2007 The IRS Research Bulletin Publication 1500

The IRS Research Bulletin

Proceedings of the 2007 IRS Research Conference

Research
Analysis
Statistics

Evaluating Preparation Accuracy of Tax Practitioners: A Bootstrap Approach

Kim M. Bloomquist, Michael F. Albert, and Ronald L. Edgerton, Internal Revenue Service

In recent years, both the number and share of individual taxpayers who rely on tax practitioners to prepare their Federal income tax returns have increased steadily. In 1996, paid practitioners prepared 63 million (53 percent) individual income tax returns. By 2005, the number of paid preparer returns topped 80 million (62 percent).[1] Even more important is the share of taxes reported on paid preparer returns. In 1996, 65 percent of total taxes were reported on returns prepared by tax practitioners. By 2005, the preparer share of total reported taxes rose to 74 percent.[2] This trend indicates the growing dependency of our nation's tax system on the tax preparation industry, and it underscores the need for the Internal Revenue Service (IRS) to understand better how commercial tax preparation influences reporting behavior.

This paper is motivated, in part, by a recent Government Accountability Office (GAO) study (GAO, 2006) that examined the return preparation accuracy of chain tax preparers in a large U.S. metropolitan area. In that study, GAO investigators posing as taxpayers submitted 19 fictitious tax returns to different offices of nationwide chain tax preparers and evaluated the accuracy of the completed returns. GAO found that preparers committed numerous errors. While many of these errors had little tax consequence, eight out of 19 returns (42 percent) had a tax discrepancy of at least $1,500 (six returns with excess refunds and two returns with overstatement of tax). GAO also found that tax preparers omitted income information from returns even when the "taxpayer" provided supporting documentation.

The GAO study did not identify a specific cause or set of causes for the high error rate but cited two possible factors. First, in most States, anyone can become a paid tax preparer regardless of education or training. The preparer population includes a diverse group that includes full-time self-employed CPAs and tax attorneys, as well as individuals employed part-time in seasonal positions by chain tax preparation companies. The broad range of experience reflected in this group and the variability in training needed

---

[1] Internal Revenue Service, Taxpayer Usage Study. See http://www.irs.gov/taxstats/article/0,,id=96629,00.html, last accessed on March 26, 2007.

[2] Analysis of Individual Return Transaction File data.

to keep up to date on a Tax Code that changes from year to year could be contributing to the large number of observed errors.

Second, while penalties exist to promote due diligence on the part of tax preparers, not all preparers are held to the same standards. For example, while paid preparers are subject to penalties for various infractions under the Internal Revenue Code (IRC) (e.g., a $1,000 penalty for aiding and abetting understatement of tax liability), only CPAs, attorneys, and enrolled agents are subject to disciplinary proceedings under Circular 230.

A third possible contributing factor not mentioned in the GAO study is a decline in the number of IRS enforcement staff available to monitor compliance by individual taxpayers and tax preparers. Between 1995 and 2005, the number of revenue agents and revenue officers fell from 24,217 to 17,817, a drop of 26 percent. In 2006, the number of full-time IRS enforcement staff rose to 18,524, an increase of 4 percent from the previous year but still well below levels of a decade ago.[3] A reduced IRS enforcement presence may lead some practitioners to conclude that the economic gain associated with obtaining larger tax refunds for clients outweighs the odds of being investigated and fined by IRS.

While ad hoc studies like GAO (2006) are useful exploratory devices, they are not statistically representative and cannot provide information on industrywide trends. Therefore, GAO recommended that IRS conduct research to determine the extent to which paid preparers contribute to inaccuracies on individuals' tax returns. This paper attempts to respond to this request in two ways. First, we present summary measures of return preparation accuracy using population data from the IRS Office of Research's Compliance Data Warehouse (CDW). Second, we develop a computational method for identifying tax practitioners who have a high percentage of returns with one or more errors. The method is demonstrated in a case study of preparers in the State of Connecticut.

We hasten to point out what this study is not. First, this paper does not propose a methodology for identifying intentional versus unintentional errors. We believe such a determination must take into consideration preparer intent which data analysis alone cannot reliably ascertain. Second, the proposed methodology cannot determine if errors are due to practitioner error or other factors (e.g., taxpayers who do not provide information documents). Ultimately, we believe such determinations can be made only after a careful review of quality control procedures used by individual practitioners.

---

[3] *IRS Data Book*, various issues. See http://www.irs.gov/taxstats/article/0,,id=102174,00.html, last accessed on March 26, 2007.

In the next section, we present some measures of tax return preparation accuracy based on analysis of IRS data. In the third section, we describe a methodology for identifying tax practitioners with a high percentage of clients with potential preparation errors, referred to as the discrepancy rate. The methodology relies on bootstrap resampling of network vertices in a two-mode network consisting of practitioners and Zip Codes. Practitioners with discrepancy rates exceeding the one-tailed 95-percent confidence interval are identified as candidates for further analysis. The fourth section demonstrates the proposed methodology in a case study for the State of Connecticut. Finally, the last section summarizes key points.

Preparation Accuracy on Federal Income Tax Returns: What the Data Reveal

In this section, we present summary population measures of return preparation accuracy. Table 1 shows two indicators of return preparation accuracy for taxpayers who self-prepare and for two categories of preparers: those with an employer identification number (EIN) and those preparers with a preparer taxpayer identification numbers (PTIN).[4] The two indicators are: (1) presence of a math error and (2) a nonzero (positive or negative) dollar amount assigned by the IRS's Automated Underreporter (AUR) program.[5] The types of errors covered by the math error and AUR programs are mutually exclusive.

For individual income tax returns received during 2006, 2.7 percent of returns had a math error. Self-prepared returns are more likely to have a math error than returns prepared by practitioners. In 2006, 5.0 percent of self-prepared returns had a math error versus 1.1 percent for all paid preparers. However, the situation is somewhat reversed for misreporting. Approximately 12.2 percent of returns prepared by a tax practitioner with an EIN are identified as potential misreporter cases versus 10.2 percent for self-prepared returns.[6] Because our interest is on practitioner errors, we focus our attention on potential misreporting errors for the remainder of this paper.

[4] Business operators are required to have an EIN if they have employees, operate as a corporation or partnership, or if certain other conditions apply. In general, firms with an EIN are established businesses as opposed to a single individual operating on a part-time basis. In Processing Year (PY) 2006, there were 328,000 unique preparer EINs and 550,000 unique PTINs that were not also associated with an EIN.

[5] Math errors include errors in addition or subtraction, incorrect dependent SSNs, errors in filing status, exemptions, and incomplete schedules and forms. See GAO (2000) and IRS (2003).

[6] Returns flagged by the AUR program are considered potential underreporters until a case review is conducted.

Table 1. Preparation Accuracy of Federal Individual Income Tax Returns, by Preparer Type, PY 2005 and 2006

Preparer	Processing Year	Total Returns (1,000s)	Returns with One or More Math Errors: Number (1,000s)	Returns with One or More Math Errors: Percent	Returns with a Potential AUR Discrepancy: Number (1,000s)	Returns with a Potential AUR Discrepancy: Percent
All Paid Preparers	2006	82,585	944	1.1%	9,816	11.9%
	2005	80,701	1,080	1.3%	9,656	12.0%
Preparers with PTIN	2006	14,600	328	2.2%	1,514	10.4%
	2005	14,632	393	2.7%	1,538	10.5%
Preparers with EIN	2006	67,985	616	0.9%	8,302	12.2%
	2005	66,069	687	1.0%	8,118	12.3%
Self-Preparers	2006	52,710	2,653	5.0%	5,390	10.2%
	2005	53,236	2,630	4.9%	5,380	10.1%
Total	2006	135,295	3,597	2.7%	15,206	11.2%
	2005	133,937	3,710	2.8%	15,036	11.2%

Source: Individual Returns Transaction File and Automated Underreporter File

Table 2 displays the number and percentage of tax returns with potential misreported amounts by size of firm for the two categories of preparers. The main finding in this table is the negative relationship between firm size and percentage of filers with a potential AUR discrepancy. In particular, the percentage of filers with a potential AUR discrepancy is 3 to 4 percentage points higher among firms with fewer than 100 clients compared to firms with more than 5,000 clients.

Table 2. Returns with a Potential AUR Discrepancy, by Type of Preparer and Firm Size, PY 2006

Preparer Type	Firm Size (No. Clients)	Total Returns (1,000s)	Returns with a Potential AUR Discrepancy: Number (1,000s)	Returns with a Potential AUR Discrepancy: Percent
Preparers with PTIN	Under 100	4,527	514	11.4%
	100-499	6,328	651	10.3%
	500-999	2,134	209	9.8%
	1000-4999	1,545	134	8.7%
	Over 5000	67	5	7.5%
Preparers with EIN	Under 100	3,152	452	14.3%
	100-499	16,227	2,257	13.9%
	500-999	12,873	1,657	12.9%
	1000-4999	18,755	2,140	11.4%
	Over 5000	16,978	1,797	10.6%

Source: Individual Returns Transaction File and Automated Underreporter File

Table 3 examines the source of potential misreporting errors by income line item for returns flagged by the AUR program. The line item with the highest overall AUR frequency is wages, salaries, and tips closely followed by State and local income tax refunds, mortgage interest, and withholding. Together, these four line items account for nearly 60 percent of all potential AUR discrepancies in 2006.

Table 3. Number (in 1,000s) of Potential AUR Discrepancies, by Line Item, PY 2006

	Preparer Type			
Line Item	**Taxpayer**	**Preparers with PTIN**	**Preparers with EIN**	**Total**
Wages, Salaries, Tips	813	223	1,401	2,437
Interest	441	145	698	1,284
Dividends	297	89	439	825
State & Local Income Tax Refunds	729	222	1,451	2,402
Capital Gains	159	45	256	460
Rents & Royalties	72	24	129	225
Taxable Pensions	531	118	664	1,313
IRAs	84	23	116	223
Taxable SSI	328	127	507	962
Other Income	238	111	596	945
Mortgage Interest	706	276	1,339	2,321
Withholding	751	197	1,187	2,135
Total	5,149	1,600	8,783	15,532

Note: Returns may have multiple AUR discrepancies. Row total includes only those categories shown.

Source: Automated Underreporter File

Table 4 displays the top ten States with the highest percentage of preparer returns with an AUR discrepancy in PY 2005 and 2006. For the U.S., an average of 12 percent of individual tax returns prepared by a tax practitioner had a potential AUR discrepancy. For the most part, States that ranked in the top 10 in 2005 also ranked highest in 2006. Interestingly, California ranked third in both years. GAO (2006) cites California as one of only two States in the U.S. (the other is Oregon) that require unenrolled preparers to register with State agencies and meet continuing education requirements. Although not shown in Table 4, nearly the same ten States are the top ten States for self-prepared tax returns, although discrepancy rates range, on average, 1-2 percentage points lower than preparer rates. This is consistent with the overall average potential AUR discrepancy rates shown in Table 1.

Table 4. Top 10 States with the Highest Percentage of Preparer Returns with a Potential AUR Discrepancy, PY 2005-06

	PY2005		PY2006	
Rank	State	Preparer Returns with a Potential AUR Discrepancy (%)	State	Preparer Returns with a Potential AUR Discrepancy (%)
1	Nevada	15.0%	Maryland	14.7%
2	Maryland	14.9%	Nevada	14.4%
3	California	14.0%	California	14.2%
4	Arizona	13.6%	Arizona	13.7%
5	Colorado	13.5%	Connecticut	13.5%
6	District of Columbia	13.4%	New Jersey	13.3%
7	New Jersey	13.4%	Colorado	13.2%
8	Georgia	13.2%	District of Columbia	13.1%
9	Connecticut	13.1%	Georgia	13.0%
10	Arkansas	13.0%	South Carolina	12.8%
	U.S. Average	12.0%	U.S. Average	11.9%

Source: Individual Returns Transaction File and Automated Underreporter File

Thus far, we have summarized return preparation accuracy on individual tax returns with respect to the number of returns with math errors or with potential misreported amounts as determined by the IRS's AUR program. We found that paid preparers commit far fewer math errors, both in absolute and relative terms than self-preparers. However, paid preparers account for a higher number and a larger percentage of tax returns with a potential AUR discrepancy. Therefore, we believe the focus on preparer-related errors mainly should be on AUR cases. We now turn to the second aim of this paper: the development of a methodology for identifying individual tax practitioners having high inaccuracy rates.

Methodology for Evaluating Preparation Accuracy of Tax Practitioners

Our aim in this section is to identify practitioners whose observed AUR discrepancy rate exceeds the rate that would be expected if clients were drawn at random from the population of taxpayers who use a preparer. The approach taken is closely related to Snijders and Borgatti (1999). They describe how resampling methods can be used to generate nonparametric statistical measures for one-mode networks. A one-mode network is where all network vertices (nodes) represent one type of object (e.g., persons). Our approach differs from Snijders and Borgatti (1999) in that we employ a two-mode network with separate vertices for tax practitioners and Zip Codes.

Bootstrap Resampling

Bootstrap resampling is only one of a number of techniques (e.g., jackknife, delta) used to evaluate the precision of sample statistics when the underlying distribution is unknown. Originally popularized by Efron (1979), the basic idea of the bootstrap method is that a sample containing N observations contains all of the information of the underlying population. These data are resampled with replacement wherein each artificial sample contains N observations. A sampling distribution is created by drawing multiple samples and computing the statistic of interest (e.g., median, mean) for each sample. Using the sampling distribution data, and assuming approximate normality, one can compute the standard error for the test statistic using conventional methods, such as a t-test.

An alternative approach that does not rely on the normality assumption is to use the bootstrap sampling distribution directly to calculate the probability of obtaining an observed density as large as actually observed assuming the null hypothesis. In this case, we evaluate proportions based on the count of bootstrap samples that have a test statistic larger than the observed value. This is sometimes referred to as the percentile method and is one of several methods for obtaining approximate confidence intervals (Efron, 1981).

In mathematical terms, given a set of observations $X = (X_1 \ldots X_n)$, we can construct a bootstrap sampling distribution (X^*) using the following two-step process:

1. Draw $i_1, \ldots, i_n$ independently from the uniform distribution on $\{1, \ldots, n\}$
2. Set $X^j\ (X^j_{i_1}, X^j_{i_2}, \ldots, X^j_{i_n})$ for $j = 1, \ldots, m$ and $X^* = (X^1 \ldots X^m)$.

The bootstrap sample X^j is constructed by drawing n observations with replacement from the original sample $X_1 \ldots X_n$. In principle, this means a bootstrap sample could consist of the same value repeated n times. However, the probability of this occurring is quite small, as the number of different bootstrap samples available is n^n (Adibi, Cohen, Morrison, 2004). The bootstrap principle assumes $X = (X_1 \ldots X_n)$ is a random sample from a distribution $\boldsymbol{P}$, and the sample statistic $\hat{\theta} = s(X)$ is an estimation for the population parameter θ. Finally, $\hat{\theta}^* = s(X^*)$ is the bootstrap replication of θ.

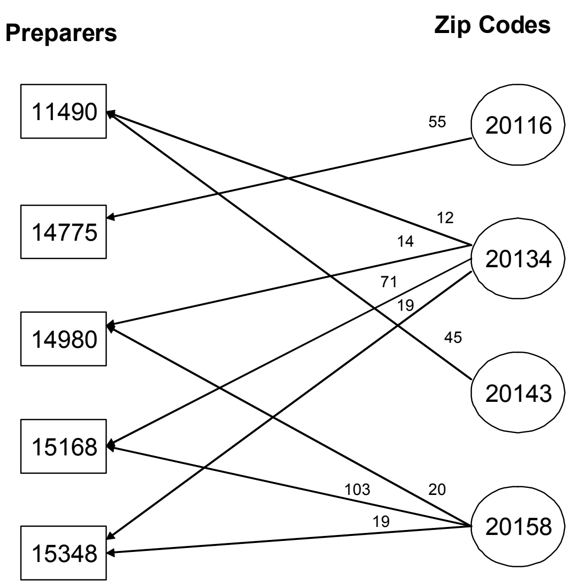
Sample Size

Efron (1979) points out that the bootstrap method correctly estimates (asymptotically) the value of a known population parameter. The minimum number of samples required to obtain reasonably accurate estimates depends on the parameter of interest. Efron and Tibshirani (1986) show a sample size of 100 is adequate to compute a coefficient of variation. However, they recommend a minimum sample size of 1,000 to compute nonparametric confidence intervals. This makes intuitive sense because confidence intervals typically are at the extremes of the distribution so that a large number of generated values are needed to adequately characterize the tail region.

Bootstrap Procedure

To perform the bootstrap procedure, a two-mode network is constructed that consists of tax practitioners and taxpayers aggregated by five-digit Zip Code. Network links represent the number of clients from each Zip Code using the services of different tax preparers. Figure 1 displays a hypothetical two-mode network consisting of five preparers (with hypothetical identifiers) and four Zip Codes.

Figure 1. Bipartite Graph of Network of Five Tax Preparers and Four Zip Codes



For each preparer in our sample, we compute the mean expected AUR discrepancy rate and the 95-percent confidence interval. To do this, we generate 1,000 bootstrap samples with each sample drawn from Zip Codes in the same proportion as the preparer's clientele. This approach, known as stratified bootstrap resampling, reduces the probability of obtaining biased estimates when bootstrap samples are generated from data not in the original sample. For example, referring to preparer 11490 in Figure 1, a single sample consists of 12 observations from Zip Code 20134 and 45 observations from Zip Code 20143. For each observation, a uniform random number ($0 \leq u < 1$) is generated. If the value of u is less than or equal to the Zip Code AUR discrepancy rate (D_{zip}), then we assign an AUR discrepancy case to the preparer; otherwise, we assume that taxpayer does not have an AUR discrepancy. For sample j, we compute the mean discrepancy rate for preparer k ($\overline{D_k^j}$) as:

$$\overline{D_k^j} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{where } x_i \in \{0,1\}$$

This procedure is repeated 1,000 times, and the bootstrap replication of the population mean for each preparer *k* (i.e., the expected discrepancy rate) is calculated as follows:

$$\hat{\theta}_k^* = \frac{1}{N}\sum_{j=1}^{N} \overline{D_k^j} \qquad \text{where } N = 1{,}000$$

The one-tailed 95-percent confidence interval is obtained by sorting the 1,000 observations in ascending order and selecting the cutoff as the value of the 950th observation.

Bootstrap versus Population Measures

An alternative to the bootstrap method is to compare preparer discrepancy rates to the Zip-weighted population average for all Zip Codes in a preparer's market area and select the preparer if it exceeds the average. While this approach is somewhat simpler to implement from a computational perspective than bootstrapping, it also likely will identify many more preparers with "significant" discrepancy rates than an approach based on selecting preparers who exceed the 95-percent confidence interval. In addition, the bootstrap method offers greater flexibility should researchers wish to use a different cutoff value, say the 99-percent confidence interval, in order to further isolate preparers with the most extreme discrepancy rates. Therefore, while population-based measures may be simpler to implement computationally, we believe bootstrap resampling offers greater flexibility by enabling the researcher to specify alternative cutoff values (confidence intervals) for identifying preparers.

Case Study

In this section, we demonstrate our methodology using PY 2006 tax return data for the State of Connecticut. Our metric of preparation accuracy is the fraction of tax returns with a potential AUR discrepancy, otherwise referred to as the discrepancy rate.[7] For this demonstration, we selected practitioners with 100 or more clients and at least 20 potential AUR cases. Our final sample included 1,178 preparers (1,014 with an EIN and 164 with a PTIN) who collectively had over 730,000 clients who filed Federal tax returns in the State of Connecticut in 2006. The median firm has 339 clients and a market area comprised of 54 Zip Codes. The average maximum number of clients from any one Zip Code was 23 percent, indicating that most preparers' clients do not reside in a single Zip Code. The maximum preparer discrepancy rate is 54 percent, and the minimum is 4 percent. Finally, the median discrepancy rate for preparers in our sample is 14 percent compared to a State average of 13.5 percent (see Table 4).

Our data also include 519 Zip Codes ranging in size from 1 to 29,344 filers. The median Zip Code has 446 filers. The market area for the largest preparer (with respect to number of filers) includes 377 Zip Codes, and the smallest market area has 10 Zip Codes. Zip Code discrepancy rates are calculated from tax returns filed by individuals located in each Zip Code without regard to preparer used.[8] The maximum Zip Code discrepancy rate is 100 percent, the minimum is 0 percent, and the median is 11 percent.

The study data are from the Entity and the AUR databases on the IRS Compliance Data Warehouse (CDW). These two data sources separately provide the total number of returns filed and the number of AUR returns by preparer and by Zip Code. The data were formatted into a comma-separated flat file with each line (record) of the file representing one practitioner. The field layout for each record is as follows:

[7] Our discrepancy rate measure includes both positive and negative discrepancies. Although one could select only discrepancies in one direction (e.g., underreporting), the intended aim here is to reduce all errors regardless of source. Therefore, we include potential discrepancies from both overreporting and underreporting.

[8] However, the Zip Code discrepancy rate is calculated only from filers who used a paid preparer.

Field 1: practitioner identifier

Field 2: practitioner discrepancy rate (with two implied decimal places)

Field 3: number of Zip Codes in the practitioner's market area

Field 4: first Zip Code identifier

Field 5: number of this practitioner's clients from this Zip Code

Field 6: Zip Code discrepancy rate (with two implied decimal places)

Fields 4-6 are repeated for each Zip Code in this practitioner's market area.

The input data file is processed by a program written in Java that performs the bootstrap resampling and computes summary statistics including mean, standard deviation, and the 95-percent confidence interval.

Results

We ran the bootstrap procedure on the 1,178 preparers in our dataset and sorted the output in descending order by number of excess AUR discrepancy cases for preparers exceeding the 95-percent confidence interval threshold. The number of excess cases is determined by multiplying each preparer's number of clients by the difference between the observed and expected preparer discrepancy rates. We summarize results for the 50 preparers with the largest number of excess cases in Table 5. Among these 50 firms, the average (mean) number of clients was 957, the smallest firm had fewer than 150 clients, and the largest firm had roughly 3,000 clients.

Table 5 shows the 10 firms with the largest number of excess AUR discrepancy cases accounted for 2.6 percent of Connecticut filers with potential misreporting in 2006, as well as 4.7 percent of the approximately $5.8 billion in net potential underreported amount.[9] The top 50 firms accounted for 8.4 percent of potential AUR discrepancy cases and 11.4 percent of the net potential underreported amount. These results show that a small number of firms accounts for a significant percentage of taxpayers with a potential AUR discrepancy and a larger share of potential underreporting.

[9] There were 39,025 unique preparer identification numbers (EINs and PTINs) in Connecticut in PY 2006.

Table 5. Bootstrap Program Results for Connecticut Preparers, PY 2006

Preparers*	Returns with Potential Misreporting		Net Potential Underreported Amount	
	Number	Percent	($M)	Percent
Top 10 Firms	3,539	2.6%	$271.6	4.7%
Top 20 Firms	6,373	4.7%	$384.1	6.6%
Top 30 Firms	8,242	6.1%	$543.4	9.3%
Top 40 Firms	9,863	7.3%	$625.6	10.8%
Top 50 Firms	11,346	8.4%	$665.3	11.4%
Total	135,878	100.0%	$5,816.5	100.0%

*Ranked in descending order by number of “excess” filers with potential misreporting.

Benefits and Limitations of Bootstrap Methodology

The primary benefit of this methodology is its low cost and ease of implementation. The required data are available on existing IRS databases and can be extracted for the entire country with only a few lines of SQL code. The algorithm used to identify individual preparers is data-driven and is applicable for most practitioners. However, because this method relies on Zip Code data to generate bootstrap sampling distributions, the analyst must take care to ensure a reasonable degree of independence between preparer and Zip Code observations. This condition is more likely to be met for midsize to larger firms. Even so, the proposed methodology should be thought of primarily as a screening tool and not as a technique for carrying out tests of statistical significance.

Summary and Conclusion

This paper investigated the extent to which the commercial tax preparation industry contributes to the number of inaccurately prepared returns. Such inaccuracies may negatively impact both the IRS and taxpayers through increased administrative costs, greater taxpayer burden, and possibly reducing the level of voluntary compliance. We presented different measures for common types of errors encountered on individual tax returns including math errors and potential misreporting. We used these measures to describe aspects of return preparation accuracy for self-preparers and two different categories of tax practitioners: those with an EIN and those with a PTIN only. Finally, we proposed a bootstrap resampling methodology to identify individual preparers with high potential AUR discrepancy rates and demonstrated its use in a case study of preparers in the State of Connecticut.

Our data analysis found math errors were committed more frequently by self-preparers, but clients of paid preparers had a higher incidence of potential misreporting. There is a negative relationship between firm size and incidence of taxpayers with potential misreporting. The line items misreported most frequently include: wages, salaries, and tips, State and local income tax refunds, mortgage interest, and withholding.

The case study of Connecticut preparers found that a significant percentage of potential AUR cases, as well as the associated potential net underreported amount, can be attributed to a small number of preparers. Given this finding, we believe that a substantial reduction in the number of AUR discrepancies could be achieved by annually monitoring tax practitioners using data-driven techniques like those proposed in this paper in combination with a program of outreach and education to the selected preparers. Assuming the program focused only on the top 2,000 tax practitioners nationwide (approximately two-tenths of 1 percent) with respect to "excess" number of AUR cases, we estimate such a program potentially could reach preparers who annually are responsible for over 715,000 potential AUR cases and $30 billion in net potential underreported income.[10]

Acknowledgments

The authors wish to thank Jim Bullis, Janet McCubbin, Mary-Helen Risler, and Professor John Scholz for comments on an earlier draft. However, the authors are responsible for any remaining errors or misinterpretations of data.

References

Adibi, J.; P. R. Cohen; and C. T. Morrison, Measuring Confidence Intervals in Link Discovery: A Bootstrap Approach, paper presented at the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, 22-25 August 2004.

Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7(1), pp. 1-26.

Efron, B. (1981), Nonparametric Standard Errors and Confidence Intervals, *The Canadian Journal of Statistics*, 9(2), pp. 139-158.

[10] This estimate assumes 7.3 percent of 9.8 million potential preparer AUR cases and 10.8 percent of $281 billion in net potential underreported income on preparer returns in PY 2006.

Efron, B., and R. Tibshirani, (1986), Bootstrap Methods for Standard Errors, Confidence Errors, and Other Measures of Statistical Accuracy, *Statistical Science*, 1(1), pp. 54-75.

Internal Revenue Service, Report to Congress: *IRS Tax Compliance Activities*, Department of the Treasury, Internal Revenue Service, July 15, 2003.

Snijders, T. A. B., and S. P. Borgatti, (1999), Non-Parametric Standard Errors and Tests for Network Statistics, *Connections*, 22(2), pp. 161-170.

U.S. General Accountability Office (March 2000), *Tax Administration: IRS Use of Nonaudit Contacts*, GAO Publication GAO/GGD-00-7, Washington, DC.

U.S. General Accountability Office (2006), *Paid Tax Return Preparers: In a Limited Study, Chain Preparers Made Serious Errors*, GAO Publication GAO-06-563T, Washington, DC.