Closing the Gap: Automated Screening of Tax Returns to Identify Egregious Tax Shelters

Dave DeBarr
The MITRE Corporation
7515 Colshire Dr
McLean, VA 22102
debarr@mitre.org

Zach Eyler-Walker
The MITRE Corporation
7515 Colshire Dr
McLean, VA 22102
zach@mitre.org

ABSTRACT

According to the most recent strategic plan for the United States Internal Revenue Service (IRS), high-income individuals are a primary contributor to the "tax gap," the difference between the amount of tax that should be collected and the amount of tax that actually is collected [1]. This case study addresses the use of machine learning and statistical analysis for the purpose of helping the IRS target high-income individuals engaging in abusive tax shelters. Kernel-based analysis of known abuse allows targeting individual taxpayers, while associative analysis allows targeting groups of taxpayers who appear to be participating in a tax shelter being promoted by a common financial advisor. Unlike many KDD applications that focus on classification or density estimation, this analysis task requires estimating risk, a weighted combination of both the likelihood of abuse and the potential revenue losses.

Keywords

Abusive Tax Shelter, Data Mining, Support Vector Machine, Link Analysis.

1. INTRODUCTION

This case study focuses on the use of data analysis techniques to identify egregious tax shelters provided by "pass-through" entities to high-income taxpayers. Trusts, partnerships and subchapter S corporations are referred to as "pass-through" entities because tax liabilities for their income are simply passed to their beneficiaries, partners or shareholders respectively. The allocation of gains and losses from a pass-through entity is recorded for each payee using Schedule K-1 [2,3,4].

Here is a brief characterization of the data available for each type of entity:

- For the purposes of this study, a high-income taxpayer is an individual who reports an annual income of $250 thousand or more. For tax year 2003, 1.9 million high-income returns were filed. The IRS maintains over 1,000 variables to describe each of these returns.
- A trust is a financial entity established to allow a trustee to manage property on behalf of another party, called the beneficiary. The most common type of trust is a grantor trust, in which income of the trust is taxed as income of the grantor. For tax year 2003, 3.5 million trust returns were filed with 4.4 million schedule K-1 records. The IRS maintains over 200 variables to describe each of these returns.
- A partnership is a business in which partners share the gains and losses from operating the business. The most common type of partnership involves leasing real estate property. For tax year 2003, 2.5 million partnership returns were filed with 14.5 million schedule K-1 records. The IRS maintains over 100 variables to describe each of these returns.
- A subchapter S corporation, hereafter referred to simply as an S corporation, is an incorporated business that meets the requirements of subchapter S of the "normal" income taxes chapter of the Internal Revenue Code [5]. For tax year 2003, 3.4 million S corporation returns were filed with 5.9 million schedule K-1 records. The IRS maintains over 100 variables to describe each of these returns.

The IRS had stopped transcribing the schedule K-1 pass-through allocations in 1995, but this practice was resumed for tax year 2000 (calendar year 2001) [6]. The MITRE Corporation was asked to investigate possible analysis methods for exploiting the information about relationships between taxpayers and these pass-through entities. The major lines of investigation included visualization of the relationships and data mining to identify and rank possibly abusive tax avoidance transactions.

The remainder of this paper is organized as follows. Section 2 discusses the use of visualization for reviewing taxpayer relationships. Section 3 addresses targeting compliance issues for high-income individuals, while section 4 addresses targeting related compliance issues among groups of high-income individuals. Section 5 covers reduction of search complexity by pruning irrelevant links. Section 6 covers filtering groups by measuring compliance risk and merging related groups. Section 7 discusses results, and section 8 presents conclusions and future work.

2. VISUALIZATION

The visualization of the relationships between trusts, partnerships, S corporations and taxpayers included both direct payer to payee relationships and indirect payer to payee relationships; e.g. linking a spouse to a primary filer, a sole proprietorship to an owner or a subsidiary to a parent corporation. Compared to having to repeatedly query a database for linked entities or having to manually switch back and forth between paper returns, this was considered a big improvement. The IRS has subsequently instantiated a prototype visualization system for use by both researchers investigating trends and compliance staff reviewing tax returns. This system now has over 200 user accounts.

Figure 1 illustrates an example of the relationships between a high-income taxpayer and his pass-through entities using a graph with directed edges. All sensitive labels have been removed from the graphs in this paper, but in use the nodes are labeled with name and Taxpayer Identification Number (TIN—either an Employer Identification Number or a Social Security Number), and the edges are labeled with the dollar amounts for gains and losses. A diamond represents a trust, an oval represents a partnership and a rectangle represents an S corporation. The rounded rectangles in the bottom-left are the taxpayer (bold border) and his spouse. The parallelogram in the middle is the taxpayer's sole proprietorship. The octagons indicate the presence of additional payees for three of the partnerships. A user can click on nodes or links to review the transcribed line items for the associated entity or relationship. Colors are used to indicate the presence of various attributes; e.g. red links indicate a net loss and black links indicate a net gain for payer to payee links. The width of a link indicates the amount of money being allocated to a payee; i.e. a thicker link indicates a larger magnitude for money.

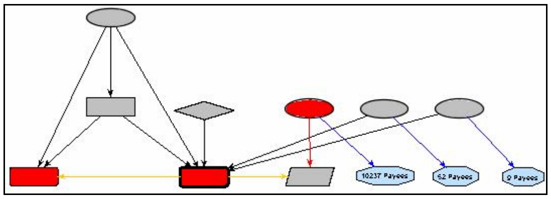

Figure 1 Visualization of a Taxpayer's Investments

Figure 2 illustrates an example of a prototypical tax shelter with all other relationships for the taxpayer "hidden." This shelter is described in IRS Notice 2000-44 [7] and is commonly referred to as a "Son of BOSS (Bond and Option Sales Strategies)" shelter.

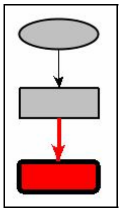

Figure 2 Abusive Son-of-BOSS Tax Shelter

The following description sketches out the salient points of the Son-of-BOSS shelter; however, it is not intended to be a technically accurate description of all related shelter activities. The taxpayer uses a "straddle" to effectively shelter their income:

1. The taxpayer obtains a large $X million gain, often associated with the sale of some asset such as a business.
2. A tax advisor tells the taxpayer he can avoid paying tax on the large gain by exploiting a "loophole" in the tax law. Instead of paying 15-30% tax on the gain, he only needs to pay a smaller fee to the tax advisor.
3. The promoter (tax advisor) sets up a partnership for financial investments, often including himself as a tax matters partner; i.e. the partner who handles tax matters for the partnership (not shown in figure 2).
4. The taxpayer buys call options for $X million; i.e. the option to purchase stock from someone.
5. The taxpayer transfers these call options to the partnership.
6. The taxpayer then sells call options for $X million to someone else; i.e. the option to purchase stock from the taxpayer.
7. The taxpayer ignores the liability of the underwritten call options because the tax advisor claims this is allowed. The fundamental equation of accounting says ASSETS = LIABILITIES + OWNER_EQUITY; so the taxpayer is claiming an inflated value for owner equity.
8. Upon sale of the call options, the taxpayer claims an $X million loss to offset his income from the large gain; i.e. a breakeven transaction is used to shelter millions of dollars of income from taxes.

The S corporation in figure 2 is being used to facilitate the loss for the taxpayer.

While legitimate tax shelters do exist, such as depreciation claimed for investment in residential property that houses low-income tenants, a loss is generally not allowed unless it results in an actual loss for the taxpayer [8].

3. MODELING SHELTER RISK FOR HIGH-INCOME INDIVIDUALS

While electronic filing is becoming more popular, the majority of the tax forms submitted by those involved with abusive tax shelters are not filed electronically. Therefore, most are manually transcribed. If all line items for every return were accurately recorded and available electronically, it would make the job of identifying potentially abusive transactions much easier. Unfortunately, only a subset of the line items are actually transcribed and available in electronic form; and the values that are available are sometimes questionable [9].

We began the modeling process by working with an IRS technical advisor to identify the type of behavior the IRS is interested in identifying. The two principal methods for abusively sheltering income from taxes include manufacturing offsetting losses [without any real loss for the taxpayer] and not reporting income. Identifying abusive offsetting losses is considered to be lower hanging fruit, however, because taxpayers are encouraged to actually record their transactions accurately. There is a three year statute of limitations for shelters that are reported, while there is no statute of limitations for income that goes unreported. Additionally, possible fines and penalties are much stronger for unreported income [10]. While MITRE has done some work in the area of identifying unreported income, here we report our work dealing with offsetting losses.

The existing system used by the IRS for targeting compliance issues is constructed by analyzing audit results for a set of randomly selected tax returns. Unfortunately, since truly egregious tax shelters are relatively rare (currently believed to occur in about 1% of the high-income taxpayer population), random selection of audits is unlikely to capture egregious transactions. This explains why some truly egregious shelters may receive a low score. In the future, weighted sampling might be used to improve coverage in this relatively rare portion of the population; e.g. computing the selection probability based on the proportion of total positive income being reported as taxable income. For the majority of high-income taxpayers, this proportion is substantial. As illustrated in figure 3, the mode of the distribution occurs around the 88th percentile. In the mean time, we explored the use of kernel-based techniques as an alternative for initial ranking of tax returns for review by a compliance expert.

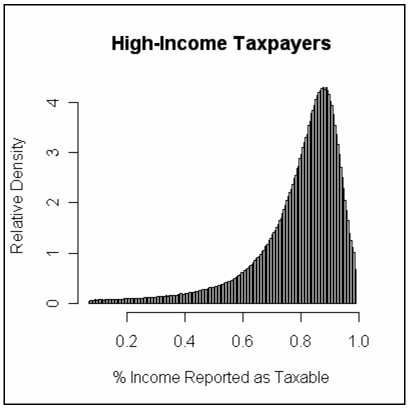

Figure 3 Taxable Income

Given a database with a few known examples of "abusive" transactions and no other "labels," we pursued constructing a single-class model to measure the similarity of high-income taxpayer relationships to known examples of abuse. We started with a half-dozen examples of abuse provided by the technical advisor, but discovered there was no data connecting the pass-through entities to the high-income taxpayers. We compensated for the lack of training examples by asking the technical advisor to describe the behavior of interest and querying the database to find matching examples. More than 50 variables were being considered for each set of entities connected to a high-income taxpayer. The query output was processed slowly, as there was a significant learning curve for the data miners trying to understand how the subset of transcribed line items from different tax returns related to one another. Note: Because the IRS wants to encourage electronic filing, the IRS does not use untranscribed line items from electronic returns for targeting compliance issues.

Our initial query involved searching for one to four high-income taxpayers receiving little income from an initial year partnership and large losses from an initial year S corporation. While the first couple of matches proved we needed to refine the targeting criteria, later examples indicated the existence of multi-million dollar shelters that had been previously undiscovered. An existing compliance project was used to support initiating audits on the discovered shelters.

After obtaining 30 examples, we constructed a single-class model using a Support Vector Machine (SVM) to produce a similarity measure for weighting the loss in question. Training the single-class SVM [11] was described to the IRS domain experts as being similar to how you might train a revenue agent. First, a few positive examples are provided in terms of the features relevant to identifying a potentially abusive transaction. The "trainee" is then asked to evaluate a new set of returns to identify similar behavior. In this case, the "trainee" is also expected to identify the source of the suspicious loss.

Somewhat redundant features were used to provide robustness against transcription errors; e.g. using line items describing short term capital losses from both schedule K-1 of the pass-through entity and the high-income taxpayer's return. Instead of having the single-class SVM produce a TRUE/FALSE class label, however, we used the raw sum of the kernel (similarity) function output to compute a risk metric for ranking; i.e. we used a Gaussian kernel to compute the similarity between each transaction and the optimal prototypes from the known abusive transactions (training data). The "nu" parameter of the SVM was used to allow a small subset of examples to be declared to be outliers, while the "gamma" parameter of the kernel was used to allow selection of the optimal prototypes (support vectors) by favoring less complex models providing the best coverage of the known abusive examples. To find the best hyper-parameter values, a hierarchical grid search was conducted over the range of feasible "nu" and "gamma" values using leave one out cross validation.

It takes only a few minutes to construct the model, and it takes only an hour to assess risk for a year of tax return data. The longest part of the process involves preprocessing the data by deriving features from the line items of the returns and normalizing the feature vectors to unit length (so a $1 million dollar offset of a $1 million dollar income is given the same weight as a $100 million dollar offset of a $100 million dollar income, and the resulting weights are then multiplied by the magnitude of the sheltered income to assess overall risk).

This model was successful for identifying and ranking a few specific types of transactions, revealing an estimated $200 million dollars of previously undiscovered shelters. This model was also useful for providing coverage of substantially similar transactions. Nevertheless, while precision for this model was around 90%, the recall was suboptimal. Transactions were being missed due to transcription issues and the use of similar transactions with different types of assets; e.g. foreign currency straddles characterized as "ordinary loss" instead of stock option straddles characterized as "short term capital loss".

Based on feedback from the domain experts, we decided to generalize the targeting strategy by relaxing the targeting criteria to review a smaller, more general set of targeting features; e.g. total positive income, largest gain, largest loss, taxable income, etc. This simplistic model is known as the Shelter Risk Function (SRF). The values associated with a transaction for a high-income taxpayer are compared to an idealized shelter using a Gaussian kernel. The kernel width was selected using a jackknife procedure [13] to identify the value that produced the highest correlation to audit results from a prior tax year.

For our initial evaluation, an idealized shelter is characterized as a single source of income being offset by a single source of loss, resulting in zero taxable income. Again, somewhat redundant features are employed to provide robustness against transcription errors. The similarity of data describing a taxpayer and an idealized shelter is used as a weight for the income that is being sheltered. The results of the risk assessments for this model are then fed to an associative analysis engine to identify groups of related shelter suspects.

4. MODELING SHELTER RISK FOR GROUPS OF HIGH-INCOME INDIVIDUALS

While SRF can provide a reasonable targeting metric for identifying potentially abusive shelter activity, it does not attempt to identify common links between shelters. Some shelters are customized for an individual and are not tightly linked to additional shelters. Other shelters, however, share a common structure and mode of operation, which having been designed once can be sold to different clients over and over by a shelter promoter [14].

Figure 4 illustrates an example of a group of approximately 40 related shelters. On the far left is the promoter that created the tax shelter, and on the far right is a pair of entities that "sell" this tax shelter to taxpayers. The outer ring of the picture is a set of high-income individuals who are sheltering their income. The ring in the middle is a set of partnerships manufacturing abusive tax shelters using straddles. The two entities in the center are foreign partners, tax indifferent parties claiming the allocation of gain from the straddles. This promotion was used to shelter hundreds of millions of dollars for high-income taxpayers.

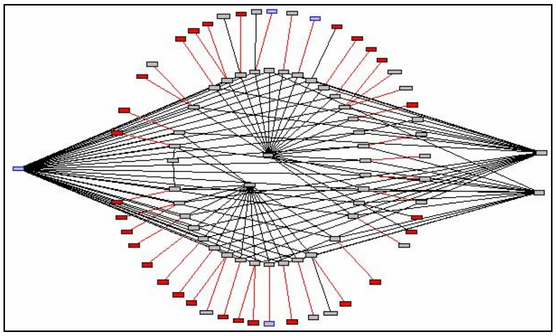

Figure 4 A Group of Related Tax Shelters

These “cookie cutter” shelter promotions are particularly interesting to the IRS because they make it necessary to argue only one instance of the shelter in court. If the IRS wins the case, or a few of these cases, it becomes more likely that the other instantiations will not go before the court, and further litigation costs are saved. When it is considered that some abusive shelters are worth tens of millions of dollars or more to their recipients, it is not surprising that they are defended vigorously. Detecting promotions rather than just single tax shelters can thus be highly advantageous for the IRS in terms of reduced cost of enforcement.

We developed the Promoter Risk Function (PRF) to be used in conjunction with SRF with the goal of grouping SRFs and other high income individuals together with businesses and individuals potentially promoting abusive tax shelters. PRF is a custom link analysis application that explores the relationship between groups of SRF suspects and various identifiers of the potential promoter nodes, including their names, addresses and Taxpayer Identification Numbers (TINs). Preparer information is included in the attributes being examined.

Promotions such as the one shown in figure 4 helped to suggest our approach. Each participant in this promotion receives his own shelter from an individual partnership, but those partnerships are in turn all connected to just a handful of promoter entities. Given a subset of this shelter’s participants, it is possible to traverse their K-1 links to discover potential promoter entities. By reversing this process—expanding the search outward from the newly discovered promoter suspects—not only will the original suspects be reached, but the other individuals associated with the promotion as well. In this way we can group a set of suspects into various suspected shelter promotions and also discover previously unsuspected shelter participants.

We generally use SRF as the starting point for promotion detection with PRF, although other targeting functions can be used as well. PRF starts with a specific target group (hereafter referred to as suspects), but as described above it also discovers other high income SSNs that are associated with suspected promoter attributes. This is beneficial in that not only does it yield better recall of promotion participants, but it also allows PRF to judge the likelihood of particular groupings of suspects and non-suspects, as discussed below.

5. PRUNING IRRELEVANT LINKS

The most naive implementations of PRF will not run to completion in a reasonable amount of time, due to combinatorial explosion. Some nodes may have as many as hundreds of thousands of connections to others. Traversing these nodes even once is too much, and caching such results is not feasible with even 4 gigabytes of system memory available. In order to reduce the run time and storage requirements, we implemented some simple pruning heuristics.

Connections between nodes were cut when the number of investors or investments was greater than threshold, unless the link represented more than 10% of the equity for that node. For the investments threshold, Chebyshev's inequality [15] was used with k = sqrt(20) to identify inbound links where the payee has an unusually large number of payers. For the investors threshold, a domain expert defined rule was used to identify outbound links in which the payer has more than 10 payees. This reduced the number of possible links to be analyzed from 24.7 million links to 16.8 million links.

Additionally, we limit depth of search to κ hops from the starting points, the input suspects. Unlike the pruning heuristics, this heavily affects the behavior of PRF in both positive and negative ways. The limited search could obscure more complicated shelter promotion schemes in which invariant promoter attributes are always more than κ hops away from the suspects. However, there is a hidden advantage here: by limiting the search depth, spurious connections between suspects are encountered less often, helping to limit false positives. A final step was to restructure the database to reduce the number of disk reads required to process the links.

These measures reduced the execution time by orders of magnitude so that the PRF tools can complete a search of a given tax year in approximately three to four hours on our hardware (a 2.8 gigahertz Intel Xeon processor with 4 gigabytes of memory).

6. FILTERING AND MERGING GROUPS

After the database has been searched for links, we take additional steps to filter and group the data. The most important filtering stage is to threshold potential promoter identifiers based on their level of support in the input group. Those identifiers linked to smaller numbers of suspects are less likely to be part of a promotion. In the degenerate case, when there is only a single suspect linked to an identifier, the identifier has no resolving power in the discovery of a shelter promotion.

If the support threshold is met for a given potential promoter, we generate the odds ratio of the number of suspects to non-suspects associated with the potential promoter, compared to the ratio of the total number of other suspects to the total number of other non-suspects in the population. The greater the odds ratio, the less likely a group with the same number of suspects and non-suspects would be generated from a random sampling of the population. While a p-value from a Chi-square or Fisher test [16] can be used to evaluate the hypothesis that P (Suspect = True | Link to X) ≤ P(Suspect = True | No Link to X), a p-value is not so useful for measuring the degree of association between the conditions "Suspect = True" and "Link to X".

Consider the contrast provided by contingency tables 1 and 2. The one-sided Fisher test p-value for table 1 is 5.7*10-183, while the p-value for table 2 is 0.1*10-183. Yet the odds ratio for table 1 is about 1868, while the odds ratio for table 2 is only 25. A smaller p-value indicates the null hypothesis is less likely to be true, but a larger odds ratio indicates a stronger association between the two factors. Intuitively, the odds ratio result makes more sense because 95% of the returns associated with preparer A are associated with shelter suspects, while less than 20% of the returns associated with preparer B are associated with shelter suspects.

		Shelter Suspect?	
		Yes	No
Prepared by A?	Yes	95	5
	No	19,905	1,979,995

Table 1: Contingency Table for Preparer A

		Shelter Suspect?	
		Yes	No
Prepared by B?	Yes	197	803
	No	19,803	1,979,197

Table 2: Contingency Table for Preparer B

A final grouping stage is run to associate promoter identifiers with other promoter identifiers by thresholding on the Jaccard similarity [17] between the identifiers' suspect groups. When the similarity between the suspects is sufficiently high, we consider the two groups to be part of the same potential promotion and to form part of the same metagroup.

7. RESULTS

With typical threshold parameter settings, PRF finds on the order of 500 metagroups of potential promoters and SSNs for every year, which combine to total a few billion dollars of sheltered income. Around 50% of this total is associated with the top 20 metagroups, as ranked by a combination of the lower bound of a confidence interval for the odds ratio [16] of the metagroup and the amount of income suspected of being sheltered.

When comparing these top 20 metagroups to a list of known shelter participants from the IRS, there is a substantial overlap between the taxpayers believed to have participated in an abusive tax shelter. Examining the PRF results more closely, there appear to be several advantages PRF can provide in addition to its goal of appropriately grouping SRF suspects with potential promoters.

First, it is able to use the initial suspect list to discover other high income individuals that, while not passing the SRF threshold, do appear to be participating in abusive shelters. This is borne out in the high proportion of non-SRFs in the PRF output that overlap the IRS's list of known shelter participants.

Beyond that, PRF is also able to automatically discover shelter participants unknown to the IRS. PRF also maintains the links associating nodes in its output clusters, reducing the effort required to verify whether a suspected participant in a promotion really is taking the shelter.

Perhaps the biggest advantage to PRF is its ease of use and relative efficiency. One of the largest promotions for tax year 2001 required roughly two weeks of work by multiple IRS agents to trace out. PRF is able to do much of that work in just a few hours, with no human intervention. Further it discovers participants that the IRS may not have otherwise found. Running PRF as soon as the data for each tax year is received has the potential to find the most egregious tax shelters efficiently and quickly, with minimal auditor labor.

8. CONCLUSIONS

Recent efforts to combat abusive tax shelters have met with some success [18]. While MITRE is certainly not solely responsible for this outcome, we did play a role in helping to identify abusive shelters. After review of the output by domain experts, audits for selected cases resulted in substantial assessments for additional tax collection by the IRS. Nevertheless, more work remains to be done, such as adjusting the risk values by accounting for differences in tax rates; e.g. the most widely paid tax rate for capital gains is 15%, while the tax rate for ordinary income is often more than 30%. Further work is also needed in assessing risk for unreported income, as well as understanding how abusive promotions change over time.

9. ACKNOWLEDGEMENTS

A number of people provided critical support for this work. The IRS compliance experts included: Mike Whalin, Susan Kerrick and Scott Emerson. The IRS research experts included: Mark Mazur, Larry May, John Davidson, Kay Wolman, Maury Harwood, Jim Needham and Christine Smith. Our colleagues at MITRE included: Anne Cady, Dan Venese, Lowell Rosen and Sarah Piekut. This work was performed under IRS contract number TIRNO-99-D-00005.

10. REFERENCES

[1] IRS Publication 3744, "IRS Strategic Plan: 2005 - 2009", Rev 6-2004, p.18, http://www.irs.gov/pub/irs-utl/strategic_plan_05-09.pdf

[2] IRS Form 1041 Schedule K-1, "Beneficiary's Share of Income, Deductions, Credits, etc." 2003, http://www.irs.gov /pub/irs-prior/f1041sk1--2003.pdf

[3] IRS Form 1065 Schedule K-1, "Partner's Share of Income, Deductions, Credits, etc." 2003, http://www.irs.gov/pub/irs-prior/f1065sk1--2003.pdf

[4] IRS Form 1120S Schedule K-1, "Shareholder's Share of Income, Deductions, Credits, etc." 2003, http://www.irs.gov /pub/irs-prior/f1120ssk--2003.pdf

[5] United States Code Title 26 Section 1362, "Subchapter S - Tax Treatment of S Corporations and their Shareholders", 1982, http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi ?dbname=browse_usc&docid=Cite:+26USC1362

[6] GAO Report GAO-02-618T, "Enhanced Efforts to Combat Abusive Tax Schemes - Challenges Remain", Apr 2002, p.10, http://www.gao.gov/new.items/d02618t.pdf

[7] IRS Notice 2000-44, "Inflated Partnership Basis Transactions (Son of BOSS)", Sep 2000, http://www.irs.gov /pub/irs-utl/notice_2000-44.pdf

[8] "How can I recognize an abusive tax shelter?", Frequently Asked Tax Questions and Answers, http://www.irs.gov /faqs/faq-kw195.html

[9] Government Accounting Office Report GAO-04-1040, "IRS Should Take Steps to Improve the Accuracy of Schedule K-1 Data", Sep 2004, pp.3-4, http://www.gao.gov /new.items/d041040.pdf

[10] United States Code Title 26 Section 6663, "Imposition of Fraud Penalty", 1989, http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi ?dbname=browse_usc&docid=Cite:+26USC6663

[11] B. Scholkopf, J. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson, "Estimating the Support of a High-Dimensional Distribution", Neural Computation, 13, Jul 2001, pp.1443-1471.

[12] IRS Notice 2002-65, "Passthrough Entity Straddle Shelters", Sep 2002, http://www.irs.gov/pub/irs-utl/notice_2002_65_(l21).pdf

[13] B. Efron and R.J. Tibshirani, "An Introduction to the Bootstrap", May 1994, pp.141-150.

[14] Senate Report 109-54, "The Role of Professional Firms in the U.S. Tax Shelter Industry", Apr 2005, http://frwebgate.access.gpo.gov/cgi-bin/useftp.cgi ?IPaddress=162.140.64.88&filename=sr054.pdf &directory=/diskb/wais/data/109_cong_reports

[15] S. Ghahramani, "Fundamentals of Probability", Prentice Hall, 2nd ed, Sep 1999, pp.437-441.

[16] A. Agresti, "Categorical Data Analysis", Wiley-Interscience, 2nd ed, Jul 2002, pp.91-101.

[17] D.J. Hand, H. Mannila, and P. Smyth, "Principles of Data Mining", The MIT Press, Aug 2001, p.37.

[18] IRS Newswire IR-2004-87, "Strong Response to 'Son of BOSS' Settlement Initiative", Jul 2004, http://www.irs.gov /newsroom/article/0,,id=124937,00.html

About the authors:

Dave and Zach work for the Knowledge Discovery Group of the MITRE Corporation, a non-profit company that operates a Federally Funded Research and Development Center for the Internal Revenue Service. Zach earned his Master of Science Degree in Computer Science from the University of Massachusetts at Amherst. Dave earned his Master of Science Degree in Computer Science from George Mason University, where he is currently working on his PhD.