



## 2024 IRS/TPC Research Conference

*Abstracts of Papers with Bios of Presenters*

### Session 1: Harnessing Data for Better Research

#### Improving Linkages to Individual Income Tax Data

**Amy O'Hara**, *Stephanie Strauss, Maanasa Vatsavayi, Nathan Wycoff (Georgetown University)*

Researchers within IRS, Treasury, and other government agencies often need to link individual taxpayer data to conduct longitudinal analyses or link external sources. Most linkages rely on Taxpayer Identification Numbers (TINs). State and local agencies also seek linkages to federal tax information to measure and improve program outcomes and benefits access. These institutions often lack Social Security Numbers (SSN) and may have unknown data quality for other personal identifying information.

**Amy O'Hara** is a Research Professor in the Massive Data Institute and Executive Director of the Federal Statistical Research Data Center at the McCourt School of Public Policy. She works on data governance, linkage, and privacy preserving technologies. Her research focuses on population measurement, eviction and debt data, and methods to improve secure data access. She received her Ph.D. in Economics from the University of Notre Dame.

---

#### A Large Scale, High Quality US Occupational Database: Results from Merged IRS and ACS Write-Ins

*Victoria Bryant, Thomas Hertz, Kevin Pierce (IRS, RAAS); Julia Beckhusen, Lynda Laughlin, Liana Christin Landivar, Carl Sanders (US Census Bureau); Josh Gagne, David Grusky, Sofia Jamesson (Stanford University); Michael Hout (New York University); Ananda Martin-Caughey (Brown University); Javier Mirand (University of Jena)*

The measurement of worker occupations is a key component in understanding long term changes in the U.S. social and economic structure. However, practical problems in measuring occupation at scale, such as non-universe data coverage, noisy survey responses, and difficulty working with raw text strings, continue to make using occupation data a roadblock to statistical agencies and researchers alike. In this project, we develop and validate an occupation database that addresses all these problems by combining the U.S. Census Bureau and Internal Revenue Service's best available measures of occupations to create a large, high quality, and linkable database of individual worker occupations.

We match Federal Tax Information (FTI) data available for all electronically filed tax returns in year 2019 to American Community Survey (ACS) data for the same year. Here, we present the similarities and differences in individual occupation responses between the two sources with the goal of understanding the strengths and weaknesses of each. First, we compare write-in responses between the IRS Form 1040 and the ACS using Token Set Ratios (TSRs), a means of approximate string matching that quantifies the distance between the characters in two text strings. Second, we develop and use a Large Language Model-based autocoder to examine the semantic relationship between write ins. Finally, we preliminarily investigate the relative sensitivity of each write-in to year-over-year occupational transitions.

We find that the overall quality of TSR matches across the two write-in data sets is bimodal, with over 50% of the sample with a high-quality match, but another 25% of the sample with what seems to be a no-match. The average write-in match score varies significantly over occupation and age. Results from a regression model indicate that quality of matches across write-ins is positively associated with higher socio-economic status and the ability of the Census Bureau's probabilistic identity matching system to identify the worker. Finally, combining data across years to create a two-year panel, we find that the IRS and ACS report significantly different occupational mobility rates, with the IRS reports more conservative and the ACS more liberal in assigning changes. Results from the Large Language Model are still awaiting disclosure clearance.

*Carl Sanders is an Economist at the Census Bureau working on measurement issues in income and occupation.*

---

## Disaggregating Tax Compliance Burden: A Comparative Study

**Bizuayehu Bedane** (IRS, RAAS)

The main objective of this research is to conduct a comparative examination of tax compliance costs incurred by individuals and business taxpayers in complying with tax regulations. This study uses administrative data and published literature to explore the concepts, methodological challenges, and structure of tax compliance costs. Several key insights have emerged by comparing taxpayers' compliance costs in the United States against those in the United Kingdom, Canada, Germany, and Australia. Firstly, the analysis reveals that tax compliance costs exhibit a regressive trend, suggesting a negative correlation of tax compliance costs with income and business size. Secondly, it highlights the challenges encountered in tax compliance research, such as non-response bias, data scarcity, and monetizing the value of tax compliance time, emphasizing the need for cautious interpretation when comparing such studies. Lastly, the study reveals that while individual taxpayers in the U.S. face higher tax compliance costs than their counterparts in Canada and Germany, small business taxpayers experience lower costs than Australia and the U.K.

*Bizuayehu Bedane is an Economist in IRS's Research, Applied Analytics, and Statistics division.*

---

*Brittany Jefferson is a Social Scientist in Taxpayer Services - Strategies & Solutions at the IRS. She attended Wayne State University in Detroit, MI and the University of Georgia in Athens, GA. She joined the IRS in October of 2020.*

*Leonard Burman is an Institute Fellow at the Urban Institute and Paul Volcker Professor Emeritus at the Maxwell School of Syracuse University. He co-founded the Tax Policy Center, a joint project of the Urban Institute and the Brookings Institution, in 2002.*

## Session 2: Discovering the Art of Avoidance

### Using a Gravity Model to Predict Cross-Border Tax Avoidance

Lori Stuntz, **Michael Udell** (IRS, RAAS)

International trade economists use gravity models to explain cross-border flows of goods and services across features of the two countries that either encourage (mass, or size) or discourage (distance) these flows. We repurpose this concept to measure the attractiveness for cross-border tax avoidance.

Our approach recasts the gravity model mass concept as a composite measure that includes the withholding tax rate (WHT) on payments between countries, ownership requirements associated with each WHT, and capital gains taxes imposed by the destination country. The distance term becomes measures of tax administration transparency across the border (with less transparency increasing the attractiveness for cross-border tax avoidance) and an index that measures regulator quality in each country. Unlike the classical gravity equation, which is measured as the attraction across two masses, this repurposed gravity equation for tax avoidance is not limited to an attraction across a single border. Instead, we develop a framework to measure the attraction across multiple borders as a sequence. An important feature of our model is that the order of the countries in a sequence matters.

We create a database of treaty dividend withholding tax rates and associated ownership percentages for qualified corporate dividends across 230 countries using treaty information from the International Bureau of Fiscal Documentation ("IBFD"). For each possible country pair, we record up to 4 different dividend withholding rates and required company ownership percentages, for a total of 59,018 possible bi-lateral cross-border dividend withholding rates. We combine these treaty WHT rates with country level data on domestic tax rates capital gains (also obtained from IBFD), World Bank measures of regulator quality and

political stability for each country, and indicators for country participation in various exchange of information programs with the United States. These data allow us to identify countries that are possible “destinations” (low or no income tax and low transparency) or “conduits” (low or no withholding tax and politically stable).

To generate sequences, we begin with all possible dividend withholding rates and associated ownership requirements for dividends between each of the 230 countries in our database (Country A – Country B) and we join in all of the dividend treaty WHTs for each of the 59,018 country B’s with the rest of the world to create over 15 million potential 3-country sequences. We calculate the gravity equation for tax avoidance for each sequence with weights estimated using a measure of foreign financial investment flows.

We develop a framework to chain gravity indexes and evaluate if it is advantageous to add an additional country to the sequence. The gravity model for cross-border tax avoidance is flexible and can be generalized so that sequences originate from any of the 230 countries in our dataset.

***Michael Udell** is a senior technical advisor in the Emerging Compliance Issues Lab in the Knowledge, Development and Application Division in IRS’s office of Research, Applied Analytics and Statistics. Mike is a boomerang employee of the IRS having begun his career in 1985 in the Research Division working with data from the Tax Compliance Measurement Program. He went on to work at the Joint Committee on Taxation through 2008 and worked for Ernst & Young before founding District Economics Group in 2013. Mike rejoined the IRS in 2023. His research interests span most federal taxes as well as the VAT in Europe. Chief among these is building statistical models of tax revenue and compliance. He currently works with Lori Stuntz in the ECI group on a gravity equation model of cross-border tax avoidance. Mike has a PhD in economics from Caltech.*

---

### Art in the Age of Tax Avoidance

***Matthew Pierson** (The Wharton School, University of Pennsylvania)*

Donations of art to non-profits may achieve both licit and illicit aims. Documenting the determinants and extent of these donations in the U.S., organizations that accept these donations differ substantially from those that do not. Non-profit organizations hold sizable art assets with 1.2% identified as holding art—valued at \$6 billion. There is significant heterogeneity in the determinants of accepting art donations, associated with both public good provision and secrecy, asset protection, and poor tax compliance. Only 17% of organizations holding art or accepting donations record its value. Yet those that do consistently write down this value after donation. I find that, conditioning on increased probability of IRS scrutiny, non-profits respond by providing valuations more often, and more accurately valuing donations of art. This behavior reveals a substantial wedge of art donations accepted at inflated values within the non-profit sector, with back of the envelope tax losses due to avoidance estimated of \$4.8 billion in income tax losses over 12 years.

***Matthew Pierson** is a researcher at Wharton Research Data Services (WRDS), The Wharton School, University of Pennsylvania. His work is in the fields of public finance, household finance, financial assets and markets, and law and finance, with a particular interest in the interplay of specific asset and legal structures and tax evasion and avoidance.*

---

### Staying on the Wagon: Estimating Indirect Deterrence Effects from Filing and Payment Compliance Programs

***Brett Collins**, Chris Wilson, Corbin Miller, Mark Payne, Sean Roh, Yan Sun, Alex Turk (IRS, RAAS)*

The Internal Revenue Service employs a comprehensive array of enforcement strategies aimed at ensuring timely tax filings and payments. Numerous studies, from both IRS and external academic sources, evaluate the effectiveness of these strategies on taxpayer compliance. These analyses typically distinguish between the direct effects—immediate behavioral changes in taxpayers treated enforcement process—and the indirect effects—adjustments in behavior among taxpayers who are not directly treated but are influenced by heightened awareness of IRS activities or revised perceptions of their own risk of IRS enforcement actions if they don’t meet their tax obligations.

Our research addresses gaps in the indirect effect literature by exploring three critical dimensions. First, we delve into the indirect effects of tax enforcement on populations that consistently demonstrate compliance, examining the preventative capabilities of IRS strategies to maintain voluntary compliance within these groups. This perspective is crucial for understanding the full spectrum of enforcement impact beyond the traditionally studied non-compliant sectors. Second, we expand our analysis to a national scale using the Facebook Social Connectedness Index, which allows us to model the transmission of enforcement perceptions through social connections. This method provides a robust framework for comparing variations in individuals' perceptions of IRS interventions, based on their connections to those directly affected by enforcement actions. Such an approach offers a nuanced understanding of the widespread indirect effects across the U.S. Third, we apply advanced statistical methods in our analysis, utilizing a two-stage least squares (2SLS) regression model with instrumental variables. This model capitalizes on the significant shifts in IRS resources and enforcement strategies from 2011 to 2019, using these fluctuations as a natural experiment to isolate the causal impacts of enforcement actions from potential confounding factors.

**Brett Collins** is an Economist with the Policy and Program Impact Lab of the Knowledge Development and Application group, part of IRS's Research, Applied Analytics, and Statistics division. He has been with RAAS for over fifteen years and has been involved with a variety of projects, including forecasting return volumes, evaluating taxpayer responses to new IRS outreach efforts, and assessing the impact of the introduction of Form 1099K information reporting. Brett holds a master's degree from the University of California, San Diego, where he studied International Economics at the School of Global Policy and Strategy.

---

**Devi McKalko** is a tax attorney and accountant at IRS's Research, Applied Analytics, and Statistics division. She received her law degree with a focus on taxation from the University of Denver and her bachelor's and master's degrees in accounting from the University of Hawaii.

**William Boning** is a Financial Economist in the Office of Tax Analysis at the U.S. Department of the Treasury. His research focuses on the role of technology in tax administration, on stimulus payments, and on tax enforcement. He received his Ph.D. from the University of Michigan.

### Session 3: Trusting the Tax Man: Metrics, AI, and Audits

#### Measuring Success: New Performance Metrics for a New Internal Revenue Service

**Janet Holtzblatt** (Urban-Brookings Tax Policy Center)

In April 2023, the Internal Revenue Service took a major step forward with its release of a strategic operating plan detailing how the agency will invest the ten-year \$80 billion boost to its budget provided under the Inflation Reduction Act (IRA). That plan had been requested by Treasury Secretary Janet Yellen immediately after the passage of IRA. Secretary Yellen also instructed the IRS to identify metrics for measuring performance.

Although the IRA funding was cut by over 25 percent a month later in the Fiscal Responsibility Act, the IRS has maintained its commitment to the full plan. The plan is a serious and comprehensive effort to bring the agency into the 21st century, but there are still difficult lingering questions that will have to be addressed within the next few years.

First, the long-term plan is not fully spelled out. The budget boost is a ten-year investment, but most of the strategic operating plan provides details for just the next few years. Many features of the plan involve recruitment, research, evaluation, and pilot programs—the first steps toward development and implementation of effective long-term strategies.

Second, the strategic plan contains objectives and a summary of “what success would look like” for each new initiative. However, it does not provide specific metrics or targets for evaluating the agency's performance in achieving many of the goals of the plan.

Given the early stages of development of the plan, holding the IRS to tough metrics would be premature. Establishing targets now could discourage efforts to test different approaches to determine which is the most effective and fair to taxpayers.

But identifying and designing serious performance metrics can begin in advance of implementation of the various initiatives. In this paper, I establish five principles for designing metrics.

*Janet Holtzblatt is a senior fellow at the Urban-Brookings Tax Policy Center. Holtzblatt's recent research focuses on tax administration, the tax treatment of families and workers, racial disparities in the individual income tax, and wealth taxes. Before joining the Tax Policy Center, Holtzblatt was the unit chief for tax policy studies in the Tax Analysis Division of the Congressional Budget Office.*

---

### Research on Audit Rates by Race and Ethnicity: 2024 Update

**Tom Hertz** (IRS, RAAS)

This presentation updates our prior report on audit rate disparities by race and ethnicity. We summarize findings on some of the drivers of disparity and discuss some of the methodological research we are conducting to improve our ability to estimate outcomes by race and ethnicity. We present estimates of audit rates for Black, Hispanic, Asian-American, and White taxpayers. We describe the work we have done to reduce the frequency of missing parental social security numbers and present some initial findings from our audit selection pilot program. For this pilot, we have developed a suite of machine-learning based audit selection models that are intended to simultaneously improve exam outcomes and reduce racial disparity. The pilot will also permit us to observe the effects of omitting nonrespondent taxpayers from model training datasets and will generate baseline data for future iterative model refinements.

*Tom Hertz is an economist who has been with IRS – RAAS since 2019. He has over 35 years' experience in empirical social science research on topics including the determinants of health outcomes, the role of race and gender in labor markets, intergenerational mobility, immigration and agricultural labor, and the economics of education. He currently leads the RAAS-RICS collaboration on Exam Disparity which is charged with building an understanding of the drivers of differences in audit rates by race.*

---

### Tools to Promote Trustworthiness in a Prototype AI System at the IRS

**Michael Szulczewski**, Mike Feldman, Steffani Silva (MITRE); Brandon Anderson, Alissa Graff (RAAS)

The Internal Revenue Service (IRS) is exploring the use of artificial intelligence (AI) to better identify the risk of tax noncompliance. While federal guidance directs agencies like the IRS to use AI in a manner that fosters public trust, there are few tools for assuring trustworthy AI that are standardized across the federal government and that can be implemented in AI projects. Here, we consider a prototype AI system we developed at the IRS and explore tools including documentation and software that promote trust in the system. We outline the system, identify stakeholders, define goals for AI trustworthiness based on their needs and federal guidance, and describe the development of tools to satisfy those goals. This study informs and advances the adoption of trustworthy AI by identifying trustworthiness tools, explaining adoption challenges, and demonstrating an approach to overcome those challenges for a real-world use case.

*Michael Szulczewski, PhD is a Principal Engineer at the MITRE Corporation and has worked in software engineering, AI/ML, and modeling and simulation for more than 15 years. He previously led the U.S. Computational Modeling Technology Team for Saudi Aramco—the 2nd-largest public company in the world in 2023—and has developed AI/ML models in a variety of fields ranging from tax assessment to thermodynamics. He holds a BS, an MS, and a PhD from the Massachusetts Institute of Technology, and has authored a patent, over 30 conference talks, and 17 peer-reviewed journal articles.*

---

**Melissa Vigil** is an Economist at IRS's Research, Applied Analytics, and Statistics division.



**Arnstein Øvrum** is an Analyst at the Norwegian Tax Administration. His work focuses on impact assessments and the use of predictive models for tax enforcement.

#### Session 4: Simplifying the Filing Burden

##### Technical Challenges in Maintaining Tax Prep Software with Large Language Models

**Sina Gogani-Khiabani**, Varsha Dewangan, Ashutosh Trivedi (CU Boulder); Nina Olson (Center for Taxpayer Rights); Saeid Tizpaz-Niari (UT El Paso)

As the US tax law evolves to adapt to ever-changing politico-economic realities, tax preparation software plays a significant role in helping taxpayers navigate these complexities. The dynamic nature of tax regulations poses a significant challenge to accurately and timely maintaining tax software artifacts. The state-of-the-art in maintaining tax prep software is time-consuming and error prone as it involves manual code analysis combined with an expert interpretation of tax law amendments. We posit that the rigor and formality of tax amendment language, as expressed in IRS publications, makes it amenable to automatic translation to executable specifications (code). Our research efforts focus on identifying, understanding, and tackling technical challenges in leveraging Large Language Models (LLMs), such as ChatGPT and Llama, to faithfully extract code differentials from IRS publications and automatically integrate them with the prior version of the code to automate tax prep software maintenance.

**Sina Gogani-Khiabani** is a Ph.D. student in Computer Science at the University of Texas at El Paso. He focuses on software engineering with an emphasis on artificial intelligence, fairness in AI, and large language models. Sina's research aims to advance the understanding and development of ethical AI software solutions.

---

##### Rethinking Tax Information: The Case for Quarterly 1099s

**Kathleen DeLaney Thomas** (UNC School of Law)

Congress has recently expanded information reporting requirements for third-party settlement organizations (“TPSOs”) by significantly lowering the reporting threshold from \$20,000 to \$600 (as well as eliminating the requirement for 200 or more transactions). Although the new \$600 threshold was enacted in 2021, the IRS has announced it will delay enforcement until at least 2025, using the old \$20,000/200 transactions threshold for 2023 and phased \$5,000 reporting threshold for 2024.

The IRS’s delayed implementation of the new \$600 reporting threshold for TPSOs illustrates a tension in tax administration. On the one hand, more information is generally better for tax enforcement. Subjecting more taxpayers to information reporting means that more individuals should be deterred from cheating and should report their income accurately. However, casting a wider net imposes costs. More information returns impose a greater burden on the IRS to process those returns, as well as greater costs on the third parties that must issue the returns. Casting a wider net among taxpayers may also increase the chances that nonreportable income shows up on information returns (for example, gross proceeds from casual sales that do not exceed basis), increasing complexity and confusion among taxpayers. Thus, in setting a threshold for information reporting, policymakers face a tradeoff between these costs and the foregone revenue that results from unreported income.

The current approach to enhancing tax enforcement through information reporting has been to expand its use through either lowering the reporting threshold (as in the recent case of TPSOs) or widening the scope of third parties required to report. Either approach generally results in more information returns issued to more taxpayers. However, there is a third approach that has received virtually no attention in the United States: policymakers could also increase the frequency and efficacy of tax information sent to taxpayers. More specifically, Congress could require information returns to be sent quarterly – on what this article calls a Form 1099-ES – to align with taxpayers’ estimated tax payment deadlines. While receiving quarterly tax information would likely help taxpayers make timely estimated tax payments, this approach is also not without costs. Third parties would have an increased burden to compile and distribute tax information four

times rather than once a year. And although the IRS would not have to process quarterly information returns (which would be sent only taxpayers), it would have to enforce a requirement to send quarterly returns (for example, by imposing penalties on third parties who fail to do so). This article will explore the tradeoffs between the current approach of expanding the scope of information reporting with an approach that requires more frequent information.

**Kathleen DeLaney Thomas** is the Aubrey L. Brooks Distinguished Professor of Law at the University of North Carolina School of Law. Her teaching and research interests include federal income taxation, tax policy, and contracts law. Prior to joining the UNC Law faculty, Thomas practiced tax law at Simpson Thacher & Bartlett LLP (2005-2007) and Cooley LLP (2007-2011), and was an Acting Assistant Professor of Tax Law at NYU School of Law (2011-2013).

---

### Investigating the Impact of Free E-File Letter Intervention on Taxpayer's Tax Filing and Preparation Methods

**Pei-Hua Chen**, Astin Cornwall, Anne D. Herlache, Scott Leary, Brenda Schafer, Melissa Vigil (IRS, RAAS); Rizwan Javaid (IRS Taxpayer Experience Office)

This study investigates the effectiveness of interventions in promoting e-filing adoption, particularly among taxpayers who were likely eligible to file using the Free file program (generally, having adjusted gross income  $\leq$  \$73,000) in tax Year 2021. Two outcomes are of interest: 1) e-filing vs. paper filing and 2) choice of tax preparation method (Free File; paid preparer; software; VITA; paper; software-prepared, paper-filed). The study focuses on two taxpayer segments, or strata: 1) Frequent filers and 2) New or infrequent filers. Participants were randomly assigned to one of three groups: Free File Letter, Checklist, or Control (no contact).

The results showed that taxpayers who received either treatment were more likely to choose e-filing compared with those who did not receive the mailings. One result of note is that younger (under 30) and older (over 75) taxpayers were more likely to e-file after receiving the letter. The former matches expectations; the latter is surprising and could potentially reflect informal assistance in e-filing. Regarding tax preparation method, we see interesting nuance across strata and covariates (e.g., return complexity, income, etc.). Among other things, the results showed that the income tax complexity affected taxpayers' tax preparation method differently depending on their filing experience. Frequent filers with potentially more complex returns tended to utilize tax preparation software, perhaps for its automated features and potential assistance with complex tax situations. Infrequent or new filers who might have less experience with the tax filing process were more likely to seek professional help from paid preparers. Overall, the study provides insights into effectively encouraging the take-up of e-filing.

*Dr. Pei-Hua Chen, a passionate behavioral scientist with more than 18 years of research experience, has always been fascinated by how people make decisions. This curiosity has translated into real-world solutions that bridge the gap between academia and policy. Her expertise spans interventions, psychometrics and statistics, algorithms, and human factors (e.g., automated test assembly for Defense Language Institute to improve efficiency and fairness in their form construction process). She is currently working on a pilot intervention to increase taxpayers' participation in e-filing. Dr. Chen also has a keen interest in exploring the human and managerial aspects of AI.*

---

**John Guyton** is the Senior Advisor for Knowledge Development and Application in Research, Applied Analytics, and Statistics at the IRS. John has a PhD from the University of Maryland, over 15 years of experience at the IRS, and prior experience in a Big 4 tax practice. John has published and led research on a wide range of tax topic areas ranging from policy to application, outreach to enforcement, low-income taxpayers to the largest corporations. John's current research and program management focus is reporting compliance and examination.

**Robert Weinberger** is a nonresident fellow at the Urban-Brookings Tax Policy Center where his research focuses on tax administration. Before joining TPC in 2018, he spent eight years as a senior fellow at the Aspen Institute's Financial Security Program following a career in government and law including 13 years as head of H&R Block's Washington Office and its Policy Research Institute.