IRS Fraud Filters
Balancing Return Integrity and Taxpayer Rights
Speakers

• Moderator: Joshua R. Beck - Taxpayer Advocate Service
  Senior Systemic Advisor

Panelists:

• Nina Olson - National Taxpayer Advocate

• Denise D. Davis - IRS Wage & Investment, Return Integrity
  Operations Director

• James Creech - Law Offices of James Creech
What Are Fraud Filters?

- Fraud filters are automated filters designed to prevent Treasury from paying fraudulent refunds by identifying potential fraud and freezing the refund before it is paid.

- Fraudulent refunds can be either refunds claimed as part of an identity theft (IDT program) or by a taxpayer fraudulently reporting income or dependents in order to obtain a larger refund (WVP program).

- While the two types of fraud are not necessarily linked, the indicators are sufficiently similar that the same filters can catch both types of fraud.

- IRS estimates that it stopped at least $10.6 billion worth of fraudulent refunds before they were paid in 2016, but paid $1.6 billion worth of fraudulent refunds that were not stopped by the filters. Since 2015 the IRS reports $24 billion in IDT refunds have been stopped.
Sophisticated Fraud Needs Sophisticated Models

• First use of automated fraud detection was the Electronic Fraud Detection System (EFDS) which was implemented in 1994

• EFDS was refined and updated until 2017 when it was superseded by the Dependent Database (DDb) and the Return Review Program (RRP)

• More efficient programs were needed to counteract more efficient criminals. In the past 10 years ID theft has moved from a simple stolen TIN to a stolen TIN plus stolen W-2 or other income information. This allows organized crime to file returns that can be essentially correct but for the bank account and routing information

• The IRS countermeasures are working, Number of ID theft claims:
  2015 - 677,000  2016 - 401,000  2017 - 242,000  2018 - 199,000
Differences Between DDb and RRP

- Dependent Database (DDb) - Consists of binary, yes or no, rules. If the rule is broken the return is held for further analysis.

- Return Review Program (RRP) - In addition to binary rules (1000+ rules) RRP uses more advanced modeling such as predictive models and clustering to identify non obvious indicators of fraud. For example, the clustering function could potentially identify batches of fraudulent refunds from the same geographic area, apply lower threshold for flagging those refunds, and place holds on more suspicious refunds from that zip code.
Data, Data, Everywhere

• The IRS looks at a number of data points but the most important are third party information returns such as W-2’s and 1099’s. GAO notes that there are 40 different information returns and 25 of them are directly matched to an individual income tax return.

• RRP contains 3 years worth of taxpayer information.

• The IRS can modify RRP during the filing season based upon new data or the RRP’s interpretation of existing data. Examples of modifications include: improving certain filters to catch an emerging fraud, responding to a third party data breach that could be used to file false returns, or increase the threshold for flagging a return in an effort to reduce the number of legitimate returns flagged as fraudulent.
Scope of the Program:

As of March 2018 the IRS had spent $419 million to develop and implement RRP

Fiscal year 2019 budget request was $106 million
What Happens If A Return Is Flagged?

- If a return is flagged due to a failed rule, or it exceed a threshold score issue by RRP, the return is reviewed by the IRS. Some returns just below the threshold are reviewed a second time.

- The capacity of the IRS to review and process the returns selected by the program is one of the factors used by RRP.

- Reviews range from asking a taxpayer to verify their identity (IDT filter) to a CI referral (fraud, frivolous arguments, etc).

- 2017 filing season, 857,438 returns flagged for IDT. 219,210 for RRP. In total this is less that 1% of the 158 million individual returns filed.
Flagging A Return Freezes The Requested Refund

• Depending on the type of flag the return is sent to TPP (IDT refund program) or WVP (non IDT program)

• For IDT refund issues TPP contacts the taxpayer by mail and asks them to verify their identity. This can be done over the phone, online, or in person by visiting a Taxpayer Assistance Center

• Guidance for TPP procedures can be found in IRM Section 25.25.6
Average Time To Resolve A TPP
Income Verification Is 40 Days

• Submission to selection: 2 days

• Notification to resolution: 24 days (includes selection to notification: 5 days)

• Resolution to refund: 14 days

• Total of 40 days For comparison average time for pre-refund wage verification is 38 days and a dual TPP and pre-wage hold is 46 days

• Worth noting that if you claim EITC or ACTC the IRS cannot issue a refund before mid February. If a taxpayer claiming a EITC files in January and is selected by IDT they may not receive their refund until early April
How Well Are The Fraud Filters Working For Taxpayers?

• The IRS definition of success of the program is measured in dollars. However, they do have two user based metrics that try to quantify how well the filters are working

• Burden on taxpayers is tracked by the False Positive Rate and Operational Performance Rate

• False Positive Rate = Legitimate returns flagged divided by total number of potentially fraudulent returns

• Operational Performance Rate = False positive rate minus number of false positives resolved within two weeks

• Regardless of operational metrics, pre refund hold TAS cases jumped significantly in 2018. 63,637 cases were opened compared to 16,432 cases in 2017
2018 Tax Filing Season Was Plagued By High Error Rates

- FPR error rate for non-IDT returns was 81% (Jan 1 - Oct 3, 2018)
- OPR error rate for non-IDT returns was 64% (Jan 1 - Oct 3, 2018)
- TAS internal metric Operation FPR which uses a 4 week cure window numerator and the 2 week IRS OPR window denominator was 77% (Jan 1 - Oct 3, 2018)
- Note that industry standards consider a 50% error rate as an acceptable error rate
The IRS Is Improving

- The IRS has improved the time it takes to check the posting of third party information returns that could prevent a false positive. In 2019 information is checked daily as opposed to weekly.

- If return is flag for an information mismatch, but the mismatch would not impact the claimed refund, the refund is released immediately.

- If a refund is flagged for both IDT and non IDT fraud concerns both will be worked simultaneously to reduce the amount of processing time.
Critiques Of Algorithmic Selection
Beyond The IRS User Metrics
Dangers Of Relying On Algorithms In General

- Studying the results of modern algorithms is an empirical science. Outcomes are generally not reproducible like physics or chemistry. This requires a higher degree of trust in the people creating the parameters, identifying and labeling the relevant data, and interpreting the outcomes (and repeating the process). The open the model is, the more stakeholders can identify output problems.

- Effect of reinforcement learning? One potential example in RRP is the clustering analysis that is performed. Does a high level of geographic flagging indicate more fraud? Or does it indicate that the RRP filters anticipate more fraud and apply the filters more rigorously to anyone in that zip code?
Dangers Continued

• The whole point of predictive technologies is to identify patterns that exist in data that may not be visible to humans. What we can’t tell is does this technology identify human behavior that already exists or is it exasperating tendencies amongst low income filers? What level of group behavior can we apply to an individual without offending our sense of justice?

• How does data play a role on this? Predictive technology is entirely dependent on data. How do we adapt to segments of society that generate more/less/different data than others? The engineers that use the data to create the models are more likely to use data that represents their own world view
IRS vs Taxpayer Burden

Current Burden Load

Hypothetical Algorithm Enhanced Burden Load

Taxpayer

IRS
How Do We Minimize Algorithm's Infringement Of Individual Rights?

- The answer is not to stop using these systems. The solution is to design the systems from the outset in a manner that does not simply solve a problem but rather embraces a balancing tests between competing needs.

- Include ombudsmen in identifying goals, as well as the data collection, data analysis, and review process. Require solutions prior to implementation that define acceptable failure rates, create procedures to remedy harms and dedicated resources to implement the procedures.

- Understand that ethics/burden is in the eye of the beholder. Inappropriate to say that the agency has the sole ability to determine how burden should be defined or that the law defines the applicable “ethics” of how to use the technology.

- As much transparency as practical.
Questions?