2022 IRS/TPC Research Conference

Abstracts of Papers with Bios of Presenters

Session 1: Balancing Audits: Enforcement vs. Measuring Non-Compliance

Improving Risk Models by Supplementing Random NRP Audits with Non-Random Operational Audits Using Statistical Controls for Bias

Ishani Roy, Brett Collins, ***Alex Turk****, Mark Payne (IRS, RAAS)*

Measuring and predicting tax reporting compliance typically requires audit data to label the reporting behavior of the taxpayer. The Internal Revenue Service (IRS) National Research Program (NRP) provides reporting compliance data from random audits that thoroughly and objectively classifies the issues on the tax return. The NRP audit data is representative of the underlying filing population, but detailed audits by revenue agents are costly and their random nature can result a relatively large share of audits result in small or no adjustments to total tax. This paper develops a framework for including labels from Operational audits (OP) to supplement the results from random National Research Program (NRP) audits to produce unbiased estimates of reporting compliance risk. Since operational audits have no additional opportunity cost, including the data from these audits promises to extend the capabilities of the risk model while keeping samples small and focused. We also investigate replacing part of the NRP data with appropriately chosen OP data to produce unbiased tax reporting predictions, which may be a useful way to extend the capabilities of the risk model in segments with low NRP coverage, or where NRP cases look like OP cases. This paper focuses on models for cases that involve claims of the Earned Income Tax Credit (EITC), which includes segments where NRP samples are already low.

Alex Turk *is the Supervisory Economist for the Policy and Program Impact Lab at the IRS's Research, Applied Analytics, and Statistics Division. An Iowa State alum, Turk earned his undergraduate economics degree (1988) and Ph.D. in economics (2000).*

Integrating Reward Maximization and Population Estimation: Sequential Decision-Making for Internal Revenue Service Audit Selection

Peter Henderson*, Ben Chugg, Kristen Altenburger, Daniel E. Ho (Stanford University); Brandon Anderson (Stanford University/IRS); John Guyton, Alex Turk (IRS, RAAS); Jacob Goldin (Stanford University/U.S. Department of the Treasury)*

We introduce a new setting, *optimize-and estimate structured bandits*. Here, a policy must select a batch of arms, each characterized by its own context, that would allow it to both maximize reward and maintain an accurate (ideally unbiased) population estimate of the reward. This setting is inherent to many public and private sector applications and often requires handling delayed feedback, small data, and distribution shifts. We demonstrate its importance on real data from the United States Internal Revenue Service (IRS). The IRS performs yearly audits of the tax base. Two of its most important objectives are to identify suspected misreporting and to estimate the "tax gap"—the global difference between the amount paid and true amount owed. We cast these two processes as a unified optimize-and-estimate structured bandit. We provide a novel mechanism for unbiased population estimation that achieves rewards comparable to baseline approaches. This approach has the potential to improve audit efficacy, while maintaining policy-relevant estimates of the tax gap. This has important social consequences given that the current tax gap is estimated at nearly half a trillion dollars. We suggest that this problem setting is fertile ground for further research, and we highlight its interesting challenges.

Peter Henderson *is a joint JD-Ph.D. (Computer Science) candidate at Stanford University. He is also an Open Philanthropy AI Fellow, Graduate Student Fellow at the Regulation, Evaluation, and Governance Lab, and Student Volunteer at the Internal Revenue Service.*

Augmenting National Research Program Tax Change Estimates by Incorporating Operational Audit Information: A New RAAS Research Initiative

***Louis Rizzo**, John Riddles, Xiaoshu Zhu, Richard Valliant (Westat); Kimberly Henry (IRS, RAAS)*

The National Research Program (NRP) is a long-standing initiative of the Internal Revenue Service (IRS) which estimates tax change using a probability sample of tax returns, but in recent years, budget reductions to the IRS have necessitated reductions in the NRP sample sizes, resulting in a reduction in the precision of tax change estimators. This research initiative studies the possibility of leveraging a large source of auxiliary information, the operational audits (OP) done each year as part of normal IRS tax compliance assurance.

We have developed several estimators for NRP and OP data using a variety of methods and assumptions, including one estimator which treats the operational audits as self-representing in the tax-return universe for a given year, given a weight of 1. For this approach we use random forest modeling and compute a 'propensity to be operationally audited' for the NRP audits, then adjust the NRP sample weights by one minus this estimated propensity, producing conservative estimators. We also develop composite estimators that compute NRP estimators and operational return estimators (unweighted), and linearly combine them, where the linear factor is based on relative precision. This first requires Kolmogorov-Smirnov (K-S) and other tests of the equality of the distributions, do determine if they are "full equating" or "partial equating" on only a subdomain within the stratum. Another estimator we have developed is a 'sliding-scale' composite estimator, which composites the NRP and OP estimators for a stratum whether or not full equating is justified, but sets composite weights by MSE rather than variance, and assumes the un-equated OP estimator may be biased.

***Louis Rizzo** has 30+ years of experience in survey research work following the completion of his PhD in mathematical statistics at the University of Chicago in 1989. He has been at Westat since 1993 after four years as a professor at the University of Iowa. In his time at Westat, he has participated in a number of research projects for the Internal Revenue Service.*

***Alan Plumley** is a Technical Advisor in the IRS Office of Research, Applied Analytics, and Statistics. His expertise is in the areas of compliance measurement and modeling for efficient workload selection and resource allocation. A 37-year veteran of IRS research, Alan earned his Ph.D. in public policy from Harvard University. He is also the editor of the annual IRS Research Bulletin.*

Session 2: Burden vs. Opportunity

The Spiderweb of Pass-Through Tax Planning

*Jacob Goldin, **Ryan Hess**, Daniel E. Ho, Rebecca Lester, Mansheej Paul (Stanford University)*

We study the tax planning choices of pass-through businesses, focusing on partnership entities. Partnerships control more than $30 trillion in assets, vastly outnumber U.S. public firms, and are increasingly used by business taxpayers. However, prior literature provides little evidence about which specific features of partnerships facilitate tax planning, due in large part to limited public data about these businesses. We examine the role of organizational complexity using confidential, anonymized administrative tax data. We first construct partnership entity network structures to evaluate the size, ownership, and connections among partnership entities. Approximately 75 percent of partnership groups are "simple" structures composed of one single partnership owned directly by individual taxpayers. In contrast, the most complex organizations resemble webs of ownership, with clusters of overlapping partners. Preliminary work leverages machine learning (ML) approaches to predict partnership non-compliance; Future analyses will include additional features and alternative ML approaches to improve upon this baseline. Our analysis of network structures and tax planning strategies helps explain why partnerships are attractive to both businesses and high net worth individuals for tax planning purposes. The findings also contribute to the broader strategy for enhancing IRS capabilities related to tax administration of flow-through entities.

***Ryan Hess** is a Postdoctoral Fellow at Stanford University. Ryan's research with Stanford's RegLab involves examining tax networks to identify ways to improve tax compliance.*

Automatic Tax Filing: Simulating a Pre-Populated Form 1040 Automatic Tax Form

*Lucas Goodman, Andrew Whitten (U.S. Department of the Treasury, Office of Tax Analysis); **Katherine Lim** (Federal Reserve Bank of Minneapolis); Bruce Sacerdote (Dartmouth College)*

Each year Americans spend over two billion hours and $30 billion preparing individual tax returns, and these filing costs are regressive. To lower and redistribute the filing burden, some commentators have proposed having the IRS pre-populate tax returns for individuals. We evaluate this hypothetical policy using a large, nationally representative sample of returns filed for the tax year 2019. Our baseline results indicate that between 62 and 73 million returns (41 to 48 percent of all returns) could be accurately pre-populated using only current-year information returns and the prior-year return. Accuracy rates decline with income and are higher for taxpayers who have fewer dependents or are unmarried. We also examine 2019 non-filers, finding that pre-populated returns tentatively indicate $9.0 billion in refunds due to 12 million (22 percent) of them.

***Katherine Lim** is an economist with the Federal Reserve Bank of Minneapolis. Her research and policy interests include self-employment, women's labor force participation, and the taxation of pass-through businesses.*

The Distribution of the Individual Income Tax Underreporting Tax Gap

***Drew Johns** (IRS, RAAS)*

This paper presents estimates of the distribution of the individual income tax and self-employment tax underreporting tax gap. Tax Gap estimates historically have been reported according to the type of tax and the major components within a type of tax have reflected sources of misreporting, such as line items or groups of line items. Recently, there has been interest in understanding the distribution of the individual income tax underreporting tax gap by tax return characteristics. Self-employment tax is reported on the individual income tax return and included in the calculation of total tax on that return. Therefore, this paper includes underreported self-employment tax in estimating the distribution of the underreporting tax gap for the individual income tax return. Estimates of noncompliance are grouped by Activity Code and compare those estimates to the allocation of examination resources, controlling for the type of examination. The distribution of noncompliance is also analyzed at three different levels of income: reported income, examiner determined income, and an econometric estimate for undetected income called the Detection Controlled Estimate (DCE). Taxpayers who underreport their income often fall into a lower income percentile when ranked by reported income relative to their percentile when ranked by examiner determined or DCE adjusted income. This fact has implications for tax administration and resource allocation since only reported income for taxpayers is observed, absent an examination. In addition, a significant portion of the estimated tax gap reflects noncompliance that would unlikely be detected on an examination. Therefore, the tax gap does not necessarily reflect the amount of direct revenue that could be recovered through increased enforcement activity.

***Drew Johns** is an Economist in the Compliance Modeling Lab within the IRS Knowledge Development and Application Division. Since joining the IRS in 2005 he has primarily been involved with research related to reporting compliance studies and measuring the tax gap. He has a B.A. in Mathematics and Economics from Susquehanna University (Selinsgrove, PA) and an M.S. in Agricultural and Applied Economics from Virginia Tech (Blacksburg, VA).*

***Steve Rosenthal** is a senior fellow in the Urban-Brookings Tax Policy Center at the Urban Institute. He researches, speaks, and writes on a range of federal income tax issues, with a particular focus on business taxes.*

Session 3: Improving Audit Outcomes: Thinking Inside the Box

Graph-Based Machine Learning Methods for Case Selection and Population Segmentation

***Matt Olson**, Ben Howard, Devika Mahoney-Nair (MITRE); Annette Portz (IRS, RAAS)*

We present the motivation and methods of active IRS research into the applicability of graph-based machine learning for enforcement on the Global High Wealth (GHW) population. Specifically, our immediate focus is graph neural network autoencoders—unsupervised models that learn fixed-length vector representations of variably sized networks of pass-through entities and taxpayers. With the fixed-length representation in hand, we will then use more 'standard' clustering methods to help IRS better stratify the GHW population. We have two near-term goals with the work: (1) assist reinforcement learning-guided case selection where the history of audits is presently too thin to support the development of fully supervised models and (2) narrow the confidence intervals on IRS's estimates of the tax gap due to noncompliant GHW taxpayers. In the long term as exam results are more systematically available on the GHW populations, we expect that the encoders will accelerate development of supervised models.

***Matt Olson** is a senior principal analyst in MITRE's IRS program. He has worked across Research, the Business, and IT for the last ten years munging data, finding signal, building models, and helping to improve operations. He holds a BS in Aerospace Engineering from MIT and a Ph.D. in Operations Research from George Mason.*

Automated Discovery of Tax Schemes Using Genetic Algorithms

*Karen Jones, Camrynn Fausey, **Eric O. Scott**, Geoff Warner, Sanith Wijesinghe (MITRE); Hahnemann Ortiz (IRS, LB&I)*

Recent estimates of the yearly tax gap exceed $600 billion in the United States, and a significant portion of this gap is due to financial transactions whose sequence and structural design results in exploiting loopholes in the Internal Revenue Code (IRC). Taken individually, these transactions may appear perfectly innocuous and routine, but considered collectively, and in the context of the financial networks in which they occur, reveal an insidious intent to reduce tax liability.

Our approach consists of an Evolutionary Behavior Discovery software framework (EBD): an iterative, two-fold computational framework for generating novel taxpayer behaviors. One portion of this framework employs an evolutionary algorithm—a search heuristic inspired by the process of Darwinian evolution, which encodes potential taxpayer behaviors as "chromosomes" that are iteratively modified via mutation, recombination, and selection. The second portion of the framework is the taxpayer behavior simulation that occurs during the evaluation phase of the evolutionary cycle (before chromosome selection). The simulation executes any viable economic transactions encoded by the chromosomes, in which financial entities exchange goods and services, and computes any taxable income incurred.

We are currently investigating behaviors that may have emerged in response to significant changes in the IRC introduced under the Tax Cuts and Jobs Act (TCJA). We are exploring several mechanisms by which large businesses might reduce their liability under specific provisions. By extending our EBD into the domains of new tax policy, we hope to enable the IRS to anticipate novel taxpayer avoidance behaviors that may have emerged following these changes to the IRC, and to aid in the development of filters, or other statistical techniques, capable of detecting these novel tax planning behaviors.

***Eric "Siggy" Scott** is a Lead Artificial Intelligence Engineer at MITRE Corporation in Northern Virginia and a PhD candidate at George Mason University. His research focuses on heuristic optimization algorithms, transfer learning, and their applications to modeling problems in a variety of fields. He holds a double B.Sc. in Computer Science and Mathematics from Andrews University in Berrien Springs, Michigan, and a M.Sc. in Computer Science from George Mason University.*

Incorporating the Specific Indirect Effect of Correspondence Audits into IRS Resource Allocation Decisions

***Alan Plumley**, Daniel Rodriguez (IRS, RAAS); Leigh Nichol (MITRE)*

The Internal Revenue Service (IRS) has long believed that after taxpayers are audited, many become more compliant in subsequent years. In recent years, IRS and MITRE have estimated the specific indirect effect associated with many categories of correspondence audit and have expressed these effects as the average additional amount of tax reported by audited taxpayers in the five years following the audit that they would not have reported had they not been audited. This paper is a first look at how the IRS could make use of such plausible, empirical estimates in its operational decision-making.

The paper provides an overview of the indirect effect estimates that we developed for all correspondence audit categories, then provides a simulation to illustrate how they can be combined with the direct effect to maximize the total impact of the audits on tax revenue. Because we found that the indirect effect varies across audit categories, the budget allocation that maximizes total (direct + indirect) revenue is different from the allocation that seeks to maximize only the direct revenue. This means that when indirect effects are considered in the budget allocation across multiple categories, the combined direct revenue will decrease relative to the allocation that maximizes direct revenue alone. However, the combined sum of direct and indirect revenue will be largest when the allocation takes both into account.

***Alan Plumley** is a Technical Advisor in the IRS Office of Research, Applied Analytics, and Statistics. His expertise is in the areas of compliance measurement and modeling for efficient workload selection and resource allocation. A 37-year veteran of IRS research, Alan earned his Ph.D. in public policy from Harvard University. His dissertation on the determinants of individual income tax compliance broke much new ground in the effort to estimate the impact of various IRS activities on the voluntary compliance behavior of the general population. He has presented research papers at many tax conferences, and has co-authored chapters for two books related to tax administration and compliance. He is also the editor of the annual IRS Research Bulletin.*

***Mike Stavrianos** is a Founding Principal of ASR Analytics (a GCOM Company), where he currently leads the firm's Federal Government practice. Mike has served as an advisor and consultant to various IRS divisions over the past 25 years, including RAAS, research groups in SB/SE and W&I, Criminal Investigations, and IRS IT. He currently supports a variety of RAAS initiatives involving taxpayer experience, fraud detection, graph analytics, and behavioral research.*

Session 4: Why Do Taxpayers Comply?

To File or Not to File? What Matters Most?

*Brian Erard (B. Erard & Associates); Tom Hertz, Pat Langetieg, **Mark Payne**, Alan Plumley (IRS, RAAS)*

Federal individual income tax returns are a key data source for understanding the drivers of many forms of taxpayer behavior. Although the information reported on these returns is not always entirely accurate, they provide granular details on many taxpayer characteristics. This information is not available for nonfilers, making this population an elusive research topic. In our earlier work on this topic, we have attempted to overcome this informational gap by comparing the characteristics of filers of federal individual income tax returns, against the characteristics of the more general population of filers and nonfilers recorded in Census survey data; specifically, the Current Population Survey Annual Social and Economic Supplement (CPS-ASEC), though this approach involved several drawbacks.

Recently, however, we have been able to link detailed IRS administrative data from tax returns and third-party information reports with Census survey data for several tax years, permitting a more definitive assessment of whether a tax filing requirement is present and, if so, whether a tax return has been filed (either timely or late). In this paper, we exploit this new data source to estimate several logit specifications of filing behavior to understand what factors distinguish whether a taxpayer is likely to file a timely return, a

late return, or no return at all. Among the key factors we explore are the types and amounts of income reported to the IRS on third-party information documents, eligibility for major tax benefits (e.g., credits), prior filing behavior, filing status, changes in economic well-being, the extent of pre-payments of tax (e.g., through withholding), and demographic changes that individuals experience over time.

Mark Payne *is an Economist in the Knowledge Development and Application Division (KDA) in the Research and Applied Analytics Department (RAAS) at the IRS. His research has focused on measuring and analyzing the extent of non-filing and underpayment of taxes by non-filers and on analyzing the effects of taxpayer preparation and submission method choices on return accuracy.*

Economic Influencers of Total Enforcement Revenue Collected and Operational Implications

Jess Grana*, Lucia Lykke, Sam Schmitz (MITRE); Ron Hodge (IRS, RAAS)*

IRS resource planning considers the average revenue per case at each step of the compliance process. Revenue per case is derived from historical data. However, changing economic conditions, IRS resources and taxpayer population may cause future revenue per case to deviate from past values. Resource allocation models that use outdated revenue per case inputs will not generate staffing allocations that optimize total enforcement revenue.

We build two proof-of-concept models to forecast enforcement revenue. First, we develop a prototype "macro level" model to estimate the effect of economic conditions, IRS resources, and taxpayer attributes on broad measures of enforcement revenue. This model helps illuminate the potential drivers of enforcement revenue at a high level. Second, we develop a "micro" model that forecasts revenue at the level of steps in the compliance process (e.g., Automated Collection System cases and Field Collection cases derived from correspondence audits). Using taxpayer characteristics, we predict taxpayer progression through the compliance process as well as their expected revenue once arrived at each step. We use the "micro" forecasts of average revenue at these steps in a workforce allocation model, which optimizes resource allocation based on expected revenue per case at each step of the compliance process. We demonstrate that using forecasted revenue per case rather than historical estimates leads to workforce allocations that can generate higher total enforcement revenue.

Jess Grana *is an economist with the MITRE Corporation. She specializes in microeconomics and conducts research related to taxation, competition, and transportation. She received her Ph.D. in Economics from American University.*

Non-Monetary Sanctions as Tax Enforcement Tools: Evaluating California's Top 500 Program

Chad Angaretis, Allen Prohofsky (California Franchise Tax Board); ***Brian Galle*** *(Georgetown University Law Center); Paul R. Organ (University of Michigan)*

Many U.S. states and countries around the world use non-monetary sanctions to encourage tax compliance, including public disclosure, license suspension, and withholding of other government-provided benefits or privileges. Little is known about the effectiveness of these programs. Using administrative tax microdata from California's "Top 500" program, we study whether notices warning of the imminent publication of a taxpayer's personal information and potential license suspension affect payment and other compliance outcomes, as well as whether these notices affect subsequent reported earnings. Exploiting variation over time in the cutoff balance for program eligibility we find evidence of strong positive compliance responses to the program, with no evidence of an impact on subsequent reported earnings. We also develop estimates of the deadweight loss caused by publication of noncompliers, and conclude that the program generates positive net social welfare. Together, these results suggest that non-monetary sanctions can be efficient tax enforcement tools, at least among the relatively high-income population we study.

Brian Galle *is a professor of law at Georgetown University Law Center, and a research associate in the IRS Joint Statistical Research Program. He is currently on leave from Georgetown while working at the Securities & Exchange Commission; all views presented here are his own.*

Alexander Yuskavage *is an economist at the Office of Tax Analysis in the Department of the Treasury. His research primarily focuses on taxpayer compliance and corporate microsimulation. He received his Ph.D. from the University of Wisconsin-Madison in 2010.*