Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Improving Linkages to Individual Income Tax Data

Amy O'Hara
Stephanie Straus
Maanasa Vatsavayi
Nathan Wycoff

Massive Data Institute
McCourt School of Public Policy
Georgetown University

IRS/TPC Joint Research Conference
June 13, 2024

Background

- Statistics of Income (SOI) and other groups within IRS and Treasury need to link tax records
  - Within and across tax years
  - Join external files
- State and local agencies seek linkages, too
  - Measure program outcomes
  - Improve benefits access
- Explore ways to standardize the process
  - Strict schema requirements
  - Seek automated and scalable linkage methods

3

Secure Query System (SQS)

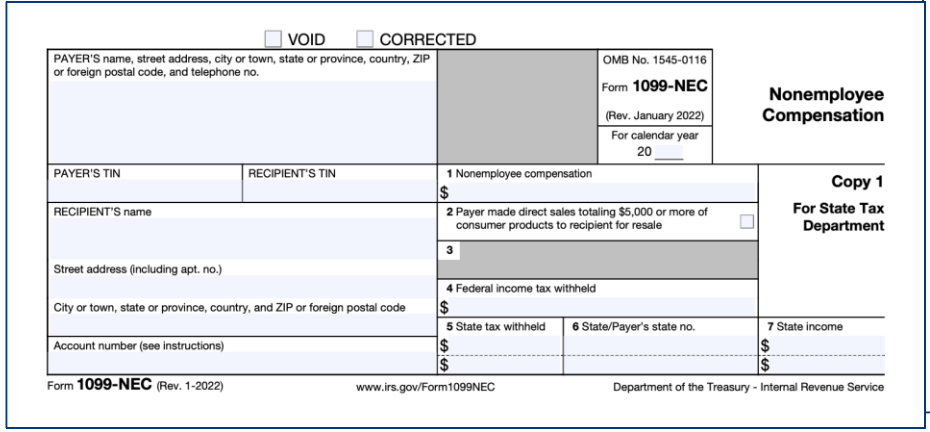
- SOI considering designs for SQS
- System linking end-users (clients) of the data, a data intermediary, and SOI, featuring:
  - Data validation on client side
  - Administrative functions handled by intermediary
  - Automated matching process within SOI, by SOI employees
  - Tabulation of pre-defined statistics
  - Automated disclosure avoidance review

Exploring Linkage Strategies to Individual Tax Data

- Research goals
  - Conduct linkages with and without SSNs
  - Using multiple combinations of personal identifiers
  - Exact and probabilistic matching methods
  - Consider range of data quality and completeness that client input files may contain

Need for Person-level Linkages

- Client data to Form 1099-NEC and Form W-2
- Client data to Form 1040s


VOID CORRECTED

PAYER'S name, street address, city or town, state or province, country, ZIP or foreign postal code, and telephone no.

OMB No. 1545-0116

Form 1099-NEC

(Rev. January 2022)

For calendar year 20___

Nonemployee Compensation

PAYER'S TIN | RECIPIENT'S TIN

1 Nonemployee compensation $

Copy 1 For State Tax Department

RECIPIENT'S name

2 Payer made direct sales totaling $5,000 or more of consumer products to recipient for resale

3

Street address (including apt. no.)

4 Federal income tax withheld $

City or town, state or province, country, and ZIP or foreign postal code

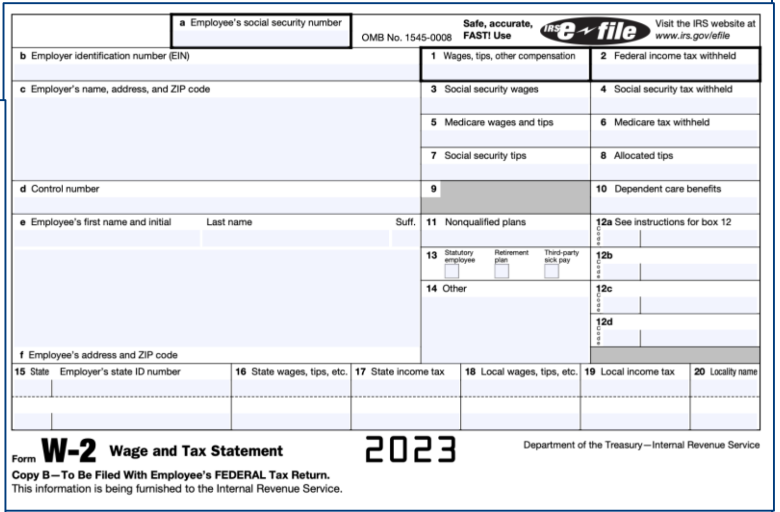
5 State tax withheld $ $

6 State/Payer's state no.

7 State income $ $

Account number (see instructions)

Form 1099-NEC (Rev. 1-2022) www.irs.gov/Form1099NEC Department of the Treasury - Internal Revenue Service


a Employee's social security number

OMB No. 1545-0008

Safe, accurate, FAST! Use IRS e-file

Visit the IRS website at www.irs.gov/efile

b Employer identification number (EIN)

1 Wages, tips, other compensation

2 Federal income tax withheld

c Employer's name, address, and ZIP code

3 Social security wages

4 Social security tax withheld

5 Medicare wages and tips

6 Medicare tax withheld

7 Social security tips

8 Allocated tips

d Control number

9

10 Dependent care benefits

e Employee's first name and initial Last name Suff.

11 Nonqualified plans

12a See instructions for box 12

13 Statutory employee Retirement plan Third-party sick pay

12b

14 Other

12c

12d

f Employee's address and ZIP code

15 State Employer's state ID number

16 State wages, tips, etc.

17 State income tax

18 Local wages, tips, etc.

19 Local income tax

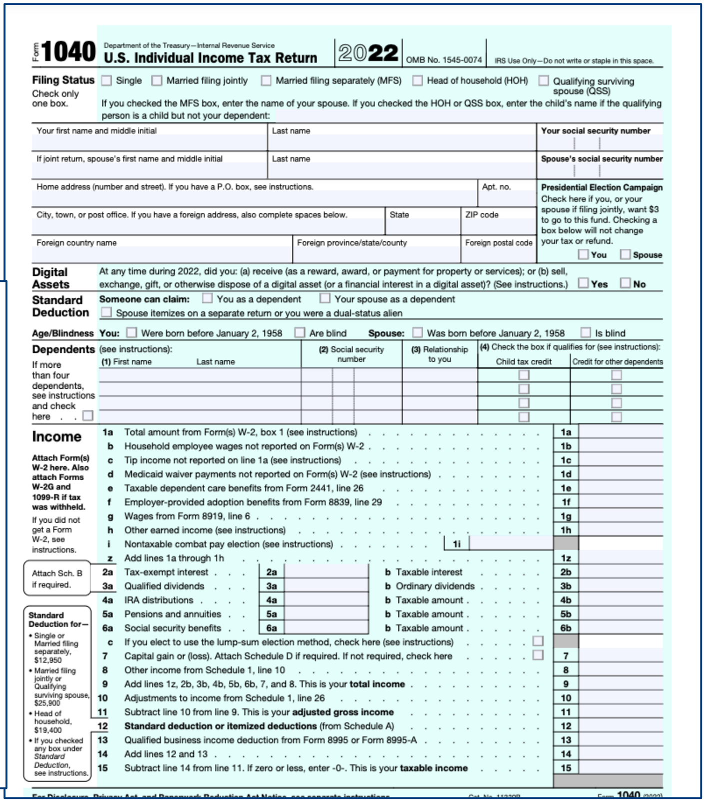
20 Locality name

Form W-2 Wage and Tax Statement 2023

Department of the Treasury—Internal Revenue Service

Copy B—To Be Filed With Employee's FEDERAL Tax Return.

This information is being furnished to the Internal Revenue Service.


Form 1040 Department of the Treasury—Internal Revenue Service U.S. Individual Income Tax Return 2022 OMB No. 1545-0074 IRS Use Only—Do not write or staple in this space.

Filing Status: Single; Married filing jointly; Married filing separately (MFS); Head of household (HOH); Qualifying surviving spouse (QSS)

Your first name and middle initial | Last name | Your social security number

If joint return, spouse's first name and middle initial | Last name | Spouse's social security number

Home address (number and street). If you have a P.O. box, see instructions. | Apt. no. | Presidential Election Campaign

City, town, or post office. If you have a foreign address, also complete spaces below. | State | ZIP code

Foreign country name | Foreign province/state/county | Foreign postal code

Digital Assets

Standard Deduction

Age/Blindness

Dependents

Income

1a Total amount from Form(s) W-2, box 1 (see instructions)

1z Add lines 1a through 1h

2a Tax-exempt interest | b Taxable interest

3a Qualified dividends | b Ordinary dividends

4a IRA distributions | b Taxable amount

5a Pensions and annuities | b Taxable amount

6a Social security benefits | b Taxable amount

7 Capital gain or (loss)

8 Other income from Schedule 1, line 10

9 Add lines 1z, 2b, 3b, 4b, 5b, 6b, 7, and 8. This is your total income

10 Adjustments to income from Schedule 1, line 26

11 Subtract line 10 from line 9. This is your adjusted gross income

12 Standard deduction or itemized deductions (from Schedule A)

13 Qualified business income deduction from Form 8995 or Form 8995-A

14 Add lines 12 and 13

15 Subtract line 14 from line 11. If zero or less, enter -0-. This is your taxable income

IRS Elements Available

- Form 1040
  - SSN, first name (FN), middle initial (MI) and last name (LN), house number and street name, apartment number, city, state, and zip code
- Form W-2
  - SSN, FN and MI, LN, and address in one field (not separating house number/street address, city, state, and zip code)
- Form 1099-NEC
  - Recipient TIN, name (in one field), street address including apartment number, and a single field for city, state, and zip code

7

Challenges with IRS Data

- Challenges
  - Amended returns
  - Late returns (current mailing address rather than their address from the earlier tax year
  - Information returns are submitted to IRS by the employer or payer, reflecting the address known to those entities
  - Multiple job holdings generate multiple W-2s and 1099-NECs with discrepant info

Expected SQS Client Data Elements

- SSN
- FN, MI, middle name (MN), LN
- Address (at time of service/participation/enrollment)
- Some organizations also have DOB, spouse, and parent/guardian information (for minors)

Expected SQS Clients

Higher Education Institutions, State and Local Education Agencies, Education Research Organizations, State and Local Workforce Agencies, Registered Apprenticeship Programs, State and Local Corrections Agencies, State and Local Health and Human Services Agencies, Public Housing Agencies, non-profit and research organizations

Synthetic Data to Test Match Strategies

- Pseudopeople dataset (Haddock et al., 2024)
  - Generated demographic dataset mimicking adult population of US at various life stages
  - Random sample of 10,000 ‘Connecticut’ records containing simulated 1099s and 1040s
  - Mild corruption – blank 20% of SSNs and corrupt 5% of remaining SSNs (fill with 0s, 9s, remove 1-2 digits, etc)
  - Moderate corruption – insertion, deletion, transposition and substitution errors; introducing misspellings in last names; miskeying/mishearing errors

Matching Program and Approach

- Splink (Linacre, 2022)
  - Open-source linkage package that uses the Fellegi-Sunter model (1969) to conduct probabilistic record linkages with user-specified blocking and matching rules.
  - Probabilistic and exact matching on 22 combinations of identifiers
    - SSN, Full LN, 4char LN, Full FN, 2char FN, FI, Full MN, MI, Street name, Age
  - Blocking – ZIP5 and ZIP3
  - Combinations based on patient matching literature (Deng et al, 2023) and National Center for Advancing Translational Sciences

Testing Match Passes

- Took two corrupted Pseudopeople 1040-like datasets
- Matched to uncorrupted 1040 and 1099 data
- Evaluated each match pass using true pair identifiers
- Evaluated performance using precision and recall (Hastie et al., 2009)

Preliminary Results

- Successful match passes
  - Exact match on SSN alone
  - Fuzzy match on LN, 2char FN, MI, age without blocking
  - Block on ZIP3, fuzzy match on LN, FN, MI
  - Block on ZIP3, fuzzy match on LN, FN, age

Next Steps

- Test match passes on larger datasets
  - Within state
  - Across states
- Test approach for lagged matches
- Propose match output statistics to produce for clients
- Assess capacity building needs for name and address standardization and parsing for state and local agencies

References

- Haddock, B., Pletcher, A., Blair-Stahn, N., Keyes, O., Kappel, M., Bachmeier, S., & Flaxman, A. (2024). Simulated data for census-scale entity resolution research without privacy restrictions: a large-scale dataset generated by individual-based modeling. Gates Open Research, 8, 36.
- Linacre, R., Lindsay, S., Manassis, T., Slade, Z., & Hepworth, T. (2022). ApplyiSplink: Free software for probabilistic record linkage at scale. International Journal of Population Data Science, 7(3).
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
- Ehrgott, M. (2005). Multicriteria optimization (Vol. 491). Springer Science & Business Media.


amy.ohara@georgetown,edu
stephanie.straus@georgetown.edu
GEORGETOWN UNIVERSITY
McCourt School of Public Policy
MASSIVE DATA INSTITUTE

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

A Large Scale, High Quality U.S. Occupational Database: Results from Merged ACS and IRS Write-Ins

Julia Beckhusen, Victoria L. Bryant, David B. Grusky, Thomas N. Hertz, Michael Hout, Liana Christin Landivar, Lynda Laughlin, Ananda Martin-Caughey, Javier Miranda, Kevin Pierce, Carl Sanders

IRS-TPC Conference, 13 June 2024

Purpose & motivation

- Worker occupation is a key driver in economic growth (Violante 2008), career progression (Yamaguchi 2011), and cross-sectional and intergenerational inequality (Card and DiNardo 2002, Long and Ferrie 2013).
- Universe-level occupation data available in some countries (e.g. Denmark), but administrative and data collection difficulties in the U.S.
- Census: American Community Survey
- IRS: Form 1040 “Occupation” field

Contribution

- Create near-universe dataset of coded worker occupations
  - Match e-filed Form 1040s and 1-Year ACS
- Evaluate quality of matched IRS/ACS write-ins
  - Token similarity
  - Semantic similarity
- Create a Large Language Model-based autocoder mapping text write-ins to Census 2018 occupation codes.
- (Preliminary) Evaluate cross-sectional and longitudinal accuracy of IRS occupational distribution

United States® Census Bureau

Data

- American Community Survey 2019 1-Year Microdata (ACS) write-ins
- IRS Tax Year 2018 Form 1040 write-ins

ACS and IRS Occupation Prompts

ACS

e. What was this person's main occupation? *(For example: 4th grade teacher, entry-level plumber)*

Driver

f. Describe this person's most important activities or duties. *(For example: instruct and evaluate students and create lesson plans, assemble and install pipe sections and review building plans for work details)*

Pick people up in my car,
drive them where they need to
go, and drop them off

F1040

Sign Here

Under penalties of perjury, I declare that I have examined this return and accompanying schedules and statements, a
belief, they are true, correct, and complete. Declaration of preparer (other than taxpayer) is based on all information o

Your signature	Date	Your occupation
Spouse's signature. If a joint return, **both** must sign.	Date	Spouse's occupation

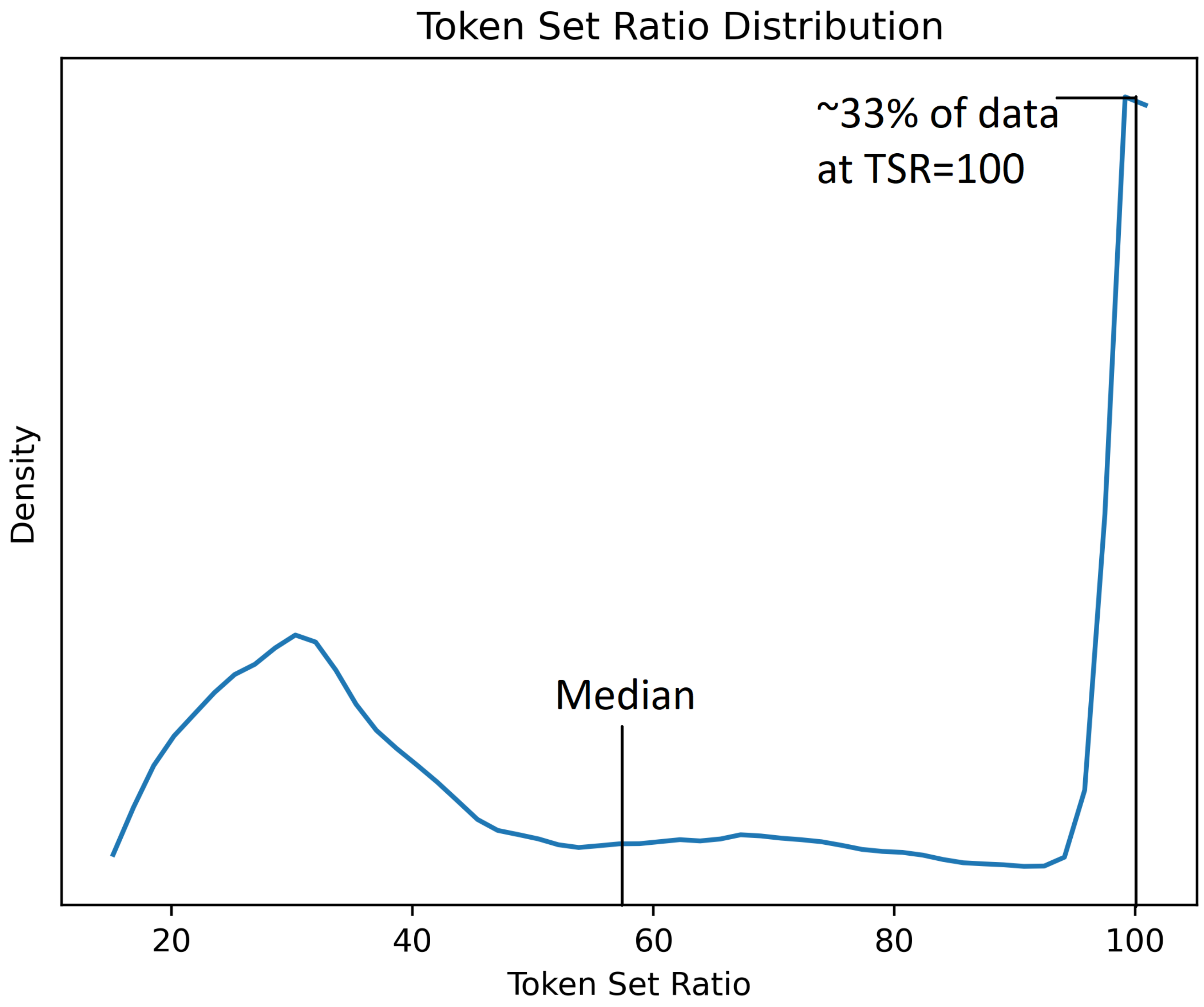
Joint return?
See instructions.
Keep a copy for your records.

United States® Census Bureau

Token Similarities

- Token Set Ratio: 0-100 score of similarity of two strings

- TSR(“Lawyer”, “Lawyer”) = 100
- TSR(“Clown”, “Teacher”) = 17
- TSR(“Lawyer”, “Attorney”) = 29
- TSR(“Paralegal”, “Paramedic”) = 56



United States® Census Bureau

Transformer-based Autocoder

- BERT (Bidirectional Encoder Representations from Transformers) architecture for Large Language Modeling
  - Open Source LLM, pretrained on Wikipedia and the Toronto BookCorpus (3.3 billion words)
  - Maps a text string to a numerical vector representation (“encoding”).
- Occupational coding problem estimated as a Multinomial Logit with 565 choices
- Inputs: text writein -> BERT encoding, industry category
- Target: assigned 2018 Census occupational code (565 categories).

United States® Census Bureau

Estimation Results

Model	Match Rate	Top 2	Top 10
ACS LLM Text + Industry	**0.81**	**0.90**	**0.97**
IRS LLM Text + Industry	**0.42**	**0.54**	**0.77**

Source: U.S. Census Bureau, 2019 American Community Survey 1-year and IRS Form 1040 Tax Year 2018

Semantic Similarity

- The ACS and IRS model each predict a probability distribution
- **Total Variation Distance** (TVD) between them measures prediction disagreement
- Results from TVD broadly agree with results from token-based analysis
- Approx. 50% paired entries semantically similar, approx. 33% high quality semantic matches

Agency Benefits

- IRS:
  - Fully coded occupational field
  - Response quality control via ACS comparisons
- Census:
  - Show feasibility of Open Source, Machine Learning-based occupation coding
  - Improved imputes for missing records

Conclusion

- Creating a near-universe file of coded occupations from Form 1040 write-ins is feasible when combined with ACS data.
- Economically significant information in IRS write-ins, but measurement challenges remain.
- Next steps: aggregation; years 2011-2018.

Funding: Russell Sage Foundation [Hout & Grusky]

Thank you!

Carl Sanders
Carl.E.Sanders@Census.gov

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Disaggregating Tax Compliance Burden: A Comparative Study

Biz Bedane - IRS-RAAS
June 13, 2024

Overview

- Introduction
- Tax Compliance: Concepts, Methods, and Challenges
- Tax Compliance Cost and Structure: Empirical Evidence
- Comparison of Individual and Business Taxpayers Compliance Cost: Case Study
- Conclusion

Introduction

- **What is tax compliance cost?:** Tax compliance cost is the sum of out-of-pocket expenses and the imputed value of time and resources (internal and external costs).
- **Objective:** To conduct a comparative examination of tax compliance costs incurred by individuals and business taxpayers.
- **Data Source**: Administrative data and published literature
- The study examined the conceptual underpinnings and methodological challenges and compared and contrasted U.S. taxpayers' tax compliance burden with that of the U.K., Australia, Canada, and Germany.

Main Findings

- Tax compliance studies face numerous challenges such as data scarcity, non-response bias, questionnaire framing issues, and monetization of compliance time.
- Tax compliance costs are regressive with firm size and income.
- Individual taxpayers’ compliance cost in the U.S. are higher than Germany and Canada, while small businesses’ compliance cost are lower than those of Australian and the U.K.

Tax Compliance: Concepts

Social Cost vs Taxpayer Compliance Costs (Tran-Nam et al., 2000)

- **Social costs** encompass efficiency loss (deadweight loss), administrative expenses, and compliance costs
- **Tax compliance costs** include out-of-pocket expenditures plus the imputed value of time and resources minus the benefits of tax compliance
- **Administrative costs** denote the government's expenses in tax collection

Total Taxpayer Burden: (Guyton et al., 2003)

- **Total burden** is tax liability and excess burden
- **Excess burden** is compliance, psychological, and efficiency costs
- **Compliance burden** comprises out-of-pocket payments, time, psychological, and efficiency costs
- **Psychological costs** refer to the dissatisfaction, frustration, and anxiety stemming from interactions with the tax system, which are challenging to quantify
- **Efficiency loss** results from tax-induced distortions, leading to a change in consumer and producer surplus, which are difficult to measure and often omitted from compliance cost assessments
- Generally, tax compliance costs include expenses by taxpayers to fulfill their tax obligations, preparing and filing time, and out-of-pocket outlays

Tax Compliance: Methods

The Standard Cost Method (SCM) - Used across the European Union and defines compliance costs to include all expenses related to adhering to regulations, except for direct financial costs and long-term structural impacts.

- **Advantages:** It is versatile for impact assessments, including cross-border transactions, relevant to all forms of taxes and legislative frameworks, supports segmentation, and facilitates comparisons between countries
- **Drawbacks:** Issues with representativeness, failure to consider temporary compliance costs, and excluding non-mandatory expenses like those for tax planning

The World Bank: Evaluates the ease of tax compliance across 189 economies. Tax burden is measured by the hours spent annually on tax preparation, filing, and payment.

- **Advantage**: Provides consistency (Pedersen et al., 2013) and a substantial volume of expert estimates (Eichfelder and Vaillancourt, 2014).
- **Drawbacks:** Data does not distinguish between micro, small, medium, and large firms, preventing any inference about how compliance costs might vary across different-sized businesses (D'Andria and Heinemann, 2023).
- In some developing countries, the methodology has faced criticism for producing unrealistically large figures (Eichfelder and Vaillancourt, 2014), and irregularities have been documented (D'Andria and Heinemann, 2023).

Tax Compliance: Methods

The Internal Revenue Service (IRS)

- Conducts the Individual Taxpayers Burden (ITB) and Business Taxpayers Burden (BTB) surveys since 1984
- **ITB** - Surveys were conducted in 1984, 1999 (for Wage and Investment taxpayers only), 2000 (specifically for self-employed taxpayers), 2007, and annually since 2010
- ITB Surveys categorized tax returns by preparation method and then further stratified within these categories based on five complexities levels
- **BTB -** Conducted in 1984, 2004, 2009, and 2012, with plans for subsequent surveys to occur annually or every three years
- The IRS conducted simulations using the ITBM, SBBM (Contos et al., 2009), and BTBM. The IRS Taxpayer Burden Model (**TBM**) was developed in 2002 and updated in 2010, employs a log-linear model specification.
- The dependent variable, the logarithm of compliance cost, is estimated as a function of various independent variables. The model controls the type and volume of taxpayer activities (Guyton et al., 2023).
- **Advantage**: Representative data and employs a robust methodology.
- **Drawback**: IRS survey is respondents' inability to differentiate the time used to prepare their federal and state tax returns.

Tax Compliance: Challenges

Data availability

- Studies rely on surveys, qualitative interviews, case studies, and administrative data. Lack of panel data make comparison over time and across observations impossible.
- Hsiao (2007 and 2022) noted that panel data increase degrees of freedom and facilitate more precise inference of model parameters, control for unobserved individual, and time heterogeneity which strengthen statistical inference.

Survey Design (framing issues)

- A study using Belgian business data found that framing temporal aspects of cost measurement (annually versus monthly) could drastically change estimates. For small businesses, estimates could be reduced by as much as 53% or increased by up to 112%, with an average change of 39% downward or 65% upward (Eichfelder and Hechtner, 2016).

Non-response Bias

- Lignier et al., (2014), Evans et al., (2013), Schoonjans et al., (2011), Brick et al. (2010), Contos et al., (2012), and Smulders et al., (2012) highlighted the critical role of addressing non-response bias, which stems from systematic differences between those who respond to surveys and those who do not.
- Evans et al., (2013) and Tran-Nam et al., (2014) employed wave analysis to tackle non-response bias.
- Slemrod and Venkatesh (2002), Blaufus et al., (2014), and Blaufus et al., (2019) calculated a set of weights.

Monetization of Compliance Time

- Constant cost based on the average market wage (Schoonjans et al., 2011)
- Applying variable monetization rates (Contos et al., 2012)
- Charging the hourly rates of external service providers as seen in the EU Standard Cost Model (Pedersen et al., 2013)
- Using valuations reported by respondents themselves (Smulders et al., 2012 ;Evans et al., 2016).

Evidence:

- Contos et al., (2012) examined using variable monetization rates ranged from $8 to $90 per hour and the fixed monetization rate was $28.73.
- They found that the average compliance cost for U.S. businesses was $11,600 using variable rate monetization and $10,300 using constant rate monetization, as estimated through the Business Taxpayers Burden Model (BTBM).

Tax Compliance Cost and Structure: Empirical Evidence

Individual Taxpayers' Compliance Costs - Selected Studies (2003-2024)

Study	Country	Sample size	Response rate	Time burden (in hours)	Average Cost	Tax year
Guyton et al (2003)	USA	All = 15447 W&I = 6366 SE= 9081	W&I = 60.6% SE = 56.4%	All=25.5 hours [a]W& I = 13.8 hours SE[b] = 59.5 hours	Total= $149 per taxpayer W& I = $75 SE = $363	W& I=1999 SE = 2000
Mathieu, L., Price, C., and Antwi, F.(2010)	UK	320	32%	8 hours	£498($329)*	2000
Marcuss et al, 2013	USA	7685	42%	12.5 hours	$373	2010 ITB survey
Blaufus, Eichfelder, and Hundsdoerfer (2014)	Germany	629		All=9.8-14.4 hours EM:7.1-8.8 SE:20.6-35.9	€298 ($217.5) to €450($328.5)**	2007
Tran_Nam, et al(2014)	Australia	517	13.4%		AUD 796.585($773)***	2011-12
Blaufus,K.,Hechtner,F., Jarzembski,J.(2019)	Germany	18,196	0.54%	9.13-10.23 hours	€228(($205) to €321($289)****	2015
Stark, K. and Smulders, S.(2019)	South Africa	556		29.5 hours	Total=ZAR6905***** ($483) EM=ZAR 3314($231.7) SE=ZAR24416($1707)	2016/17
Vaillancourt and Li(2024)	Canada	1523		1.5 hours	$130	2023

*Annual Average exchange rate:1pound = 0.661 USD(source data.oecd.org)(2000), ** 1 USD = 1.37 euro(2007), ***1USD = 1.03 AUD(2011), ****1 USD = 1.11euro(2015),*****1 USD = 14.3 ZAR(2016/17).

[a]W&I=Wage and Investestment, [b]SE= self employed

Source: https://www.imf.org/external/np/fin/ert/GUI/Pages/CountryDataBase.aspx

Individual Tax Compliance Costs

- Response rate ranges: 0.54% (Germany) to 60.6% (U.S.)
- Time burden: 1.5 hours (Canada) to 29.5 hours (SA)
- Average cost: $130 (Canada) to $773(Australia)
- Sample size: 320 (U.K.) to 18,196 (Germany)

Business Taxpayers' Compliance Costs - Selected Studies (2002-2017)

Study	Country	Sample size	Firm size	Cost per turnover	Average Cost per firm	Response rate
Slemrod, J. and Venkatesh, V.(2002)	USA	443	Large and medium		$134,954	
Contos et al (2009)	USA	7049	Small		$6644	
Contos et al (2012)	USA	22,000	All businesses		$11,600	31.5%
Smulders, S.et al(2012)	South Africa	5865	Small Business		R63 328* ($8722)	6.7%
Hansfor,A. and Hasseldine, J. (2012)	UK	41	Small and medium		£21,362** ($13,330)	<1%
Sapiei,S.,Adbullah,M., Sulaiman.(2014)	Malaysia	98	Small, medium, large	Avg=0.01% Small= 0.057%, large =0.001%	MYR47126*** ($14,411.6)	20.7%
Lingier,P.,Evans,, and Tran-Nam(2014)	Australia	682	Small, micro and medium	14%	AUD 11,004**** ($10,683.5)	7.5%
Evans, Lignier, and Tran-Nam(2016)	Australia	79	Large	0.04%	AUD 1,802,785 ($1,750,276.7)****	42%
Yesegat, W., Coolidge, J, and Corthay,L.(2017)	Ethiopia	1003	All businesses	4.7%	$406	
Stamatopoulos, Hadjidema,Eleftheriou(2017)	Greek	285	Large		€9571($12,710) *****	27.9%

*Annual Average exchange rate 1USD =7.26ZAR(FY2014), **1pound = 0.624 USD(2011)(source data.oecd.org),*** 1USD = 3.27MYR(2014), ****1USD = 1.03 AUD (FY2011), *****1USD = 0.753 euro(2013)(source data.oecd.org)

Source: https://www.imf.org/external/np/fin/ert/GUI/Pages/CountryDataBase.aspx

Business Taxpayers' Compliance Costs

- Sample size: 41 to 22,000.
- Response rate: <1% to 42%.
- Average cost: $406 - $1.75 million.

Tax Compliance Cost and Structure: Empirical Evidence

Estimated Share of Tax Compliance Activities

Handling tax compliance obligation	VAT				CIT			
	Micro	Small	Medium	LSE	Micro	Small	Medium	LSE
Internally	26%	24%	33%	28%	16%	20%	29%	24%
Outsourced(fully+ partially)	74%	76%	67%	72%	84%	80%	71%	76%

N=2 479; Source: VVA / KPMG (2021) in D'ANDRIA, D., & HEINEMANN, M. (2023).

Internal, External, and Non-labor Costs for selected countries

Country	Internal cost	External Cost	Non-labor cost
US	58.7%	24.8%	16.5%
Australia	45.7%	34.2%	20.5%
Greek	30.2%	52.6%	17%

Allocation of Compliance Time for Different Activities by U.S. Businesses in percent (2010 - 2023)

Year	Record-keeping(%)	Tax Planning (%)	Form completion and submission time (%)	All other time(%)
2010	53.1	12.5	21.9	12.5
2011	50.0	12.5	21.9	12.5
2012	56.5	13.0	26.1	4.3
2013	54.2	16.7	20.8	8.3
2014	54.2	12.5	25.0	8.3
2015	54.5	18.2	22.7	9.1
2016	54.5	18.2	22.7	4.5
2017	52.4	14.3	23.8	4.8
2018	52.6	15.8	26.3	5.3
2019	50.0	15.0	25.0	5.0
2020	52.4	14.3	23.8	9.5
2021	54.5	18.2	22.7	9.1
2022	48.0	20.0	24.0	8.0
2023	50.0	16.7	25.0	8.3
Average	52.6	15.6	23.7	7.8

Source:Compiled from 1040 instructions https://www.irs.gov/pub/irs-pdf

The average record-keeping time (2010-2023) allocated by U.S. business taxpayers took half of the total time.

The average form completion and submission time (2010 - 2023) allocated by U.S. individual taxpayers is 37% followed by record keeping at 36%.

Evans, et al. (2014) findings suggest that SMEs from the U.K. and Australia spend two-thirds of their time on recording information, while Canadian and South African businesses spend roughly half of their time on this function.

Tax Compliance Costs and Structure: Empirical Evidence

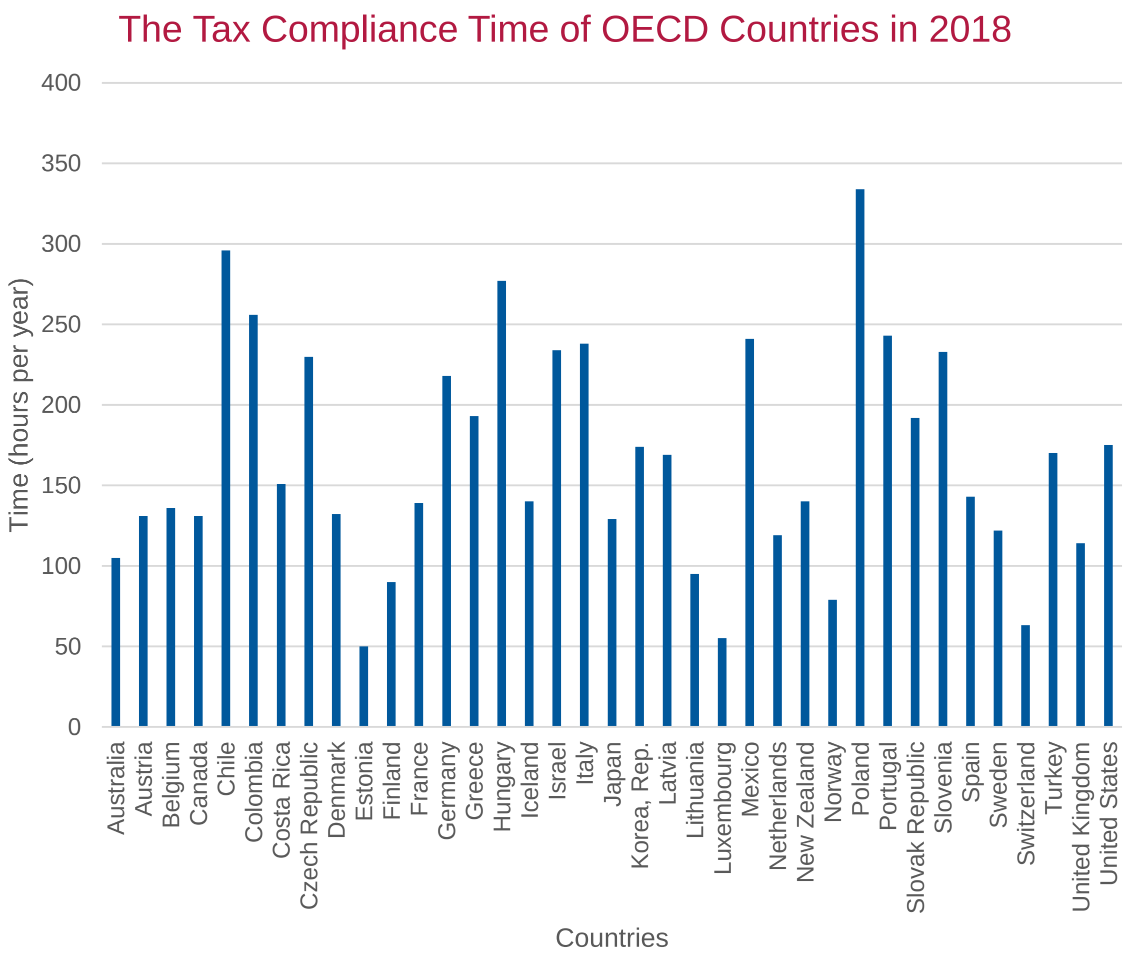
- **Drivers of Compliance Costs**: Income, tax code complexity, and firm size
- **Income**: Berger et al. (2017) confirmed compliance costs, as a percentage of pretax income, are highest for individuals in the lowest income quintile
- **Tax code complexity** increase tax compliance costs (Evans et al., 2016; Blaufus et al., 2019; Lazos et al., 2022; Marcuss et al., 2013)
- Berger et al. (2017) estimated that the tax code's complexity costs individuals over $104 billion in Tax Year 2017, averaging $596 per taxpayer.
- Benzarti (2020) discovered compliance costs influence taxpayers' decisions between itemized and standard deductions
- **Firm size** negatively related to tax compliance costs (Evans et al. 2016; Contos et al., 2012; Evans et al., 2014)

Comparison of Individual and Business Taxpayers Compliance Cost: Case Study

Location	Payments (number per year)	Time (hours per year)
Australia	11	105
Belgium	11	136
Canada	8	131
France	9	139
Germany	9	218
Italy	14	238
Japan	19	129
Netherlands	9	119
United Kingdom	9	114
United States	11	175
East Asia & Pacific	20.6	173.0
Europe & Central Asia	14.4	213.1
Latin America & Caribbean	28.2	317.1
Middle East & North Africa	16.5	202.6
OECD high income	10.3	158.8
South Asia	26.7	273.5
Sub-Saharan Africa	36.6	280.6

Source: World Bank (2018)

The average medium-sized U.S. firm spends 175 hours on tax compliance, higher than the OECD high-income average of 158.8 hours and more than the U.K., Australia, Canada, and Japan, yet less than Germany and Italy.



Case Study: Individual Taxpayers

Country	USA	Australia	Canada	Germany
Year of Study	2013(TY 2010)	2014(FY 2011/12)	2024(TY 2023)	2019(2015 tax year)
Sample size	7685	517	1523	18,196
Response rate	43%	13.4%	NA	0.54%
Compliance cost per taxpayer	$373	A$796.85($773.6)*	$130	€106($96)**
Compliance cost per tax revenue		4.84%	1.2%	
Time	12.5 hours	8.3 hours	1.5 hours	10.6 hours

U.S. compliance costs are shown to be lower than Australia's, but higher than those of Germany and Canada.

*1 USD= 1.03 AUD(2011 average) , **1USD = 1.6 pound (2011)

Source: https://www.imf.org/external/np/fin/ert/GUI/Pages/CountryDataBase.aspx

Case Study: Business Taxpayers

Country	USA	Australia	Australia	UK	Canada	UK	Australia
Year of Study	2009(2012) (T.Y. 2009)	2014 (TY2011)	2016 (FY2011/12)	2012 (TY2011)	2014	2014	2014
Business Type	All businesses	SMEs	Large	SMEs	Small	Small	Small
Sample size	22,000	682	79	41	2449	4420	3500
Response rate	31.5%	7.5%	42%	<1%	1.35%	0.9%	4.5%
Compliance cost per taxpayer	$11600	A$11,004 ($10,683.5)*	A$1,802,785 ($1,750,276.7)*	£21,362 ($13,351)**	$50,286	$36500	$34640
Compliance cost per tax revenue		14%	0.04%				

U.S. SMEs incur higher costs than their Australian counterparts, but lower than those in the U.K.

*1 USD= 1.03 AUD(2011 average) , **1USD = 1.6 pound (2011)

Source: https://www.imf.org/external/np/fin/ert/GUI/Pages/CountryDataBase.aspx

Conclusions

- Tax compliance costs are determined by firm size, income, and tax code complexity.
- Tax compliance studies face numerous challenges, including data scarcity, non-response bias, and variability in the valuation of tax compliance time. Consequently, comparisons between tax compliance studies should be approached with caution.
- This study indicates that tax compliance costs exhibit a regressive pattern, with firm size and income negatively correlated with compliance burdens.
- Individual taxpayers in the U.S. shoulder higher tax compliance costs compared to the countries examined in this study (Germany and Canada). Conversely, compliance costs for small businesses in the U.S. are lower than those in Australia and the U.K.

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

A Gravity Model of Cross-Border Tax Avoidance

13 June 2024

Presented by Internal Revenue Service
Research, Applied Analytics & Statistics

Lori Stuntz, Economist
Michael Udell, Economist

Cross-Border Tax Avoidance

What do we mean by cross-border tax avoidance?

- Relocation of taxable income from locations with taxable real economic activity to locations with very limited economic activity and little or no tax.
- For the gravity model, it does not matter whether this shifting is legal tax avoidance or illegal tax evasion.

How large is cross-border tax avoidance?

- Although there is a lack of consensus on the *amount* of cross-border tax avoidance, there is consensus on its *existence*
    - Beer, de Mooij, and Liu (2020) report avg profits decrease by 1.59% for each 1 percentage point increase in domestic corporate tax rates (avg of 37 papers).
    - Lejour (2021) surveys estimates of annual worldwide corporate tax revenue losses due to avoidance of between $123 and $180 billion during the past decade.
    - Johanessen, Reck, Risch, Slemrod, Guyton and Langetieg (2023) estimate that approximately $2 trillion (2.5%) of US household wealth held in tax haven countries in 2018.

Presentation Outline

Gravity Model for cross-border tax avoidance

- Develop a model based on international trade literature gravity models to identify pathways of financial flows across countries that could facilitate tax avoidance

Data Construction and Sources

- Walk through the various country level data and discuss how we construct country sequences

Index Variable Construction

- How we construct the variables used in the gravity model for cross-border tax avoidance

Weighting the Gravity Model using Foreign Direct Investment (FDI) financial flows

What is a Gravity Model?

General Gravity Model

Used to explain the force of attraction between two bodies

- Attraction might increase with size of each body and decrease with distance

Gravity Models in International Trade

Used to predict bilateral trade flows between two entities

- Attraction (or trade) might increase with economic size and decrease with geographic distance

- could be GDP
- Distance could be miles

Gravity Model for Tax Avoidance

We develop a structural Gravity Model to measure the attractiveness of cross-border tax avoidance

- Ideal dependent variable: measure of tax avoidance across borders
- Current dependent variable: measures of financial flows across borders (FDI)
- Explanatory variables: tax rates in each country; treaty withholding taxes between countries; regulator quality; and measures of tax administrator transparency

Other Important Contributions:

- Uses readily available country level measures that can be updated annually
- Any sequence of countries in any order can be considered

- Adapt the gravity equation to multiple borders (not just two)
  - Allows us to look at sequences of countries with any number of border crossings
  - Model provides distinct measures of attractiveness for each sequence.
  - For example, A -> B -> C -> D could have a different gravity index score than A -> C -> B -> D

Gravity Model Index

Use the index to identify best potential **conduit** or **destination** given a set of observed countries

- **Conduits** facilitate the flow into and out of a country with the lowest tax burden but the greatest number of tax linkages to
- **Destinations** have low (or no) taxes and the least tax transparency / information sharing

We first consider only sequences that originate in the United States (USA origin), and later expand the analysis to sequences that can originate in any country in the world (Worldwide).

Gravity Model for Tax Avoidance

Methodology

Country Data
- Treaty Dividend Withholding Taxes (WHT) for all Country Pairs
- World Bank Data
- Capital Gains Tax Rates
- Exchange of Information Variables

Create Sequences
- Link Countries into Sequences
- Create Index Variables

Weighted Index
- Estimate Gravity Model Index Weights
- Use coefficients to weight the Gravity Index

Stopping Rules
- Eliminate sequences with $0 of Adjusted FDI
- Remove when better to stay in Country 2

Benefits of this approach

- Can be updated each year with new data to detect potential hotspots of activity in real time
- Model only uses aggregate country level data and treaty data
- Does not rely on private taxpayer information
- Can consider any sequence of countries originating in any of the 228 countries in the model

Gravity Model for Tax Avoidance

- Gravity equation for tax avoidance:

$$\frac{DIV.OWN.path^{\beta_1}(1 - CG.ratio_n)^{\beta_2}}{\frac{1}{RQ.path}^{\beta_3}(1 + EOI.ratio)^{\beta_4}}$$

  - **DIV.OWN.path** product of (1 – Dividend withholding tax rate) * (1 – required ownership percentage) across the sequence
  - **CG.ratio** ratio of capital gains tax rate for non-residents in the destination country to the capital gains tax rate for non-residents in the origin country
  - **RQ.path** average value of World Bank Regulator Quality index across all countries in a sequence
  - **EOI.ratio** an index that sorts all 8 possible EOI paths across a three-country sequence; discussed in detail on slide 22.

Gravity Model for Tax Avoidance

- To estimate the $\beta' s$, or the weights, we take the log of the gravity model equation:

$$log(Tax\ Avoidance) = \beta_0 + \beta_1 \log(DIV.OWN.path) +$$

$$\beta_2 \log(1 - CG_n) + \beta_3 \log(\mathrm{RQ.path}) + \beta_4 log \frac{1}{(1+EOI.path)} + \varepsilon$$

- We currently use Inward Foreign Direct Investment (FDI) across the entire path to measure financial flows across each sequence. Inward FDI is investment flowing into a country from a foreign source.
  - **Total.Inward.FDI** is the simple sum of Inward FDI across a sequence
  - **Inward.Adjust** is a concept we derive to measure amounts that could actually flow across an entire sequence (discussed in depth later)
  - In this presentation we focus on **Inward.Adjust**

Data Construction and Sources

International Bureau of Fiscal Documentation (IBFD)

https://research.ibfd.org

- Country level tax features for 230 countries, including the capital gains rate for non-resident individuals (CG) from the Country Tax Guides
- Tax treaty dividend withholding tax (WHT) rates and required minimum ownership percentages from Country Treaty Tables
  - For each country pair, we code up to 4 dividend WHTs and required ownership percentages
  - This yields 59,143 pairs of dividend WHTs

World Bank Global Indicators of Regulatory Governance

Reports governance indicators for six dimensions of governance, including Regulator Quality (RQ) and Political Stability (PS)

Indicator	Count Better than USA	Count Worse than USA	Total Count
RQ	15	194	209
PS	91	120	211

- We normalize values so that each indicator ranges from 0 to 1.
- RQ.path is the average of the index across each country in a sequence

Data Construction and Sources

We consider 4 Exchange of Information (EOI) Variables, each coded as an indicator equal to 1 if the country is a participant

FATCA – Foreign Account Tax Compliance

- Requires foreign financial institutions (FFIs) to report to the IRS information about financial accounts held by U.S. taxpayers, or by foreign entities in which U.S. taxpayers hold a substantial ownership interest.
  - Source: https://home.treasury.gov/policy-issues/tax-policy/foreign-account-tax-compliance-act

EOIR – Exchange of Information upon Request

- Countries with which the U.S. has in effect an income tax or other convention or bilateral agreement relating to the exchange of tax information
  - Source: Rev. Proc. 2021-32, Section 3 (page 3)

Data Construction and Sources: EOI

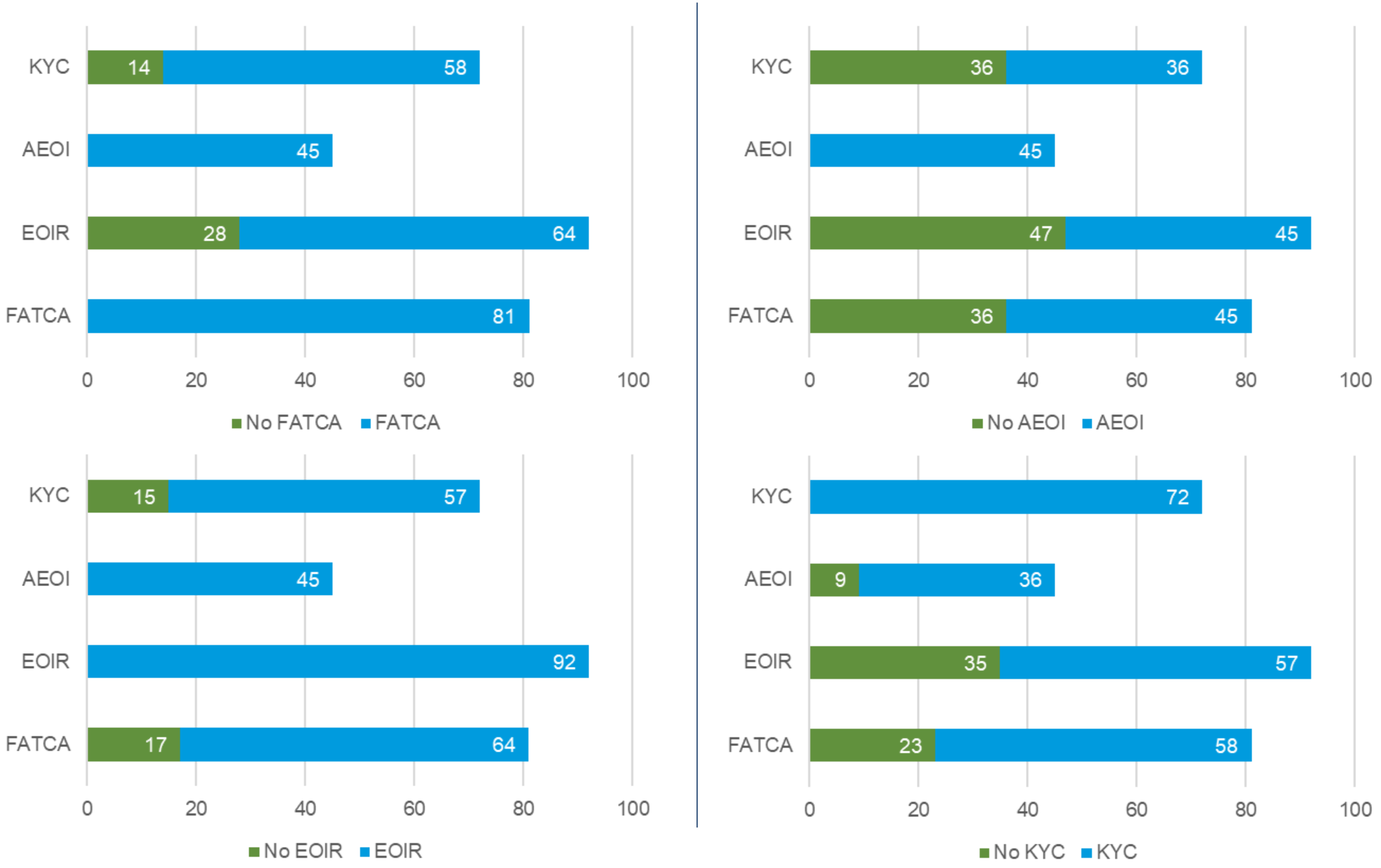
AEOI – Automatic Exchange of Information

- Countries with which the Treasury Department and the IRS have determined that automatic exchange of deposit information is appropriate.
  - Source: Rev. Proc. 2021-32, Section 4 (page 6)

KYC – Know Your Customer

- Country level agreements that require foreign financial institutions to obtain identity documents from clients
  - General source: https://www.irs.gov/businesses/international-businesses/list-of-approved-kyc-rules

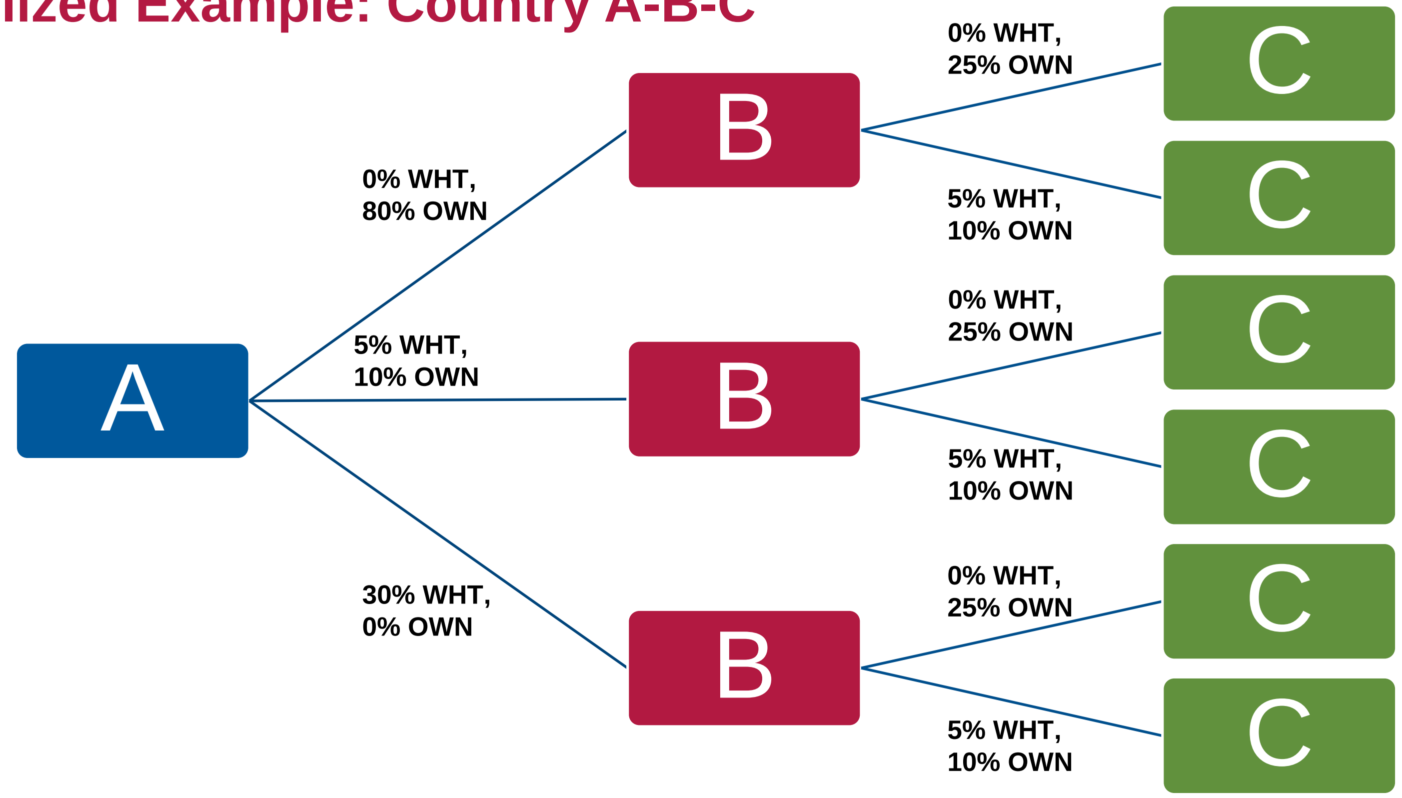
Variation in EOI Indicators



Sequence Construction

- Take 59,143 pairs of dividend WHTs and create 3 country sequences by linking together treaty rates.
  - 80,564 USA origin sequences (Country A = USA)
  - 15,178,094 Worldwide sequences (Country A = Any country in Gravity Model)

Stylized Example: Country A-B-C



Index Variable Construction: DIV.OWN.path

- Index variable *DIV.OWN* is constructed across a path by multiplying (1 – *DIV WHT) * (1- OWN)* for each hop across a sequence.
- For example, consider a 3-country sequence, A -> B -> C, with 9 possible sets of dividend withholding rates across the full sequence. Solely based on these rates, our model would call the top row the "BEST" option out of these 9 and the bottom row would be deemed the "WORST".

DIV WHT A-B	OWN A–B	DIV WHT B-C	OWN B-C	(1 - DIV) path	Implied OWN A–C	(1 - OWN) path	*DIV.OWN. path*
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
				(1 - [1]) * (1 - [3])	[4] * [2]	(1- [4]) * (1- [6])	[5] * [7]
0	.80	0	.05	1	.04	0.912	0.912
0	.80	0	.10	1	.08	0.828	0.828
0	.80	.15	0	0.85	0	1	0.850
.05	.10	0	.05	0.95	0.005	0.94525	0.899
.05	.10	0	.10	0.95	.01	0.891	0.846
.05	.10	.15	0	0.8075	0	1	0.808
.30	0	0	.05	0.7	0	0.95	0.665
.30	0	0	.10	0.7	0	0.9	0.630
.30	0	.15	0	0.595	0	1	0.595

- A **"BEST"** path is one with the lowest possible withholding tax rate and minimal ownership requirements across the sequence.
- A **"WORST"** path is one with the largest combination of withholding tax rates and ownership rates.

Index Variable Construction: Capital Gains – origin and destination countries

CG

- For USA origin sequences, we could use the capital gains tax rate for non-resident individuals in Country 3 (the destination) as a proxy for the tax cost to gain access to the cross-border financial flow

CG_ratio

- For Worldwide sequences, we introduce a measure with direction:

$$cg\_ratio = \frac{(1 - CG_3)}{(1 - CG_1)}$$

- A CG rate of 0% in the destination country is more attractive to someone leaving a country with a high CG rate than it is to someone leaving a country that also has a 0% CG rate.
  - ratio greater than 1 indicates improvement in the CG rate
  - ratio less than 1 indicates the taxpayer is worse off along this dimension
  - ratio equal to 1 indicates no change between Country 3 and Country 1

Index Variable Construction: Regulator Quality, 3-Country Sequences

RQ

- For Regulator Quality, we use the simple average for RQ across each country in the sequence.
- We do impute missing values for 19 countries
  - Imputation regression uses GDP per capita, FATCA and AEOI indicators, participation in various multilateral treaties, and indicators for whether the country is a territory of France, the Netherlands, the UK, or the U.S.
  - R^2 was 0.7391
- Imputed Countries
  - Curacao, Gibraltar, Monaco, Guadeloupe, St Maarten, San Marino, Bonaire, Isle of Man, Faroe Islands, British Virgin Islands, Guernsey, Turks and Caicos, New Caledonia, Falkland Islands, Northern Mariana Islands, French Polynesia, Cook Islands, Montserrat, Niue

Index Variable Construction: Regulator Quality Imputation

ISOCode	RQ	RQ.imputed	ISOCode	RQ	RQ.imputed	ISOCode	RQ	RQ.imputed	ISOCode	RQ	RQ.imputed
HKG	2.167	2.167	JPN	1.377	1.377	BES	NA	1.101	NCL	NA	0.595
SGP	2.118	2.118	GIB	NA	1.376	CYP	1.032	1.032	SVN	0.58	0.58
NZL	2.092	2.092	TWN	1.372	1.372	MUS	1.029	1.029	FLK	NA	0.525
NLD	2.051	2.051	CHL	1.35	1.35	ARE	1.015	1.015	BRB	0.495	0.495
AUS	1.933	1.933	GRL	1.324	1.324	ESP	0.945	0.945	ROU	0.488	0.488
CAN	1.89	1.89	MCO	NA	1.322	PRT	0.911	0.911	URY	0.476	0.476
CHE	1.887	1.887	MLT	1.285	1.285	IMN	NA	0.894	MNP	NA	0.425
FIN	1.823	1.823	GUF	1.282	1.282	POL	0.881	0.881	HRV	0.424	0.424
NOR	1.816	1.816	ISR	1.274	1.274	PRI	0.872	0.872	OMN	0.423	0.423
SWE	1.801	1.801	GLP	NA	1.252	FRO	NA	0.865	QAT	0.42	0.42
DEU	1.786	1.786	BEL	1.247	1.247	VGB	NA	0.85	BHR	0.416	0.416
MAC	1.76	1.76	CZE	1.235	1.235	BMU	0.844	0.844	PAN	0.388	0.388
GBR	1.717	1.717	MTQ	1.21	1.21	VIR	0.844	0.844	PYF	NA	0.378
LUX	1.694	1.694	REU	1.21	1.21	SVK	0.826	0.826	COK	NA	0.35
EST	1.645	1.645	AND	1.21	1.21	GGY	NA	0.768	COL	0.341	0.341
USA	1.631	1.631	SXM	NA	1.195	CYM	0.756	0.756	LCA	0.307	0.307
DNK	1.624	1.624	ABW	1.194	1.194	BRN	0.718	0.718	KNA	0.293	0.293
IRL	1.588	1.588	FRA	1.16	1.16	ITA	0.706	0.706	MSR	NA	0.284
LIE	1.497	1.497	LTU	1.159	1.159	JEY	0.683	0.683	MEX	0.279	0.279
AUT	1.44	1.44	LVA	1.157	1.157	HUN	0.652	0.652	GRC	0.24	0.24
ISL	1.435	1.435	SMR	NA	1.141	TCA	NA	0.634	ZAF	0.234	0.234
CUW	NA	1.405	KOR	1.108	1.108	BGR	0.626	0.626	NIU	NA	0.228

Shaded values are imputed. Countries with smaller values of RQ than NIU are not shown. All imputed countries are contained on this chart.

Index Variable Construction: EOI, 3-Country Sequences

EOI.ratio

- To introduce direction, we take EOI.path and divide by the ratio of EOI in Country C to EOI in Country A.

	Country A	Country B	Country C	EOI.path	(1 + EOI C) / (1 + EOI A)	EOI.ratio
	0	1	1	0.6	2	0.3
	0	0	1	0.75	2	0.375
ALL EOI	1	1	1	0.5	1	0.5
	1	0	1	0.6	1	0.6
	0	1	0	0.75	1	0.75
NO EOI	0	0	0	1	1	1
	1	1	0	0.6	0.5	1.2
	1	0	0	0.75	0.5	1.5

- 8 possible outcomes for Worldwide Sequences. Shaded rows are the 4 possible outcomes for USA origin Sequences.
  - The largest value is the most attractive for tax evasion: leaving a country with EOI participation and hopping to two counties with no EOI participation. The smallest value is the least attractive: starting in a country with no EOI and hopping to two countries with EOI participation

Dependent Variable Construction

Inward.Adjust

- We calculate an Adjusted FDI for each sequence that is the portion of Inward FDI from Country C into Country B that could possibly make it into Country A.

Country	Country	Country	Inward.1	Inward.2	Inward.Total.1	Inward.Total.2	adjust	Inward.Adjust
A	B	C	**100**	200	1000	**735**	0.136054	**27**
A	B	D	100	50	1000	**735**	0.136054	7
A	B	E	100	25	1000	**735**	0.136054	3
A	B	F	100	300	1000	**735**	0.136054	41
A	B	G	100	10	1000	**735**	0.136054	1
A	B	H	100	150	1000	**735**	0.136054	20

- Suppose **$100** of Inward FDI in Country A comes from Country B (Inward.1) and that total Inward FDI into Country A is $1,000 (Inward.Total.1). Inward.2 is the amount of Inward FDI from the country listed in the third column into Country B and this amount totals **$735** (Inward.Total.2). We know that only a maximum of $100 out of this $735 of Inward FDI into Country B is invested into Country A.

- We adjust all amounts proportionally by the ratio of Inward.1 to Inward.Total.2 (adjust) and multiply this factor times the amounts in Inward.2 to derive what we are calling **Inward.Adjust**. This is the maximum amount of Inward FDI from each country into Country B that could eventually become Inward FDI into Country A. Notice that Inward.Adjust sums up to 100, and for the first sequence it is **$27**

Index Variables:
Worldwide 3-Country Sequences

Variable or Stat	ALL	BEST	With FDI	No FDI	Reg Sample
Path: (1 - DIV WHT)	0.786	0.796	0.796	0.795	0.828
Path: (1 - OWN)	0.979	0.995	0.992	0.997	0.982
DIV.OWN	0.768	0.791	0.790	0.792	0.812
CG ratio	1.036	1.032	1.034	1.031	1.040
Path: WB RQ	0.532	0.514	0.543	0.498	0.600
FATCA w/ direction	0.828	0.849	0.804	0.873	0.691
EOIR w/ direction	0.809	0.828	0.771	0.859	0.659
AEOI w/ direction	0.905	0.920	0.875	0.859	0.789
KYC w/ direction	0.851	0.868	0.834	0.885	0.729
Count:	14,837,452	11,696,856	4,075,933	7,620,923	670,281
Count w/ FDI	6,067,599	4,075,933	4,075,933	0	670,281
Adjusted FDI ($B)	24.1	8.1	8.1	---	49.8

- 14,837,452 sequences have complete data
- Of these, 11,696,856 represent the BEST (least dividend withholding taxes) sequences within a 3-country path.
- 4,075,933 BEST paths have no missing FDI data
- 670,281 paths have non-zero FDI and can be used for the estimation

Gravity Model Index Weights: Worldwide

- Dependent Variable: log(Inward.Adjust)
- Higher index values are associated with
  - lower dividend withholding taxes across the path (DIV.OWN.path),
  - improvement in the capital gains tax rate from origin to destination,
  - high average regulator quality
  - moving from a country with information sharing to one or more countries without information sharing

	Worldwide w/ Directionality
Constant	1.321 ***
	[62.820]
log DIV.OWN.path	3.704 ***
	[88.328]
log cg_ratio	0.1533 ***
	[6.543]
log RQ.path	11.554 ***
	[364.916]
log FATCA.ratio	-1.146 ***
	[-98.966]
Observations	670,281
R2	0.225
Adjusted R2	0.225
F statistic	48,669.05

*** p < 0.001; ** p < 0.01; * p < 0.05.

Weighting the Gravity Index

- We take the estimated coefficients from our preferred model (Worldwide with ALL EOI measures) and plug them into the structural equation of the gravity model, which with some rearranging looks like this:

$$DIV.OWN.path^{\beta_1}(1 - CG.ratio)^{\beta_2} RQ.path^{\beta_3}(1 + EOI.path)^{-\beta_4}$$

- This allows us to weight the gravity index for all 3-country sequences (triplets) with complete data

- Also generate a weighted index for each pair of countries using the same weights for a simple stopping rule:

  - If the weighted index for the triplet is larger than the weighted index for the pair, then move on to Country 3. Otherwise, stay in Country 2.

  - This rule drastically reduces the set of potential triplets and will make it possible to construct longer sequences.

Constructing Longer Sequences

Country Data
- Treaty WHTs for all Country Pairs
- World Bank Data
- Capital Gains Taxes
- Exchange of Information Variables

Create Sequences
- Link Countries into Sequences
- Create Index Variables

Weighted Index
- Estimate Gravity Model Index Weights
- Use coefficients to weight the Gravity Index

Stopping Rules
- Eliminate sequences with $0 of Adjusted FDI
- Remove when better to stay in Country 2

Create Longer Sequences

- A 4-Country sequence is constructed from two 3-Country sequences
- We chain sequences by multiplying the Index for each triplet and implementing a move-to-D or stay-at-C test

A-B-C	B-C-D	A-B-C-D
Index1	Index2	Chained.Index = Index1 * Index 2

- If Chained.Index >= Index1 * Index1, then advantageous to move to Country D
- If Chained.Index < Index1 * Index1, then best to stay in Country C
- Using the weights from the Worldwide sequences we can link countries indefinitely

Applications

Predict best conduits and destinations

- For a particular origin country, look at the sequences with the largest index values to find the most attractive destinations.
    - Can string together sets of 3-Country sequences and look at predictions
- For a given set of countries, what are the most likely next two countries in a sequence (i.e., a conduit and a destination)
    - Take a set of countries and link each country to the set of 3-Country sequences to get the next two potential countries
    - Pick the largest possible index for each destination country
    - For the set of possible destinations, choose the conduits with the largest indexes to get the most attractive conduits
- Can make these predictions for any country or set of countries contained in the gravity model

Future model expansions

- Adding additional years
  - Current model is only for calendar year 2017
  - Working on adding treaty data for additional years
- Model other types of withholding
  - This model is all based on dividend withholding tax rates.
  - Could expand to withholding tax rates on interest or royalties.
- Other possibilities??

References

- Beer, S., de Mooij R. & Liu, Li. (2020). *International Corporate Tax Avoidance: a Review of the Channels, Magnitudes, and Blind Spots*. Journal of Economic Surveys 34(3), 660-688. https://doi.org/10.1111/joes.12305

- Ferwerda, Joras, van Saase, Alexander, Unger, Brigitte, & Getzner, Michael. (2020) *Estimating money laundering flows with a gravity model-based simulation*. Scientific Reports (2020) 10:18552. https://doi.org/10.1038/s41598-020-75653-x

- Johanessen, N., Reck, D., Risch, M., Slemrod, J., Guyton, J. & Langetieg, P. (2023). *The Offshore World According to FATCA: New Evidence on the Foreign Wealth of US Households*. Memo.

- Lejour, A. (2021). *The Role of Conduit Countries and Tax Havens in Corporate Tax Avoidance*. Tilburg Center for Economic Research Discussion Paper, 2021-014.

- Walker, John & Unger, Brigitte. *Measuring Global Money Laundering: "The Walker Gravity Model".* Review of Law and Economics. (2009) 5:821-853.

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Art in the Age of Tax Avoidance

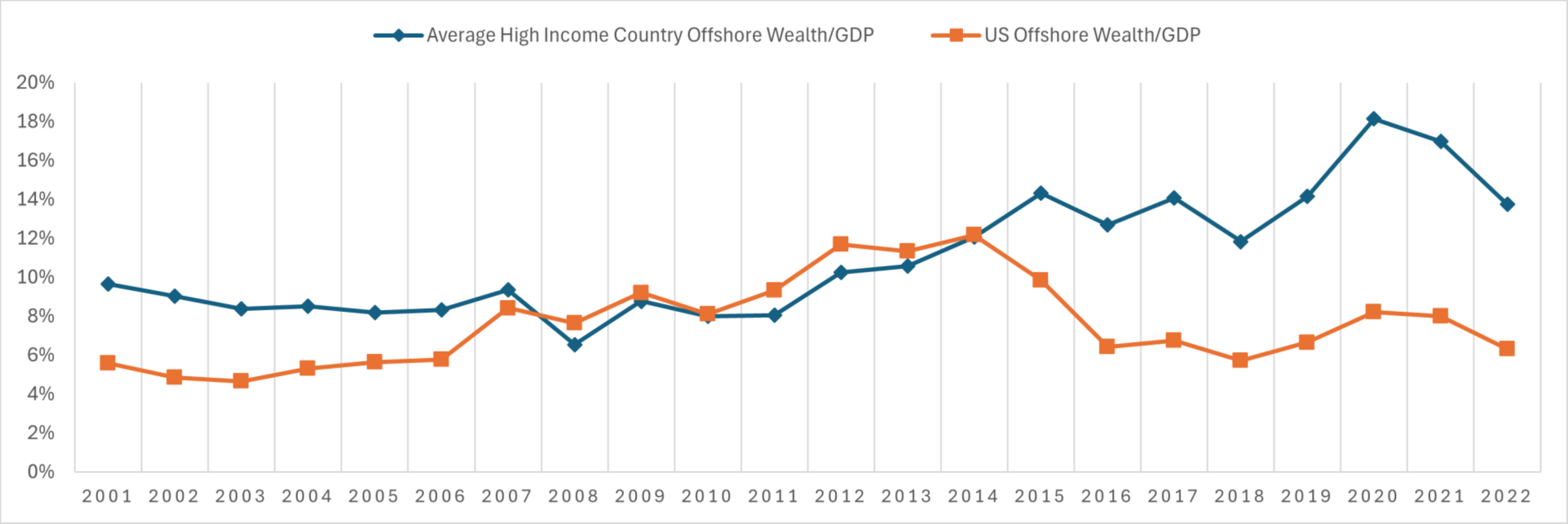
Matthew Pierson
WRDS, The Wharton School, University of Pennsylvania

May 16, 2024

Motivation

Tax Avoidance/Evasion

- We know a good deal about tax evasion via offshore tax havens
  - 8% of global financial assets in tax havens
  - U.S. uses tax havens less than other developed countries



Motivation

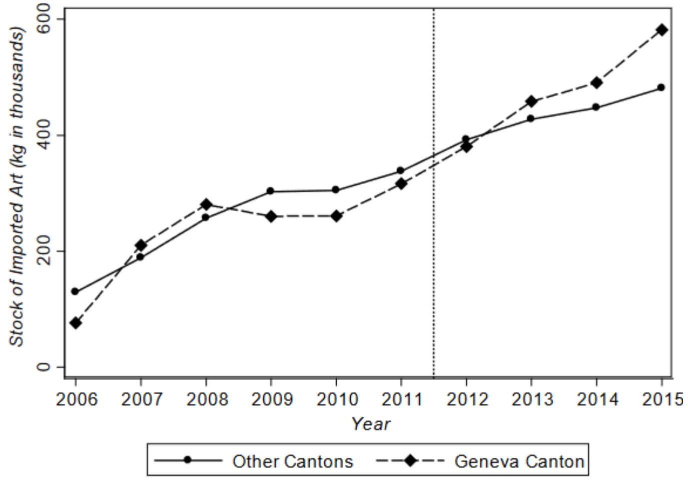
Is the U.S. special?

- One hypothesis: Enough tax avoidance opportunities domestically
  - Noted fuzzy line between charitable giving and non-profits as a tax avoidance vehicle
    - Charitable donations more responsive to taxes than other countries (Fack and Landais, 2016)
    - However charities are, by definition, charities and provide public goods (Gee and Meer, 2019)

- How do we disentangle charitable activity from tax avoidance by non-profits?

Laboratory

- Donations of an asset with known simultaneous illicit and legitimate uses
  - Will be reported/not directly self-incriminating
  - Not necessarily legitimate purposes
- Solution: Art donations to non-profits
  - Art is an historically opaque market, often used for illicit purposes
    - Ang (2020), U.S. Senate (2020), Helgadóttir (2023)
  - Art sometimes used to evade tax
    - De Simone, et al (2020), Londoño-Vélez & Ávila-Mahecha (2023)



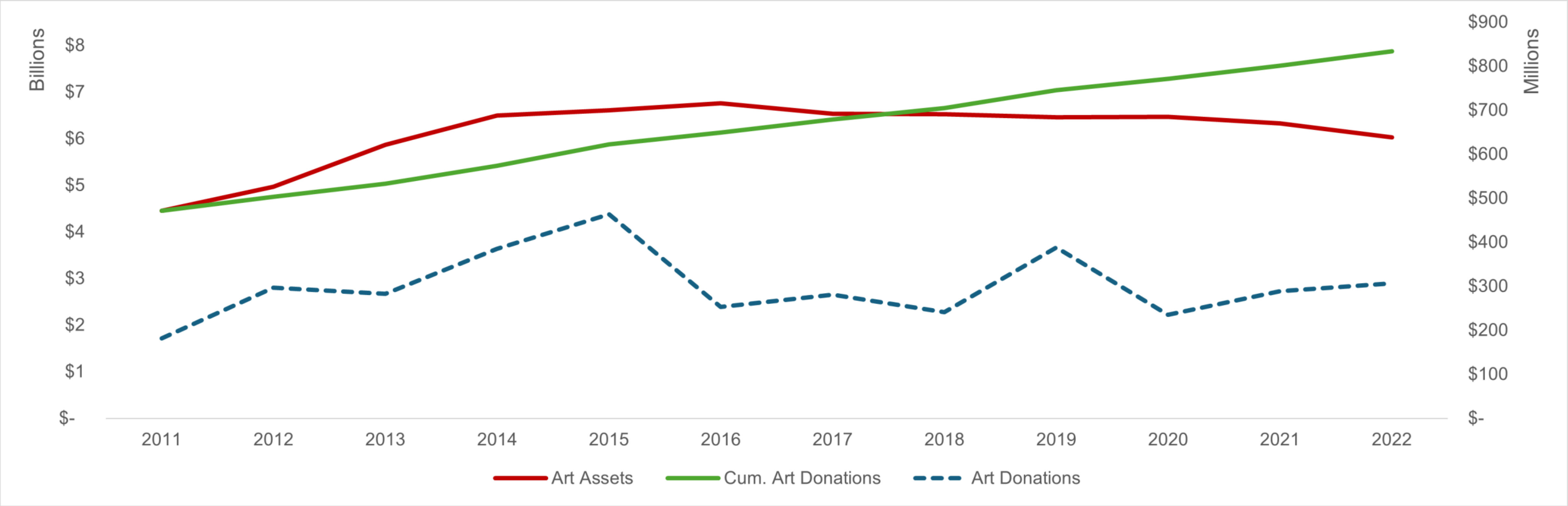
Contribution

- Examine non-profit art donations and assets
  - First to do so
- Disclosing & re-valuing art assets and donations a function of audit risk
  - Audit threat and tax compliance: Kleven, et al (2011)
- Audit risk reveals tax motivated behavior and potential tax losses
  - Value of audit: Boning, Hendren, and Sprung-Keyser (2023)
  - Tax losses to non-financial assets: Johannesen, et al (2022), Alstadsæter, et al (2022)

Non-Profits and Art

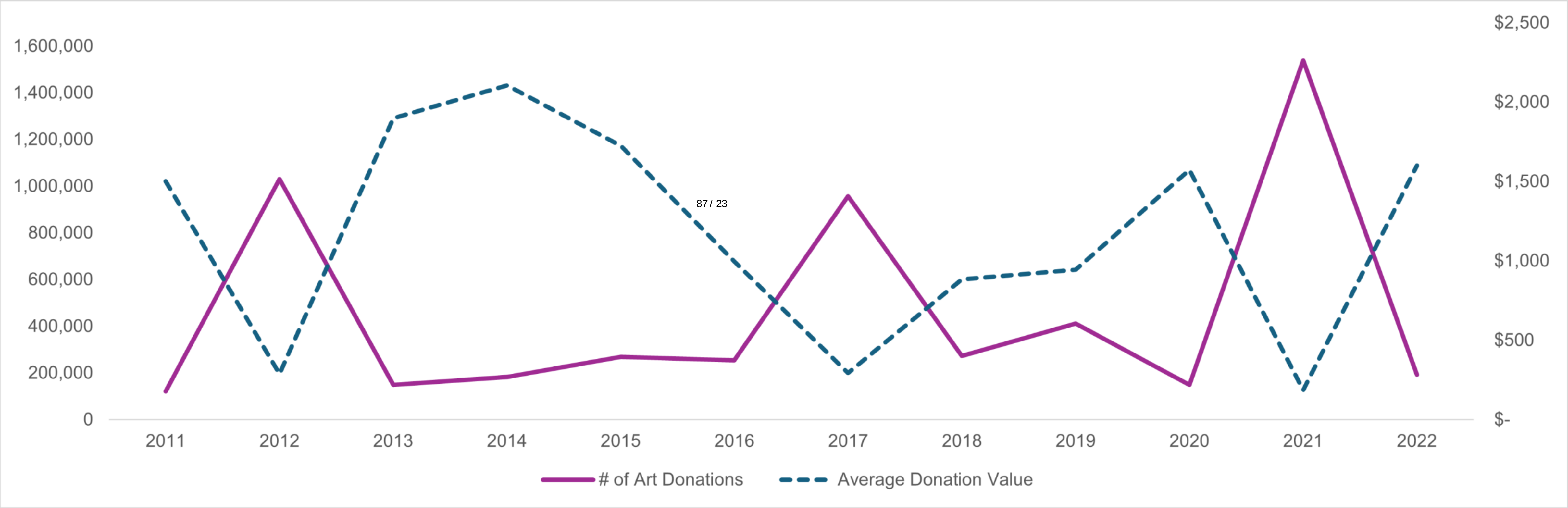
- Build sample of all e-filed non-profit (Form 990) filing orgs ▸ Form 990 Background
    - 5.3 million organization-years from 2011 to 2022
- Non-profits hold $12.4T in assets in 2022
- 1.2% of non-profits hold art, only 17% of these (0.2% of total) record asset and donation value
- Art assets worth at least $6B in 2022, donations of $300M

Art Assets



Billions
$8
$7
$6
$5
$4
$3
$2
$1
$-
$900
$800
$700
$600
$500
$400
$300
$200
$100
$-
Millions
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Art Assets
Cum. Art Donations
Art Donations

Art Donations



Comparative Statistics: Art Filings

	Art Filing	No Art Filing	Difference
Organization Type			
Private Foundation	2.06%	17.757%	-15.694%***
Education	19.821%	5.931%	13.889%***
Religious	0.163%	1.409%	-1.246%***
Library	2.103%	0.492%	1.610%***
Museum	15.990%	0.334%	15.656%***
Medical	2.530%	1.706%	0.824%***
Other	57.332%	72.371%	-15.039%***
Organization Characteristics			
Audit Flag	25.056%	7.776%	17.280%***
Charity Nav. Rating	92.263%	35.963%	56.299%***
Charity Nav. Stars	3.035	1.038	1.997***
Foreign Operated	3.08%	0.381%	2.695%***
Family Foundation	31.531%	34.248%	-2.717%***
log(Total Assets)	16.660	13.029	3.632***
Total Revenue (millions)	147.452	9.761	137.690***
Salary Expense (millions)	5.954	0.311	5.644***
Contributions/Total Revenue	60.155%	44.292%	15.863%***

Comparative Statistics: Art Value

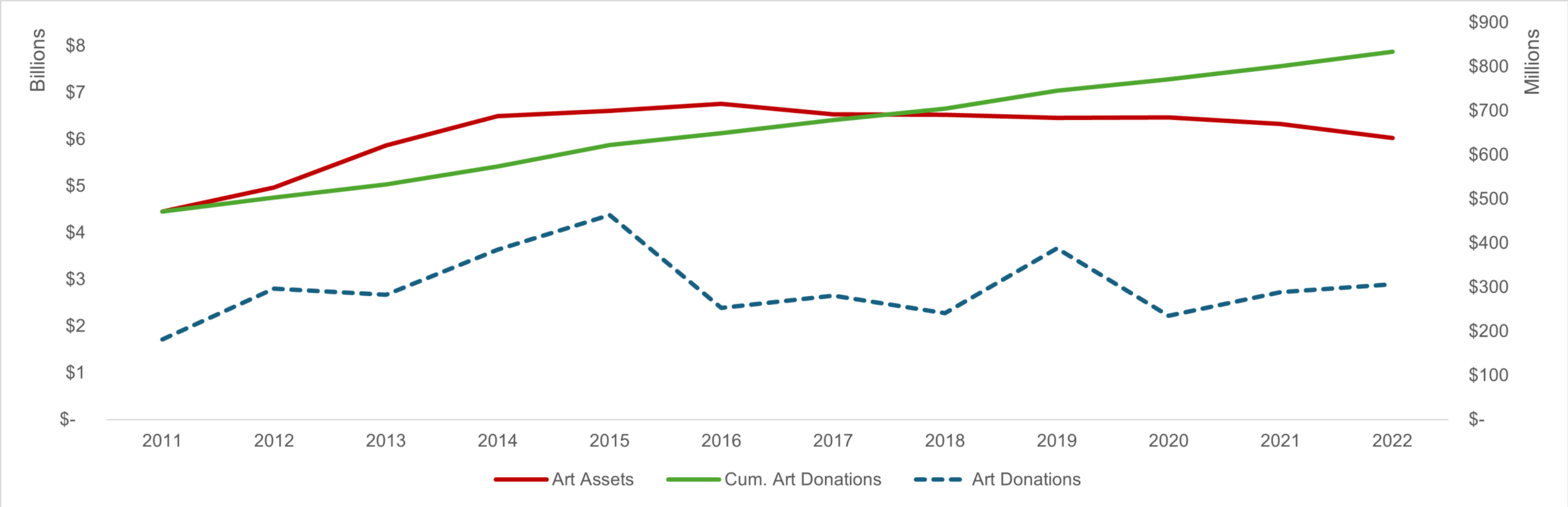
	Art Value	No Art Value	Difference
Organization Type			
Private Foundation	1.245%	2.215%	-0.970%***
Education	48.689%	14.456%	34.233%***
Religious	0.071%	0.180%	-0.109%***
Library	0.347%	2.429%	-2.082%***
Museum	9.672%	17.163%	-7.491%***
Medical	3.714%	2.309%	-1.404%***
Other	36.262%	61.248%	-24.986%***
Organization Characteristics			
Audit Flag	41.169%	22.061%	19.108%***
Charity Nav. Rating	96.378%	91.498%	4.880%***
Charity Nav. Stars	3.467	2.955	0.512***
Foreign Operated	7.081%	2.332%	4.749%***
Family Foundation	36.925%	30.529%	6.396%***
log(Total Assets)	18.264	16.362	1.902***
Total Revenue (millions)	319.729	115.436	204.293***
Salary Expense (millions)	13.363	4.578	8.785***
Contributions/Total Revenue	50.599%	61.931%	11.332%***

Comparative Statistics: Art Overvalue

	Overvalued Donation	No Overvalued Donation	Difference
Organization Type			
Private Foundation	1.586%	1.101%	0.485%*
Education	52.138%	47.240%	4.898%***
Religious	0.000%	0.101%	-0.101%***
Library	0.655%	0.217%	0.438%***
Museum	9.138%	9.897%	-0.759%
Medical	3.380%	3.855%	-0.475%*
Other	33.103%	37.589%	-4.485%***
Organization Characteristics			
Audit Flag	44.586%	39.733%	4.853%***
Charity Nav. Rating	97.069%	96.088%	0.981%**
Charity Nav. Stars	3.552	3.431	0.121***
Foreign Operated	9.344%	6.130%	3.215%***
Family Foundation	35.966%	37.328%	-1.362%**
log(Total Assets)	18.453	18.184	0.269***
Total Revenue (millions)	346.294	308.566	37.728
Salary Expense (thousands)	0.150	0.127	0.023*
Contributions/Total Revenue	47.682%	51.824%	-4.142%***

By Art Stated Use
By Donation Valuation Method

Overvaluation



$$V_{i,T} = A_{i,2011} + \left(\sum_{t=2011}^{T} D_{i,t} - S_{i,t}\right) - A_{i,T} > 0$$

where, for years t to T and non-profit i, art donations D, art sales S, art assets A, such that $V_{i,T}$ is the overvaluation for non-profit i in year T

Audit Risk

- Audit Flag $\Rightarrow$ 1.8% more likely to report, 3% more likely to value, and 4.5% more likely to re-value art

- Instrumenting on prior year non-profit audit rates, 9.3%, -18% and 36%
  - Under some strong assumptions, 36% of art donated to non-profits is overvalued

- Making sense of this: classic evasion model what responses does a non-profit have?

  ▸ A-S 1972 Refresher

  - Evasion $\downarrow$ audit probability $\uparrow$
  - Costs of decreased ”evasion”? Mechanical $\uparrow$ in compliance costs
  - Keep audit probability constant by $\downarrow$ probability in other ways
    - $\uparrow$ paper filings, which have no reporting improvements

Audit Flag

- What is an audit flag?
  - Diversion of assets
  - Political activity
  - Unrelated business income
  - Excess benefit transactions
  - Loans to disqualified persons
  - Excess compensation
  - Foreign grant activity
  - Fundraising income/expense discrepancies

- Non-profit advisory services that $\uparrow$ audit probability

- A measure of audit risk
  - Downside: behavior selection by non-profit

Audit Flags

	(1)	(2)	(3)
	Art Filing	Art Value	Overvalued Art
Audit Flag$_{(t-1)}$	1.806***	2.941***	4.660**
	(0.067)	(0.836)	(2.211)
Year F.E.	Yes	Yes	Yes
Art Use & Don Val F.E.	-	Yes	Yes
Observations	5,364,313	62,541	9,801
R-squared	0.090	0.125	0.028

▸ NTEE ▸ Logit ▸ Ordinal

Instrumental Variables

	(1)	(2)	(3)
	Art Filing	Art Value	Overvalued Art
Audit Flag* $_{(t-1)}$	9.324***	-18.512***	36.303***
	(0.232)	(1.939)	(6.506)
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Org FE	Yes	Yes	Yes
Observations	5,362,930	62,541	9,500
Number of EIN	779,842	9,404	1,317
F-Test	3560.28***	238.15***	38.54***

Responses to Audit Flags

- Evasion ↓ audit probability

↑

- Costs of decreased ”evasion”? Mechanical ↑ in compliance costs

▸ Compliance Costs

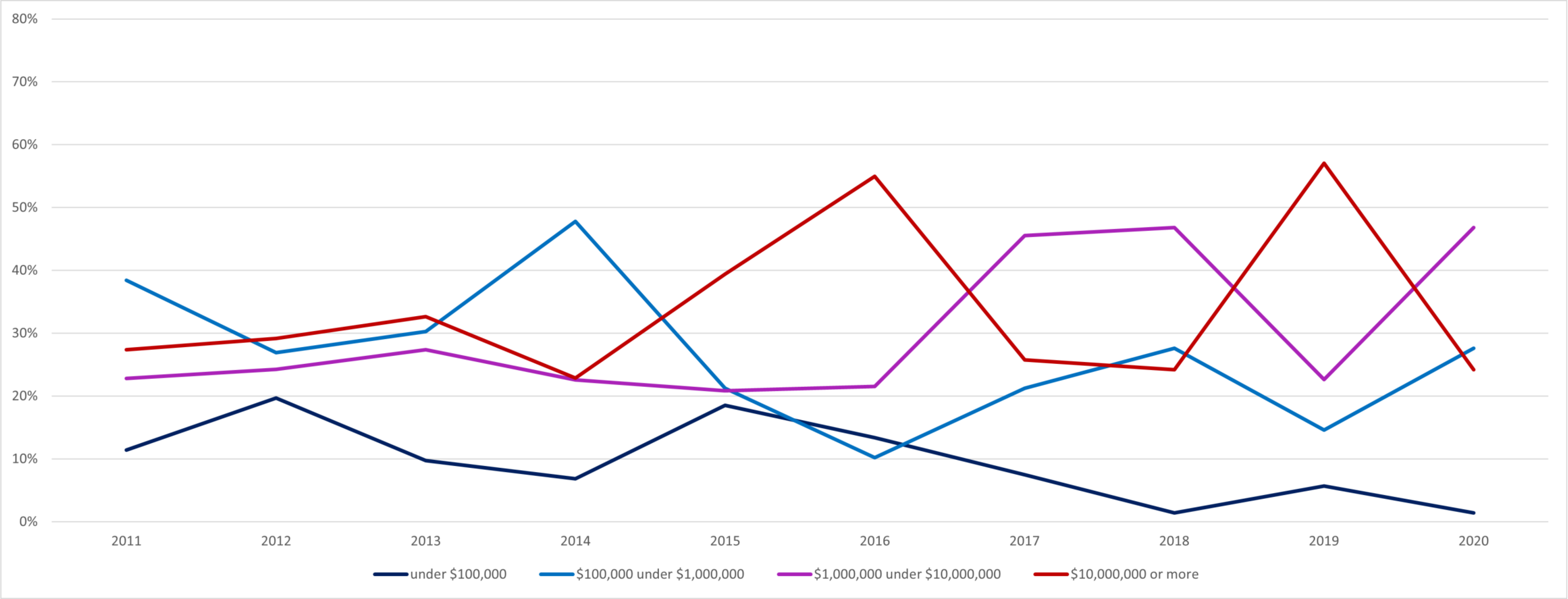
- Keep audit probability constant by ↓ probability in other ways
  - ↑ paper filings, which have no reporting improvements

▸ Paper Filings

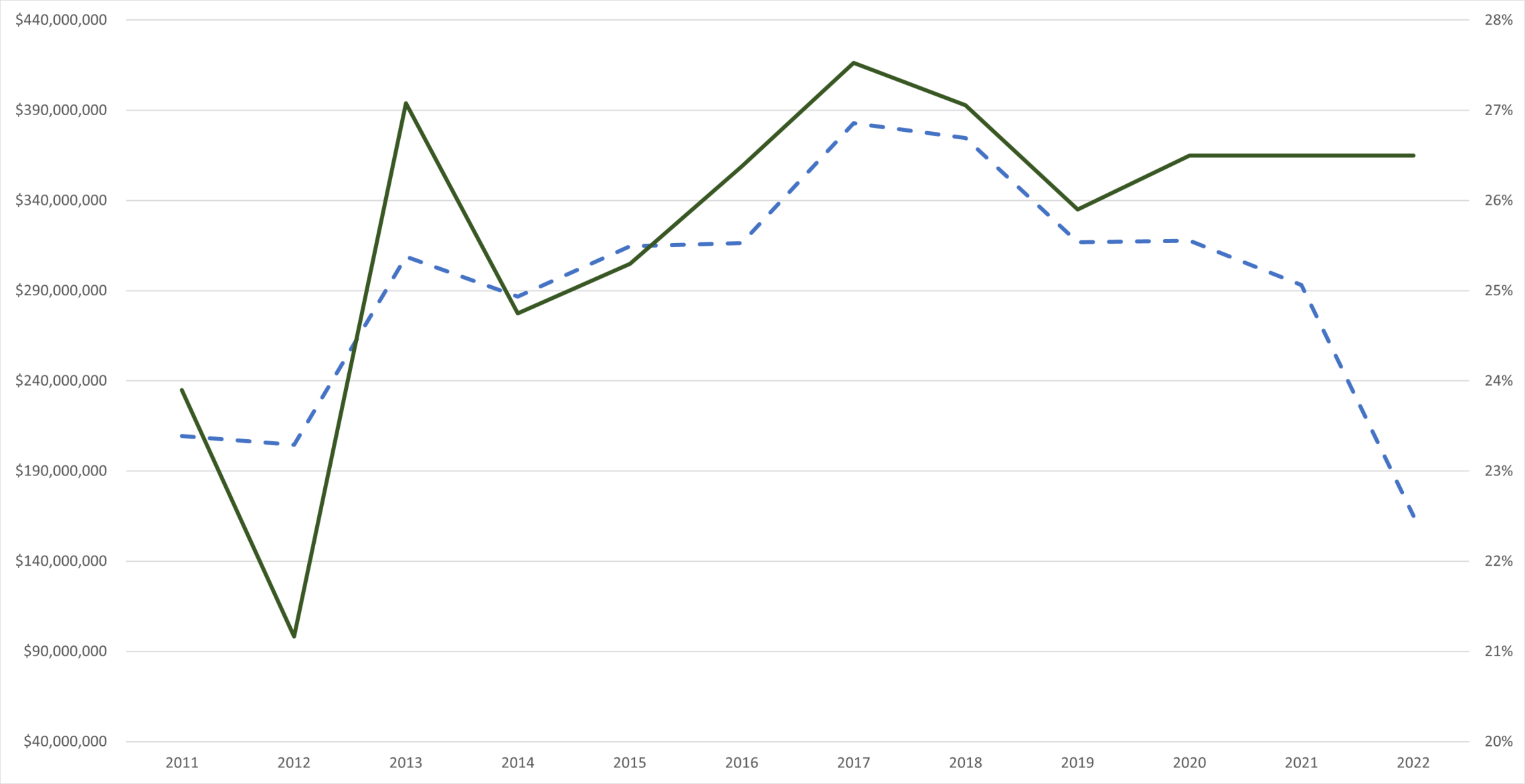
Tax Avoidance

- Extrapolating at audit flag rate for the full art filing sample up to $28 billion in tax losses
  - Audit flags indicate up to 36% of art is overvalued
- Minimum tax losses of $4.8B across full sample (2022 $)

Art Donations by AGI



Time Series of Tax Losses and ETRs



▸ Calculating Weighted Average ETR

Estimated Tax Loss

Organizations with Art Value	
Art Write-down Value	$11,294.84
Estimated Income Tax Loss	**$4,806.07**
Organizations with Art Filing	
Predicted Art Write-down Value	$110,657.21
Predicted Income Tax Loss	**$28,678.73**

Conclusion

- Describe art holdings by U.S. non-profits
  - 1.2% of all Form 990 filing orgs
    - Extensive assets ($6B), despite consistent lack of reporting (76%)
  - Variation by organization, stated use, and donation valuation types

- ↑ in audit probability leads to ↑ art disclosure, value, and valuation accuracy
  - IV using audit rates to address selection
    - Causes ↑ in art disclosure and re-valuation, causes ↓ in valuation
  - Mechanical ↑ in compliance costs
  - Use of paper filings to negate ↑ in audit probability from audit flags

- Estimated tax losses from overvalued donations worth at least $400M/year (2022 $), up to $2.4B/year

Form 990 Filing Details

- Art disclosed on basic Form 990, Questions 8 and 30
- Art values on Schedule D (assets) and M (donations)
- Donation valuation methods listed on column (d) of Schedule M
- Stated use of art holdings listed on Question 3 of Schedule D

Back

8 Did the organization maintain collections of works of art, historical treasures, or other similar assets? *If "Yes," complete Schedule D, Part III* . . . .

30 Did the organization receive contributions of art, historical treasures, or other similar assets, or qualified conservation contributions? *If "Yes," complete Schedule M* . . . .

Part III **Organizations Maintaining Collections of Art, Historical Treasures, or Other Similar Assets**
Complete if the organization answered "Yes" on Form 990, Part IV, line 8.

Part I **Types of Property**

	(a) Check if applicable	(b) Number of contributions or items contributed	(c) Noncash contribution amounts reported on Form 990, Part VIII, line 1g	(d) Method of determining noncash contribution amounts
1 Art—Works of art				
2 Art—Historical treasures . . .				
3 Art—Fractional interests . . .				

3 Using the organization's acquisition, accession, and other records, check any of the following that make significant use of its collection items (check all that apply).

- **a** ☐ Public exhibition
- **b** ☐ Scholarly research
- **c** ☐ Preservation for future generations
- **d** ☐ Loan or exchange program
- **e** ☐ Other

Overvaluation by Art Stated Use

Art Stated Uses	Overvalued Donation	No Overvalued Donation	Difference
Public Exhibit	64.931%	64.498%	0.433%
Preservation	60.655%	60.904%	0.249%
Research	43.828%	38.574%	5.253%***
Loan	28.241%	19.939%	8.302%***
Other Use/Unknown	3.069%	1.942%	1.127***%

Back

Overvaluation by Donation Valuation Method

Donation Valuation Methods	Overvalued Donation	No Overvalued Donation	Difference
Donor Auction	0.414%	0.101%	0.312%**
Comparable Sales	1.414%	1.406%	0.008%
Cost	3.207%	1.493%	1.714%***
Donor Supplied	2.310%	2.043%	0.267%
Org. Estimate	1.172%	1.072%	0.100%
Market Value	14.310%	12.665%	1.646%**
Insurance	0.552%	0.840%	-0.289%
Appraisal	18.655%	16.737%	1.918%**
Artist	0.172%	0.232%	-0.059%
Other Valuation Method/Unknown	57.793%	63.411%	-5.618%***

Back

Allingham-Sandmo 1972

$$\max_{\bar{w}} (1 - p) \cdot u(w - \tau \cdot \bar{w}) + p \cdot u(w - \tau \cdot \bar{w} - \tau(w - \bar{w})(1 + \theta))$$

where w is true income, $\bar{w}$ is reported income, τ is tax rate, p is audit probability, θ is the percentage penalty, and $u(.)$ is a concave utility function

FOC in $\bar{w}$: $\Rightarrow \frac{u'(c^{Audit})}{u'(c^{NoAudit})} = \frac{1-p}{p\theta}$

- Individual taxpayer problem $\Rightarrow$ generalize informally to non-profit problem

- Tax evasion $(w - \bar{w})$ decreases with fine size and audit probability.

Back

Robustness: NTEE Classification

	(1) Art Filing	(2) Art Value	(3) Overvalued Art
Audit Flag$_{(t-1)}$	0.388***	3.536***	4.050*
	(0.038)	(0.771)	(2.111)
Controls	Yes	Yes	Yes
Don Val, & Art Use Dummies	-	Yes	Yes
NTEE & Year F.E.	Yes	Yes	Yes
Observations	5,295,137	72,147	10,413
R-squared	0.642	0.167	0.046

Back

Robustness: Logit

	(1) Art Filing	(2) Art Value	(3) Overvalued Art
Audit Flag$_{(t-1)}$	0.488***	0.193***	0.231***
	(0.022)	(0.032)	(0.059)
Controls	Yes	Yes	Yes
Don Val F.E.	-	Yes	Yes
Art Use F.E.	-	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	5,295,137	62,541	9,794

Back

Robustness: Ordinal Breakdown

	(1) Art Filing	(2) Art Value	(3) Overvalued Art
# Audit Flags$_{(t-1)}$=1	0.464***	2.963***	4.445**
	(0.037)	(0.804)	(2.206)
# Audit Flags$_{(t-1)}$=2	0.615***	4.030*	7.196
	(0.164)	(2.264)	(4.673)
# Audit Flags$_{(t-1)}$=3	-0.480	-7.478	-12.681
	(0.585)	(6.360)	(17.779)
# Audit Flags$_{(t-1)}$=4	-7.732**	-21.826***	-
	(3.604)	(1.960)	-
# Audit Flags$_{(t-1)}$=5	10.251***	-9.744***	-
	(0.392)	(1.656)	-
Controls & Year F.E.	Yes	Yes	Yes
Don Val & Art Use F.E.	-	Yes	Yes
Observations	5,295,137	72,147	10,413
R-squared	0.652	0.172	0.028

Compliance Outcomes

	(1) Audit Committee	(2) log(Accounting Fees)	(3) log(Legal Fees)
Audit Flag $_{(t-1)}$	0.012*** (0.002)	0.317*** (0.006)	0.379*** (0.012)
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Don Val FE	Yes	Yes	Yes
Art Use FE	Yes	Yes	Yes
Observations	1,547,312	2,084,595	897,598
R-squared	0.085	0.429	0.297

Back

Paper Filings

	Art Filing		Art Value		Overvalued Art	
	>$10M eFile (1)	<$10M eFile (2)	>$10M eFile (3)	<$10M eFile (4)	>$10M eFile (5)	<$10M eFile (6)
Audit Flag$_{(t-1)}$	7.204***	-0.081	8.934***	1.409***	6.088***	-1.718
	(0.114)	(0.051)	(0.355)	(0.414)	(0.874)	(1.523)
Observations	461,847	1,696,638	51,977	52,842	9,162	5,146
R-squared	0.009	0.000	0.012	0.000	0.005	0.000

Paper Filings

Panel B.

	Paper	Art	Filing	Paper	Art	Value	Paper	Overvalued Art	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Audit Flag$_{(t-1)}$	2.506***			4.533***			0.925		
	(0.107)			(0.620)			(2.515)		
Paper Filing$_{(t-1)}$		-0.038			-0.469			-2.570**	
		(0.036)			(0.286)			(1.015)	
Audit Flag$_{(t-2)}$ w/ Paper Filing$_{(t-1)}$			0.128			0.758			-0.891
			(0.113)			(0.887)			(2.911)
Observations	1,696,638	1,696,638	1,696,638	52,842	52,842	52,842	5,146	5,146	5,146
R-squared	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000

Back

Calculating Weighted Average ETR

- ETR by AGI group
- Art non-cash donations by AGI group
- Weight ETR by % of art donation value

Table 1D. All Individual Returns With Noncash Charitable Contributions Reported on Form 8283, by Donation Type and Size of Adjusted Gross Income, Tax Year 2021

[All figures are estimates based on samples—money amounts are in thousands of dollars]

Size of adjusted gross income	Returns with donations of art and collectibles				
	Number of returns [1]	Number of donations	Donor's cost [2]	Fair market value	Amount carried to Schedule A [3]
	(1)	(2)	(3)	(4)	(5)
All returns	**41,492**	**68,331**	**581,682**	**1,908,640**	**1,402,501**
Under $25,000	0	0	0	0	0
$25,000 under $50,000	* 4,038	* 5,045	* 13,059	* 11,053	* 11,053
$50,000 under $75,000	* 7,343	* 9,344	* 18,302	* 8,592	* 8,592
$75,000 under $100,000	* 3,005	* 4,006	* 4,407	* 4,707	* 4,707
$100,000 under $200,000	10,401	24,418	152,357	474,367	159,698
$200,000 under $500,000	10,894	13,880	45,642	35,412	35,020
$500,000 under $1,000,000	2,576	4,100	48,154	144,410	137,604
$1,000,000 under $1,500,000	1,090	1,846	50,558	69,013	64,650
$1,500,000 under $2,000,000	596	2,804	22,158	75,551	75,586
$2,000,000 under $5,000,000	893	1,682	59,093	149,519	140,180
$5,000,000 under $10,000,000	341	592	32,429	102,274	102,108
$10,000,000 or more	315	613	135,523	833,743	663,303

Back

IRS Research, Applied Analytics & Statistics

TPC TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

June 13th, 2024

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, and Alex Turk*

(IRS Research, Applied Analytics, and Statistics)

Introduction

In the past 20 years the IRS has experienced a reduction in monetary resources but has faced expanded responsibilities. This has forced the IRS to be selective in the use of use of these resources and constrained the coverage in filing and payment compliance programs.

These declines provide a novel opportunity to use this natural experiment to conduct an analysis to estimate the direct and indirect effects of filing and payment compliance programs to support tax administration.

Effect Definitions

Direct effects are changes in the behavior for the treated taxpayer (e.g. ACS letters, field contacts levies, etc.). Behavioral responses can occur on various outcomes/margins, such as:

- Resolution of prior year delinquencies
- Improved compliance in current year
- Improved future compliance.

Indirect effects are changes in the behavior of a taxpayer not subject to the treatment, but are the result of: Knowledge of the IRS Action/treatment or

- Updated belief in their possibility/likelihood of future treatment.

There are various channels to propagate these effects, such as:

- Public data
- Preparers
- Treated taxpayer in social network, such as friends and family

Past Literature

Previous research that estimates indirect effects has been narrowly focused on specific programs, but those studies have demonstrated indirect effects for:

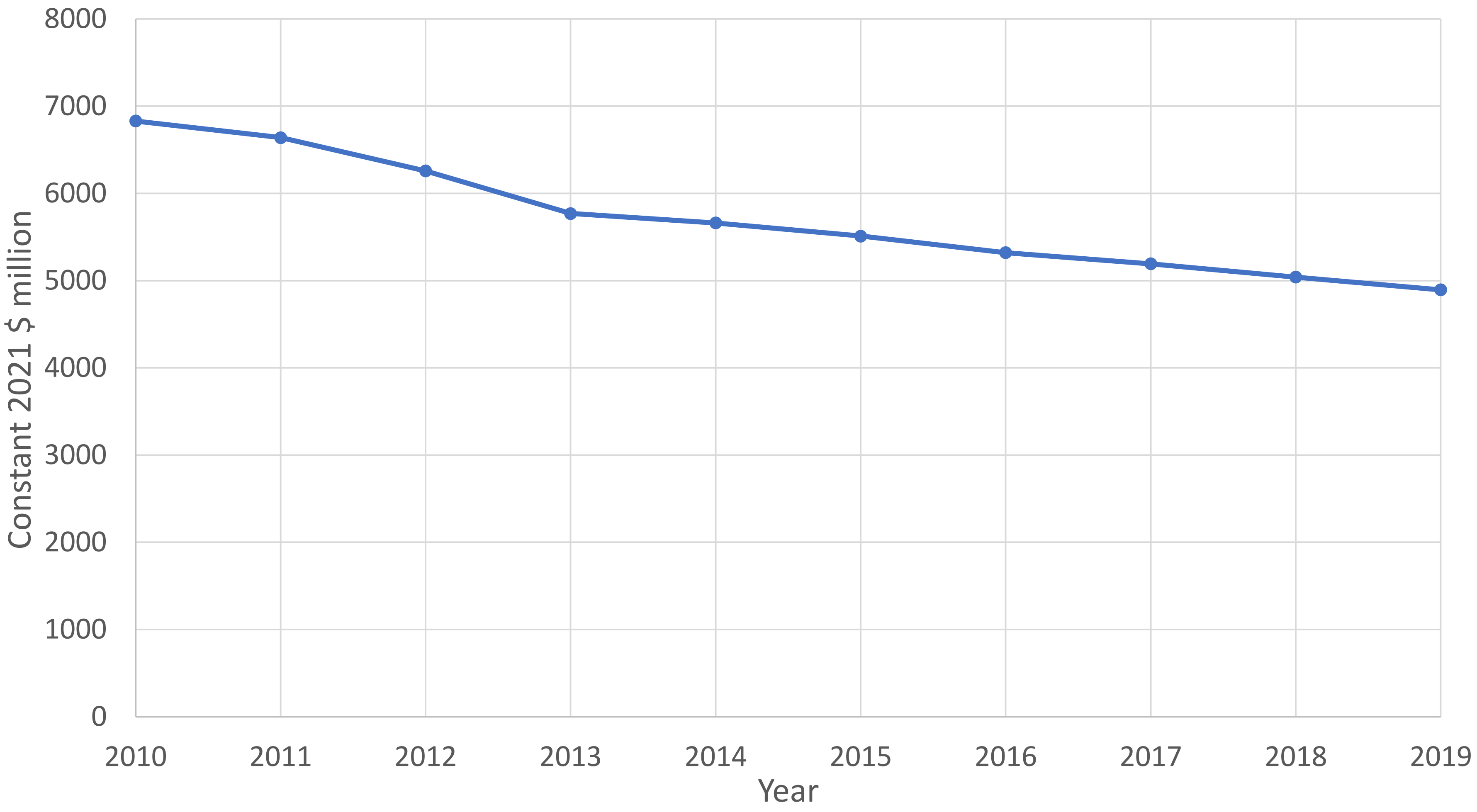
- Field FTD Alert visits
- ASFR
- Notice of Federal Tax Lien filing

These studies have found indirect effects of roughly 1-2 times the magnitude of direct effect of the IRS action. These studies also have certain limitations:

- The estimated effects are limited to similarly noncompliant taxpayers. This offers a lower bound for our analysis.
- There are no estimates of the effects on taxpayers who are currently filing and paying.

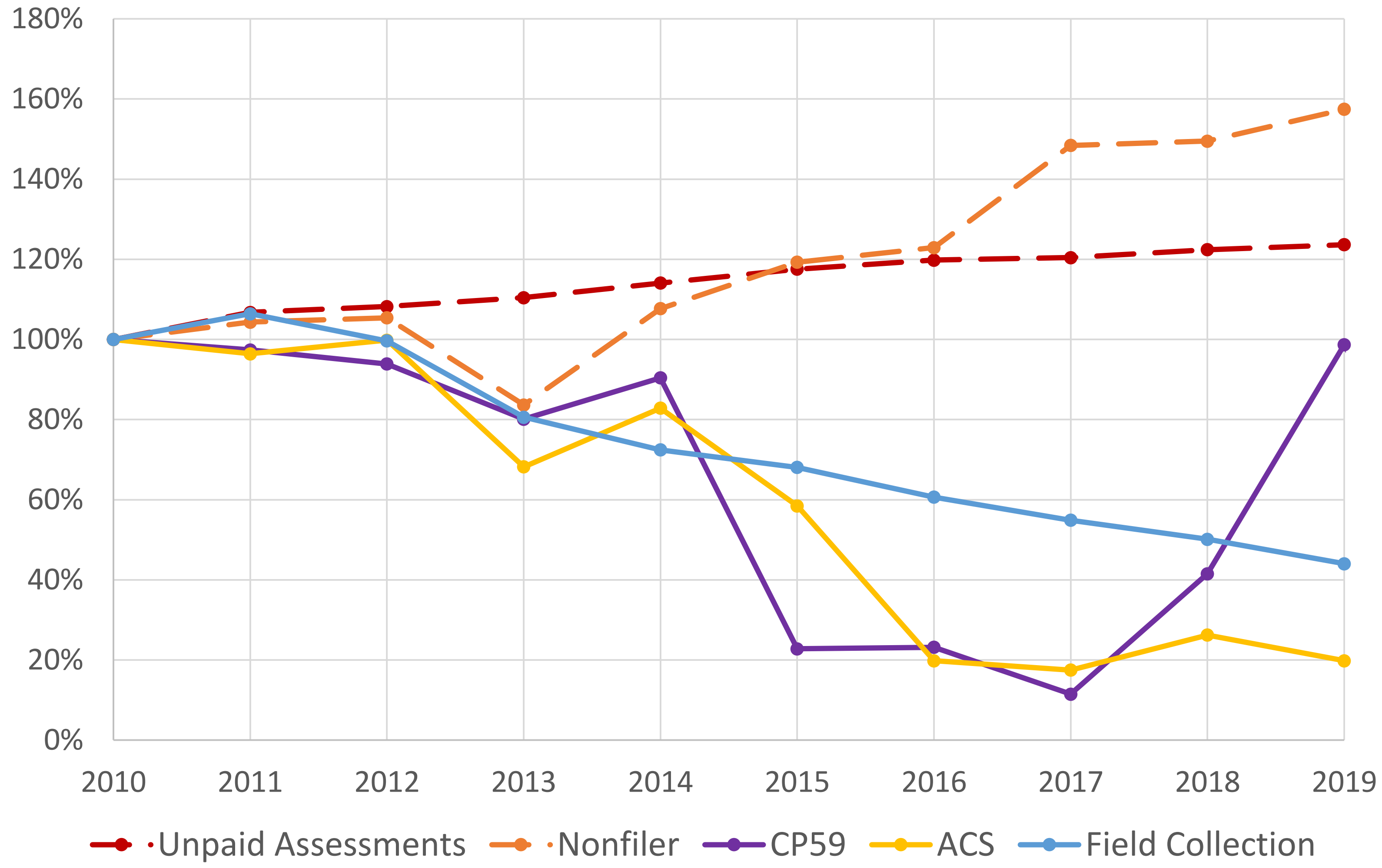
Our study will address this gap by focusing on taxpayers who previously filed and paid on time.

IRS Enforcement Budget in 2021$



source: SOI Data Book, Table 30, Bureau of Labor Statistics CPI-U

Compliance Trends 2010=100%



source: Compliance Data Warehouse

Data

- **Study Population**
  - 1% sample of individual taxpayer population who fully paid and timely filed in the previous year.
  - Use 2011-2019, a period of cuts to compliance programs (natural experiment), but before the COVID-19 pandemic

- **Sample Size**
  - 1% sample leads to a repeated cross-section of compliant taxpayers, about 1.2-1.3 million each year, for a total of 11.6 million

- **Supplemental Sample**
  - About 95% of the compliant population in the sample remains compliant each year, so we sample an additional 10% of previously compliant taxpayers who did not fully pay in the current year for our models that focus on this group
  - The 10% sample for these taxpayers totals about 6 million

Models

- **Two-Stage Logistic Model for Filing and Payment Compliance**
  - To allow for geographic variation in enforcement levels, we use a model that measures aggregate treatments at the zip code level, weighted by the social connectedness index (SCI). To minimize endogeneity in the ACS, CP59, and Field variables—arising from the likelihood that higher rates of non-compliance in a region prompt greater enforcement in the region—we utilize a two-stage least squares (2SLS) method using the IRS budget as an instrumental variable.

- **Multinomial Logistic Model for Filing and Payment Compliance**
  - To differentiate between taxpayers who only file late and those who also fail to pay, we use a multinomial framework with the two-stage approach.

- **Linear Model for Change in Balance Due**
  - Model the indirect effects of compliance programs on the magnitude of their change in balance due if compliance is not reached.

Initial Results

Study Context (2011-2019)

- Annually: 149M returns filed, 125M compliant.
- 2% of previously compliant taxpayers (2.6M) filed late and 3% (3.7M) became delinquent in the subsequent year.

Impact on Interventions:

- ACS Letters: Had the largest impact, significantly reducing both late/non-filings and payment delinquencies.
- Delinquent Return Notices (CP 59): Effectively decreased both late/non-filing and payment delinquencies, though less so than ACS.
- RO Field Contacts: Provided modest but meaningful reductions in non-compliance.
- Additionally, interventions significantly lowered debt amounts for delinquent taxpayers.

We also find that taxpayers with higher reporting compliance risk in the prior year are also at a higher risk of not filing and or paying on time.

Social Connectedness Index Weights

- Bailey et al. (2018) used anonymized linking data from social media to create an objective measure of the intensity of connections between zip code pairs. The social connectedness index (SCI) reflects the density of Facebook friendships between every pair of zip codes in the United States.

- We use the SCI to build weights for each zip code to better reflect the intensity of compliance programs that may result in indirect effects

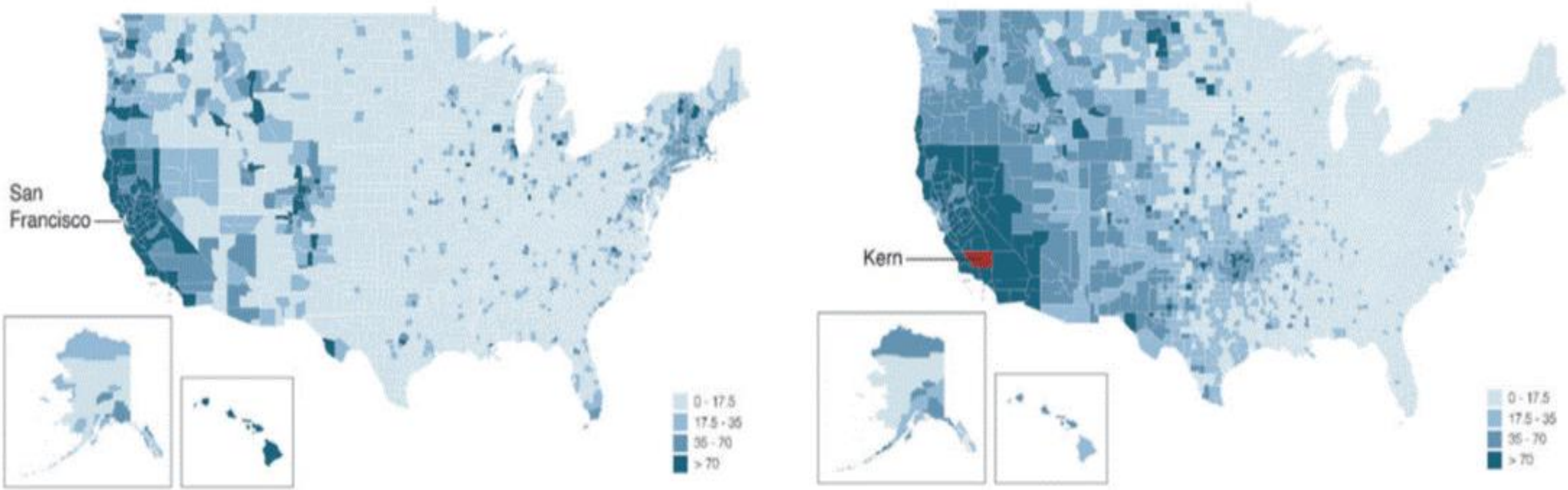
Example of Building SCI weights for ACS letter treatments

$$ACS_{jt} = \sum_{k} w_{jk} ACS_{kt}^{raw}$$

Where ACS_{jt} is the weighted average of ACS letters at zip code *j* in year *t*, w_{jk} is the social connection measure between zip code *j* and *k*, and ACS_{kt}^{raw} is the number of ACS letters sent to zip code *k* in year *t*.

Social Connectedness Index Insights

- Figures show San Francisco’s widespread social connections vs. Kern County’s localized ties, demonstrating SCI’s nuanced approach beyond geographic distance.

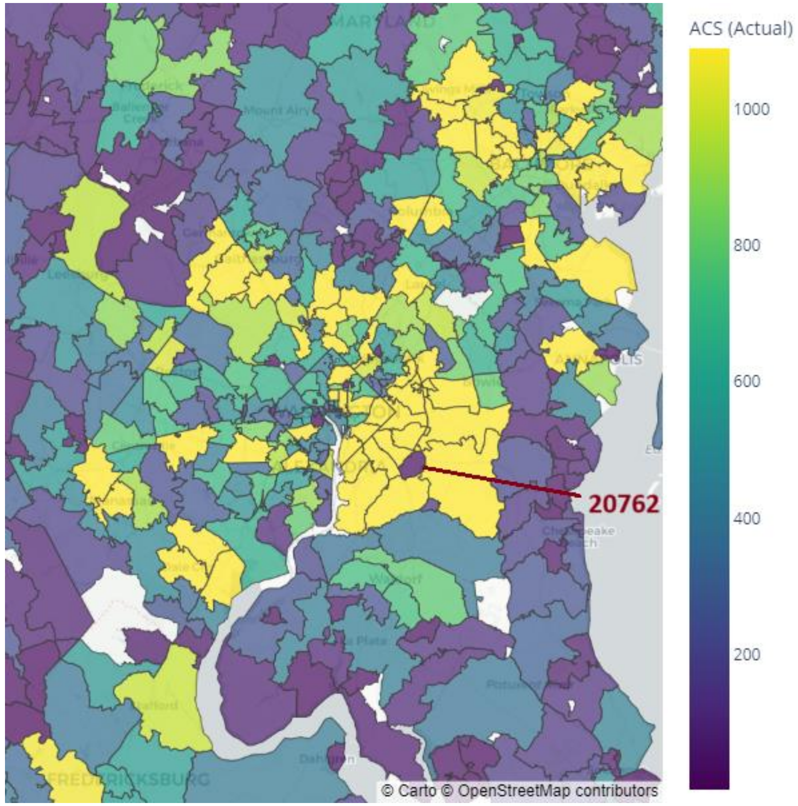


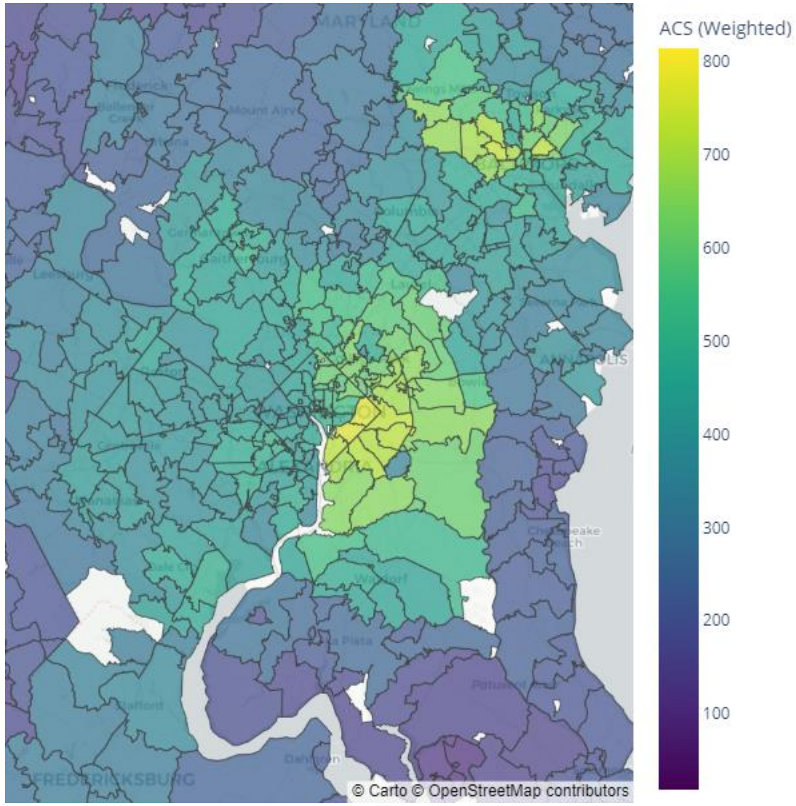
Source: Bailey et al. (2018)

Benefits of SCI Index

Comparison of Raw and SCI-Weighted ACS Letters sent to the Washington D.C. Area in 2011.

- SCI transformation smooths ACS notice distribution, revealing true social dynamics in indirect effect analysis.
- Zip code 20762 (military base) shows fewer notices even after SCI transformation due to social isolation.





Source: Compliance Data Warehouse

Stage One with Instrumental Variable

We focus on three compliance programs, including both campus and field

- Select Automated Compliance System (ACS) letters sent to delinquent taxpayers
- CP59 notices sent to nonfilers
- Field collection cases

Use IRS Enforcement Budget as Instrument to Address Endogeneity

- Model program levels by year and zip code
- Include zip code and year fixed effects

1. $ACS_{jt} = \alpha + \beta Z_t + \gamma_{zip} + \eta_{year} + v$

2. $CP59_{jt} = \alpha + \beta Z_t + \gamma_{zip} + \eta_{year} + v$

3. $Field_{jt} = \alpha + \beta Z_t + \gamma_{zip} + \eta_{year} + v$

The endogenous variables are regressed for each year t and zip code j, ACS_{jt}, $CP59_{jt}$, and $Field_{jt}$ on the IV, which is the annual IRS enforcement budget Z_t, including zip code fixed effects γ_{zip} and year fixed effects η_{year}.

Stage Two Multinomial Logistic Model

In the second stage, the probability of a taxpayer *i* in zip code *j* not filing and paying taxes on time in year *t*, denoted by P_{ijt}, is regressed on the predicted values of endogenous variables, $\widehat{ACS}_{jt-1}$, $\widehat{CP59}_{jt-1}$, and $\widehat{Field}_{jt-1}$ with other control variables X_{ijt-1}, along with zip code (γ_{zip}) and year (η_{year}) fixed effects to account for omitted variables that may influence P_{ijt}

Model Filing and Payment Compliance Using Stage One Predictors and Controls

$$P_{ijt} = \alpha_i + \beta_1 \widehat{ACS}_{jt-1} + \beta_2 \widehat{CP59}_{jt-1} + \beta_3 \widehat{Field}_{jt-1} + \sum_k \theta_k X_{ijt-1} + \gamma_{zip} + \eta_{year} + e$$

- P_{ijt}=2 if the taxpayer has an outstanding balance due at the end of time *t* (taxpayer did not fully pay)
- P_{ijt}=1 if the taxpayer did not file on time but did not accumulate an outstanding balance due (taxpayer paid on time, but filed late)
- P_{ijt}=0 (taxpayer was fully compliant, filing and paying on time)

OLS Model of Change in Balance Due

Following the same general two-stage approach as the overall compliance model, we use the larger (10%) sample of previously compliant taxpayers who became non-compliant and for those who ended the year with new tax debts (P_{ijt}=2), we model the change in their outstanding tax debts after one year, as follows:

Stage 1:

1. $ACS_{jt} = \alpha + \beta Z_t + \gamma_{zip} + \eta_{year} + v$
2. $CP59_{jt} = \alpha + \beta Z_t + \gamma_{zip} + \eta_{year} + v$
3. $Field_{jt} = \alpha + \beta Z_t + \gamma_{zip} + \eta_{year} + v$

Stage 2:

$$\log(U_{ijt}) = \alpha_i + \beta_1 \widehat{ACS}_{jt-1} + \beta_2 \widehat{CP59}_{jt-1} + \beta_3 \widehat{Field}_{jt-1} + \sum_k \theta_k X_{ijt-1} + \gamma_{zip} + \eta_{year} + e$$

Where:

U_{ijt} = the amount of tax not timely filed and paid

For filers, this amount is the total balance on the first notice sent to the taxpayer, for nonfilers it is the balance due on a potential substitute for return (SFR)

Control Variables in X_{ijt-1}

We incorporate a comprehensive set of taxpayer characteristics from their most recent return filed in the previous year *t*-1, including:

- Filing Status
- Log total positive income
- Track record for timely filing
- Balance due (before remittance)
- Under-withholding as a percent of total positive income
- Proportion of income subject to withholding
- Activity code/audit class
- Interaction terms for activity code and discriminant index function (DIF) score (captures numerous risk characteristics)

Comparing SCI and Distance Weights

Variable	SCI Weighted (N=11.6 million)	Distance Weighted (N=11.6 million)	Unweighted (N=11.6 million)
Intercept	-4.762 *** (0.027)	-5.236 *** (0.029)	-4.744 *** (0.008)
ACS weighted average	-1.367 *** (0.009)	-0.081 *** (0.007)	-0.037 *** (0.002)
CP59 weighted average	-0.753 *** (0.005)	-0.039 *** (0.004)	-0.033 *** (0.002)
Field collection weighted average	-0.066 *** (0.000)	-0.002 *** (0.000)	-0.015 *** (0.002)
Married filing jointly	-0.260 *** (0.003)	-0.243 *** (0.004)	-0.237 *** (0.003)
Log total positive income	0.236 *** (0.001)	-0.235 *** (0.002)	-0.233 *** (0.002)
Timely filed in past four years	-0.884 *** (0.003)	-0.873 *** (0.003)	-0.876 *** (0.003)
Balance due (before remittance)	0.232 *** (0.004)	0.232 *** (0.004)	0.234 *** (0.004)
% of income under-withheld	2.576 *** (0.016)	2.603 *** (0.016)	2.582 *** (0.016)
50% or more of income not subject to withholding	0.210 *** (0.005)	0.202 *** (0.005)	0.208 *** (0.005)

Response Variable: P_{ijt} *(0:compliant, 1:non-compliant)*

Multinomial Model Results

Variable	P = 1 (Late filers) (N=11.6 million)	P = 2 (Not fully paid) (N=11.6 million)
Intercept	-0.072 *** (0.026)	-0.135 *** (0.022)
ACS weighted average	-2.095 *** (0.009)	-3.452 *** (0.008)
CP59 weighted average	-1.152 *** (0.005)	-1.926 *** (0.004)
Field collection weighted average	-0.100 *** (0.000)	-0.175 *** (0.000)
Married filing jointly	-0.179 *** (0.004)	-0.008 ** (0.003)
Log total positive income	0.010 *** (0.002)	0.118 *** (0.001)
Timely filed in past four years	-0.322 *** (0.003)	-0.270 *** (0.003)
Balance due (before remittance)	-0.044 *** (0.005)	0.190 *** (0.004)
% of income under-withheld	-0.153 *** (0.016)	1.649 *** (0.015)
50% or more of income not subject to withholding	-0.008 (0.005)	0.099 *** (0.005)

Response Variable: P_{ijt} ***(0: compliant, 1: non-compliant no balance due, 2: non-compliant with balance due)***

Average Marginal Effects

Variable	P = 1 (Late filers) (N=11.6 million)	P = 2 (Not fully paid) (N=11.6 million)
ACS weighted average	-0.148 *** (0.002)	-0.249 *** (0.002)
CP59 weighted average	-0.082 *** (0.000)	-0.139 *** (0.001)
Field collection weighted average	-0.007 *** (0.000)	-0.012 *** (0.000)
Married filing jointly	-0.007 *** (0.000)	-0.005 ** (0.000)
Log total positive income	0.003 *** (0.000)	0.007 *** (0.000)
Timely filed in past four years	-0.021 *** (0.000)	-0.028 *** (0.000)
Balance due (before remittance)	0.002 *** (0.000)	0.010 *** (0.000)
% of income under-withheld	0.025 *** (0.000)	0.088 *** (0.002)
50% or more of income not subject to withholding	0.002 *** (0.000)	0.006 *** (0.000)

Response Variable: P_{ijt} ***(0: compliant, 1: non-compliant no balance due, 2: non-compliant with balance due)***

Averages reflect the expected change for an increase of 1,000 notices or field collection cases

Marginal Effect Estimates for 10% Increase

Compliance Program	Late Filers (P=1)		Delinquent Cases (P=2)	
	Change in Probability	Overall Decrease	Change in Probability	Overall Decrease
ACS Letters	-0.3	15%	-0.5	17%
CP59 Notices	-0.1	5%	-0.2	6%
Field Collection	-0.0007	-	-0.001	-

To estimate national impacts, we calculate the effects of a 10% increase in each program relative to the average level over the study period.

- 10% increase in ACS letters is associated with reductions of approximately 0.3 percentage points in the incidence of late filings and 0.5 percentage points in delinquencies, equating to decreases of 15% and 17%, respectively
- For CP59 notices, a 10% increase results in a 0.1 percentage point decrease in late filings, reflecting a 5% improvement, and a 0.2 percentage point reduction in delinquencies, translating to a 6% decrease among the non-compliant population.
- The impact of field collections is statistically significant, but much more modest

OLS Model Results (Δ Balance Due)

Variable	Parameter Estimate (N=3.5 million)
Intercept	5.503 *** (0.017)
ACS weighted average	-0.019 *** (0.006)
CP59 weighted average	-0.009 *** (0.003)
Field collection weighted average	-0.000 * (0.000)
Married filing jointly	0.068 *** (0.002)
Log total positive income	0.202 *** (0.001)
Timely filed in past four years	-0.177 *** (0.002)
Balance due (before remittance)	-0.185 *** (0.003)
% of income under-withheld	0.337 *** (0.011)
50% or more of income not subject to withholding	0.030 *** (0.003)

Response Variable: $\boldsymbol{log(U_{ijt})}$, ***Change in Outstanding Balance Due***

Parameters reflect the expected change for an increase of 1,000 notices or field collection cases

Estimating Total Change in Balance Due

Compliance Program	Reduction in Outstanding Tax Debts	Percentage Decrease
ACS Letters	$3.5 billion	18%
CP59 Notices	$1.1 billion	6%
Field Collection	$9.8 million	0.05%

To estimate national impacts, we calculate the effects of a 10% increase in each program relative to the average level over the study period.

- 10% increase in ACS letters interventions is associated with a decrease of approximately $3.5 billion in the national balance, representing an 18% reduction.
- 10% increase in CP59 notices results in a $1.1 billion decrease, or 6%.
- 10% increase in field visits correlates with a $9.8 million decrease, or 0.05%.

Conclusions

- Largest impacts from ACS letters
  - May reflect more extensive coverage, reaching 45,000 zip codes and averaging 176 letters per zip code annually
- Through indirect effects even modest increases in program levels may result in significant improvements to compliance
  - These effects are substantial even for taxpayers who have demonstrated previous compliance
- Well distributed enforcement activities that align with the natural communication flows between communities may improve the effectiveness of compliance programs

References

Bailey, Michael et. al., “Social Connectedness: Measurement, Determinants, and Effects” *Journal of Economic Perspectives* vol. 32, no. 3, pp. 259-280

Bloomquist, Kim, “Incorporating Indirect Effects in Audit Case Selection: An Agent-Based Approach” Internal Revenue Service (2012)

Boning, William C., Guyton, John, Hodge, Ronald, and Joel Slemrod, “Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms.” *Journal of Public Economics* vol. 190, issue C (2019)

Datta, Saurabh, Orlett, Stacy, and Alex Turk, “Individual Nonfilers and IRS-Generated Tax Assessments: Revenue and Compliance Impacts of IRS Substitute Assessments When Taxpayers Don’t File” Internal Revenue Service (2015)

López-Luzuriaga, Andrea & Scartascini, Carlos, "Willing but Unable to Pay?: The Role of Gender in Tax Compliance," IDB Publications (Working Papers) 12983, Inter-American Development Bank (2023)

Turk, Alex and Terry Ashley, “Accounts Receivable Resolution and the Impact of Lien Filing Policy On Sole Proprietor Businesses, 2002 Federal Forecasters Conference Proceedings, pp. 323-332 (2002)

Turk, Alex, Iuranich, John, Orlett, Stacy, and Saurabh Datta, “Resolving Unpaid Taxes and the Notice of Federal Tax Lien: Evidence from the Fresh Start Initiative” Internal Revenue Service (2016)

IRS Research, Applied Analytics & Statistics

TPC TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Comments on “Using a Gravity Model to Predict Cross-Border Tax Avoidance,” “Art in the Age of Tax-Avoidance,” and “Staying on the Wagon”

William Boning, U.S. Department of the Treasury

IRS-TPC Conference, June 2024

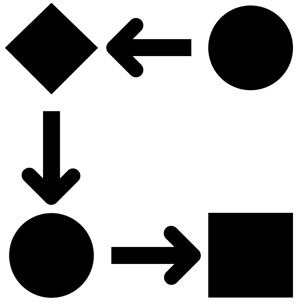
Any opinions and conclusions expressed herein are those of the discussant and do not necessarily represent the views of the Department of the Treasury.

Gravity I: Escaping the Gravity Well

$$Tax.Avoidance = \frac{DIV.OWN.path^{\beta_1}(1-CG.ratio)^{\beta_2}}{\left(\frac{1}{RQ.path}\right)^{\beta_3}(1+EOI.ratio)^{\beta_4}}$$

Is this really a gravity model?

- Missing a mass term
  - Combine with regulator quality
- Lots of distance terms:
  - Ownership rules, tax rates, disclosure rules
  - But distance has a direction here, and is defined over a network graph



Gravity II: Evaluating Model Fit

This model seems well-suited to the context, so prove it:

- Pick a measure of goodness-of-fit (adjusted R^2?, something else?)
- Compare the goodness-of-fit of various models:
  - Vanilla gravity model
  - Your model constrained to sequences of length 2
  - Your model with sequences of length 3

Gravity III: Motivation

The model currently answers the question:

Which sequences of countries are advantageous for tax avoidance?

- But readers will ask: So what? Why is this answer useful?

- Discuss the problems the model is built to solve
  - You do mention identifying foreign countries that may be omitted from returns
  - Behavior of individuals? Businesses? Big MNEs?

Gravity IV: Counterfactuals

Why build a structural model like this one?

Counterfactuals!

1. What would happen if all countries eventually adopted a minimum level of exchange of information?
2. What if all countries had full transparency?
3. What if there were a minimum capital gains tax rate?
4. What if there were no withholding taxes?

Art I: The Art of Reporting Obligations

- What are the reporting obligations…
  - Under tax law?
  - Under accounting standards?
  - Form 990 instructions:

    "Museums and other organizations that elect not to capitalize their collections (according to ASC 958-360-45) shouldn't report an amount on line 1g for works of art and other collection items donated to them."
  - Likewise for Schedule D and Schedule M.
- What are the potential consequences for failure to report?

Art II: Filling in the Background

The appendix defines an audit flag as any of:

- Diversion of assets
- Political activities
- Unrelated business income
- Excess benefit transactions or loans to disqualified persons
- Excess compensation
- Foreign grant activity
- Fundraising expense discrepancies

This is a wide range of things. Discuss them and consider whether to treat some of them differently from the others.

Art III: Apples to Apples

Compare like with like:

- Flow of non-profit-received art vs. flow of art sales
- Stock of tax-exempt art assets vs. stock of art assets
- Annual estimated tax revenue lost vs. annual income taxes

Art IV: What's Trendy?

Exclusion restriction:

Last year's audit rate for non-profits only affects reporting this year through the change in audit probability due to audit flags

- Probability of audit even without audit flags also changes
- Then non-profits respond accordingly
- And many other things that can go wrong

More broadly: audit rates fell steadily over time

- Misbehavior could be trending up or down for various reasons
- The time trends are still interesting and worth discussing

Wagon I: A Bumpy Road to Travel

Estimating network effects is ***hard***

Wagon I: A Bumpy Road to Travel

Estimating network effects is ~~***hard***~~ ***exponentially harder than estimating direct effects***

Wagon I: A Bumpy Road to Travel

Estimating network effects is ~~hard~~ ***exponentially harder than estimating direct effects***

- Even RCTs run into subtle problems and biases
- Without randomization, selection into network treatment is a nasty issue
- Aronow and Samii (2017 AAS):
    - Aggregation bias with heterogeneous treatment effects
    - Even in RCTs controlling for network degree
    - Provide a simulation example and alternative estimators

Wagon II: All aboard

Write for a broader audience

- Define and perhaps rename the treatments: What is an ACS notice and who gets one? What's a CP-59?
- Recap what happens before and after these steps in the enforcement process (flow charts?)
- Simplify language and avoid IRS jargon and acronyms

Wagon III: Assembling the Wagon

Build up from simpler estimates to your network effects

- Direct effects
- Effects of intensity of treatment in the same ZIP
- Then add in the connectedness-across-ZIPs network treatment

Wagon IV: Falling Off the (IV) Wagon

Exclusion restriction

Declines in IRS' enforcement budget affect behavior only through changes in the intensity of these treatments in connected ZIP codes

- Really? Requires that other enforcement has no deterrent effect
- Time trends problem
- Does leadership choose which activities and places have deeper budget cuts? Selection concerns
- ZIPs where enforcement falls most had high pre-period enforcement, so might have different trends
- Connected ZIPs experience common economic shocks (urban, rural)

Run an experiment!

This is a great context for an experiment

- Block randomize
- Pre-register the design and analysis plan

This would be a **huge** contribution

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

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

Janet Holtzblatt

June 13, 2024

Funding, Rescissions, and Responsibilities

What policymakers provided the IRS

Inflation Reduction Act: $79 billion over 10 years

What policymakers rescinded from the IRS

Fiscal Responsibility Act: $1.4 billion

Further Consolidated Appropriations Act: $20.2 billion

What's left? **$57.4 billion**

What's required?

A week after IRA's passage, Secretary Yellen directed the IRS to develop an operating plan with **metrics** and targets.

IRA Goals and Metrics: 2023 Strategic Operating Plan (SOP)

- Established 5 objectives
  1. Support taxpayers—to achieve accuracy in returns and receipt of tax incentives
  2. Quickly resolve taxpayer issues when they arise
  3. Focus expanded enforcement on complex returns, high-dollar noncompliance
  4. Deliver cutting-edge technology, data, analytics for greater effectiveness
  5. More diverse workforce with more service-oriented culture
- Metrics are outcomes (for objectives) and “measure of success” (for initiatives)
  - Sometimes vague and circular
  - Example: *Objective to support taxpayers. Success is levels of service increase.*

IRA Goals and Metrics: 2024 Update to SOP

- Strategic operating plan has evolved, and so have metrics.
- *Objective: Dramatically improve services.*
    - *Outcome: In 2024, 85 percent rate of answered phone calls on the IRS helpline during the filing season with an average wait time of less than five minutes.*
- But still can be circular.
- *Objective: Focus expanded enforcement on complex returns, high-dollar noncompliance*
    - *Outcome: Increase in audit coverage and other types of enforcement of large corporations, partnerships, and high-income, high wealth-taxpayers*

What Do or Should Performance Metrics Measure?

- Effectiveness and shortcomings of allocating funds to IRS relative to other agencies or reducing the deficit
- Improvement (or not) relative to some benchmark
- Allocation of funds—between different IRS budget categories or between types of activities

What Are the Types of Performance Measures?

- Government Performance and Results Act distinguishes between three types of metrics;
  - *Outcome*: Assessment of how well the program achieved its goals
  - *Output*: Tabulation, calculation, or recording of an activity or effort
  - *Service levels*
- OMB encourages agencies to use *outcome* measures when feasible and appropriate but adds two more possible metrics in instructions to agencies:
  - *Inputs* (time or monetary costs)
  - *Efficiency* (the ratio of the inputs to its outputs or outcomes).

Government-wide Annual Requirements

- Government Performance and Results Act of 1993 (amended 2010)
  - Annual measures of agency-wide *outcomes*, *outputs*, *service levels*, *inputs*, and *efficiency*

- Paperwork Reduction Act of 1980 (amended 1995)
  - *Burdens* imposed on individuals and businesses by filling out forms

- Improper Payments Information Act of 2002 (amended 2019)
  - Annual estimates of *improper payments* for programs most susceptible to erroneous payments

IRS and Government Performance and Results Act

- Annual measures of *outcomes*, *outputs*, *service levels*, *inputs*, and *efficiency*
- 25 performance measures included on the IRS list, but list changes over time
- Only two are *outcome* measures
  - Taxpayer satisfaction
  - Repeat non-compliance rate
- Most are *output* measures, such as:
  - Percent of calls to customer service representatives that are answered
  - Number of audits of high-income taxpayers, partnerships, and big businesses *(new)*
- Examples of other measures:
  - Rentable square feet per person (*input*)
  - Costs to collect $100 (*efficiency*)

IRS and Paperwork Reduction Act

- Measures burden imposed on taxpayers from filling out paperwork (*output*)
  - Hours spent on each of the following categories—recordkeeping, tax planning, and form completing and submission
  - Total out-of-pocket expenditures, ranging from payments to preparers and purchases of tax return preparation software to much smaller items such as copying costs and postage
- Reported on forms or instructions
- Doesn't account for other costs incurred by taxpayers in interactions with IRS

IRS and Improper Payments Information Act

- Agency must identify programs and activities that "may be susceptible to significant improper payments." (*output*)
  - Any payment that should not have been made or was made in the incorrect amount (either too much or too little) under the law
- Originally, only earned income tax credit included in IRS's list
- Extended to three other refundable tax credits
- IRS is not required to report on noncompliance for any other tax provision

Principles for Performance Measures

- Outcome measures should be aligned with IRS's mission statement:

  *…provide America's taxpayers top quality **service** by helping them understand and meet their tax and **enforce** the law with integrity and **fairness** to all…*

- Within outcome categories, include measures for output, input, efficiency
- Distinguish between the IRS's role and factors beyond its control
- Consider metrics in context and trade-offs between metrics
- Be explicit about what should but isn't measured
- Numbers don't tell the whole story

Distinguish Between the IRS's Role and Factors Beyond its Control

- It's not just about the IRS
  - Difficult-to-administer tax laws
  - Recessions, natural disasters, and pandemics
  - Budget cuts

- And as with all estimates, methodology always evolving (hopefully, improving)
  - If the tax gap methodology changes, the IRS provides alternative estimate under old approach--but the residual difference can be larger

- Why not adopt the approach taken by OMB/Treasury and CBO in analyzing budget baseline

Example: Changes in CBO's Baseline Projections of Deficit Since May 2023
Trillions of dollars

Legislative changes	-2.6
Economic changes	0.2
Technical changes	1.1
Total deficit changes	-1.4

Metrics in Context

- Too often, focus on one measure
  - How many calls answered?
  - How big is the tax gap?
  - What is the audit rate?

- More emphasis should be placed on combination of metrics
  - Fuller picture of performance of activity
  - Trade-off between activities
  - Links (or broken links) between the mission goals

Outcome Measure—Taxpayer Services
Taxpayer Satisfaction Is Little-Known Metric Without Context

- A GPRA performance measure, derived from private company's survey
- In 2023, 65 percent of Americans satisfied with IRS—but what does that mean?
- Context matters
  - No details on sources of satisfaction or how varies by type of taxpayer
  - No way to link to the specifics of the taxpayer's interaction with IRS
- Alternatives
  - Comprehensive Taxpayer Attitude Survey
  - GSA's Touchpoints Survey (user experience for Direct File pilot)
    - What are costs to extending to other IRS products?

Output Measure 1—Taxpayer Services
Percent of Calls Answered May Be the Misunderstood Metric
Too Many Telephone Numbers?

`When Announced	Percent of calls answered
April 2023	85
Aprii 2023	87
May 2023	52
March 2024	52
	66
April 2024	84

Output Measure 1—Taxpayer Services
Why is a Measure of Answered Phone Calls so Confusing?

- For 2023, the IRS's *level of service* ratios been as high as 87% and as low as 52%.
  - Definition matters—higher rates tend to include calls with automated responses
  - Timing matters—higher rates tend to cover just filing season
  - Only about half of telephone calls to IRS are included in any LOS metric
- Context matters
  - The annual rates give a better perspective on trade-offs between goals
  - Need to also consider:
    - Combined effect with other existing performance metrics (accuracy)
    - Other aspects of phone service—such as hang-ups (Taxpayer Advocate)
    - Other aspects of responsibilities of taxpayer service (answering mail)
    - When lower scores for telephone calls are a good thing.

Output Measure 1—Taxpayer Services
Percent of Calls Answered May Be the Misunderstood Metric
Too Many Telephone Numbers?

`When Announced	Percent of calls answered	What's happening?
April 2023	85	Werfel testimony on April 27 (thru April 14)
April 2023	87	Treasury release on April 17 (filing season)
May 2023	52	TIGTA (thru May 13; only calls during working hours)
March 2024	52	IRS budget justification (entire year)—no automated calls
	66	Same as above, with automated calls
April 2024	84	IRS press release (filing season)

Output Measure 2—Taxpayer Services Compliance Burden Is the Misnamed Measure

- Kudos to the IRS for the surveys and microsimulation models for individual and business taxpayers
- What needs to be done?
  - More detail on costs incurred by different types of taxpayers
  - Meets requirements of PRA—but what about other costs beyond filling out form
  - What about burden incurred by taxpayers who try but don't use form—or file a return
- Context matters: "Compliance burden" is a misnomer. How much does it cost to fill out:
  - Return with no errors?
  - Return with inadvertent errors?
  - Return with intentional errors?

Outcome Measure—Enforcement Tax Gap is the Metric without a Mandate

- Source of data: Random sample of individual taxpayers (NRP), administrative data, household surveys
- Issues are well-known
  - Undetected income and detection control model (DCM)
  - Undetected errors to the advantage of the taxpayer
  - Stops with auditor’s recommendations
  - Gray areas of tax code
- Context matters
  - Perhaps more than other metrics, need to decompose sources of tax gap
  - Link to compliance burden

Output Measure—Enforcement
Number of Audits Are the New Performance Measure

- Audits joined the ranks of performance measures in 2022—just the number and just for certain groups of taxpayers
- Context matters
  - Ultimately what matters is the audit rate, not the number of audits, but info on audit rates lag
  - No-change rate
  - Non-response rate
  - Factors associated with noncompliance
  - Burdens on compliant taxpayers

Audit Rates, Closure Rates, and No-Change Rates
2018 Individual Income Tax Returns, as of end of FY 2023

Positive Income1	Audit Rate (%)	Closure Rate (%)	No-Change Rate (%)
Under $100,000*	0.3	99	12
$100,000 to $500,000	0.2	97	15
$500,000 to $1 million	0.4	87	22
$1 million or more	1.6	77	33
Total	0.3	95	12
Addendum *$1 million or more, assuming all remaining cases result in a change in tax liability*	*1.6*	*100*	*25*
** EITC (included in under $100,000)*	*0.9*	*100*	*13*

Efficiency Measure—Enforcement
Return on investment (ROI)—Metric du jour?

- ROI used to estimate the amount of revenue raised by increasing IRS funding
- Flew under radar for many years.
  - Because of budget scorekeeping rules, can’t score revenues from increase in IRS funding.
  - Historically, estimated just for program integrity programs
- Treasury and CBO used similar methodology for many years
  - ROI derived from ERIS
  - Adjustments for learning curbs (for new employees and—just CBO—for would-be noncompliant taxpayers)
  - Collection rates over time from other IRS data

Efficiency Measure—Enforcement ROIs Present New Estimating Challenges

- Even though scorekeeping rules did not change, ROIs and IRS revenues took center stage in IRA debate—perhaps for obvious reasons
  - Initial gross estimates: Treasury at **$320 billion** and CBO at **$220**
  - Post-IRA enactment: Treasury at **$390 billion** and CBO at **$160 billion**
- CBO and Treasury have been revising methodology—with Treasury now at **$497** billion—though with different parameters
  - Voluntary direct compliance
  - Impact of capital investment on productivity
  - Start-up lags (CBO only)
- If scope of ROIs included certain other activities, Treasury estimates revenues up to **$851 billion**

Efficiency Measure—Enforcement ROIs in Perspective

- Do ROIs underestimate costs?
- What should be the scope of ROIs?
- Do we lose sight of other goals with focus on ROIs?

Outcomes and Outputs Measures: Equity
Equity is the Emerging Metric

- New focus for tax administration
  - Gale: Compliance costs by AGI
  - Tax gap by income: Johns and Slemrod, DeBacker, Guyton et al, Auten and Splinter
  - Number of audits by income (IRS performance measure)
  - Audit rates by race and ethnicity: Stanford study

- Underlying data limits scope of the equity studies
- Added challenge of imputing the unobserved-undetected income and race and ethnicity
- Focus on disparities for some groups without comparable analysis of other groups

Conclusion

- Aspirational
- Metrics should provide insight not just on success but also on trade-offs
- Metrics not sufficient without digging deep into reasons
- Transformation of IRS is opportunity to transform performance measures—or at least, increase awareness of limitations of the metrics

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Research on Audit Rates by Race & Ethnicity: 2024 Update

Presented to IRS-TPC conference, June 13, 2024

By the RAAS-RICS Collaboration on Exam Disparity:

Brandon Anderson, Keegan Brown, Damon Frezza (MITRE), Zane Girouard, Alissa Graff, Tom Hertz, Navya Kambalapally (MITRE), Piper Kurtz, Kara Leibel, Justin Nave (MITRE), Mark Payne, Daniel Rodriguez, Brian Sartain

Thanks also to Ish Alejo, Andrea Cannon, James Chou, Jim Clifford, Frank Cousin, Elena Derby, Holly Donnelly, Connor Dowd, Hadi Elzayn, Robin Fisher, JC Garnish, Geoff Gee, Jacob Goldin, Cam Guage, John Guyton, Joe Hancuch, Anne Herlache, Chris Hess, Daniel Ho, Ron Hodge, Wesley Janson, Drew Johns, Barry Johnson, Nancy Jones, Kye Lippold, Jake Mortenson, Stephanie Needham, Kevin Pierce, Dean Plueger, Alan Plumley, Arun Ramesh, Pete Rose, Dan Rosenbaum, Evelyn Smith, Ben Swartz, Alex Turk, Masanue Vah, Chloe Zheng.

Research on Audit Rates by Race & Ethnicity: 2024 Update

Disclaimers:

- The IRS does not collect data on taxpayer race. Instead, race was imputed using Bayesian Improved First name Surname Geocoding (BIFSG), which assigns each taxpayer a probability of belonging to each of six race/ethnicity categories by matching taxpayers' names and addresses to race/ethnicity distributions drawn from public sources. These estimated race data are used for research purposes only; the IRS does not and will not consider an individual's imputed race as part of the case selection process.

- This document reflects the views of the authors, one of whom (Hertz) is also an author of the paper by Elzayn et al. (2023). This work is preliminary and pre-decisional and is being shared in the interest of eliciting constructive feedback to improve our understanding of the issues. The perspectives and findings expressed herein should not be taken to represent IRS or Treasury Department Policy.

Recap of previous findings

- Audit rates for Black taxpayers in TY2014 were 3x to 5x higher than for non-Black taxpayers.
- This disparate impact was driven both by differences in audit rates by race *among* EITC claimants and by the fact that audit rates overall were higher for EITC claimants than for non-claimants. Similar disparities have been found in all years examined from TY2010 through TY2022.
- These disparities cannot fully be explained by group differences in rates of noncompliance:
  - If noncompliance is defined in terms of the *total tax understatement* on an EITC return, rather than the portion that is related to *overclaimed refundable credits*, then gross-revenue-maximizing models would select Black EITC claimants at *lower* rates than other claimants.
  - However, correspondence audits cannot determine this total tax understatement – that requires a full scope field exam, which increases the cost of the audit.
- Historically, data limitations such as missing parental social security numbers have made it difficult to determine the eligibility of dependents claimed for EITC; and racial differences in the effects of these limitations have raised the relative audit rate for Black taxpayers.

Recap: Use of high-risk preparers correlates with audit rate disparity

- The Refundable Credits Return Preparer Strategy (RCRPS) program has identified 87,000 high-risk registered return preparers since 2005. In TY2019, preparers on this list submitted 17M tax returns.
- By applying the BIFSG race data to the full population, we can show that clients of RCRPS-identified preparers are disproportionately drawn from minority communities.

	All taxpayers			EITC Claimants		
	RCRPS treated preparer?			RCRPS treated preparer?		
Race/ethnicity*	Yes	No	All	Yes	No	All
Black	19%	11%	12%	26%	18%	20%
Hispanic	40%	14%	16%	39%	21%	25%
White	30%	67%	63%	25%	53%	47%
Other	11%	8%	8%	10%	8%	8%
Count of returns (M)	17.1	140.8	157.9	5.8	20.5	26.3

- Audit rates are higher for clients of high-risk preparers and that raises the relative audit rate for Black taxpayers. If we isolate returns not generated by high-risk preparers and recalculate overall disparity, it falls by 13% among EITC claimants and by 21% overall.
- This may reflect the effects of these preparers improperly advising their clients, or of differences in client characteristics.

	All taxpayers			EITC claimants		
	RCRPS-identified preparer?			RCRPS-identified preparer?		
Audit rates by race	Yes	No	All	Yes	No	All
Black	1.63%	0.47%	0.67%	3.05%	1.45%	1.90%
Nonblack	0.57%	0.18%	0.22%	1.33%	0.65%	0.79%
Disparity ratio	2.9	2.7	3.1	2.3	2.2	2.4
Change in disparity if drop clients of identified preparers			-21%			-13%

Note: These calculations, based on the BIFSG race/ethnicity probabilities, likely understate the share of EITC claimants who are Black, and the share of preparer clients who are Black. The -21% change in disparity is calculated after first subtracting one from the disparity ratios, so is given by (2.7-1) / (3.1-1).

Summary of recent accomplishments

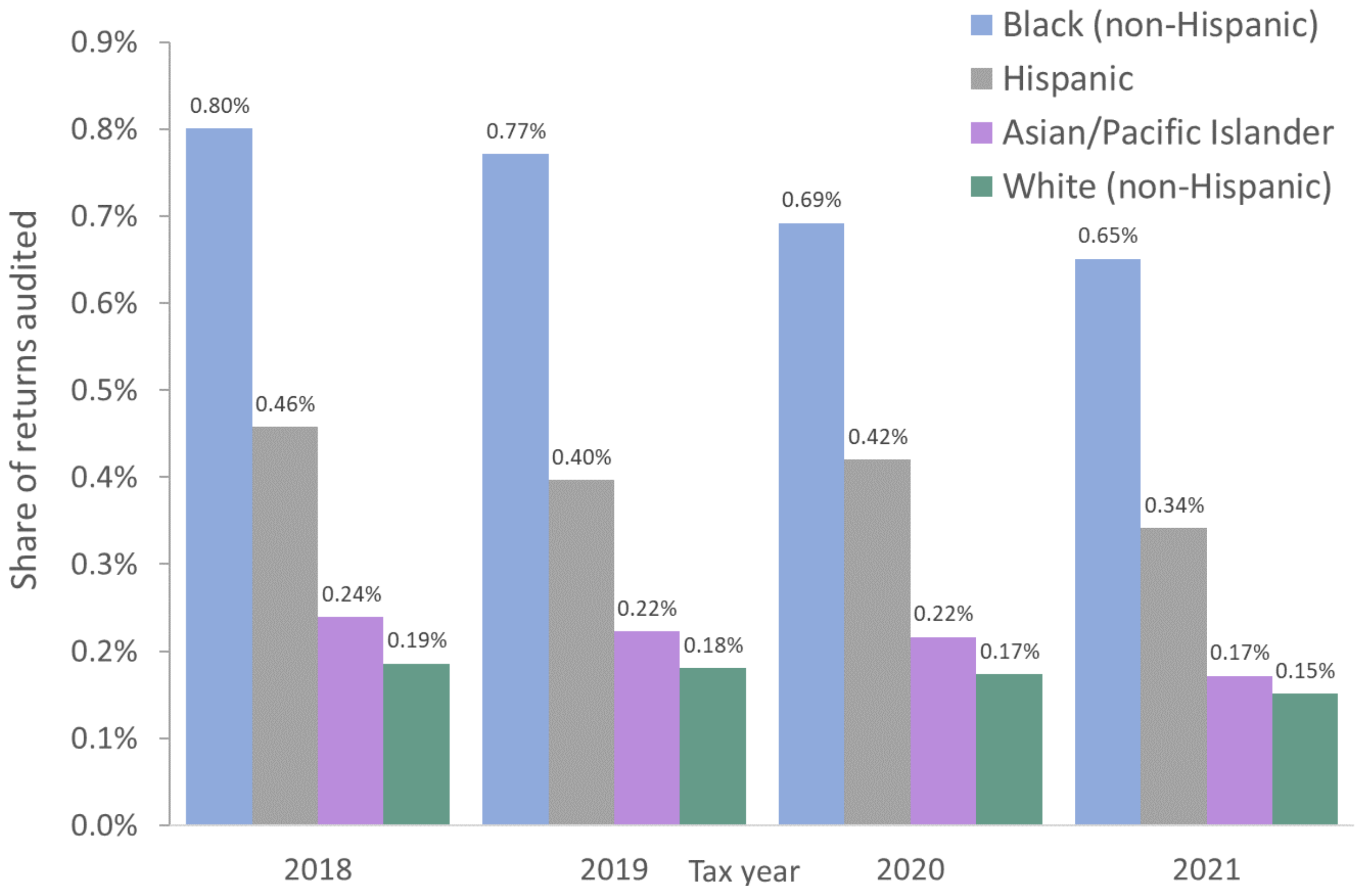
Since last year's IRS-TPC Conference, the IRS has:

- Finalized an agreement with the Census Bureau that may improve our ability to estimate tax and enforcement outcomes by race and ethnicity
- Made significant progress in reducing missing data on parental SSNs, which have contributed to racial disparity in EITC audit rates
- Introduced a new EITC risk scoring system that we hope will better align audit rates with noncompliance risk. Audits started in the first quarter of 2024
- Finalized an innovative pilot EITC audit model that is expected to both improve audit outcomes in dollar terms and reduce racial disparity in audit selection rates, compared to status quo methods, while generating the data needed for further iterative improvements.
- *Today:* Reported audit rates for additional demographic groups

Improved methods for race imputation

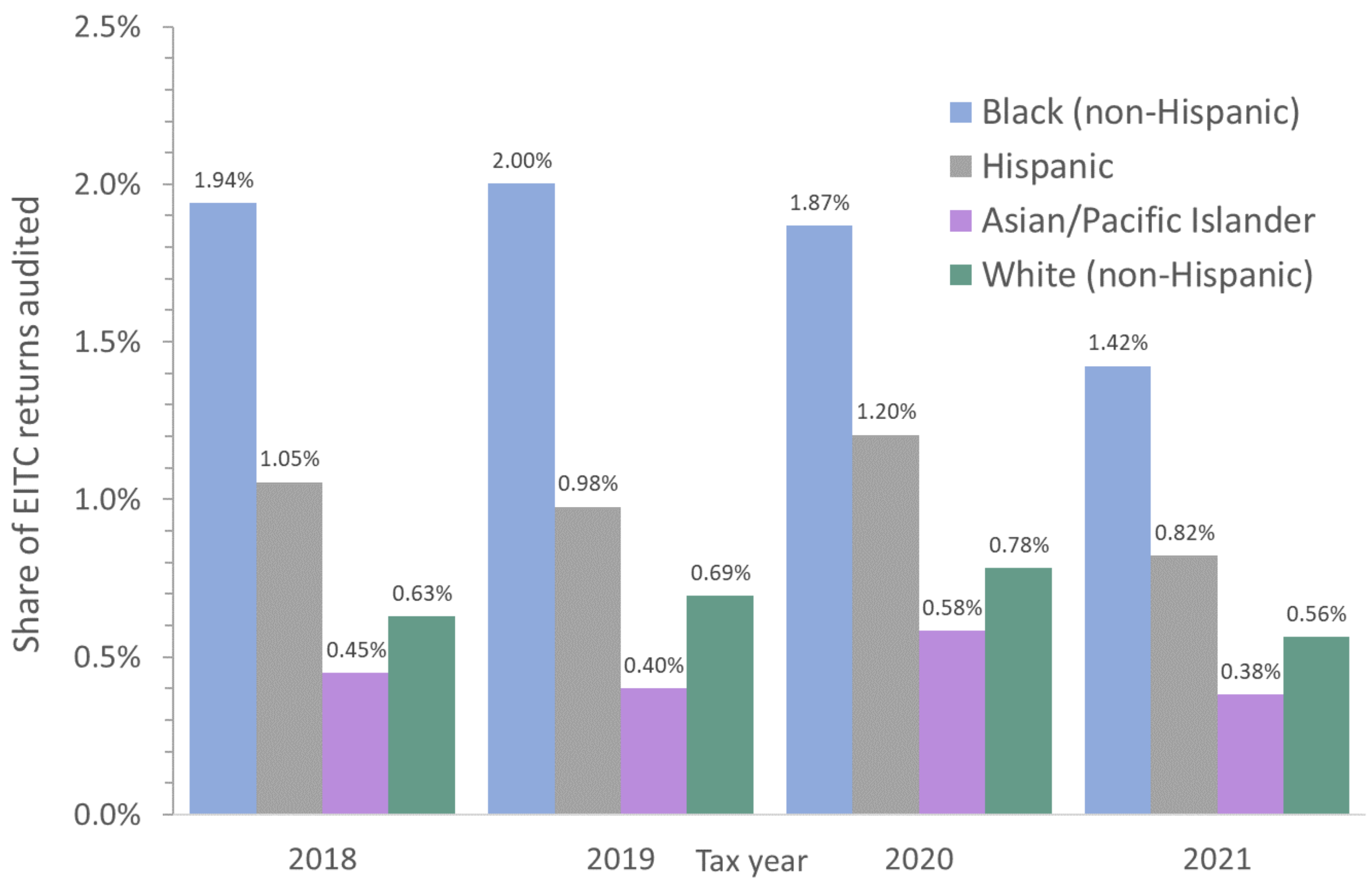
- The BIFSG method of inferring race/ethnicity probabilities cannot yield statistically unbiased estimates of racial differences in mean outcomes. Under certain conditions, however, these biases may be bounded (see Elzayn et al, 2023).
- The Census Bureau has agreed to provide differential-privacy-compliant noise-infused race/ethnicity data to IRS, at the population level. These data will permit the unbiased estimation of outcomes by race/ethnicity. We hope this method will replace BIFSG in future.
- We are also working to improve BIFSG using better data on the race/ethnicity distribution of first names, and by linking taxpayers to returns filed in 2010 and 2020, which are Decennial Census years. This works for the 97% of taxpayers who can be found on a return filed in one of those two years.
- This should reduce both bias and variance in the BIFSG probability estimates:
  - It permits us to estimate neighborhood demographics at the block level, rather than the block-group level (the finest geography in the American Community Survey (ACS), our previous source).
  - It eliminates the component of variance that is due to the ACS being a relatively small sample of the population.
- *Note*: The estimates reported today do not yet reflect either of these methodological improvements. They are based on the methods outlined in the paper by Elzayn et al.

New findings: Audit rates by race & ethnicity, all returns, TY18-21



Notes: (1) Includes open exams. (2) Race/ethnicity data are missing for 5% of taxpayers in TY20 and 7% in TY21. (3) The BIFSG method tends to understate the difference in audit rates between White taxpayers and members of racial and ethnic minority groups. (4) These data will be updated using more accurate race imputations.

Audit rates by race & ethnicity, EITC returns, TY18-21



Notes: (1) Includes open exams. (2) Race/ethnicity data are missing for 3% of EITC claimants in TY20 and 8% in TY21. (3) The BIFSG method tends to understate the difference in audit rates between White taxpayers and members of racial and ethnic minority groups. (4) These data will be updated using more accurate race imputations.

New scoring model implemented

- A new EITC risk-scoring system was developed and tested starting in 2020 (prior to the discovery of racial disparity in EITC audit rates). This regulates the primary EITC pre-refund audit workstreams.

- In pilot testing, it was found to raise revenue and reduce the no-change rate compared to the prior scoring system. After these successful tests, the new score was implemented in January of 2024.

- We will evaluate the impacts of this new model and use the results to inform future updates to improve both equity and revenue outcomes.

Filling in missing data on parental social security numbers

- Errors in the imputation of child residency and relationship status have a disparate impact on Black EITC claimants, raising their audit rates relative to non-Black claimants.
- One source of these errors is the fact that parental Social Security Numbers (SSN) are often missing, particularly for Black and Hispanic fathers, and for Hispanic mothers.
- IRS's IT division has been working with SSA to backfill these missing parental SSNs and has been able to reduce missing values considerably, particularly for mothers' SSNs. This should improve accuracy of audit selection.
- However, missing data rates remain high for fathers, and are still higher for Black and Hispanic fathers than for White fathers claiming EITC with dependents, so there is more work to be done.

	Missing Father's SSN				Missing Mother's SSN			
	TY21	**TY22**	**TY23**	**% Change in missing data**	**TY21**	**TY22**	**TY23**	**% Change in missing data**
Black	45%	42%	40%	-11%	7%	3%	2%	-71%
Hispanic	42%	32%	31%	-27%	21%	8%	7%	-64%
White	27%	23%	22%	-19%	6%	2%	1%	-78%
All	36%	30%	29%	-21%	12%	4%	4%	-69%

Notes: This looks at children born in 2005 or later and claimed for EITC in either TY21 or TY22. For this fixed cohort, we then calculated the share whose parental SSNs were non-missing and available for use at the time audit selections were made for each of the tax years shown.

Pilot EITC audit selection models

- Two pilot audit programs have been initiated with the joint goal of improving audit outcomes and reducing racial disparity:
    1. A model designed to detect erroneous Schedule C expense deductions among EITC claimants is now in field testing.
    2. A second model, designed to better identify EITC compliance issues of all kinds (including invalid residency and relationship status of dependents) will be fielded later this year.
- The pilots make use of new data sources (such as Forms 1095 which shed light on household composition) and improved machine learning methods to increase model accuracy.
- The pilot projects will shed light on a range of issues:
    - Do models trained in past operational audit data perform better or worse than models trained in nationally representative random audit data?
    - Does the exclusion of nonrespondents from operational audit training datasets lead to less biased models? (GAO have emphasized the issue of nonresponse biases).
    - Do models that omit features that are both highly correlated with race and highly influential in the determination of the predicted outcomes achieve lower bias with comparable performance (as found by Elzayn et al, in a non-operational context)?
    - Do models that are trained to detect large dollar values of noncompliance, at the cost of a higher no-change rate, result in fewer Black taxpayers being selected (as found by Elzayn et al, in some but not all non-operational contexts)? (GAO have emphasized this issue as well).
    - What is the relation between the size of the Schedule C businesses audited, the demographics of the taxpayers affected, and the resources needed to perform the audits?

Lessons learned to date from pilot model development work

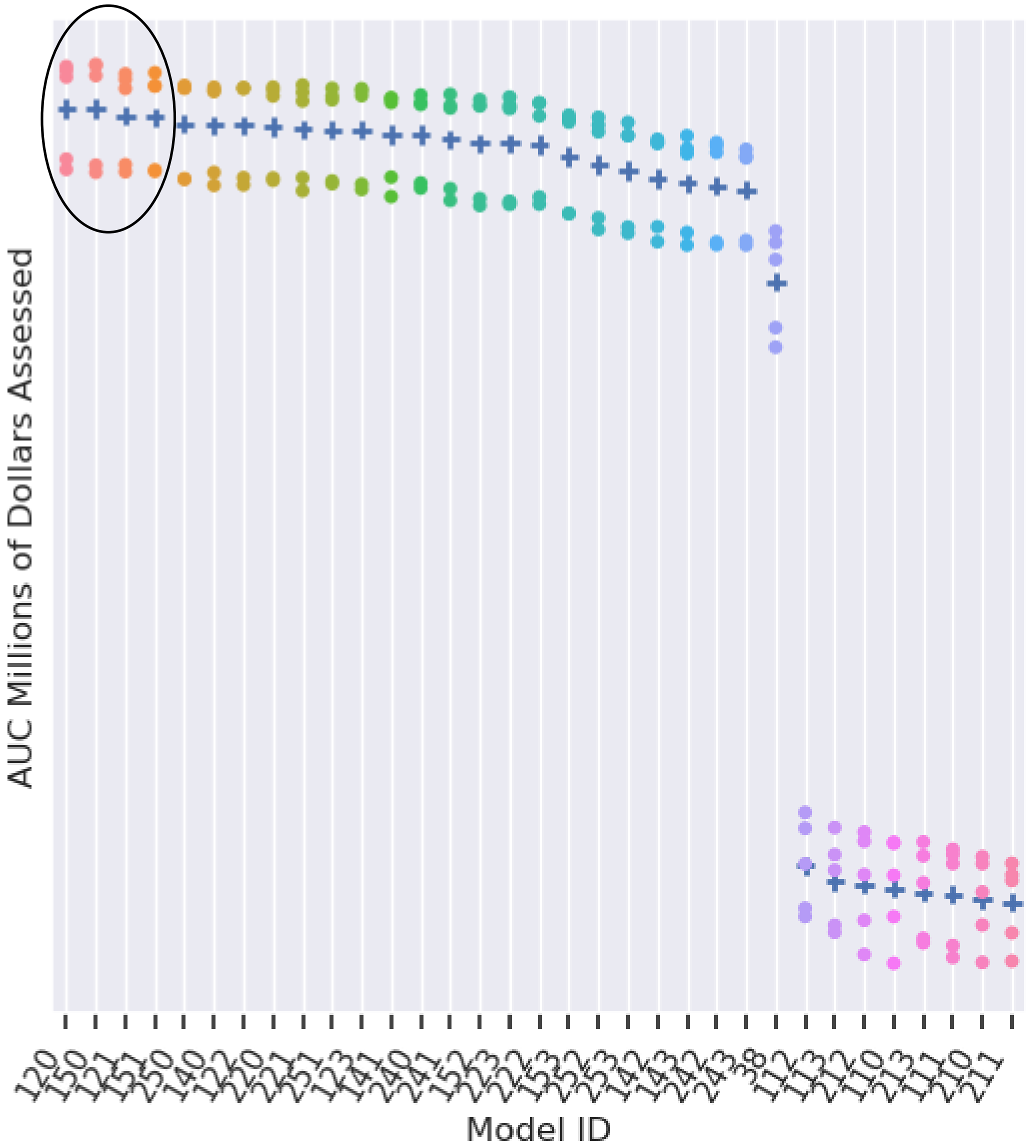
- The analysis of past operational audit data has already provided some important insights regarding the relation between business size, taxpayer demographics, and the exam durations.
- The following table illustrates that prioritizing larger Schedule C businesses reduces the share of Black taxpayers in the audit-eligible population, but also requires longer exam durations.

Gross Receipts + Other Income	Expected share Black	Avg. exam duration (hrs)
Low	40%	1.4
Medium	25%	2.7
High	20%	3.4
Highest	17%	9.8

Note: These are *forecasts* based on past operational data from similar audit projects, not the actual pilot program results; the pilot program is designed to permit better estimates of these relationships, while also reducing disparity and raising revenue compared to status quo processes.

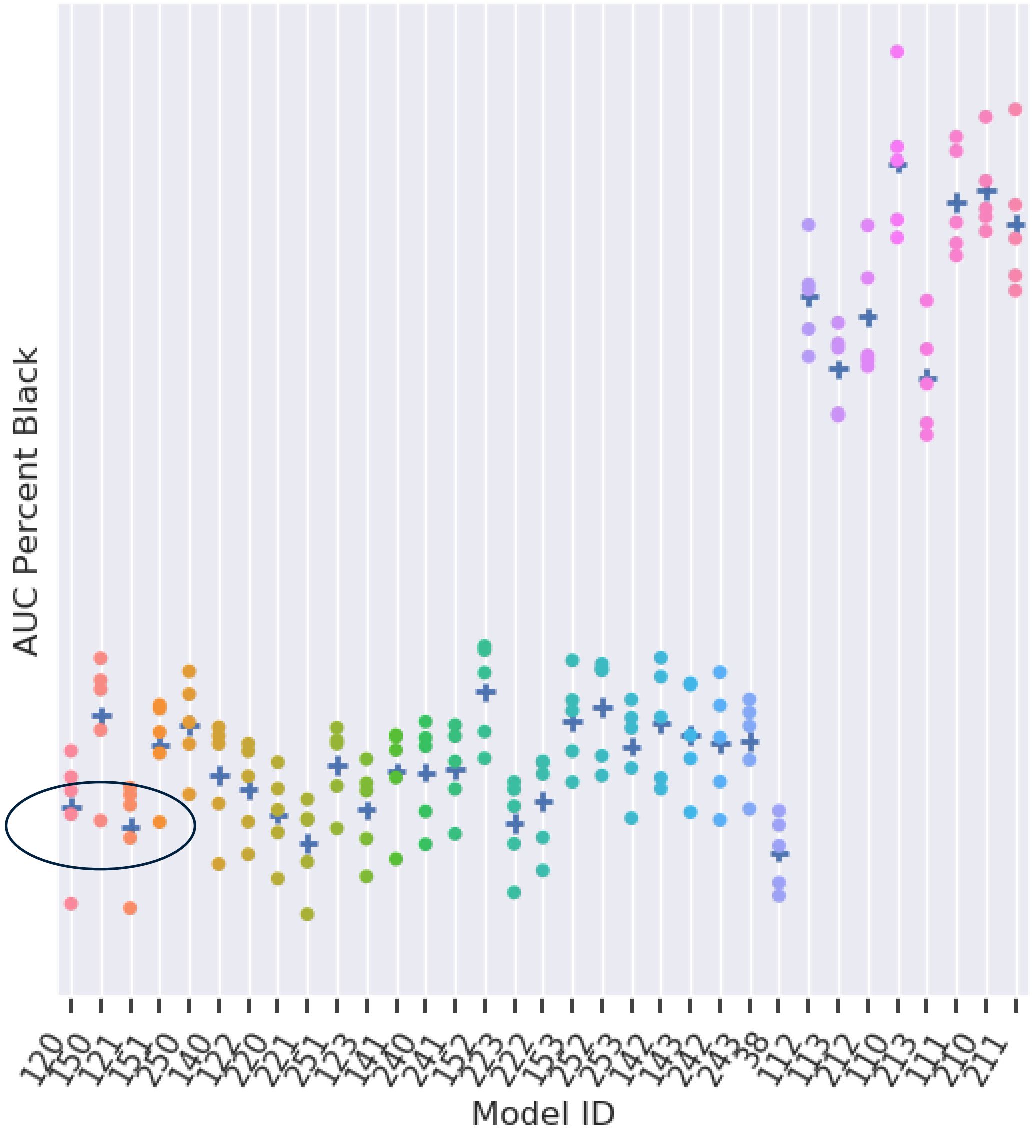
Log Models and Two Stage Models: Maximize revenue per case

The four data points circled represent the top four performing models, in terms of dollars per case.



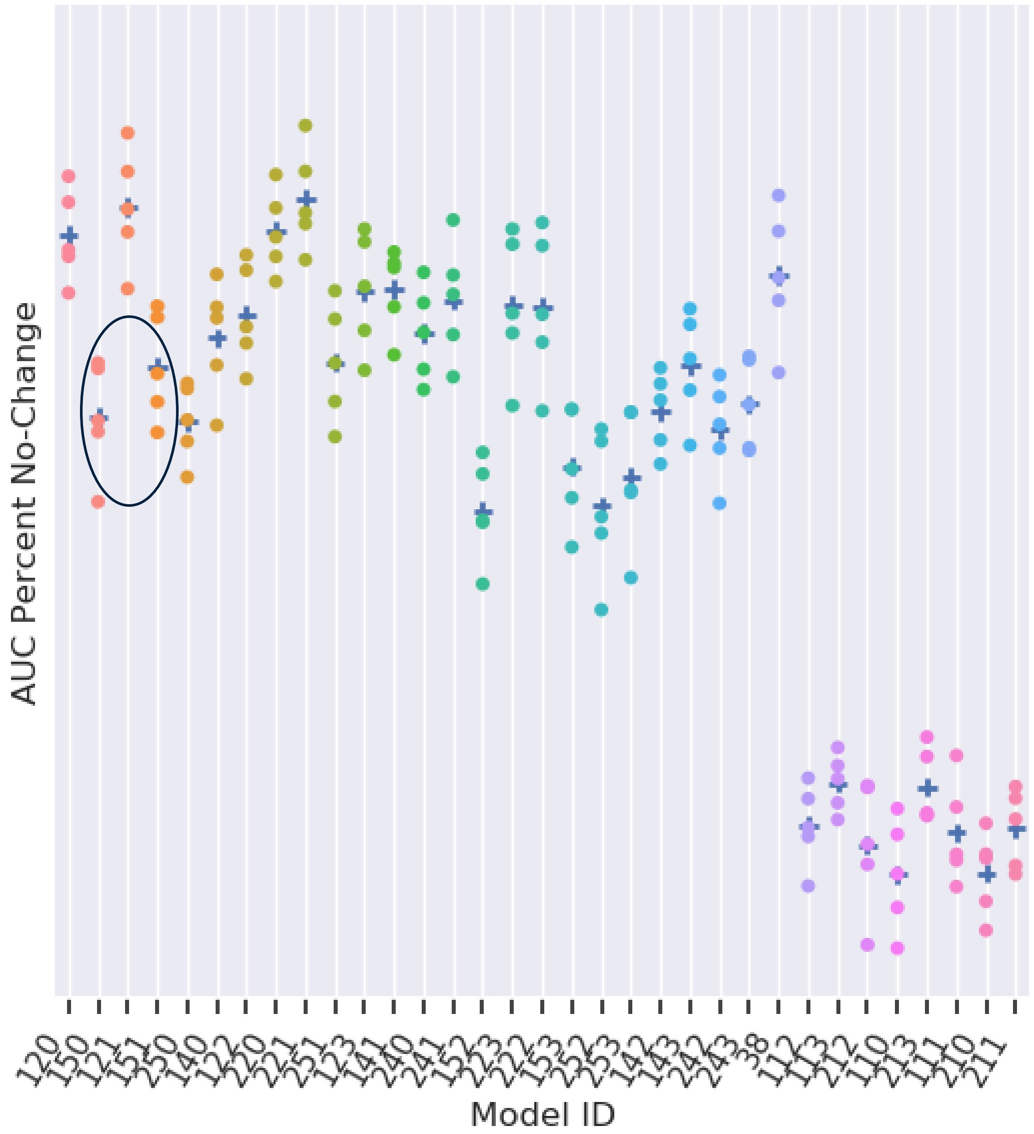
Log Models: Lowest disparity among high-yield models

Of the four high-performing models from the previous slide, the two models circled here (models 120 & 121) selected the lowest share of Black taxpayers.



Two Stage Models: Lowest no change rates among high-yield models

However, the two models shown here (models 150 and 151) achieve the lowest no-change rates.



Summary

The IRS' Strategic Operating Plan commits us to designing "enforcement actions that appropriately reflect risk and level of noncompliance and address enforcement disparities."

The research presented today reflects that commitment in the following ways:

- We are working to improve the quality of the data used to impute race and ethnicity categories, with help from the Census Bureau.
- We have made significant progress in improving the data used to impute relationship and residency status, and expect further improvements via the incorporation of the 1095 data.
- Our first pilot project will allow us to test the hypothesis that models trained only on data from respondent taxpayers can significantly reduce racial disparity in audit selection without loss of revenue in some contexts.
- Both pilot models are also designed to provide the data needed for iterative improvements.

We are working as quickly as possible to improve our audit selection algorithms, but the development and testing of new models takes time. As this work progresses, we will continue to learn, continue to share our findings, and continue to improve our processes.

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

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

IRS-TPC Research Conference on Tax Administration

June 13, 2024

M. L. Szulczewski, M. Feldman, S. Silva
MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

A. Graff, B. Anderson
IRS Research, Applied Analytics & Statistics
KNOWLEDGE DEVELOPMENT & APPLICATION

The findings, interpretations, and conclusions expressed in this presentation are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury or the Internal Revenue Service. All results have been reviewed to ensure that no confidential information is disclosed.

Overview

EXECUTIVE ORDERS

Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

"Agencies must therefore design, develop, acquire, and use AI in a manner that fosters public trust"

We explored, developed, and tested 3 tools to foster trust for a prototype AI system

Data Card

[Write the dataset name here (and any aliases)]

1 Dataset Identification

1.1 **Summary.** Briefly describe the dataset in plain language.

1.2 **Creator.** Who created the dataset (i.e., what research group, agency, or division)? Provide contact information if available (e.g., name, email, etc.)

1.3 **Points of Contact (POCs).** Identify POCs who can answer questions about the model or provide assistance (list names, email addresses, etc.).

1.4 **Release date.** When was the dataset made available?

YYYY-MM-DD

1.5 **Size.** What is the size of the dataset in bytes?

1

Model Card

[Write the model name here (and any aliases)]

1 Model Identification

1.1 **Model type.** What is the model type (e.g., convolutional neural network)?

1.2 **Task.** What is the model task (e.g., regression, classification, anomaly detection, etc.)?

1.3 **Creator.** Who created the model (i.e., what research group, agency, or division)? Provide contact information if available (e.g., name, email).

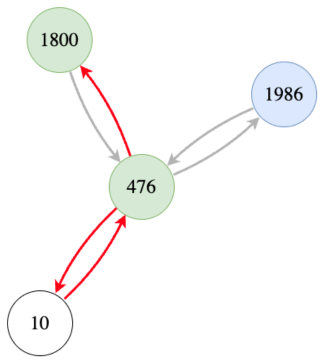
1.4 **Points of Contact (POCs).** Identify POCs who can answer questions about the model or provide assistance (list names, email addresses, etc.).

1.5 **Creation date.** When was the model created?

YYYY-MM-DD

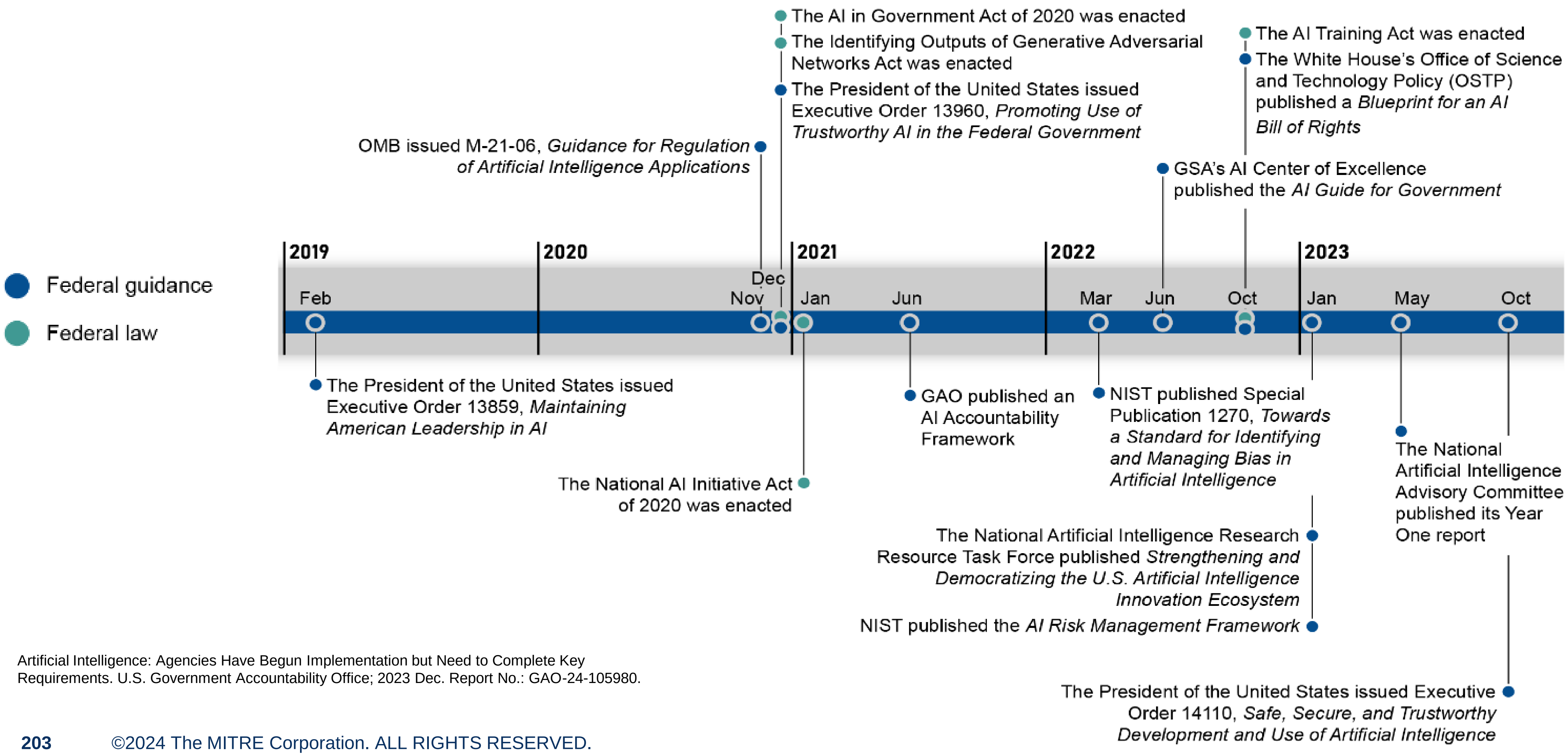
2

Explainability Tool



3

Federal government is addressing AI trustworthiness



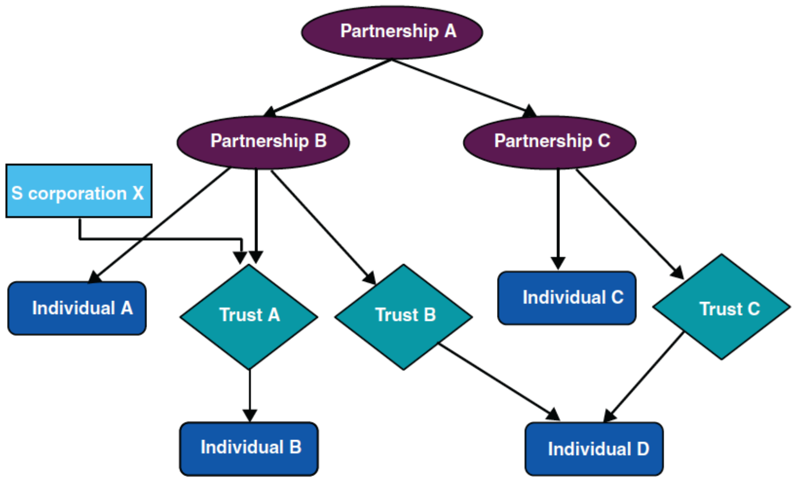
Artificial Intelligence: Agencies Have Begun Implementation but Need to Complete Key Requirements. U.S. Government Accountability Office; 2023 Dec. Report No.: GAO-24-105980.

Implementation Gap

Our prototype AI system predicts potential tax non-compliance for enterprises

input

enterprise: a network of flow-through entities



Tax Gap: IRS Can Improve Efforts to Address Tax Evasion by Networks of Businesses and Related Entities (Report to the Committee on Finance, U.S. Senate No. GAO-10-968), 2010. . U.S. Government Accountability Office.

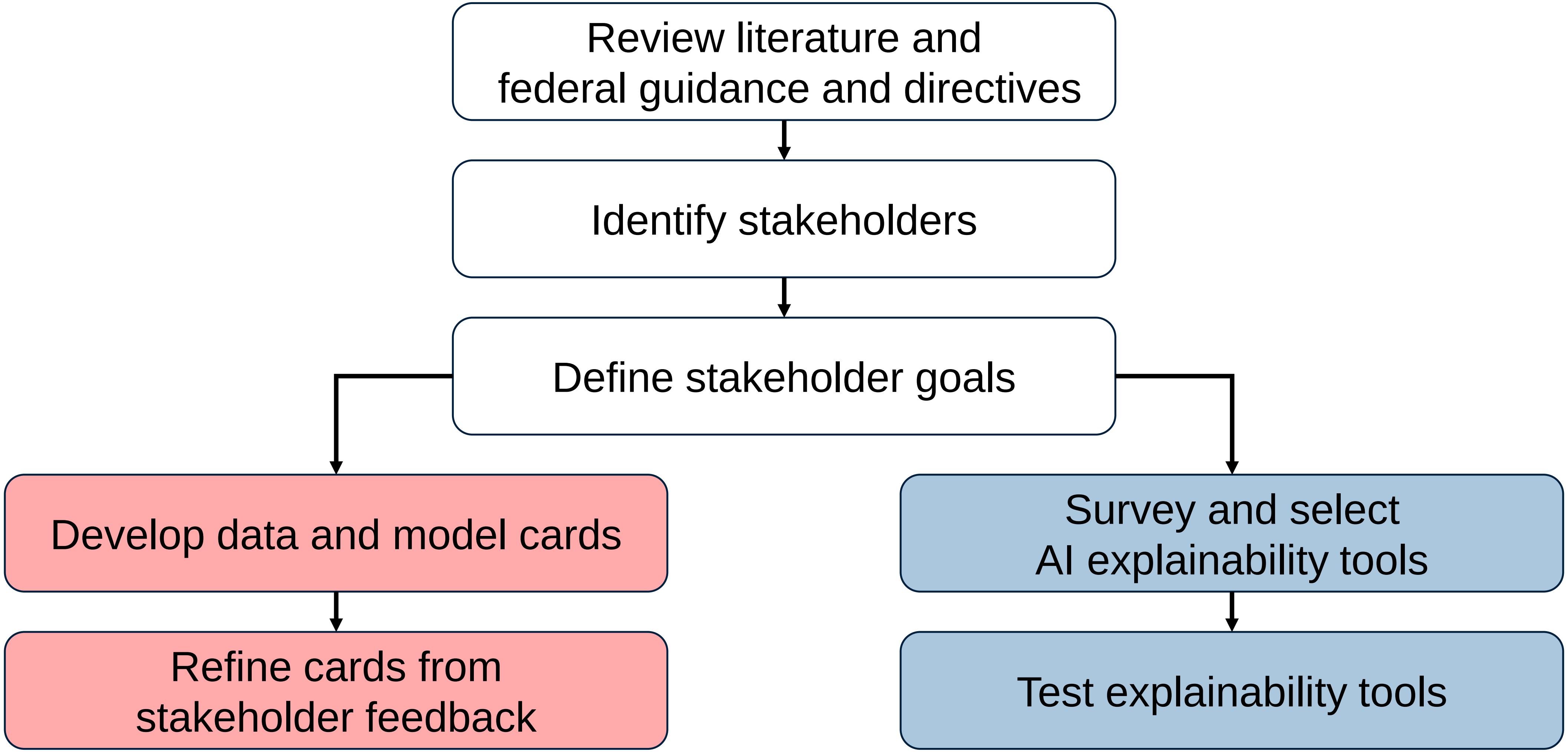
model

graph neural network

output

additional tax an exam of the controlling owner would recommend

Our Approach to Promote AI Trustworthiness

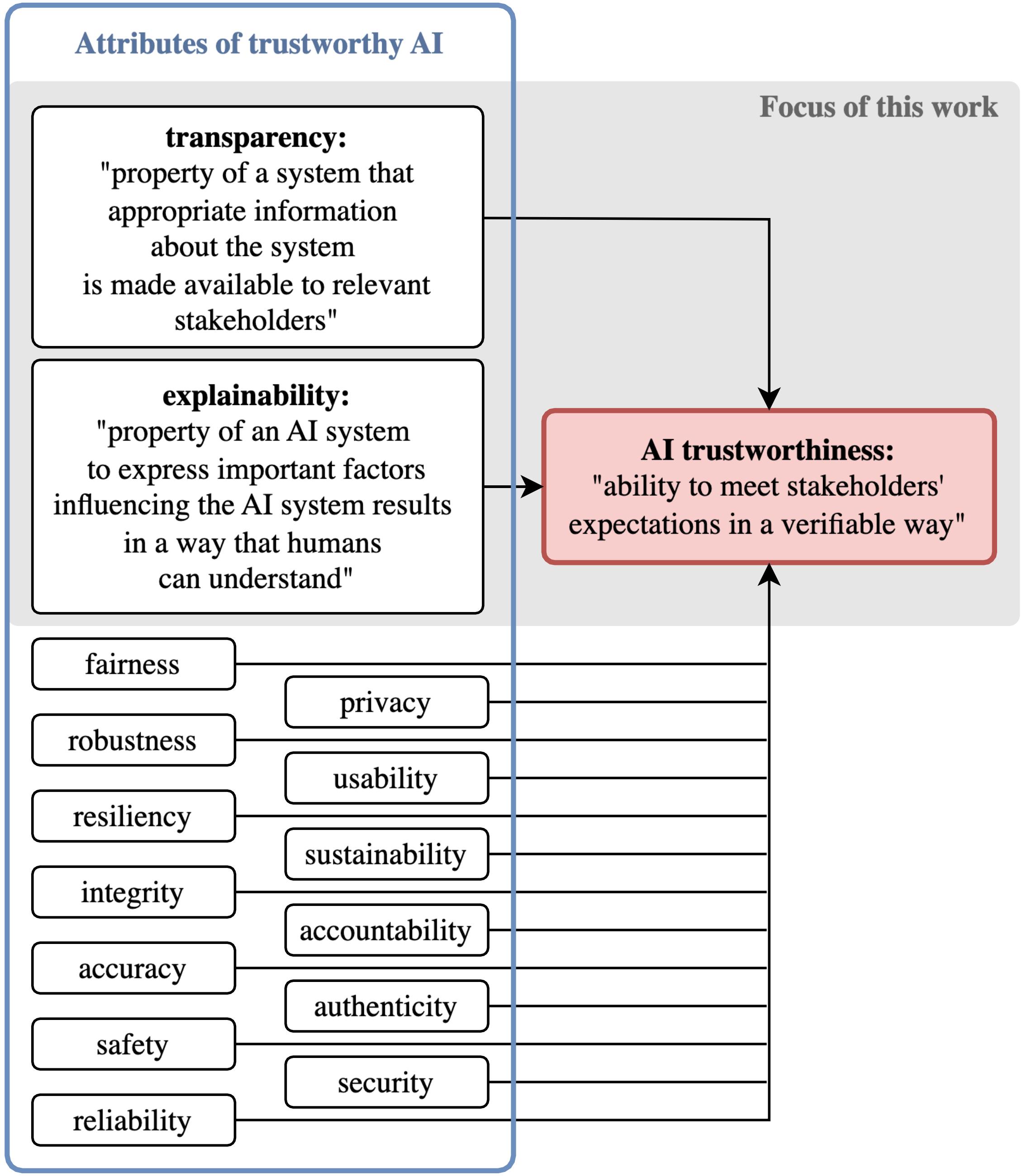


We reviewed 60 journal articles and 23 sources of federal guidance

"AI Risk Management Framework"	2023	NIST
"Towards a Standard for Identifying and Managing Bias in Artificial Intelligence"	2022	NIST
"Four Principles of Explainable Artificial Intelligence"	2021	NIST
"Trust and Artificial Intelligence"	2020	NIST
"U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools"	2019	NIST
"Artificial Intelligence: Emerging Opportunities, Challenges, and Implications"	2018	GAO
"Artificial Intelligence: Agencies Have Begun Implementation but Need to Complete Key Requirements"	2023	GAO
"Proposed Memorandum for the Heads of Executive Departments and Agencies"	2023	OMB
"Guidance for Regulation of Artificial Intelligence Applications"	2020	OMB
"Open Data Policy-Managing Information as an Asset"	2013	OMB
"Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government"	2020	EOP

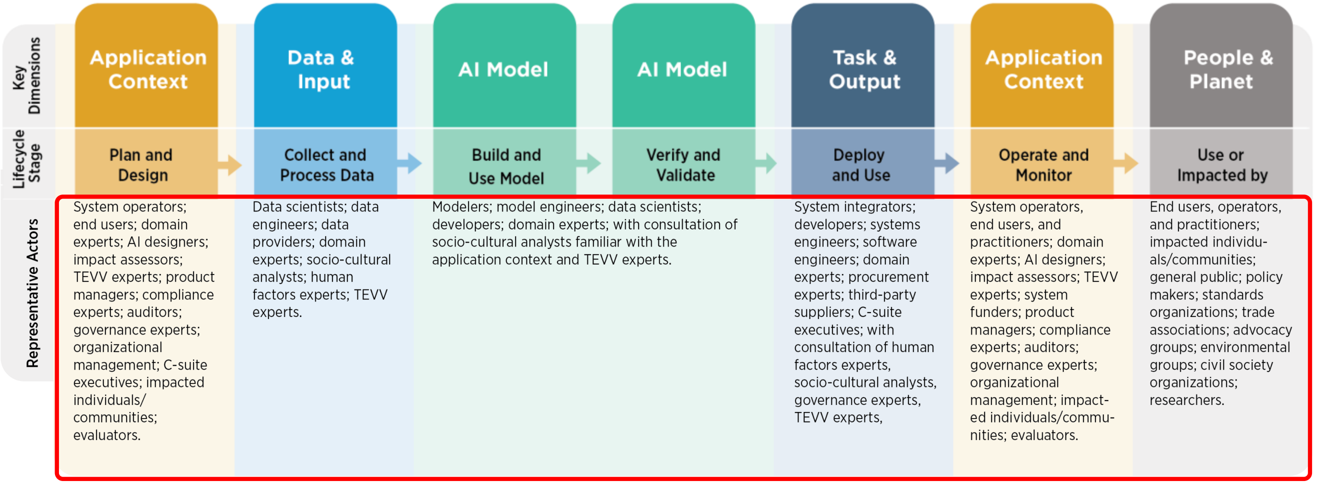
"Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"	2023	EOP
"Maintaining American Leadership in Artificial Intelligence"	2019	EOP
"Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People"	2022	OSTP
"AI Guide for Government"	2022	GSA
"Treasury Strategic Plan 2022-2026"	2023	TREAS
"Federal Data Strategy 2021 Action Plan"	2021	OMB, OSTP, DOC, SBA
"Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem"	2023	NAIRR
"National Artificial Intelligence Advisory Committee Year 1 Report"	2023	NAIAC
"Data, Analytics, and Artificial Intelligence Adoption Strategy Accelerating Decision Advantage"	2023	DOD
"Ethical Principles for Artificial Intelligence"	2020	DOD
"National Artificial Intelligence Initiative Act of 2020"	2020	Congress
"AI in Government Act of 2020"	2020	Congress

We focused on two attributes of trustworthy AI



We identified stakeholders based on NIST’s AI Risk Management Framework

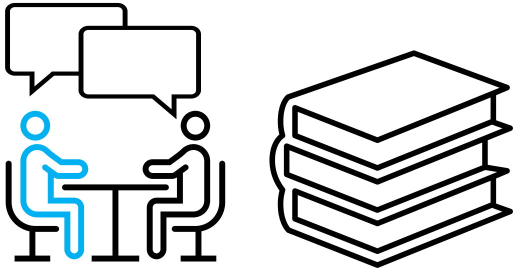
Stakeholders in AI lifecycle stages (NIST)

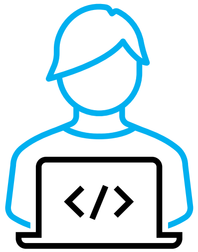


Adapted from:
Artificial Intelligence Risk Management Framework. National Institute of Standards and Technology; 2023 Jan. Report No.: NIST AI 100-1.

Our Stakeholders
organization leadership
domain experts (e.g., pass-through entity experts)
AI model developers
model-development managers
operations & monitoring engineers
operations managers
users (e.g., classifiers, auditors)
internal AI impact assessors
outside entities (e.g., TIGTA, GAO)

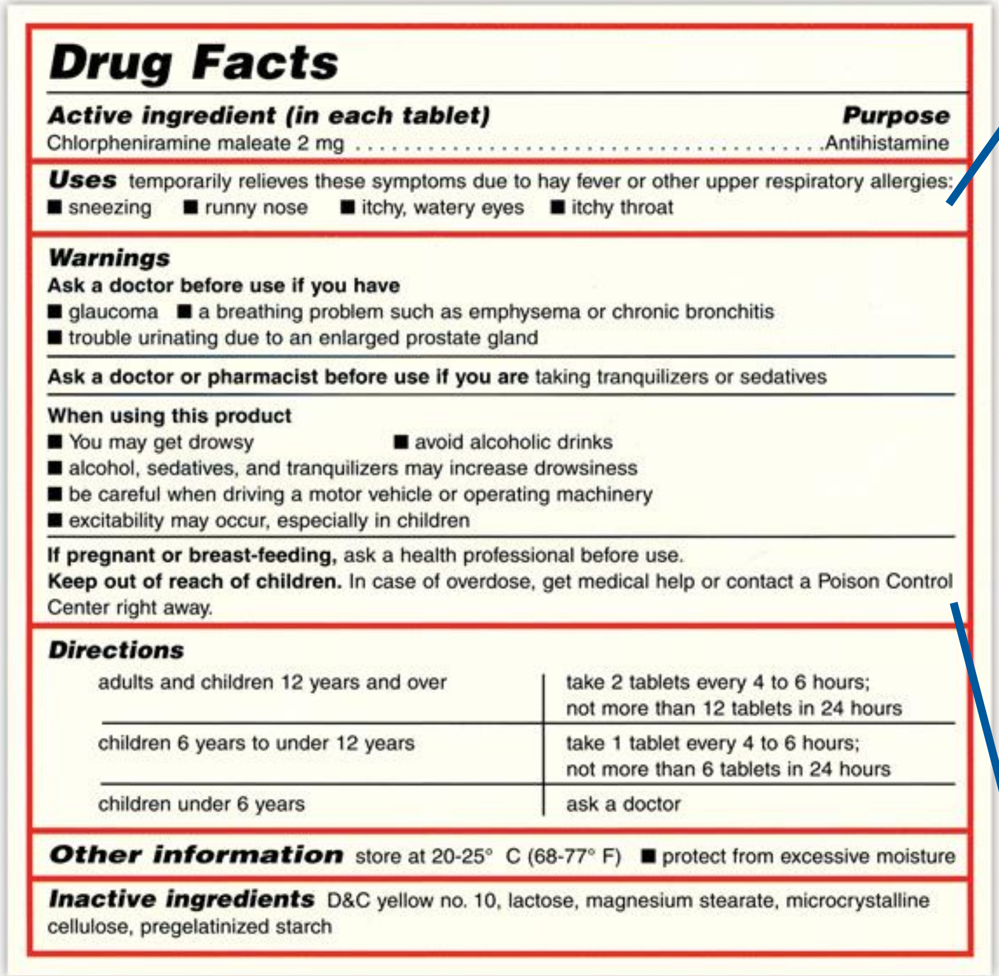
We explored, developed, and tested 1 tool for each goal.





Stakeholder goals for transparency and explainability	Tools
Document the dataset in plain-language (e.g., composition, quality issues, intended uses & users)	data card
Document the model in plain-language (e.g., inputs, outputs, performance, risks, mitigations)	model card
Provide explanations of why specific inputs create specific predictions that are meaningful to stakeholders.	explainability model

Cards are like drug fact labels



Drug Facts

Active ingredient (in each tablet) — **Purpose**
Chlorpheniramine maleate 2 mg .... Antihistamine

Uses temporarily relieves these symptoms due to hay fever or other upper respiratory allergies:
■ sneezing ■ runny nose ■ itchy, watery eyes ■ itchy throat

Warnings
Ask a doctor before use if you have
■ glaucoma ■ a breathing problem such as emphysema or chronic bronchitis
■ trouble urinating due to an enlarged prostate gland

Ask a doctor or pharmacist before use if you are taking tranquilizers or sedatives

When using this product
■ You may get drowsy ■ avoid alcoholic drinks
■ alcohol, sedatives, and tranquilizers may increase drowsiness
■ be careful when driving a motor vehicle or operating machinery
■ excitability may occur, especially in children

If pregnant or breast-feeding, ask a health professional before use.
Keep out of reach of children. In case of overdose, get medical help or contact a Poison Control Center right away.

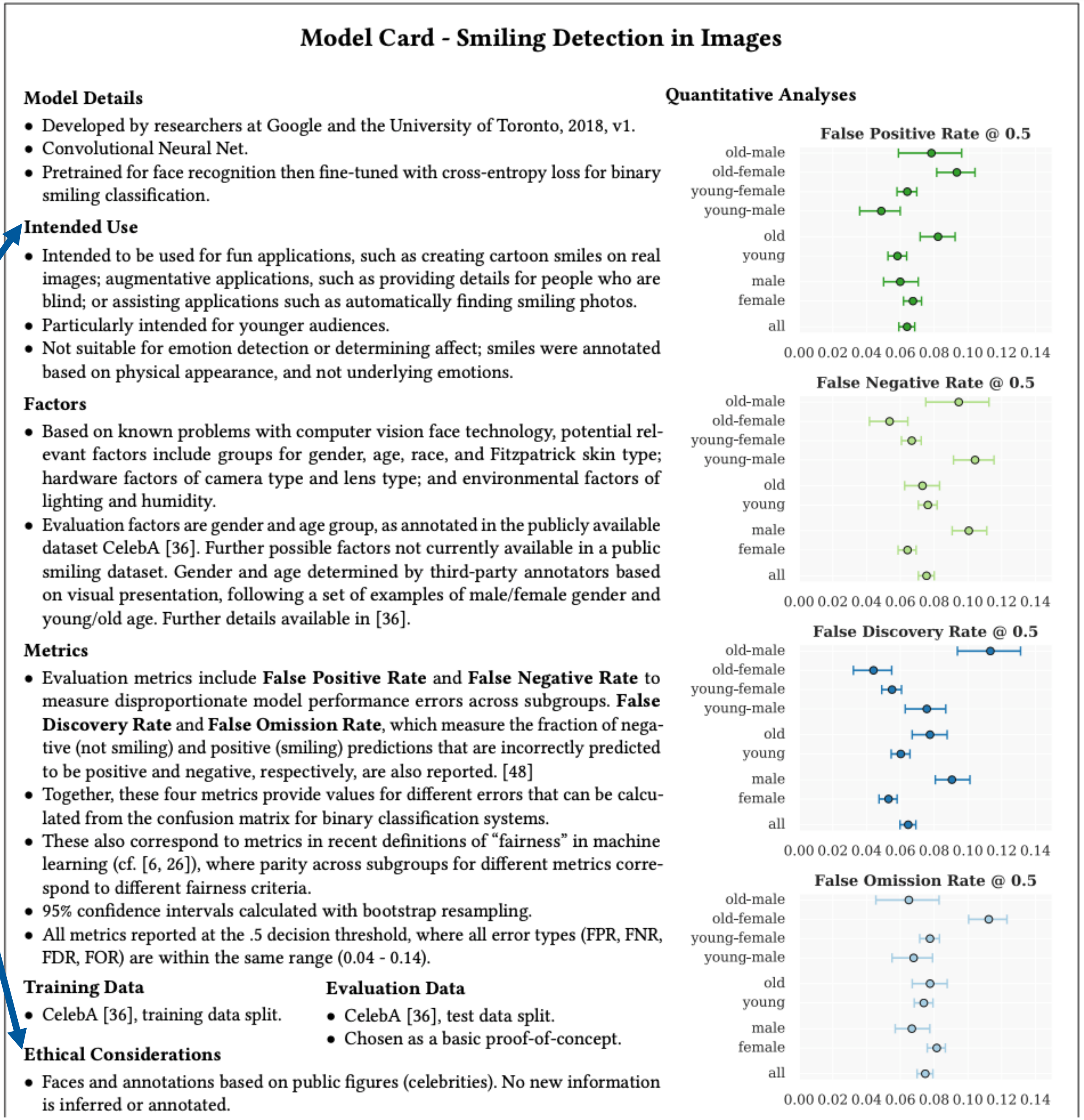
Directions

adults and children 12 years and over	take 2 tablets every 4 to 6 hours; not more than 12 tablets in 24 hours
children 6 years to under 12 years	take 1 tablet every 4 to 6 hours; not more than 6 tablets in 24 hours
children under 6 years	ask a doctor

Other information store at 20-25° C (68-77° F) ■ protect from excessive moisture

Inactive ingredients D&C yellow no. 10, lactose, magnesium stearate, microcrystalline cellulose, pregelatinized starch

https://www.fda.gov/drugs/information-consumers-and-patients-drugs/otc-drug-facts-label.



Model Card - Smiling Detection in Images

Model Details
- Developed by researchers at Google and the University of Toronto, 2018, v1.
- Convolutional Neural Net.
- Pretrained for face recognition then fine-tuned with cross-entropy loss for binary smiling classification.

Intended Use
- Intended to be used for fun applications, such as creating cartoon smiles on real images; augmentative applications, such as providing details for people who are blind; or assisting applications such as automatically finding smiling photos.
- Particularly intended for younger audiences.
- Not suitable for emotion detection or determining affect; smiles were annotated based on physical appearance, and not underlying emotions.

Factors
- Based on known problems with computer vision face technology, potential relevant factors include groups for gender, age, race, and Fitzpatrick skin type; hardware factors of camera type and lens type; and environmental factors of lighting and humidity.
- Evaluation factors are gender and age group, as annotated in the publicly available dataset CelebA [36]. Further possible factors not currently available in a public smiling dataset. Gender and age determined by third-party annotators based on visual presentation, following a set of examples of male/female gender and young/old age. Further details available in [36].

Metrics
- Evaluation metrics include **False Positive Rate** and **False Negative Rate** to measure disproportionate model performance errors across subgroups. **False Discovery Rate** and **False Omission Rate**, which measure the fraction of negative (not smiling) and positive (smiling) predictions that are incorrectly predicted to be positive and negative, respectively, are also reported. [48]
- Together, these four metrics provide values for different errors that can be calculated from the confusion matrix for binary classification systems.
- These also correspond to metrics in recent definitions of "fairness" in machine learning (cf. [6, 26]), where parity across subgroups for different metrics correspond to different fairness criteria.
- 95% confidence intervals calculated with bootstrap resampling.
- All metrics reported at the .5 decision threshold, where all error types (FPR, FNR, FDR, FOR) are within the same range (0.04 - 0.14).

Training Data
- CelebA [36], training data split.

Evaluation Data
- CelebA [36], test data split.
- Chosen as a basic proof-of-concept.

Ethical Considerations
- Faces and annotations based on public figures (celebrities). No new information is inferred or annotated.

Quantitative Analyses

Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model Cards for Model Reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency [Internet]. 2019 [cited 2024 Jan 14]. p. 220–9. Available from: http://arxiv.org/abs/1810.03993

Our Cards

Data card authors(s)	[name(s)]
Contact info.	[email address, phone number, etc.]
Data card release date	[year-month-day]
Data card version	[VX.X]

Data Card

[Write the dataset name here (and any aliases or acronyms)]

1. Dataset Identification

1.1. **Summary.** Briefly describe the dataset in plain language.

1.2. **Creator.** Who created the dataset (i.e., what research group, agency, or division)? Provide contact information if available (e.g., name, affiliation, email address, website).

1.3. **Points of Contact (POCs).** Identify POCs who can answer questions about the model or provide assistance (list names, email addresses, departments, etc.).

1.4. **Release date.** When was the dataset made available?
YYYY-MM-DD

1.5. **Size.** What is the size of the dataset in bytes?

1.6. **Version.** Provide the version number of the dataset or other identifying information.

1.7. **Format.** Describe the format of the dataset.
Example: The dataset consists of 3 CSV files and 1 JSON file.

1.8. **Sensitivity.**

Does the dataset contain sensitive data[12]?	Does this datasheet contain sensitive data?
☐ No ☐ Yes	☐ No ☐ Yes

Model card authors(s)	[name(s)]
Contact info.	[email address, phone number, etc.]
Model card release date	[year-month-day]
Model card version	[VX.X]

Model Card

[Write the model name here (and any aliases or acronyms)]

1. Model Identification

1.1. **Model type.** What is the model type (e.g., linear regression, convolutional neural network, etc.)?

1.2. **Task.** What is the model task (e.g., regression, classification, anomaly detection, etc.)?

1.3. **Creator.** Who created the model (i.e., what research group, agency, or division)? Provide contact information if available (e.g., name, affiliation, email address, website).

1.4. **Points of Contact (POCs).** Identify POCs who can answer questions about the model or provide assistance (list names, email addresses, departments, etc.).

1.5. **Creation date.** When was the model created?
YYYY-MM-DD

1.6. **Version.** Provide the version number of the model or other identifying information.

Is there a commit ID for the model in a version-control system?
☐ No
☐ Yes. Provide the ID.

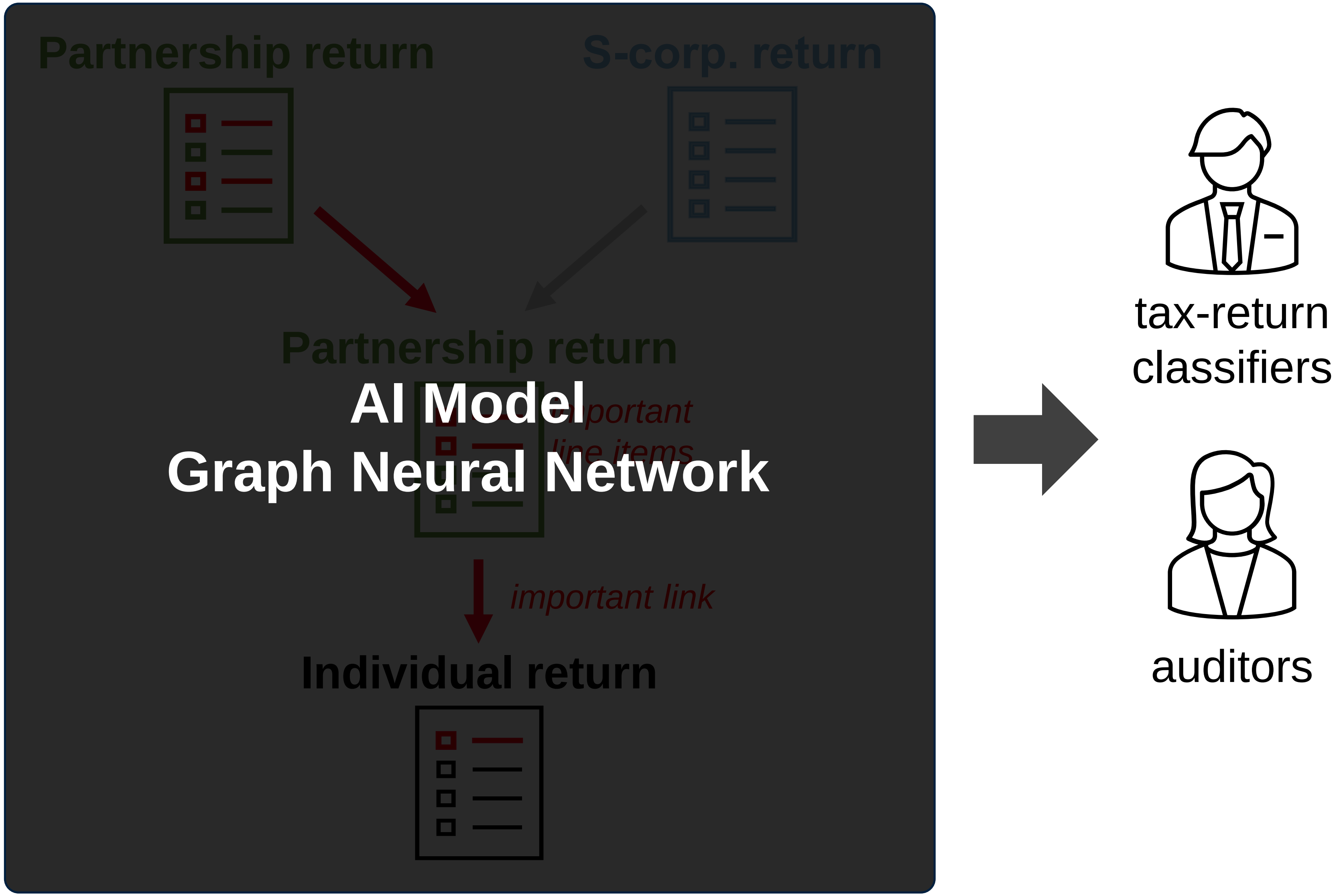
2. Motivation

2.1. **Initiator.** What entity (e.g., division, team, agency, or external party) ordered the creation of the model?

2.2. **Purpose.** Why was the model created? What were the intended uses?

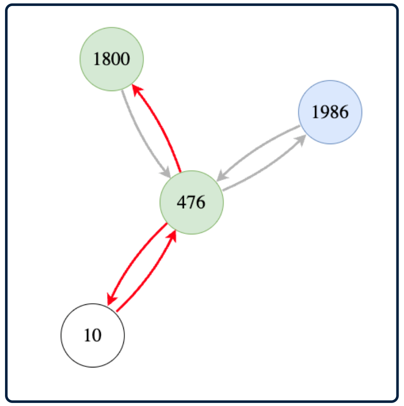
2.3. **Users.** Who were the intended users of the model (i.e., what person, research group, agency, or division)?

Tool for explaining model predictions

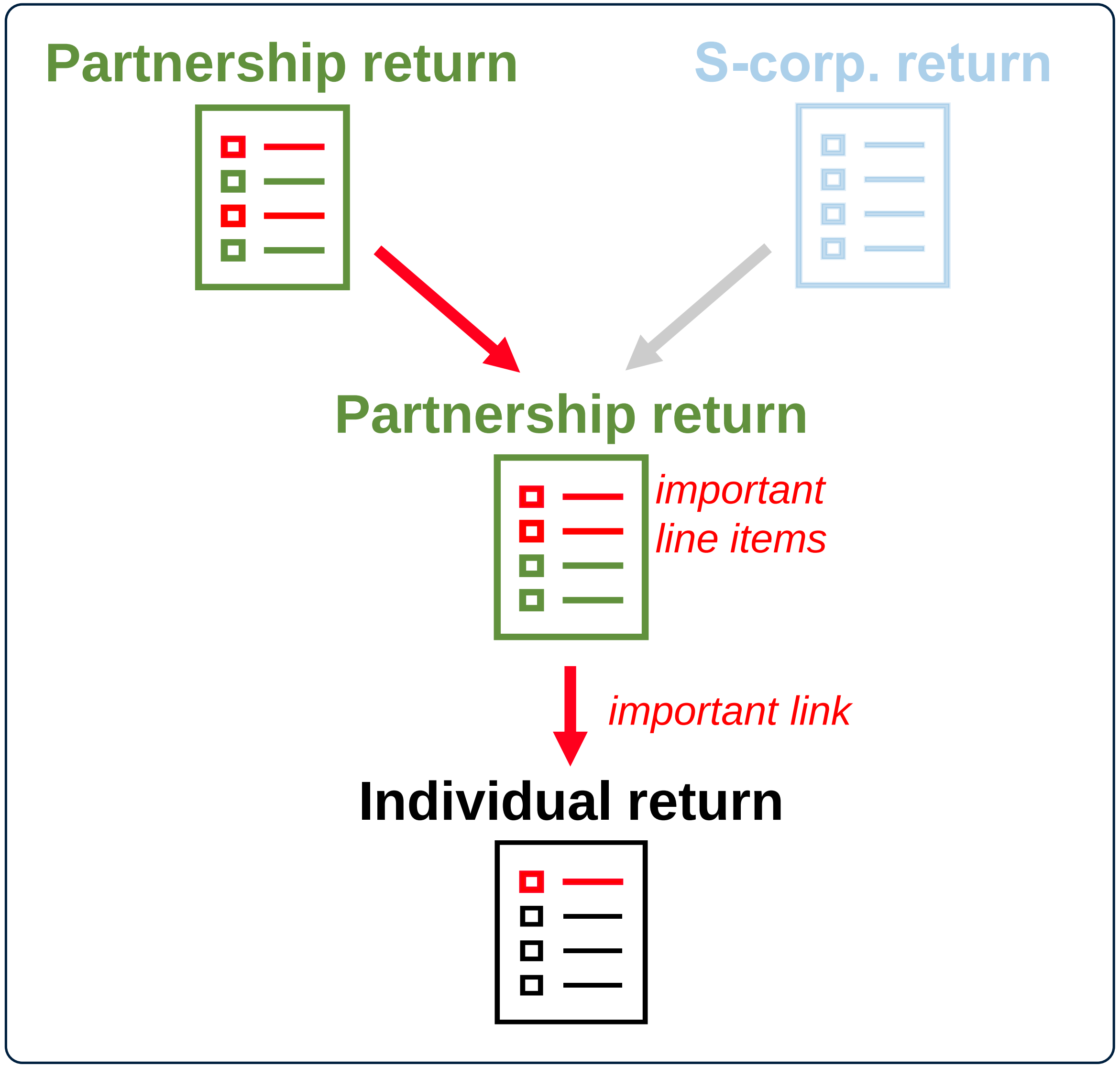


Tool for explaining model predictions

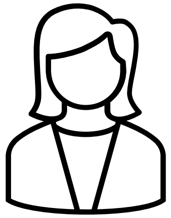
GNN Explainer



Ying R, Bourgeois D, You J, Zitnik M, Leskovec J. GNNExplainer: Generating Explanations for Graph Neural Networks. arXiv:190303894 [cs, stat] [Internet]. 2019 Nov 13 [cited 2023 Mar 20]; Available from: http://arxiv.org/abs/1903.03894

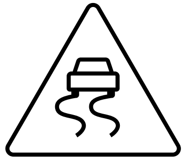


tax-return classifiers

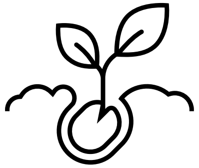


auditors

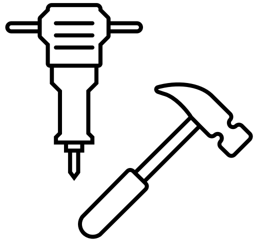
Key takeaways for practitioners



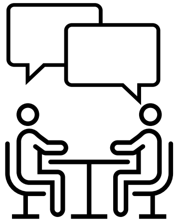
Define terms



Start filling out data and model cards at the beginning

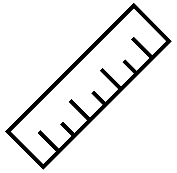


Choose explainability tools with appropriate usability and meaningful explanations for decision makers

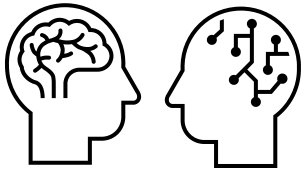


Use cards to communicate between groups of stakeholders (e.g., engineers and managers)

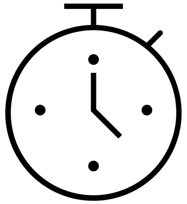
Challenges for practitioners



Setting measurable standards for AI trustworthiness



Balancing tradeoffs between self-interpretable and black-box models



Balancing efforts to improve transparency and explainability with time and resource constraints

Thank you

Backup Slides

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

M. L. Szulczewski[a,c], M. Feldman[a], S. Silva[a], A. Graff[b] B. Anderson[b]

[a] The MITRE Corporation

[b] Internal Revenue Service (IRS); Research, Applied Analytics and Statistics Division (RAAS)

[c] Corresponding author: mszulczewski@mitre.org, 781-223-5492

Abstract

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.

Stakeholder roles	Tasks
Data, software, and model engineers	Process data, write software, develop models, and test models
Operations and monitoring engineers	Operate and monitor AI systems
Domain experts	Provide deep knowledge about a field
Project managers	Ensure data, software, and model engineering meet requirements and communicate with stakeholders
Operations managers	Manage the deployment and use of an AI system
Leadership	Ensure alignment of AI projects with organizational goals
AI impact assessors	Evaluate AI assurance
External entities	Provide guidance or directives for specifying, managing, or reporting AI risks

Stakeholders

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Adapted from:
Artificial Intelligence Risk Management Framework. National Institute of Standards and Technology; 2023 Jan. Report No.: NIST AI 100-1.

OSHA® QUICK CARD™

Hazard Communication Safety Data Sheets

Section 1, Identification includes product identifier; manufacturer or distributor name, address, phone number; emergency phone number; recommended use; restrictions on use.

Section 2, Hazard(s) identification includes all hazards regarding the chemical; required label elements.

Section 3, Composition/information on ingredients includes information on chemical ingredients; trade secret claims.

Section 4, First-aid measures includes important symptoms/effects, acute, delayed; required treatment.

Section 5, Fire-fighting measures lists suitable extinguishing techniques, equipment; chemical hazards from fire.

Section 6, Accidental release measures lists emergency procedures; protective equipment; proper methods of containment and cleanup.

Section 7, Handling and storage lists precautions for safe handling and storage, including incompatibilities.

https://www.osha.gov/sites/default/files/publications/OSHA3493QuickCard SafetyDataSheet.pdf

data card

Open Images Extended - More Inclusively Annotated People (MIAP)

Dataset Download • Related Publication

This dataset was created for fairness research and fairness evaluations in person detection. This dataset contains 100,000 images sampled from Open Images V6 with additional annotations added. Annotations include the image coordinates of bounding boxes for each visible person. Each box is annotated with attributes for perceived gender presentation and age range presentation. It can be used in conjunction with Open Images V6.

Authorship

PUBLISHER(S)
Google LLC

INDUSTRY TYPE
Corporate - Tech

DATASET AUTHORS
Candice Schumann, Google, 2021
Susanna Ricco, Google, 2021
Utsav Prabhu, Google, 2021
Vittorio Ferrari, Google, 2021
Caroline Pantofaru, Google, 2021

FUNDING
Google LLC

FUNDING TYPE
Private Funding

DATASET CONTACT
open-images-extended@google.com

Motivations

DATASET PURPOSE(S)
Research Purposes
Machine Learning
Training, testing, and validation

KEY APPLICATION(S)
Machine Learning | Object Recognition | Machine Learning Fairness

PROBLEM SPACE
This dataset was created for fairness research and fairness evaluation with respect to person detection.
See accompanying article

PRIMARY MOTIVATION(S)
- Provide more complete ground-truth for bounding boxes around people.
- Provide a standard fairness evaluation set for the broader fairness community.

INTENDED AND/OR SUITABLE USE CASE(S)
- **ML Model Evaluation for:** Person detection, Fairness evaluation
- **ML Model Training for:** Person detection, Object detection

Additionally:
- **Person detection:** Without specifying gender or age presentations
- **Fairness evaluations:** Over gender and age presentations
- **Fairness research:** Without building gender presentation or age classifiers

Use of Dataset

SAFETY OF USE
Conditional Use
There are some known unsafe applications.

UNSAFE APPLICATION(S)
⚠ Gender classification | Age classification

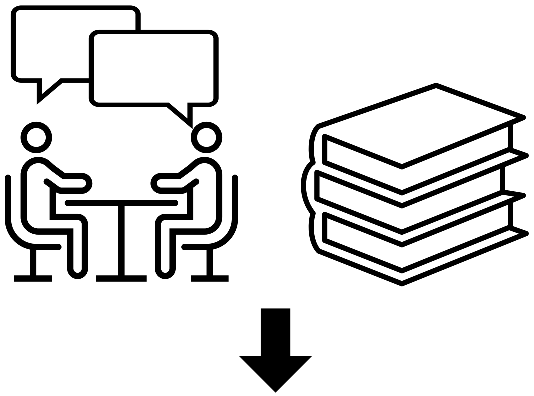
UNSAFE USE CASE(S)
This dataset **should not** be used to create gender or age classifiers. The intention of percieved gender and age labels is to capture gender and age presentation as assessed by a third party based on visual cues alone, rather than an individual's self-identified gender or actual age.

CONJUNCTIONAL USE
Safe to use with other datasets

KNOWN CONJUNCTIONAL DATASET(S)
- The data in this dataset can be combined with Open Images V6

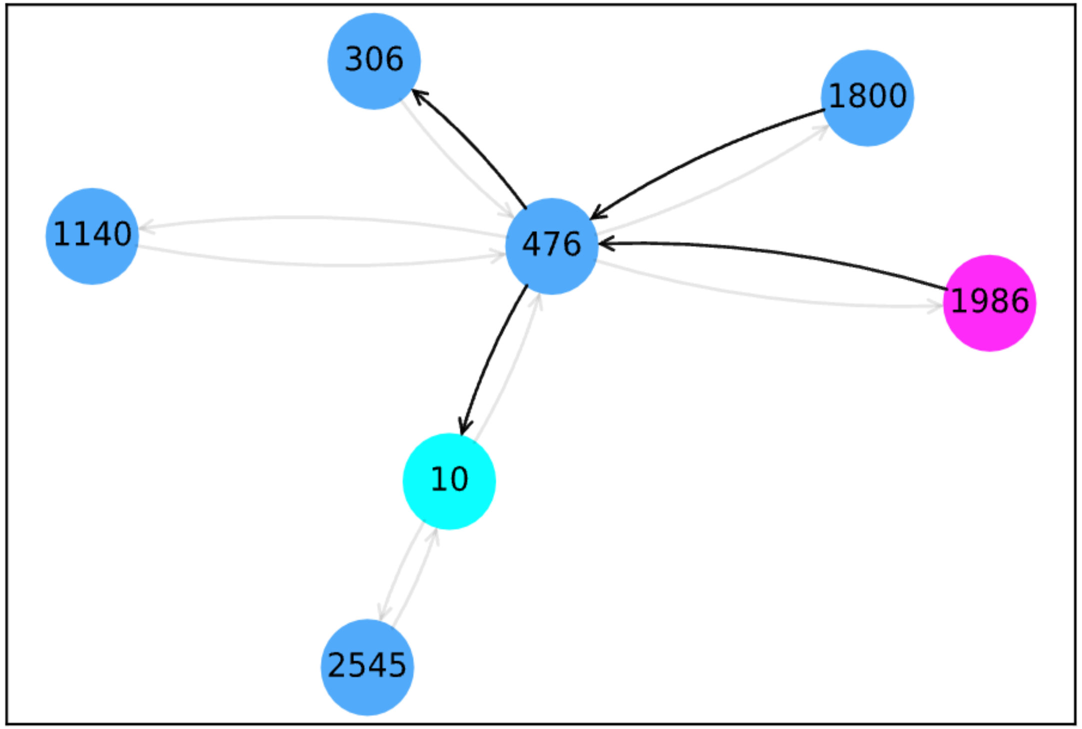
KNOWN CONJUNCTIONAL USES
Analyzing bounding box annotations not annotated under the Open Images V6 procedure.

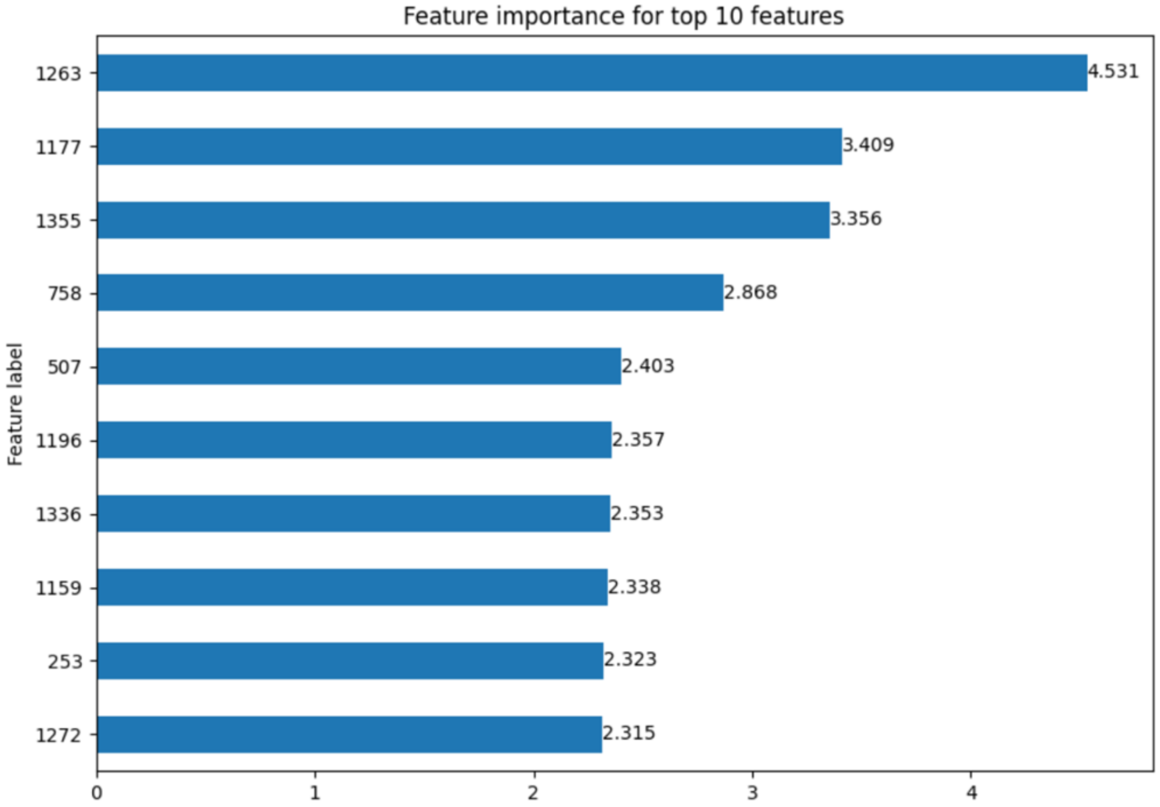
Pushkarna M, Zaldivar A, Kjartansson O. Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency [Internet]. New York, NY, USA: Association for Computing Machinery; 2022 p. 1776–826. (FAccT '22).



Stakeholder goals for transparency and explainability
Use plain language.
Provide brief summaries of the dataset and model.
For the dataset, describe the collection, composition, quality issues, processing, maintenance, and intended uses and users.
For the model, describe the inputs, outputs, risks, mitigations, performance, limitations, and intended uses and users.
Explain why specific inputs create specific predictions.
Ensure explanations are meaningful to stakeholders.

GNN Explainer





Ying R, Bourgeois D, You J, Zitnik M, Leskovec J. GNNExplainer: Generating Explanations for Graph Neural Networks. arXiv:190303894 [cs, stat] [Internet]. 2019 Nov 13 [cited 2023 Mar 20]; Available from: http://arxiv.org/abs/1903.03894

IRS Research, Applied Analytics & Statistics

TPC TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

The Norwegian
Tax Administration

SESSION 3:
TRUSTING THE TAX MAN: METRICS, AI, AND AUDITS

DISCUSSION

IRS-TPC Research Conference, June 13, 2024

Arnstein Øvrum | The Norwegian Tax Administration

RELEVANCE AND COMMONALITIES

Holtzblatt

Performance Metrics for Taxpayer Services, **Enforcement**, and **Equitable Treatment**

Hertz et al.

Research on **Audit Rates** by **Race** & **Ethnisity:** 2024 Update:

Szulczewski et al.

Tools to Promote **Trustworthiness** in AI Systems → **Transparancy** and **Explainability**

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

- Performance metrics for a ten-year $80 billion boost = Not easy!
- Metrics for overall IRS performance, or for the impact of the IRA budget boost?
- «Current IRS metrics are a patchwork of measures»
- Need more wholistic metrics consistent with outcomes in IRS mission statement (services, enforcement, equity)
- At the same time, IRA Strategic Operating Plan has many initiatives – how to simplify metrics?
- Outcomes due to IRS vs. external factors – more randomized controlled trials (RCTs) for impact assessment?
- Compliance burden 2023 (1040): 13 hours + $270. Should decline gradually over next 10 years?
- Taxpayer Satisfaction: ACSI and CTAS. Alternatives to boost survey samples? (e.g., survey when submitting tax return)

Hertz et al.: Research on Audit Rates by Race & Ethnisity: 2024 Update

- Impressed by your thorough approach on an important topic
- Useful insights on sources of (unintended) racial disparities in audit rates (audit objectives, missing data, etc.)
- Intriguing that new audit seletion models seem to both reduce racial disparities in audit rates and improve audit results
- Models that maximize revenue per case – are they also the most cost-efficeint (revenue per exam hour)?
- IRS Strategic Operating Plan: How will expanded enforcement for high-income indivuals affect audit rates by race?
- Disparities in audit rates vs. audit revenue: What is total audit revenue / total taxes, by race?
- Look forward to following your ongoing work – and motivated to address this topic also in Norwegian Tax Administration

Szulczewski et al.: Tools to Promote Trustworthiness in a Prototype AI System at the IRS

- **Documentation of AI systems is an important, but (thus far) often neglected topic**
- **Paper demonstrates benefits beyond documentation itself**
  - Improved communication between stakeholders
  - Explainability from model predictions (GNNexplainer) that is useful for the stakeholders
- **Data Cards and Model Cards: How to avoid overlapping documentation? (e.g., Microsoft Word and Data Science tools)**
- **Many existing tools and guidelines for AI Systems: Highlight what is novel with your approach?**
- **Trustworthiness at full scale: Could/should Data and Model Cards be made publicly available at e.g. IRS website?**
- **What about other AI Systems at the IRS – will they be required to do the same, guided by your example?**

IRS Research, Applied Analytics & Statistics

TPC TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Technical Challenges in Maintaining Tax Prep Software with Large Language Models

Sina Gogani Khiabani
Computer Science Department
University of Texas at El Paso

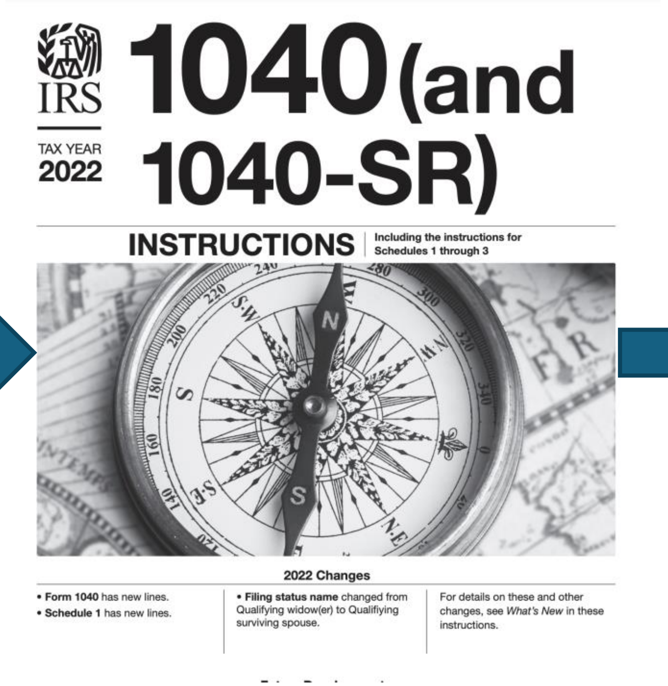
Varsha Dewangan
Computer Science Department
University of Colorado Boulder

Nina Olson
Executive Director
Center for Taxpayer Rights

Ashutosh Trivedi
Computer Science Department
University of Colorado Boulder

Saeid Tizpaz-Niari
Computer Science Department
University of Texas at El Paso

U.S. Tax 101: Manual Tax Filling



IRS

TAX YEAR 2022

1040 (and 1040-SR)

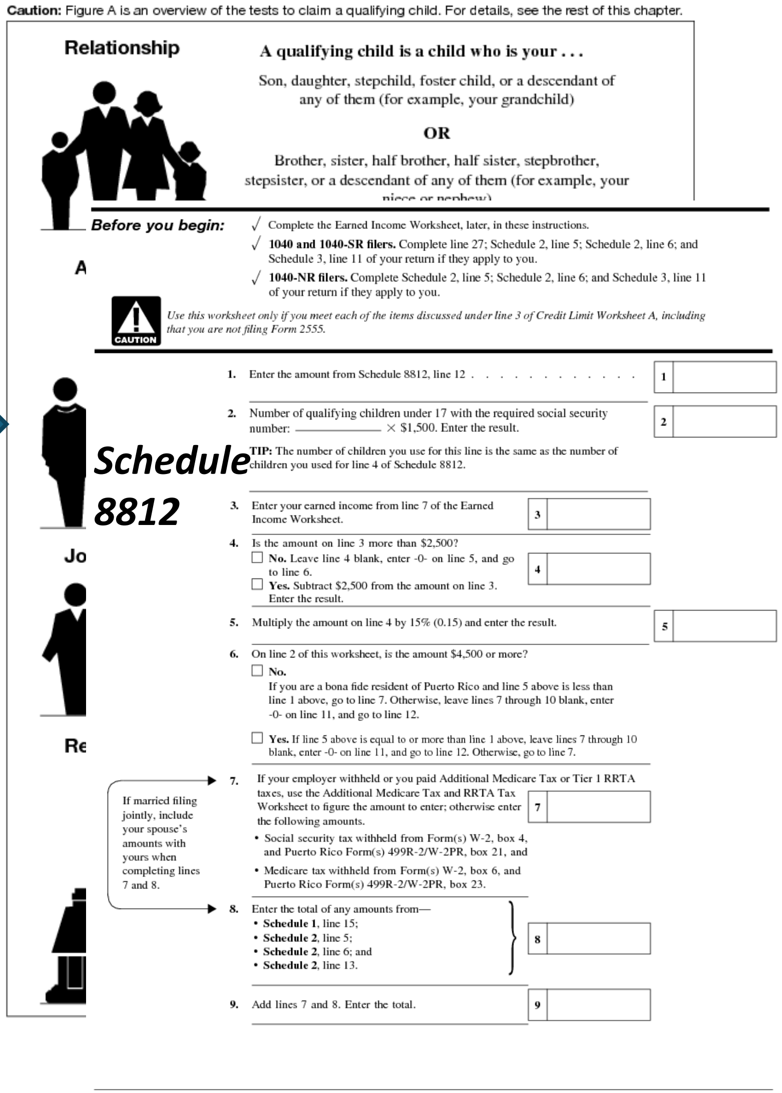
INSTRUCTIONS Including the instructions for Schedules 1 through 3

2022 Changes

- **Form 1040** has new lines.
- **Schedule 1** has new lines.
- **Filing status name** changed from Qualifying widow(er) to Qualifying surviving spouse.

For details on these and other changes, see *What's New* in these instructions.

Publication 596 (EITC)



Caution: Figure A is an overview of the tests to claim a qualifying child. For details, see the rest of this chapter.

Relationship

A qualifying child is a child who is your . . .

Son, daughter, stepchild, foster child, or a descendant of any of them (for example, your grandchild)

OR

Brother, sister, half brother, half sister, stepbrother, stepsister, or a descendant of any of them (for example, your niece or nephew)

A

Jo

Re

Schedule 8812

Before you begin:

√ Complete the Earned Income Worksheet, later, in these instructions.

√ **1040 and 1040-SR filers.** Complete line 27; Schedule 2, line 5; Schedule 2, line 6; and Schedule 3, line 11 of your return if they apply to you.

√ **1040-NR filers.** Complete Schedule 2, line 5; Schedule 2, line 6; and Schedule 3, line 11 of your return if they apply to you.

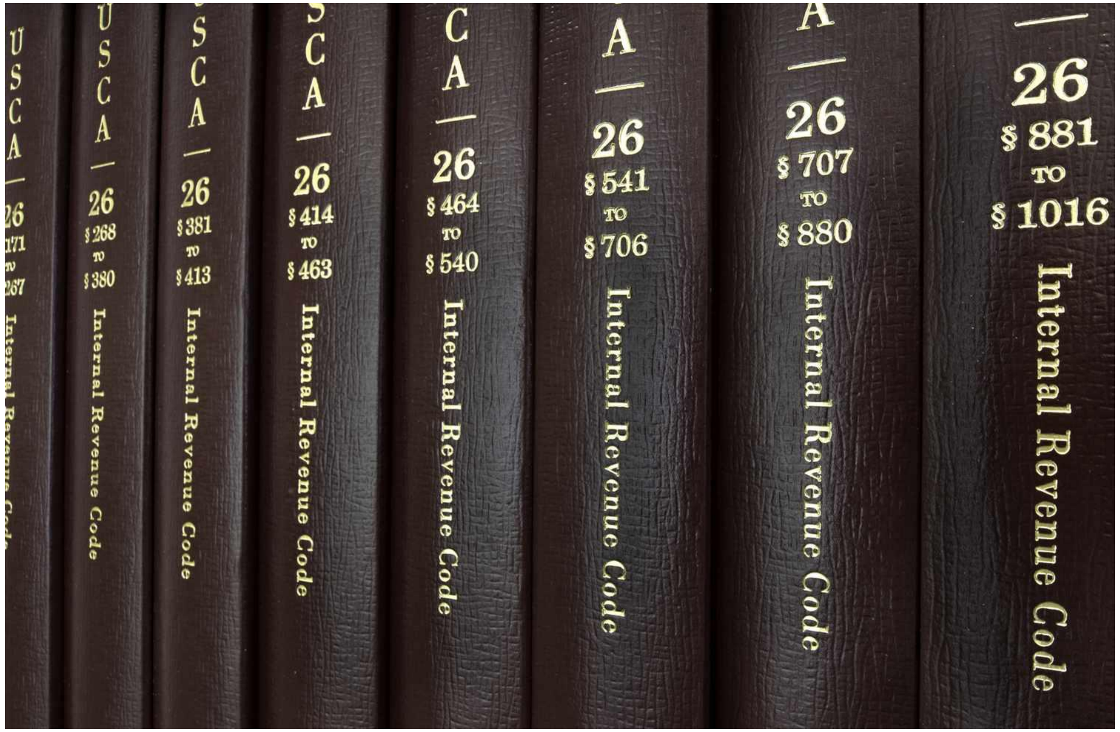
CAUTION *Use this worksheet only if you meet each of the items discussed under line 3 of Credit Limit Worksheet A, including that you are not filing Form 2555.*

1. Enter the amount from Schedule 8812, line 12 1
2. Number of qualifying children under 17 with the required social security number: ________ × $1,500. Enter the result. 2
   TIP: The number of children you use for this line is the same as the number of children you used for line 4 of Schedule 8812.
3. Enter your earned income from line 7 of the Earned Income Worksheet. 3
4. Is the amount on line 3 more than $2,500?
   ☐ **No.** Leave line 4 blank, enter -0- on line 5, and go to line 6.
   ☐ **Yes.** Subtract $2,500 from the amount on line 3. Enter the result. 4
5. Multiply the amount on line 4 by 15% (0.15) and enter the result. 5
6. On line 2 of this worksheet, is the amount $4,500 or more?
   ☐ **No.** If you are a bona fide resident of Puerto Rico and line 5 above is less than line 1 above, go to line 7. Otherwise, leave lines 7 through 10 blank, enter -0- on line 11, and go to line 12.
   ☐ **Yes.** If line 5 above is equal to or more than line 1 above, leave lines 7 through 10 blank, enter -0- on line 11, and go to line 12. Otherwise, go to line 7.

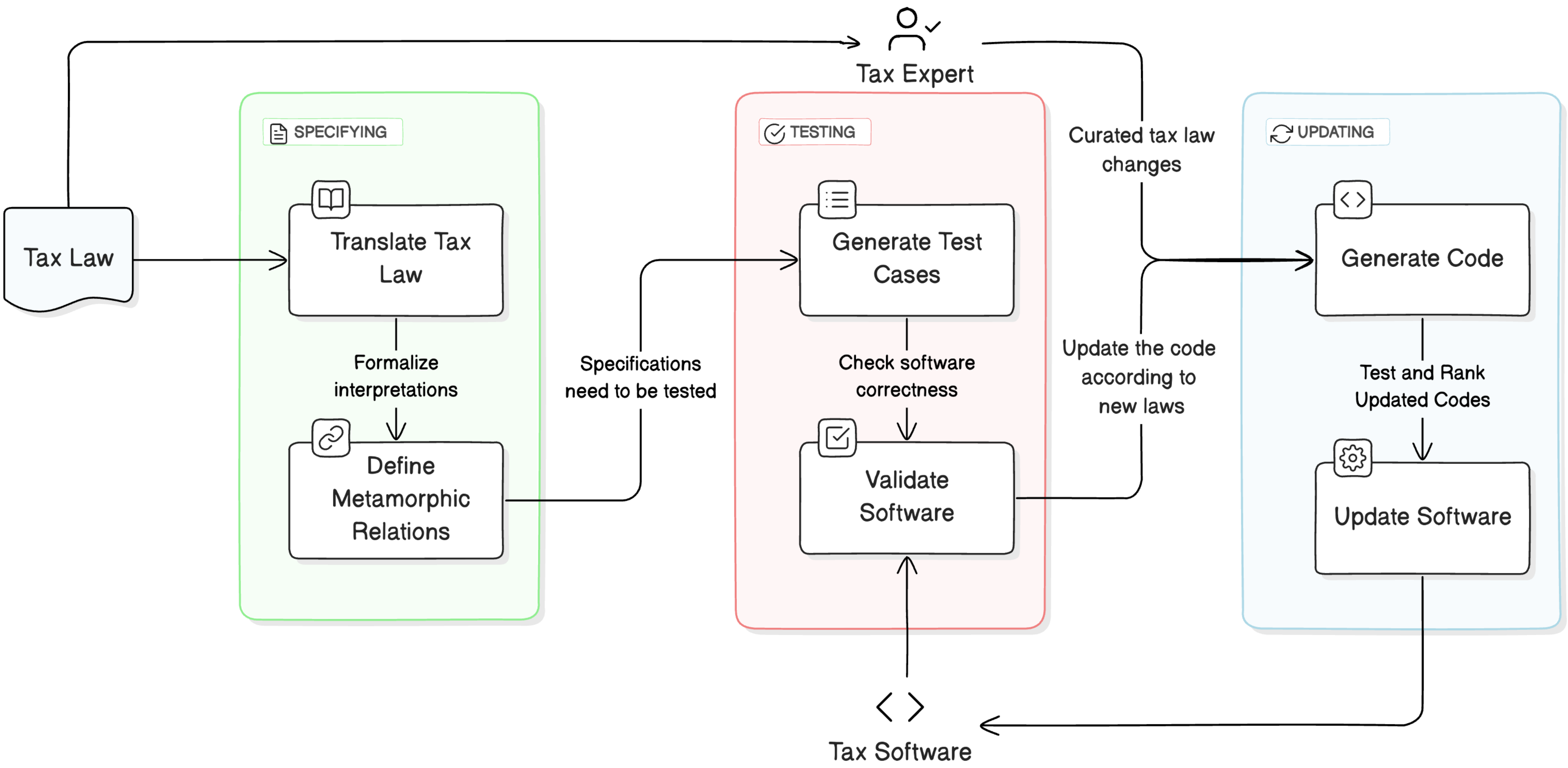
If married filing jointly, include your spouse's amounts with yours when completing lines 7 and 8.

7. If your employer withheld or you paid Additional Medicare Tax or Tier 1 RRTA taxes, use the Additional Medicare Tax and RRTA Tax Worksheet to figure the amount to enter; otherwise enter the following amounts. 7
   - Social security tax withheld from Form(s) W-2, box 4, and Puerto Rico Form(s) 499R-2/W-2PR, box 21, and
   - Medicare tax withheld from Form(s) W-2, box 6, and Puerto Rico Form(s) 499R-2/W-2PR, box 23.
8. Enter the total of any amounts from— 8
   - **Schedule 1**, line 15;
   - **Schedule 2**, line 5;
   - **Schedule 2**, line 6; and
   - **Schedule 2**, line 13.
9. Add lines 7 and 8. Enter the total. 9

The Growing Need for Trustworthy Tax Software



Meeting the Challenge: A Three-Pronged Approach



Tax Expert
Tax Law
SPECIFYING
Translate Tax Law
Formalize interpretations
Define Metamorphic Relations
Specifications need to be tested
TESTING
Generate Test Cases
Check software correctness
Validate Software
Curated tax law changes
Update the code according to new laws
UPDATING
Generate Code
Test and Rank Updated Codes
Update Software
Tax Software

Translating Tax Rules into Formal Specifications

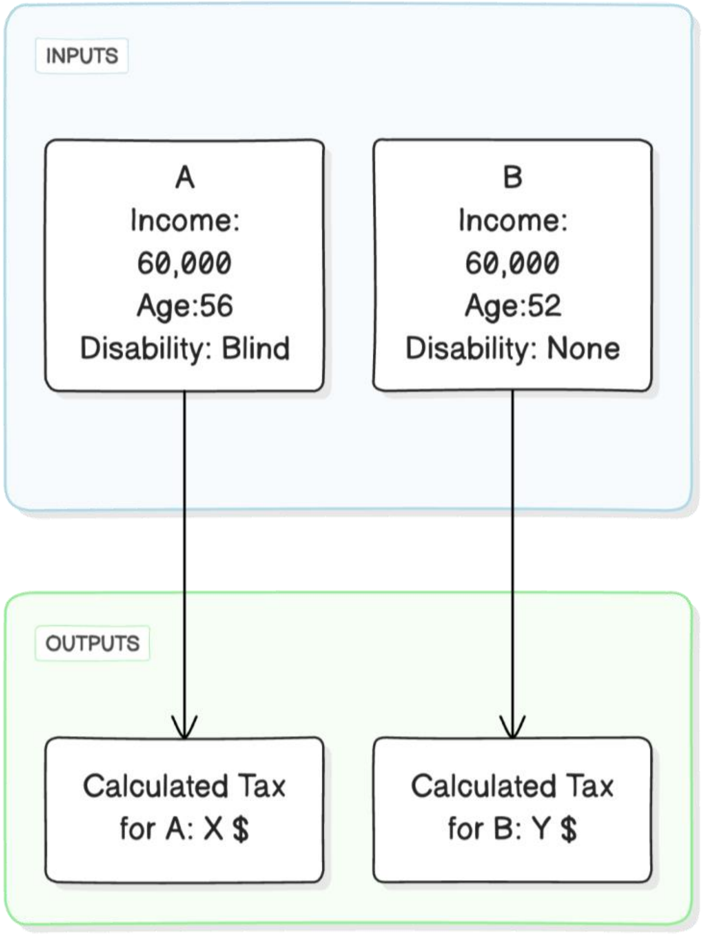
"An individual with a disability (e.g., blindness) should receive higher standard deduction."

$$\boxed{\forall\, x_1, x_2.\ x_2 \equiv_{blind}\ x_1 \wedge x1.blind\ \wedge \neg x2.blind\ \Rightarrow Return(x_1) \geq Return(x_2)}$$

Among two individuals who are exactly the same, but one is blind and another is not, the blind taxpayer should receive higher tax benefits.

Srinivas, Dananjay & Das, Rohan & Tizpaz, Saeid & Trivedi, Ashutosh & Pacheco, Maria. (2023). **On the Potential and Limitations of Few-Shot In-Context Learning to Generate Metamorphic Specifications for Tax Preparation Software.** 230-243. 10.18653/v1/2023.nllp-1.23.

Metamorphic Testing

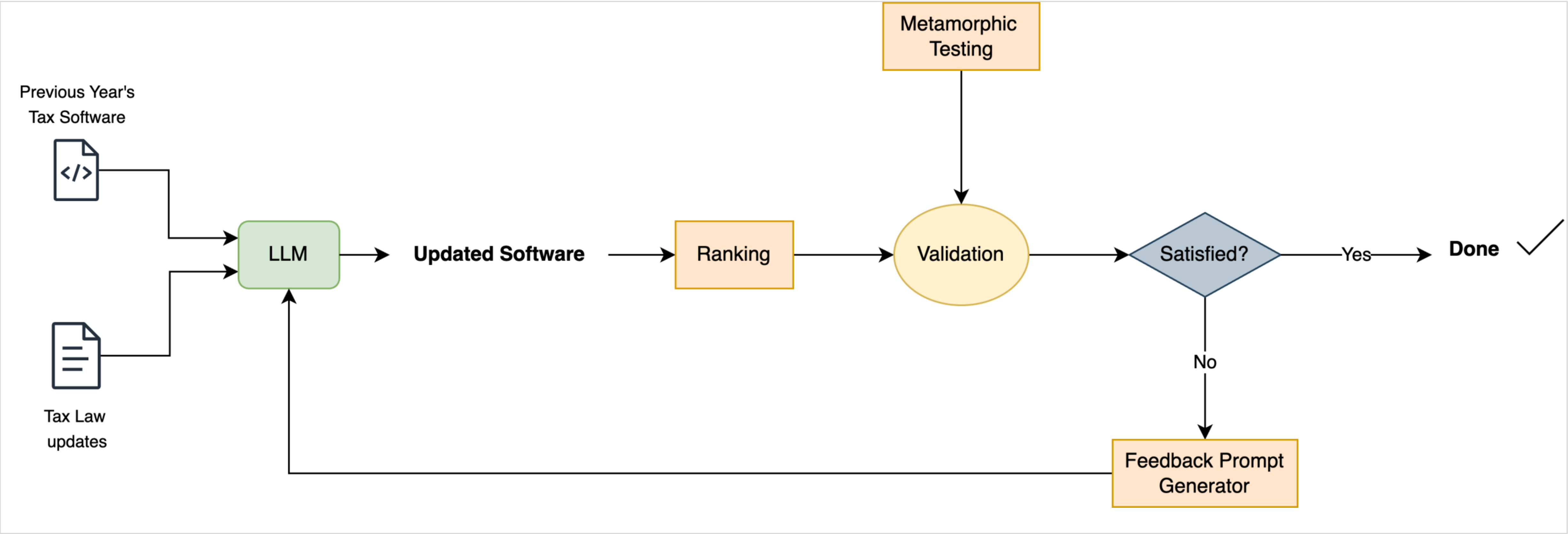


Tizpaz-Niari, S., Wagner, M., Darian, S., Reed, K., & Trivedi, A. (2022). **Metamorphic Testing and Debugging of Tax Preparation Software.** 2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), 138-149.

The Challenge of Keeping Tax Software Up-to-Date

- Manual coding and interpretation of IRS publications.
- Complex tax law changes.
- Error prone
- Alignment with IRS regulations.

Can Large Language Models(LLM) automate this update process?


Previous Year's Tax Software
Tax Law updates
LLM
Updated Software
Ranking
Metamorphic Testing
Validation
Satisfied?
Yes
Done
No
Feedback Prompt Generator

Teaching AI the Language of Tax Code

```python
def calculate_tax(income, status, brackets):
    tax_due = 0.0
    previous_limit = 0
    for limit, rate in brackets[status]:
        if income > previous_limit:
            taxable_amount = min(income, limit) - previous_limit
            tax_due += taxable_amount * rate
            previous_limit = limit
        else:
            break
    return tax_due

# Function to get deduction based on filing status, age, and blindness
def get_deduction(status, is_blind, is_older):
    return DEDUCTIONS[status][(is_blind, is_older)]
```

Prior code

Deductions Logic:
Deductions vary based on marital status, age, and blindness:
Single:
Younger than 65: $12,550
65 or Older: $14,250
Blind: additional $1,700
Head of Household:
Younger than 65: $18,800
65 or Older: $20,500
Blind: additional $1,700
Married Filing Separately:
Younger than 65: $12,550
65 or Older: $14,250
Blind: additional $1,700
Married Filing Jointly:
Younger than 65: $25,100
65 or Older: $26,450
Blind: additional $1,350

Context

```python
def calculate_tax_alternative(income, status, brackets):
    tax_due = 0.0
    previous_limit = 0
    for limit, rate in brackets[status]:
        if income > previous_limit:
            # Calculate tax only for the amount within the current bracket
            taxable_amount = min(income, limit) - previous_limit
            tax_due += taxable_amount * rate
            previous_limit = limit
        else:
            break
    return tax_due

def get_deduction_alternative(status, is_blind, is_older):
    key = (is_blind, is_older)
    return DEDUCTIONS[status][key]
```

Correct Generation

```python
def calculate_tax(income, filing_status, age, is_blind):
    deduction = deductions[filing_status][age]
    if is_blind:
        deduction += deductions[filing_status]['Blind']

    taxable_income = income - deduction
    status_index = {'Single': 0, 'Married Filing Jointly': 1, 'Married Filing Separately': 2,
    'Head of Household': 3}[filing_status]

    tax = 0.0
    for i in range(len(brkpt[status_index])):
        if taxable_income <= brkpt[status_index][i]:
            tax += taxable_income * txrt[status_index][i]
            break
        else:
            tax += (brkpt[status_index][i+1] - brkpt[status_index][i]) * txrt[status_index][i]

    return tax
```

Faulty Generation

Finding the Needle in the Code Haystack

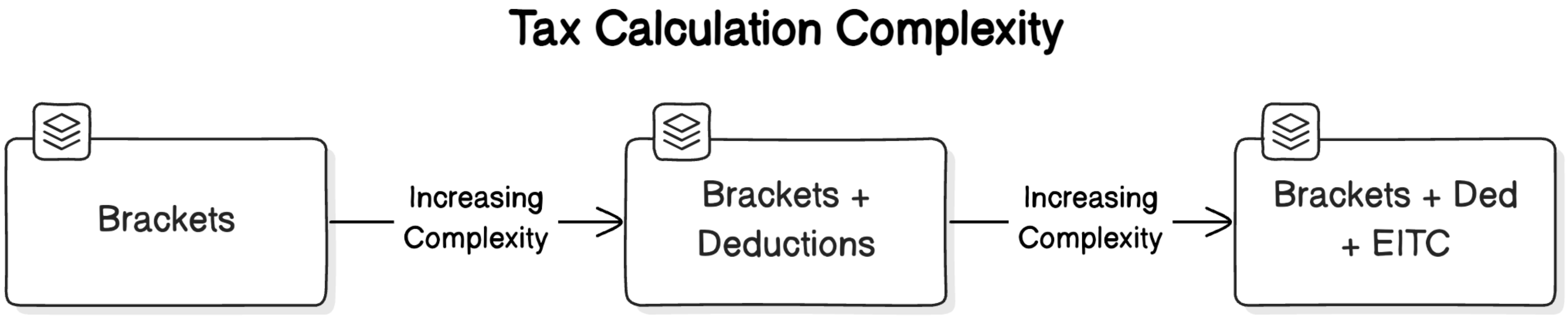
Our Metrics

- CodeBertScore
- MajorityVoteScore
- WeightedScore

Scenario	LLM	Versions	CodeBertScore	MajorityVoteScore	WeightedScore	Ground Truth Score
Brackets	GPT 4	Version 3	0.914	1	0.944	100/100
		Version 5	0.911	1	0.942	100/100
		Version 9	0.911	1	0.592	100/100
		Version 4	0.91	1	0.941	100/100
	GPT 3.5	Version 1	0.941	1	0.962	100/100
		Version 2	0.939	1	0.96	100/100
		Version 7	0.937	1	0.959	100/100
		Version 8	0.936	0.59	0.815	59/100

Can LLMs Handle Tax Code?

- We have three scenarios. Each is tested with and without prior code:



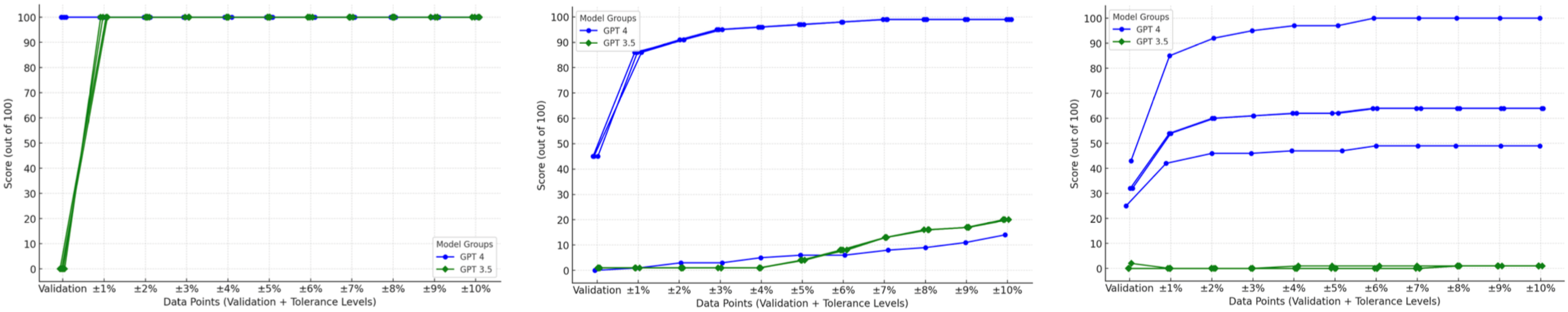
Starting from Scratch: Can LLMs Generate Code Without a Head Start?

Scenarios	LLM	Versions	CodeBertScore	MajorityVoteScore	WeightedScore	Ground Truth Score
Brackets	GPT 4	Version 7	0.9	1	0.935	100/100
		Version 2	0.899	1	0.934	100/100
		Version 4	0.899	1	0.934	100/100
		Version 5	0.899	1	0.934	100/100
	GPT 3.5	Version 9	0.894	0.94	0.91	0/100
		Version 2	0.892	0.94	0.909	0/100
		Version 6	0.903	0.06	0.608	0/100
		Version 8	0.894	0.06	0.602	0/100
Brackets + Deductions	GPT 4	Version 3	0.871	0.88	0.875	45/100
		Version 2	0.866	0.88	0.871	45/100
		Version 5	0.861	0.88	0.869	45/100
		Version 10	0.887	0.12	0.58	0/100
	GPT 3.5	Version 2	0.859	1	0.916	1/100
		Version 1	0.859	1	0.916	1/100
		Version 6	0.858	1	0.915	1/100
		Version 10	0.858	1	0.915	1/100
Brackets+Ded+EITC	GPT 4	Version 7	0.883	0.79	0.827	43/100
		Version 1	0.863	0.7	0.765	25/100
		Version 5	0.87	0.61	0.714	32/100
		Version 6	0.857	0.61	0.709	32/100
	GPT 3.5	Version 6	0.852	1	0.941	2/100
		Version 2	0.851	0.98	0.929	0/100
		Version 10	0.845	0.98	0.926	0/100
		Version 3	0.845	0.5	0.638	0/100

A Helping Hand: How Does Prior Code Affect Performance?

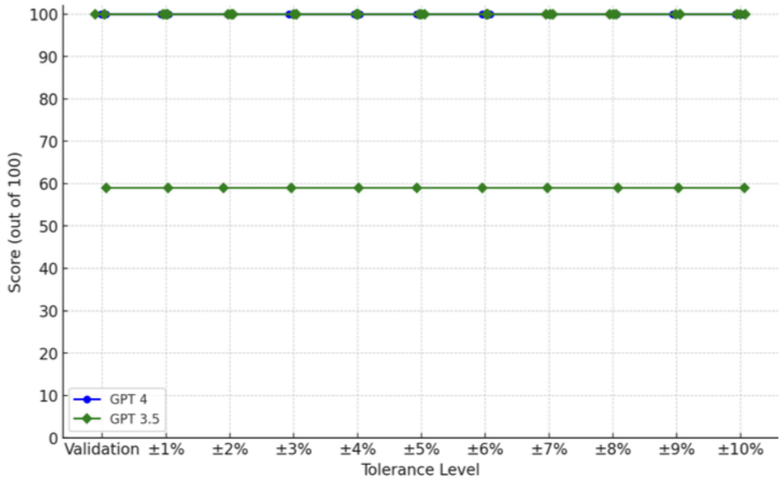
Scenario	LLM	Versions	CodeBertScore	MajorityVoteScore	WeightedScore	Ground Truth Score
Brackets	GPT 4	Version 3	0.914	1	0.944	100/100
		Version 5	0.911	1	0.942	100/100
		Version 9	0.911	1	0.592	100/100
		Version 4	0.91	1	0.941	100/100
	GPT 3.5	Version 1	0.941	1	0.962	100/100
		Version 2	0.939	1	0.96	100/100
		Version 7	0.937	1	0.959	100/100
		Version 8	0.936	0.59	0.815	59/100
Brackets + Deductions	GPT 4	Version 7	0.972	1	0.983	51/100
		Version 5	0.972	1	0.983	51/100
		Version 3	0.972	1	0.983	51/100
		Version 6	0.972	1	0.983	51/100
	GPT 3.5	Version 4	0.976	1	0.99	21/100
		Version 3	0.976	1	0.99	21/100
		Version 6	0.975	1	0.99	21/100
		Version 5	0.975	1	0.99	21/100
Brackets+Ded+EITC	GPT 4	Version 6	0.978	1	0.991	48/100
		Version 8	0.978	1	0.991	48/100
		Version 10	0.976	1	0.991	48/100
		Version 3	0.976	1	0.991	48/100
	GPT 3.5	Version 1	0.986	1	0.994	56/100
		Version 2	0.977	0.92	0.943	56/100
		Version 7	0.977	0.56	0.727	35/100
		Version 3	0.977	0.56	0.727	35/100

Beyond Perfect Matches: How 'Close' Are the LLMs Getting?

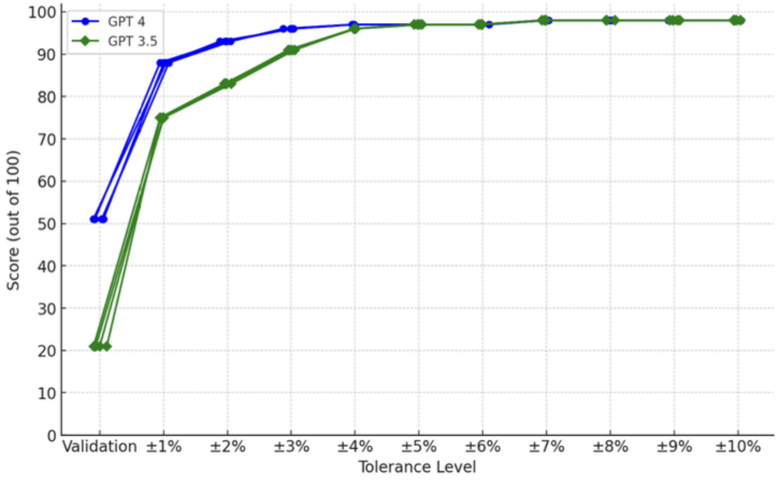


(a) Brackets (b) Brackets+Deductions (c) Brackets+Deductions+EITC

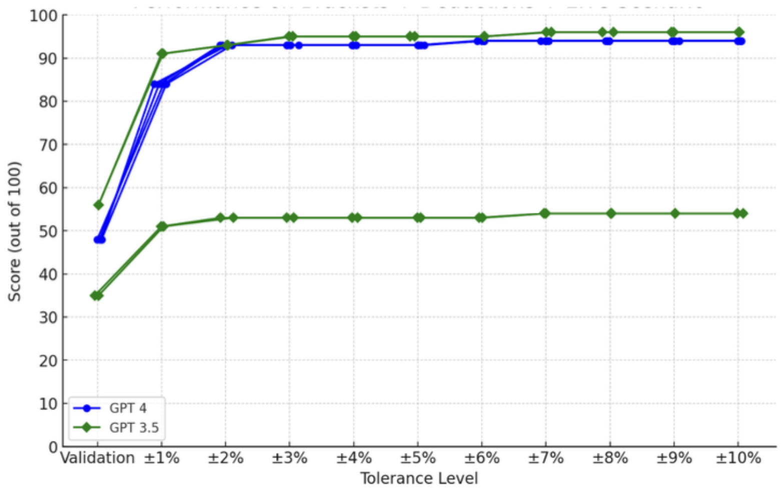
Prior Code = Higher Accuracy and Consistency



(a) Brackets



(b) Brackets+Deductions



(c) Brackets+Deductions+EITC

LLMs and the Future of Tax Software

- LLMs can automate tax software updates but need human expertise.
- Prior code context improves LLM accuracy and consistency.
- Complex tax logic (e.g., EITC) still challenges LLMs, needing refinement.
- Integrate robust testing (e.g., metamorphic) for code reliability.
- Develop feedback loops for continuous LLM improvement.

Smarter Tax Software: A Future Powered by AI

- **Enhanced Tax Compliance**
- **Collaboration of AI and Experts**
- **Our Mission to Improve Taxpayer Experience**

Thanks!

Any Questions?

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

More Information or More Frequent Information?

A Proposal For Quarterly 1099s

Kathleen DeLaney Thomas

UNC SCHOOL OF LAW

Today's Presentation

. . .

Based on earlier work:

**Rethinking Tax Information: The Case for Quarterly 1099s* (forthcoming So. Cal. Law Review 2024)

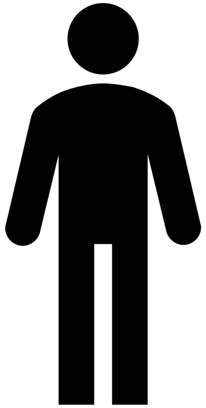
**Improving the Tax System for Independent Contractors: Quarterly 1099s* (Tax Notes, Jan. 1, 2024)

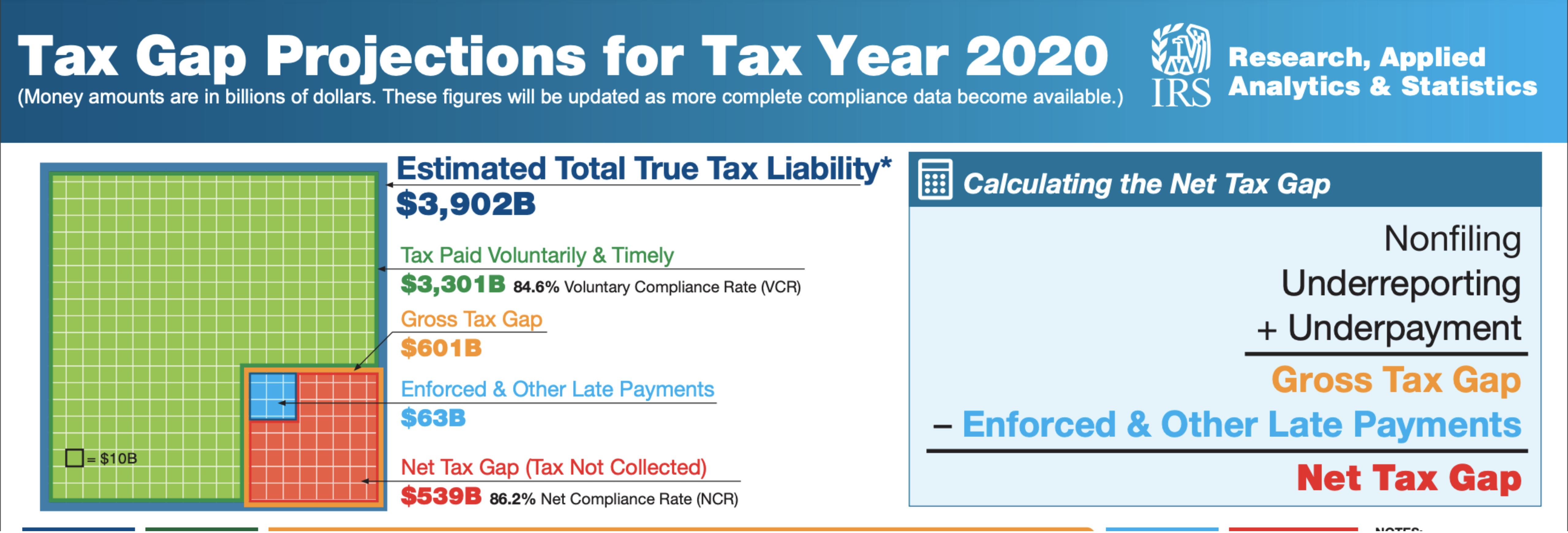
How should we think about the path forward for 1099-K reporting, particularly for online platforms?

UNC SCHOOL OF LAW

Focusing on taxpayers who receive business income from Third-Party Settlement Organizations (TPSOs)

Online Platform → Taxpayer (Sole Proprietor)





Compliance rate when employer withholding is present = 99%

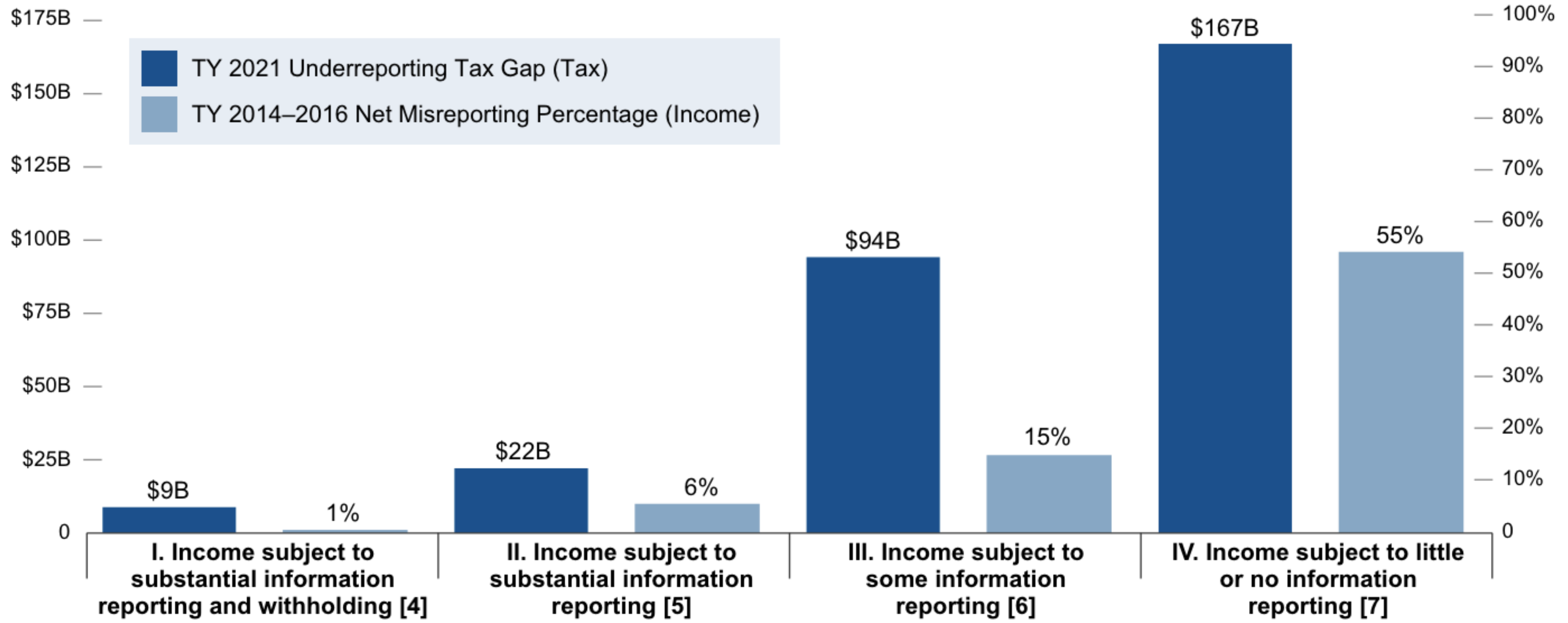
Compliance with no information reporting/withholding < 50%

Third-Party Information Reporting Is Effective

Tax Gap Projections for Tax Years 2020 and 2021

Figure 4. Effect of Information Reporting on Individual Income Tax Reporting Compliance, Tax Years 2014–2016 NMP Estimates and TY 2021 Projections

"Visibilty" Chart: Tax Year 2021 [1] Individual Income Tax Underreporting Tax Gap Projections and Tax Year 2014–2016 [2] Net Misreporting Percentage [3] Estimates by "Visibility" Category of Income Items



[1] These figures will be updated as more complete compliance data become available.
[2] The TY 2014–2016 estimate is the annual average for the TY 2014, 2015, and 2016 timeframe. This chart displays the tax gap attributable to the underreported income category and the rate at which that income is misreported as measured by the Net Misreporting Percentage.
[3] The Net Misreporting Percentage is the ratio of the net misreported amount to the sum of the absolute values of the amounts that should have been reported, expressed as a

Source:
IRS Pub. 5869

AMERICAN RESCUE PLAN (2021)

Lowered 1099-K Reporting Threshold for TPSOs

- Old rule: > $20,000 in payments + > 200 transactions
- New rule: > $600 in payments (no transaction minimum)

IRS has delayed enforcement: Old ($20k/200) threshold for 2023
Phased $5000 threshold for 2024

First Issue

Setting the Right 1099-K Threshold: Information Reporting is Not Without Costs

- **Administrative costs to third parties who must issue 1099s** (although some studies indicate these are relatively low)

- **Costs to IRS to process information returns**

- **Lower threshold for TPSO reporting may capture more nontaxable transactions and create confusion/complexity**

Second Issue

Compliance Challenges For Platform Workers…

- **Many platform workers do not receive a 1099-K** because they do not meet the $20,000/200 transaction threshold.
- **But even if they receive a Form 1099-K, platform workers struggle with saving for and remitting estimated taxes.**
  - Note, new $600 threshold doesn't necessarily solve this problem

Estimated Tax Compliance Is An Issue

- GAO report identified saving for and remitting estimated taxes as a top compliance challenge of platform workers (2020)
- A third of surveyed gig workers did not know whether they had to pay estimated taxes and nearly half did not set aside money for taxes (Bruckner 2016)
- TIGTA report found that 25% of taxpayers who received a Form 1099-K and filed 1040 did not report correctly and 13% did not report and pay self-employment taxes (2019)

Proposal

Estimated Taxes are Due Quarterly: Why Not Require TPSOs to Send 1099s Quarterly?

Form **1099-ES**

Box 1a. Gross Payments for the Payment Period January 1- March 31* $____________ *Or adjusted accordingly for the relevant payment period
Box 1b. Gross Payments Year to Date $__________________

Proposal

Quarterly 1099s (Form 1099-ES)

- Sent to taxpayer once a certain dollar threshold is reached for the quarter (e.g., $600, $5000, $10,000)
- Sent only to taxpayer; IRS receives year-end Form 1099-K only
- Sent after quarter ends but before estimated taxes are due
- Provide simple instructions + safe harbor calculation for paying estimated taxes

Proposal

Quarterly 1099s (Form 1099-ES)

Sample Schedule for Quarterly Form 1099-ES Deadlines

End of Payment Period	1099-ES Due Date	Estimated Tax Payment Due Date
March 31	April 5	April 15
May 31	June 5	June 15
August 31	September 5	September 15
December 31	January 5	January 15

Proposal

Form 1099-ES: Sample Instructions/Safe Harbor

Your estimated taxes for the period ending March 31 are due **April 15**.

You can pay your estimated taxes at https://www.irs.gov/account.

You may elect to calculate your payment for this payment period as **5% of the gross amount reported in Box 1a.**

*You may also elect to use other methods to calculate your estimated taxes. For more information see Publication 505, Tax Withholding and Estimated Tax.

Considerations for 1099-K and/or 1099-ES

- Compliance Benefit and Revenue
- Costs to Third Parties
- Costs to IRS
- Perceptions of Fairness

Compliance Benefit/Revenue

JCX estimates new $600 threshold for 1099-K will generate $8.4 billion (from 2021-31)

- Does this assume 94% compliance rate?
- Will new rules capture a less compliant group of taxpayers who don’t budget for taxes?

Revenue benefit of quarterly 1099s is uncertain –what does this add to 1099-K reporting?

- IRS “Estimated Tax Payments Program” (generic reminder notices, see Pub. 5901) indicates substantial revenue potential - **$53B (from 2028-2034, $7.5B/year)**
- Reminders from IRS v. third parties, which is more cost-effective?

Costs Imposed on Third Parties & IRS

Lower threshold for Annual Form 1099-K

- More information returns; higher costs to third parties
- Possible added complexity for taxpayers
- More returns for IRS to process

Quarterly Form 1099-ES

- Higher costs to third parties – new requirements, but already have the info
- Possible reduced complexity for taxpayers
- Modest additional cost to IRS – no additional returns to process; but must enforce quarterly requirement

Perceptions of Fairness?

- More information reporting may enhance perceptions of fairness – that everyone is paying their “fair share”
- The new $600 reporting threshold has received a lot of negative attention – taxpayers who should have been reporting income may perceive it has a new tax increase (and it has been falsely portrayed this way)
- TPSOs/interest groups generally oppose more information reporting
- Quarterly 1099s may enhance perceptions of fairness -> they are aimed at helping taxpayers pay estimated taxes; no new tax info is going to the government

Path Forward for 1099 Reporting for TPSOs?

Further study needed:

- Impact of the phased $5000 threshold for 1099-K in 2024
- How big of a burden on third parties would quarterly 1099 reporting impose?
- Impact of generic reminder notices v. quarterly 1099s (with taxpayer-specific info)

A possible compromise that could generate revenue and enhance fairness:

A "compromise" annual threshold for Form 1099-K
(e.g., $5000 or $10,000)
+
Quarterly 1099s

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

Investigating the Impact of a Free File Letter Intervention on Taxpayers' Return Filing and Preparation Methods

Pei-Hua Chen, Astin Cornwall, Anne D. Herlache, Scott Leary, Brenda Schafer, Melissa Vigil (Research, Applied Analytics, & Statistics) &
Rizwan Javaid (Office of the National Taxpayer Advocate)

Internal Revenue Service

IRS TPC June 13th, 2024

“Our new Constitution is now established; everything seems to promise it will be durable; but in this world nothing is certain except death and taxes.”

by Benjamin Franklin (1789).

Advantages of E-filing Returns

- The IRS is the federal agency with which a vast majority of citizens and businesses interact. Modernization initiatives like e-filing can improve the service delivery to the public.
- E-filing can be a win-win situation for both taxpayers and the IRS.
- Taxpayers can enjoy the convenience of filing electronically anytime, anywhere, with reduced errors and faster refunds.
- For the IRS, e-filing translates to streamlined administration, improved data accuracy, and shorter processing times.
- Despite these advantages, approximately 9% of taxpayers still chose paper filing in TY2022 (IRS, 2023a).

Factors related to E-Filing behavior

- Pippin and Tosun (2014) found that e-filing rates are lower in rural counties, counties with low population size, counties with a higher share of Hispanics and Asians, and counties with a higher share of the elderly population.
- A taxpayer experience survey (IRS, 2023b) found that taxpayers who are younger, self-prepare their returns, or have limited English proficiency were more likely to be interested in e-fling. Participants in the study also indicated that cost and privacy were key factors in their decision to use an online filing platform.

Factors related to E-filing behavior

- Wang (2003) studied the factors affecting the adoption of e-filing and found that computer self-efficacy had significant effect on adoption intention.
- Parsad, Jones & Greene (2005) showed that the percentage of public schools with internet access increased from 35% in 1994 to 95% in 2005. The number of Americans with internet access at home was 67% in 2001 (Perrin & Duggan, 2015).
- Generational (or age) differences in computer self-efficacy, influenced by the late prevalence of internet access after the 2000s, likely impact e-filing adoption.

Purpose of the Study

- Our research focused on understanding how to increase e-filing adoption, especially among lower income taxpayers who qualify for IRS Free File.
- By removing the cost barrier associated with e-filing software, we aimed to see if making filing essentially free would influence taxpayers' filing decision.
- This study utilized an intervention strategy, sending either an informational letter or a filing checklist to 125,000 taxpayers whose 2021 adjusted gross income (AGI) was less than $73,000 to evaluate the impact of the outreach on their choice of filing and preparation method.

Research Questions

1. How does the provision of a Free File letter influence taxpayers' tax filing choice between e-filing and paper filing?
2. Are there any demographic differences in how the treatments influence the decision to e-file or their tax preparation method (age, urban/rural, income tax complexity, filing experience)?
3. How does the provision of a Free File letter influence taxpayers' tax preparation choice (i.e., paper, free file, software, paid preparer, VITA, software-prepared paper-filed returns)?

Method

- Population: Taxpayers who self-prepared a paper return in TY2021 with income of $73K or less, excluding habitual paper filers (those who paper filed for the prior three tax years).
- Sample: 125,000 taxpayers during filing season 2023, broken into two strata.

Table 1: Descriptions of the two strata in our sample population

Stratum	Median Age	Median Income	Total Taxpayers
Repeat Filers	54	$14,287	652,027
New/Infrequent Filers	25	$4,143	105,300

Research Design and Mailing

- 125,000 taxpayers were randomly assigned to each of the two treatment groups with 5 mailings (25,000 each) based on the timing of their TY2021 return filing.
- After removing undeliverable mail, there were approximately 53,000 in each treatment group. A control group sample of 53,600 was randomly chosen across strata.

Group No.	Group Type	Correspondence Content	N (delivered)
1	No-Contact Control	None	53,600
2	Treatment Group 1	Free File Letter (Letter 6171): You may be qualified for Free File: fast refund, fewer errors and free	62,500 (53,473)
3	Treatment Group 2	Checklist to file tax (Publication 5732)	62,500 (53,370)

Treatment 1: Free File Letter (L6171)

Faster refund? ✔ Fewer errors? ✔ Free? ✔
Check your eligibility for IRS Free File today!

What you need to know
There are many potential advantages to free online tax preparation:

- Free electronic filing of your federal tax return.
- Getting your refund faster.
- Access to free commercial software for federal and state returns.
- Less chance of making a mistake on your tax return or missing a tax benefit, like the Earned Income Tax Credit (EITC).

Read below for information about free IRS-sponsored programs.

Free File program
What is the Free File Program?

- Free File provides free commercial software to help prepare your return online.
- Most taxpayers qualify if they earned $73,000 or less in 2022.
- You will need only your 2021 tax return, 2022 tax documents, and a valid email address to begin.
- **For more information, visit www.irs.gov/FreeFile.**

Other information

- If you have questions about this letter, you can call 888-525-6797 (toll-free).
- You don't need to respond to this letter.

Treatment 2: Tax Filing Checklist (P5732)

Tax Filing Checklist

The checklist below will assist you in properly filing your federal income tax return and help you avoid costly penalties for filing incorrectly.

#	Action	✓
1.	**I used the correct filing status.** • If you are married and living with your spouse, neither of you may file a Head of Household return. • For help selecting the correct filing status, visit irs.gov/help/ita/what-is-my-filing-status.	
2.	**I used my correct address.** • The IRS must be able to contact you by mail if there is a question about your return. This would be the address where you live or regularly receive your mail.	
3.	**I reported all of my income.** • You must report all taxable income as well as tax-exempt interest. *Note:* Generally, all income you receive is taxable, including income from bartering. Money and assets that you receive as a gift or inheritance are not taxable to you.	
4.	**I claimed only the deductions to which I am entitled.** • Be sure to claim all allowable expenses. Maintain records of those expenses for at least three years. • If you are self-employed, see Publication 535 for information on expenses you may claim for your business. To view the publication, go to irs.gov/publications/p535.	
5.	**I claimed only the credits to which I am entitled.** • For more information about the Earned Income Tax Credit (EITC), visit irs.gov/eitc. • For more information about the Child Tax Credit, visit irs.gov/ctc.	
6.	**I signed my return.** *Important:* If you are married filing a joint return, both spouses must sign the return.	
7.	**I saved a copy of my return.** You should keep a copy of your tax return for at least three years.	
8.	I filed my return on____________________(enter date)	

Tips to remember when selecting a preparer:

- **Ask about Service Fees.** Avoid preparers who base fees on a percentage of the refund or who boast bigger refunds than their competition. When asking about a preparer's services and fees, don't give them tax documents, Social Security numbers or other information before you decide to hire the preparer.
- **Make Sure the Preparer is Available.** Make sure your preparer will be available after your return is filed to

Dependent Variables

- Income tax submission method (binary): E-filing vs. paper filed.
- Tax preparation methods (categorical): Free file, VITA, paid preparer, self-on-paper, and software-prepared paper filed returns.

Demographic and Socio-Economic Variables

Age or Age Groups

Age is treated as a continuous and a control variable in the analysis. Age is categorized into distinct groups to create interaction terms and mitigate potential multicollinearity issues. Age groups are as follows:

Group	Age Range
1	Under 30
2	30 - 44
3	45 - 59
4	60 -74
5	75 and over

Demographic and Social-Economic Variables

- **Income**: Adjusted Gross Income (AGI) is treated as a control variable. AGI was standardized using AGI-(mean(AGI))/SD(AGI). The imputation of missing values was created from the median of each combination of strata, treatment group, age group, and urbanicity.
- **Income Tax Complexity Score**: Tax returns were assigned a complexity score (1 to 5) based on the types of income, deductions, and credits reported. A higher score indicates a more complex return.
- **Urbanicity**: Participants' zip codes were matched with the zip code tabulation area (ZCTA) population density data from the 2020 Census to create the urbanicity variable. The minimum population to be classified as an urban area is 5,000. The urbanicity variable is binary and is equal to 1 if urban and 0 for rural areas.

General Form of Logistic Regression Model

- A logistic regression model predicts the likelihood of tax preparation methods (multiple categorical outcomes) or e-file adoption (coded as 1 for e-filed and 0 for paper filed) and is represented as follows:

$$\log\left(\frac{P(Y=j)}{P(Y=m)}\right) = \beta_{0j} + \beta_{1j}X_1 + \beta_{2j}X_2 + \cdots + \beta_{kj}X_k$$

For $j = 1, 2, \ldots, m - 1$

- Y is the categorical outcome variable with m categories. (m = 2 for binary outcome)
- $X_1, X_2, \ldots, X_k$ are predictor variables.
- $\beta_{0j}, \beta_{1j}, \ldots, \beta_{kj}$ are coefficients for category *j*.
- $P(Y = j)$ is the likelihood of choosing category *j.*

General Form of Logistic Regression Model

- The likelihood of choosing a specific category (*j*) for the tax preparation method or e-file (*Y*) compared to a chosen reference category (self on paper) can be represented as a logistic regression function as follows:

$$\mathrm{P}(Y = j|X) = \frac{\exp(\boldsymbol{X'\beta_j})}{\sum_{k=1}^{J} \exp(\boldsymbol{X'\beta_k})}$$

- $\mathbf{P}(\boldsymbol{Y} = \boldsymbol{j}|\boldsymbol{X})$: Represents the probability of a taxpayer choosing category j (e.g., Free File) for their tax preparation method or for e-file given the set of independent variables (*X*).
- $\boldsymbol{X'}$: Represents the vector of independent variables transposed.
- $\boldsymbol{\beta_{ij}}$: Represents the vector of coefficients associated with each independent variable for category *j*. These coefficients indicate the magnitude and direction of the effect of each variable on the odds of choosing category j compared to the reference category.

Descriptive Results of E-Filing

Group	Filing Rate (%)	E-File Rate (%)	E-File Rate (%) Repeat Filers	E-File Rate (%) New/Infrequent Filers
No-Contact Control	60.0	38.7	37.1	43.0
Free File Letter	68.5	39.7	37.7	44.6
Checklist	69.2	38.2	36.6	42.2

Results: E-filed vs. Paper Filed

- Repeat filers are 1.12 times more likely to e-file than new filers.
- As complexity increases one unit, a taxpayer is 1.14 times more likely to e-file their tax return.
- For each additional year of age, taxpayers are 1.6% less likely to e-file.
- Rural residents are 1.6% less likely to e-file.
- As AGI increases by one S.D., taxpayers are 5.15 times more likely to e-file.
- Taxpayers who received either the free file letter or checklist were more likely to e-file compared with the no contact group.

Independent Variable	B	SE	Odds of E-filing
Strata	0.115***	0.00384	1.122
Complexity	0.133***	0.00175	1.142
Age	-0.016***	0.00014	0.984
Rural	-0.017***	0.00170	0.984
AGI_S	1.640***	0.01010	5.153
Checklist	0.288***	0.00667	1.333
Letter	0.328***	0.00665	1.389
Constant	-0.717	0.00781	0.488
Chi-square	259.574		
Degrees of freedom	18		

Note: AGI_S is standardized Adjusted Gross Income.

Reference group is strata 2 (new filers) for the variable strata**; Age group < 30 is the** reference group for Age Group; **The Control group** is the reference category for Treatment. The reference group of Rural is **urban**.

*p < .05. **p < .01. ***p < .001.

Results: E-filed vs. Paper Filed (Interactions)

- The checklist was more effective in encouraging e-filing among individuals over 75 years of age compared with those under 30.
- Both the youngest and oldest age groups were most likely to e-file, while other age groups were less responsive to the mailings.

Independent Variable	B	SE	Odds of E-filing
Checklist x Age Group: 30-44	-0.532***	0.009	0.587
Checklist x Age Group: 45-59	-0.313***	0.009	0.731
Checklist x Age Group: 60-74	-0.077***	0.009	0.926
Checklist x Age Group: >=75	0.114***	0.012	**1.121**
Letter x Age Group: 30-44	-0.475***	0.009	0.622
Letter x Age Group: 45-59	-0.348***	0.009	0.706
Letter x Age Group: 60-74	-0.092***	0.009	0.912
Letter x Age Group: >=75	0.0002	0.012	**1.000**

Note: AGI_S is standardized Adjusted Gross Income.

Reference group is strata 2 (new filers) for the variable strata; **Age group <30 is the reference group** for Age Group; The Control group is the reference category for Treatment. The reference group of Rural is **urban**.

*p < .05. **p < .01. ***p < .001.

Results: Tax Preparation Method (Repeat Filers)

- Repeat filers who received the free file letter were 1.43 times more likely to choose Free File over self-prepared on paper.
- Repeat filers with a one-S.D. increase in income were 2.79 times more likely to use tax preparation software than to file on paper.
- Repeat filers with potentially more complex returns preferred tax preparation software.

Odds	Effect	Estimated	SE	exp(b)	p-value
Free File vs. Self on Paper	Intercept	-1.587	0.014	0.204	0.000
	AGI_S	0.046	0.026	1.047	0.074
	Age	-0.017	0.000	0.983	0.000
	Urban	0.002	0.009	1.002	0.793
	Complexity	-0.076	0.005	0.927	0.000
	Letter	0.357	0.011	**1.429**	0.000
	Checklist	-0.048	0.011	0.953	0.000
Software vs. Self on Paper	Intercept	0.372	0.007	1.451	0.000
	AGI_S	1.027	0.011	**2.792**	0.000
	Age	-0.023	0.000	0.977	0.000
	Urban	0.052	0.004	1.054	0.000
	Complexity	0.125	0.002	**1.133**	0.000
	Letter	-0.006	0.005	0.994	0.192
	Checklist	-0.010	0.005	0.990	0.044

Results: Tax Preparation Method (Repeat Filers)

- Repeat filers who received the checklist were 1.1 times more likely to go to a VITA center than self prepared on paper.
- Repeat filers with one standard deviation increase in income were 5.6 times more likely to use a paid preparer than self prepared on paper.

Odds	Effect	Estimated	SE	exp(*b*)	*p*-value
VITA vs. Self on Paper	Intercept	-4.451	0.023	0.012	0.000
	AGI_S	-0.168	0.030	0.845	0.000
	Age	0.032	0.000	1.033	0.000
	Urban	0.090	0.011	1.094	0.000
	Complexity	-0.245	0.007	0.783	0.000
	Letter	0.045	0.013	1.046	0.000
	Checklist	0.098	0.012	**1.103**	0.000
Paid Preparer vs. Self on Paper	Intercept	-0.821	0.009	0.440	0.000
	AGI_S	1.720	0.014	**5.587**	0.000
	Age	-0.014	0.000	0.986	0.000
	Urban	0.039	0.006	1.039	0.000
	Complexity	0.048	0.003	1.049	0.000
	Letter	-0.016	0.007	0.984	0.017
	Checklist	-0.041	0.007	0.960	0.000

Results: Tax Preparation Method (New/Infrequent Filers)

- New filers who received the free file letter were 1.45 times more likely to use free over self-prepared on paper.
- New filers with one S.D. increase in income were 2.76 times more likely to use tax preparation software instead of self prepared on paper.

Odds	Effect	Estimated	SE	exp(*b*)	*p*-value
Free File vs. Self on Paper	Intercept	-1.678	0.062	0.187	0.000
	AGI_S	-0.262	0.173	0.769	0.131
	Age	-0.024	0.001	0.976	0.000
	Urban	-0.053	0.040	0.948	0.188
	Complexity	0.089	0.027	1.093	0.001
	Letter	0.374	0.047	**1.454**	0.000
	Checklist	-0.007	0.050	0.993	0.894
Software vs. Self on Paper	Intercept	0.469	0.030	1.598	0.000
	AGI_S	1.016	0.077	**2.762**	0.000
	Age	-0.024	0.001	0.976	0.000
	Urban	-0.028	0.021	0.972	0.176
	Complexity	0.022	0.014	1.022	0.126
	Letter	-0.052	0.024	0.949	0.032
	Checklist	-0.119	0.024	0.888	0.000

Results: Tax Preparation Method (New/Infrequent Filers)

- New filers living in urban areas were more likely to use a VITA center or a paid preparer instead of self-prepared on paper.
- New filers with one S.D. increase in income were 4.45 times more likely to use a paid preparer instead of self-prepared on paper.
- As the complexity score increased one unit, new filers were 1.18 times more likely to use a paid preparer than self-prepared on paper.

Odds	Effect	Estimated	SE	exp(*b*)	*p*-value
VITA vs. Self on Paper	Intercept	-4.501	0.124	0.011	0.000
	AGI_S	-1.375	0.325	0.253	0.000
	Age	0.030	0.002	1.031	0.000
	Urban	0.118	0.081	**1.125**	0.147
	Complexity	-0.307	0.063	0.735	0.000
	Letter	-0.152	0.092	0.859	0.099
	Checklist	-0.347	0.096	0.707	0.000
Paid Preparer vs. Self on Paper	Intercept	-0.936	0.040	0.392	0.000
	AGI_S	1.492	0.092	**4.447**	0.000
	Age	-0.015	0.001	0.986	0.000
	Urban	0.105	0.028	**1.111**	0.000
	Complexity	0.165	0.017	**1.180**	0.000
	Letter	-0.056	0.033	0.946	0.086
	Checklist	-0.118	0.033	0.888	0.000

Results: E-Filed vs. Paper-Filed

- Taxpayers who received either the free file letter or checklist were more likely to choose e-filing compared with the no contact group.
- Surprisingly, people 75 and over who received either a letter or a checklist were more likely to e-file their taxes.
- One possible explanation is that taxpayers over 75 may be less likely to prepare their taxes by themselves (i.e., they may have sought informal assistance) after they received the letter or checklist.

Results: Tax Preparation Method

- People who received the Free File letter were more likely to choose Free File compared to the control group.
- People with higher income tended to prepare their taxes using software or seek professional help from a paid preparer.
- The results showed that the income tax complexity affected taxpayers' tax preparation method differently depending on their filing experience:
  - Repeat filers with potentially more complex returns tended to utilize tax preparation software possibly 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.

Discussion

- This study offers valuable insights for promoting electronic filing adoption, particularly among taxpayers who qualify for the Free File program.
- Recognizing the impact of demographics on filing preferences can help tailor future initiatives.
- The Free File letter’s success indicates that broader public awareness campaigns, possibly with partners like tax software providers, public libraries, or IRS taxpayer service centers, can expand Free File to a wider audience.

Discussion

- The study focus only on taxpayers eligible for free e-filing which limits generalizability to the entire taxpayer population.
- The timing of the study (during pandemic) may limit its applicability in different tax years or under different economic conditions.
- Our follow-up study incorporates more comprehensive benefits, addresses concerns about e-filing in the modified letters, and provides a better understanding of adoption across the entire taxpayer income spectrum.

References

Internal Revenue Service. (2023a) *IRS Data Book 2022*.

Internal Revenue Service. (2023b) IRS Report to Congress: Inflation Reduction Act §10301(1)(B) IRS-run Direct e-File Tax Return System (Publication 5788 (5-2023))[Report]. Department of the Treasury. Retrieved from https://www.irs.gov/pub/irs-pdf/p5788.pdf

Parsad, B. & Jones, J. & Greene, B. (2005) Internet access in U.S. Public Schools and Classrooms: 1994-2003. E.D. Tab. NCES 2005-015. US Department of Education.

Perrin, A. & Duggan, M. (2015, June 26). *Americans' Internet Access: 2000-2015.* Pew Research Center. https://www.pewresearch.org/internet/2015/06/26/americans-internet-access-2000-2015/

Pippin, S. E., & Tosun, M. S. (2014). Electronic tax filing in the United States: An analysis of possible success factors. *Electronic Journal of e-Government, 12*(1), 20-36.

Wang, Y. S. (2003). The adoption of electronic tax filing systems: an empirical study. *Government Information Quarterly, 20*(4), 333-352.

Thank you

Sample Selection Criteria

Strata 1: Frequent Filers

- Taxpayers who self-prepared a paper return in TY2021 with income of $73K or less
- Taxpayers who filed at least one return between TY2018 to TY2020
- Taxpayers who did not file a paper return every year between TY2018 to TY2021

Strata 2: New or Infrequent Filers

- Taxpayers who self-prepared a paper return in TY2021 with income of $73K or less
- Taxpayers who did not file any returns between TY2018 to TY2020

Crosstabulation of Tax Preparation Methods: Frequent Filers

	Treatment Group		
Tax Preparation Method	**Control Group**	**Checklist Letter**	**Letter**
Free File	866	940	1365
Paid Preparer	2910	3108	3094
Paper	12049	13600	13160
Software	7063	7830	7630
V_CODE	0	1	1
VITA	665	838	770
Not Filed	12197	9310	9657
Total	**35750**	**35627**	**35677**

Crosstabulation of Tax Preparation Methods: New or Infrequent Filers

		Treatment Group	
Tax Preparation Method	**Control Group**	**Checklist Letter**	**Letter**
Free File	393	552	765
Paid Preparer	1171	1355	1358
Paper	4120	5140	4855
Software	2783	3479	3509
V_CODE	0	0	1
VITA	136	101	116
Not Filed	9247	7116	7192
Total	**17850**	**17743**	**17796**

Correlation Matrix

Variable	E-Filed	Age	AGI	URBAN	complexity
E-Filed	1.000				
Age	-0.190	1.000			
AGI	0.065	0.125	1.000		
URBAN	-0.006	-0.027	0.031	1.000	
Complexity	-0.098	0.353	0.103	0.010	1.000

Research, Applied Analytics & Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

14th Annual IRS/TPC Joint Research Conference on Tax Administration

#LiveAtUrban

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

IRS-TPC Joint Research Conference

Session 4: Simplifying the Filing Burden

June 13, 2024

Robert Weinberger

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

Sina Gogani-Khiabani, *et al.*

Technical Challenges in Maintaining Tax-Prep Software with LLMs

1. Premise: Tax Code and Regs are constantly changing
2. Tax preparation software needs yearly updates
3. Present manual updates are time-consuming, error-prone
4. Can AI-LLMs automate the process?
5. Test several scenarios of increasing complexity

Paper's Conclusions

1. LLMs work better when we build on prior software vs. starting fresh
2. GPT-4.0 is more accurate and consistent than 3.5
3. Learning grows through repeated testing and feedback
4. More complexity increases errors
5. Human expertise is still needed

Comments

1. It's worth detailing what tax software developers now do.
2. The authors may underestimate the Tax Code's complexity and need for interpretation.
3. The paper needs to address hallucinations.
4. Evaluating accuracy using a model of similarly-situated taxpayers differs from optimizing the outcome for an individual taxpayer, a tougher challenge; it may also miss vulnerabilities to fraud.
5. What are acceptable margins of error—tolerances?
6. ➡ Has potential but not yet ready for prime-time

Rethinking Tax Information: The Case for Quarterly 1099s

Kathleen DeLaney Thomas

☐ CORRECTED (if checked)

FILER'S name, street address, city or town, state or province, country, ZIP or foreign postal code, and telephone no.

FILER'S TIN

PAYEE'S TIN

OMB No. 1545-2205

Form **1099-K**

(Rev. March 2024)

For calendar year ______

Payment Card and Third Party Network Transactions

1a Gross amount of payment card/third party network transactions
$

1b Card Not Present transactions
$

2 Merchant category code

Check to indicate if FILER is a (an):
- Payment settlement entity (PSE) ☐
- Electronic Payment Facilitator (EPF)/Other third party ☐

Check to indicate transactions reported are:
- Payment card ☐
- Third party network ☐

3 Number of payment transactions

4 Federal income tax withheld
$

Copy B

For Payee

This is important tax information and is being furnished to the IRS. If you are required to file a return, a negligence penalty or other sanction may be imposed on you if taxable income results from this transaction and the IRS determines that it has not been reported.

PAYEE'S name

Street address (including apt. no.)

City or town, state or province, country, and ZIP or foreign postal code

PSE'S name and telephone number

Account number (see instructions)

5a January $	**5b** February $
5c March $	**5d** April $
5e May $	**5f** June $
5g July $	**5h** August $
5i September $	**5j** October $
5k November $	**5l** December $

6 State	**7** State identification no.	**8** State income tax withheld
		$
		$

Form **1099-K** (Rev. 3-2024) (Keep for your records) www.irs.gov/Form1099K Department of the Treasury - Internal Revenue Service

Problems & Solutions

1. IRS needs more information to improve compliance and close the tax gap, especially where there is limited or no third-party income reporting

   ➡ ***2021 law lowers the 1099-K reporting threshold from $20,000 and > 200 transactions to $600 for all transactions. It will raise >$500m/yr.***

2. Gig workers, sellers of goods, providers of services, and renters of property using payment cards, apps, or online marketplaces may be uncertain as to their income, their employment and income tax obligations, and their need to save enough for quarterly estimated payments

   ➡ ***IRS delayed the 2023 effective date and plans a $5,000 phase-in for 2024 to address complaints, reduce confusion, improve planning, saving, and accurate filing, and give the IRS time to modify forms***

Quarterly 1099 Proposal

1. The proposal addresses IRS information needs; could help with confusion/errors
2. But burdens would increase: Under existing law number of 1099-Ks sent would jump from 14m to 46m in 2025 (~84,000 would be filed on paper); this would add 184m quarterly 1099-ESs
3. 1099-Ks would be sent to many more without a tax obligation
4. Stakeholder outcry would likely increase
5. Congressional approval is possible – if the threshold is raised >$600
6. Illustrates trade-offs of competing goals: improving compliance vs. reducing burdens

Comments

1. Excellent paper, well presented, creative
2. Might also explore whether, with taxpayer consent, payors could share information with tax practitioners to educate their clients
3. Might it deter economic activity if sellers decide it’s not worth it?
4. IRS enforcement would be needed, but likely?
5. Will safe harbors help—5% proposed?
6. Reporting threshold compromise legislation seems likely

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

Pei-Hua Chen, *et al.*

The Impact of Letters on E-filing and Tax Preparation

- **Problem:** Still too many paper return filers (15m or 9%)
- **Solution:** Convert them to e-filers via persuasive outreach
- **Study Focus:** Frequent and new/infrequent paper filers
  - Include those eligible for Free File (to remove the cost obstacle)
  - Exclude habitual paper filers (to hit those more likely to change)
- **Treatment:** 125,000 taxpayers in 5 waves timed to match 2021 filing date
  - Send an IRS letter / Send IRS checklist / No contact = control
  - Sort by age; urbanicity; filing history; return complexity; and AGI

Faster refund? ✔ Fewer errors? ✔ Free? ✔
Check your eligibility for IRS Free File today!

What you need to know

There are many potential advantages to free online tax preparation:

- Free electronic filing of your federal tax return.
- Getting your refund faster.
- Access to free commercial software for federal and state returns.
- Less chance of making a mistake on your tax return or missing a tax benefit, like the Earned Income Tax Credit (EITC).

Read below for information about free IRS-sponsored programs.

Free File program

What is the Free File Program?

- Free File provides free commercial software to help prepare your return online.
- Most taxpayers qualify if they earned $73,000 or less in 2022.
- You will need only your 2021 tax return, 2022 tax documents, and a valid email address to begin.
- **For more information, visit www.irs.gov/FreeFile.**

Other information

- If you have questions about this letter, you can call 888-525-6797 (toll-free).
- You don't need to respond to this letter.

E-filing Context

- **Advantages:** (1) **Speedier refunds**; (2) lower processing costs; (3) fewer errors
- **Barriers:** (1) Cost (state or federal); (2) unable to e-file many forms, schedules, attachments; (3) e-file rejections; (4) overriding software blocks e-filing; (5) fear of increased audit risk; (6) security and privacy concerns; (7) confusion about how e-file works; (8) unaware e-file is more accurate; (9) no need for faster refund or balance-due; (10) lack of technology; (11) taxpayer preference; (12) initially, preparer resistance
- **Progress:** *e.g.*, PINs vs. 8453; 2-D bar coding/OCR scanners; CADE; postcards
  - Most individual returns are e-filed (**91%** in 2023)
  - All returns are eligible for free filing
  - Half of DIY returns are already filed free

E-file Milestones

- 1986 first e-file tests
- 1992 TeleFile starts (1040EZs)
- 1994 CERCA formed
- 1998 IRS RRA (goal: 80% by 2007)
- 1999 IRS reinstates Debt Indicator
- 2000 CADE starts
- 2003 Free File starts
- 2004 Modernized e-file debuts
- 2005 e-filing = 50%+; TeleFile halts
- 2008 first MITRE e-file study
- 2008 PIN replaces paper signature for e-file
- 2009 phased preparer e-filing mandated
- 2009 CADE 2 starts
- 2010 IRS halts mailing 1040 booklets
- 2017 80% of major returns e-filed
- 2020 Pandemic disruptions, paper backlogs
- 2023 91% of 163m returns e-filed
- 2024 Free File extended to 2029
- 2024 IRS advances scanning technology
- 2025 Direct File expands, made permanent
- 1986-2024 GAO, TIGTA, TAS, ETAAC, MITRE, etc. studies; IRS Blueprints, Strategic Plans

https://www.irs.gov/pub/irs-news/fs-11-10.pdf

Conclusions

- The study finds dozens of results but the bottom line is:
  - Taxpayers who received either mailing were more likely to choose e-filing compared with those who did not. (But not by much.)
  - Few surprises. In most respects, the paper validates other studies.

Comments

1. Needs a discussion of past studies. Does it advance insights over what we already know? Are the results statistically significant?
2. The letter is not persuasive or compelling. Doesn’t reflect behavioral insights on what is most motivating or lessons from advertising. The checklist doesn’t mention e-filing.
3. Those studied had very low incomes. Did they possibly not need to file?
4. What’s really motivating? **Refunds**.
5. A common reason for not e-filing is a lack of awareness that can be addressed through IRS marketing/advertising.
6. Should the IRS focus on reducing barriers, declare victory, let nature take its course, and use its 900 new scanners to capture the stubborn holdouts?