitioners is unlikely to accomplish anything more than the existing system targeting individual taxpayers can accomplish. In order to offer some added advantage, the practitioner-based system must find those differences attributed to the practitioner and not to the practitioner's clients. It remains to be determined whether the proposed Zip Code comparison will be sufficient to provide the "similarity" needed to identify these differences.

The current analysis is limited to the currently-available AUR measure and is, therefore, very limited as a general tool for targeting a wider set of enforcement actions. However, there is no reason that it could not be extended to utilize other information derived from the extensive data available from audits. Such audit data need to be treated carefully because they provide observations on only a limited number of returns. But detection-controlled estimation techniques could be used to control for these problems in order to analyze the audit records of practitioners, to seek the same pattern of audit-measured deviations that are higher than expected. In short, this paper provides one example of the kinds of models that can be used to analyze patterns of compliance associated with individual practitioners rather than individual taxpayers. Since each practitioner is associated with many returns, the data patterns to be explored are richer and the potential impact on compliance is broader in scope.

Using Practitioners To Detect Evolving Compliance Problems

NSTP suggests how tax practitioners can be used as an additional means for detecting compliance problems. The members of NSTP were asked prior to the 2007 filing season to gather and report anecdotal evidence identifying why their clients "were having difficulty in voluntarily complying with tax code." By the end of April 2007, 100 of the 5,000 members of NSTP responded. The article reports the relative frequencies of typical responses, providing a summary of NSTP members' perspectives on the main obstacles to voluntary compliance.

The study suggests that the responses from 100 members represent summary observations from the approximately 65,000 returns (including joint returns) prepared by those members. The frequency for each type of compliance obstacle is calculated on an assumption that the anecdotal evidence reflects the true proportion of obstacles encountered. Although the actual frequency among taxpayers is likely to differ (in part because practitioners only know the obstacles that are visible to them), the reported

frequencies can at least represent a measure of beliefs among practitioners who reported.

The potential value of such surveys is to provide policymakers and administrators information immediately after the tax filing season about compliance obstacles encountered. NTSP confirms some of the longstanding and well-known obstacles:

- Procedural difficulty with IRS, including communication;
- Burdensome reporting and tax filings;
- Unreasonable penalty and interest assessments; and
- Insufficient encouragement to file and pay timely.

Perhaps more helpfully, it highlights a somewhat more surprising perceived lack of importance or priority among taxpayers who “would get to it [filing return] when other more important issues no longer took priority.”

Surveys of practitioners may become most valuable to the extent that they could be used as early warning detectors of new compliance obstacles that emerge each year, which might then be dealt with before they become worse. But to convince policymakers and administrators about the importance of any new problem, more systematic methods of providing reliable frequency counts would be needed. Are the 100 reporting practitioners typical of NTSP members, or do they represent some particular set of clients more likely to complain about compliance obstacles? To what extent do the reported obstacles truly represent the frequency among represented taxpayers? Do the reports reflect strategic concerns of practitioners rather than taxpayers?

Although these questions would be difficult to answer with the current survey, the next survey could be designed to reduce these problems and provide a clearer picture of the relative frequency of newly discovered problems among specific types of taxpayers. The challenge is to devise transparent survey techniques and analytic methods that could pinpoint real problems without triggering an audit response from the IRS, since such responses would discourage practitioners from openly identifying auditable problems. For example, a survey of practitioners representing the EIC clients studied in TASA would have long ago identified the problems raised in that paper about the compliance difficulties—and could have probably done so more effectively than a survey of the taxpayers. By working together, the IRS and NSTP or other organizations of professionals may be able to design mechanisms to utilize the extensive experience of practitioners in order to provide timely information about changing compliance problems. Such a survey-based mechanism could potentially provide an alternative means of detecting

new compliance issues years before they would be uncovered through audit programs.

In sum, the three studies in this section on Tax Practitioners provide excellent examples of research that can help improve tax administration and policy. TASA and NSTP show how the expertise and experience of practitioners can be used to improve audits and other aspects of tax administration, while BAE shows how research might develop better means of identifying the individual practitioners in the dysfunctional sector of tax practitioners. In each case, I have suggested some remaining research issues that if solved could strengthen these efforts. The level of sophistication already achieved, as exhibited particularly in the two articles representing IRS research, demonstrates the great progress that continues to be made in applying existing research techniques capable of enhancing the relevance and power of tax compliance and enforcement studies.