States Follow the IRS in Joining the Big Data Revolution

Data analysis is not a new concept, but it nonetheless is the center of a social and technological revolution. In recent months, we have gained significant insight into how the IRS and state taxing authorities are leveraging advanced technology and machine learning to mine the petabytes of taxpayer data that they collect and retain.

IRS Use of Data

The IRS is not a newcomer to the field of data analysis. In 1962, the IRS began using computers to randomly select tax returns for audit, and by 1969 it used computers to automate a process by which returns were selected for audit based on criteria weighted for probability of error or fraud, a product of years of data collection. IRS use of computerized data analysis has become more and more sophisticated over the years, to the point that many cases of taxpayer error or fraud, such as 1099 or W-2 underreporting or even failure to file, are automatically caught and corrected without the need for any human involvement.[1]

Not content to rest on its laurels in the era of big data, the IRS has made significant investments in data analysis, including the creation of the Nationally Coordinated Investigation Unit ("NCIU"), an arm focused on using IRS and external data to select criminal investigations, and signing a 7-year, $99 million deal with Palantir Technologies in September. The Palantir system allows the IRS to search and analyze vast quantities of data from both internal and external data sources in a single, unified research platform.[2] During a December 5 ABA Tax Section webinar, Todd Egaas, Director of Technology Operations and Investigative Services at the IRS Criminal Investigation Division ("CI"), described how powerful new data analytics tools, such as those provided by Palantir, allow investigators to automate much of the investigative process and catch complex patterns of non-compliance that would be difficult, if not impossible, for a human agent to catch alone. One such victory came in November 2017, when U.S. Magistrate Judge Jacqueline Scott Corley, sitting in San Francisco, granted the IRS' request to compel Coinbase Inc., one of the largest virtual currency exchanges, to turn over information on the accounts of 14,000 Coinbase users who had bought, sold, sent, or received at least $20,000 in a given year between 2013 and 2015. As its data showed that only 800 to 900 U.S. taxpayers reported Bitcoin gains from 2013 to 2015, the IRS argued that many Coinbase users were not reporting their Bitcoin gains.[3]

Don Fort, IRS CI's Chief, has been quick to note that IRS advances in data analysis bear no resemblance to a "Big Brother" situation, saying on December 13 at the ABA's National Institutes on Criminal Tax Fraud and Tax Controversy in Las Vegas that "this is using the data that we have, leveraging that data to find areas of noncompliance, and sending those cases out into the field." Further, he noted that IRS CI maintains stringent policies on privacy and disclosure of information.[4]

---

[1] ABA Tax Section Webinar Series, The IRS' Modern Use of Artificial Intelligence and Big Data for Tax Enforcement (2018)

[2] https://www.irs.gov/pub/irs-utl/lca-pia.pdf

[3] https://www.bloomberg.com/news/articles/2018-02-08/irs-cops-scouring-crypto-accounts-to-build-tax-evasion-cases

[4] https://www.taxnotes.com/tax-notes-today/criminal-violations/irs-big-data-analysis-doesnt-make-it-big-brother-official-says/2018/12/17/28pjq

The IRS is not limiting its data mining efforts to detection of criminal fraud. The agency also uses datamining to identify potentially fruitful cases to pursue on civil audit, as well as to predict the outcome of cases that are referred to the IRS Office of Appeals. During the December 5 webinar, IRS Chief Analytics Officer Benjamin Herndon pointed to new analytics tools as invaluable to identifying taxpayer responses to changes in notices and policies, allowing the agency to efficiently maximize desired taxpayer responses. For an agency facing an increasingly data-driven world and frequent budget cuts, these powerful tools are essential.

Recent State Developments

Where the IRS has gone, the States have begun to follow. In recent years, state departments of revenue have begun to make sizeable investments in data analysis, and these investments have already borne fruit. In 2015, the Utah Tax Commission discovered $11 million worth of fraudulent returns, but due to Utah's use of computerized data analysis to catch fraudulent returns, scammers got away with less than $20,000.[5] In 2017, Massachusetts rolled out its new $50 million return-processing system, and taxpayers immediately noticed a significant uptick in the number of returns flagged for additional fraud protection screening.[6] At the November State and Local Tax Forum, Commissioner of the Massachusetts Department of Revenue Christopher C. Harding explained that Massachusetts is collaborating with other states and vendors on fraud prevention, and that while under the old system 30% of returns were run through fraud filters, the new system runs every single return through the filters. Investments in such fraud filters have already been shown to pay sizeable dividends: the Arizona Department of Revenue stated in its 2017 Annual Report that enhanced fraud detection, including advanced analytics and machine-learning algorithms, has protected over $100 million in state revenue from being stolen since 2015.[7]

Like the IRS, many state departments of revenue have faced significant budgetary pressure in recent years, as governments have tried to cut down the size and cost of government, and have turned to technology to fill the gap. As powerful as data analytics are, however, there is a limit to the extent they can replace human investigators. In 2016, for example, the Arizona Department of Revenue began to lay off dozens of auditors and tax collectors, citing budget cuts. The result was a catastrophe, as audit collections dropped nearly 47 percent—$82 million—in 2017.[8] The IRS itself has taken a markedly different approach: IRS CI has recently announced a hiring blitz, in the course of which it will hire 250 special agents, a number of data scientists, and over 100 professional staff.[9]

Conclusion

The IRS and state taxing authorities will only grow their reliance on big data to identify taxpayers for civil audits and criminal investigation in the coming years. Using machine learning, they will become increasingly sophisticated in spotting patterns of irregularity, as well as honing in on

---

[5] http://archive.sltrib.com/article.php?id=3462795&itype=CMSID

[6] https://www.bostonglobe.com/business/2017/05/14/not-audit-but-your-refund-may-depend/DKeCuUYqVlJYrQpeYRApTL/story.html

[7] https://www.azdor.gov/sites/default/files/REPORTS_ANNUAL_2017_ASSETS_fy17_annual_report.pdf, pg. 11

[8] https://www.azcentral.com/story/news/local/arizona-investigations/2018/01/30/if-arizona-had-not-fired-so-many-tax-collectors-would-there-money-teacher-raises/1067745001/

[9] https://www.taxnotes.com/tax-notes-today/criminal-violations/irs-big-data-analysis-doesnt-make-it-big-brother-official-says/2018/12/17/28pjq

taxpayers for civil audit that have taken tax positions that, if successfully contested, will result in substantial adjustments, penalties, and interest in the hands of the government. Data analysis may also allow for more audits and criminal investigations in the coming years, as part of the investigative work done by agents and auditors shifts to machines.