The IRS Research Bulletin

Proceedings of the 2023 IRS / TPC Research Conference

Research, Applied Analytics & Statistics

Publication 1500 (Rev. 5-2024) Catalog Number 11546J Department of the Treasury **Internal Revenue Service** www.irs.gov

Papers given at the

13th Annual Joint Research Conference on Tax Administration

Cosponsored by the IRS and the Urban-Brookings Tax Policy Center

June 22, 2023

Compiled and edited by Alan Plumley*
Research, Applied Analytics, and Statistics, Internal Revenue Service

*Prepared under the direction of Barry W. Johnson, IRS Chief Research and Analytics Officer

Foreword

This edition of the IRS Research Bulletin (Publication 1500) features selected papers from the IRS-Tax Policy Center (TPC) Research Conference held on June 22, 2023, at the Brookings Institution in Washington, DC. Conference presenters and attendees included researchers from many areas of the IRS, officials from other government agencies, and academic and private sector experts on tax policy, tax administration, and tax compliance. Many people participated in this, our first in-person conference in several years. Videos of the presentations are archived on the Tax Policy Center website to enable additional participation.

The conference began with welcoming remarks by Wendy Edelberg, Director of the Hamilton Project at the Brookings Institution, Eric Toder, Institute Fellow at the Urban-Brookings Tax Policy Center, and Barry Johnson, then Deputy Chief Data and Analytics Officer in the IRS Office of Research, Applied Analytics and Statistics. The remainder of the conference included sessions on taxpayer service, estimating audit aftershocks, understanding contemporary taxpayers, and hidden assets and networks. The keynote speaker was Washington Post columnist Catherine Rampell, who offered her insights on contemporary tax policy and tax administration issues.

We trust that this volume will enable IRS executives, managers, employees, stakeholders, and tax administrators elsewhere to stay abreast of the latest trends and research findings affecting tax administration. We anticipate that the research featured here will stimulate improved tax administration, additional helpful research, and even greater cooperation among tax administration researchers worldwide.

Acknowledgments

This IRS-TPC Research Conference was the result of preparation over a number of months by many people. The conference program was assembled by a committee representing research organizations throughout the IRS. Members of the program committee included: Alan Plumley, Brett Collins, Kelly Dauberman, and Valentina Kachanovskaya (Research, Applied Analytics, and Statistics); Anne Dayton (SB/SE Division); Brittany Jefferson (W&I Division); and Rob McClelland (Tax Policy Center). In addition, Megan Waring, of the Brookings Institution, and John Buhl, of the Urban Institute, oversaw numerous details to ensure that the conference ran smoothly.

This volume was prepared by Lisa Smith (layout and graphics), Anne McDonough (editor), and Beth Kilss (contractor), all of the IRS Statistics of Income Division. The authors of the papers are responsible for their content, and views expressed in these papers do not necessarily represent the views of the Department of the Treasury or the Internal Revenue Service.

We appreciate the contributions of everyone who helped make this conference a success.

Barry Johnson
IRS Chief Data and Analytics Officer

13th Annual IRS-TPC Joint Research Conference on Tax Administration

Contents

3. Understanding Contemporary Taxpayers (Continued)

4. Hidden Assets, Hidden Networks

5. Appendix

1

▽

Service Is Our Surname

Millard ◆ Faruqi ◆ Flateman ◆ Mirabito
Smolenski ◆ Stavrianos ◆ Szczerbinski

Framinan ◆ Greco ◆ Murphy ◆ Rasey
Alvarez ◆ Colona

Maag ◆ Airi ◆ Hunter

Looking Beyond Level of Service: Using Behavioral Insights To Improve Taxpayer Experience

Jan Millard (IRS, RAAS), Omar Faruqi, Jonah Flateman, Jamil Mirabito, Sarah Smolenski, Michael Stavrianos, and Lauren Szczerbinski (ASR Analytics LLC)

In 2016, as part of the Servicewide Future State Initiative, the IRS initiated a notice redesign effort focusing on Collection notices issued through the Automated Collection System (ACS) (e.g., LT11, LT16) as well as those issued prior to ACS entry (e.g., CP14, CP501, CP503, CP504). The redesigned notices included changes to wording and format which collectively guide taxpayers towards desired behaviors and away from undesired behaviors. These "behavioral nudges" were designed based on a robust and rapidly growing body of research from the behavioral sciences (e.g., psychology, neuroscience, behavioral economics), which examines how individuals absorb, process, and react to information, and applies this knowledge to design practical policies and interventions with human behavior in mind.[1] IRS has explored the application of behavioral nudges through other taxpayer communication channels, including recorded announcements in the Customer Voice Portal (CVP) system.[2] This initiative was prompted, in part, by a 2019 study conducted by the United Kingdom's tax authority, which found taxpayers were much more likely to "channel shift" (i.e., abandon a telephone call in favor of web-based self-service) after hearing certain recorded messages containing behavioral nudges.

This paper discusses a pilot test conducted to evaluate the efficacy of a sequence of CVP message prompts redesigned using behavioral insights to encourage callers routed to ACS Application 75 (App 75) to abandon the call queue and shift to online service channels.[3] The study team used behavioral design techniques to develop an alternative sequence of voice prompts with the taxpayer experience in mind, aiming to increase awareness of online resources relevant to specific tax issues (e.g., establishing an online account to access tax return information and view payment history) and provide callers with information necessary to consider self-service channels to resolve their issue rather than continue to wait on hold for a Customer Service Representative (CSR). As the IRS seeks to reduce costs, improve taxpayer compliance, and enhance the overall taxpayer experience, redesigning CVP announcements based on research from behavioral sciences provides an opportunity to achieve all three objectives.

The pilot results suggest using behavioral insights to design voice prompts can improve taxpayer experience. Using voice prompts to provide salient details of the benefits of using online tools enables taxpayers to opt-in to preferred service channels, thus saving both time and money. Taxpayers who called the IRS and heard redesigned messages were more likely to abandon their call and shift to online channels compared with callers who heard the existing messages. By informing taxpayers of the availability of relevant online alternatives, phone assistors were freed up to answer calls from taxpayers who prefer or require CSR assistance to resolve their tax issue.

---

1 The name of this discipline stems from the Behavioral Insights Team, an organization established in 2010 within the government of the United Kingdom to improve government policy and services and save money using behavioral nudges. The concept of using behavioral insights to improve performance has become pervasive throughout government. In 2015, a U.S. Executive Order encouraged all Federal departments and agencies to develop strategies for applying behavioral science insights to programs and, where possible, rigorously test and evaluate the impact of these insights.

2 CVP is a component of the IRS Unified Contact Center Enterprise (UCCE), an IP–based technology for call distribution and management, to support taxpayers and IRS partners who want to communicate with the IRS by phone. UCCE routes calls to applications which encompass a distinct product line or service and may have a dedicated call queue and group of trained CSRs.

3 Calls are routed to App 75 when the caller provides a TIN and UCCE evaluates the TIN for an ACS indicator. A call enters the CVP queue once the call is routed to the application and remains in the queue until 1) the caller hangs up, abandoning the call; 2) the CVP executes a courtesy disconnect due to extreme call volume; or 3) the caller connects with a CSR.

Findings from the study also point to opportunities to use behavioral methods to redesign message prompts on other IRS phone applications. Designing effective nudges requires an understanding of the reasons taxpayers may call the IRS to ensure voice prompts provide information most relevant to those circumstances. As such, this paper builds on findings from the CVP pilot and introduces an approach to attribute outcomes, such as taxpayer phone calls, to events. This approach can inform future opportunities to tailor call queue messages to caller profiles, thereby increasing the effectiveness of messages in encouraging callers to self-resolve tax issues through online service channels. As the IRS continues to expand online service offerings available to taxpayers, behavioral insights can be used to promote adoption by informing taxpayers of relevant tools and explaining how to use them.

The IRS uses Level of Service (LOS) to evaluate its ability to assist callers, measuring the proportion of calls routed and connected with a live assistor. LOS excludes calls routed to automated assistance and callers who hang up before connecting with an assistor. By this measure, answering as many calls as possible is the optimal outcome, while the value of issue resolution via self-service alternatives may not be considered. Using more comprehensive metrics could help the IRS evaluate its ability to provide "top quality service" to taxpayers. For example, CVP pilot results showed using voice message prompts to encourage channel shift can accelerate issue resolution, both by enabling callers to self-serve online and by freeing up live assistors to handle callers with more complex issues unable to be resolved online. This paper will discuss alternative measures to evaluate taxpayer experience beyond LOS as the IRS continues to expand online services available to taxpayers.

Methodology and Design

Lessons learned through prior notice redesign pilots informed the approach to test the impact of designing voice prompts to encourage callers routed to ACS App 75 to use self-service channels rather than remain on hold in the call queue.

Message Design

A growing body of research conducted by the IRS and other tax authorities demonstrates Behavioral Insights can be used to improve tax administration by nudging taxpayers towards desirable actions and away from undesirable actions. To develop improved versions of the CVP announcements, the IRS leveraged insights derived from previous notice redesign efforts and related behavioral research input from IRS stakeholders and additional research on the use of behavioral nudges to influence customer contact channels, including a 2019 study conducted by the United Kingdom's tax authority, Her Majesty's Revenue and Customs (HMRC).[4] The HMRC study is highly analogous to the CVP test as it used recorded messages to encourage callers who could self-serve to use online service tools by applying behavioral nudging techniques. HMRC increased channel shift rates by using behavioral insights to enhance high-traffic messages with the greatest potential for improvement.[5]

- **Be definite and clear** where possible, adding "if you can do this online, please hang up now."
- **Give precise instructions.** Provide taxpayers with a specific digital resource rather than a general resource like IRS.gov. Callers may have previously tried to self-serve online and may be less likely to do so again unless provided with new or more helpful information.
- **Shorten announcements.** Keep messages to 30 seconds or less and ensure each message contains no more than seven pieces of information.
- **Prime individuals for lists.** Use leading language to alert callers to a forthcoming list (e.g., "There are several kinds of income you will need to tell us about. These are…")

[4] Her Majesty's Revenue and Customs. Behaviour, Insight, and Research Team (BIR), 2019.

[5] The HMRC paper used the following behavioral insights techniques to improve announcements for callers: definite and clear language, precise instructions for completing tasks, shorter announcement length, list priming, and incentives for using self-service channels like online resources.

- **Provide incentives to move away from the phone**. Taxpayers who call are already tied to the telephone response channel. Highlighting incentives can nudge callers to try digital self-service tools.

Callers entering the ACS App 75 queue hear prerecorded announcements followed by hold music until the call is abandoned, disconnected, or routed to a CSR. Announcements provide general information regarding potential actions to resolve issues. Announcement themes include relevant information pertaining to making payments or payment plans online or provide guidance for preparing account documentation prior to connecting with a CSR. Should the caller wish to connect with a CSR, the CSR can assist the caller with the following: making full payments (by assisting taxpayers with online payment applications or sending a payment via mail); establishing payment plans; obtaining levy sources; reviewing liability disagreements; evaluating eligibility for Currently Not Collectible (CNC) determination; and resolving inquiries related to these issues.

The study team developed a sequence of five prototype announcements for App 75, which were evaluated in the pilot . The prototype messages address specific taxpayer concerns, highlight benefits of self-service tools, and acknowledge resource constraints associated with IRS phone resources. The redesigned messages aimed to nudge callers to abandon earlier in the sequence, freeing up space in the call queue for callers with more complex issues. To achieve this, the order of message themes in the sequence address issues expected to be most common among callers first. Table 1 summarizes themes employed by control and redesign announcement sequences. The first two announcements aim to nudge those calling to make a payment or establish or modify a payment plan to use IRS online tools by highlighting the salient benefits. The remaining messages in the sequence reiterate the availability of online services and remind callers to have their documentation ready if they intend to speak with a CSR.

TABLE 1. Control and Prototype Announcement Themes

Message #	Control Announcements	Redesigned Announcements
Message 1	Make sure you're prepared when your call is answered. Have details on hand related to the cause of balance due, your financial situation, and any unfiled returns	If you're calling to make a payment, online is your best option. IRS cannot accept payments over the phone, but there is a quick, easy, and secure option online
Message 2	Online options are available. Go to irs.gov/payments to explore a variety of online service options, such as accessing account information or making a payment	If you're interested in a payment plan, OPA is the best choice. Benefits of using OPA include reduced user fees for new and modified plans, instant confirmation, and the ability to explore a variety of plan options
Message 3	Go to IRS.gov and use the search feature to find services. There are many services available online that don't require waiting	Use Online Account to view up-to-date account information. Assure "comfort callers" the most current information about their account is accessible through OLA
Message 4	Payment plan options may be available if you can't pay now. Visit irs.gov/payments–you may be able to pay a portion of your balance or make payments with credit card	While you're waiting, check out online services. Acknowledge the wait time and suggest checking out new and improved features available online
Message 5	Check out safe and secure services on IRS.gov. You don't have to wait–you can go to IRS.gov to explore online services	If you choose to wait, make sure your information is ready. Recap online service options (before indefinite hold) and remind callers to have information ready

Test Methodology

To test the effectiveness of the redesigned message sequences, the pilot alternated the control and redesigned message sequences played to callers routed to the App 75 call queue. The pilot included callers entering the ACS App 75 queue by inputting a valid Taxpayer Identification Number (TIN) with an ACS indicator present on their account. Results were tracked for 30 days after the final pilot call.

Due to CVP system constraints, calls could not be randomly assigned to either redesigned or control message sequences.[6] As such, the test could not be implemented as a true randomized control trial. Control and redesigned message sequences were alternated each day during the six-week test period: July 12 through August 20, 2021. A comparison of characteristics of the two samples verified they were similar in all key respects, other than the treatment received.

To measure the effectiveness of redesigned CVP announcements, we evaluated caller behavior after reaching the App 75 call queue and compared outcomes for callers who heard control messages with callers who heard the redesigned messages. Metrics describing call outcomes (e.g., call abandon rate) were evaluated over the time interval in which the taxpayer was in the application. Metrics describing channel shift and online service access were evaluated over the 30 days following the caller entering the CVP queue. We compared metrics observed among the treatment group to those observed among the control group and tested the statistical significance of any differences. The following outcomes were evaluated across control and redesign callers in the pilot sample:

- Channel Shift: Callers who channel shift abandon their call before connecting with a CSR and access IRS online resources within 30 days of the call.
- Online Resource Access: Online resource access evaluates use of Online Account (OLA), Online Payment Agreement (OPA), or Get/View transcript applications occurring after a pilot call.
- Average Speed to Answer (ASA): Time callers spent in queue before connecting with a CSR. Increasing callers who channel shift should reduce the amount of time callers remaining in the queue must wait to connect with a CSR.
- Abandoned Calls: Proportion of calls abandoned while in the queue. Increasing rates of call abandonment among callers who can self-serve online will free up CSR capacity to assist callers with more complex tax issues. [7]

Sample Selection

The primary goal of the CVP pilot was to encourage callers able to self-service to abandon the call queue and shift to online service channels to resolve their tax issue. Therefore, the channel shift rate was the preferred metric for determining sample size requirements. Historic channel shift rates are not available for App 75 due to an inability to connect associated TIN-level data for abandoned call records. For this reason, we used historic channel shift rates from a prior pilot evaluating channel shift rates among LT11 notice recipients.[8] Individuals receiving this notice would be a subset of App 75 callers due to their account being in ACS. Because the LT11 taxpayer population received a notice of intent to levy, this population may be more apt to contact the IRS. For this reason, we used the channel shift rate provided by HMRC as a comparison. By reviewing variance in these two channel shift rates, we determined the sample sizes required to detect meaningful differences in channel shift rates across message groups. Achieving 90% power in identifying a 1% difference using the 1.59% channel shift rate reported in the HMRC study required a minimum sample of 3,500 calls. A more conservative approach to identify a 1% difference at 90% power using the channel shift rate from the LT11 population required a minimum sample of roughly 30,000 calls.

---

[6] Each call queue uses a single sequence of announcements at any given time, so all callers in the queue at the same time must hear the same announcements, meaning the control and treatment groups could not be tested concurrently.

[7] LT11 Notice Redesign Pilot Test (2019). Internal Revenue Service.

[8] Data retrieved from Intelligent Contact Management/Customer Voice Portal (ICM/CVP) platform. TIN-level data later acquired via UWR 2021-099 and merged with ICM data.

During the pilot, 307,837 calls were routed to ACS App 75.[9] About 63% of calls were exposed to at least one control or redesigned message, meaning the caller remained on the call long enough to hear at least one message in the announcement sequence before connecting with a CSR, receiving a courtesy disconnect (due to high call volume), or abandoning their call while waiting in queue. After exclusions, the resulting population consisted of 85,102 taxpayers and 103,512 calls.[10] Given the large sample size, the test was sufficiently powered (90%) to detect a 1% difference in the channel shift rate attributable to redesigned messages.

Pilot Analysis Groups

To analyze the results of the pilot, callers were assigned to one of three groups based on the number of calls made during the pilot test period:

- **Group 1** includes taxpayers who were routed to App 75, remained on the line to hear at least the first announcement in the message sequence, and did not call again within the pilot period.
- Group 2 includes individuals who called more than once during the pilot, but only during one call attempt were they on the line to hear at least one message in the sequence. All other call attempts were abandoned or disconnected prior to hearing the first announcement in the message sequence.
- **Grou**p 3 includes callers who heard at least one message in the sequence, called back at least once more and again heard at least one message in the sequence. Repeat callers may: have called back after a 2-hour courtesy disconnect; not have had time to remain on hold; be seeking assistance after attempting to use self-service options; be following up with a CSR; or be listening to queue messages again.

Table 2 summarizes the number of callers in each group who heard control and redesigned messages.

TABLE 2. Pilot Callers per Prototype by Analysis Group

Group	Message Sequence	Pilot Callers
Group 1	Control	30,580
	Redesign	31,146
Group 2	Control	3,606
	Redesign	3,437
Group 3	Control	7,884
	Redesign	8,449

Results and Discussion

Increase Channel Shift

The primary goal for the CVP pilot was to redesign voice prompt messages to nudge taxpayers able to self-serve to abandon their call while in the queue and shift to IRS online channels. The primary metric used to evaluate channel shift was the rate at which taxpayers abandoned their call after hearing at least one voice prompt and accessed an IRS online application to address their issue.[11] The channel shift rate considers a variety of actions that do not require CSR support and can be performed using IRS self-service tools, such as making a one-time payment, establishing or modifying a payment plan, confirming payment history, checking

[9] Table 25 lists exclusionary criteria and the number of calls and callers removed from the study for meeting one or more of the exclusionary criteria.

[10] See Table 27 in the Appendix for a list of exclusionary criteria, and the resulting number of calls and number of callers removed from the study.

[11] Channel shift actions include self-service payments, accessing OPA, requesting a return transcript, or accessing Online Account within 30 days of the abandoned call. Self-service payments are identified using the first and second positions of the EFT number associated with the payment transaction.

account balance, and viewing a transcript. Abandoning the call queue to accomplish any of these tasks using IRS online resources would be considered a successful outcome for the redesigned messages.

Callers exposed to at least one message in the redesigned sequence appeared more likely to channel shift than callers who heard the control messages. Table 3 shows the channel shift rate aggregated across the three pilot call groups. Across groups, redesigned messages increased the channel shift rate by about 13% relative to the control messages.

TABLE 3. Channel Shift Rate

Prototype	Channel Shift Rate	Relative Uplift (Percentage Change)
Control	12.51%	
Redesign	14.11%	+ 12.83% ***

***p-value < 0.001.

Table 4 summarizes differences in the channel shift rate for control and redesign callers by analysis group. Redesigned messages increased the channel shift rate by over 14% relative to the control messages for Group 1 callers and improved the channel shift rate by nearly 11% for Group 2 callers. Redesigned messages increased the channel shift rate by 16% relative to the existing messages for Group 3 callers.[12]

TABLE 4. Channel Shift Rate by Group

Group	Prototype	Channel Shift Rate	Relative Uplift (Percentage Change)
Group 1	Control	15.29%	
	Redesign	17.47%	+ 14.24%***
Group 2	Control	19.52%	
	Redesign	21.62%	+ 10.73%*
Group 3	Control	6.43%	
	Redesign	7.45%	+ 15.93%***

*p-value < 0.05; ***p-value < 0.001.

Table 5 shows the days between call and channel shift action for pilot callers who channel shift. While the channel shift rate considered self-service actions within 30 days of a call, between 80 and 85% of channel shifters complete their channel shift actions in the first 7 days following their call. Most callers who channel shift do so on the same day as their call. Across groups, a larger proportion of callers who heard the redesigned messages channel on the day of their call compared with callers who heard the existing messages. About 70% of Group 1 callers who heard redesigned messages channel shift on the same day as their pilot call, while 68% of callers who heard the control messages channel shifted on the same day as the call. About 65% of Group 2 callers who heard redesigned messages and channel shifted did so on the same day as the call, while nearly 62% of callers who heard control messages did so. Group 3 comprises repeat callers where outcomes are calculated at the call level.[13] Just over 64% of callers who heard the redesigned messages and channel shifted did so on the same day, whereas 59% of callers who channel shifted after hearing control messages did so on the same day as their call.

[12] This analysis assumes if a Group 3 caller called twice and channel shifted after each call, both actions would be included in the analysis. However, if they called twice and only channel shifted after the second call regardless of whether it occurred in the same 30-day window, this action would only be counted once.

TABLE 5. Days Between Call and Channel Shift

Days to Channel Shift	Group 1		Group 2		Group 3	
	Control	Redesign	Control	Redesign	Control	Redesign
Same Day	67.84%	70.43%	61.65%	65.41%	58.94%	64.32%
1–7 Days	15.27%	14.28%	19.18%	16.02%	24.12%	20.55%
8–30 Days	16.89%	15.29%	19.18%	18.57%	16.94%	15.14%

Actions taken by channel-shift callers include self-service payments, accessing IRS OLA, using OPA, and actions such as viewing a return transcript. Table 6 summarizes the first self-service action taken by callers who channel shift. OPA access appears to be the most common self-service outcome following call abandonment. About 31% of control and 38% of redesign callers who channel shift access OPA. The higher rate of OPA access among redesign callers can perhaps be attributed to the reminder of increased costs associated with establishing a payment plan over the phone.

TABLE 6. First Self-Service Action-Channel Shift Callers

Prototype	Channel Shift Callers	Self-Service Payment*	OPA Access	OLA	Get / View Transcript
Control	6,550	28.21%	31.21%	25.01%	15.57%
Redesign	7,645	28.82%	37.59%	20.30%	13.29%

* Self-service payments include Direct Pay, and other forms of electronic payment such ACH Debit, credit card, e-file debit. See Table 24 in the Appendix for a summary of Direct Pay application use by App 75 callers.

Individual Message Success

To examine the efficacy of individual messages in encouraging App 75 callers to channel shift, we evaluated outcomes for each group by last message heard prior to call abandonment.

Callers who abandoned the queue before the final message in the sequence most often did so after hearing the second message. Across groups, a larger proportion of callers who heard the redesigned messages abandoned after the second message compared with callers who heard the control messages. The second message in the redesign sequence highlights the salient benefits of OPA, such as reduced user fees for setting up a payment plan through OPA as opposed to over the phone. As shown in Table 7, a larger proportion of callers who heard the redesigned messages abandoned before the fifth message in the sequence compared with callers who heard the control messages.

TABLE 7. Callers Who Abandon in Queue–Distribution of Last Message Heard

Group	Prototype	Last Message Heard				
		1	2	3	4	5
Group 1	Control	4.29%	9.32%	5.08%	2.81%	78.50%
	Redesign	6.36%	13.94%	4.19%	5.14%	70.37%
Group 2	Control	5.25%	10.23%	5.93%	2.57%	76.02%
	Redesign	6.45%	14.37%	4.81%	5.26%	69.10%
Group 3	Control	3.66%	8.15%	4.58%	2.77%	80.84%
	Redesign	5.70%	12.09%	4.26%	4.97%	72.99%

Because announcements are designed to encourage callers to shift to online service channels, we expect taxpayers who abandon their call during the announcement sequence to exhibit higher rates of self-service than those who abandon during the indefinite hold. Table 8 shows the channel shift rate for taxpayers who abandoned their call by the last message in the sequence they heard. Among callers who abandon, callers who heard redesigned messages were more likely to channel shift after most of the announcements in the sequence.

The proportion of Group 3 callers who abandon and subsequently channel shift after hearing messages in the control sequence is lower than the channel shift rate for callers who heard redesigned messages for all but the fourth message in the sequence. Group 3 callers, who called multiple times during the pilot, may prefer to speak with a CSR and therefore may be less likely to channel shift, as shown below. Among Group 3 callers who channel shift after abandoning their call, it is possible callers may have called multiple times to listen to the voice prompts if information presented by the messages was missed initially.[14]

TABLE 8. Proportion of Abandon Callers Who Channel Shift by Last Message Heard

Group	Prototype	Last Message Heard				
		1	2	3	4	5
Group 1	Control	31.20%	36.19%	37.52%	40.80%	38.19%
	Redesign	45.41%	45.22%	39.06%	40.41%	40.26%
Group 2	Control	35.11%	40.44%	45.28%	42.48%	38.90%
	Redesign	48.25%	47.24%	48.24%	49.46%	39.39%
Group 3	Control	7.67%	9.67%	8.49%	10.89%	13.43%
	Redesign	12.73%	11.04%	11.59%	10.51%	15.13%

Increase Use of Online Services

Redesigned messages are aimed to increase callers' use of IRS online services, which include self-service payments, use of OPA for establishing or modifying payment plans, requests for prior return transcripts, and OLA access.[15]

Callers across groups appear more likely to access IRS online resources when exposed to the redesigned message compared with callers exposed to control messages. Table 9 shows the proportion of callers accessing IRS online services within the 30 days following their pilot call. Group 1 callers realized over an 8% improvement in the online service access rate relative to the existing messages. About 33% of Group 2 callers who heard redesigned messages accessed IRS online services in the 30 days following their phone call, compared with roughly 31% of callers who heard control messages. About 15% of callers in Group 3 who heard the control messages accessed IRS online resources compared with roughly 17% of callers who heard the redesigned messages. The redesigned messages achieved more than a 14% improvement in the rate of online service access relative to the existing messages. Results are significant for each group.

[14] If a Group 3 caller called twice within a 30-day period and channel shifted after each call, both channel shift actions are included in this analysis. However, if they called twice and only channel shifted after the second call, regardless of whether if occurred in the same 30-day window, this entry would only be counted once.

[15] The CVP announcement sequence is only played once before an indefinite hold with music. If a taxpayer misses a certain component of the announcement sequence, they will have to call again to hear the announcements again.

TABLE 9. IRS Online Service Access Rate[a]

Group	Prototype	IRS Online Service Access	Relative Uplift
Group 1	Control	29.24%	
	Redesign	31.65%	+ 8.25% ***
Group 2	Control	30.70%	
	Redesign	33.02%	+ 7.57%*
Group 3[b]	Control	14.63%	
	Redesign	16.74%	+ 14.45%***

*p-value < 0.05; ***p-value < 0.001.

[a] Online resources accessed within 30 days of call.

[b] For repeat callers in Group 3, 30-day outcomes are shown only for the call where the online action occurred most recently after. If a caller in Group 3 called twice and accessed online services after the second call, their action would be counted once and be associated with the outcomes of the second call. If they called another time and accessed online services again after the call, but still within the 30-day window of the first and second calls, the action would be associated with the outcomes of the third call for a total number of two access outcomes for the individual.

Most callers who use IRS resources following their call appear to access online services on the day of their call or within the first seven days. Table 10 shows callers who used IRS online tools after their pilot call typically accessed those resources on the same day as their call. About 73% of callers who heard redesigned messages and roughly 70% of callers who heard control messages used IRS online resources on the same day as the call. A smaller proportion of callers who accessed online services did so within seven days of their call.

TABLE 10. Days Between Call and First IRS Online Service Action

Days to Channel Shift	Control	Redesigned
Same Day	70.24%	73.28%
1 – 7 Days	13.46%	11.96%
8 – 30 Days	16.17%	14.65%

Many App 75 callers seeking to make a payment or establish or modify a payment plan have the option to use IRS online tools, such as OPA and IRS Direct Pay, to accomplish the task.[16] Individual taxpayers can establish or modify a payment plan, change the due date of their payments, change their bank account information, and change their monthly installment amount using OPA. Table 11 shows the OPA access rate for OPA-eligible pilot callers.[17] OPA access is defined as taxpayer entry into the OPA application. Pilot callers who heard the redesigned messages appear to access OPA at a higher rate than callers who heard the control messages. Among the three groups, Group 3 callers realized the largest improvement in the OPA access rate relative to callers hearing the control messages (42%). Callers who heard redesigned messages in Group 1 and Group 2 saw 26% and 21% improvements in the OPA access rate, respectively, relative to callers exposed to control messages. All group-level results are significant.

[16] All analyses in this section observe outcomes in the 30 days following each call.

[17] IRS Direct Pay is identified by EFT numbers with the first position in 2, the second position in 2 (ACH Debit), and the third position in 2 (IRS Debit); OPA Rate includes taxpayers who access the OPA app and had an associated IA or pending IA transaction.

TABLE 11. OPA Rate by Group

Group	Prototype	OPA-Eligible Callers	OPA Access	Relative Uplift
Group 1	Control	24,641	14.74%	
	Redesign	25,066	18.63%	+ 26.36%***
Group 2	Control	3,062	16.13%	
	Redesign	2,890	19.45%	+ 20.54%***
Group 3	Control	14,575	7.90%	
	Redesign	16,000	11.19%	+ 41.75%***

***p-value < 0.001.

Redesigned Messages Can Increase Taxpayer Savings by Encouraging Use of OPA. Taxpayers who use OPA to set up a payment plan or modify a payment plan rather than doing so over the phone with a CSR will save between $76 and $95 per payment plan. Over the course of the six-week pilot, 1,069 taxpayers who heard the redesigned messages and 684 taxpayers who heard the control messages abandoned their call and set up a payment plan through OPA. The second message in the redesigned sequence encourages callers interested in establishing or modifying a payment plan to use OPA, highlighting potential cost savings from lower user fees charged by OPA compared with an over the phone.

To estimate yearly taxpayer savings resulting from the redesign messages, we will only consider taxpayers who either abandoned or were disconnected, since the callers who connected would likely have received additional information from the CSR. Among callers who abandoned their call or disconnected, 395 additional callers in the redesign group established a payment plan in the month after their call. If each of these pilot taxpayers saved between $76–$95 per call, over the course of a month the redesign pilot population would have saved a total of $30,020–$37,525. If the redesigned messages were scaled to the entire App 75 population, we estimate taxpayers could save between $86,264–$107,830 in one month, or $1,035,168–$1,293,960 in one year.[18]

Table 12 shows the rate of self-service payments for callers who heard redesigned messages and callers who heard the control messages.[19] Across pilot groups, self-service payment rate was generally comparable for callers who heard redesigned and control messages. The redesigned messages appear to have had a positive effect on the self-service payment rate for callers in Group 2 who abandoned their calls. Among callers who connect with a CSR, self-service payments are still a possible outcome. The IRS cannot process payments over the phone, and therefore CSRs may guide callers interested in making a payment to do so through an IRS payment application.

[18] OPA allows individuals and businesses with an outstanding balance in aggregate assessed tax, penalties, and interest, to request a payment plan. Individual taxpayers are eligible to use OPA to full pay or set up a short-term plan if their outstanding balance is less than $100,000. To use OPA to set up an IA, the total balance must be less than $50,000.

[19] This analysis estimates taxpayer savings through OPA if the redesign message sequence were implemented on App 75. It assumes callers who provided their TIN (i.e., callers in the analysis group) behave similar to callers who did not provide their TIN, but this may not be the case. Callers in the analysis group were not a random subset of 68.7% of the population, but rather may include a self-selected subset choosing to provide their TIN. It is difficult to predict whether callers who did not provide a TIN would react to the redesign messages similarly as the TIN callers and, if not, the rate with which they differed. Some proportion of call records without TIN information may be a result of routing processes which limit the IRS's ability to associate TIN information with call records, rather than the taxpayer's decision to provide a TIN (e.g., if calls do not pass-through TIN Entry, TIN information is not captured in the Integrated Customer Contact Environment database).

TABLE 12. Self-Service Payment Rate by Group

Group	Call Outcome	Prototype	Payment Rate	Relative Uplift
Group 1	Connected	Control	13.91%	-
		Redesign	13.66%	- 1.81%
	Abandoned	Control	14.90%	-
		Redesign	15.13%	+ 1.55%
Group 2	Connected	Control	14.22%	-
		Redesign	14.30%	+ 0.56%
	Abandoned	Control	13.84%	-
		Redesign	16.38%	+ 18.33%*
Group 3a	Both	Control	6.68%	-
		Redesign	6.86%	+ 2.78%

*p-value < 0.05.

[a] Outcomes for Group 3 callers were not broken out by connected versus abandoned because it would be unclear whether the action could be attributed to the announcement sequence or direction from a CSR if they both connected and abandoned their calls.

Improve Call Resource Allocation

In Calendar Year (CY) 2019, more than 2.7 million phone calls reached App 75 and 51.2% connected with a CSR. Monthly call characteristics showed the abandon rate for App 75 callers generally fluctuated between 40% and 55%. The average abandon rate and ASA tend to be correlated, with longer wait times leading to higher abandon rates. [20]

The redesigned CVP announcement sequence was developed to nudge callers who could self-serve to resolve their issues online, reducing wait times for callers who require CSR support or cannot self-serve. The abandon rate measures the proportion of callers who abandon the call by hanging up after the start of the announcement sequence. Table 13 shows App 75 callers who heard the redesigned messages were more likely to abandon their call while in queue than callers who heard the control messages.

TABLE 13. Abandon Rate

Prototype	Abandon Rate	Relative Uplift
Control	44.76%	
Redesign	46.70%	+ 4.33%***

***p-value < 0.001.

Table 14 shows the abandon rate for each analysis group. Group 1 callers who heard the redesigned messages saw close to a 5% increase in the abandon rate compared with callers who heard the control messages. Group 2 and Group 3 callers who heard redesigned messages realized a nearly 4% increase in the abandon rate compared with callers who heard the control messages. Group 1 and Group 3 results suggest redesigned messages increased the rate of callers abandoning while in queue relative to the control messages.

[20] This analysis is restricted to taxpayers with an outstanding balance at the time of their call.

TABLE 14. Abandon Rate by Group

Group	Prototype	Abandon Rate	Relative Uplift
Group 1	Control	40.52%	
	Redesign	42.37%	+ 4.57%***
Group 2	Control	49.61%	
	Redesign	51.41%	+ 3.63%
Group 3	Control	50.93%	
	Redesign	52.76%	+ 3.59%***

***p-value < 0.001.

Table 15 shows the ASA for all App 75 callers as measured by average time in the call queue in minutes. Among callers who connected with a CSR, redesign callers spent, on average, roughly three fewer minutes in the call queue relative to callers who heard the control messages. Redesigned messages nudged callers able to self-service to abandon their call at a higher rate than the control messages, which helped free up space in the queue for callers with issues requiring CSR support.

TABLE 15. ASA—Connected Callers

Prototype	ASA (mm:ss)	Difference
Control	87:22	
Redesign	85:49	- 3:17***

***p-value < 0.001.

Table 16 shows the ASA for connected callers in each analysis group. Group 1 callers who heard redesigned messages waited, on average, two minutes less in the queue than callers who heard the control messages. Group 2 callers who heard redesigned message spent nearly five and a half minutes less, on average, in the queue than callers who heard control messages. For Group 3 callers, there was no significant difference in the amount of time spent in the queue for callers who heard redesigned or control messages.

TABLE 16. ASA by Group—Connected Callers

Group	Prototype	ASA (mm:ss)	Difference
Group 1	Control	88:00	
	Redesign	85:56	–2:04***
Group 2	Control	89:56	
	Redesign	84:31	–5:27***
Group 3	Control	85:34	
	Redesign	85:48	+ 0:14

***p-value < 0.001.

Redesigned Messages Reduce Queue Time for Callers Who Speak with a CSR. Callers who can resolve their tax issues online and choose to abandon their call sooner can reduce resource utilization of IRS call-systems and shorten the wait time for other callers to reach a CSR. To evaluate the redesigned messages' ability to improve call center resource allocation, the pilot tracked callers who abandoned prior to speaking with a CSR and observed the actions taken after the point of abandonment. In general, callers who heard at least one redesign message and abandoned their call, did so earlier in the message sequence than callers who heard the control messages. By the end of the five-message announcement sequence, a larger proportion of redesign callers had abandoned (e.g., roughly 30% for Group 1 callers) the queue compared with control callers (e.g., roughly 22% for Group 1 callers). Further, redesign callers generally spent less time waiting in queue compared with callers who heard control messages. Among callers who abandoned, callers who heard redesigned messages spent, on average, 3 to 5 fewer minutes waiting in the call queue compared to callers who heard the control messages. Callers in Groups 1 and 2 who heard the redesigned messages and remained on the line to connect with a CSR also experienced shorter wait times–roughly 2 to 5 fewer minutes than callers who heard the control messages.

Recommendations for Future Research

Deeper Understanding of Taxpayer Reasons for Calling Can Inform Improvements to Message Design

Understanding taxpayers' reasons for calling the IRS may inform further improvements to voice messages. Insight into taxpayers' possible motivations for choosing to wait in the queue to speak with a CSR can help identify how queue messages may be refined to assist taxpayers with specific issues by informing them of self-service resources most relevant to their circumstance or offer guidance for how to prepare information for speaking with a CSR while waiting on hold. We analyzed taxpayer journeys over the 30 days prior to their pilot call, and using event groups (e.g., notices issued, online authentication events, account transactions), identified specific events and actions within each group to calculate common pathways leading to a call.[21]

Evaluating notices issued to taxpayers within 30 days of their pilot call suggests the type of notice or number of notices issued could influence taxpayers' willingness to channel shift. Table 17 summarizes channel shift rates for callers sent one of the notices issued most frequently to taxpayers prior to their pilot call. Channel shift rates for pilot calls attributable to the CP14, the first notification of a balance due, were highest for both redesigned and control messages. Among those notice types issued most frequently to pilot taxpayers, the CP49, which notifies taxpayers their refund has been applied to pay an outstanding tax debt, saw lower channel shift rates for both control and redesigned messages. Taxpayers may opt to remain in the queue to connect with a CSR in response to notices or other circumstances which may not by addressed specifically by either the control or redesigned messages. Taxpayers who were sent multiple notices within 30 days prior to calling may prefer to wait to speak with a CSR –in particular, in instances where the notices issued appear to present conflicting information (e.g., notices with different balance due amounts).

[21] CY 2019 data retrieved from Enterprise Telephone Data Aspect Application Activity Report.

TABLE 17. Channel Shift Rate for Most Commonly Issued Notices Prior to Pilot Call

Notice Type	Prototype	Channel Shift Rate
CP504, Final/3rd Balance Due	Control	14.8%
	Redesign	15.6%
CP14, Balance Due	Control	16.5%
	Redesign	20.1%
CP90, Final Notice – Levy, Right to CDP Hearing	Control	13.8%
	Redesign	15.5%
LT11, Final Notice – Notice of Intent to Levy	Control	13.1%
	Redesign	16.7%
CP49 – Refund Applied to Other Tax Liability	Control	12.5%
	Redesign	14.1%
Multiple Notices	Control	14.1%
	Redesign	15.5%

Over 51,000, or 60% of pilot taxpayers were sent at least one notice in the 30 days prior to their pilot call. The CP504 was the most common notice issued to pilot taxpayers. More than 16,500 CP504 notices resulted in a pilot phone call. Some taxpayers were issued multiple notices in succession; for example, pilot taxpayers issued the CP504 had over 148 other distinct notice types issued, in addition to a CP504 within 30 days prior to call. About 6,300 pilot taxpayers issued a CP504 were sent a CP14 notice prior to the CP504, both within the same 30-day window prior to call. On average these CP504s were issued just 21 days after the CP14. The standard interval between Balance Due notices is at typically 35 days, this scenario may have motivated additional taxpayers to call with concerns or seeking clarification.

TABLE 18. Call Outcomes for Pilot Callers Issued More than 1 Notice 30 Days Prior to Call

# Notices Issued	Prototype	Call Outcome	
		Connected	Abandoned
Two Notices	Control	51.8%	47.0%
	Redesign	49.4%	48.7%
Three Notices	Control	55.6%	43.6%
	Redesign	54.2%	44.2%
Four or More Notices	Control	58.0%	40.2%
	Redesign	52.8%	45.1%

Over 20,000 pilot taxpayers were issued more than one notice within the 30 days prior to calling the IRS and 3,700 of these taxpayers received multiple notices within seven days of making a phone call. Taxpayers may be issued more than one notice of the same type if the notice presents information specific to a given tax year. For example, over 1,200 taxpayers were sent multiple CP71C notices (Annual Reminder of Balance Due) if they had outstanding tax debt for more than one prior tax year. Taxpayers who were issued multiple notices prior to calling appeared to show less tendency to abandon. Regardless of whether callers heard redesigned or control messages, the connected call rate increased by 3.8–4.8 percentage points for callers issued three notices compared with callers issued two notices in the 30 days prior. While concern or confusion stemming from conflicting information across notices issued in close proximity may not be feasible to address via information in a voice prompt, redesigning queue messages to encourage callers who can self-serve to use online services can reduce wait time for callers requiring CSR assistance. Further exploration of the effect of notices issued in close proximity on phone calls may help IRS to identify changes to underlying business processes which could improve taxpayer experience.

Call outcomes following specific events suggest taxpayers may be more inclined to stay in the queue if current messages do not adequately address the specific issue motivating the call. Table 19 summarizes events immediately preceding pilot calls effectively acknowledged by the redesigned messages which experienced higher abandon rates. Calls attributable to notices requesting payment, such as Balance Due or Annual Reminder notices, or delivery of Collection Due Process notice (determined by return receipt) saw higher abandon rates among callers who heard the redesigned messages. Redesigned messages highlighted the availability and benefit of self-service payment tools relative to continuing to wait on hold to speak with a CSR.

TABLE 19. Most Common Events Immediately Before Call[a] (Higher Abandon Rates with Redesign Messages)

Description	Prototype	Call Outcome	
		Connected	Abandoned
CP501/CP504, Subsequent Balance Due	Control	58.7%	40.5%
	Redesign	54.6%	44.1%
CP14 – Balance Due	Control	62.7%	36.4%
	Redesign	50.3%	48.5%
CP90/CP91 (Notice of Intent to Levy, Right to CDP Hearing)	Control	42.0%	53.9%
	Redesign	37.1%	57.6%
Additional Tax Assessment	Control	61.3%	38.1%
	Redesign	56.5%	42.1%
Payment	Control	56.2%	42.2%
	Redesign	53.4%	44.9%
Certified Mail Return Receipt Signed	Control	53.5%	44.1%
	Redesign	47.8%	50.2%
CP71C/LT39, Annual Balance Due Reminder	Control	56.0%	41.7%
	Redesign	51.0%	47.6%

[a] Table 20 shows outcomes for Group 1 callers only.

Table 20 summarizes events immediately preceding pilot calls where the abandon rate is comparable for callers who heard the control and redesigned messages. The circumstances surrounding some of these events may not be acknowledged specifically by either the existing or redesigned messages. For example, taxpayers who accessed IRS online tools (e.g., OPA) and proceeded to call the IRS, remained on hold to connect with a CSR more often than they abandoned their call in the queue. Most taxpayers called within five days after going online. Taxpayers who attempt to self-serve using online tools but are unable to resolve their issue may call the IRS and wait in the queue to connect with a CSR for assistance. Further exploration of challenges encountered by taxpayers attempting to use online tools could help identify opportunities to improve user experience and help users in resolving issues on the first attempt.

Some circumstances or events preceding a phone call may be best addressed by speaking with a CSR. For example, issues related to changes to the Advanced Child Credit for Tax Year 2021 as part of the American Rescue Plan Act or a Bureau of Fiscal Service (BFS) Levy implemented with the Federal Payment Levy Program may require CSR guidance to resolve. These circumstances may not be practical to address via call queue voice messages due to limited potential for self-service resolution.

TABLE 20. Most Common Events Immediately Before Call[a]–Comparable Abandon Rates for Redesigned and Control Messages

Description	Prototype	Call Outcome	
		Connected	Abandoned
Advanced Child Credit	Control	57.6%	42.0%
	Redesign	59.2%	39.1%
IRS Online Service Access (Get/View Transcript)	Control	58.5%	40.2%
	Redesign	58.5%	40.0%
OLA Access	Control	66.1%	32.7%
	Redesign	66.0%	32.5%
OPA Access	Control	68.9%	30.3%
	Redesign	66.8%	32.5%
BFS Levy Program	Control	58.2%	40.4%
	Redesign	56.6%	41.7%

[a] Table 21 shows outcomes for Group 1 callers only.

Future research may explore analyzing call transcripts and pursuing opportunities to collect information from CSRs about the nature of handled calls. Research in these areas could help provide deeper understanding of taxpayer motivations for calling the IRS and common issues taxpayers may be facing when they choose to stay and connect with a CSR. Future research efforts may also explore the construction of taxpayer profiles via event clusters and applications of attribution modeling, These profiles could determine the events driving calls and demand for CSR support, which can inform strategies for further improving taxpayer interactions with the IRS.

Level of Service Measures May Understate the Caller Experience

The IRS uses LOS to evaluate its ability to answer taxpayer questions and assist taxpayers in meeting their tax obligations.[22] LOS is a budget-level measure required by the Congressional Budget Justification and Annual Performance Report and Plan. It is defined as the success rate of taxpayers calling the Accounts Management (AM) office of the IRS in connecting with a CSR. While inbound calls to AM applications inform the LOS estimate and calls to ACS applications may not directly, consideration should be given to the potential effect of applying lessons learned through redesigning App 75 messages to improve the existing AM App 10 messages.[23] The LOS formula is shown below: [24]

$$LOS = \frac{(CSR\ Answered + Automated\ Answered)}{(CSR\ Answered + Automated\ Answered + Abandoned\ + Busy\ + Disconnected)}$$

LOS alone may be limited in its ability to evaluate taxpayer access to assistance from the IRS. LOS measures the proportion of calls answered, but may not fully capture other aspects of the customer experience. For example, a taxpayer may connect with a CSR, but LOS does not capture whether the taxpayer resolved their

[22] This analysis did not consider or evaluate demographic information (e.g., age, income level) which may have affected a callers' ability to channel shift.

[23] National Taxpayer Advocate. (2018). *Measuring the Taxpayer Experience—The IRS Level of Service Measure Fails to Adequately Show the Experience of Taxpayers Seeking Assistance Over the Phone*. NTA Blog.

[24] App 10 is an AM application similar to ACS App 75. Callers routed to App 10 are individual taxpayers with a balance due.

issue during the call or if additional contact with the IRS was necessary. As the Taxpayer Advocate Service (TAS) notes, "[a]chieving a high LOS does not mean much if the IRS is unable to answer taxpayers' questions over the phone or guide them to an appropriate resolution of their issues." Further, an additional aspect of customer experience not captured by LOS is wait time. TAS found time spent waiting on the phone was a primary factor in customers' satisfaction with telephone service.[25]

Channel shift measures the rate at which callers abandon the call queue and shift to use online self-service tools to resolve their tax issue. According to the LOS formula shown above, an increase in the number of channel shift callers would increase the number of abandoned calls in the denominator, resulting in a lower LOS. In this sense, LOS would not capture the benefit to taxpayers who choose to abandon the call queue to use IRS online tools to self-serve.

Using a combination of metrics may offer a more comprehensive assessment of level of LOS provided to taxpayers and effort required to resolve taxpayer issues. Additional metrics which could be considered to evaluate service include:

ASA quantifies the amount of time callers spend waiting to connect with a CSR. As shown by CVP pilot results, ASA was shorter for callers in the redesign group due primarily to the higher rate of callers deciding to channel shift and abandon their calls to use self-service tools. Those who were able to self-serve abandoned and did so online, reducing the average time spent in queue for callers who were either unable to or preferred to speak with a CSR. As mentioned in the Taxpayer Advocate Service 2021 Annual Report to Congress, time spent waiting to connect with a CSR is a primary factor in satisfaction with telephone service. Strategies which help to reduce wait times can improve taxpayers' phone experience and should contribute positively to the IRS's service performance measure.

Level of Access (LOA) measures the proportion of calls received during business hours which were connected with a CSR. In response to observed LOS limitations, Treasury Inspector General for Tax Administration (TIGTA) and TAS proposed alternative success metrics. TIGTA noted the Social Security Administration (SSA) and tax agencies in several states use LOA and describe it as a more accurate measure of callers who receive assistance from IRS. IRS agreed to add LOA as a supplementary metric for evaluating phone line performance. The formula for LOA is shown below.

$$LOA = \frac{(CSR\ Answered + Automated\ Answered)}{(CSR\ Answered + Automated\ Answered + Abandoned\ + Busy\ + Disconnected) - x}$$

LOA is similar to LOS but excludes calls made to IRS outside of business hours () from the denominator. LOA does not capture taxpayer satisfaction from self-servicing online. Like LOS, an increase in abandon rates due to callers channel shifting would likely have a negative effect on this customer service measure.

First Contact Resolution (FCR) measures the proportion of taxpayer engagements successfully resolving a taxpayer issue and resulting in no follow-up high-touch engagements, e.g., phone calls or Taxpayer Assistance Center (TAC) visits. FCR represents the rate of calls resolved on the first attempt without the need for the customer to be transferred or called back. FCR was strongly tied to customer satisfaction in a 2020 survey of customer satisfaction on the AM line.[26] The formula for FCR is shown below.

$$FCR = \frac{\#\ Taxpayers\ with\ Issue\ Resolved\ on\ First\ Call}{Total\ Connected\ Calls}$$

FCR quantifies the proportion of callers who contact the IRS for assistance and do not contact the IRS again following their initial contact. Evaluating FCR for inbound calls assesses whether callers who connect with a CSR successfully resolve their issues on the first attempt. FCR, as defined above, does not consider taxpayer interaction with IRS through other channels, such as TAC visits or online engagement.

[25] National Taxpayer Advocate. (2022). *Annual Report to Congress*. National Taxpayer Advocate. 36-37.

[26] National Taxpayer Advocate. (2021). *Annual Report to Congress*. National Taxpayer Advocate. 66-80.

The IRS Customer Experience Visualization Tool provides performance metrics, such as the First Touch Resolution (FTR). FTR is similar to FCR but includes TAC visits in the equation. FTR quantifies the proportion of taxpayers who call the IRS or visit a TAC and do not call or visit a TAC again in the 90 days after their engagement. The equation for FTR is shown below.

$$FTR = \frac{\#\ of\ Calls\ or\ TAC\ Visits\ with\ no\ Subsequent\ Calls\ or\ Visits}{Total\ \#\ of\ Calls\ and\ TAC\ Visits}$$

Like FCR, FTR does not consider taxpayers who abandon phone calls to use self-service tools and do not follow up with IRS in the subsequent 90 days. FTR or FCR, as currently defined, would not be affected by improvements in the channel shift rate achievable through efforts such as the CVP message redesign pilot. FTR may be a useful indicator to monitor the efficacy of service provided to taxpayers who pursue high-touch engagements with IRS, such as phone calls or TAC visits.

Taxpayer Effort (TE) applies weights to taxpayer interactions (i.e., online resource access, phone calls, TAC visits, TAS engagements, and inbound correspondence), to estimate TE exerted in issue resolution. To capture taxpayer interactions in a digital age, IRS must leverage available data to track taxpayer interactions via online resources. The equation below illustrates how weights are applied to IRS interactions to estimate TE .[27]

$$TE = (1 * OLS) + (2 * Correspondence) + (3 * Connected\ Calls) + (4 * TAC) + (4 * TAS)$$

Table 21 shows estimated TE for pilot callers in the 30 days following their first pilot call. Callers in Groups 1 and 3 who heard redesigned messages showed slightly lower estimated TE compared with callers who heard control messages. Callers in the redesign group abandoned calls in favor of self-service channels a higher rate than callers in the control group, likely contributing to the lower effort estimated.

TABLE 21. Estimated TE within 30 Days of First Call

Group	Prototype	Average TE	Relative Uplift
Group 1	Control	3.35	
	Redesign	3.27	-2.34%**
Group 2	Control	2.19	
	Redesign	2.19	-0.10%
Group 3	Control	3.85	
	Redesign	3.67	-4.62%**

**p-value < 0.01.

Effort to Serve (ETS) measures the effort required by IRS to resolve taxpayer issues. This metric applies weights to inbound mail, inbound phone calls connected with a CSR, and TAC visits to estimate level of effort. The formula for ETS is shown below.

$$ETS = (17 * Mail) + (41 * Connected\ Calls) + (67 * TAC)$$

Table 22 shows the estimated ETS applied to CVP calls in the 30 days following the first pilot call. The average ETS for callers in Groups 1 and 3 who heard redesigned messages appears slightly lower than the estimated average ETS for callers who heard the control messages. Redesign callers were more likely to abandon calls

[27] IRS, W&I, Accounts Management Toll-Free Customer Satisfaction Survey FY 2020 Semiannual Report 8-9 (July 30, 2020). 37% of respondents reported “other” to improve their experience and 35% said that resolving their issue would improve their experience.

and self-serve online, likely contributing to the lesser average ETS compared with callers who heard control messages.

Table 22 shows the estimated ETS applied to CVP calls in the 30 days following the first pilot call. The average ETS for callers in Groups 1 and 3 who heard redesigned messages appears slightly lower than the estimated average ETS for callers who heard the control messages. Redesign callers were more likely to abandon calls and self-serve online, likely contributing to the lesser average ETS compared with callers who heard control messages.

TABLE 22. Estimated ETS Within 30 Days of First Call

Group	Prototype	Average ETS	Relative Uplift
Group 1	Control	30.8	
	Redesign	29.7	-3.42%***
Group 2	Control	17.9	
	Redesign	17.3	-3.44%
Group 3	Control	38.3	
	Redesign	36.3	-5.26%***

***p-value < 0.001.

Measures such as ASA, LOA, FCR, TE, and ETS provide additional insight into IRS service performance and effort required to resolve taxpayer issues. As discussed in this paper, abandoned calls may not always signal shortcomings in service (i.e., callers who abandon the call queue to shift to use self-service tools) and connected call rates may not fully reflect taxpayers' phone experience (e.g., amount of wait time or number of touches required for resolution). Choosing to abandon the call queue and shift to self-service tools can save taxpayers time otherwise spent waiting in queue and in some cases, can save money due to lower user fees associated with self-service platforms like OPA. Improving awareness of digital resources and self-service tools relevant to specific situations enables taxpayers to elect the option best suited to their preferences and needs in resolving outstanding tax issues. The CVP pilot showed the potential benefit of incorporating behavioral nudges to enhance call queue messages and increase use of web-based channels for specific needs, helping reduce demand on phones and reduce wait time for taxpayers who need to speak with a CSR. Looking forward, using a combination of service indicator measures can offer the IRS a more comprehensive view of the level and quality of service provided to taxpayers. Further, a comprehensive set of service metrics could help capture the impact of IRS efforts to apply behavioral research to improve taxpayer experience.

Appendix

TABLE 23. Acronym List

Acronym	Definition
ACS	Automated Collection System
AMS	Account Management Services
ARDI	Account Receivable Dollar Inventory
ASA	Average Speed to Answer
ATTS	Automated Time Tracking System
BMF	Business Master File
BOD	Business Operating Division
CDP	Collection Due Process
CDW	Compliance Data Warehouse
CNC	Currently Not Collectable
CSR	Customer Service Representative
CVP	Customer Voice Portal
DNIS	Dialed Number Identification Service
EIN	Employer Identification Number
ETS	Effort to Serve
FCR	First Contact Resolution
FTR	First Touch Resolution
HMRC	Her Majesty's Revenue and Customs
ICCE	Integrated Customer Contact Environment
ICM	Indirect Channel Management
IMF	Individual Master File
IRS	Internal Revenue Service
IRTF	Individual Returns Transaction File
IVR	Interactive Voice Response
OIC	Offer in Compromise
OLA	Online Account
OLS	IRS Office of Online Services
OPA	Online Payment Application
PSA	Public Service Announcement
RAAS	Research, Applied Analytics, and Statistics
RCT	Randomized Control Trial
SSA	Social Security Administration
TAC	Taxpayer Assistance Center
TBRM	Topic Based Routing Menu
TDA	Taxpayer Delinquent Account
TE	Taxpayer Effort
TERC	Total Enforcement Revenue Collected
TIGTA	Treasury Inspector General for Tax Administration
TIN	Taxpayer Identification Number
TRIS	Telephone Routing Interactive System
TTS	Text to Speech
UCCE	Unified Contact Center Enterprise
URL	Uniform Resource Locator
UWR	Unified Work Request

To provide clear guidance and nudge taxpayers to use online resources where possible, the IRS redesigned the ACS App 75 call queue announcements. Table 24 describes Behavioral Insights techniques from the HMRC study and prior notice redesigns used in redesigning ACS App 75 announcements.

TABLE 24. Techniques Used in ACS App 75 Announcement Redesigns

Audience Awareness	• Clearly define the target caller for each announcement and only provide relevant information. Focus on addressing common inquires and a single high-level topic. • Consider callers' perspective, prior journey, and potential frustration, including the inertia and sunk cost associated with reaching the queue.
Clarity and Simplicity	• Simplify language using plain, unambiguous English and an active voice. • Provide no more than seven pieces of information and make them count. • Give clear, brief next steps with precise instructions. • Prime people for lists by hinting at their impending presentation. • Harmonize language and keyword usage with the proposed self-service website.
Behavioral Nudges	• Give callers an incentive (e.g., loss aversion) to go online. Make incentive announcements truthful and helpful, not promotional or assertive. • Provide reassurance the self-service tool provides achievable benefits, including averting losses. • Emphasize and positively frame the self-service's immediacy, convenience, simplicity, completeness, and security. • Appeal to social norms with a focus on descriptive norms and the minority frame. • Encourage callers who can self-serve to go online while they remain on hold for a CSR.
Attention Management	• Limit announcement length to 30 seconds with sequenced announcements separated by 30 second intervals and a brief pause prior to the start of each announcement. • Repeat the core announcement and action (i.e., URL) at the end of the announcement while avoiding verbatim repetition. • Craft announcements in a conversational style and consider beginning with a question. • Do not repeat the same announcement on the same call. • Employ a clear, friendly, upbeat, and slowly paced human voice.

Redesigned and Control Message Announcements

TABLE 25. Redesigned App 75 Messages

Announcement	Prototype Announcement Language
#8548	Are you calling to make a one-time payment? We cannot process payments over the phone. Visit irs.gov/payments to securely make a payment from your bank account, credit card, or debit card. Explore payment plan options if you cannot pay your balance in full. Again, visit irs.gov/payments.
#8549	Are you calling to set up a payment plan? The fee to set up a plan over the phone can be as much as $225, compared to just $31 if you use the Online Payment Agreement tool at irs.gov/OPA. If owe less than $50,000 you can save time and money by visiting irs.gov/OPA to apply for a payment plan. You can choose from a variety of plan options and get instant confirmation if you qualify. Are you calling to revise an existing payment plan? You can do that online for less too. Avoid the wait to speak with a representative and save money by hanging up and visiting irs.gov/OPA.
#8550	Are you calling to check on the status of a payment or with a question about your account? Most individual taxpayers can sign up to view their account information online at irs.gov/account. You can securely check your account balance, view payment history and any scheduled or pending payments, and access tax records by registering at irs.gov/account.
#8551	At this time of year, the average wait to speak with a representative is about 45 minutes. Most taxpayers find it is more convenient to use our online self-service options. Online services provide step-by-step instructions to securely make a payment, set up a payment plan, or check your account status. Our online services frequently feature a chat option to receive live assistance from one of our representatives. Avoid the wait and visit irs.gov/payments and choose the option right for you. Again, that's irs.gov/payments.
#8552	Are you experiencing financial hardship and calling to request a temporary delay in collection until your financial situation improves? If so, have your income and expense information available to share with the customer service agent. We'd like to help resolve your tax issue today. Be ready to discuss the cause for your balance due—or, if you have unfiled returns, the date you expect to file, and the amount due. While waiting, consider visiting irs.gov/individuals to explore the new and improved online services.

TABLE 26. App 75 Control Messages

Announcement	Existing Message Language
#8065	So that we can better assist you today, please be ready to discuss the cause for the balance due. If you are unable to pay now, you may be asked to provide income and expense information to determine your ability to pay. If you have unfiled returns, provide the date you expect to file the return and the amount due.
#8058	We are experiencing high call volumes. You can complete many account actions on-line using www.irs.gov/payments. IRS offers many safe and secure on-line services such as arranging for payments or getting your account information and much more. Sign-up using our two-step authentication process to gain access to your account information, including account balances, payment history, transcripts, and other information. You can make payments online and establish an installment agreement. Consider hanging up now and going to www.irs.gov/payments to explore on-line options available to you.
#8062	We are experiencing lengthy wait times. You don't have to wait! You can arrange for payments, get records of your account by going to www.irs.gov and select the payment button or look under the tool section for payments and many other services offered to you online without waiting!
#8059	Can't pay now? Did you know that you may be able to pay all or a portion of your balance or make monthly payments by using your credit card? Please visit us at www.irs.gov/payments.
#8064	Tired of waiting? You don't have to! IRS offers many safe and secure on-line services such as arranging for payments or getting your account information and much more. Go to www.irs.gov and check it out!

Exclusionary Criteria

307,837 calls were routed to App 75 during the test period. For a variety of reasons, some calls or callers were excluded from subsequent analysis, as shown in Table 27. The most common exclusionary condition was the absence of Taxpayer Identification Number (TIN) information to associate with the call. A total of 96,304 calls were dropped from the pilot sample due to the absence of TIN information.

TABLE 27. CVP Redesign Pilot Exclusionary Criteria

Exclusion	Rationale	App 75 Count
Calls without TIN information *	Calls without TIN information include: 1) calls not routed through the Interactive Voice Response (IVR), and 2) calls routed through IVR where the caller did not input their TIN and abandoned before connecting with a CSR. Without TIN Information, it is not possible to measure outcomes associated with control or redesigned messages.	96,304 Calls
No prototype message exposure	If callers routed to Apps 75 did not hear any announcement in the sequence because they abandoned or were disconnected, it is not possible to measure the effect of messages of observed outcomes.	11,456 Callers 13,062 Calls
Callers exposed to different message prototypes	If taxpayers called on multiple days and were exposed to both announcement message sequences, it is not possible to determine which prototype influenced outcomes.	21,796 Callers 78,479 Calls
Exposure to both Apps 75 and 85	Callers who called multiple times and enter both applications' call queues may experience differing announcements given App 85 is intended for BMF entities.	309 Callers 786 Calls
BMF Taxpayer in App 75	App 75 is intended for IMF callers in Collection. Any BMF callers routed to App 75 are inconsistent with the intent of the pilot.	1 Caller 1 Call

* Without TIN information, it is not possible to determine the number of unique callers associated with these call records. Therefore, only a count of pilot calls dropped from the sample is provided for each App 75 and 85.

After dropping calls without TIN information, the pilot sample comprised 211,515 calls associated with 147,481 taxpayers. However, an additional 33,562 App 75 callers were dropped from the test sample for meeting at least one of the other four exclusionary criteria summarized in the table above. Table 28 summarizes the number of callers excluded from subsequent analysis by message sequence prototype for each application.

TABLE 28. Total Exclusions by Message Prototype

Message	Total Callers
Control	16,485
Redesign	17,077

Note: 21,796 callers were exposed to both the control and redesign prototypes. These callers are shown in the table based on the first version of the announcements sequence they were exposed to.

The Balance Due Taxpayer: How Do We Reduce IRS Cost and Taxpayer Burden for Resolving Balance Due Accounts?

Javier Framinan, Frank Greco, Shannon Murphy, and Howard Rasey (IRS Wage & Investment Strategies and Solutions), Javier Alvarez and Angela Colona (IRS Taxpayer Experience Office)

Introduction

Taxpayers with a balance due tax return create a significant cost to the IRS, especially when it must issue a CP14 notice in efforts to collect unpaid amounts. The IRS sends CP14 notices to taxpayers who have a balance due but do not fully pay by the filing deadline. Using Tax Year (TY) 2019 data, Wage and Investment Strategies and Solutions (WISS) estimates IRS's total monetary cost for issuance of CP14 notices is $90.8 million. This comes out to approximately $12.10 per CP14 from issuance to resolution. These numbers provide the basis of potential cost savings discussed throughout this report. Table 1 gives a breakdown of the costs associated with resolving balance due accounts. The calculation is likely an underestimate, as it considers only the direct costs of issuance and initial resolution of a CP14 and leaves out the costs associated with Taxpayer Assistance Center visits and additional correspondence to the IRS. Furthermore, not all balance due taxpayers receive a CP14 notice. But there are still other costs such as those associated with installment agreements for balance due taxpayers who cannot pay their balance in full by the filing deadline.

TABLE 1. Estimated Annual Cost of Issuing CP14 Notices

	Count (Millions)	Cost (Each)	Cost (Total)
Total cost to mail a CP14 notice	7.5	$.51	$3,825,000
CP14 Outcomes			
Full pay	1.2	-	-
Installment agreement	3.3	$6.12	$20,196,000
Ignore (receive CP501)	2.6	$0.51	$1,326,000
Call	0.9	$72.73	$65,457,000
Grand Total Cost			$90,804,000

Note: We excluded Taxpayer Assistance Center ($251.38 per visit) and written correspondence ($95.47 per response) due to data limitations.

The first purpose of this study is to identify taxpayers most vulnerable to an unintentional shift into a balance due position with the goal of balance due prevention. Taxpayers can shift to this unfavorable[1] position because of personal characteristics or intentional changes they make to their withholding and exemption selections. But in many instances, the shift can occur unintentionally as the result of a significant life event. Such life events make these shifts largely predictable, but only if the taxpayer recognizes the tax implications of the event. The second purpose of the present study is to identify the information currently available, or lack thereof (i.e., the messaging gap), to help taxpayers avoid unintentional shifts to balance due. We conducted analyses on both to help IRS's newly formed Taxpayer Experience Office (TXO) design targeted interventions that inform potential balance due taxpayers, save IRS resources (i.e., through fewer notices, fewer phone calls, less collection activity), and reduce taxpayer burden. We divide this report into three correspondingly enumerated sections.

[1] Balance due accounts are disadvantageous for the IRS, and we assume taxpayers prefer to receive a refund, break even, or have a balance due they can pay by the filing deadline.

1. Taxpayers Most Vulnerable to Balance Due

Method

Our outcome variable is a categorical measure of change in balance due. As shown in Table 2, we grouped taxpayers according to change in balance due from TY 2016 to TY 2017. We provide counts of taxpayer returns whose balance due did not change, shifted in a favorable direction, or shifted in an unfavorable direction. For example, we group a taxpayer who received a refund in TY 2016 and then had a balance due with a CP14 in TY 2017 in the unfavorable shift category.

TABLE 2. Changes in Balance Due from Tax Year 2016 to Tax Year 2017

	Volume of Returns	Percent of Total
No Change from Tax Year 2016 to Tax Year 2017	**110,688,283**	**82.6%**
Refund or even → Refund or even	97,511,566	72.8%
Balance due without CP14 → Balance due without CP14	10,904,331	8.1%
Balance due with CP14 → Balance due with CP14	2,272,386	1.7%
Favorable Shift from Tax Year 2016 to Tax Year 2017	**10,717,643**	**8.0%**
Balance due without CP14 → Refund or even	7,618,951	5.7%
Balance due with CP14 → Refund or even	1,922,187	1.4%
Balance due with CP14 → Balance due without CP14	1,176,505	0.9%
Unfavorable Shift from Tax Year 2016 to Tax Year 2017	**12,539,414**	**9.4%**
Refund or even → Balance due without CP14	9,132,071	6.8%
Refund or even → Balance due with CP14	2,221,312	1.7%
Balance due without CP14 → Balance due with CP14	1,186,031	0.9%
Total	**133,945,340**	**100.0%**

We used older TY 2016 and TY 2017 data for three reasons. The 2018 Tax Cuts and Jobs Act (TCJA) increased the standard deduction for taxpayers, which resulted in a dramatic reduction in Schedule A filings as shown in Figure 1. We used data prior to the implementation of this act because it is more reflective of tax return volumes with Schedule A the IRS can expect in 2026 once the changes implemented by the TCJA expire. Additionally, the 2020 COVID pandemic likely led to a temporary reduction in Schedule C filings. As shown in Figure 2, prior to the pandemic Schedule C filings had been steadily increasing. We assume the upward trend will resume moving forward. Finally, the pandemic also had an impact on marriage and divorce trends. A 2022 study[2] showed while marriage and divorce rates had been in decline prior to the pandemic, the pandemic magnified the trend to a steeper decline. As with Schedule C filings, we assume the trends in marriage and divorce to return to pre-pandemic levels.

[2] Westrick-Payne, K.K., Manning, W.D., & Carlson, L. (2022). Pandemic Shortfall in Marriages and Divorces in the United States. Socius, 8. https://doi.org/10.1177/23780231221090192

FIGURE 1. Schedule A Volume by Tax Year

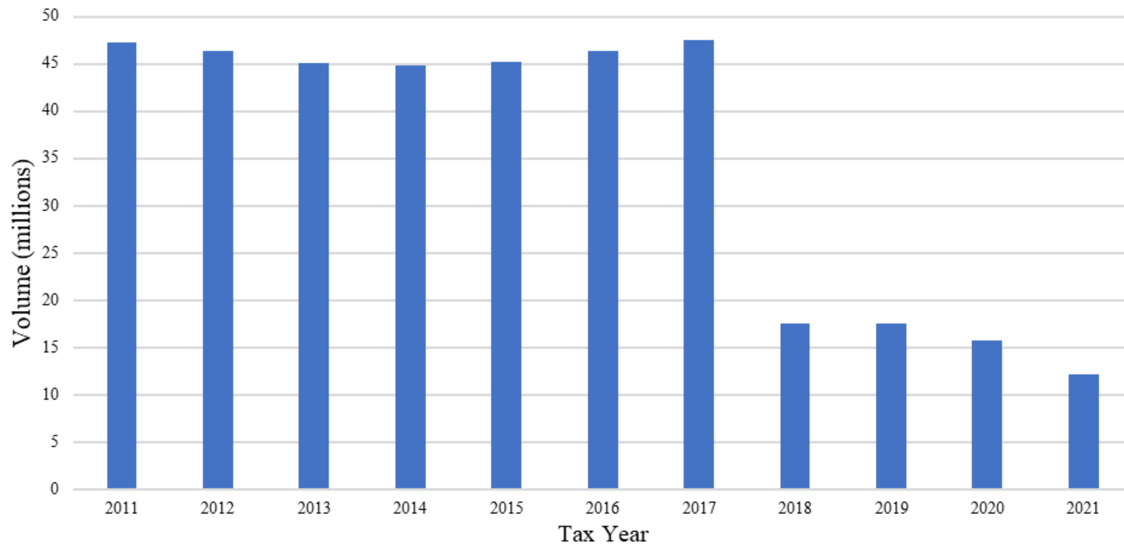
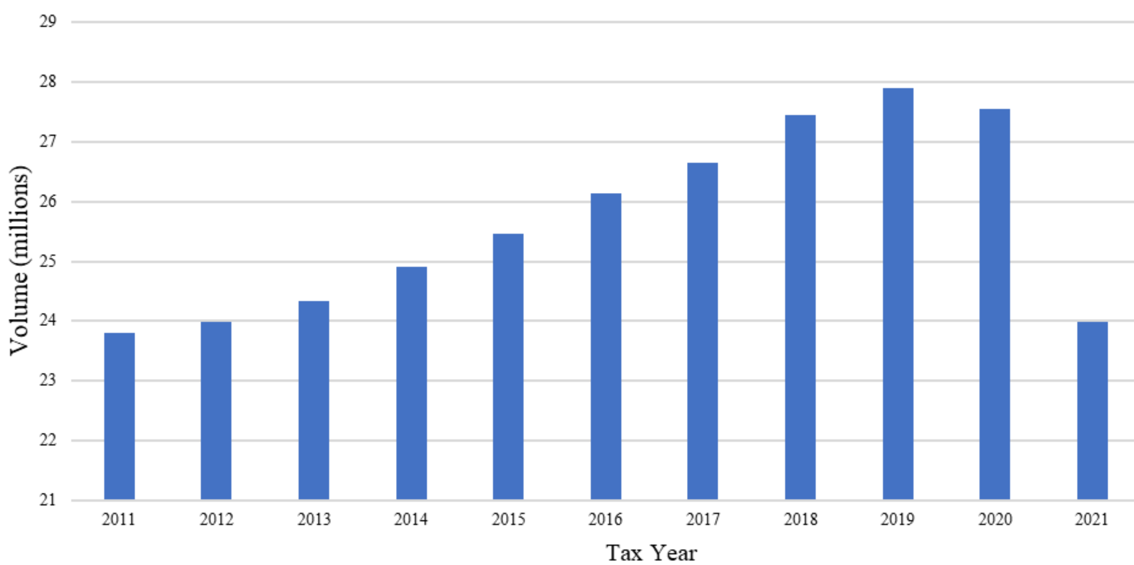


FIGURE 2. Schedule C Volume by Tax Year



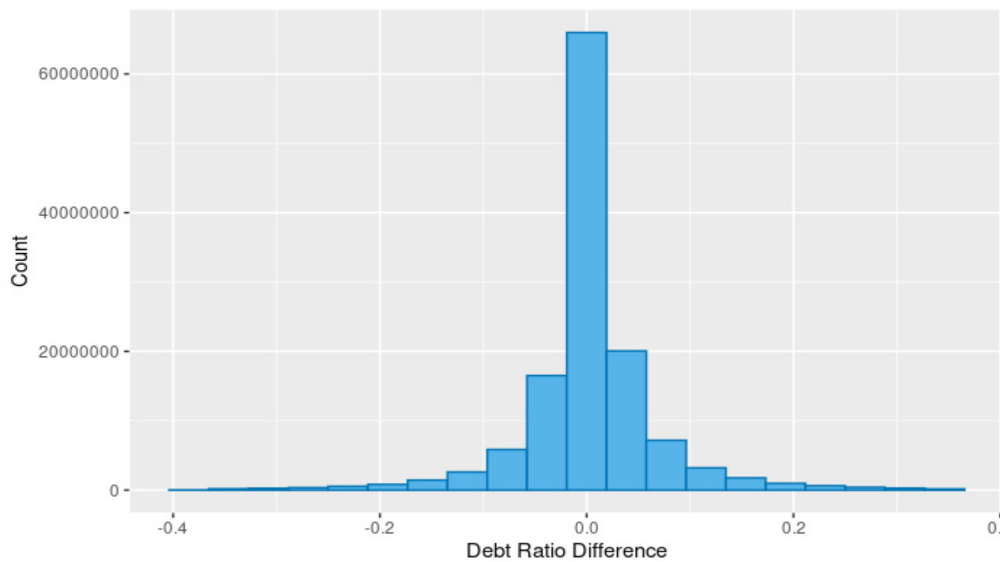
We first conducted a series of descriptive statistics and chi square analyses to identify key 1040 characteristics associated with unfavorable balance due changes. We then conducted a series of logistic regressions[3] to determine the change in predicted probability of a taxpayer shift into unfavorable balance due given changes in their personal characteristics (i.e., a change in filing status) or a change in their tax return (i.e., attaching or removing a schedule). Predicted probabilities range from 0 (impossible) to 1 (happens with certainty). Since we focused on unfavorable shifts into balance due categories, moving forward we refer to an increase in predicted probability as "risk." Additionally, we use the simpler term "balance due" to describe unfavorable shifts in balance due in the context of our logistic regression analyses. Note that Figures 3–7 plot total positive income on the X axis. The same pattern emerged when we plotted age on the X axis; therefore, we present our findings

[3] Due to limitations in computing power, to conduct the logistic regression analyses we created an analytic dataset by randomly sampling five million cases from our original dataset using the sample_n function in R's dplyr package.

with age held constant at the median age. We highlight and discuss changes in personal characteristics and tax returns that have a particularly large impact on the predicted probability of balance due.[4] Finally, to better understand the interactive effect divorce and changes to Schedules A and C have on balance due, we conducted a series of crosstabulations to investigate the relationship between divorce, Schedule A, and Schedule C changes.

Initially, we conceptualized our outcome variable as a continuous measure of "debt ratio difference." We calculated debt ratio[5] in TY 2016 and TY 2017 as total refund or balance due divided by total positive income.[6] Then, we calculated debt ratio difference as percentage point difference between TY 2016 and TY 2017 debt ratios. Debt ratio difference would indicate amount of a taxpayer's balance due in relation to their ability to pay each year. Unfortunately, as shown in Figure 3, our debt ratio difference data was leptokurtic[7] and unsuitable for use in analysis of variance or standard regression analyses.

FIGURE 3. Leptokurtic Distribution of Debt Ratio Difference Variable



Although our debt ratio variable was not suitable for inferential statistics, basic descriptives confirm intuition as shown in Table 3. As taxpayers' balance due increases relative to their total positive income, inability to pay the balance due increases. For example, the biggest percentage point difference between TY 2016 to TY 2017 (+5.67) was for the group of taxpayers who went from refund or even in TY 2016 (-2.95%, meaning their refund was 2.95% of their total positive income) to a balance due with a CP14 in TY 2017 (2.72%, meaning their balance due was 2.72% of their total positive income).

4 Given a large enough sample size, even very small effect sizes can produce significant p-values. Our sample was large, and all logistic regressions were statistically significant at $p < .001$. Therefore, we focus our discussion on interpretation of our large effect sizes.

5 Negative debt ratios indicate refunds and positive debt ratios indicate balance due. This is a function of the way IRS databases capture refunds as negative values and balance due as positive values.

6 We chose total positive income, as opposed to adjusted gross income, because it more accurately represents the amount of money a taxpayer has available to pay a balance due.

7 Kurtosis is a statistic that measures the extent to which a distribution contains outliers. Leptokurtic distributions have higher kurtosis than the normal distribution, which means they have "heavy tails" and contain more outliers.

TABLE 3. Median Debt Ratio and Median Debt Ratio Difference

	Median Debt Ratio		Difference in percentage points
	Tax Year 2016 (%)	Tax Year 2017 (%)	
No Change from Tax Year 2016 to Tax Year 2017			
Refund or even → Refund or even	-5.17%	-4.30%	+0.87
Balance due without CP14 → Balance due without CP14	2.53%	2.63%	+0.10
Balance due with CP14 → Balance due with CP14	4.91%	4.57%	-0.34
Favorable Shift from Tax Year 2016 to Tax Year 2017			
Balance due without CP14 → Refund or even	1.75%	-2.11%	-3.85
Balance due with CP14 → Refund or even	2.97%	-2.62%	-5.59
Balance due with CP14 → Balance due without CP14	4.72%	3.86%	-0.86
Unfavorable Shift from Tax Year 2016 to Tax Year 2017			
Refund or even → Balance due without CP14	-2.29%	1.63%	+3.92
Refund or even → Balance due with CP14	-2.95%	2.72%	+5.67
Balance due without CP14 → Balance due with CP14	3.93%	4.40%	+0.47

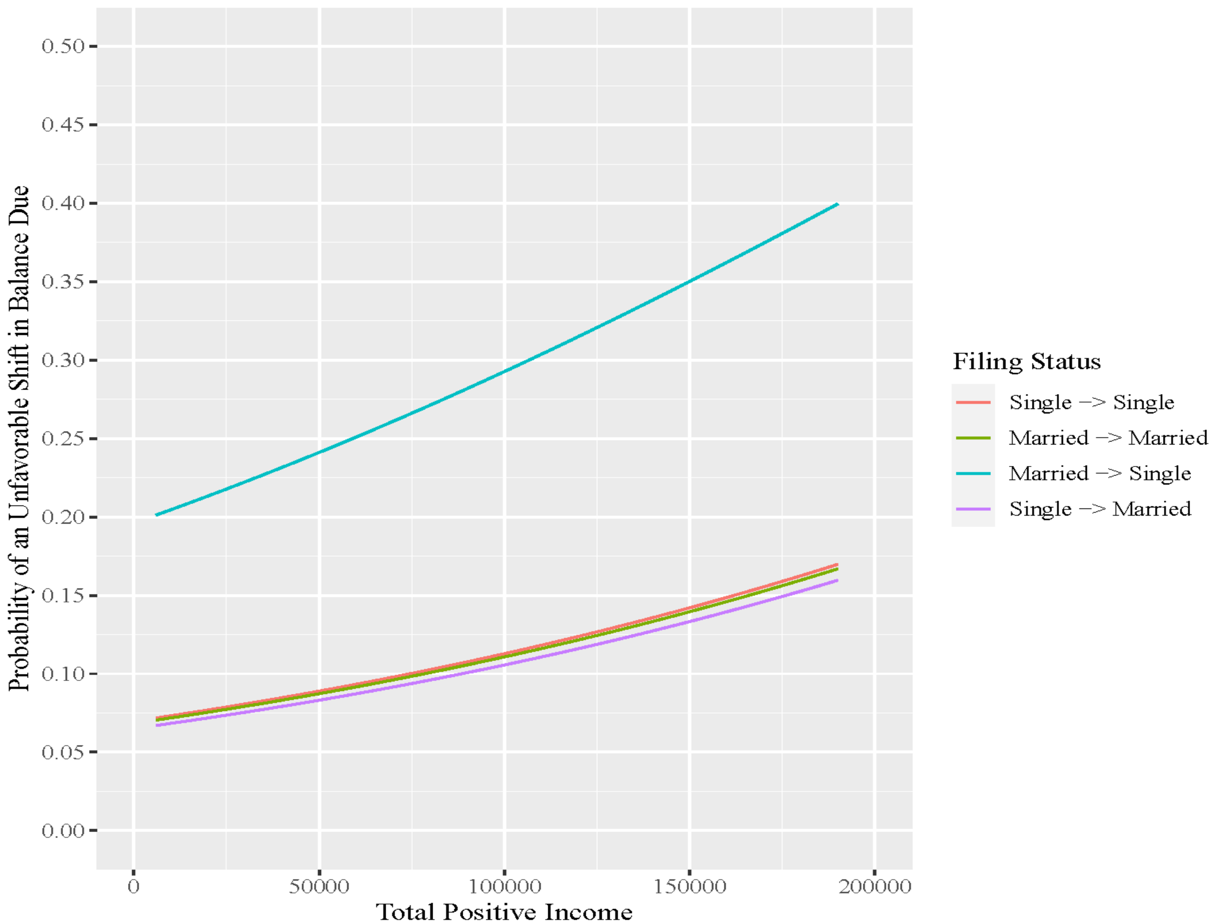
Note: Negative debt ratios indicate refunds and positive debt ratios indicate balance due.

Results

Filing Status Logistic Regressions

Divorce (moving from Married Filing Jointly [MFJ] or Married Filing Separately [MFS] status to Single filing status) creates the most risk for balance due. As shown in Figure 4, the impact of divorce is consistent and large, even when we hold age and total positive income constant or consider the impact of Schedule A and Schedule C. The one exception (discussed below) is single taxpayers who add Schedule C. We hypothesize this is due to the multifaceted impact divorce has on personal finances and tax returns. After a divorce, a taxpayer may lose the ability to claim children as dependents or attach Schedule A to deduct medical expenses and mortgage interest. Expenses associated with divorce also may compel taxpayers to withdraw money from retirement savings, which results in both a 10% early withdrawal fee and the requirement to report that amount as income. As the graph below shows, the balance due risk of a recently divorced taxpayer is approximately triple that of other taxpayers across all levels of total positive income. While the impact of divorce is dramatic, the number of taxpayers who get divorced is small. There were 578,475 returns in TY 2017 (0.5% of all returns) from divorced taxpayers who were married in TY 2016. Of the returns associated with taxpayers who got divorced, 7.8% (45,287) shifted to balance due that required a CP14 (which cost the IRS approximately $548,000).

FIGURE 4. Effect of Filing Status on Balance Due

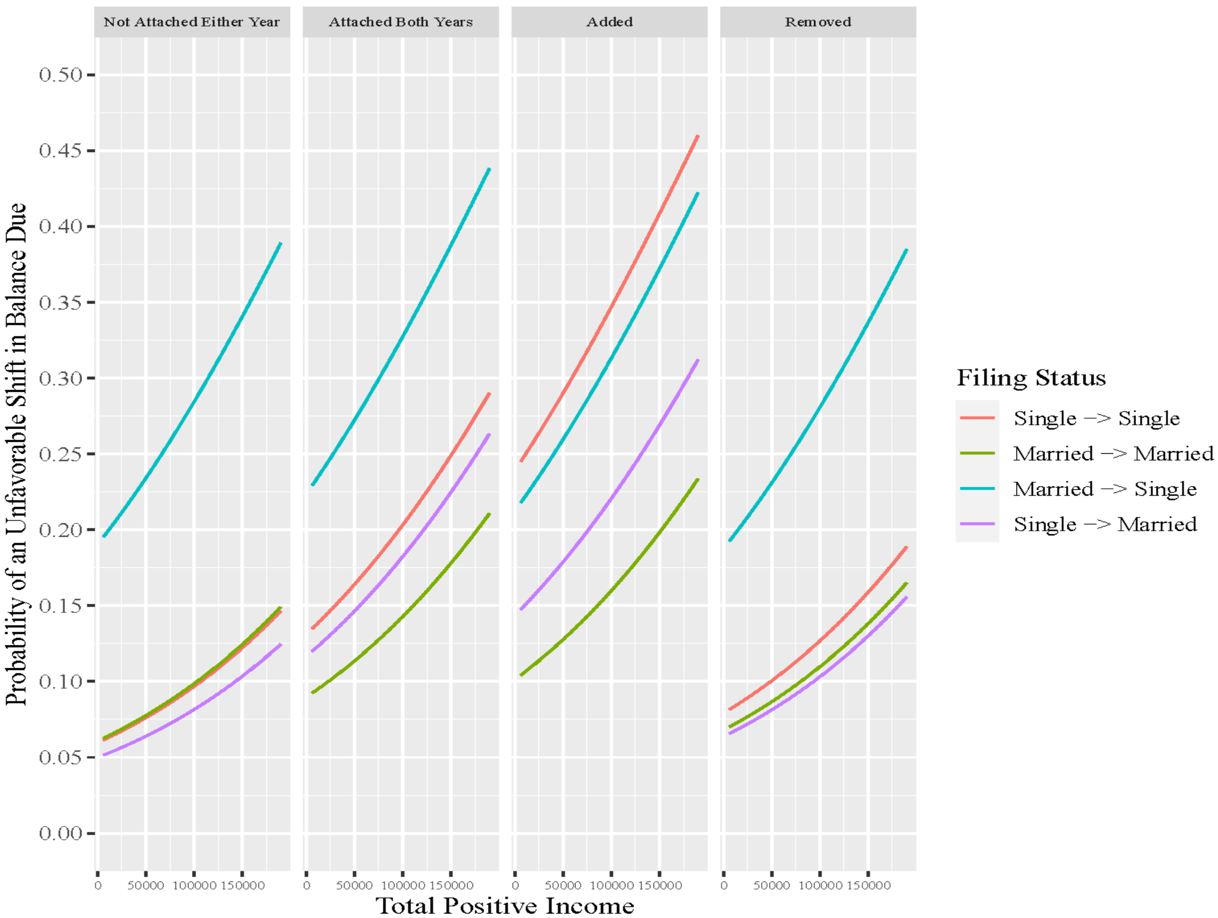


Notes:
Age held constant at median.
An unfavorable shift in balance due is defined as either a change from refund/even to a balance due with or without a CP14 or having a balance due without a CP14 followed by a balance due with a CP14.

The impact of divorce on balance due is so large that it completely suppresses the effects of schedule attachment except in one specific circumstance. As shown in Figure 5, single taxpayers who add a Schedule C have a higher risk of balance due than divorced taxpayers who add a Schedule C. The difference in risk between single and divorced taxpayers who add a Schedule C is small, but noteworthy as the only instance where divorced taxpayers do not have the highest risk by a substantial margin. We hypothesize that single filing status may function as a proxy for demographic characteristics associated with workers in the gig economy (i.e., younger workers with fewer family ties for whom the gig job is the only source of income). A 2018 survey of gig economy workers[8] finds that for those whose gig job is the primary source of income, 80% state it would be difficult to pay an unexpected expense of $1,000. For these taxpayers, a balance due on their tax return may count as an unexpected expense.

[8] Gig-Economy-2018-Marketplace-Edison-Research-Poll-FINAL.pdf (edisonresearch.com)

FIGURE 5. Effect of Filing Status and Schedule C on Balance Due

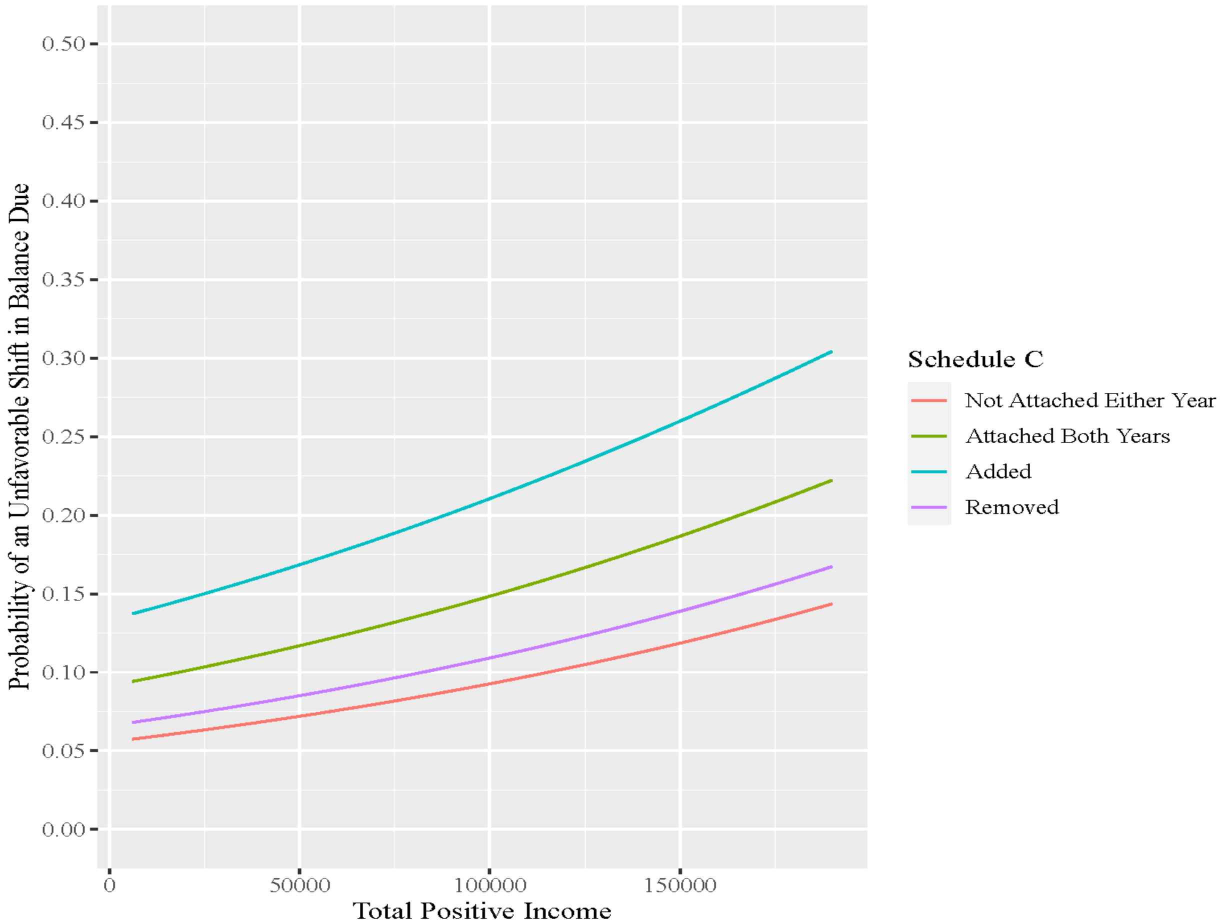


Notes:
Age held constant at median.
An unfavorable shift in balance due is defined as either a change from refund/even to a balance due with or without a CP14 or having a balance due without a CP14 followed by a balance due with a CP14.

Schedule C Logistic Regressions

In general, adding Schedule C or having it attached both tax years studied increases the risk of balance due compared to not having it attached in either year. Figure 6 shows, compared to taxpayers who did not have Schedule C attached in either TY 2016 or TY 2017, the increased risk of adding Schedule C was more than double and the increased risk for having it attached both years was a little less than double. The effect of adding Schedule C or having it attached both years is not as strong as the impact of divorce; however, the number of associated returns in these two groups is larger. Approximately 23 million taxpayer returns either added Schedule C in TY 2017 or had it attached in both TY 2016 and TY 2017. Of these returns, around 1.2 million (5.2%) experienced a shift in balance due that resulted with receipt of a CP14. The total cost to the IRS for the CP14 notices sent to balance due taxpayers who either had Schedule C attached both years or added it in TY 2017 was approximately $14.5 million dollars.

FIGURE 6. Effect of Schedule C on Balance Due

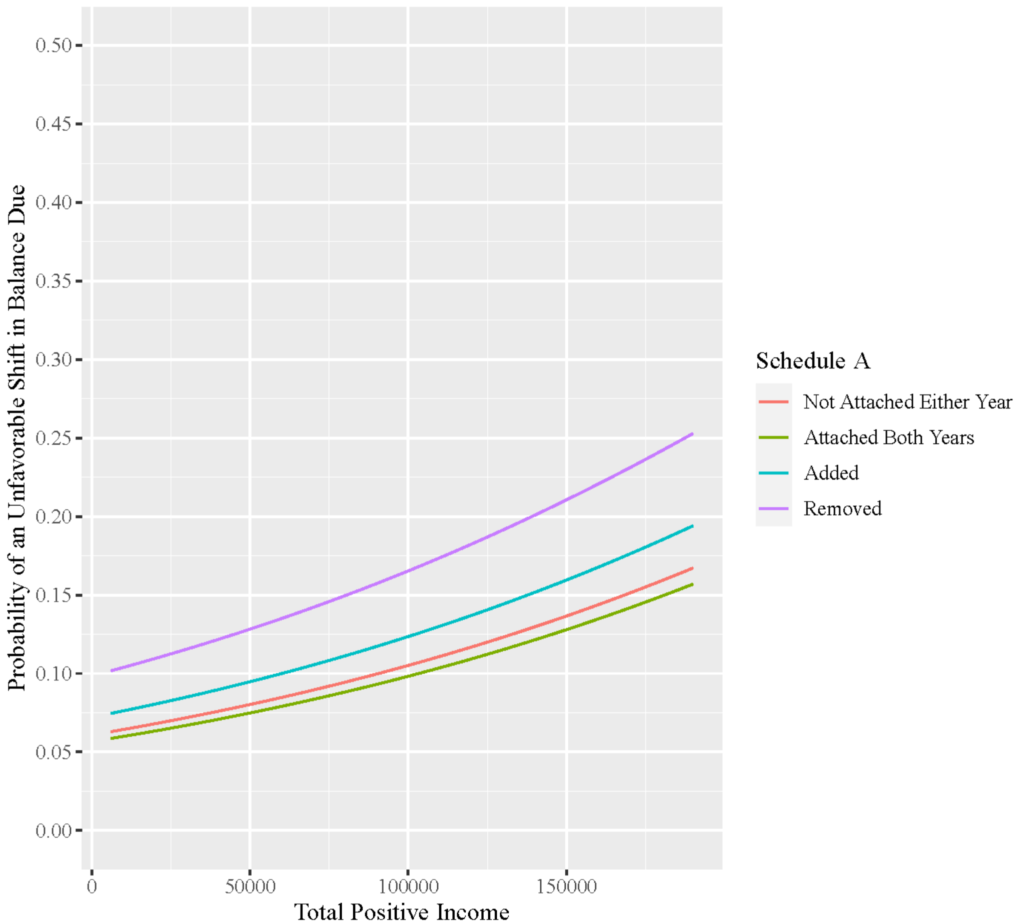


Notes:
Age held constant at median.
An unfavorable shift in balance due is defined as either a change from refund/even to a balance due with or without a CP14 or having a balance due without a CP14 followed by a balance due with a CP14.

Schedule A Logistic Regressions

As shown in Figure 7, compared to not having it attached either year, removing Schedule A increased the risk of a balance due category by a little less than double. Like Schedule C, the increased risk is not as strong as that of divorce, but it impacts significantly more taxpayers. Between TY 2016 and TY 2017, approximately 5 million returns dropped Schedule A. Of those, around 200,000 (4.1percent) received a CP14, which cost the IRS a little more than $2.4 million dollars.

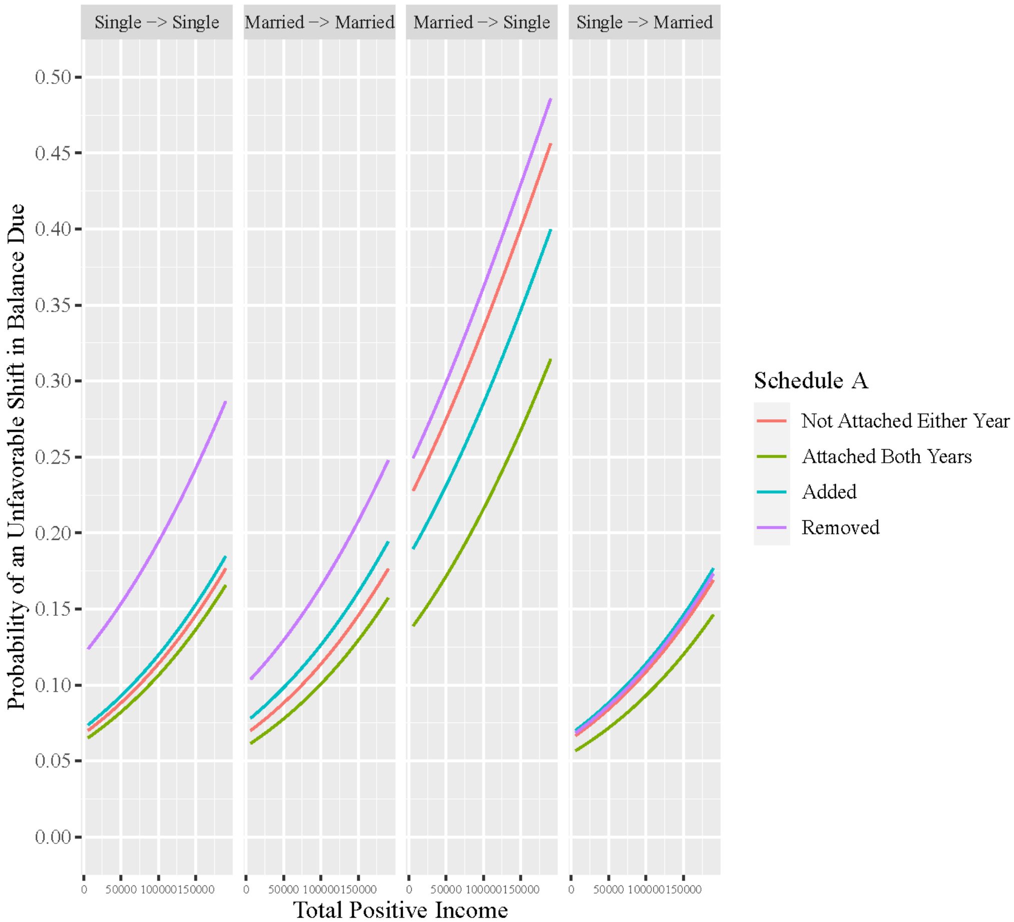
Figure 7. Effect of Schedule A on Balance Due



Notes:
Age held constant at median.
An unfavorable shift in balance due is defined as either a change from refund/even to a balance due with or without a CP14 or having a balance due without a CP14 followed by a balance due with a CP14.

Removing Schedule A had a particularly strong impact on single taxpayers. As shown in Figure 8, removing Schedule A almost doubled single taxpayers' risk of a balance due category whereas for other taxpayers the impact of removing Schedule A was less (for married taxpayers) or nonexistent (for taxpayers who got married).

FIGURE 8. Effect of Filing Status and Schedule A on Balance Due



Notes:
Age held constant at median.
An unfavorable shift in balance due is defined as either a change from refund/even to a balance due with or without a CP14 or having a balance due without a CP14 followed by a balance due with a CP14.

Interaction between divorce and making specific changes to tax returns

As the logistic regression analyses demonstrate, the impact of divorce is large and somewhat independent of tax return or demographic characteristics like total positive income and age. We conducted crosstab analyses to further investigate the interactive effect between divorce and specific changes a taxpayer might make to their tax return during a divorce. The rate at which taxpayers experience an unfavorable balance due shift is 26% for taxpayers who get divorced compared to 9% overall. The impact of additional changes to tax returns does not differ greatly between all taxpayers and those who get divorced except in two cases. As shown in Figure 8, removing the mortgage interest deduction has a larger additional negative impact on all taxpayers than it does on taxpayers who get divorced. Similarly, as shown in Figure 9, having Schedule C attached or adding it has a larger additional negative impact on all taxpayers than it does on those who get divorced.

FIGURE 9. Percentage of Taxpayers Who Experienced an Unfavorable Balance Due Shift After Making the Following Changes

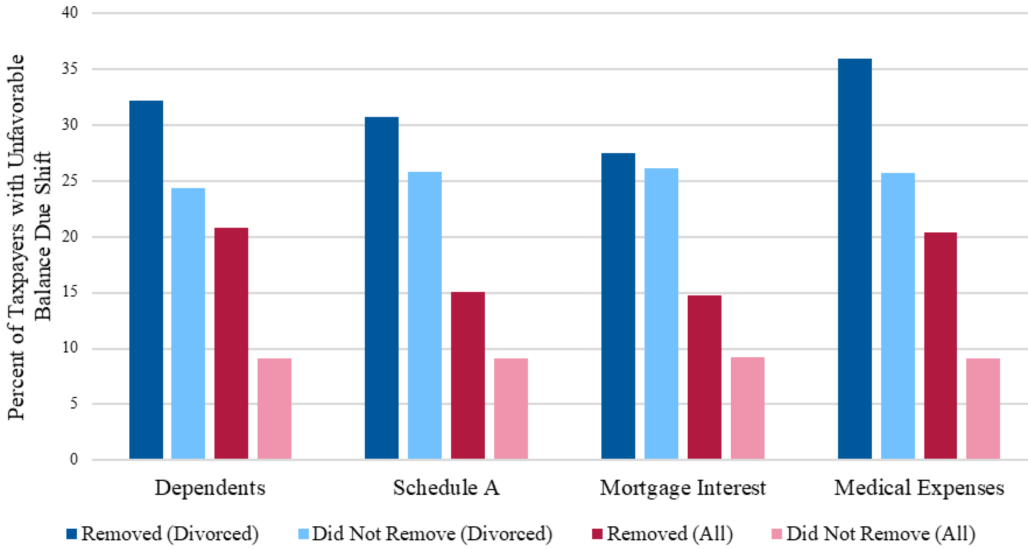
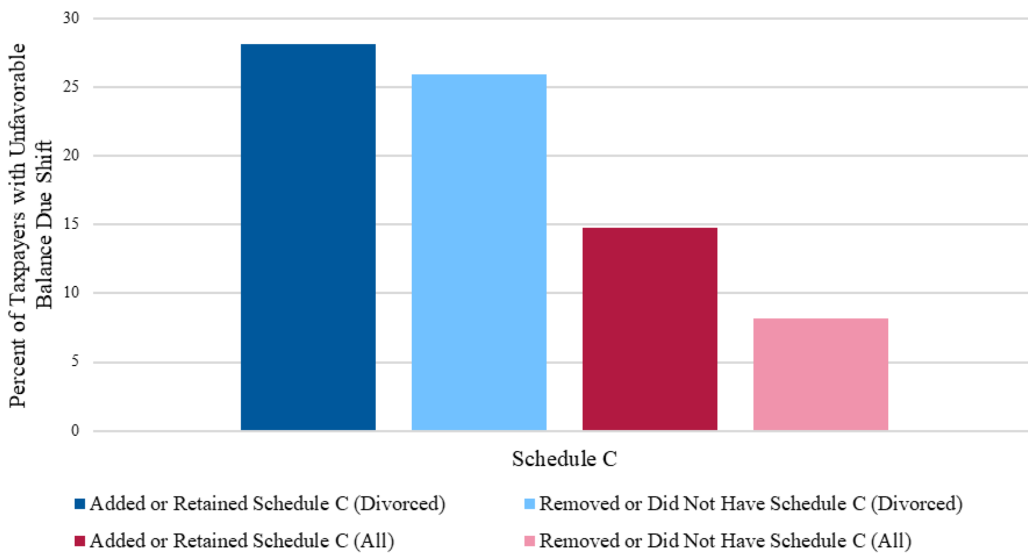


FIGURE 10. Percentage of Taxpayers Who Got Divorced and Experienced an Unfavorable Balance Due Shift After Making the Following Changes



Conclusions

A significant number of taxpayer (and associated tax return) shifts into unfavorable balance due categories is not due to commonly occurring changes on a taxpayer's return, such as increased income. Rather, many result from specific life events associated with filing status, small business activity, and deduction qualifications. Somewhat surprisingly, the financial impact of an unfavorable shift to balance due is not income specific. Divorce, removing Schedule A, attaching Schedule C (or having it attached already) affects all income levels and can happen to any taxpayer. As such, IRS and other stakeholders in the tax administration ecosystem might consider preemptive mitigation strategies aimed at these groups of taxpayers to reduce the occurrence of unfavorable balance due positions.

2. Balance Due Gap Analysis

Method

The results of the above statistical analysis make clear that certain groups of taxpayers are particularly vulnerable to an unfavorable balance due shift: those who get divorced, those who remove Schedule A, and those who either add Schedule C or attach in both tax years. These findings suggest there might be opportunties for improvement in IRS outreach efforts to ensure taxpayers can easily find the educational and other resource materials they need even when not necessarily looking for them. To explore the potential gap, we first reviewed current messaging available to taxpayers on IRS.gov. We used specific keywords affected taxpayers might use to conduct their own research. We chose search terms assuming taxpayers vaguely know divorce or starting a business has some tax implications. We used the following keywords in the IRS.gov search tool: divorce, starting a small business, and gig economy.

Second, we reviewed current messaging available to taxpayers using the Google search engine. We chose Google because it controls more than 92% of the search engine market share worldwide.[9] We conducted our Google search of keywords below assuming a naïve taxpayer may not have considered the tax implications of such life events. We then added the phrase "and taxes" to the end of each keyword/phrase to account for taxpayers who recognize there are tax implications to certain life events. We used the following keywords/phrases:

- Divorce
- Getting divorced
- Starting a new business
- How to start a new business
- Independent contractor
- Driving for Uber
- Driving for Lyft

Third, we conducted an online search for in-person support to capture messaging available to taxpayers who may not rely on the internet for information. We searched the following groups:

- Divorce attorney organizations
- Divorce support groups
- Tax preparation organizations, CPAs, and accountants
- Tax workshops

We limited our internet searches to the use of an IRS-issued computer. When compared to a similar search on a non-IRS computer, the results were similar. However, there were differences in the order results appeared. Additionally, we limited our Google search to the first two pages of results. According to an article from Small Business Genius,[10] "75% of searchers never click past the first page of results." So we assume a typical taxpayer does not likely continue to search internet results beyond page two. We did not include search results marked "Ad." In some searches, pages one and two contained minimal information. In such cases, we then proceeded to page three of the search results.

Results

The search of IRS.gov found several publications and assistance to help meet a taxpayer's tax obligations (Appendix A). However, the technical nature of the publications suggest they cater primarily to tax professionals and those with knowledge of filing taxes. Also, the IRS.gov front page does not specifically address the issue of avoiding a balance due.

Searching the term "divorce" in Google yielded hundreds of results. Most of the webpages surveyed did not produce results that would help a taxpayer find information to avoid a balance due situation (Appendix B).

---

[9] telanganatoday.com/google-dominates-search-engine-market-holds-92-per-cent-share-report.

[10] Small Business Genius.

At page two of the search results, we found a government-related website,[11] which has a link to IRS Publication 504, Divorced or Separated Individuals, at IRS.gov. When the keyword phrase included "and taxes," results produced several websites offering basic tax advice and included links to divorce attorneys. However, there were no extensive explanations of taxes related to balance due.

Searching the phrase "starting a new business," we found basic tax information included in a list of general advice for those starting a business (Appendix C). Page one contained results with links to the Small Business Administration[12] and the U.S. Government website.[13] Both sites briefly discussed filing self-employment taxes and their related forms. However, neither website contained in-depth information about the potential tax consequences of having a balance due account. Page two of the search results contained a link to the IRS's Small Business/Self-Employed webpage. When we added the keyword phrase, "and taxes," results included more beneficial websites, including IRS.gov.

A search of topics related to the gig economy (Appendix D) and specific companies in the rideshare industry (e.g., Uber, Lyft) produced results with general guidance about working as an independent contractor. These companies' websites contained minimal guidance for meeting tax obligations. Primarily, the sites emphasized that drivers receive an IRS Form 1099 to document income earned and the need to keep accurate records for tax filing. Searches for "independent contractor" and "gig economy" produced sites with limited tax advice. Adding the keyword phrase "and taxes" to the initial searches produced links to IRS.gov and sites with more tax-related topics.

In searching for information about offline tax advice about divorce, starting a business, and working in a gig economy (see Appendix E), most results were for local support groups and workshops.

Conclusions

A simple Google search on divorce, starting a business, and working in the gig economy provide minimal or no guidance to naïve taxpayers who may not realize the tax implications of these life events. Searches that add the phrase "and taxes" produce somewhat more helpful information. However, they do not provide sufficient guidance as an "early intervention." Taxpayers must know the specific keywords to use to get helpful results.

IRS research has consistently shown individual taxpayers think about taxes and interact with the IRS only when time to file their tax return. The 2020 Taxpayer Experience Survey Free File Focus Groups[14] found taxpayers "start thinking about preparing and filing their tax return in late January or early February." Taxpayers give little consideration to their potential tax liability earlier in the year. Additionally, 2021 Taxpayer Experience Survey[15] results indicate 60% of taxpayers consult only one source for assistance with any tax issues. Finally, results of the 2021 Comprehensive Taxpayer Attitude Survey[16] show a majority (66%) of taxpayers "trust the IRS to help me understand my tax obligation" and a remarkable 86% of taxpayers agree "the more information and guidance the IRS provides, the more likely people are to correctly file their tax returns." Our findings suggest IRS attention to early warnings affords a great opportunity to work with partners and stakeholders to fill the communication and messaging void.

3. Intervention

The results of the two above studies indicate the IRS should review current messaging to encourage taxpayers (particularly those vulnerable to the balance due shift) to review their tax withholdings, exemptions, etc. immediately upon a significant life event. The enactment of the Taxpayer First Act[17] (TFA) in 2019 provided the

[11] www.benefits.gov.

[12] www.SBA.gov.

[13] www.USA.gov.

[14] Internal Revenue Service, Wage and Investment Strategies and Solutions. (July 2020). 2020 Taxpayer Experience Survey (TES) Free File Focus Groups Top Line Summary Report.

[15] Internal Revenue Service, Wage and Investment Strategies and Solutions. (March 2022). Taxpayer Experience Survey (TES) 2021 National Report.

[16] Internal Revenue Service, Research, Applied Analytics, and Statistics. (April 2022). Comprehensive Taxpayer Attitude Survey: Past, Present, and Future.

[17] Taxpayer First Act | Internal Revenue Service.

IRS an opportunity to reimagine and improve services based on the needs of the taxpayer. Additionally, the 2021 Presidential Executive Order (EO) on Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government[18] directs Federal agencies to "design and deliver services with a focus on the actual experience of the people whom it is meant to serve." To put our findings to work in service of the TFA and the 2021 EO, WISS and newly formed Taxpayer Experience Office (TXO) jointly designed data-driven interventions to help taxpayers avoid the shift to an unplanned balance due.

The TXO formed a Balance Due Team (BDT) to find proactive solutions to help prevent balance due accounts. Instead of reacting to undesired balance due accounts, taxpayers will employ pre-filing interventions and strategies to mitigate the underlying causes. This team will collaborate with internal and external stakeholders to develop, implement, and evaluate plans to reduce the number of taxpayers who have a balance due at the time of filing. The team aims to accomplish this goal by creating and implementing a phased strategy divided into three areas of focus:

- Phase 1 (in progress): Divorced taxpayers (change in filing status)
- Phase 2 (tentative): Itemized deductions (Schedule A)
- Phase 3 (tentative): "Side hustle" and gig economy workers (Schedule C)

Since Phase 1 is in progress, we discuss it in detail below.

Phase 1: Divorced Taxpayers

The BDT shares a collective vision to decrease the number of balance due occurrences in divorced taxpayers by proactively providing education and outreach, services and products, and technology platforms/applications that meet the needs of the recently divorced taxpayer in the language, timing, and method preferred (choice and access). The key strategic goals for Phase 1 are as follows.

- **Development of IRS.gov/divorce:** Information on how divorce can impact taxes is scattered throughout IRS.gov and can be difficult to find. As the gap analysis concludes, search engines and the internet have insufficient information readily available. Most sites have little value or a referral to a tax professional for additional information. The BDT's goal is to develop IRS.gov/divorce, which will serve as a landing page for divorced taxpayers and provide a one-stop shop where taxpayers and tax professionals can easily access divorce related tax material.
- **Development of new material:** Much of the information on divorce and taxes are in publications and articles difficult for the average taxpayer to understand. Taxpayers going through a divorce are already stressed and may abandon their search for relative tax impact information if too difficult. The BDT seeks to design and develop easy-to-read material to help taxpayers avoid a balance due as they navigate their divorce. One-page flyers such as "How to not owe taxes after a divorce" or "5 things to know about divorce and taxes" can grab the attention of a taxpayer going through divorce and provide information to help them avoid a balance due prior to filing.
- **Development of an external communication campaign:** The BDT will help create content for external outreach. The IRS can share this content through social media, online (IRS.gov) and/or directly with partners. They will leverage external networks and technology to develop an outreach campaign to drive traffic to IRS.gov/divorce. The IRS can distribute easy-to-understand content on social media posts across multiple platforms and in multiple languages and otherwise share the new BDT material with external stakeholders.

[18] Executive Order on Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government | The White House.

Follow-on Efforts

A foundation of the BDT intervention plan is the WISS statistical analysis and gap analysis that inform efforts reaching the recently divorced taxpayer, the taxpayer who lost itemized deductions, and the taxpayer with a "side hustle" or working in the gig economy. Starting in summer 2023, WISS and the BDT leveraged focus groups with tax preparers at the IRS Nationwide Tax Forums and Latino Tax Fest that gather further input to support and refine the BDT's intervention strategy. Tax preparers are a valuable IRS partner and their extensive knowledge regarding the tax-specific information, resources and support taxpayers need to meet their tax obligations promise to provide valuable insight.

Appendix A. Search Results from IRS.gov

Search term(s)	Results	Comments
Divorce	Page 1. "What You Need to Know Before Getting a Divorce." https://www.nolo.com/legal-encyclopedia/getting-started-with-your-divorce.html (retrieved 7/8/2022)	Looked like it would be useful, but when clicked there was no information on taxes.
	Page 1. https://www.findlaw.com/family/divorce/how-to-divorce.html (retrieved 7/8/2022) >>> Followed link on page: https://www.findlaw.com/family/divorce/marriage-divorce-taxes-and-your-social-security-number.html (retrieved 7/8/2022)	Page focused on changing one's name with SSA after divorce. No information on taxes; the page recommends contacting a divorce attorney. Page provides attorney referral service based on Zip Code.
	Page 3. https://www.hg.org/divorce-law-center.html. (Retrieved 7/11/2022) >>> Followed link to "Divorce Law Basics." https://www.hg.org/divorce-law-center.html (retrieved 7/11/2022)	Provides basic information about divorce. No information on taxes. Page provides attorney referrals.
Getting divorced	Page 1. "What Happens in a Divorce?" https://www.alllaw.com/articles/family/divorce/article64.asp (retrieved 7/11/2022). IRS laptop	Provides information on the process of divorce. No tax considerations given.
	Page 1. "Divorce Advice Every Woman Getting a Divorce Needs To Hear." https://www.brides.com/pieces-of-divorce-advice-for-women-1102751 (retrieved 7/11/2022)	Recommends gathering financial information related to assets and liabilities. No discussion of taxes.
	Page 1. "Should I Get a Divorce?" https://www.oprahdaily.com/life/a26040141/should-i-get-a-divorce/ (retrieved 7/11/2022)	From Oprah Winfrey's website. Because of her notoriety and reach, though it would contain useful information. Site contains "signs" that it is time for divorce. No tax information provided.
	Page 2. https://myguidance.fidelity.com/ftgw/pna/public/lifeevents/content/divorce/getting-divorced (retrieved 7/11/2022) >>> Clicked link on page "Finances after Divorce." https://myguidance.fidelity.com/ftgw/pna/public/lifeevents/content/divorce/divorce-and-finances	Website states to "to consider any tax consequences associated with selling investments in a taxable account." No further information provided.
	Page 2. "12 Mistakes to Avoid When Divorcing Over 50." https://www.investopedia.com/personal-finance/mistakes-avoid-when-divorcing-over-50/ (retrieved 7/11/2022)	Contains a warning not to ignore tax consequences. Information is about the implications of making/receiving alimony and child support payments. Refers the reader to www.benefits.gov.
	>>> Clicked hyperlink on page "program can help you" Tax Relief for Divorced or Separated Individuals. https://www.benefits.gov/benefit/946 (retrieved 7/12/2022).	Benefits.gov links to Pub 504 at IRS.gov.
	Page 2. What Older Adults Should Know about Getting Divorced and (Maybe) Remarried. https://www.kiplinger.com/personal-finance/604696/what-older-adults-should-know-about-getting-divorced-and-maybe-remarried (retrieved 7/11/2022)	Seemed like it would be useful. Contained no information on taxes.

Appendix A (continued). Search Results from IRS.gov

Search term(s)	Results	Comments
	Page 3. “Divorce.” https://www.legalzoom.com/articles/divorce (retrieved 7/11/2022) >>> Followed link “Considering Divorce? 10 Things to Think About” https://www.legalzoom.com/articles/considering-divorce-10-things-to-think-about (retrieved 7/12/2022)	Thought it could have useful information, but it only contained articles on how to file for divorce and forms need to file for divorce. No discussion of tax implications.
	Page 3. “Getting Divorced.” https://turbotax.intuit.com/tax-tips/marriage/getting-divorced/L20NC66cf (retrieved 7/11/2022)	Good explanation of filing status, claiming dependents, medical expenses, tax credits, payments to an ex-spouse, transfer of assets, home sale, and transfer of retirement assets.
Getting divorced and taxes	Page 1. Divorce & Taxes 101: Filing Taxes After a Divorce. https://blog.turbotax.intuit.com/tax-tips/divorce-and-taxes-4018/ (retrieved 7/22/2022)	Brief discussion about filing status, claiming child support, requirement for filing as HOH, Child and Dependent Care credit. The article is followed by people posting their questions, comments. Most recent post was from 2018.
	Page 1. Filing Taxes After Divorce: A Practical Guide. https://smartasset.com/taxes/filing-taxes-after-divorce (retrieved 7/22/2022)	Advice on choosing the right filing status, updating one’s W-4, claiming dependents, and deducting legal fees.
	Page 1. Most-Overlooked Tax Breaks for the Newly Divorced. https://www.kiplinger.com/taxes/tax-deductions/602038/most-overlooked-tax-breaks-for-the-newly-divorced (retrieved 7/22/2022)	Reminds taxpayers to update W-4, how to determine if one qualifies for filing as HOH, alimony payments for divorce decrees before end of 2018. Noncustodial parents must complete Form 8332 if they claim child tax credit.
	Page 1. Filing Taxes After Divorce. https://www.hrblock.com/tax-center/filing/personal-tax-planning/divorce-and-taxes/ (retrieved 7/22/2022)	Alimony payments no longer deductible. Refers taxpayers to Form 8332 noncustodial parent claiming the child/children. Discusses how IRAs are handled.
	Page 1. Tax Complications to Watch Out for During and After a Divorce. https://familylaw.lyttlelaw.com/tax-complications-to-watch-out-for-during-and-after-a-divorce.html (retrieved 7/22/2022)	Austin, TX, divorce attorney page. Explains how to divide tax refunds, filing tax returns in TX, requirement for filing HOH.
	Page 1. What Getting Divorced or Separated Means for Your 2021 Tax Return. https://www.thebalance.com/what-divorced-or-separated-means-for-taxes-4125740 (retrieved 7/22/2022)	Divorce must be final before end of year for IRS to recognize. Refers to Pub 504; filing as HOH; claiming the children; alimony no longer deductible; paying back taxes and property taxes.
	Page 1. Your Taxes After Divorce. https://www.investopedia.com/taxes-after-divorce-5192868 (retrieved 7/22/2022)	Discusses filing status and qualifying for HOH. Explains rules related to Earned Income Credit, American Opportunity Tax Credit, and child and dependent care credit. After 2018 alimony payments were no longer deducted from taxable income. Gains on the sale of primary home not taxable up to $250K; claiming HOH.
	Page 2. Tax Tips for Women Going Through Divorce. https://www.forbes.com/sites/jefflanders/2012/03/07/tax-tips-for-women-going-through-divorce/ (retrieved 7/22/2022)	Warns about filing joint returns, how to handle over/underpayments, filing HOH, claiming children, child support and alimony considerations, capital gains tax on high ticket assets held for a long time

Appendix B. Internet Search Results for "Divorce" Topics

Search term(s)	Results	Comments
Divorce	Page 1. "What You Need to Know Before Getting a Divorce." https://www.nolo.com/legal-encyclopedia/getting-started-with-your-divorce.html (retrieved 7/8/2022).	Looked like it would be useful, but when clicked there was no information on taxes.
	Page 1. https://www.findlaw.com/family/divorce/how-to-divorce.html (retrieved 7/8/2022). >>> Followed link on page: https://www.findlaw.com/family/divorce/marriage-divorce-taxes-and-your-social-security-number.html (retrieved 7/8/2022).	Page focused on changing one's name with SSA after divorce. No information on taxes; the page recommends contacting a divorce attorney. Page provides attorney referral service based on Zip Code
	Page 3. https://www.hg.org/divorce-law-center.html (Retrieved 7/11/2022). >>> Followed link to "Divorce Law Basics." https://www.hg.org/divorce-law-center.html (retrieved 7/11/2022).	Provides basic information about divorce. No information on taxes. Page provides attorney referrals.
Getting divorced	Page 1. "What Happens in a Divorce?" https://www.alllaw.com/articles/family/divorce/article64.asp (retrieved 7/11/2022). IRS laptop.	Provides information on the process of divorce. No tax considerations given.
	Page 1. "Divorce Advice Every Woman Getting a Divorce Needs To Hear." https://www.brides.com/pieces-of-divorce-advice-for-women-1102751 (retrieved 7/11/2022).	Recommends gathering financial information related to assets and liabilities. No discussion of taxes.
	Page 1. "Should I Get a Divorce?" https://www.oprahdaily.com/life/a26040141/should-i-get-a-divorce/ (retrieved 7/11/2022).	From Oprah Winfrey's website. Because of her notoriety and reach, though it would contain useful information. Site contains "signs" that it is time for divorce. No tax information provided.
	Page 2. https://myguidance.fidelity.com/ftgw/pna/public/lifeevents/content/divorce/getting-divorced (retrieved 7/11/2022). >>> Clicked link on page "Finances after Divorce." https://myguidance.fidelity.com/ftgw/pna/public/lifeevents/content/divorce/divorce-and-finances	Website states to "to consider any tax consequences associated with selling investments in a taxable account." No further information provided.
	Page 2. "12 Mistakes to Avoid When Divorcing Over 50." https://www.investopedia.com/personal-finance/mistakes-avoid-when-divorcing-over-50/ (retrieved 7/11/2022).	Contains a warning not to ignore tax consequences. Information is about the implications of making/receiving alimony and child support payments. Refers the reader to www.benefits.gov
	>>> Clicked hyperlink on page "program can help you" Tax Relief for Divorced or Separated Individuals. https://www.benefits.gov/benefit/946 (retrieved 7/12/2022).	Benefits.gov links to Pub 504 at IRS.gov.
	Page 2. What Older Adults Should Know about Getting Divorced and (Maybe) Remarried. https://www.kiplinger.com/personal-finance/604696/what-older-adults-should-know-about-getting-divorced-and-maybe-remarried (retrieved 7/11/2022).	Seemed like it would be useful. Contained no information on taxes.

Appendix B (continued). Internet Search Results for "Divorce" Topics

Search term(s)	Results	Comments
	Page 3. "Divorce." https://www.legalzoom.com/articles/divorce (retrieved 7/11/2022). >>> Followed link "Considering Divorce? 10 Things to Think About." https://www.legalzoom.com/articles/considering-divorce-10-things-to-think-about (retrieved 7/12/2022).	Thought it could have useful information, but it only contained articles on how to file for divorce and forms need to file for divorce. No discussion of tax implications.
	Page 3. "Getting Divorced." https://turbotax.intuit.com/tax-tips/marriage/getting-divorced/L20NC66cf (retrieved 7/11/2022).	Good explanation of filing status, claiming dependents, medical expenses, tax credits, payments to an ex-spouse, transfer of assets, home sale, and transfer of retirement assets.
Getting divorced and taxes	Page 1. Divorce & Taxes 101: Filing Taxes After a Divorce. https://blog.turbotax.intuit.com/tax-tips/divorce-and-taxes-4018/ (retrieved 7/22/2022).	Brief discussion about filing status, claiming child support, requirement for filing as HOH, Child and Dependent Care credit. The article is followed by people posting their questions, comments. Most recent post was from 2018.
	Page 1. Filing Taxes After Divorce: A Practical Guide. https://smartasset.com/taxes/filing-taxes-after-divorce (retrieved 7/22/2022).	Advice on choosing the right filing status, updating one's W-4, claiming dependents, and deducting legal fees.
	Page 1. Most-Overlooked Tax Breaks for the Newly Divorced. https://www.kiplinger.com/taxes/tax-deductions/602038/most-overlooked-tax-breaks-for-the-newly-divorced (retrieved 7/22/2022).	Reminds taxpayers to update W-4, how to determine if one qualifies for filing as HOH, alimony payments for divorce decrees before end of 2018. Noncustodial parents must complete Form 8332 if they claim child tax credit.
	Page 1. Filing Taxes After Divorce. https://www.hrblock.com/tax-center/filing/personal-tax-planning/divorce-and-taxes/ (retrieved 7/22/2022).	Alimony payments no longer deductible. Refers taxpayers to Form 8332 noncustodial parent claiming the child/children. Discusses how IRAs are handled.
	Page 1. Tax Complications to Watch Out for During and After a Divorce. https://familylaw.lyttlelaw.com/tax-complications-to-watch-out-for-during-and-after-a-divorce.html (retrieved 7/22/2022)	Austin, TX, divorce attorney page. Explains how to divide tax refunds, filing tax returns in TX, requirement for filing HOH.
	Page 1. What Getting Divorced or Separated Means for Your 2021 Tax Return. https://www.thebalance.com/what-divorced-or-separated-means-for-taxes-4125740 (retrieved 7/22/2022).	Divorce must be final before end of year for IRS to recognize. Refers to Pub 504; filing as HOH; claiming the children; alimony no longer deductible; paying back taxes and property taxes.
	Page 1. Your Taxes After Divorce. https://www.investopedia.com/taxes-after-divorce-5192868 (retrieved 7/22/2022).	Discusses filing status and qualifying for HOH. Explains rules related to Earned Income Credit (EIC), American Opportunity Tax Credit (AOTC), and child and dependent care credit. After 2018 alimony payments were no longer deducted from taxable income. Gains on the sale of primary home not taxable up to $250K; claiming HOH.
	Page 2. Tax Tips for Women Going Through Divorce. https://www.forbes.com/sites/jefflanders/2012/03/07/tax-tips-for-women-going-through-divorce/ (retrieved 7/22/2022).	Warns about filing joint returns, how to handle over/underpayments, filing HOH, claiming children, child support and alimony considerations, capital gains tax on high ticket assets held for a long time.
	Page 2. Divorced or Separated and Income Taxes. https://www.efile.com/divorce-or-separated-and-taxes/ (retrieved 7/22/2022).	Discusses filing status as HOH, enrolling in health insurance plan and calculating Premium Tax Credit, retirement contributions, handling alimony and child support.

Appendix C. Internet Search Results for "Starting a New Business" Topics

Search term(s)	Results	Comments
Starting a new business	Page 1. How to Start a Business: A Step-by-Step Guide. https://www.businessnewsdaily.com/4686-how-to-start-a-business.html (retrieved 7/13/2022).	Looked like it would have beneficial information. Only mentioned how to apply for an EIN and register business with the state. No discussion of paying taxes.
	Page 1. 10 steps to start your business. https://www.sba.gov/business-guide/10-steps-start-your-business (retrieved 7/13/2022).	Thought a government website would have information on paying taxes. The site explains how to apply for an EIN with IRS and how to get a state tax ID number, but no mention of paying taxes.
	Page 1. Start Your Own Business. https://www.usa.gov/start-business (retrieved 7/13/2022).	Thought a government website would have information on paying taxes. The site explains how to apply for an EIN with IRS and how to get a state tax ID number, but no mention of paying taxes.
	Page 1. How To Start A Small Business In 2022: Complete Step-By-Step Guide. https://www.forbes.com/advisor/business/how-to-start-a-business/ (retrieved 7/13/2022).	One paragraph explaining that it is important to start planning for taxes—income, self-employment, etc. No other mention of paying taxes.
	Page 1. How to Start a Business in 13 Steps. https://www.nerdwallet.com/article/small-business/how-to-start-a-business (retrieved 7/13/2022).	Has a section with links to other tax-related articles.
	>>>Followed link A Tax Guide for Small-Business Owners. https://www.nerdwallet.com/article/small-business/small-business-tax-preparation (retrieved 7/13/2022).	Brief description of Schedule C, Form 1120, Schedule K-1. Recommends contacting a tax professional.
	>>>Followed link 15 Self-Employment Tax Deductions in 2022. https://www.nerdwallet.com/article/taxes/self-employment-tax-deductions (retrieved 7/13/2022).	Provides a list of potential deductions with references to various IRS publications (436—travel expenses, 535—business expenses)
	>>>Followed link How Estimated Quarterly Taxes Work. https://www.nerdwallet.com/article/taxes/estimated-quarterly-taxes (retrieved 7/13/2022).	Explains how to calculate and pay estimated taxes
	Page 2. The Complete, 12-Step Guide to Starting a Business. https://www.entrepreneur.com/article/297899 (retrieved 7/13/2022).	Looked like it would have beneficial information. No discussion of paying taxes.
	Page 2. How to Start a Business From Scratch. https://www.thehartford.com/business-insurance/strategy/how-to-start-a-business (retrieved 7/13/2022).	Looked like it would have beneficial information. No discussion of paying taxes.
	Page 2. How to Start a Business. https://howtostartanllc.com/start-a-business (retrieved 7/13/2022).	Looked like it would have beneficial information. Only mentioned how to apply for an EIN and register business with the state. No discussion of paying taxes.
	Page 2. How to Start a Business: A Guide to Starting a Business. https://www.oberlo.com/blog/how-to-start-a-business (retrieved 7/13/2022).	General information about starting a business (coming up with a business idea, writing a business plan). No information about taxes.

Appendix C (continued). Internet Search Results for "Starting a New Business" Topics

Search term(s)	Results	Comments
	Page 2. Checklist for Starting a Business. https://www.irs.gov/businesses/small-businesses-self-employed/checklist-for-starting-a-business (retrieved 7/13/2022).	IRS.gov
	>>>Followed link Business Taxes. https://www.irs.gov/businesses/small-businesses-self-employed/business-taxes (retrieved 7/13/2022).	Explanation of income tax, estimated tax, self-employment tax, employment taxes
How to start a small business	Page 1. How to Start a Small Business. https://www.adp.com/resources/articles-and-insights/articles/h/how-to-start-a-small-business-a-step-by-step-guide.aspx (retrieved 7/13/2022).	Contained information on how to apply for an EIN with the IRS.
	>>>Followed link How to Do Payroll. https://www.adp.com/resources/articles-and-insights/articles/h/how-to-do-payroll.aspx (retrieved 7/13/2022).	Contained information about calculating payroll taxes and filing forms 941 (quarterly withholding) and 940 (Federal unemployment tax).
	Page 1. How to Start a Business: A Step-by-Step Guide. https://www.businessnewsdaily.com/4686-how-to-start-a-business.html (retrieved 7/13/2022).	Looked like it would be beneficial. Only mentioned how to apply for an EIN and register business with the state. No discussion of paying taxes.
	Page 1. How To Start A Small Business In 2022: Complete Step-By-Step Guide. https://www.forbes.com/advisor/business/how-to-start-a-business/ (retrieved 7/13/2022).	Thought it would have good information. Only mentioned how to apply for an EIN and register business with the state. No discussion of paying taxes.
	Page 2. How to Start a Business. https://howtostartanllc.com/start-a-business (retrieved 7/13/2022).	Thought it would have good information. Only mentioned how to apply for an EIN and register business with the state. No discussion of paying taxes.
	Page 2. How to Start a Small Business at Home. https://www.uschamber.com/co/start/startup/starting-small-business-at-home (retrieved 7/13/2022).	Thought it would have good information. Only mentioned how to apply for an EIN and register business with the state. Encourages hiring a tax professional. No discussion of paying taxes.
	Page 2. How To Start A Business When You Have Literally No Money. https://girlboss.com/blogs/read/start-a-business-without-money (retrieved 7/13/2022).	Female entrepreneur-targeted website. Thought would have good information. No discussion of taxes. Only information about how to apply for an EIN. The site encourages hiring an employment attorney. Encourages visiting SBA.gov
	Page 2. How to Start a Small Business. https://www.wikihow.com/Start-a-Small-Business (retrieved 7/13/2022).	Looked like it would be beneficial. Only recommended hiring an accountant or attorney to help with tax matters. No discussion of paying taxes.
	Page 2. How to Grow a Successful Business. https://www.investopedia.com/articles/pf/08/make-money-in-business.asp (retrieved 7/13/2022).	Thought the site would have information about keeping a business in compliance but there was no information related to taxes.

Appendix C (continued). Internet Search Results for "Starting a New Business" Topics

Search term(s)	Results	Comments
Starting a business and taxes	Page 1. Checklist for Starting a New Business. https://www.irs.gov/businesses/small-businesses-self-employed/checklist-for-starting-a-business (retrieved 7/22/2022).	IRS website with useful information for small business owners
	Page 1. Starting a Business. https://turbotax.intuit.com/tax-tips/small-business-taxes/starting-a-business/L7PBcAdVh (retrieved 7/22/2022).	Information about choosing an accounting method, filing quarterly taxes, paying employment taxes, keeping records for Schedule C, determining whether taxpayer is an employee or independent contractor, keeping track of expenses, and home office deductions
	Page 1. Small Business Tax Information. https://www.usa.gov/business-taxes (retrieved 7/22/2022).	Government website with link to back to IRS.gov
	Page 1. Tax and Business Forms You'll Need to Start a Small Business. https://www.businessnewsdaily.com/9-tax-and-business-forms-needed-to-start-a-small-business.html (retrieved 7/22/2022).	Contained guidance on paying estimated taxes, paying employment taxes if a company has employees, completing tax Form Schedule C, and information about business tax deductions.

Appendix D. Internet Search Results for "Gig Economy" Topics

Search term(s)	Results	Comments
Driving for Uber	Page 1. Flexible driving opportunities with Uber. https://www.uber.com/us/en/drive/ (retrieved .7/14/2022)	Main recruiting page for Uber. Describes benefits of driving with the company. No information about taxes.
	Page 1. Uber Driver Requirements: A Step-by-Step Guide. https://www.investopedia.com/articles/personal-finance/120315/how-become-uber-driver-step-step-guide.asp (retrieved 7/14/2022).	Subheading "How Do Uber Drivers Pay Taxes?" Explains self-employment tax rate of 15.3% (Medicare and SS). Advises to talk with a CPA about what can be written off as expenses.
	Page 1. I'm a driver for Uber and Lyft — here are 10 things I wish I knew before starting the job. https://www.businessinsider.com/uber-lyft-drivers-job-advice-car-2019-8 (retrieved 7/14/2022).	Looked like it would be helpful but did not mention anything about taxes or tax planning.
	Page 1. How to Become an Uber Driver: A Beginner's Guide. https://www.nerdwallet.com/article/finance/how-to-become-an-uber-driver (retrieved 7/14/2022).	Reminds gig workers that they are responsible for 12.4% to Social Security and 2.9% to Medicare, for a total of 15.3%
	>>> Followed link to "What Gig Workers Need to Know About Taxes." https://www.nerdwallet.com/article/finance/what-gig-workers-need-to-know-about-taxes (retrieved 7/14/2022).	Brief mention of deducting retirement plan contributions. Warns about underpayment penalties. Encourages setting up a payment plan if one cannot afford to pay their taxes.
	Page 2. Is Driving for Uber Worth It in 2022? https://millennialmoneyman.com/driving-for-uber/ (retrieved 7/14/2022).	Brief comment about self-employment taxes. Encourages downloading mileage tracking applications.
	Page 2. Make Money Driving For Uber: The Ultimate Side Hustle. https://www.goodfinancialcents.com/how-to-become-an-uber-driver-requirements/ (retrieved 7/14/2022).	Reminds drivers they are responsible for (1) their tax bill, including paying quarterly tax payments if applicable; (2) keeping track the mileage; and (3) collecting receipts for gas and vehicle upkeep.
	Page 2. How Much Do Uber Drivers Make? Is It Worth Your Time? https://www.gobankingrates.com/money/side-gigs/how-much-do-uber-drivers-make/ (retrieved 7/14/2022).	Informs drivers they will need to pay self-employment taxes.
	>>> Followed link to How Does an Independent Contractor Pay Taxes? https://www.gobankingrates.com/taxes/filing/independent-contractor-taxes/ (retrieved 7/14/2022).	Talks about self-employment taxes.
	>>> Followed link to Got a Side Hustle? Here's How To Calculate Estimated Taxes. https://www.gobankingrates.com/taxes/filing/deadline-countdown-self-employment-guide-filing/ (retrieved 7/14/2022).	Explains when and how to file quarterly estimated taxes.
Driving for Lyft	Page 1. It pays (a lot) to drive right now. https://www.lyft.com/drive-with-lyft (retrieved 7/14/2022).	Main recruiting page for Lyft. Describes benefits of driving with the company. No information about taxes.

Appendix D (continued). Internet Search Results for "Gig Economy" Topics

Search term(s)	Results	Comments
	Page 1. How much can you make driving for Lyft in Atlanta? https://www.quora.com/How-much-can-you-make-driving-for-Lyft-in-Atlanta (retrieved 7/14/2022).	Discussion board about driving for Lyft
	>>>Searched in Quora "Lyft+Taxes." https://www.quora.com/search?q=lyft%20taxes (retrieved 7/14/2022).	General information about taxes. Most post refer the reader to IRS.gov
	Page 2. How Much Do Lyft Drivers Make? https://www.gobankingrates.com/money/side-gigs/how-much-do-lyft-drivers-make/ (retrieved 7/14/2022).	Thought the site would have information regarding filing and reporting taxes related to earnings. No information about taxes.
	Page 2. Your Step-By-Step Guide to Becoming a Lyft Driver [2022 Update]. https://www.ridester.com/drive-for-lyft/ (retrieved 7/14/2022).	Site offers information only on how to initially get set up with Lyft. No information about taxes.
	Page 2. Lyft vs. Uber: What's the Difference? https://www.investopedia.com/articles/personal-finance/010715/key-differences-between-uber-and-lyft.asp (retrieved 7/14/2022).	Thought this would have information since it appeared under Investopedia's "Personal Finance" section. No information about taxes.
	>>>Followed link to "Gig Economy" https://www.investopedia.com/terms/g/gig-economy.asp (retrieved 7/14/2022).	Provides basic information about the gig economy. No information about taxes.
	Page 2. How to Become a Lyft Driver. https://gigworker.com/become-lyft-driver/ (retrieved 7/14/2022).	Site offers information only on how to initially get set up with Lyft. No information about taxes.
Gig Economy	Page 1. Gig Economy. https://www.investopedia.com/terms/g/gig-economy.asp (retrieved 7/14/2022).	Provides basic information about the gig economy. No information about taxes.
	Page 1. Thriving in the Gig Economy. https://hbr.org/2018/03/thriving-in-the-gig-economy (retrieved 7/14/2022).	Provides general information. No mention of taxes.
	Page 2. What is the gig economy and what's the deal for gig workers? https://www.weforum.org/agenda/2021/05/what-gig-economy-workers/ (retrieved 7/14/2022).	Provides a definition of the gig economy. No information about taxes.
	What is the Gig Economy? The Complete Guide for 2022. https://www.oberlo.com/blog/what-is-the-gig-economy (retrieved 7/14/2022).	Provides pros and cons to the gig economy. No information about taxes.
Independent Contractor	Page 1. Independent Contractor (Self-Employed) or Employee? https://www.irs.gov/businesses/small-businesses-self-employed/independent-contractor-self-employed-or-employee (retrieved 7/14/2022).	IRS.gov. See IRS search results.
	Page 1. Independent Contractor. https://www.investopedia.com/terms/i/independent-contractor.asp (retrieved 7/14/2022).	Provides independent contractor information about their self-employment tax responsible for Social Security and Medicare. Brief mention of deducting retirement plan contributions. Lists common tax forms for independent contractors (Schedule C, 1040, 1040-ES)

Appendix D (continued). Internet Search Results for "Gig Economy" Topics

Search term(s)	Results	Comments
	Page 1. Minimum Requirements for Working as an Independent Contractor https://www.nolo.com/legal-encyclopedia/minimum-requirements-working-independent-contractor-29978.html (retrieved 7/14/2022).	A section of the article describes registering for an EIN with IRS and completing a Schedule C.
	Page 1. Understanding What an Independent Contractor Is. https://www.businessnewsdaily.com/15853-independent-contractor-employee-differences.html (retrieved 7/14/2022).	Describes the definition of an independent contractor. No information about filing taxes.
	Page 1. Employee or Independent Contractor? UGA. https://www.georgiasbdc.org/employee-or-independent-contractor/ (retrieved 7/14/2022).	Describes the IRS definition of an independent contractor and refers the reader to Pub 15a (link was broken http://www.irs.gov/pub/irs-pdf/p15a.pdf) Also refers readers to SBA for a definition of independent contractor. No tax information provided.
	Page 2. What is an Independent Contractor? https://andersonadvisors.com/independent-contractor/ (retrieved 7/14/2022).	Explains that independent contractors have to pay their own Social Security taxes. Explains the requirements for issuing 1099s
	Page 2. Fair Labor Standards Act Advisor—Independent Contractor. https://webapps.dol.gov/elaws/whd/flsa/docs/contractors.asp (retrieved 7/14/2022).	DOL.gov. Defines an independent contractor. No tax information provided.
	Page 2. What Is An Independent Contractor? Here's Why It Matters Under the Trump Tax Law. https://www.forbes.com/sites/alangassman/2018/10/05/what-is-an-independent-contractor/?sh=37b9c2871692 (retrieved 7/14/2022).	Describes the IRS "Pass Through Deduction" (Tax Cuts and Jobs Act). No information about filing taxes.
Driving for Uber and taxes	Page 1. How Do Rideshare (Uber and Lyft) Drivers Pay Taxes? https://www.taxoutreach.org/rideshare/how-do-rideshare-uber-and-lyft-drivers-pay-taxes-2/ (retrieved 7/22/2022).	Discussed self-employment taxes, tracking deductions, filing quarterly taxes, and completing a Schedule C and Schedule SE.
	Page 1. Tax Deductions for Rideshare (Uber and Lyft) Drivers and Food Couriers. https://www.taxoutreach.org/rideshare/tax-deductions-for-rideshare-uber-and-lyft-drivers/ (retrieved 7/22/2022).	Explains standard vehicle deductions and provided an infographic on choosing the Standard Mileage Deduction vs. Actual Expenses. The site provides a sample of a 1099 and how to complete a Schedule C.
	Page 1. Tax Tips for Uber Driver-Partners: Understanding Your Taxes. https://turbotax.intuit.com/tax-tips/self-employment-taxes/tax-tips-for-uber-drivers-understanding-your-taxes-/L7sbLCSc4 (retrieved 7/22/2022).	Contains information on 1099s, being self-employed, completing Schedule C, deducting mileage, and examples of other tax-deductible business expenses.
	Page 1. Your tax questions, answered https://www.uber.com/us/en/drive/tax-information/ (retrieved 7/22/2022).	Provides a brief explanation of tax documents and promotes the use of TurboTax.
	Page 1. The Uber & Lyft Driver's Guide to Taxes https://bench.co/blog/tax-tips/uber-driver-taxes/ (retrieved 7/22/2022).	Discusses self-employment taxes, filing quarterly estimates, and completing Schedule C. Offers a subscription service to maintain tax records.

Appendix D (continued). Internet Search Results for "Gig Economy" Topics

Search term(s)	Results	Comments
	Page 2. 5 Things to Know About Rideshare Driver Taxes https://www.morningbrew.com/daily/stories/2021/04/22/5-things-know-rideshare-driver-taxes (retrieved 7/22/2022).	Explains Form 1099, the concept of "ordinary and necessary," vehicle deductions, and the Qualified Business Income (QBI) deduction.
	Page 2. Tax Tips for Lyft and Uber Drivers: What to Know for 2021. https://www.picnictax.com/blog/lyft-uber-rideshare-driver-taxes/ (retrieved 7/22/2022).	Contains information on 1099s, being self-employed, completing Schedule C, and deducting mileage.
	Page 3. Uber Tax Information: Essential Tax Forms, Documents, & Checklists. https://www.ridester.com/uber-tax-information/ (retrieved 7/22/2022).	Discussed self-employment taxes, tracking deductions, filing quarterly taxes, and completing a Schedule C and Schedule SE.
	Page 3. Learn How to File Taxes for Uber and Lyft Drivers https://www.udemy.com/course/learn-how-to-file-taxes-from-uber-lyft/ (retrieved 7/22/2022).	Offers a course for $39.99 on how to file taxes for Uber and Lyft drivers.
	Page 3. Taxes for Rideshare/Uber Drivers https://www.solvable.com/tax-help/business-taxes/taxes-for-rideshare-uber-drivers/ (retrieved 7/22/2022).	Provides information on acting as an independent contractor, filing appropriate tax forms, completing Schedules C and SE, filing quarterly estimated taxes.

Appendix E. Search Results for Offline Resources

Search term(s)	Results	Comments
Divorce support groups	Page 1. Divorce Support Groups—Atlanta https://www.psychologytoday.com/us/groups/ga/atlanta?category=divorce (retrieved 7/18/2022).	Support group and psychologist referrals service.
	Page 1. Oasis—Buckhead Church—Divorce Recovery https://buckheadchurch.org/oasis (retrieved 7/18/2022).	Atlanta-based church that offers divorce support groups.
Divorce attorney organizations	Page 1. American Academy of Matrimonial Lawyers. https://www.aaml.org/ (retrieved 7/18/2022). >>>Searched AAML Journal "taxes." Article: Divorce and Taxes: Fifty Years of Change Volume 24, 2012, Number 2, p. 489 https://aaml.org/resource/collection/3BDEDFA9-B18B-4C53-B875-2CF630D-DAD9C/Wilder.pdf (retrieved 7/18/2022).	Article discussing filing status rules.
Divorce workshops	Page 1. Second Saturday. https://www.secondsaturday.com/ (retrieved 7/19/2022). >>>Click "Find a Workshop" >>>Choose State	Provides a list of online in-person workshops for divorced people.
	Page 1. Family Law Workshop Information https://www.fultoncourt.org/family/family-workshop.php (retrieved 7/19/2022).	Fulton County (Georgia) Court provides workshops for people going through/considering divorce.
Tax advice for small businesses	No valuable results	
Tax workshops	Page 1. Small Business Tax Workshops, Meetings and Seminars https://www.irs.gov/businesses/small-businesses-self-employed/small-business-tax-workshops-meetings-and-seminars (retrieved 7/19/2022).	IRS.gov website. Taxpayer chooses their state for a list of tax workshops in their area.
	Page 1. TaxworkShop.com. https://www.taxworkshop.com/ (retrieved 7/19/2022).	Workshops for tax practitioners. Last scheduled workshop was September 2021.
Gig economy workshops	No valuable results.	

The Impact of Annual Changes in Family Structure and Income on Tax Credits

Elaine Maag, Nikhita Airi, and Lillian Hunter (Urban-Brookings Tax Policy Center) [1]

I. Introduction

Refundable tax credits, those that can exceed federal income taxes owed, provide an important source of financial support for many low- and middle-income families with children. The largest of these are the Earned Income Tax Credit (EITC) and Child Tax Credit (CTC). The credits are determined on an annual basis and are often received as a single payment as part of a family's tax refund. For low-income families, it is often the family's most significant financial event of the year (Morduch and Schneider (2017)).

Tax credits accrue to tax units—the group of individuals who appear on a tax return together based on legal relationships, child residency, and support. Though families may change throughout the year, only one adult or married couple will likely be able to benefit from the EITC and CTC for any one child (and often it is the same person for both credits), even when several adults provide significant amounts of support to a child throughout the year.

Because most families file taxes once a year, after the tax year has ended, families that change throughout the year may have difficulties correctly determining their filing status and who can properly claim a child for the purpose of receiving child-related benefits. According to the Internal Revenue Service (IRS), the most common error filers make when claiming the EITC is claiming a child who is not actually a qualifying child (IRS (2022)).

Families often report that the amount of tax credits they receive at tax time are a surprise (T. Anderson *et al.* (2022); Romich and Weisner (2000)). The complexity of the credits along with family and income changes throughout the year may contribute to not knowing what credits are likely to be delivered at tax time (Maag *et al.* (2017); Maag *et al.* (2016)).

This analysis reviews how trends in changing family structures diverge from how tax credits are delivered. We then briefly describe the EITC and CTC and the role income and family composition play in their calculations. Finally, we explore how well data in one year predict EITC and CTC receipt in a subsequent year.

Understanding the predictability of tax credits is important for two reasons. First, because the EITC and CTC are significant sources of economic support for families, it is important to gain a better understanding of how much credit amounts vary from year to year. Second, recent experience with advancing up to half of the CTC in 2021 has re-energized calls for delivering tax credits on a monthly basis in advance of families filing a tax return.

If credits are advanced based on information from the prior year, families that experience drops in credits for which they are eligible may be required to repay any credit they received in error when they complete their tax return. This may create a hardship for some families. Uncertainty about tax credits can cause tax filers to

[1] This brief was funded by Intuit Financial Freedom Foundation and the Annie E. Casey Foundation through the Innovations in Cash Assistance for Children initiative. The authors are grateful to them and to all our funders, who make it possible for the Urban-Brookings Tax Policy Center to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban-Brookings Tax Policy Center, the Urban Institute, the Brookings Institution, their trustees, or their funders.

The authors gratefully thank Joyce Morton, who developed and created the datasets underlying this analysis, Kevin Werner, who performed microsimulation modeling runs, and Laura Wheaton who advised the construction of the analysis file and reviewed an earlier draft of the paper. We also thank Donald Marron, Tracy Gordon, Krista Holub, and Charlie Leonard for thoughtful comments on an earlier draft, Alexandria Dallman for editorial review, and Muskan Jha for publication support. TRIM3 is maintained and developed by the Urban Institute, under primary funding from the Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation (HHS/ASPE).

take unnecessary precautions, such as borrowing less ahead of receiving tax refunds, decreasing the overall benefits of tax credit programs (Caldwell *et al.* (2023)). Understanding to what extent prior-year data can be useful in predicting eligibility is critical to designing an advanced payment program that can sufficiently protect families from potentially needing to pay back credits determined to be in error.

On the other hand, people who find out they are eligible for larger credits when they fill out their tax return than predicted by prior-year data may miss out on the full impact of having advance credits. Because both issues are likely to affect low-income families more, we focus most of our analysis on families with income below 200 percent of the federal poverty level.

Among low-income families, 39 percent see the amount of their EITC drop at least $500 from one year to the next, 39 percent see their EITC change by less than $500 (almost half of this group receives no EITC in either year), and the remaining 22 percent see their EITC rise by at least $500. CTC amounts are more stable from one year to the next. About 49 percent, see their CTC change by less than $500, 20 percent see their credit drop by at least $500, and the remaining 31 percent see their CTC rise by at least $500.

Year-over-year changes in income drive much of the change in credits, though changes in the number of children and marital status drive some credit change. Almost all benefits from the EITC accrue to families with children in the bottom 40 percent of the income distribution. Initially, benefits phase in with earnings. Once earnings reach about $10,000 for families with one child or about $14,000 for families with at least two children, they remain flat as income increases. Benefits begin to phase out once income increases beyond about $19,000. Among the 39 percent of low-income families that see their EITC decrease from one year to the next, almost three-quarters experience a decrease in their EITC because their income rises. In other words, their credit begins to phase out or phases out completely because they have more income in the second year we observe them than in the first year.

In contrast, the CTC delivers benefits to all but the highest income families with children. Benefits generally increase with earnings until the maximum benefit is reached. A single parent with two children needs about $32,000 in income to receive the full CTC benefit. The Tax Policy Center estimates that 19 million children under age 17 live in families that do not receive the full $2,000 per child CTC benefit because their families do not earn enough.[2]

The credit does not begin to phase out until income reaches $200,000 for single parents and $400,000 for married couples. As a result, low-income families rarely lose credits because their income increases. Because the credit initially rises with earnings, low-income families often qualify for higher credits when their earnings increase from one year to the next. For the 31 percent of families that see their CTC rise by at least $500, about two-thirds of the time that increase is driven by increases in income. Families move from receiving no CTC or only part of the $2,000 per child credit to receive more or all of the credit.

If the IRS were to advance credits based on prior year filing information, Congress would need to consider how accurate advance payments are likely to be and what sort of protections should reasonably be put in place (and how many people would likely need those protections). Although income changes drive most changes, families that change throughout the year because of marriage, divorce, or change in the number of children in the tax unit are most likely to see large (at least $2,000) CTC swings—and these types of families are becoming a greater share of all families with children. Program administrators should consider the larger context of changes to the American family when designing policy going forward.

II. Changes to the American Family

The tax system was designed at a time when marriage rates were high and children tended to grow up in families with two biological parents. By 2019, 41 percent of children lived in a household arrangement other than with two married biological parents (L. Anderson *et al.* (2022)). The decline in marriage rates alongside an increase in births outside of marriage is also reflected in tax data. In 1962, 59 percent of all tax returns were

2 "Distribution of Tax Units and Qualifying Children by Amount of Child Tax Credit (CTC), 2022," table T22-0123, Tax Policy Center, October 18, 2022, https://www.taxpolicycenter.org/model-estimates/children-and-other-dependents-receipt-child-tax-credit-and-other-dependent-tax.

filed by married couples. By 2018, the share of tax returns filed by married couples dropped to 35 percent. Over the same period, the share of returns filed by single parents with custody of their children (head of household) increased from about 2 percent to over 10 percent, with single filers without children on their tax return making up almost all of the rest of the filing population (CBO (2019)).

By far, the most common filing status associated with receipt of the EITC is head of household. In 2017, nearly 56 percent of EITC claimants were unmarried filers with children (National Taxpayer Advocate (2020), p. 52), or people who typically file as head of household.

Child custody is shared in a growing number of cases, both in households headed by single parents and by married couples where at least one partner has an additional child outside their current marriage. Sharing custody complicates the process of determining who should receive the CTC and EITC on behalf of that child. More than one parent or caregiver may reasonably feel entitled to the credit, even if the law does not define them as eligible. In recent years, over half of divorces have resulted in shared custody agreements (Meyer *et al.* (2017)). Children from lower-income families are more likely to live in families with tax filing ambiguities that complicate their ability to claim tax credits: as many as 60 percent of lower-income families, compared to 40 percent overall (Michelmore and Pilkauskas (2022)).

III. Description of Earned Income and Child Tax Credits

The EITC and CTC together lift more children out of poverty than any other income support program in a typical year (Fox and Burns (2020)). Benefits from the EITC are concentrated among low- and moderate-income families, while benefits from the CTC cover almost every family with children. We describe each credit's structure briefly to understand better why it may be difficult to predict the credit eligibility in advance of filing a tax return.

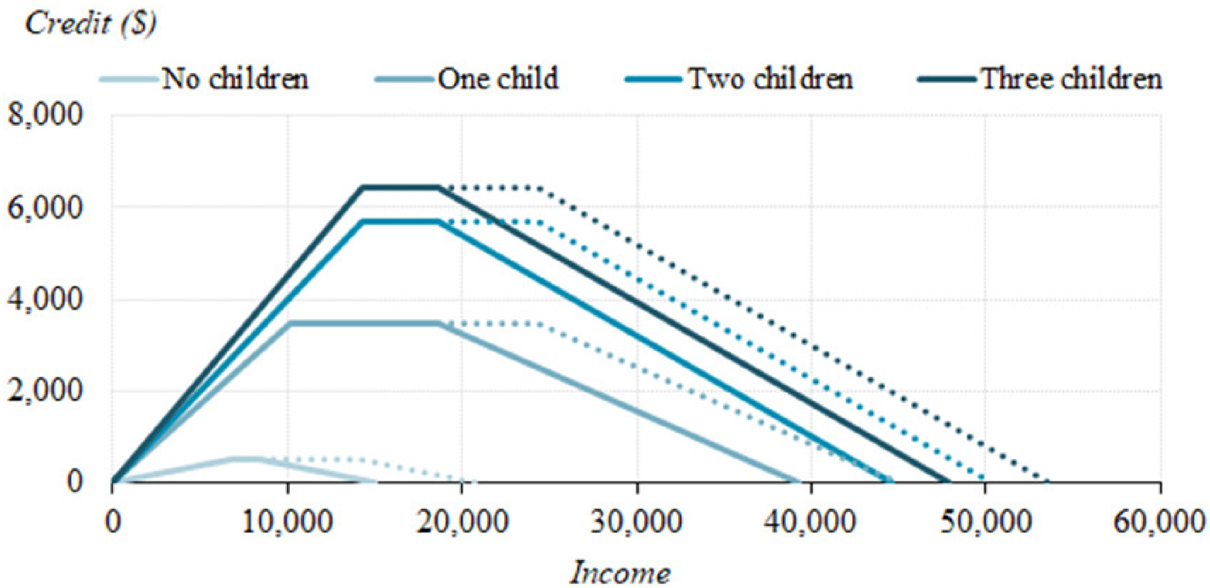
A. *Earned Income Tax Credit*

The EITC provides substantial support to low- and moderate-income working parents. Workers receive a credit equal to a percentage of their earnings up to a maximum credit (Figure 1). Both the credit rate and the maximum credit vary by family size, with larger credits available to families with more children. From 2015 to 2018, the years of our analysis, the maximum credit for families with one child varied from \$3,359 to \$3,461, while the maximum credit for families with three or more children varied from \$6,242 to \$6,431. A much smaller credit is available to some workers without children living at home (about \$500). After the credit reaches its maximum value, it remains flat until income reaches the point where the credit begins to phase out. Thereafter, it declines with each additional dollar of income until no credit is available (Figure 1).[3] The EITC is a refundable tax credit—if a family qualifies for a credit worth more than the taxes they owe, they may receive it as a tax refund. Each year, the credit grows with inflation.

In cases where a child is supported by people in more than one tax unit, the tax unit where the child lives for the majority of the year is the intended beneficiary of the EITC. In a multigenerational household, a parent has the option to claim the child, the child's grandparent in the household can claim the child if that grandparent has a higher income than the child's parent. One tax unit may benefit from the EITC on behalf of a child and generally the same tax unit will also benefit from the CTC. If two parents of a child cohabit—live together without marrying—they may choose which parent will claim the child. If both cohabiting parents claim the child on a tax return, the one with the higher income will be determined eligible.

[3] The EITC begins to decrease whenever a family's earnings or adjusted gross income, whichever is higher, exceeds the phaseout threshold.

FIGURE 1. Earned Income Tax Credit, 2018



Source: Urban-Brookings Tax Policy Center calculations.
Notes: Assumes all income comes from earnings. Assumes children meets all tests to be EITC-qualifying children. Dotted lines represent married couples. All credit amounts are indexed annually for inflation.

The group of people who benefit from the EITC is not stagnant. Prior analysis using tax data showed that over a 10-year period, 61 percent of claimants claimed the EITC for one or two years and about 20 percent of EITC recipients claimed the credit for more than five years (Dowd and Horowitz (2011)). Credit eligibility relies on both the income of the taxpayers and the composition of the tax unit. A tax filer needs to know who will live in their household, their marital status, and taxpayers' income to anticipate their EITC.

Understanding who should claim a child for the EITC creates confusion. The Treasury Department has estimated that 70 percent of all improper payments of the EITC stem from the incorrect person claiming the child for credit purposes (Department of the Treasury (2018); Holtzblatt and McCubbin (2002)). The IRS indicates that another two of the five most common errors with respect to claiming the EITC are claiming a child that does not qualify for the benefit and more than one person claiming the child (IRS (2022)). Determining whether an advance credit should be based on the presence of a child will presumably also be difficult for families.

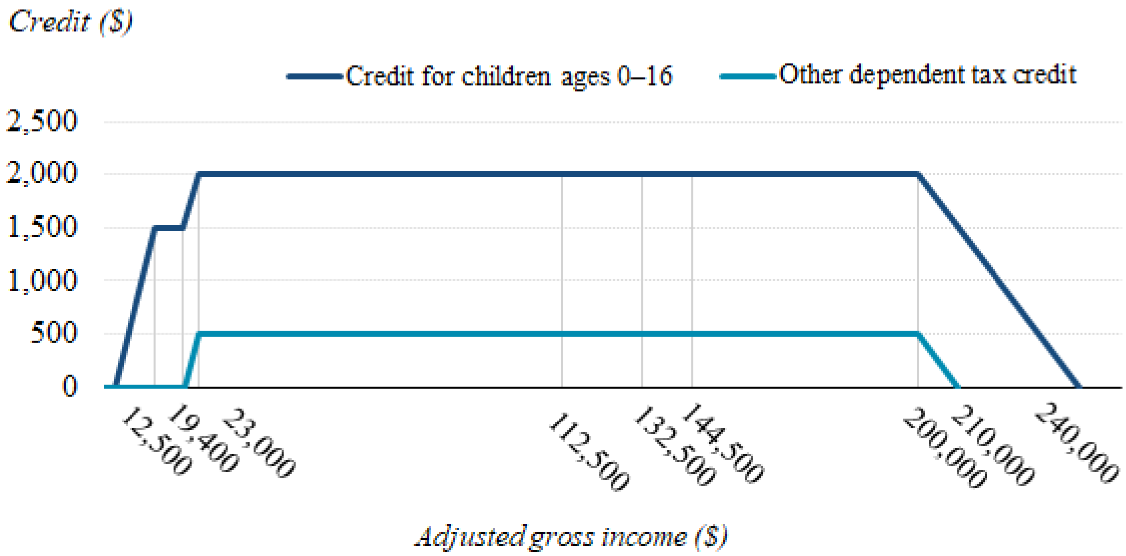
B. Child Tax Credit

The CTC offsets part of the cost of raising children for working families. Expanded as part of the Tax Cuts and Jobs Act of 2017 (TCJA), the CTC provides a benefit of up to $2,000 per child under age 17 (Figure 2). After first being used to offset taxes owed, part of the CTC can be received as a tax refund. The refundable portion of the credit is calculated as 15 percent of earnings over $2,500. Prior to 2021,[4] the refundable portion of the credit was limited to $1,400 per child. How much of the credit can be received as a refund is the only CTC parameter indexed for inflation. In 2022, the refundable portion rose to $1,500. Over 90 percent of families with children benefit from the CTC.[5]

[4] In 2021, the American Rescue Plan Act created a temporary expansion of the CTC, making the credit fully refundable. It also increased the size of the credit to up to $3,600 per child up to age 5 and up to $3,000 per child ages 6 to 17. For more information see https://taxpolicycenter.org/briefing-book/what-child-tax-credit.

[5] "Tax Benefit of the Child Tax Credit (CTC) Current Law, by Expanded Cash Income Percentile, 2022," table T21-0225, Tax Policy Center Microsimulation Model (version 0721-1), September 2021, https://www.taxpolicycenter.org/model-estimates/tax-benefits-child-tax-credit-september-2021/t21-0225-tax-benefit-child-tax-credit.

FIGURE 2. Child Tax Credit, Single Parent with One Child, 2018



Source: Urban-Brookings Tax Policy Center calculations.
Notes: Assumes all income comes from earnings. Assumes child meets all tests to be a CTC-qualifying dependent. Credit phases out beginning at $400,000 of income for married parents. Only children with Social Security numbers qualify for the CTC. Noncitizens under age 18 who meet the dependency tests for eligibility can qualify for the other dependent tax credit.

In addition to the refundable and nonrefundable portions of the CTC, there is also a credit for other dependents (ODTC). This credit is worth up to $500 and can only be used to offset taxes owed. Generally, the credit is available to families with dependents who do not qualify for the CTC. Dependents of any age can qualify for the ODTC this includes children aged 17 or 18, full-time college students up to age 24, and children who do not have Social Security numbers.

Claiming the CTC is less studied than the EITC. Divorced and never married parents may alternate years of claiming a child, regardless of where the child lives the majority of the year. Because the rules for claiming a child are less strict for the CTC than the EITC, there are likely fewer errors in who claims the credit. For example, a child does not have to live with the claiming parent for a given number of months for the parent to claim the child—but only one parent (or other relative) may claim the child each year.

IV. Why Do EITC and CTC Amounts Change From One Year to the Next?

A. *EITC*

EITC amounts depend on three main characteristics of the tax unit: the number of eligible children, earnings and income, and marital status.[6] EITC amounts increase annually with inflation, so even with no other changes, many families will see their EITC increase from one year to the next. In the years of our study, these changes were modest, causing the maximum credit for a family with one child to grow $102—from $3,359 to $3,461.

If the number of qualifying children in a tax unit changes, a family's EITC will also likely change. The number of eligible children can increase from birth, adoption, or other arrival of a new child. The number of eligible children can decrease if a child moves to another home for more than half the year, turns 17 during the year, becomes the qualifying child of another tax unit in a household with cohabiting parents or multiple generations, or dies. Credit amounts increase for each additional child up to three. Increases beyond three have no

[6] To a lesser extent, EITC amounts depend on investment income (which in 2018 could not exceed $3,500) and a variety of other qualifying characteristics. For example, married couples must file a joint return; the taxpayer and spouse (if applicable) must have SSNs valid for work as do any qualifying children; taxpayers cannot claim a foreign earned income exclusion or be the qualifying child of another person. Taxpayers without qualifying children have additional restrictions. Internal Revenue Service, 2019. *IRS Publication 596, Earned Income Credit (EIC) For use in preparing 2018 Returns.* Washington, DC. Department of the Treasury.

effect and decreases above three have no effect. In many cases, a family will be able to predict these changes are coming the next tax year, but not in all cases. Families will not necessarily know how changes in the number of children will affect their benefit. In general, a change in the number of children is the most dramatic effect. In 2018, increasing from no children to one child increases the maximum credit from $519 to $3,461. Increasing from two to three children increases the benefit by a max of $715. There are no further adjustments for children beyond the third.

Earnings change from one year to the next for a variety of reasons. These include changes in wage rates, changes in the number of hours worked, changes in bonus income, changes in jobs, irregular schedules, moving in and out of the labor market. Prior research shows that among low-income families, those with income below twice poverty, almost two-thirds have income that for at least one month of the year will spike above or dip below their average monthly income by at least 25 percent (Maag *et al.* (2017)). Earnings can also change when marital status changes because the tax unit will now include income from both partners in the couple for a newly married couple or only one partner from the couple in the case of a divorce. How the EITC changes with income depends on whether a family has income in the phase-in period of the credit, the range where the credit delivers a flat benefit, or in the phase-out range of the credit (Williams and Maag (2008)).

Marital status changes when people marry, divorce, or become widowed. In the case of the EITC, married couples can earn more income before the credit begins to phase out than single people, so changing marital status can change credit amounts—even if income amounts do not change. In particular, a couple may be able to receive the maximum credit rather than have it partially phase down with their additional income or may be able to receive a higher credit amount if they are in the phase-out range of the credit than when they were single.

B. CTC

CTC amounts depend on the number of children in the family and to a lesser extent, earnings and income amounts. The maximum benefit does not change annually with inflation, but instead is set at $2,000 per child under age 17 until 2025, at which point it will drop to $1,000.

There is no maximum number of children that can benefit. The reasons for child changes in the CTC can be the same as for the EITC, but unlike the EITC, only specific changes inform a change in the number of children claimed for the CTC. In the case of the CTC, unmarried parents can designate who will claim the credit. It is not necessary that the child pass the same residency test required by the EITC. In some cases, for example, parents who do not live together have made an agreement to shift who claims the child annually. This change is predictable for those parents. In other cases, decisions on who will claim the CTC may be made on an annual basis and would not be any easier to predict than child residency.

The CTC phases out at relatively high-income levels ($200,000 for single parents, $400,000 for married couples). Low- and moderate-income families may see their credit increase if earnings increase—or they may see their credit decrease if earnings decrease. Unlike the EITC, they are unlikely to experience a credit decrease when earnings increase because of the relatively high point at which the CTC begins phasing out. In 2018, about 2 percent of children received no credit because their parents did not earn enough. Only increases in earnings can change their credit. About 12 percent had earnings too low to be eligible for the full credit—an earnings increase could increase their credit and a decrease could decrease their credit. Over 60 percent of children received the full credit and the vast majority would be unaffected by modest changes in earnings.[7]

The credit begins to phase out at double the income level for married couples as single parents. Changes in marital status may affect the credit amount but are unlikely to be a large factor—except to the extent that parents with low income marry low or moderate earners, increasing their tax unit's total income.

[7] "Distribution of Tax Units and Qualifying Children by Amount of Child Tax Credit (CTC), 2018," table T17-0228, Tax Policy Center, October 18, 2017, https://www.taxpolicycenter.org/model-estimates/distribution-amount-child-tax-credit-october-2017/t17-0228-distribution-tax-units.

V. Data and Methods

We use the Annual Social and Economic Supplement of the Current Population Survey (CPS ASEC) to estimate year-over-year eligibility for tax credit. The CPS ASEC collects data on a representative sample of households throughout the year on a monthly basis. Households are in the survey for four consecutive months, are out of the survey for the next eight months, and then return to the survey for another four months before leaving the sample permanently. The design means that some households will be in the survey for two consecutive years in March, the month that income data are collected, which can be used to estimate taxes including refundable tax credits. We use this feature of the survey to follow households with at least one child under age 18 who appear in the survey in two consecutive years. We use the Transfer Income Model, version 3 (TRIM3), to estimate changes in the EITC and CTC from one year to the next.[8] We exclude from our sample households where income was imputed because imputations are not designed to show changes from one year to the next.

Our analysis uses data from 2016 through 2019, which represent income amounts from 2015 through 2018. We pair observations in 2015 and 2016, 2016 and 2017, and 2017 and 2018. We apply 2018 tax law in all years: our calculations were therefore unaffected by the Tax Cuts and Jobs Act's changes to the CTC that went into effect in 2018. A household must have a child in at least one year to be part of our sample.[9]

VI. Results

We compare the EITC and CTC that a tax unit appears eligible for in year two with the credit they appear eligible for in year one.[10] All of the families in our analysis have a child in either the first or second year they are observed. For each credit, we group tax units into those with an increase in the credit of at least $500 from year one to year two, those that have a change of less than $500, and those with a credit decrease of at least $500 between the two years. Low-income families are defined as those with incomes beneath 200 percent of the official poverty measure.

A. Earned Income Tax Credit

Not all families are eligible for an EITC in both years. In our data, about 40 percent of families with children receive an EITC in at least one year and 60 percent receive no EITC in both years. We estimate that 16 percent of families experience a drop in their EITC of at least $500, over two-thirds of families (69 percent) have no major change in eligibility, and the remaining 15 percent of families appear eligible for an EITC that is at least $500 larger in year two than in year one (Figure 3).

Among low-income families with children, those with income below twice the poverty level in year one, we observe that 79 percent receive an EITC in at least one year and 21 percent receive no EITC in both years. We estimate that 39 percent of families experience a drop in their EITC of at least $500, some 39 percent have no major change in eligibility, and the remaining 22 percent see their credits increase by at least $500.

Income changes drive earned income tax credit changes and large earned income tax credit changes are most common

If a family's income is in the phase-in range of the credit, a decrease in income results in a year-over-year decrease in the EITC. A sufficiently large decrease in income from the plateau range of the credit can also

[8] Information presented here is derived in part from the Transfer Income Model, version 3 (TRIM3), and associated databases. TRIM3 requires users to input assumptions and/or interpretations about economic behavior and the rules governing federal programs. Therefore, the conclusions presented here are attributable only to the authors of this report.

[9] Our data do not allow us to implement the rules that all persons in the tax unit must have a Social Security number (SSN) eligible for work to be eligible for the EITC. Although TRIM3 models SSN requirements, the TRIM3 imputation of whether a person has an SSN is not necessarily consistent in two consecutive years of CPS data and so we do not use those imputations for this analysis. We allow the EITC parameters to adjust with inflation but deliver a maximum CTC benefit of $2,000 per child with up to $1,400 allowed as a refund, consistent with 2018 law.

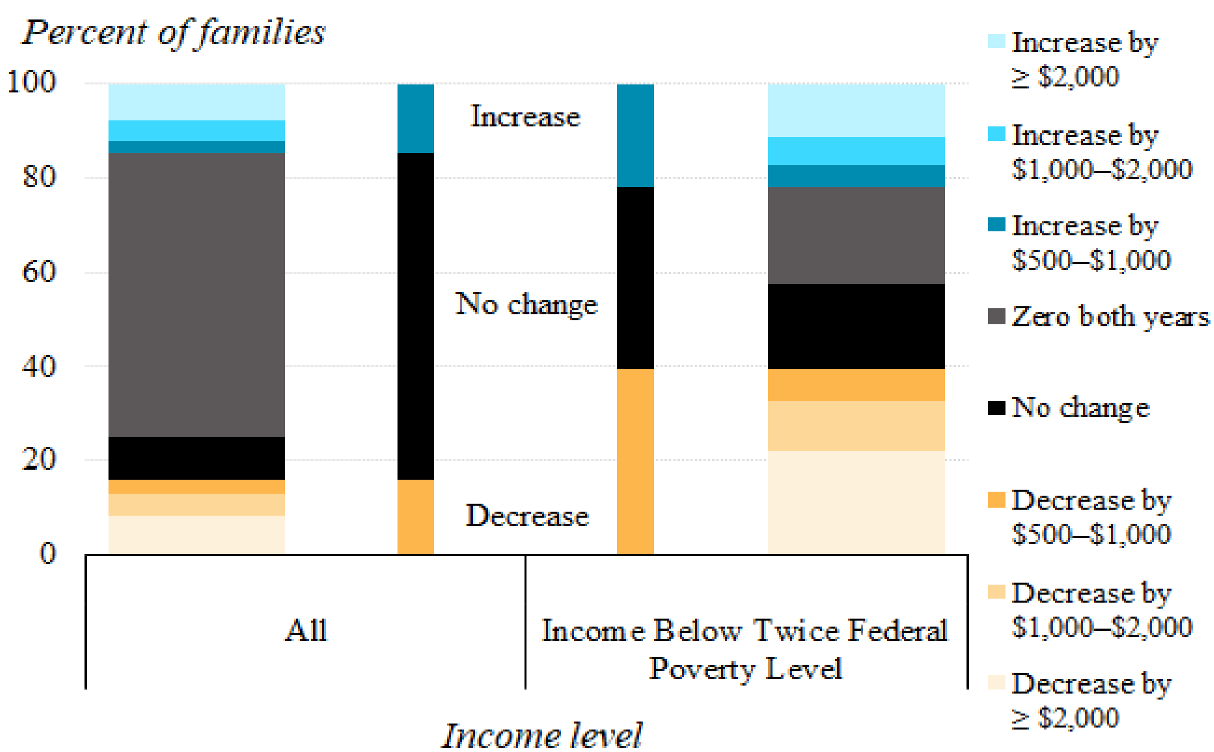
[10] Because the CPS is a household survey, we cannot track people who move households. We can track changes that happen to a tax unit if the tax unit stays in the same household. For example, if a couple divorces, and one partner remains in the household, we can compare the EITC and CTC the partner who stayed in the household qualified for in year two with the EITC and CTC the partner was eligible for as part of a couple in year one. In this way, our estimates likely overstate stability in the tax credits because people moving homes are probably more likely to experience a change in credits than people remaining in the same home.

decrease the credit. Income increases can also have the opposite effect. An increase in income can decrease benefits if a person's income moves into or farther into phase-out range of the credit.

An increase in income can happen because a person works or earns more—but also when couples marry, and additional income may become available to the tax unit. We find that over 70 percent of the low-income families that experience a decrease in their EITC experience that decrease because of an increase in earnings (representing 28 percent of all low-income families in our sample).[11] The remaining families that see an EITC drop of at least $500 from one year to the next are split roughly evenly between families where the number of children decreased and families where income decreased and caused the EITC to decrease. Changes in other household characteristics such as marital status not accompanied by changes in income or children affect under 1 percent of low-income families.

In some cases, drops in the EITC from one year to the next can be dramatic. Among low-income families, about 22 percent see a drop of at least $2,000 and another 11 percent see a drop of between $1,000 and $2,000 (Figure 3). Drops of at least $2,000 are caused by income increasing 64 percent of the time, children decreasing 20 percent of the time, and income decreasing 16 percent of the time.

FIGURE 3. Year-to-Year Changes in Earned Income Tax Credit Amounts for Families with Children, 2018 Tax Law



Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Notes: Sample includes households with one dependent child under age 18 in either year. No change is defined as a change of less than $500.

About 22 percent of low-income families with children become eligible for a larger EITC in the second year than in the first. Changes in income drive increases in the EITC for these low-income families about 75 percent of the time (9 percent of low-income families see their EITC increase because their income decreased, and another 8 percent of low-income families see their EITC increase because their income increased). For 5 percent of low-income families, we observe an EITC increase driven by the number of children in the tax unit

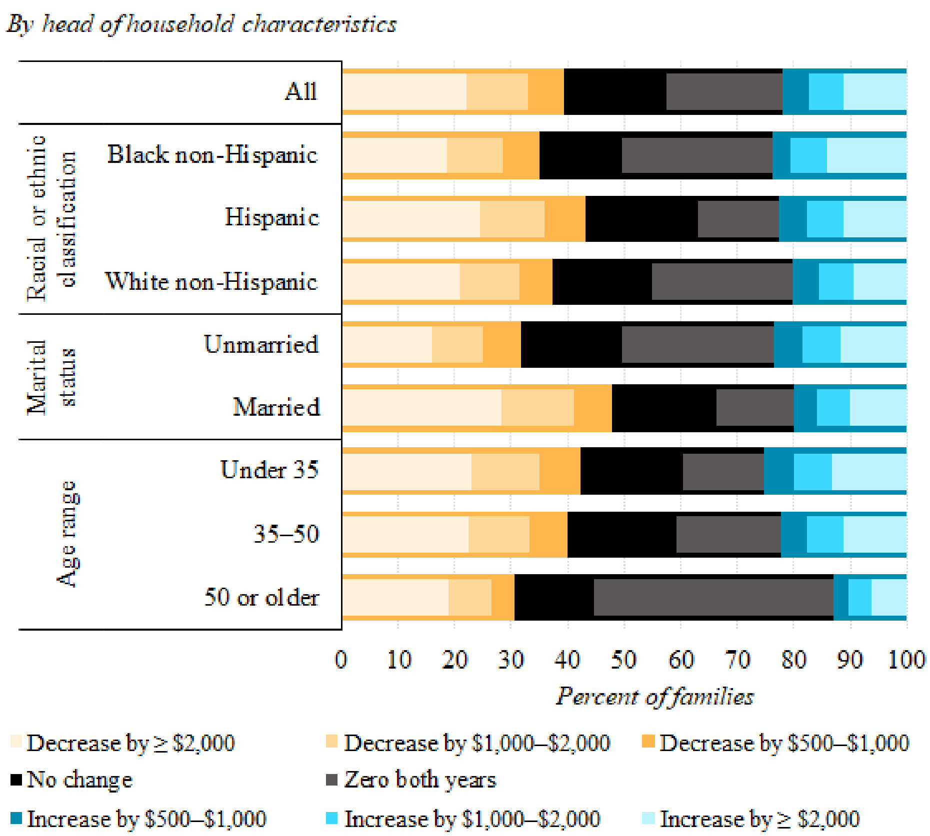
[11] Decrease in earnings is defined as a decrease in earnings without a change in the number of children over the same period.

increasing. About half of all low-income families with an EITC increase see an increase of more than $2,000. A small share of low-income families become newly eligible for an EITC in the second year we observe them after having no earnings in the first year.

Demographic variations in earned income tax credit changes among low-income households

We estimate whether the likelihood of an EITC increase or decrease varies by demographic characteristics for those with income below 200 percent of the federal poverty level in year one (Figure 4). About 42 percent of white non-Hispanic and 41 percent of Black non-Hispanic households receive no credit in both years, or have no major change in their credit amount from year one to year two. Hispanic families were less likely to experience no change in eligibility, about 34 percent. While the share of families, by race and ethnicity, that experienced an increase in the amount of EITC they were eligible for from year one to year two was roughly the same, Hispanic families were more likely to see their EITC drop than white, non-Hispanic or Black, non-Hispanic families.

FIGURE 4. Year-to-Year Changes in Earned Income Tax Credit Amounts, 2018 Tax Law



Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Notes: Sample includes households with one dependent child under age 18 in either year with incomes below twice federal poverty level in the first year observed. No change is defined as a change of less than $500. Marital status is shown only for those with same marital status in both years. Families with marital status changes excluded due to small sample size.

People who were low-income and unmarried in both years of our sample were more likely to maintain similar EITC eligibility in year two (45 percent) than people who were married (32 percent). This may be because with only one potential earner in the tax unit, there is less opportunity for variation. In a married couple, two people may be exposed to changes in employment. About 32 percent of unmarried adults with low incomes in our sample saw the amount of credit they were eligible for decline by at least $500 in year two and the remaining 23 percent saw their credit amount increase.

Younger adults with low incomes were more likely to see their EITC change from year one to year two than households where the survey respondent was either under 35 or over age 50. While just 33 percent of adults under age 35 and 35 percent of adults ages 35 to 50 experienced no change in credit eligibility, over half of adults in our older group had no major change in credit eligibility. In many cases, it was because older tax filers were more likely to be ineligible for a credit in both years. Most changes in predicted eligibility were greater than $2000.

B. Child Tax Credit

Over 90 percent of all families with children received a CTC in either year, compared with 82 percent of families with incomes under 200 percent of the federal poverty level. Among families with low incomes that received no CTC, some had no earnings or earnings below $2,500 and others had children aged out of eligibility for the program. Children must be under age 17 to qualify for the CTC. Older limits apply for children to qualify for the EITC. Just over 58 percent of all families with children saw no major change in CTC eligibility for from year to year (a change of less than $500). Forty-nine percent of low-income families with children saw changes of less than $500 from one year to the next. Among those that saw their credits change by at least $500, roughly half saw their credits increase and the other half experienced a credit decrease. Among families with low-incomes, more saw an increase in their CTC from one year to the next than saw a credit decrease.

Year-over-year CTC Decrease

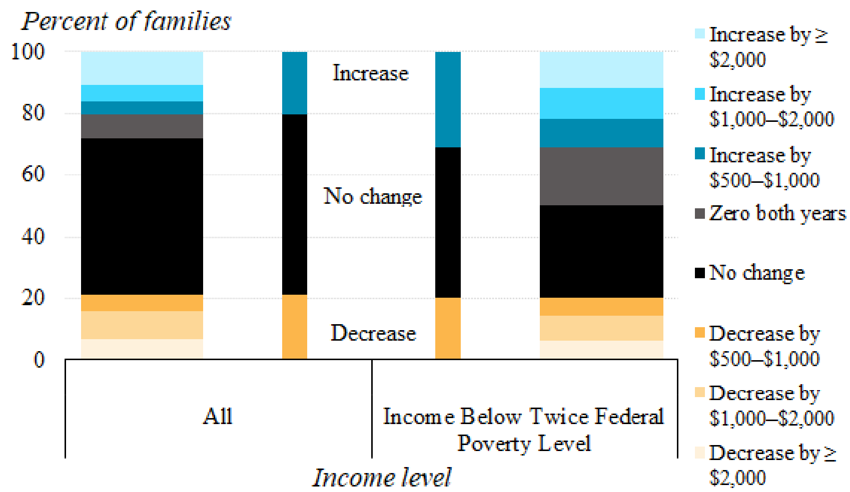
Just over 21 percent of all families with children and 20 percent of low-income families with children experienced a drop in credit eligibility of at least $500 from year one to year two. A drop in the number of children in the tax unit was associated with 65 percent of CTC decreases over $500. Among those 65 percent, about half of families with CTC decreases of at least $500 had a child age out of CTC eligibility. Among low-income families, income drops and reductions in the number of children contributed similarly to declines in credit eligibility.

Year-over-year CTC Increase

Among families with children, 21 percent of all families became eligible for a credit of at least $500 higher in the second year and 31 percent of low-income families saw the same. Just under half of credit increases were driven by an increase in the number of qualifying children, and nearly a quarter by an increase in children because of the birth or adoption of a child between years one and two. Among low-income families, credit increases most often stemmed from an increase in income. This allowed families to either move further up the phase-in of the credit or have additional tax liability that could be offset with the nonrefundable portion of the CTC. Among low-income families, a smaller share of CTC increases was attributable to increases in the number of children. For families with incomes too low to qualify for any CTC, an increase in children has no effect on their credit.

Most families whose credit decreased did so by amounts between $1,000 and $2,000. About 9 percent of all families and 8 percent of families with incomes below 200 percent of the federal poverty level in year one fell by this amount (Figure 5). For families experiencing a credit increase, the majority (10 percent of all families) had an increase of at least $2,000. This likely indicates a new child joining the tax unit.

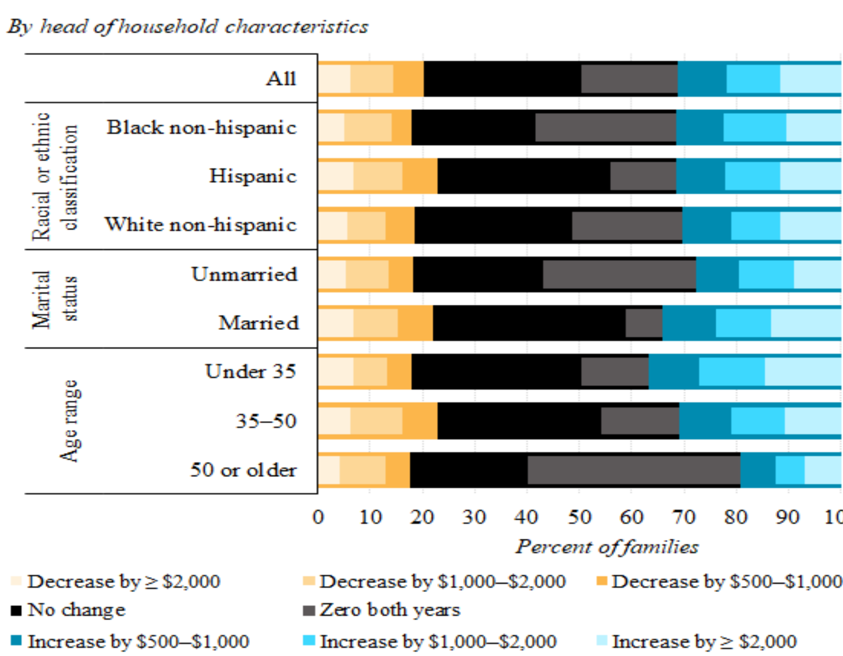
FIGURE 5. Year-to-Year Changes in Child Tax Credit Amounts for Families With Children, 2018 Tax Law



Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Notes: Sample includes households with one dependent child under age 18 in either year. No change is defined as a change of less than $500.

Comparing changes in credit by race and marital status among families with children with income below 200 percent of the federal poverty level, we see few differences. In general, families with income below 200 percent of the federal poverty level are more likely to see their credit increase (31 percent) than decrease (20 percent). We observe greater volatility in credit amounts among families where the parent that responded to the survey was between ages 35 and 50 (Figure 6).

FIGURE 6. Year-to-Year Changes in Child Tax Credit Amounts, 2018 Tax Law



Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Notes: Sample includes households with one dependent child under age 18 in either year with incomes below twice federal poverty level in the first year observed. No change is defined as a change of less than $500. Marital status is shown only for those with same marital status in both years. Families with marital status changes excluded due to small sample size.

VII. Discussion

Refundable tax credits like the EITC and CTC provide substantial support to families with children. Low- and moderate-income families often receive the credits as a single payment at tax time—but there is interest in delivering credit through the year, building on the experience of a temporary expansion of the CTC in 2021. But advancing credits is not without risk if families must pay them back if they end up receiving them errantly.

Our analysis estimates the size of year-over-year changes in the EITC and CTC to understand better how feasible it might be to deliver a tax credit based on information from the prior year's tax return. Credits delivered in advance must be based on some information. If families or the IRS were to use information from a current tax return to predict their next year's credit, our analysis shows how often they are likely to make a prediction within $500. Because credits are based on income, qualifying children (and in the case of the EITC where they reside most of the year), and marital status, families (or the IRS) would need to guess at some factors. No administrative data exist with this information, though prior research has retrospectively examined patterns of EITC participation by constructing panel datasets with tax return administrative data (Ackerman *et al.* (2009); Dowd and Horowitz (2011)). In prior work, we explored using administrative data to determine eligibility for credits and it was largely insufficient (Pergamit *et al.* (2014)). There are also shortcomings in the survey data used in this analysis: unstable households that move addresses are least likely to remain in the CPS sample. Consequently, our results could understate volatility in household arrangements, income, and tax credits.

We are more concerned with credit changes for low-income families who likely would have more difficulty paying back errantly delivered tax credits than high-income families. Moreover, they are likely to be harmed more by not getting advance credits, and recent evidence following the monthly delivery of the CTC from July to December 2021 suggests lower-income households are more interested in advance monthly payments than others (Maag and Karpman (2022)).

Among low-income families with children, those with incomes below 200 percent of the federal poverty level, 21 percent received no EITC in either the first or second year we observed them in national data, and 18 percent saw their EITC change by less than $500 from one year to the next. Advance credits could be designed to not deliver the entire benefit in advance or limit the amount of errantly delivered credit that would need to be repaid—though this would cost the government revenue. We find that 39 percent of low-income families see their EITC drop by at least $500 and 22 percent see their EITC increase by at least $500 from one year to the next. About 28 percent of low-income families saw their EITC drop from one year to the next because their income increased (about 72 percent of all families that saw an EITC drop).

Among low-income families with children, we estimate 18 percent receive no CTC in either year—often because they have no earnings or their children are over age 16, the oldest qualifying age for the credit. Another 30 percent see their CTC change by less than $500. About 20 percent of low-income families see their CTC drop by at least $500 and the remaining 31 percent see their CTC increase by at least $500. The largest groups that see credit changes are related to an increase in earnings. In the case of the CTC, 21 percent of low-income families saw their CTC increase because their earnings increased (about two-thirds of the low-income families that saw their CTC increase from one year to the next).

Families also see credits change because the number of children in the tax unit change or their marital status changes. These changes are less common than income changes but still affect a significant group of people. We estimate that 6 percent of all low-income families experienced a CTC increase because the number of children in the tax unit increased and almost 8 percent saw their CTC decrease because the number of children in the tax unit decreased. About 10 percent of EITC changed because the number of children in the household changed—divided evenly by EITC increases and decreases. Families that change from one year to the next (through marriage, divorce, or change in the number of children) are likely to see credit changes of at least $2,000.

Tax credits increasing and decreasing year over year introduces a source of income volatility among low-income families that in some cases can be positive—family income jumps more than expected because tax credits are higher than expected. In other cases, it can present a negative shock as tax credits drop. Analysis shows that these changes are often a surprise to families (T. Anderson *et al.* (2022)).

A. *Implications of Volatility*

Not all income volatility is experienced in the same way by families—and some is more predictable than others. When income increases, which often happens at tax time when tax refunds are delivered, families have new opportunities present. Families might invest in items such as a used car or childcare that can help with increasing employment or might invest in a relatively large purchase like a refrigerator. Evidence also suggests that families are more likely to go to the doctor following receipt of a tax refund (Hamad and Niedzwiecki (2019)) and families with older children are more likely to enroll in school (Manoli and Turner (2018)). If advance credits are delivered throughout the year, presumably refunds would be smaller—but it also might be the case that refunds could disappear altogether or families could unexpectedly owe taxes if too much credit is delivered in advance.

When changes in tax credits are foreseeable because children are aging out of eligibility, the IRS, tax preparers, and community organizations can work to educate taxpayers about coming changes. Trusted messengers, or third parties regarded as credible by families that face barriers to navigating the tax system, can increase tax benefit participation and are well-situated to relay tax information throughout the year (Cox *et al.* (2021); Airi *et al.* (2022); Godinez-Puig *et al.* (2022)). In some cases, families will know they are likely to add another child either through birth or adoption. Other times, changes can be more difficult to predict, and it's unlikely that families could be warned appropriately. It may be the case that custody of a child changes abruptly and who lives together can also change. The IRS would know about these changes only if families or third-party assistants were able to apprise the IRS of the changes, in which case advance credits that were being delivered could be stopped or started, as appropriate. In the recent experience with the IRS portal for the expanded CTC, few families updated information through the portal (GAO (2022)).

Decreases in the credits are more concerning because that could put families in the vulnerable position of needing to repay the IRS. In many cases, income swings that appear to be driving a lot of the year-over-year changes we observe would not be predictable, absent new reporting requirements. And, as with family changes, only if the IRS were notified quickly could advance payments be stopped to limit a family's potential liability.

VIII. Conclusion

Refundable tax credits provide an important source of income for families with low incomes. Determining who can claim a child can be complicated by family structure and living arrangements. As the share of married couples with only biological children declines and the share of children in shared custody arrangements rises, filing a tax return can become more complex. Families must determine what tax unit a child should be properly assigned to—and how that decision is made can have a dramatic impact on who will benefit from the EITC and CTC. Importantly, the determination is made on an annual basis and only one family can get a tax benefit for a given child—even when multiple families share custody of a child. Changes in the number of children that a family can claim, income, and marital status (which can itself affect income) can all drive large credit changes from one year to the next.

Most families experience year-over-year changes in their EITC or CTC of less than $500. When calculating the EITC, this often happens because families are eligible for no credit in either year one or year two. Among low-income families, those with income below twice the federal poverty level, 39 percent have a change of less than $500 in their EITC, and about half have a change of less than $500 in their CTC. A virtually identical share of families with no change in marital status or number of children have a change of less than $500 in their EITC and 56 percent of these families have a change of less than $500 in their CTC.

Among low-income families with children, about 40 percent see their EITC decrease by more than $500 and 19 percent see their CTC decrease by more than $500. Most often, when families experience a drop in their EITC, it is because their earnings increase. In some cases, this is because a single parent marries, bringing a new source of income into the tax unit. Conversely, a decrease in earnings is the most common reason for an increase in the EITC (families are moving from beyond the phase-out or in the phase-out range to the maximum credit range), which shows how the credit can mitigate a loss in income in some cases. When a low-income family's CTC increases, it often does so because of an increase in income rather than a change in the number of children in the tax unit. This is because low-income families have their credit limited by their earnings not being enough to access more credit—but very few will see earnings increase large enough to result in the credit beginning to phase out.

Helping families understand how credits are calculated might help them predict when a credit will increase or decrease. This is important because it could help families understand the financial position that they will be in at tax time the following year. Understanding how credits change from year to year can also help policy makers design advance credits that can be delivered without putting families at risk. For example, policy makers can design provisions that protect a certain amount of credit from being clawed back at tax time if too much credit has been delivered in advance.

Our research suggests that protecting about $500 of each advance credit would protect most families from needing to repay credits at tax time. These protections would cover a smaller share of low-income families, especially with respect to the EITC, which suggests that further measures are needed to protect low-income families in particular. Otherwise, an advance credit based on last year's tax return could result in a disruptive tax bill. While steps might be available to mediate changes in credit stemming from changes in income (depending on how soon they were reported), family changes present additional challenges.

References

Ackerman, Deena, Janet Holtzblatt, and Karen Masken. 2009. "The Pattern of EITC Claims Over Time: A Panel Data Analysis." *Recent Research on Tax Administration and Compliance: Selected Papers Given at the 2009 IRS Research Conference*, IRS Research Bulletin. Washington, DC: Internal Revenue Service.

Airi, Nikhita, Luisa Godinez-Puig, and Kim Rueben. 2022. "Helping New Mothers Understand the Benefits of Filing Taxes." Washington, DC: Tax Policy Center.

Anderson, Lydia, Paul F. Hemez, and Rose M. Kreider. 2022. "Living Arrangements of Children: 2019." Suitland, Suitland-Silver Hill, MD: US Census Bureau.

Anderson, Theresa, Amelia Coffey, Hannah Daly, Heather Hahn, Elaine Maag, and Kevin Werner. 2022. Balancing at the Edge of the Cliff: Experiences and Calculations of Benefit Cliffs, Plateaus, and Trade-Offs. Washington, DC: Urban Institute.

Caldwell, Sydnee, Scott Nelson, and Daniel Waldinger. 2023. "Tax Refund Uncertainty: Evidence and Welfare Implications." *American Economic Journal: Applied Economics* 15 (2): 352–76.

CBO (Congressional Budget Office). 2019. "Marginal Federal Tax Rates on Labor Income: 1962 to 2028." Washington, DC: CBO. https://www.cbo.gov/system/files/2019-01/54911-MTRchartbook.pdf.

Cox, Kris, Roxy Caines, Arloc Sherman, and Dottie Rosenbaum. 2021. "State and Local Child Tax Credit Outreach Needed to Help Lift Hardest-to-Reach Children Out of Poverty." Washington, DC: Center on Budget and Policy Priorities.

Dowd, Tim, and John B. Horowitz. 2011. "Income Mobility and the Earned Income Tax Credit: Short-Term Safety Net or Long-Term Income Support." *Public Finance Review* 39 (5): 619–52. https://doi.org/10.1177/1091142111401008.

Department of the Treasury. 2018. "Financial Report of the United States Government: Fiscal Year 2017." Washington, DC: Department of the Treasury.

Fox, Liana, and Kalee Burns. 2020. "The Supplemental Poverty Measure: 2019." Suitland, Suitland-Silver Hill, MD: U.S. Census Bureau.

GAO. 2022. "Current and Future Federal Preparedness Requires Fixes to Improve Health Data and Address Improper Payments." Washington, DC: Government Accountability Office.

Godinez-Puig, Luisa, Aravind Boddupalli, and Livia Mucciolo. 2022. "Lessons Learned from Expanded Child Tax Credit Outreach to Immigrant Communities in Boston." Washington, DC: Tax Policy Center.

Hamad, Rita, and Matthew J. Niedzwiecki. 2019. "The Short-Term Effects of the Earned Income Tax Credit on Health Care Expenditures among US Adults." *Health Services Research* 54 (6): 1295–1304. https://doi.org/10.1111/1475-6773.13204.

Holtzblatt, Janet and Janet McCubbin. 2002. "Issues Affecting Low-Income Filers." In *The Crisis in Tax Administration* edited by Henry Aaron and Joel Slemrod, 148–200. Washington, DC: Brookings Institution Press.

IRS (Internal Revenue Service). 2022. "Common Errors for the Earned Income Tax Credit (EITC)." Washington, DC: IRS.

IRS (Internal Revenue Service). 2022. "Handling the Most Common Errors." Washington, DC: IRS.

Maag, Elaine, H. Elizabeth Peters, Sara Edelstein. 2016. Increasing Family Complexity and Volatility: The Difficulty in Determining Child Tax Benefits. Washington, DC: Urban Institute.

Maag, Elaine, H. Elizabeth Peters, Anthony Hannagan, Cary Lou, Julie Siwicki. 2017. Income Volatility: New Research Results with Implications for Income Tax Filing and Liabilities. Washington, DC: Urban Institute.

Maag, Elaine, and Michael Karpman. 2022. "Many Adults with Lower Income Prefer Monthly Child Tax Credit Payments." Washington, DC: Urban Institute.

Manoli, Day, and Nicholas Turner. 2018. "Cash-on-Hand and College Enrollment: Evidence from Population Tax Data and the Earned Income Tax Credit." *American Economic Journal: Economic Policy* 10 (2): 242–271. https://doi.org/10.1257/pol.20160298.

Meyer, Daniel R., Maria Cancian, and Steven T. Cook. 2017. "The Growth in Shared Custody in the United States: Patterns and Implications." *Family Court Review* 55: 500–12. https://doi.org/10.1111/fcre.12300.

Michelmore, Katherine M., and Natasha V. Pilkauskas. 2022. "The Earned Income Tax Credit, Family Complexity, and Children's Living Arrangements." *The Russell Sage Foundation Journal of the Social Sciences* 8 (5): 143–65. https://doi.og/10.7758/RSF.2022.8.5.07.

Morduch, Johnathan, and Rachel Schneider. 2017. *The Financial Diaries: How American Families Cope in a World of Uncertainty.* Princeton: Princeton University Press.

National Taxpayer Advocate. 2020. "Earned Income Tax Credit: Making the EITC Work for Taxpayers and Government." Washington, DC: National Taxpayer Advocate.

Pergamit, Michael R., Elaine Maag, Devlin Hanson, Caroline Ratcliffe, Sara Edelstein, and Sarah Minton. 2014. 2014 Pilot Project to Assess Validation of EITC Eligibility with State Data: Final Report. Washington, DC. Urban Institute.

Romich, Jennifer L., and Thomas Weisner. 2000. "How Families View and Use the EITC: Advanced Payment versus Lump-sum Delivery." *National Tax Journal* 4 (2). https://doi.org/10.17310/ntj.2000.4S1.09.

Williams, Roberton, and Elaine Maag. 2008. "The Recession and the Earned Income Tax Credit." Washington, DC: Urban Institute.

2

∇

Estimating Audit Aftershocks

Besnek ◆ Partington

Lindsay ◆ Grana ◆ McGlothlin ◆ Plumley

Plumley ◆ Rodriguez ◆ Grana ◆ McGlothlin

Changes to Voluntary Compliance Following the Random Enquiry Program on Income Tax Returns

Murat Besnek and Allan Partington (Australian Taxation Office)

1. Introduction

There are three ways audits impact revenue collected by tax administrators. First, through adjustments, penalties, and interest payments made during the audit process when correcting the taxpayer's initial misreported liability (audit yield). Second, through changes to future voluntary compliance of audited taxpayers where the audit influences the subsequent reported liabilities (direct deterrent effect). Third, through spillover effects on non-audited taxpayers whose reported liabilities are influenced in part by their expected probability of an audit, based on their observation of the tax administration's activities (indirect deterrent effect). Tax administrators have precise information about audit yields; however, less is known about the direct and indirect deterrent effects. Moreover, even though there is extensive literature on tax evasion,[1] the literature on behavioural changes of taxpayers is limited (Advani *et al.* (2019); Beer *et al.* (2015); Gemmell and Ratto (2012)).

Activities like audits are commonly used by tax administrators to increase tax compliance. Without these strategies, taxpayer contributions would be expected to be limited in the absence of strong altruistic motivations. We know—as do rational taxpayers—that it is not financially possible to pursue every taxpayer who is noncompliant because of audit costs. So, the payoff for noncompliance is an expected value—the payoff multiplied by the probability of not being caught. The more credible the threat of an audit, the lower the payoff for noncompliance, making it more beneficial to comply with the tax system (Bergolo *et al.* (2020)).

In this paper, we estimate the direct deterrent effect of Random Enquiry Program (REP) audits on tax returns, performed by the Australian Taxation Office (ATO) in 2015, 2016, and 2017. The audits include taxpayers from the individuals not in business (INIB); small business–individuals in business (SB-IIB); and small business–small company (SB-SC) population types. The estimate can be used to determine the intertemporal benefits (costs) of audits, potentially influencing the number of audits being allocated to certain populations and/or to audits in general. Estimates of the indirect deterrent effect are beyond the scope of this study.

One recommendation we adopt from Gemmell and Ratto (2012) is to separately test the responses of the so called "compliant" and "noncompliant" taxpayers.[2] This enables us to see the heterogenous treatment effects without allowing them to cancel each other out. As a point of difference during the estimation phase, we use the Poisson Pseudo Maximum Likelihood (PPML) estimator rather than ordinary least squares (OLS). We believe that the PPML estimator has two distinct advantages: (1) it removes the need to alter zero-inflated datasets[3] and (2) the results do not rely on the normal distribution assumption (Bellego *et al.* (2021)).

While we find that the audits change voluntary compliance, the results depend on the population type of the taxpayer, and if the audit finds them to be compliant or noncompliant. For instance, we find that the noncompliant taxpayers in the INIB population have a negative direct deterrent effect.[4] In contrast, we find that the noncompliant taxpayers in the SB-IIB and SB-SC populations have a positive direct deterrent effect.

[1] For instance, the tax gap estimates provided by the Australian Taxation Office.

[2] Being noncompliant hinges upon an error being detected during the audit process.

[3] For example, in Gemmell and Ratto (2012), all the observations where $y_i = 0$ is removed.

[4] That is to say that they report lower liabilities in the post-audit years than the control group.

Seventy-eight percent of taxpayers in the INIB population are noncompliant, so obtaining statistically significant estimates for the compliant taxpayers in this population are not possible. The compliant taxpayers in the SB-IIB (SB-SC) population have a positive (negative) direct deterrent effect. The largest (smallest) overall audit yield is found in the SB-IIB (INIB) population. The audit treatment effects appear to remain steady for all population types. There is no indication of returning to the levels displayed by their control groups over the multiple years following the audit allocation date,[5] which were covered by the study.

Our future research will involve extending the analysis by using 2018 audits on income tax returns from the REP. In addition, possibly incorporating the risk-based audits from the operational data to see if the risk-based audits produce different results. Estimating the indirect deterrent effect may also happen at some point in the future. The remainder of the paper is organised as follows: Section 2 describes the data and discusses who is considered noncompliant; Section 3 explains the empirical methodology; Section 4 presents the results by considering what the estimated intertemporal benefits (costs) of audits imply to the current compliance activities undertaken by the ATO; and Section 5 concludes the paper.

2. Data

The REP involves auditing the returns of randomly selected taxpayers from the INIB, SB-IIB and SB-SC populations. For this study, there are three REP datasets available,[6] separated by the financial year of the taxpayers' net tax amount[7] under investigation. Each dataset is analysed individually, but we also provide a pooled estimate. The data only includes taxpayers that are contacted by the ATO, using allocation date as a proxy for the date the taxpayers are contacted. Due to internal profiling, some taxpayers are not contacted by the ATO, and as a result, are not included in this study (see below for more details about internal profiling).

For each dataset, using the same sampling approach, we randomly select a control group that is approximately ten times larger than the treated. We check that the control groups do not include any taxpayers that are contacted by the ATO for other reasons during the period of the study. A Wilcoxon rank-sum test is then applied to confirm that the net tax amounts of the treated and control groups are similar in the pre-audited periods. If the Wilcoxon rank-sum test fails, a new control group is selected until the test passes. so that the treated and control groups are comparable. Once the taxpayers are comparable, we acquire the net tax amount for each taxpayer between the financial years of 2011–2020. We ensure that our analysis only focuses on voluntary compliance. For instance, we remove the 2015 net tax amount for taxpayers that are a part of the 2015 sample. This applies to taxpayers in the treated and control groups.

The INIB population consists of taxpayers with no business connection. They are typically individual entities other than those identified as being in or linked to small business, high wealth or wealthy Australians or recipients of passive or personal services income (PSI). The small business population focuses on small businesses, SB-IIB being individuals and SB-SC being companies. The individuals in the small business population include taxpayers with (i) turnover less than \$10 million (ii) exclusive connections to small businesses with aggregated turnover less than \$10 million, (iii) links to small business entities in the capacity of being a shareholder, director, trustee or partner, and (iv) PSI recipients. The companies in the small business population have an aggregated turnover less than \$10 million and are not controlled by high wealth groups (being groups with net assets greater than \$50 million with an ownership level greater than 40%).

The taxpayers in the REP are subject to internal profiling when they are selected. To minimise the burden on taxpayers, where income can be matched to a third-party dataset on the ATO system and the amounts that cannot be verified are immaterial, these returns are not investigated further. Such taxpayers are accepted as having no (or immaterial) tax adjustments. The remainder of the taxpayers in REP are then escalated to a review to determine material amounts that could not be verified. The review covers, but may not be limited to:

[5] Allocation date refers to the date when the taxpayer is allocated to an auditor to commence the case.

[6] 2015, 2016 and 2017.

[7] Net tax amount is tax on taxable income plus Medicare levy minus non-refundable offsets.

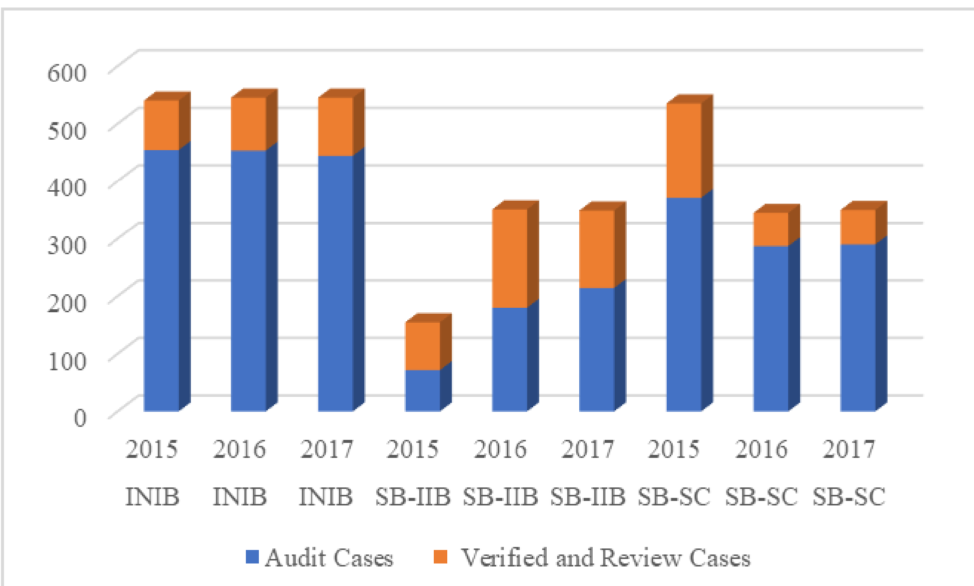
- Compliance history of the taxpayer;
- Recent financial performance of the business;
- Comparisons of declared income and expenditure;
- Checking merchant activity for credit card sales;
- Comparison with industry benchmarks;
- ATO risk flags; and
- Property ownership.

As well as reviewing the affairs of the taxpayers, any directly associated individuals and entities are also reviewed. These associates may include:

- Spouse and family members;
- Partners and partnerships;
- Companies, directors and shareholders; or
- Trusts.

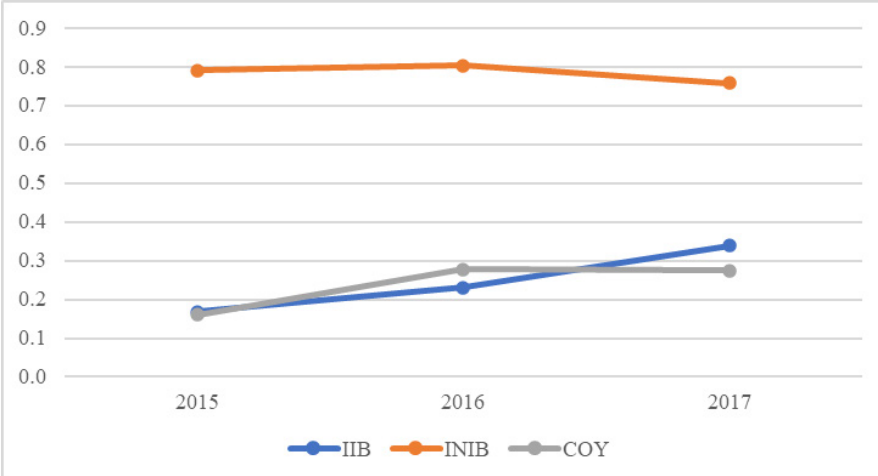
Where issues are found, taxpayers are taken to the next stage, which is an audit. Only the taxpayers that are escalated to an audit are included in this study as there is no taxpayer contact during the review stage. We would expect the behaviour of the verified and reviewed taxpayers to be no different than the control group. Approximately 17% of the INIB, 46% of SB-IIB and 23% of the SB-SC populations are removed from REP samples. Leaving behind 1,351 audits for the INIB, 466 audits for the SB-IIB, and 948 audits for the SB-SC populations. Figure 1 provides a breakdown across the three financial years.

FIGURE 1. REP Sample Size



Financial Year 2015 is the first time the REP took place, so the sample size is a little smaller. Sample size between INIB and small business (SB-IIB+SB-SC) is comparable over time; though, one thing to note is that the small business population contains more companies than individuals. The size of the datasets is capped due to resources allocated to the REP, even though it would be beneficial to have more treated observations for this study. In terms of noncompliance, individuals seem to perform worse than companies. The INIB population is approximately 78% noncompliant. Companies, on the other hand, seem to conform a lot better with their tax obligations, never exceeding 30% in noncompliance, which can be observed from Figure 2.

FIGURE 2. Noncompliance as a Percentage of REP Sample Size



As expected, the audit yields are lower for the INIB population than for small businesses. The average audit yield for INIB is equal to $1,071 in 2015, $1,098 in 2016 and $881 in 2017. If we use the pooled dataset, the average audit yield for INIB equals $1,018. The average audit yield for SB-IIB is equal to $3,914 in 2015, $2,001 in 2016 and $12,253 in 2017. If we use the pooled dataset, the average audit yield for SB-IIB equals $6,936. The average audit yield for SB-SC is equal to $900 in 2015, $2,705 in 2016 and $4,129 in 2017. If we use the pooled dataset, the average audit yield for SB-SC equals $2,433. It is also apparent from Table 1 that the audit yields are not enough to cover the costs of running the REP.

TABLE 1. Average Audit Yield

AVERAGE	2015	2016	2017	POOLED
INIB	$1,071	$1,098	$881	$1,018
SB-IIB	$3,914	$2,001	$12,253	$6,936
SB-SC	$900	$2,705	$4,129	$2,433
REP AVERAGE	$1,962	$1,935	$5,754	$3,462

3. Empirical Methodology

Tax administration research frequently uses positively skewed datasets where the dependent variable equals zero on a regular basis. Under these circumstances, using OLS for statistical inference is not appropriate, due to the violation of the normality assumption. The common solution is to use a log-transformed dependent variable. However, logging the dependent variable is not ideal due to Jenny's inequality. Jenny's inequality implies that $E(\ln(y))\neq \ln E(y)$, so retransforming the log terms will result in a biased estimate (Motta (2019)). This estimate will then need to be adjusted for heteroscedasticity.

Another major issue with logging the dependent variable is the inability to log zeros. If we decide to take this approach, we need to add a positive constant to all the observations where the dependent variable equals zero or delete them altogether like Gemmell and Ratto (2012). However, removing the zeros or giving them a small positive value can worsen the heteroscedasticity (Motta (2019)). Moreover, the size of the positive constant will depend on the data at hand, adding the smallest possible value (for example, the value of 1) is not the least harmful choice. Bellego *et al.* (2021) show that the best value for the positive constant is not necessarily small, nor equal to 1, contrary to common belief.

Using the PPML estimator is a better alternative to correcting the bias of a log-transformed dependent variable because it can handle observations where the dependent variable equals zero (Silva and Tenreyro

(2006); Correia *et al.* (2019)). This estimator is popular because the only condition required for consistency is the correct specification of the conditional mean. Therefore, the data does not need to have a Poisson distribution, nor does the dependent variable need to be an integer (Gourieroux *et al.* (1984)). Although, with continuous variables, the assumption that the conditional mean equals the conditional variance is unlikely to hold. For this reason, the standard errors need to be based on the Eicker- Huber-White robust covariance estimator.

To measure the change in voluntary compliance of audited taxpayers, we use a difference in differences (DID) model. We begin by subtracting the pre-audit net tax amount from the post-audit for treated taxpayers.[8] We denote this difference d_1 . Any difference in d_1 can be a result of the REP, but also other possible events. To account for this, we repeat the same process for the control group. We denote this difference d_2. Subtracting d_2 from d_1 produces the standard DID model and it can be estimated using the following regression:

$$y_{it}=\beta_0+\beta_1 D_{PostAudit}+\beta_2 D_{Treated}+\beta_3 D_{PostAudit}\times D_{Treated}+ e_t,$$

where y_{it} is the net tax amount for taxpayer $_i$ in year t, $D_{PostAudit}$, is a dummy variable for the post-audit observations $D_{Treated}$, is a dummy variable for the treated taxpayers, and e is a random error term.

The interpretation of the coefficients is as follows: β_0 equals the average pre-audit and $\beta_0+\beta_1$ equals the average post-audit net tax amounts for the control group. $\beta_0+\beta_2$ equals the average pre-audit and $\beta_0+\beta_1+\beta_2+\beta_3$ equals the average post-audit net tax amounts for treated taxpayers. β_3 is the DID parameter that quantifies the impact of the audit.

In Gemmell and Ratto (2012), a modified version of the DID model is specified. The purpose of the specification is to avoid combining the positive and negative impacts of audits. If we do not separately test the responses of compliant and noncompliant taxpayers, there is a chance that we incorrectly conclude that audits do not impact taxpayer behaviour. The REP does keep records of other personal information. However, due to the small sample size, it is not possible to include them in the DID regression. Other than year, population type, and taxpayer compliance, we could not control for any other taxpayer characteristics. The version of the DID model we use in this study is specified below:

$$y_{it}=\beta_0+\beta_1 D_{PostAudit}+\beta_2 D_{PostAudit}\times D_{Compliant}+\beta_3 D_{PostAudit}\ D_{NonCompliant}+\delta_i+e_t$$

with notation as described before and where $D_{Compliant}$ is a dummy variable for the compliant taxpayers, $D_{NonCompliant}$ is a dummy variable for the noncompliant taxpayers, δ_i are individual fixed effects.

Interpretation of the coefficients is as follows: $\beta_0+\beta_1$ equals the average post-audit net tax amount for the control group. $\beta_0+\beta_1 + \beta_2$ equals the average post-audit net tax amount for the compliant taxpayers. $\beta_0+\beta_1+\beta_3$ equals the average post-audit net tax amount for the noncompliant taxpayers. β_2 (β_3) is the DID parameter that quantifies the impact of the audit on compliant (noncompliant) taxpayers.

One thing to note is that when the model does not separate taxpayers based on their compliance, we do not need to control for unobservable characteristics, as random sampling ensures that there are no systematic differences across the groups. However, when we run the DID model where the treated taxpayers are divided into subgroups, we need to run the PPML estimator with individual fixed effects. This is to allow for the possibility that there are systematic differences across the groups. As long as these differences stay consistent (fixed) between the time periods of interest, the individual fixed effects (δ_i) coefficient will control for these differences, even if they are unobservable.

4. Results

Each subsection below provides a detailed breakdown of the direct deterrent effect for a specific population arranged by financial year. We obtain these results using reliable audit data sourced from the ATO. We employ the industry standard during the estimation phase, that being the PPML estimator (following the advice of Jeffery Wooldridge and many other academics). The model we use has no impact on the direction of the direct

[8] For instance, for the taxpayers that are a part of the 2015 financial year, we subtract their 2011-14 net tax amount from their 2016-20.

deterrent effect, which we confirm by comparing the results to the standard DID model that can be computed without the need of a regression. We report each financial year independently, but we base our final conclusions on the coefficients acquired from the pooled estimates for the reason that they have more observations and combining the financial years do not seem to violate any of the assumptions of the model. The coefficients in the tables are in percentage form, and to compute the wider revenue effects (WRE), we multiply these coefficients by the average pre-audit net tax amounts of the population.

i. Individuals Not in Business

Table 2 presents the results for the INIB population. Recall from Section 2 that the average audit yield is equal to $1,071 in 2015, $1,098 in 2016 and $881 in 2017. If we use the pooled dataset, the average audit yield equals$1,018. The average direct deterrent effect for noncompliant taxpayers is equal to $1,043 in 2015, -$2,740 in 2016, and $543 in 2017. If we use the pooled dataset, the average direct deterrent effect for noncompliant taxpayers is -$475. Due to the small sample size, we are unable to provide a reliable estimate of the direct deterrent effect for compliant taxpayers. The audit treatment effects seem to remain strong over the period of the study, with no indication of returning to the levels displayed by their control groups.

The INIB population has the smallest overall audit yield with a negative direct deterrent effect for noncompliant taxpayers. Compliant taxpayers seem to be rare in this population (22% of the sample). The results suggest that the intertemporal benefits of the audits depend heavily on the indirect deterrent effect, as voluntary compliance of audited taxpayers appear to deteriorate in the post-audit years. Note that the low WRE amounts become much larger when multiplied by the number of audited taxpayers and the years they lodge post-audit. The audits allocated to this population should be predominantly random given the low probability of finding large amendments (small audit yields). For this population, the use of information and edu- cational strategies may also be more cost-effective in changing the taxpayers' perceived probability of an audit than running actual audits.

TABLE 2. Results for Individuals Not in Business

		WRE PPML	COEFFICENTS	P-VALUE
	POST-AUDIT	-	0.287	0.000
2015	POST-AUDIT*COMPLIANT	-	0.050	0.792
	POST-AUDIT*NONCOMPLIANT	$1,043	0.079	0.008
	POST-AUDIT	-	0.309	0.000
2016	POST-AUDIT*COMPLIANT	-	0.177	0.249
	POST-AUDIT*NONCOMPLIANT	-$2,740	-0.189	0.000
	POST-AUDIT	-	0.322	0.000
2017	POST-AUDIT*COMPLIANT	-	-0.160	0.199
	POST-AUDIT*NONCOMPLIANT	$543	0.045	0.054
	POST-AUDIT	-	0.307	0.000
POOLED	POST-AUDIT*COMPLIANT	-	-0.006	0.942
	POST-AUDIT*NONCOMPLIANT	-$475	-0.036	0.073

ii. Small Business-Individuals in Business

Table 3 presents the results for the SB-IIB population. Recall from Section 2 that the average audit yield is equal to $3,914 in 2015, $2,001 in 2016 and $12,253 in 2017. If we use the pooled dataset, the average audit yield equals $6,936. In 2016, the average direct deterrent effect for compliant taxpayers equals -$2,720. As for 2015 and 2017, we are not able provide reliable estimates. If we use the pooled dataset, the average direct deterrent effect for compliant taxpayers equals -$1,898. The average direct deterrent effect for noncompliant taxpayers is equal to $3,077 in 2015 and $5,554 in 2016. We are not able to provide a reliable estimate for 2017. If we use the pooled dataset, the average direct deterrent effect for noncompliant taxpayers equals $2,616. The audit treatment ef- fects seem to remain strong over the period of the study, with no indication of returning to the levels displayed by their control groups.

The SB-IIB population has the largest overall audit yield with the direct deterrent effect depending on the compliance of the treated population. Voluntary compliance of compliant taxpayers deteriorates, while non-compliant taxpayers improve. Audits in this population should be mainly risk-based for two reasons: (1) it is likely to uncover large amendments if the selection model is developed correctly; and (2) to avoid the risk of randomly choosing compliant taxpayers that can worsen voluntary compliance. The intertemporal benefits of audits have the potential to be large for this population, provided that the treated include a large percentage of noncompliant taxpayers.

TABLE 3. Results for Small Business-Individuals in Business

		WRE PPML	COEFFICENTS	P-VALUE
	POST-AUDIT	-	0.264	0.000
2015	POST-AUDIT*COMPLIANT	-	-0.118	0.408
	POST-AUDIT*NONCOMPLIANT	$3,077	0.292	0.016
	POST-AUDIT	-	0.217	0.000
2016	POST-AUDIT*COMPLIANT	-$2,720	-0.193	0.019
	POST-AUDIT*NONCOMPLIANT	$5,554	0.394	0.000
	POST-AUDIT	-	0.269	0.000
2017	POST-AUDIT*COMPLIANT	-	-0.094	0.224
	POST-AUDIT*NONCOMPLIANT	-	0.032	0.723
	POST-AUDIT	-	0.248	0.000
POOLED	POST-AUDIT*COMPLIANT	-$1,898	-0.148	0.007
	POST-AUDIT*NONCOMPLIANT	$2,616	0.204	0.002

iii. Small Company

Table 4 presents the results for the SB-SC population. Recall from Section 2 that the average audit yield is equal to $900 in 2015, $2,705 in 2016 and $4,129 in 2017. If we use the pooled dataset, the average audit yield equals $2,433. The average direct deterrent effect for compliant taxpayers is equal to $3,742 in 2015, $4,981 in 2016 and $5,529 in 2017. If we use the pooled dataset, the average direct deterrent effect for compliant taxpayers equals $4,848. In 2016, the average direct deterrent effect for noncompliant taxpayers equals $18,130. As for 2015 and 2017, we are not able provide reliable estimates. If we use the pooled dataset, the average direct deterrent effect for noncompliant taxpayers equals $5,955. The audit treatment effects seem to remain strong over the period of the study, with no indication of returning to the levels displayed by their control groups.

In the small business population, SB-SCs provide a lower overall audit yield than SB-IIBs. Given the costs of running audits on companies, this is not a favourable result; especially if we want to increase audits on SB-SC taxpayers. However, the positive direct deterrent effect more than makes up for the lower audit yield. The SB-SC population displays the largest improvement in voluntary compliance following audits. Both compliant and noncompliant taxpayers have large positive treatment effects. Audits allocated to this population can be random if preferred, as there seems to be no risk of worsening voluntary compliance due to poor selection. Although, there is no valid reason to believe that risk-based audits would perform differently.

TABLE 4. Results for Small Business-Small Company

		WRE PPML	COEFFICENTS	P-VALUE
	POST-AUDIT	-	0.164	0.000
2015	POST-AUDIT*COMPLIANT	$3,742	0.189	0.100
	POST-AUDIT*NONCOMPLIANT	-	0.001	0.995
	POST-AUDIT	-	0.120	0.000
2016	POST-AUDIT*COMPLIANT	$4,981	0.200	0.023
	POST-AUDIT*NONCOMPLIANT	$18,130	0.728	0.004
	POST-AUDIT	-	0.185	0.000
2017	POST-AUDIT*COMPLIANT	$5,529	0.195	0.008
	POST-AUDIT*NONCOMPLIANT	-	0.056	0.651
	POST-AUDIT	-	0.154	0.000
POOLED	POST-AUDIT*COMPLIANT	$4,848	0.197	0.000
	POST-AUDIT*NONCOMPLIANT	$5,955	0.242	0.039

5. Conclusion

This paper shows that audits on income tax returns in the REP conducted by the ATO alters taxpayers' perceptions of the probability of being audited. This in turn, for better or worse, changes the future voluntary compliance of taxpayers depending on the type (compliant or noncompliant) and which population (INIB, SB-IIB or SB-SC) from which they are chosen. Understanding the direct deterrent effects of audits is important because it quantifies the intertemporal benefits (costs), which subsequently helps the ATO make better decisions when choosing between different compliance activities, and how to allocate resources across different populations.

By comparing the treated and untreated taxpayers from the INIB, SB-IIB, and SB-SC populations, we estimate the change in voluntary compliance that occurs in the periods immediately after a taxpayer is audited. The results highlight the fact that audits influence future taxpayer behaviour, and that separate population types respond to them differently. It also underlines the importance of separating the responses of compliant and noncompliant taxpayers. We find that noncompliant taxpayers in the INIB population have a negative direct deterrent effect. In comparison, we find that noncompliant taxpayers in the SB-IIB and SB-SC populations have a positive direct deterrent effect. Due to the small sample size, we are not able to obtain statistically significant estimates for compliant taxpayers in the INIB population. We find that compliant taxpayers in the SB-IIB (SB-SC) population have a negative (positive) direct deterrent effect. All the audit treatment effects seem to remain steady during the period covered by the study. The indirect deterrent effect is beyond the scope of this paper.

The INIB (SB-IIB) population has the smallest (largest) overall audit yield. For the INIB population, information and educational strategies may be more suitable than running actual audits. Audits allocated to this population should be predominantly random, as the intertemporal benefits of audits will rely heavily on the spillover effects on non-audited taxpayers. As for the SB-IIB population, audits should be mainly risk-based to uncover the large, misreported liabilities in the population, and to avoid the risk of randomly choosing compliant taxpayers to worsen voluntary compliance. The audits allocated to the SB-SC population can be random or risk-based, as there seems to be no risk of worsening voluntary compliance due to poor selection. Both compliant and noncompliant taxpayers have large positive treatment effects.

Lastly, instead of truncating the datasets by removing all the observations where the dependent variable equals zero or trying to correct for the biasedness of a log-transformed dependent variable, we use a PPML estimator which can better handle the observations where the dependent variable equals zero. Our upcoming research will involve incorporating the 2018 REP dataset to this study. In addition, we may attempt to test the risk-based audits to see if they change the results. Developing a model that can estimate the indirect deterrent effect is also on the agenda.

References

Advani, A., W. Elming, and J. Shaw (2019). "The Dynamic Effects of Tax Audits." *CAGE Online Working Paper Series 414, Competitive Advantage in the Global Economy (CAGE)*.

Beer, S., M. Kasper; E. Kirchler; and B. Erard (2015). "Audit Impact Study." National Taxpayer Advocate 2015 Annual Report to Congress, Volume 2: *TAS Research and Related Studies*, Washington, DC, pp. 67–99.

Bergolo, M., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2020). "Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment." *NBER Working Paper Series*, 23631.

Correia, S., P. Guimarães, and T. Zylkin (2019). "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects." arXiv e-prints.

Gemmell, N. and M.Ratto (2012) "Behavioural Responses to Taxpayer Audits: Evidence From Random Taxpayer Inquiries." *National Tax Journal*, Vol. 65(1), pp. 33-58.

Slemrod, J. (2019). "Tax Compliance and Enforcement." *Journal of Economic Literature*, 57, 904–54.

Gourieroux, C., A. Monfort, and A. Trognon (1984). "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica*, 52(3): 701–20.

Silvia J.M.C., and S. Tenreyro (2006). "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

The Long-Term Impact of Audits on Nonfiling Taxpayers[1]

India Lindsay, Jess Grana, and Alexander McGlothlin (MITRE), Alan Plumley (IRS, RAAS)

1. Introduction

Based on current estimates, nonfiling taxpayers contribute 9%, or $32 billion, towards the individual income tax gap (IRS (2022)). In recent years, there has been an increase in the number of potential nonfiler cases and a simultaneous decline in resources allocated to audit these taxpayers. The resulting decline in audit rate is correlated with a loss of direct revenue from nonfiler audits (the assessed taxed owed, interest, and penalties from audited taxpayers). However, little is known about the long-term or indirect effect of audits on this group of taxpayers and whether consistent declines in audits have resulted in lower voluntary compliance. This paper considers nonfiling taxpayers with at least $100,000 in reported income and estimates the indirect effect of Field (in-person) audits on their future filing behavior.

The IRS Small Business/Self Employed (SBSE) division conducts audits of nonfiling taxpayers. These in-person audits are comprehensive and likely to leave a lasting impression on audited taxpayers that may alter their future compliance behavior. We use administrative data from the IRS for Tax Years (TYs) 2009-2014 on audited nonfiling taxpayers to compare their behavior over time to a group of eligible-but-unaudited taxpayers.

This research will enhance the IRS's ability to efficiently allocate audit resources, inform policymakers of the impact of the IRS's efforts to promote compliance, and contribute towards the literature on tax policy by highlighting the factors influencing the filing behavior of nonfilers. This is ongoing work. In addition to the impact of audits on filing behavior, future work is planned to evaluate the impact on reported total tax. The paper is organized as follows: Section 2 summarizes the relevant literature, Section 3 describes the audit selection process for nonfiler Field audits, Section 4 describes our data, Section 5 lays out the estimation approach, Section 6 presents results, and Section 7 concludes.

2. Literature Review

The literature on nonfilers primarily seeks to understand nonenforcement-related determinants of filing, such as a taxpayer's employment situation and demographic characteristics. Most of the literature studies the general nonfiling population and does not focus specifically on taxpayers earning more than $100,000, who typically have more complex tax situations than the median earning taxpayer, but a clear requirement to file a tax return. Further, to our knowledge, only three papers in the nonfiler literature evaluate the effect of past enforcement on future filing behavior.

2.1 Determinants of Nonfiling

For this study, we define nonfilers as taxpayers with a filing obligation who fail to file a return. Prior studies find that taxpayers with more easily concealed income are more likely to be nonfilers (Erard and Ho (2001)). For example, taxpayers with Schedule C business income and those employed in certain occupations (such as mechanics and helpers) were the least likely to file. Taxpayers working in industries such as construction, extraction, and production were the most likely to file. The authors find that nonfiling behavior is persistent; those who fail to file tend to continue to do so, and vice versa.

[1] Approved for Public Release; Distribution Unlimited. Public Release Case Number 23-2728. This paper was produced for the U. S. government under Contract Number TIRNO-99-D-00005, and is subject to Federal Acquisition Regulation Clause 52.227-14, Rights in Data—General, Alt. II, III and IV (DEC 2007) [Reference 27.409(a)]. No other use other than that granted to the U. S. government, or to those acting on behalf of the U. S. government under that Clause is authorized without the express written permission of The MITRE Corporation.

The persistence of nonfiling behavior also extends to the timeliness of filing. Erard *et al.* (2020) find that individuals who file in the prior year are 45 percentage points more likely to file in a timely manner the next year than those who do not. Certain demographic characteristics, such as older age and higher income, are associated with timely filing, while taxpayers with a higher filing burden, who are married, and who have income near the filing threshold are less likely to file on time. Furthermore, they find that taxpayers eligible for refundable credits are more likely to file and that there is regional variation in filing compliance.

The literature on nonfiling behavior identifies additional determinants of filing behavior: whether the taxpayer lives in a state that taxes individual income, the number of third-party forms reported to the IRS for the taxpayer, and their number of dependents. Other literature points to more abstract determinants of filing, such as a taxpayer's perception of government and sense of moral duty (Santoro *et al.* (2020); Gangl *et al.* (2015); Robson *et al.* (2020)).

2.1.1 Higher Earning Nonfilers

Erard *et al.* (2022) model higher earning taxpayers separately and find that taxpayers with investment, retirement, and self-employment income are less likely to file than those with wage income alone. Langetieg *et al.* (2017) reach a similar conclusion. Erard *et al.* (2022) also identify persistence in the filing behavior of higher-earning nonfilers, like the general nonfiler population.

2.2 Indirect Effects of Enforcement

To our knowledge, only three studies exist on the indirect effects of enforcement on nonfilers. One study, conducted in collaboration with the Bank of Greece, estimates the direct and indirect effects of audits of high wealth individuals and nonfilers (Tagkalakis (2014)). The paper finds that a 1% increase in the number of audits increases direct revenue by 0.4% and indirect revenue by 0.1%. A drawback of this paper is that the authors lack access to return-level data so can conduct their analysis only at the aggregate level.

Datta *et al.* (2015) estimate the effect of certain IRS enforcement activities on future filing behavior by evaluating nonfiler cases treated by the Automated Substitute for Return (ASFR) program. Compared to Field audits of nonfilers, ASFR handles a much larger volume of cases and works cases with simpler returns and a lower balance due. Datta *et al.* (2015) find that ASFR treatment increases the likelihood of filing by 11, 21, and 27 percentage points in the 2 through 4 years post treatment. They also cite past compliance behavior as an important predictor of future compliance.

Herlache *et al.* (2019) consider the impact of various mailed reminder-to-file notices on nonfilers' prior-year noncompliance and future filing behavior, from TYs 2016–2018. This research observes a 6.7% increase in filing of past noncompliant returns from TY 2016, a 4.6% increase in filing of returns for TY 2017, and a 4.1% increase in filing of returns for TY 2018. The impact of treatment on filing behavior was stronger for nonfilers (taxpayers exhibiting continuous nonfiling behavior) compared to stopfilers (compliant taxpayers predicted to be at risk of becoming a nonfiler in future years).

Overall, there is a gap in the literature analyzing both the behavior of higher-earning nonfilers and the role of IRS enforcement on filing behavior. In fact, Langetieg *et al.* (2017) cite the need for a longitudinal study of filing behavior using individual level IRS data. This research aims to begin to fill this gap.

3. Background on Audit Selection

In-person audits on nonfilers are conducted by either tax compliance officers or revenue agents who work in the SBSE division at the IRS. The Individual Master File (IMF) Case Creation Nonfiler Identification Process (CCNIP) is the selection process for identifying the majority of nonfiling taxpayers eligible for SBSE Field audit. Nonfilers may also be selected for audit via alternate processes, such as referral programs. For this research, the CCNIP selection process was obtained through interviews with SBSE officials.

First, the Information Returns Program (IRP) compiles reported income information from third parties for all taxpayers. Forms reported by third parties include wages, tips, and other compensation paid to employees and reported to the IRS by their employers, in addition to forms furnished to the IRS from other entities, such as banks, and other financial institutions. This IRP information is combined with available information from taxpayers' prior tax returns to estimate total income.

If a taxpayer is identified to likely have a tax liability yet has not voluntarily filed a return, the Return Delinquency Program may initiate the process of notifying the taxpayer. Up to two notices may be sent to the nonfiling taxpayer, informing them of their delinquency in filing and requesting their tax return. If the taxpayer responds to either of these letters, their return may be accepted as filed or their case is assigned to an auditor to verify information. Taxpayers who remain as nonfilers are grouped into the Taxpayer Delinquency Investigations (TDI) inventory. The IRS applies screening criteria and may distribute TDI taxpayers, based on specific taxpayer and tax return characteristics, to one of three enforcement functions: ASFR, Collection, or Field audit. To be eligible for Field audit, taxpayers typically must have an estimated tax liability above an IRS-specified threshold and total reported income typically exceeding $100,000.

The SBSE Field office receives the lists of eligible TDI cases, estimates everyone's tax liability, and assigns a priority score. The priority score is a function of the taxpayers' estimated balance due and the likelihood of securing the balance due. Cases are assigned codes specific to the type of audit required, sorted by region, and ranked by priority. Top priority cases are sent to regional field offices and assigned to tax auditors according to each region's workplan and available resources. In addition to auditing nonfiling TDI cases, Field audits may be conducted on returns that are filed in response to delinquency notices and meet selection criteria.

4. Data

4.1 Sample Construction

This study uses taxpayer and audit record data obtained from the IRS Compliance Data Warehouse (CDW) using primary Taxpayer Identification Numbers. The treatment group consists of nonfiling taxpayers in the TDI Lists workstream who were subject to audits between TYs 2009 to 2014 and is compared to a control group that consists of taxpayers who were not audited but were in the TDI Lists workstream during the same period. Each taxpayer is assigned a "baseline year" that is defined as the tax year the taxpayer entered the sample, either because they were audited in that tax year (treatment group) or because they were eligible-but-unaudited in that tax year (control group). We analyze these taxpayers' tax reporting behavior 5 years before this baseline year through 8 years after, creating a dataset measuring taxpayer behavior from 2004–2022.

4.1.1 Treatment Group

The treatment group consists of all nonfiling taxpayers audited from the TDI Lists workstream during the baseline period (TYs 2009–2014). Data on taxpayers audited from TDI Lists were obtained from IRS audit record data. We include the primary returns selected for audit in each tax year and exclude "pickups"—returns from the same taxpayer for other tax years that were audited because of the primary audit.

4.1.2 Control Group

The control group includes unaudited nonfilers identified by CCNIP during TYs 2009–2014. Taxpayer data were extracted from the CCNIP database using queries designed to replicate the SBSE screening criteria for this population. We verified income, tax due, and filing status to ensure taxpayers met the primary selection criteria. The control group excludes taxpayers who filed late, were secondary filers, or who filed in response to a delinquency notice in the baseline year.

4.1.3 Sample Cleaning

Among audited taxpayers, we drop those with missing or unmatched audit record data. We also remove taxpayers who were selected for audit outside of the CCNIP process (such as the State Audit Reporting Program and various referral programs). Because we exclusively source the control group from CCNIP, we do so for the treatment group as well. Our sample contained audited taxpayers that had a baseline year tax return on file.

These individuals either filed a Form 1040 for the tax year of the audit prior to audit start or filed in response to delinquency notices. Regardless of late filing, these individuals may still have been audited by the nonfiling Field audit group if their return met selection criteria. These individuals are dropped from the sample.

Among the control group, taxpayers with audits in the 6 years before and after baseline are dropped from the sample so we estimate the impact of the baseline audit only. We also remove any taxpayers in the control group that were identified by CCNIP but later deemed not to have a tax liability.

For both the treatment and control groups, taxpayers are removed from the sample if they died within 8 years of their baseline year. During the 2009–2014 period, some taxpayers were candidates for audit in multiple baseline years. To assign each taxpayer only one baseline year, we applied de-duplication rules. Taxpayers audited more than once are assigned their first audit year as their baseline year. Taxpayers considered eligible more than once (but never audited during our baseline period) are also assigned their first eligible year as their baseline year. Taxpayers appearing in both the treatment and control group during this period are assigned to the treatment group with their first audit year as their baseline year.

4.2 Dependent Variable

Our model's dependent variable is *fact of filing*, a binary variable that equals 1 if the taxpayer files a return and 0 otherwise. This variable is created for each taxpayer in our sample for the 5 years preceding their baseline year to the 8 years following. For baseline years, all taxpayers in our sample have a *fact of filing* equal to 0, by definition. We construct *fact of filing* for off-baseline years by consolidating filing information on a taxpayer from the Information Returns Transaction File (IRTF). For a given tax year, a taxpayer is considered to have filed if they filed a Form 1040 (timely or late), is listed as a secondary filer on another taxpayer's Form 1040, is selected for a nonfiling audit but is later deemed to not have a tax liability, or is selected for a nonfiling audit but filed prior to the audit.

A taxpayer has *fact of filing* set to 0 if they are a known nonfiler or a "ghost." In the context of this study, the term nonfiler refers to a taxpayer who *was identified by CCNIP* as having a tax liability but did not voluntarily file a return. The term nonfiler includes nonfiling taxpayers experiencing a nonfiler audit or some other treatment by an IRS nonfiler program. Though these individuals may file in response to treatment, their *fact of filing* is not voluntary. A "ghost" refers to a taxpayer who did not file and is not a known nonfiler for a given tax year. A taxpayer could become a ghost via two mechanisms:

1. The taxpayer does not have enough income to have an income tax liability. This could occur if a taxpayer suddenly becomes unemployed and has no income. If the taxpayer had no income to report to the IRS (either by themselves or by third parties), they would not appear in IRS records.
2. The taxpayer has income but does not report the income and it is not covered by third-party reporting. This could occur if a taxpayer is self-employed, and the IRS does not have a means of verifying a tax liability in advance of an audit.

Since taxpayers in either case would not be in IRS records, we cannot distinguish between the first case (no tax liability) and the second case (owes taxes); all we know about them is that they were nonfilers in the baseline year and in an off-baseline year, they did not file, and they were not identified by the IRS as a nonfiler. Based on conversations with nonfiler subject-matter experts in the IRS Research, Applied Analytics, and Statistics organization, we made the decision to assume ghosts were nonfilers. This assumes that income for higher earning nonfilers tends to be persistent and that such taxpayers are more likely to conceal their income than to have no income.

4.3 Independent Variables

In addition to audit status, we control for characteristics identified from prior literature as important determinants of the decision of whether to file. These include demographic and financial information, as well as past filing behavior. In the current model, control variables come from the CCNIP database and are time-invariant because they are derived solely from baseline year data. This approach was chosen because taxpayers in our dataset tend to become ghosts, meaning off-baseline year data are not always available. In future research, we hope to overcome this barrier by compiling time-varying control variables from a variety of data sources.

4.3.1 Demographic Variables

We control for taxpayer level demographics, including Census region of residence,[2] whether the taxpayer resides in a state taxing individual income, whether the taxpayer was over 65 in the baseline year, whether the taxpayer was under 30 in the baseline year, and their filing status in the prior year. We treat filing status as a binary variable, with 1 being Married Filing Jointly and the reference level being other filing statuses (Single, Married Filing Separately, Widow/er, Head of Household) collapsed into one category. We also include an indicator for whether the taxpayer claimed the Earned Income Tax Credit (EITC) in the prior year. Unfortunately, we are inconsistent with prior literature in that we have not yet controlled for the number of dependent children, which is not in the CCNIP database.

4.3.2 Financial Variables

We further construct a set of variables related to the taxpayers' financial status. Total IRP income is defined as the sum of all income reported on third-party forms, without subtracting possible losses or deductions. We also construct an indicator for whether the taxpayer met the $100,000 threshold in the baseline year, since some taxpayers in the treatment group did not.[3] We also include the number of IRP forms submitted by third parties to the IRS, since this captures both the filing burden felt by the taxpayer and the complexity of income sources, and because each additional report to the IRS from third parties increases the "visibility" of that taxpayer.

We include the difference in income between the prior year and the baseline year. Changes in income have not been considered by the previous nonfiling literature but help to capture the volatility of taxpayer income. It also serves as an indicator of one source of potential financial burden that may alter filing behavior. A positive value indicates the taxpayer's reported income grew in their baseline year compared to the prior year. Additionally, we construct indicators to capture the presence of various income sources listed on third-party documents, including self-employment income,[4] investment income,[5] retirement income,[6] broker transaction income,[7] as well as other types of reported income.[8]

4.3.3 Past Filing Behavior Variables

Lastly, we construct variables related to taxpayers' past compliance behavior. These variables include whether the taxpayer filed in the prior year, whether the taxpayer was a ghost in the prior year, and whether the taxpayer was audited in the 6 years prior to baseline. We also control for the operational priority score, an IRS-internal metric used to rank taxpayers for audit selection.

4.4 Data Summary

Our final sample includes a total of 5,516 taxpayers in the treatment group and 2,383 taxpayers in the control group. Figure 1 summarizes sample size by baseline year. While the control group grows throughout the sample period, the treatment group remains around 500 taxpayers in 2009, 2012, 2013, and 2014 but increases in size in 2010 and 2011. During this time period, the overall number of potential nonfiler cases in CCNIP increased from 7.1 million in 2009 to 8.4 million in 2014.[9]

[2] Census region of residence was determined from the state derived from the taxpayer's address line or ZIP code, listed on third-party forms. If Census region of residence was not present, region was set to "None."

[3] By sample design, all taxpayers in our control group met the $100,000 threshold in the baseline year.

[4] Self-employment income is restricted to the types of self-employment income required to be reported to the IRS by third parties: barter income, crop insurance, attorney fees, fishing income, medical payments, nonemployee compensation, and patronage income.

[5] Investment income includes income from distribution shares (Schedule K1), dividends (Schedule 1099-DIV), interest income (Schedule 1099-INT), and passive income (Schedule K1).

[6] Retirement income includes pension and Social Security payments.

[7] Broker transaction income is defined as income from mediating the sale or purchase of property, services, or investments (Schedule 1099-B).

[8] Other income is defined as income reported on Schedule 1099-MISC, real estate and rental income, lottery income, and business income.

[9] A glitch in CCNIP computer processing occurred in 2012, resulting in a drop in the total nonfiling taxpayers identified in that year.

FIGURE 1. Sample Size by Baseline Year

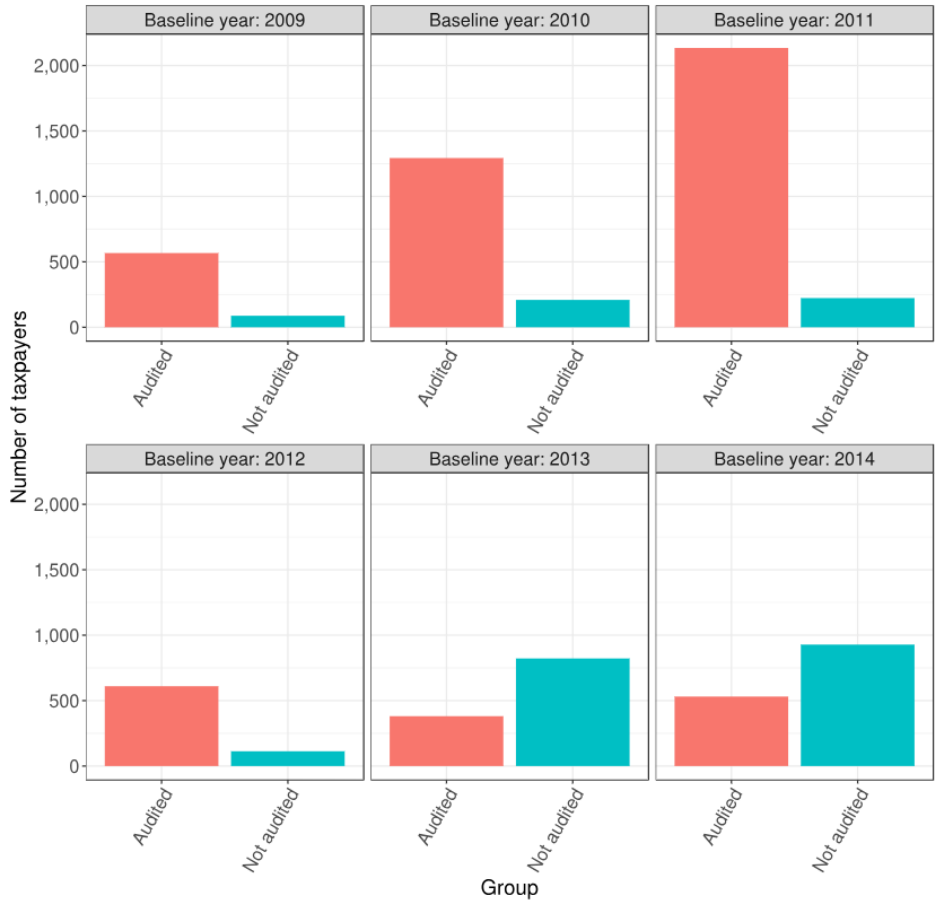
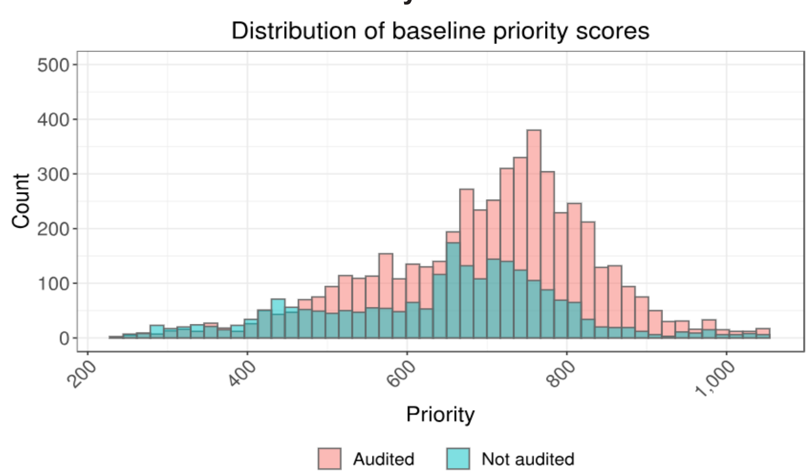


Figure 2 plots the distribution of the priority score for the treatment and control groups. While the audited group includes many more taxpayers with a priority score of 800 and above, there is common support in priority scores across both groups. This provides evidence towards the validity of our sample construction.

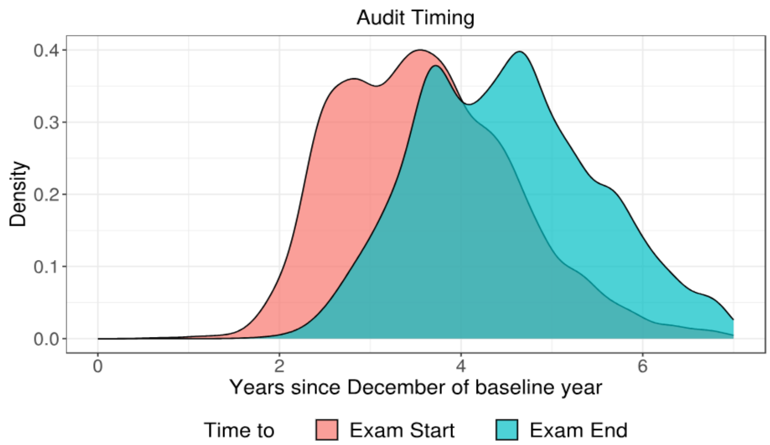
FIGURE 2. Distribution of Priority



Note: Priority is truncated for readability.

Figure 3 plots the distribution of audit start date and end date for the treatment group. The majority of audits start 2–5 years after the baseline year and end 3–6 years after the baseline year. Given this distribution of audit timing, we would not expect to see an indirect effect from an audit until at least 2 years after the baseline year.

FIGURE 3. Audit Timing



Note: Audit timing is truncated for readability.

Table 1 summarizes the control variables in our model for the treatment and control groups. As mentioned above, all control variables are sourced from the CCNIP database at the relevant baseline year. Dollar-denominated variables (Total IRP Income and Income Difference from PY) are adjusted to reflect 2018 U.S. dollars, scaled by $100,000. In terms of demographic characteristics, most taxpayers in both groups are between the ages of 30 and 65, have a filing status of single/other, and reside in a state taxing income. In terms of financial characteristics, most taxpayers in our sample earned more reported income in their baseline year than in the prior year (about $500,000 more for both groups). Taxpayers selected for audit have an average of 36 IRP documents compared to about 45 for the control group. Most taxpayers in the control group have investment income present and the majority of taxpayers in the treatment group have self-employment income. Considering past filing behavior, about 39% of the treatment group filed in the prior year while about 50% of the control group did so. More than half of the treatment group were audited prior to their baseline year, although this variable captures audits of any kind (not just TDI Lists audits). The largest differences between groups occur in the baseline priority score, whether the taxpayer met the $100,000 threshold for reported income, and whether a taxpayer was audited in the 6 years prior to baseline year.

TABLE 1. Summary Statistics for Treatment and Control Variables, TYs 2009–2014

Variable	Mean for Treatment Group	Mean for Control Group	Difference in Means
Demographic Variables			
Census Region			
East North Central	11%	8%	3%
East South Central	7%	4%	3%
Mid Atlantic	12%	12%	0%
Mountain	7%	7%	0%
New England	5%	4%	1%
Pacific	15%	16%	1%
South Atlantic	17%	20%	3%
West North Central	4%	3%	1%
West South Central	20%	16%	4%
Not Available	1%	11%	10%
Income Tax State	73%	75%	2%
Over 65	5%	7%	2%
Under 30	7%	13%	6%
PY Filing Status			
Single/other	70%	85%	15%
Married filing jointly	30%	15%	15%
PY EITC	9%	4%	5%
Financial Variables			
Total IRP Income	$5.71 (45.91)	$5.62 (28.23)	$0.09
$100,000 Threshold Indicator	54%	100%	46%
Number of IRP Forms	36.06 (232.96)	44.84 (183.65)	8.78
Income Difference from PY	$4.97 (46.36)	$5.02 (29.16)	$0.05
SE Income	69%	45%	24%
Investment Income	45%	72%	27%
Retirement Income	21%	21%	0%
Broker Transaction Income	19%	35%	16%
Other Income	29%	59%	30%
Past Filing Behavior Variables			
Filed in PY	39%	50%	11%
Ghost in PY	6%	20%	14%
Any Audit Last 6 TYs	53%	4%	49%
Baseline Priority	727 (166)	650 (163)	77

Note: Dollar-denominated variables (Total IRP Income and Income Difference from PY) are expressed in terms of 2018 dollars and are scaled by $100,000. The standard deviation for continuous variables is displayed in parenthesis. Other than Baseline Priority and Number of IRP forms, all variables reflect percentages.

5. Methodology

Our main methodological approach estimates a linear probability model of a taxpayer's fact of filing in an event-study type model.[10] This model specifies taxpayer *i*'s filing behavior in year *t* as follows:

$$Fact\ of\ Filing_{it} = \beta_0 + \beta_1 Audit_i + \beta_{2-14} Year\ from\ Baseline_{it} + \beta_{15-27} Audit_i * Year\ from\ Baseline_{it} + \alpha Taxpayer\ Controls_i + \tau Tax\ Year_t + \epsilon_{it} \quad (1)$$

Audit is a time-invariant variable set to 1 for the treatment group and 0 for the control group. The coefficient on *Audit* measures the difference in the average filing behavior across the treatment and control groups in all years. Year from baseline is a set of indicator variables that separately control for fact of filing behavior across both groups for each year, 5 years pre-baseline through 8 years post-baseline. The baseline year is the reference category and is excluded from the regression.

The primary regressors of interest are the interactions between Audit and Year from Baseline. This set of variables captures the time path of filing compliance for the audited group. We hypothesize that the estimates in pre-baseline years will be negative, indicating decreased filing compliance up to the baseline year, and that estimates in post-baseline years (beginning in the 2nd or third year from baseline) will be positive. This would indicate that audits increase the likelihood of subsequent filing and aligns with audit start dates (Figure 3). Finally, consistent with prior literature on indirect effects, we expect estimates to attenuate over time in the post-baseline period.

Taxpayer Controls are the set of demographic, financial, and filing behavior variables discussed in Section 4.3. These variables are drawn from baseline year CCNIP data and are time-invariant. Finally, *Tax Year* is a set of fixed effects capturing yearly fluctuations in fact of filing common across all taxpayers in our sample.

6. Results

6.1 Descriptive Analysis

Figure 4 visualizes the distribution of the *Fact of Filing* variable by year from baseline for taxpayers in the treatment and control groups. For the treatment group, the proportion of taxpayers filing a return decreases in the years leading up to audit and increases until the 5th year after the audited Tax Year. For the control group, the proportion of taxpayers filing a return remains around 63% in years five through two prior to baseline and decreases in the 2 years leading up to the baseline year. After the baseline year, the proportion of taxpayers in the control group filing remains between 25–40%. Despite slight differences in the proportion of taxpayers filing in either group, the treatment and control groups exhibit similar filing behavior in the years leading up to baseline. By the construction of our sample design, no taxpayer in either group filed a return in their baseline year. These plots suggest that:

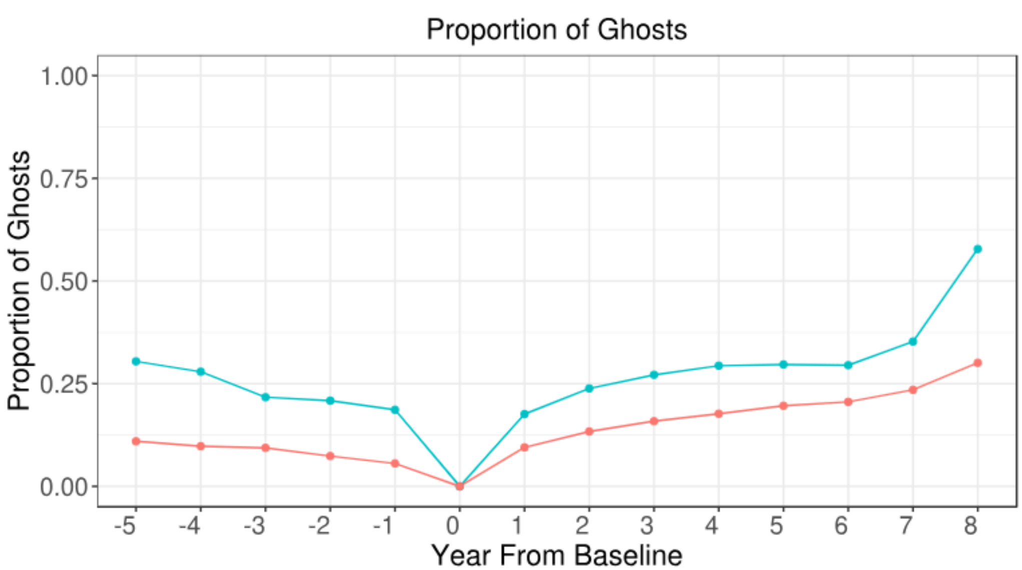
- Taxpayers in both groups trend towards delinquency in the years prior to baseline;
- The baseline year represents an outlier year for a significant portion of both the treatment and control groups, as signified by the 25–75% of taxpayers who filed a return in other years; and
- Being audited increases the likelihood of filing a return in later years (starting in the second year from baseline).

[10] A linear probability model was selected for its interpretability, although we plan to explore alternate specifications (e.g., logistic regression) in future work.

FIGURE 4. Taxpayers Filing Over Time

Figure 5 plots the distribution of ghost taxpayers in the control group (whom we consider as nonfilers in our model) and in the treatment group. The proportion of ghost taxpayers in the control group ranges from approximately 20–60%. A smaller proportion of ghosts are present in the treatment group. Both groups exhibit similar trends, with the proportion of ghosts decreasing in the years leading up to baseline year and increasing in the years following baseline year.

FIGURE 5. Ghost Taxpayers in the Treatment and Control Groups



6.2 *Econometric Analysis*

In this section, we summarize results from our main model. Full results are displayed in Table 2.

6.2.1 Audit, Year from Baseline, and Interactions

The interaction term *Audit*Year from Baseline* captures the indirect effect of an audit on filing behavior; the estimates show the time path of the fact of filing for the audited group relative to the control group. Figure 6 visualizes the coefficients and their standard errors, for the *Audit* variable and the interaction terms.

Based on audit timing (Figure 3) and descriptive analysis (see Figure 4), we hypothesized that the indirect effect would begin to appear in the second year from baseline. Table 2 confirms that this effect does begin in the second year from baseline: the audited group is 5.3–13.8 percentage points more likely to file in the 2–8 years from baseline than the control group. The magnitude of this effect increases during the 2–4 years post-baseline and decreases thereafter (an attenuation seen in much of the prior specific indirect effects literature).

The negative effect for *Year from Baseline* -1 indicates the audited group is less like to file than the control group in the year prior to their Tax Year of audit. Besides the year prior to baseline, the insignificance of interaction terms in years two and three prior to baseline and year one after baseline suggest general similarities in filing behavior across groups in the years surrounding baseline. Interestingly, the coefficient on *Audit* is not statistically significant, suggesting that irrespective of treatment and baseline year, taxpayers in the control and treatment groups exhibit similar filing behavior across Tax Years, on average. Finally, the positive coefficients estimated for the *Year from Baseline* variables indicate that both groups are more likely to file a return in off-baseline years (i.e., the baseline year is an outlier year).

FIGURE 6. Coefficients and Standard Errors for *Audit* Variables.

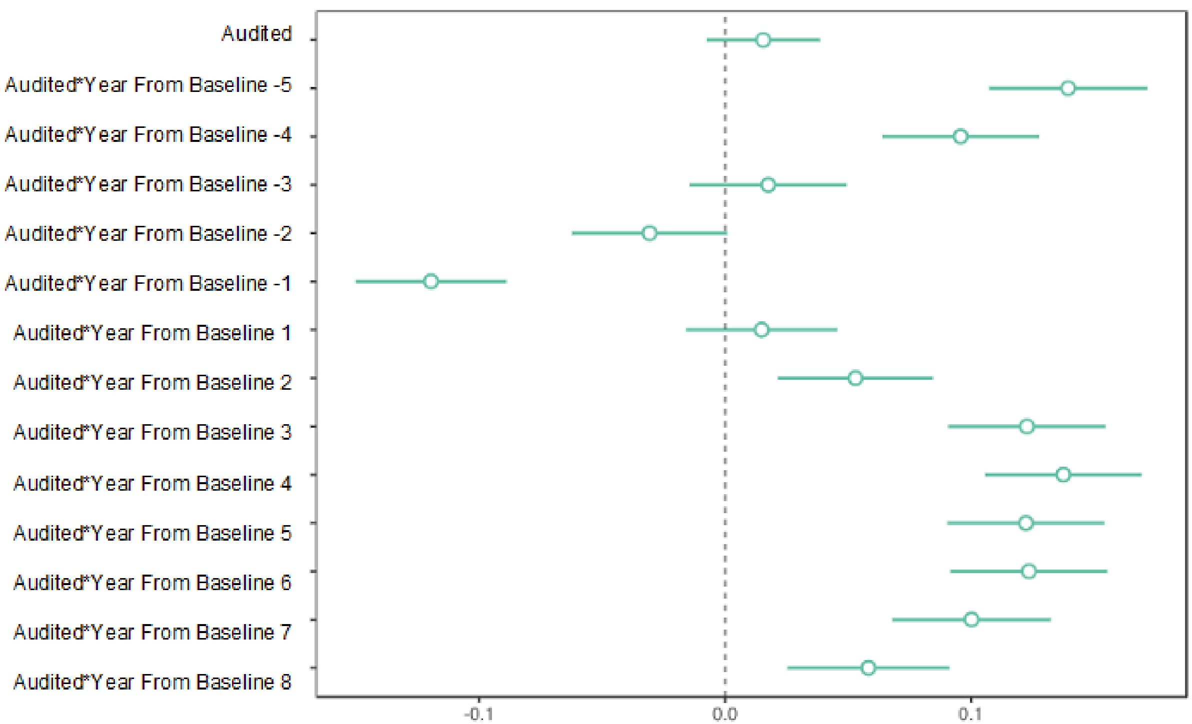


TABLE 2. Regression Results
Dependent Variable: Fact of Filing

Variable	Parameter Estimate	Standard Error
Audited	0.015	(0.012)
Year from Baseline-5	0.622***	(0.015)
Year from Baseline-4	0.635***	(0.014)
Year from Baseline-3	0.648***	(0.014)
Year from Baseline-2	0.615***	(0.014)
Year from Baseline-1	0.518***	(0.013)
Year from Baseline+1	0.306***	(0.013)
Year from Baseline+2	0.338***	(0.014)
Year from Baseline+3	0.353***	(0.014)
Year from Baseline+4	0.379***	(0.014)
Year from Baseline+5	0.398***	(0.015)
Year from Baseline+6	0.389***	(0.016)
Year from Baseline+7	0.41***	(0.016)
Year from Baseline+8	0.434***	(0.017)
Audited*Year from Baseline-5	0.139***	(0.016)
Audited*Year from Baseline-4	0.096***	(0.016)
Audited*Year from Baseline-3	0.018	(0.016)
Audited*Year from Baseline-2	-0.031	(0.016)
Audited*Year from Baseline-1	-0.12***	(0.016)
Audited*Year from Baseline+1	0.015	(0.016)
Audited*Year from Baseline+2	0.053***	(0.016)
Audited*Year from Baseline+3	0.123***	(0.016)
Audited*Year from Baseline+4	0.138***	(0.016)
Audited*Year from Baseline+5	0.122***	(0.016)
Audited*Year from Baseline+6	0.124***	(0.016)
Audited*Year from Baseline+7	0.1***	(0.016)
Audited*Year from Baseline+8	0.058***	(0.017)
Over 65	-0.031***	(0.006)
Under 30	-0.015***	(0.005)
PY EITC	0.033***	(0.005)
PY Filing Status = Married Filing Jointly	0.086***	(0.003)
Census Region: East South Central	-0.012	(0.007)
Census Region: Mid Atlantic	0.02***	(0.006)
Census Region: Mountain	-0.004	(0.007)
Census Region: New England	0.034***	(0.008)
Census Region: Not Available	-0.064***	(0.008)
Census Region: Pacific	0.022***	(0.005)
Census Region: South Atlantic	0.003	(0.005)
Census Region: West North Central	-0.02**	(0.008)
Census Region: West South Central	-0.015**	(0.006)
Income Tax State	-0.011***	(0.004)
Number of IRP Forms	0.000***	(0.000)

TABLE 2. Regression Results (Continued)
Dependent Variable: Fact of Filing

Variable	Parameter Estimate	Standard Error
Broker Transaction Income	0.023***	(0.004)
Other Income	0.01***	(0.003)
Retirement Income	0.026***	(0.004)
Investment Income	0.078***	(0.003)
SE Income	-0.007**	(0.003)
$100,000 Threshold Indicator	-0.026***	(0.004)
Income Difference from PY	0.000**	(0.000)
Total IRP Income	0.000***	(0.000)
Baseline Priority	0.000***	(0.000)
Any Audit Last 6 TYs	0.006	(0.004)
Filed in PY	0.218***	(0.003)
Ghost in PY	-0.073***	(0.005)
Constant	-0.239***	(0.024)
Observations	110,586	
R-squared	0.211	
Adjusted R-squared	0.211	
F Statistics	411.224***	(df = 72; 110,513)

Note: **p<0.05, ***p<0.01.

6.2.2 Demographic Variables

Taxpayers over 65 and under 30 have a lower probability of filing. Taxpayers claiming EITC in the year prior to baseline are 3.3 percentage points more likely to file in other years, and taxpayers with a filing status of Married Filing Jointly are 8.6 percentage points more likely to file in other years. Results across Census regions confirm prior literature findings that compliance varies across geography. Finally, living in a state that imposes an income tax decreases the likelihood of filing by 1.1 percentage points.

6.2.3 Financial Variables

Having certain types of income, such as broker transaction income, retirement income, investment income and other income increases the likelihood of filing. The presence of investment income has the strongest effect compared with other types of income sources, increasing the likelihood of filing by 7.8 percentage points. Having self-employment income *decreases* the likelihood of filing. However, since only certain categories of self-employment income are covered by third-party reporting, measurement error may obscure a true positive effect. Taxpayers who met the $100,000 threshold for reported income in their baseline year are 2.9 percentage points less likely to file, suggesting a negative relationship between income and propensity to file. However, other income-related variables, *Income Difference from PY* and *Total IRP Income,* do not have an effect on the likelihood of filing.

6.2.4 Past Filing Behavior Variables

Interestingly, a taxpayer's priority score and the presence of any type of audit in the 6 years prior to baseline do not have an effect on the probability of filing. Taxpayers who filed a return in the year prior to baseline are 21.8 percentage points more likely to file in other years. If the taxpayer is a ghost in the year prior to baseline, their expected probability of filing in other years decreases by 7.3 percentage points. Together, these estimates suggest there is persistence in filing behavior.

6.3 *Robustness Checks*

6.3.1 Definition of Ghosts

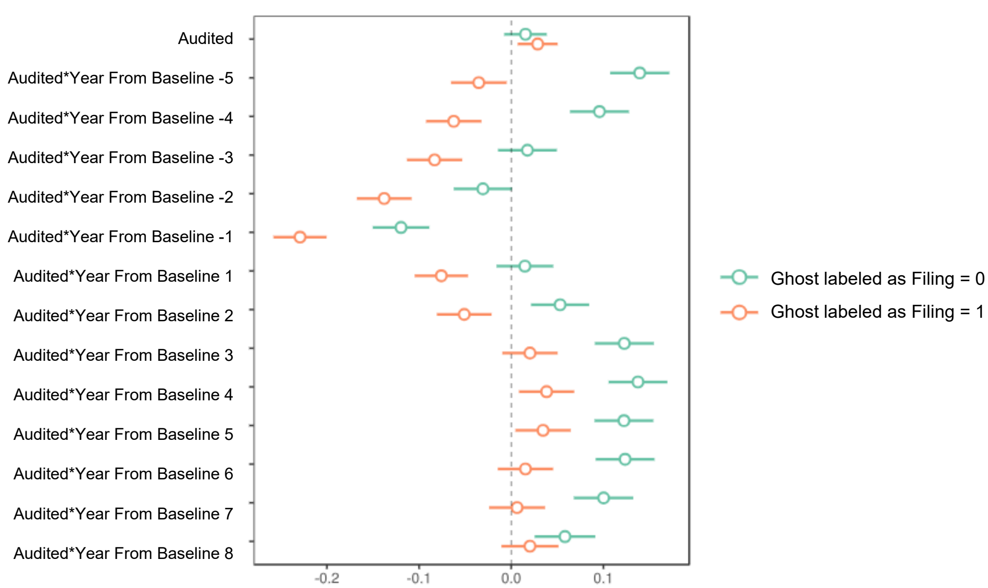
Our construction of the *fact of filing* variable assigns ghosts to the nonfiling status on the assumption that individuals who did not file and do not have third-party reported income still have a tax liability. We explore the validity of this assumption by rerunning our regression, labeling ghosts as *fact of filing* = 1. Table 3 in the Appendix contains the regression results.

Figure 7 compares the coefficients and standard errors for *Audit* variables between the linear probability model categorizing ghosts as *fact of filing* = 0 (Table 2) and the linear probability model categorizing ghosts as *fact of filing* = 1 (Table 3). The relationship between *Audit* and *Year from Baseline* on a taxpayer's *fact of filing* is smaller, although similar trends are observed. The negative effect for *Audit*Year from Baseline* values of -5 through 2 indicates the audited group is less likely to file in all years leading up to knowledge of the audit start, compared to the control group. While an indirect effect of an audit on a taxpayer's future filing behavior is observed, it is significant only in years 4 and 5 from baseline, with an expected increase in the likelihood of filing ranging from 3.5–3.8 percentage points.

By changing this assumption, the *Audit* variable becomes statistically significant at the 95-percent level, indicating taxpayers in the audited group are 2.9 percentage points more likely to file across all years. Likewise, the presence of any audit in the 6 years prior to baseline now has a negative and significant effect. The relationship between whether a taxpayer was a ghost in the year prior to baseline and their filing behavior across all years changes to a positive relationship, increasing the likelihood of filing by 14 percentage points. The influence of a taxpayer's filing in the year prior to baseline decreases, while still being positive.

Overall, the magnitude and levels of our estimates are sensitive to the definition of ghost taxpayers, however, the trends and overall takeaways seem to be consistent despite different assumptions around these taxpayers.

FIGURE 7. Coefficients and Standard Errors for *Audit* Variables Depending on the Definition of Ghosts.



7. Discussion

Understanding nonfiling behavior is difficult due to the inherent lack of information about these taxpayers. It is especially curious why higher earning nonfilers do not file a return, given that they may be more visible to the tax authority and because they are identified by third party data as potentially owing substantial taxes. As our analysis reveals, taxpayers who do not file a return often fail to do so in other years as well. These challenges require creative approaches to sample selection and model design to analyze these taxpayers' fact of filing. In this paper, we provide one of the first attempts to understand the role of audits on future filing of audited nonfilers—a contribution to the nonfiling literature, which has largely focused on the role of non-enforcement related filing determinants.

Our results indicate that in-person audits of nonfilers improve subsequent filing compliance. Nonfilers subject to a Field audit are 5.3–13.8 percentage points more likely to file in the 2–8 years after baseline (i.e., the tax year of the audited return), compared to similar but unaudited taxpayers. This effect peaks in the 4th year after baseline. This finding is qualitatively similar to findings from other studies on the effect of audits on future tax compliance.

The sensitivity analysis on our assumption of ghosts owing a tax liability confirms the presence of an indirect effect irrespective of our assumption. However, if all ghosts in our sample are missing from IRS records because they do not have a tax liability, our estimates of an indirect effect are overstated.

This study also sheds light on the relationship between prior and future filing behavior. Taxpayers who filed in the year prior to baseline year are 21.8 percentage points more likely to file in other years (regardless of being audited). On the other hand, taxpayers who were unable to be identified through IRS data processes (ghosts) in the year prior to the baseline year are 7.3 percentage points less likely to file in other years.

7.1 Comparison with Other Research

We compare our contributions to other research estimating the influence of IRS treatment on nonfiling taxpayer's filing behavior.

Datta *et al.* (2015) modeled the filing behavior of nonfilers over a 4-year period after treatment by the ASFR program between TYs 2007 and 2009. The ASFR program similarly treats nonfilers identified by CCNIP, but focuses on lower income individuals with simpler returns—individuals with primarily wage income and fewer forms reported to the IRS by third parties. The ASFR program differs from an in-person Field audit as it treats nonfilers by estimating their balance due and interacts with taxpayers via mailed correspondence. The study estimated treatment by ASFR resulted in an 11, 21, and 27 percentage point increase in the likelihood of filing 2 to 4 years after treatment, respectively. Comparatively, our estimated effects of a Field audit are smaller in magnitude. However, as our study monitored filing behavior through 8 years after baseline, we observed an effect persisting through the 8th year.

Herlache *et. al* (2019) conducted a randomized control trial to explore the influence of various mailed notice programs on nonfilers' prior year noncompliance in TY 2016 and filing behavior in TYs 2017 and 2018. Their population similarly consisted of nonfilers identified from CCNIP in TY 2016, but did not have any additional selection criteria from this population. Across the various treatments employed, they identified a 6.7% increase in filing of prior-year returns for TY 2016 and a 4.1–4.6% increase in filing for current and future returns, as a result of mailed reminder-to-filer notices. This study extended analysis to consider the role of notices on expected stopfilers, taxpayers compliant in prior years yet predicted to be at risk of becoming noncompliant, identifying a 1.6 to 2.4 percentage increase in filing.

Both mailed reminder-to-file notices and treatment by ASFR are less expensive and require fewer resources than in-person Field audits, enabling them to treat a larger population of nonfilers. Despite differences in type of treatment, tax years, and populations considered, the impact of mailed outreach was smaller and the impact of ASFR treatment was larger, compared with our estimates. The study by Herlache *et. al* (2019) provides insight into the role of notices as an instrument of deterrence prior to noncompliance behavior. This type of treatment may be an effective form of treatment when nonfilers are unaware of either the need to file, the consequences from not filing, or their detectability by the IRS. Datta, *et. al* (2015) illustrates the influence of correspondence treatment on identified nonfilers. As ASFR focuses on a different population than Field audits, the difference in estimates may be indicative of higher compliance rates in lower income and single source of income populations.

Nonfiling individuals audited by Field tend to owe a substantial tax liability, have complex returns, and remain noncompliant, despite being mailed up to two delinquency notices. These high dollar noncompliance individuals are a strategic focus of IRS enforcement moving forward (IRS (2023)). Nonfilers are responsible for a substantial portion of the tax gap, and these results highlight the value of audits as a policy tool to promote compliance. While audits of this group can be costlier than audits of simpler returns, maintaining audit coverage of the nonfiling population provides an opportunity to promote voluntary compliance as well as generate direct revenue.

7.2 Limitations and Future Research

There are several near-term extensions we plan to address. Foremost, we plan to estimate the effect of a nonfiler audit on *total tax* reported by the taxpayer in subsequent years. This extension will provide dollar-valued estimates of the indirect effect that can be compared against direct revenue as well as the cost to the IRS of conducting these audits. Other extensions include additional robustness checks, collecting a richer set of control variables, refining sample design, and considering a multinomial outcome variable, which are described in more detail below.

7.2.1 Robustness Checks

To further validate model estimates and to assist with removing confounding, future work will employ propensity-score matching or inverse probability weighting methods to the baseline year variables. These techniques will allow for an alternative and possibly more comparable treatment and control group when estimating the casual effect of an audit.

7.2.2 Richer Control Variables

Our current analysis relies on third-party reported data available only for the taxpayer's baseline year. This limits the ability to control for time-varying characteristics that influence filing behavior. Future work may consider constructing control variables for the years leading up to the baseline year. This extension poses data availability challenges since a significant portion of these taxpayers are ghosts in off-baseline years.

Future work may also incorporate additional control variables not considered here. By capturing taxpayers' reported income in off-baseline years, we could verify the presence of a tax liability. As discussed in Section 3, CCNIP and the Returns Delinquency Program issue notices to inform taxpayers of their delinquent tax liability. Additional control variables can include whether the taxpayer received a notice and the type of notice received. Alternatively, receiving a delinquency notice could serve as a treatment variable itself (instead of a TDI Lists audit). We could also control for the pattern of nonfiling: whether the taxpayer was a one-time nonfiler, a nonfiler with a pattern of interrupted nonfiling, or a nonfiler with a continuous pattern of nonfiling.

7.2.3 Sample Design

A limitation of this study is in the handling of taxpayers with multiple audits. Seventy-two percent of our treatment group received at least one other TDI Lists audit in off-baseline years,[11] while 5.2% experienced another type of audit in off-baseline years. We retain these multi-audited taxpayers in the current version of our analysis to preserve sample size, but audits occurring before the baseline year likely cause us to underestimate the indirect effect of audit. Future research can include sensitivity analysis of the indirect effect on single-audited taxpayers versus multiple-audited taxpayers.

[11] Our deduplication rules assign multi-audited taxpayers to their first year of audit within the baseline period 2009–2014. However, taxpayers assigned to baseline year 2009 could still have been audited prior to 2009. 2,120 taxpayers had a TDI Lists audit in both the 6 years prior to baseline and 6 years post baseline. 864 taxpayers had a TDI Lists audit in the 6 years prior to baseline only, and 1,142 taxpayers had a TDI Lists audit in the 6 years after baseline only.

7.2.4 Multinomial Outcome

The fact of filing outcome variable captures whether a taxpayer files for a given year. However, this variable does not capture *compliance*. A compliant taxpayer files voluntarily and timely. Future research can consider a three-level outcome variable to capture the distinctions between timely filing, late filing, and nonfiling. This would extend this analysis to estimate the indirect effect of a Field audit on a nonfiler's future compliance.

References

Datta, Saurahh, Stacy Orlett, and Alex Turk. 2015. "Individual Nonfilers and IRS-Generated Tax Assessments: Revenue and Compliance Impacts of IRS Substitute Assessments When Taxpayers Don't File." Internal Revenue Service. https://www.irs.gov/pub/irs-soi/15rescondatta.pdf.

Erard, Brian and Chih-Chin Ho. 2001. "Searching for Ghosts: Who Are the Nonfilers and How Much Tax Do They Owe?" *Journal of Public Economics* 81(1): 25-50. https://doi.org/10.1016/S0047-2727(00)00132-8.

Erard, Brian, Patrick Langetieg, Mark Payne, and Alan Plumley. 2020. "Ghost in the Income Tax Machinery". *Munich Personal RePEc Archive* 100036. https://mpra.ub.uni-muenchen.de/100036/.

Erard, Brian, Tom Hertz, Pat Langetieg, Mark Payne, and Alan Plumley. 2022 "To File or Not To File? What Matters Most?" Working Paper.

Gangl, Katharina, Erich Kirchler, Christian Lorenz, and Benno Torgler. 2015. "Wealthy Tax Non-Filers in a Developing Country: Taxpayer Knowledge, Perceived Corruption, and Service Orientation in Pakistan." In B. Peeters, H. Gribnau, & J. Badisco (Authors), Building Trust in Taxation (pp. 355-376). Intersentia. https://doi.org/10.1017/9781780684734.

Government Accountability Office. 2015. "IRS Return Selection: Certain Internal Controls for Audits in the Small Business and Self-Employed Division Should Be Strengthened." GAO-16-103. https://www.gao.gov/products/gao-16-103.

Herlache, Anne, Ishani Roy, Alex Turk, and Stacy Orlett. 2019. "Enforcement Versus Outreach – Impacts on Tax Filing Compliance" *The IRS Research Bulletin* Publication 1500 (Rev. 6-2020). https://www.irs.gov/pub/irs-prior/p1500--2020.pdf.

Internal Revenue Service. 2019. "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013." Publication 1415 (Rev. 09-2019). https://www.irs.gov/pub/irs-prior/p1415--2019.pdf.

Internal Revenue Service. 2022. "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2014-2016." Publication 1415 (Rev. 08-2022). https://www.irs.gov/pub/irs-pdf/p1415.pdf.

Internal Revenue Service. 2023. "Internal Revenue Service Inflation Reduction Act Strategic Operation Plan FY 2023-2013." Publication 3744 (Rev. 04-2023). https://www.irs.gov/pub/irs-pdf/p3744.pdf.

Langetieg, Patrick, Mark Payne, and Alan Plumley. "Counting Elusive Nonfilers Using IRS Rather Than Census Data". 2017. Internal Revenue Service. https://www.irs.gov/pub/irs-soi/17resconpayne.pdf.

Robson, Jennifer, and Saul Schwartz. 2020. "Who Doesn't File a Tax Return? A Portrait of Non-Filers." *Canadian Public Policy* 46(3): 323-339. https://doi.org/10.3138/cpp.2019-063.

Santoro, Fabrizio, Edward Groening, Winnie Mdluli, and Mbongeni Shongwe. 2021. "To File or Not To File? Another Dimension of Tax Compliance – the Eswatini Taxpayer's Survey." *Journal of Behavioral and Experimental Economics* 95 101760. https://doi.org/10.1016/j.socec.2021.101760.

Tagkalakis, Athanasios O. 2014. "The Direct and Indirect Effect of Audits on the Tax Revenue in Greece." *Economics Bulletin* 34(2): 984-1001. http://www.accessecon.com/Pubs/EB/2014/Volume34/EB-14-V34-I2-P91.pdf.

Appendix

FIGURE 8. Proportion of Audited Taxpayers Meeting $100,000 Income Threshold

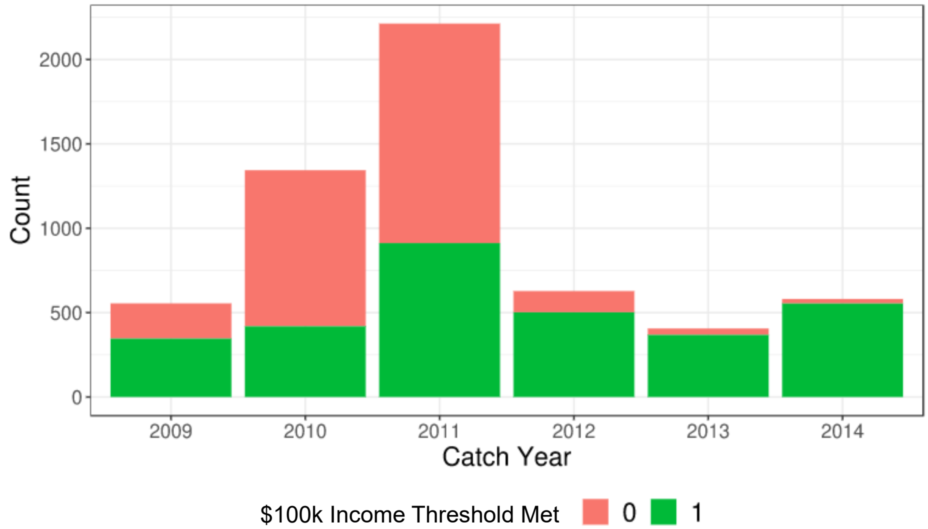


TABLE 3. Regression Results: Labeling Ghosts as Filing = 1

Dependent Variable: Fact of Filing

Variable	Parameter Estimate	Standard Error
Audited	0.029***	(0.011)
Year from Baseline-5	0.89***	(0.014)
Year from Baseline-4	0.879***	(0.014)
Year from Baseline-3	0.837***	(0.013)
Year from Baseline-2	0.785***	(0.013)
Year from Baseline-1	0.679***	(0.012)
Year from Baseline+1	0.49***	(0.012)
Year from Baseline+2	0.594***	(0.013)
Year from Baseline+3	0.65***	(0.013)
Year from Baseline+4	0.704***	(0.014)
Year from Baseline+5	0.74***	(0.014)
Year from Baseline+6	0.774***	(0.015)
Year from Baseline+7	0.8***	(0.015)
Year from Baseline+8	0.793***	(0.016)
Audited*Year from Baseline-5	-0.035**	(0.015)
Audited*Year from Baseline-4	-0.062***	(0.015)
Audited*Year from Baseline-3	-0.083***	(0.015)
Audited*Year from Baseline-2	-0.138***	(0.015)
Audited*Year from Baseline-1	-0.229***	(0.015)
Audited*Year from Baseline+1	-0.076***	(0.015)
Audited*Year from Baseline+2	-0.051***	(0.015)
Audited*Year from Baseline+3	0.02	(0.015)
Audited*Year from Baseline+4	0.038**	(0.015)
Audited*Year from Baseline+5	0.035**	(0.015)
Audited*Year from Baseline+6	0.015	(0.015)

TABLE 3. Regression Results: Labeling Ghosts as Filing = 1 (Continued)
Dependent Variable: Fact of Filing

Variable	Parameter Estimate	Standard Error
Audited*Year from Baseline+7	0.006	(0.015)
Audited*Year from Baseline+8	0.02	(0.016)
Over 65	0.012**	(0.006)
Under 30	0.058***	(0.005)
PY EITC	0.029***	(0.005)
PY Filing Status = Married Filing Jointly	0.046***	(0.003)
Census Region: East South Central	-0.006	(0.007)
Census Region: Mid Atlantic	0.03***	(0.005)
Census Region: Mountain	0.009	(0.006)
Census Region: New England	0.037***	(0.007)
Census Region: Not Available	0.073***	(0.007)
Census Region: Pacific	0.035***	(0.005)
Census Region: South Atlantic	0.023***	(0.005)
Census Region: West North Central	-0.013	(0.008)
Census Region: West South Central	0.003	(0.006)
Income Tax State	-0.006	(0.004)
Number of IRP Forms	0.000***	(0.000)
Broker Transaction Income	-0.02***	(0.004)
Other Income	-0.019***	(0.003)
Retirement Income	-0.012***	(0.003)
Investment Income	0.012***	(0.003)
SE Income	-0.014***	(0.003)
$100,000 Threshold Indicator	-0.018***	(0.003)
Income Difference from PY	0.000**	(0.000)
Total IRP Income	0.000***	(0.000)
Baseline Priority	0.000***	(0.000)
Any Audit Last 6 TYs	-0.023***	(0.003)
Filed in PY	0.172***	(0.003)
Ghost in PY	0.14***	(0.005)
Constant	-0.129***	(0.023)
Year Fixed Effects	Y	
Observations	110,586	
R-squared	0.25	
Adjusted R-squared	0.249	
F Statistics	510.736***	(df = 72; 110513)

***p<0.05, ***p<0.01.*

Silver Lining: Estimating the Compliance Response to Declining Audit Coverage[1]

Alan Plumley and Daniel Rodriguez (IRS, RAAS), Jess Grana and Alexander McGlothlin (MITRE)

I. Introduction

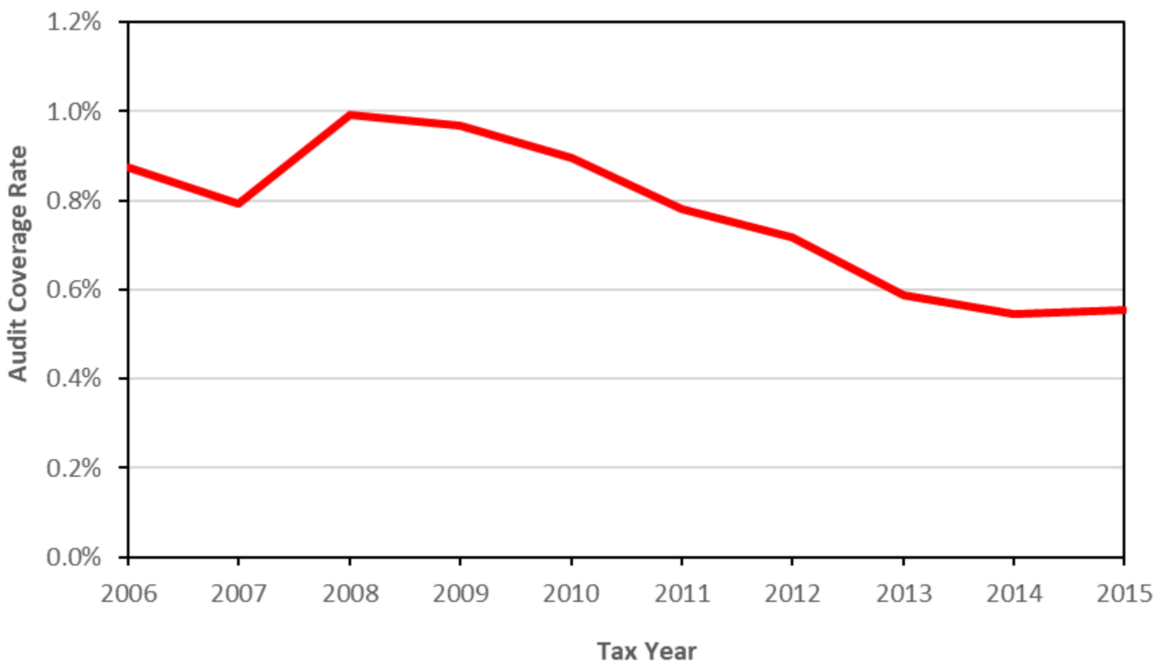
How much additional revenue would likely be generated if the IRS enforcement budget were increased by $X per year? The answer to that question is far from simple. It likely depends on things such as the size of the current budget and how it is allocated to enforcement, services, IT investments, etc. It will also depend on how the current and additional enforcement budgets are allocated to the various enforcement programs. Certainly, one impact on overall revenue would come in the form of increased direct enforcement revenue (i.e., the additional tax paid by those audited for the tax year that was subjected to the enforcement). However, it is also likely that this direct effect of increased enforcement would be accompanied by some *indirect* revenue effects—whether due to a subsequent change in compliance behavior among the specific taxpayers who were the subjects of the enforcement (known as the "specific indirect effect"), or due to a change in compliance behavior among taxpayers in the general population who were *not* the subjects of the enforcement (known as the "general indirect effect").

There have been numerous attempts over the last 40 years to estimate the general indirect effect of changes in IRS enforcement—particularly changes in audit coverage rates. Unlike the direct effect of audits, the indirect effects are not directly observable. In principle, any indirect effect on voluntary compliance is the difference between the tax that taxpayers pay given the audits and the tax they *would have* paid had the audits not happened. Because the counterfactual amount of tax that would have been paid in the absence of the audits cannot be observed, it must be estimated.

There have been two major approaches to estimating the general indirect effect of audits: (1) demonstration models; and (2) comprehensive models. Demonstration models attempt to demonstrate that a general indirect effect exists, at least in a given context—typically within a particular segment of the population (e.g., sole proprietors) through a specific type of network (such as the network of taxpayers who are clients of the same tax preparer) and according to a particular behavioral mechanism (e.g., deterrence). However, even if such analyses do demonstrate that taxpayers in audited networks behave differently than taxpayers in unaudited networks (for example), it would seem likely that many taxpayers participate in multiple networks simultaneously (e.g., employer networks, professional networks, community networks, etc.), and it is not clear that the separate impact of these multiple networks on such a taxpayer would be additive; taxpayers undoubtedly form their perceptions in a more subtle way based on all the factors in their environment. Although demonstration models do lend themselves to theoretical premises and practical experimentation, such narrowly defined analyses would not answer the question posed at the beginning of this paper about the impact of an increase (or decrease) in the overall enforcement budget. That is one rationale for a comprehensive model. Such models are agnostic about the mechanism(s) affecting taxpayer behavior and are generally not restricted to a narrow subset of the population. However, they depend heavily on being able to control for all the main drivers of behavior in addition to the enforcement activity in question.

This paper represents an initial attempt to estimate comprehensive models of the impact of individual income tax audits on the general population. It is motivated by the observation that, due to a steady decline in IRS budgets over the last 12 or so years, individual income tax audit coverage rates (the percentages of any given subpopulations that are audited) have declined substantially (see Figure 1). Although the decline in IRS budgets that precipitated this decline in audit rates has been a dark cloud over tax administration throughout this period, this dark cloud may have a silver lining: it provides us with an excellent natural experiment for determining whether that sustained decline in audit coverage might have prompted an increase in noncompliance (perhaps with the hope that a recovery of audit coverage might regain some or all of any loss in compliance). Fortunately, the IRS conducted National Research Program (NRP) audits on a separate stratified random sample of individual income tax returns each tax year (TY) from 2006–2014. The results of these audits, weighted up to represent the entire population of tax returns in each year, allow us to determine whether the population (either as a whole or with respect to subpopulations therein) increased their noncompliance as audit rates fell.

FIGURE 1. Audit Coverage* Trend Among Individual Income Tax Returns, TYs 2006–2015



* Coverage rate = (number of returns audited) / (total number of returns filed) for the tax year

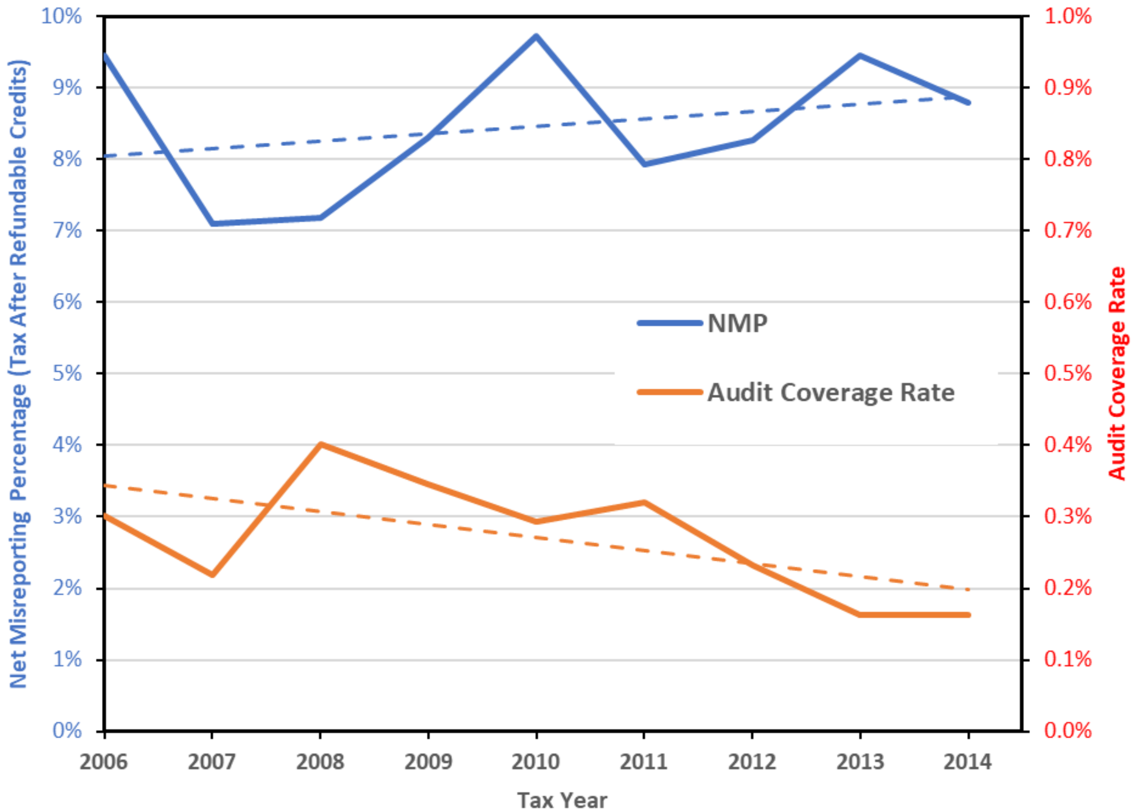
Figure 2 illustrates the relationship during this time period between the audit coverage rate trend and the trend in noncompliance—measured by the Net Misreporting Percentage (NMP)[2] on Tax After Refundable Credits (TARC)—for IRS Examination Activity Code 272,[3] which includes over half of all individual income tax returns.[4] Figure 2 shows an overall upward trend in noncompliance contemporaneous with a declining trend in the audit coverage rates over these years, suggesting the presence of a general indirect effect among this large group of taxpayers.

[2] The NMP is defined as the aggregate net amount misreported on a given line item across a group of returns divided by the sum of the absolute values of the corresponding amounts that should have been reported. The absolute values are used in the denominator to ensure that negative amounts do not distort the aggregates.

[3] IRS divides the population into mutually exclusive activity code categories based on characteristics like tax return type, the amount of income, gross receipts, or assets, and whether certain tax benefits are claimed. See Table 5.

[4] Schedules C and F are used to report nonfarm and farm sole proprietor income and expenses, respectively; Schedule E is used to report income from rental real estate, royalties, partnerships, S corporations, estates, trusts, or residual interests in real estate mortgage investment conduits; and Form 2106 is used to report employee business expenses.

FIGURE 2. Audit Coverage and NMP Trends, Activity Code 272,[5] TYs 2006–2014



This finding seems to support the widely held presumption that IRS enforcement conducted on a few taxpayers (who are thought to be noncompliant) indirectly has a positive impact on the compliance behavior of the general population, although it is far from conclusive. Consider, for example, that:

- Similar plots for other segments of the population do not exhibit a clearly negative relationship between audit rate and noncompliance trends (see Figure 14 of the Appendix).
- Taxpayers likely do not react to (or even know about) contemporaneous trends in audit coverage; their perceptions may form over time.
- A negative correlation between audit coverage and noncompliance does not prove causation. There are likely many other factors—including other IRS actions, tax policy changes, and societal trends—that influence taxpayer behavior.

For these and other reasons, this paper applies econometric techniques to the NRP data to isolate any general indirect effects from other factors that influence voluntary compliance. This is ongoing work. We have focused on the impact of audit coverage rates for this paper, but we intend to expand the scope eventually to include other IRS actions as data and statistical considerations permit. The paper is organized as follows: Section 2 reviews the relevant empirical literature; Section 3 describes our data; Section 4 summarizes our estimation methods; Section 5 presents our empirical results; and Section 6 concludes the paper.

2. Literature Review

The literature on the impact of deterrence on taxpayer noncompliance includes the specific indirect effects and general indirect effects of IRS enforcement—both a taxpayer's prior audits and their knowledge of other audits

[5] Total Positive Income < $200,000 and No EITC or Schedules C, E, F, or Form 2106. These represent 55.3% of the population over this time period.

affect their perceived probability of being audited, which may influence their compliance choices. This paper is relevant to general indirect effects—the effect of IRS contacts (such as audits) on those who are mostly[6] not contacted themselves. Most studies of the general indirect effect are demonstration models that focus on one context and indirect mechanism (e.g., communication from a tax return preparer within his network of clients about IRS audits conducted on a subset of his clients). While this is conceptually straightforward and may demonstrate the *existence* of a general indirect effect, there are bound to be many other contexts and mechanisms that produce other general indirect effects. Indeed, any given taxpayer may be a member of multiple such networks, and their combined effect may not be the simple sum of their separately estimated effects. We, therefore, consider in this paper comprehensive models that capture the indirect effect of *all* enforcement and service activities across the taxpayer population regardless of the many behavioral mechanisms that may be involved.

2.1 Demonstration Models

Most papers in the general indirect effects literature trace 1) the effect of certain audits or contacts (either operational or experimentally assigned), 2) on a subpopulation (e.g., Earned Income Tax Credit (EITC) claimants), and 3) within defined networks (e.g., geographic, preparer, supply chain). Using well-identified networks supports strong identification strategies whereby a treatment group (i.e., a network that had an audited member) is compared against a similar, but untreated group. However, the disadvantage of this approach is that findings may not be generalizable outside of a specific context or behavioral mechanism.

For example, Boning *et al.* (2020) find that the indirect effect of audits propagates through tax preparer networks; when a firm is audited, other firms sharing the same tax preparer also remit more tax thereafter. Bohne and Nimczik (2018) find that tax avoidance behaviors follow managers and tax experts as they transfer between firms. Pomeranz (2015) finds that after a firm is audited, tax compliance also improves among that firm's suppliers. Chetty *et al.* (2013) find that EITC knowledge (as proxied by income bunching) diffuses through geographic networks. Regarding individual audits, some papers find evidence of geographic spillovers (Drago, Mengel, and Traxler (2020)) and spillovers within family networks (Alstadsæter, Kopczuk, and Telle (2019)), while others find mixed or no evidence of an indirect effect (Meiselman (2018); Perez-Truglia and Troiano (2018); Grana *et al.* (2022)). These mixed results indicate that context matters,: whether there is an indirect effect depends on the quasi-experimental research design choices, the specific network or community studied, or even the country of focus.

2.2 Comprehensive Models

While useful for qualitative understanding of the nature of deterrence, these prior findings do not directly translate into operational applications such as budget justification. For example, IRS budget requests to Congress cite a return on investment (ROI) of about \$6 in direct revenue for every \$1 of enforcement funding, "before considering the significant deterrence effects" (IRS (2023)). The "significant deterrence effects" in theory include effects arising 1) from all IRS enforcement activities, 2) across the general taxpayer population, and 3) across all possible (or as many as possible) networks of propagation. In other words, the estimated indirect effects should be as comprehensive as possible.

A handful of papers target this "comprehensive indirect effect" by evaluating the effect of audit *rates* on the general population. Instead of constructing taxpayer-level networks, these papers typically model compliance in the aggregate (e.g., state or zip code level). A common approach evaluates the effect of the contemporaneous audit rate (as a proxy for audit probability) on compliance using an instrumental variable estimation method. This method is often applied because of the endogeneity problem that arises when not only do audit rates impact taxpayer compliance behavior, but that behavior also influences audit rates. Findings using this approach also are mixed.

For example, using state-level panel data, Dubin, Graetz and Wilde (1990), Plumley (1996), and Dubin (2007) find that the indirect effect of audits are six, eleven, and nine times that of the direct effect, respectively. Dubin and Wilde (1988) and Grana *et al.* (2022) use ZIP code level panel data and find mixed evidence of an

[6] For a review of the literature on the specific indirect effect of audits, see Nicholl *et al.* (2020).

indirect effect, varying across taxpayer subpopulations and audit categories. In some cases, these papers find an unexpected positive association between audit rates and compliance (such as among high income nonbusiness taxpayers).

Tauchen, Witte, and Beron (1993) use microdata from the IRS Taxpayer Compliance Measurement Program (TCMP)[7] and find that the indirect effect of audits is twice the size of the direct effect—but is only statistically significant for high income taxpayers. Hoopes, Mescall and Pitman (2012) take a similar approach using corporate returns and find that doubling the audit rate increases effective tax rates by 7%. Notably, they survey corporate tax executives and find that many take note of historical audit rates. Due to these mixed findings, a conservative estimate put forward by the U.S. Treasury is an indirect effect that is three times the direct effect (Treasury (2019)).

This paper adds to this literature by using alternative model specifications and exploiting new data. Like Tauchen, Witte, and Beron (1993), we use individual microdata from the successor to TCMP to capture noncompliance. We also conduct a parallel analysis using compliance measures captured by the automated document-matching program. We differ from prior research in our econometric specification: instead of the contemporaneous audit rate, we evaluate the effect of lagged audit rates on compliance. As we discuss in Section 4, lagged audit rates are more likely to be the correct specification of taxpayer knowledge about IRS enforcement, and it eliminates the possibility that the audit rate is endogenous with taxpayer compliance.

3. Data

Our methodology relies on modeling individual level compliance as a function of IRS audit rates, while controlling for other drivers of compliance. Our primary compliance measure is derived from NRP microdata, which allows us to control for return-level variables. However, we interpret the behavior of the individuals in the NRP sample as being representative of similar taxpayers in the general population. Therefore, we are interested in the aggregate audit rate faced by the segment of the population represented by the NRP taxpayer—not the audit probability of the taxpayer in the NRP sample. Audit rates are constructed by aggregating IRS enforcement data.

3.1 Dependent Variables

We use data on returns audited through the NRP, which selects a stratified random sample of individual income tax returns for examination for a given tax year. Because the NRP sample is designed to be representative of the population, audits through the NRP examine taxpayers who might not have been examined under normal operational audit procedures. These audits potentially encompass the whole tax return, as opposed to targeting specific areas of noncompliance, as in operational audits. The program provides useful information about noncompliance among the general population and the insights it reveals are used to update operational audit selection procedures, improve resource allocation, and provide estimates of the tax gap (IRS (2022)).

We select all returns audited through the NRP for TYs 2006−2014.[8] For each return, we use the reported amounts and NRP-corrected amounts of certain line items. Our primary outcome variable is the Net Misreported Amount (NMA), a concept used throughout tax gap studies (IRS (2022)). It is calculated for a given set of line items as the difference between the correct amounts and reported amounts for each return. We calculate seven measures of NMA based on categories of line items at the return level that span different types of taxes, income, and offsets. For income and tax categories, NMA is calculated as *Corrected Income (or Tax)—Reported Income (or Tax)*, and positive NMA values indicate underreporting of income or taxes. For offset categories, NMA is calculated as *Reported Offsets—Corrected Offsets*, and positive NMA values indicate overstatement of offsets.

[7] TCMP, a precursor to IRS's NRP, contained detailed information on compliance (resulting from detailed audits) for a random sample from the population.

[8] 2015 NRP data was released at the time of the writing of this report, and we are adding these data to our sample in ongoing work.

Table 1 summarizes these NMA measures. Our first measure of the NMA is based on the return's total TARC. For each return, we derive the amounts of total tax after all refundable credits.[9] This results in the TARC for each return (i.e., total tax less refundable credits). Both total tax and refundable credits are considered because misreporting can occur in either category. TARC is a measure that applies to all taxpayers in our sample, regardless of types of income or offsets.

For each return, we compute the NMA for six groups of tax return line items based on how visible they are to the IRS. Four of the line-item groups relate to different types of income (Visibility Groups 1-4 of Table 1), while the remaining two groups combine offsets to income (Visibility Group 5) or offsets to tax (Visibility Group 6), as defined in Table 1. We define visibility as the degree to which income or offsets are subject to withholding and/or information reporting. Compliance on income reporting varies with the "visibility" of the income. Income subject to little or no information, such as sole proprietor income, makes up the largest portion of the underreporting tax gap (IRS (2022)).

Visibility Group 1 is the income category subject to the most information reporting and withholding, while Visibility Group 4 is subject to the least. We hypothesize that compliance on certain line items may be more responsive to IRS audit rates than others. For example, rising audit rates may induce taxpayers to accurately report income that is subject to substantial information reporting (since such income would be discoverable by an audit). Whether taxpayers behave similarly for income with *no* information reporting is unclear since such income can be difficult to validate, whether through NRP or operational audits. In our analysis, we evaluate each NMA measure as the dependent variable in separate analyses.

Unlike for the TARC outcome, the Visibility Group outcomes evaluate only taxpayers to whom the relevant income or offsets apply. We remove from each Visibility Group analysis taxpayers who report zero amount *and* have zero true (corrected) amount of those line items. This ensures that a zero NMA value corresponds to compliance behavior and not to irrelevance of line items for the given taxpayer.

TABLE 1. NMA Measures

NMA Measure	Category	Line Items Included	Visibility
TARC	Tax	Total tax and refundable credits	Mixed
Visibility Group 1	Income	Wages & Salaries	High: subject to substantial information reporting and withholding
Visibility Group 2	Income	Pensions and annuities, unemployment compensation, dividend income, interest income, state income tax refunds, and taxable social security	Substantial: subject to substantial information reporting
Visibility Group 3	Income	Partnerships/S corp. income, capital gains, and alimony income	Limited: subject to some information reporting
Visibility Group 4	Income	Nonfarm proprietor income, other income, rents and royalties, farm income, and form 4797 income	Low: subject to little or no information reporting
Visibility Group 5	Offsets to income	Adjustments, deductions, and exemptions	Mixed: subject to varying amounts of information reporting
Visibility Group 6	Offsets to tax	Refundable and nonrefundable credits	Mixed: subject to varying amounts of information reporting

[9] Tax credits that are either fully or partially refundable are the Earned Income Tax Credit, the Child Tax Credit, the Education Credits, and the Health Insurance Premium Tax Credit.

3.2 *Independent Variables*

3.2.1 Audit Rates

The primary regressors of interest are audit rates. We construct the audit rate for a given tax year from IRS enforcement data as the number of unique tax returns that were audited for that year divided by the total number of unique returns filed for that year. We also create separate audit rates for each activity code. Thus, while our dependent variable and other control variables are specified at the return level, audit rates are specified at the aggregate level for the activity code of the taxpayer.

Table 5 of the Appendix summarizes the 12 activity codes that categorize individual returns. Activity codes are delineated by income thresholds, the claiming of certain credits (like EITC), and the reporting of certain income or expenses (like Schedule C for nonfarm sole proprietors and Schedule F for farm sole proprietors). As the third column of Table 5 shows, the majority of the taxpayer population falls in Activity Codes 272 and 273—those with modest annual income[10] (below $200,000) and with no active business income or expenses.

Our baseline specification "assigns" each return the audit rate for its activity code. This decision reflects the likelihood that taxpayers are most responsive to audits of similarly situated taxpayers. However, it is possible that taxpayers are not so discerning and that a higher-level audit rate is more salient. To address this, we aggregate activity codes into four groups: EITC, Non-Business Mid-Income, Business, and Non-Business High-Income. These groupings are used to construct audit rates for sensitivity analysis in Section 5.2.3.

3.2.2 Control Variables

For each NRP return, our control variables are constructed from tax characteristics that may help explain compliance behavior. These include filing status (whether the taxpayer filed as Married Filing Jointly), the total exemptions claimed by the taxpayer, the presence of wage income, the claiming of the child tax credit, whether the taxpayer itemized deductions, whether mortgage interest was deducted, an indicator for taxpayers over 65 years of age, whether the taxpayer used a paid preparer, and an indicator for electronic filing. We base these variables on the taxpayer's reported information on their return.

We also control for the correct amount on the return corresponding to the NMA variable of interest. For example, when TARC NMA is the dependent variable, we include the corrected TARC amount as a regressor. When Visibility Group 6 (credits) NMA is the dependent variable, we include the correct amount of credits as the regressor. This construction allows us to model changes in NMA that arise from compliance behavior and that are not due to changes in the underlying true tax, income, or offsets.

3.3 *Data Summary*

Figure 3 summarizes sample size by activity code. We remove outliers by trimming the bottom and top 5% from the distribution of total reported income in each activity code. We also remove observations with negative NMAs, since we are primarily interested in taxpayers who underreport their total tax (or overstate refundable credits).[11] Except for Activity Code 271, our sample includes at least 4,000 returns for each activity code during TYs 2006–2014.

[10] Income is measured as Total Positive Income (TPI), the sum of all positive amounts of income (and excluding income losses, such as from investments).

[11] Note that descriptive statistics and plots reflect trimming of taxpayers with negative *TARC* NMA, for consistency. In our regressions, each Visibility Group is evaluated in a separate model, and we trim negative values of that Visibility Group's NMA. This approach allows us to focus on taxpayers who underreport income or overstate offsets for each set of line items.

FIGURE 3. Counts of NRP Returns Before and After Trimming (TYs 2006–2014)

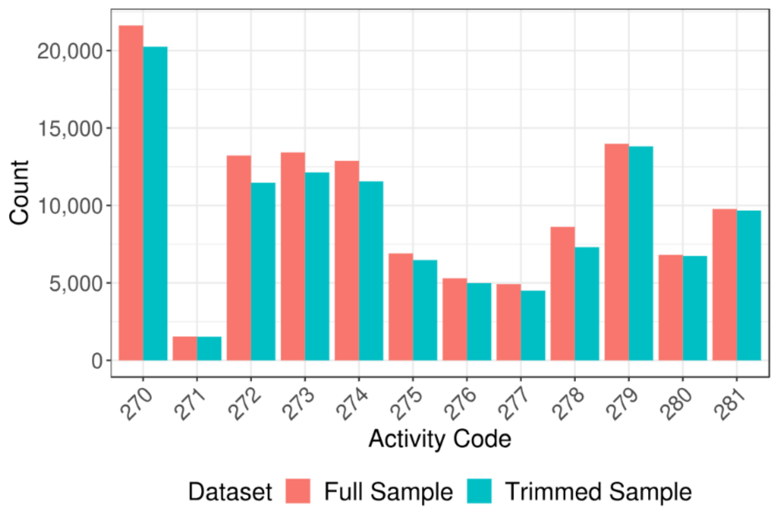


Figure 4 plots audit rates by activity code during TYs 2006–2014. Audit rates have declined across the board, although the decline is most noticeable at the high end of the income distribution. For Activity Code 281, the audit rate fell from 10.8% in 2008 to 2.6% in 2014. Audit rates for other activity codes experienced proportional declines, albeit from a lower starting level of audit coverage rate.

FIGURE 4. Audit Rates by Activity Code

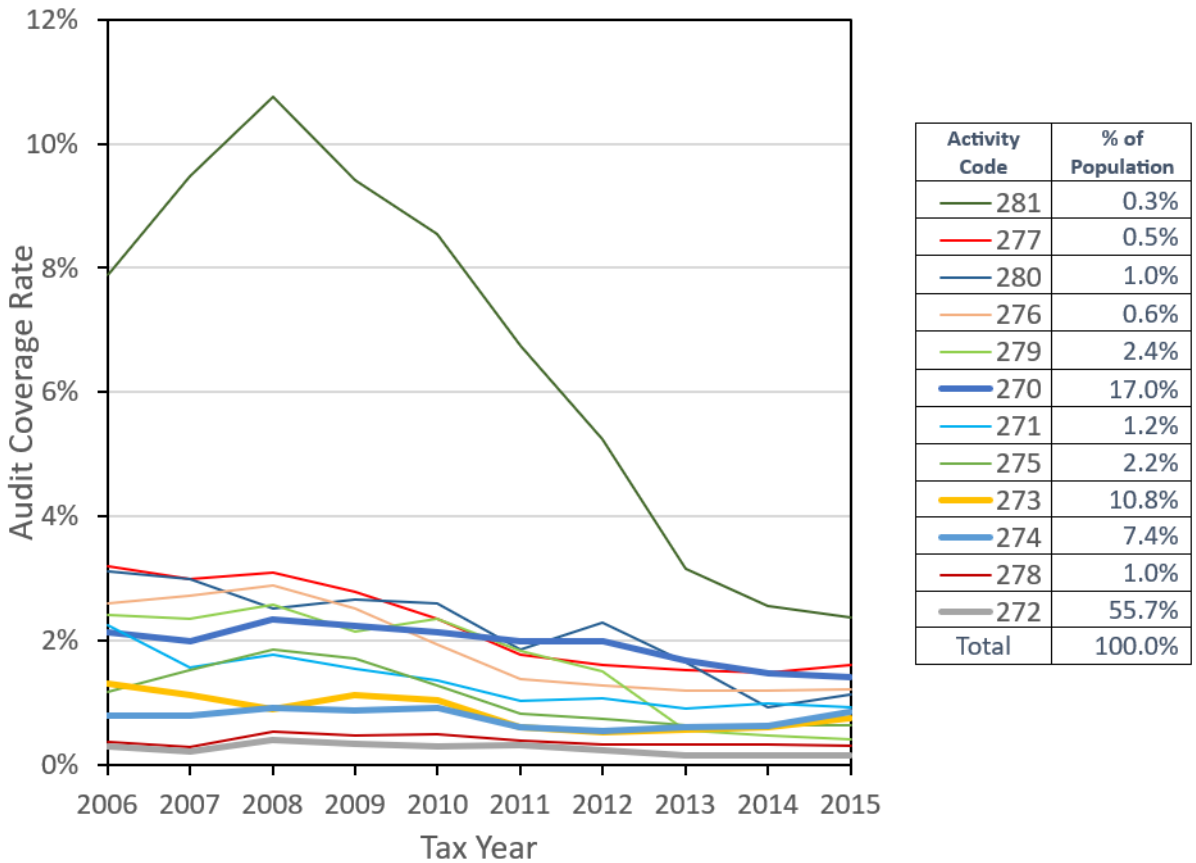


Figure 5 summarizes the aggregate NMA over time by visibility group. The total NMA for the line items included in the visibility group is calculated by weighting each return-level NMA in our NRP sample (using NRP sampling weights) and summing across all returns. Aggregate NMA fell and then increased during this time for Visibility Group 4. The totals for Visibility Groups 3 and 5 fell and plateaued somewhat.

FIGURE 5. Aggregate NMA over Time, by Visibility Group (Weighted)

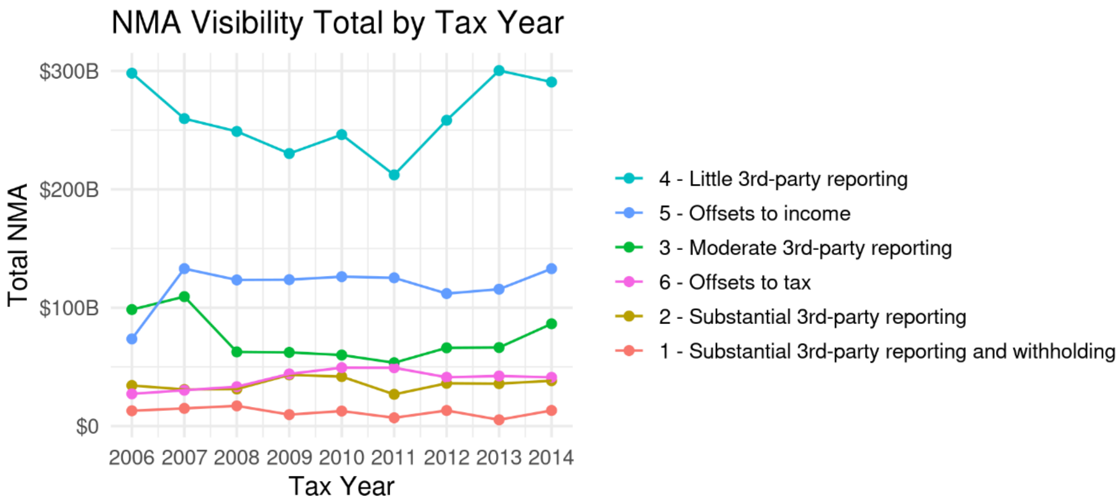
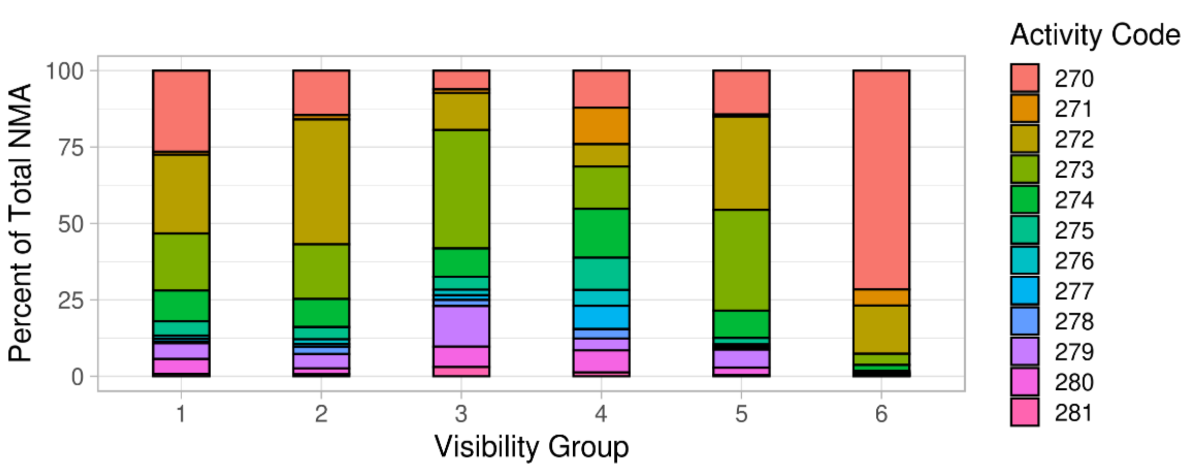


Figure 6 disaggregates NMA totals by activity code. Certain types of taxpayers are more likely to have certain types of income and offsets and are thus more likely to contribute to NMA on those items. For example, Activity Code 270 makes up a large portion of misreporting on credits (Visibility Group 6) but a much smaller portion of misreporting on partnership/S corporation income, capital gains, and alimony income (Visibility Group 3). Activity Codes 279–281, despite comprising only 3.7% of the population (per Table 5), contribute almost 25% of misreporting on Visibility Group 3 income. Activity Code 272, which includes over 55% of the population, contributes the largest portion of misreporting in Visibility Groups 1 and 2, but much less for 3 and 4.

FIGURE 6. Aggregate NMA by Activity Code (Weighted)



As stated above, our econometric specification employs NRP microdata. Table 2 summarizes the dependent and independent variables in our model (excluding audit rates) by tax year. These summary statistics apply to our trimmed data, and observations are weighted by NRP sampling weights. Dollar-denominated variables (NMAs and Correct Amounts) are adjusted to 2018 dollars. For the average return in our sample, TARC NMA drops slightly then increases during this time. This trend also holds for the average NMA for most visibility groups. Correct amounts of TARC and Visibility Groups 1, 3, and 4 also drop slightly, then increase during this time. Commensurate with decreasing marriage rates and our aging population, the proportion of NRP taxpayers filing as Single/other statuses increase somewhat, as does the proportion of taxpayers over 65. Variables declining during this time are the proportion of taxpayers with wage income, claiming a child tax credit, itemizing, and deducting mortgage interest. The use of a paid preparer fell over time, while electronic filing rose dramatically until 2012 then slightly declined.

4. Methods

Our baseline specification models taxpayer *i*'s compliance in tax year *t* as a function of IRS enforcement and other drivers of compliance:[12]

$$\begin{aligned} &\boldsymbol{log(NMA_{it}+1)} \\ &\quad \boldsymbol{= \beta_0 + \beta_1\, Audit\ Rate_{(g,t\text{-}2)} + \beta_2\, Correct\ Amount_{it} + \beta Taxpayer\ Controls_{it}} \\ &\quad \boldsymbol{+\ \alpha Tax\ Year_{t+\delta} Activity\ Code_g + \varepsilon_{it}} \end{aligned} \tag{1}$$

Return-level NMA (i.e., the NMA for TARC) is our main dependent variable, but we also evaluate separately the different NMA measures for the various visibility groups from Table 1. We take a logarithmic transformation of NMA to compensate for skewness. Audit rate is the primary variable of interest. Each taxpayer is assigned the audit rate for their activity code group g in our baseline model (other groupings are evaluated in sensitivity analyses). We lag the audit rate by 2 years to reflect the likely delay between when audits are closed and when other taxpayers become aware of them. Audit start rates (and closure rates) for the tax year at hand are not finalized until all returns for that tax year have been selected for audit (or closed). This process takes several years to resolve internally, with additional time for audit rates to be made public. In sensitivity analyses, we estimate the effect of different lags. We hypothesize that will be negative—a decrease in audit rates should lead to an increase in noncompliance.

We control for the correct amount of the line items in question, depending on the dependent variable (TARC or visibility group). Additional taxpayer control variables refer to the variables described in Section 3.2.2. Finally, we include fixed effects for tax year and activity code. Tax year fixed effects capture yearly fluctuations in compliance that are common across all taxpayers, regardless of activity code (such as due to tax policy changes).[13] Activity code fixed effects capture time-invariant determinants of compliance that are unique to each activity code, unrelated to audit rate changes. Finally, all regressions are weighted by NRP sampling weights.

Our econometric approach is most similar to Tauchen, Witte, and Beron (1993) and Hoopes, Mescall, and Pitman (2012), who evaluate the effect of aggregate audit rates on compliance at the micro level. We differ from their approach by using lagged audit rates instead of contemporaneous ones. While a contemporaneous audit rate reflects audit probability for the return being filed, it is unlikely that the taxpayer knows the contemporaneous audit rate or their audit probability until the audit cycle for that year has completed. Rather, they are more likely to be aware of historical audit rates. To the extent that audit rates change over time (which they have), contemporaneous audit rates are not suitable replacements for historical ones.

Another departure from Tauchen, Witte, and Beron (1993) and Hoopes, Mescall, and Pitman (2012) is in the treatment of the audit rates econometrically. They use an instrumental variable approach, which we opt out of for two reasons. First, lagged audit rates do not suffer from reverse causality, as taxpayers cannot influence

[12] Since NRP samples different taxpayers each year, our data are pooled cross-sections rather than panel/longitudinal.

[13] Our model currently does not control for tax policy changes that are specific to certain taxpayer groups, such as through the inclusion of activity code-tax tear fixed effects. Such fixed effects would be collinear with our audit rate variables, which do not vary within an activity code and tax year. In future work, we hope to include dummy variables capturing known policy changes for certain activity codes.

TABLE 2. Weighted Summary Statistics for NRP Sample by Tax Year

Variable	2006	2007	2008	2009	2010	2011	2012	2013	2014
Dependent Variables									
TARC NMA	$1,680	$1,722	$1,538	$1,584	$1,704	$1,556	$1,547	$1,802	$1,856
Visibility Group 1 NMA	$157	$175	$189	$107	$140	$78	$144	$57	$143
Visibility Group 2 NMA	$416	$363	$347	$478	$462	$300	$395	$387	$416
Visibility Group 3 NMA	$1,194	$1,283	$695	$688	$663	$600	$723	$717	$938
Visibility Group 4 NMA	$3,617	$3,048	$2,762	$2,544	$2,720	$2,379	$2,827	$3,247	$3,156
Visibility Group 5 NMA	$893	$1,561	$1,370	$1,367	$1,395	$1,404	$1,225	$1,250	$1,443
Visibility Group 6 NMA	$330	$355	$369	$487	$545	$552	$451	$457	$447
Independent Variables									
Correct Amount									
TARC	$10,017	$10,196	$8,915	$7,981	$8,590	$8,653	$9,679	$10,109	$10,937
Visibility Group 1	$52,400	$52,745	$50,424	$49,944	$48,948	$47,822	$50,503	$49,492	$50,665
Visibility Group 2	$9,715	$10,367	$9,745	$9,836	$10,191	$9,972	$9,728	$9,697	$10,141
Visibility Group 3	$10,290	$10,028	$6,477	$5,009	$6,135	$6,507	$8,382	$8,008	$9,623
Visibility Group 4	$11,556	$10,538	$9,613	$8,680	$9,770	$9,467	$11,211	$11,308	$11,846
Visibility Group 5	$18,495	$18,005	$17,431	$17,023	$16,528	$15,965	$16,250	$15,990	$15,803
Visibility Group 6	$1,091	$1,059	$1,221	$1,316	$1,213	$1,126	$1,116	$1,100	$1,149
Filing Status									
Single/other	57%	58%	57%	57%	59%	60%	60%	61%	60%
Married filing jointly	43%	42%	43%	43%	41%	40%	40%	39%	40%
Total Exemptions									
0 or NA	2%	1%	2%	2%	2%	2%	2%	2%	2%
1	32%	31%	31%	33%	33%	33%	34%	34%	35%
2	32%	33%	31%	32%	32%	32%	31%	32%	28%
3	17%	17%	17%	15%	16%	15%	15%	15%	16%
4	12%	11%	12%	12%	11%	11%	12%	11%	12%
5+	6%	7%	7%	6%	6%	6%	7%	6%	7%
Had wage income	85%	85%	85%	85%	84%	83%	85%	83%	83%
Claimed child tax credit	24%	23%	23%	21%	22%	19%	19%	19%	19%
Itemized	46%	46%	41%	39%	41%	40%	40%	39%	38%
Deducted mortgage interest	36%	37%	33%	31%	32%	30%	30%	29%	27%
Over 65	12%	13%	14%	14%	13%	13%	14%	15%	15%
Used paid preparer	66%	66%	65%	62%	63%	62%	62%	62%	59%
Filed electronically	50%	65%	71%	73%	80%	84%	84%	70%	70%

Note: These summary statistics apply to our trimmed NRP sample. Statistics are weighted by NRP sampling weights. Means are displayed for NMAs and Corrected TARC, while proportions are displayed for all other variables. Dollar-denominated variables are expressed in terms of 2018 dollars.

past audit rates through current reporting behavior. Second, audit rates have declined across the board at varying rates due to declining resources and shifts in allocation (but not in response to improved compliance), thereby creating a natural experiment for evaluating the causal effect of audit rates.

5. Results

5.1 Descriptive Analysis

For many groups of taxpayers, noncompliance has increased while audit rates have declined. Figure 14 of the Appendix plots the audit rate against the aggregate NMP on TARC by activity code. We calculate the aggregate NMP for each activity code by summing TARC NMA across those taxpayers, summing the absolute value of Corrected TARC across those taxpayers, and dividing the former by the latter. As such, NMP captures the extent of tax noncompliance relative to the audit-determined amount of total TARC. The plots represent NMP across our trimmed dataset (i.e., outliers and negative TARC NMAs removed).

As Figure 14 shows, noncompliance clearly trended upwards during 2006–2014 for certain groups of taxpayers: taxpayers with annual income below $200,000 who either 1) claimed EITC (Activity Codes 270 and 271) or 2) operated a business or sole proprietorship (Activity Codes 274–278). Interestingly, long-term compliance trends are less clear for taxpayers earning $200,000 and above. These plots provide suggestive evidence that not all taxpayers may respond to declining audit rates. Those taxpayers with modest total income who report certain types of income (Schedule C or F) or claim certain credits (EITC) may respond the most. High-income taxpayers appear to be less affected by audit rates but potentially for a different reason: they have more income at stake (and more resources to weather an audit) and thus could be unmoved by a perceived change in audit probability.

Alternatively, there may be a unique type of measurement error in the NRP estimate of noncompliance for high income taxpayers. For example, a large portion of income from pass-through entities and offshore accounts eludes IRS detection (Guyton *et al.* (2021)). It is likely that NRP estimates of noncompliance, such as an NMP of 3–12% for taxpayers with at least $1 million income (per Figure 14), grossly understate noncompliance at the high end. In this case, it would be unknown to the IRS whether truthful reporting of this income has improved over time (or not), limiting our ability to test for the comprehensive indirect effect within this group.

5.2 Modeling Results

Our main analysis estimates the comprehensive indirect effect of audit rates on various NMA measures and taxpayer subsamples. We also conduct a sensitivity analysis in the specification of the audit rate variable.

5.2.1 Baseline Results

Table 3 presents baseline results for our full sample, in which the audit rate variable is specified as a two-year lag of the audit rate for each taxpayer's activity code. Column 1 presents the effect on TARC NMA, while Columns 2–7 present the effects on NMA by visibility group. Since we rely heavily on audit rate variation over time, tax year fixed effects could subsume some of the effect of the audit rate. Table 6 in the Appendix presents the same analysis without tax year fixed effects. For the purposes of this discussion, we consider statistical significance to be at the 5% level.

As Table 3 shows, the effect of audit rate on compliance varies depending on the line items being evaluated. Audit rate has an unexpected positive effect on TARC NMA: a one percentage point increase in the audit rate increases NMA by 4.9% (9.9% when omitting tax year fixed effects). This unexpected result is likely driven by Activity Code 272, which represents an outsized portion of the population (55.3% as shown in Table 5) and for which our subsequent subsample analysis produces an unexpected positive effect. The unexpected result on TARC NMA suggests we should evaluate activity code subpopulations separately (which we do in Section 5.2.2).

Audit rates have the expected negative effect on noncompliance for all Visibility Groups except for Group 4. For line items with high visibility (Visibility Group 1), a one percentage point increase in audit rates decreases NMA on wages and salaries by 2.2% (5.1% when tax year fixed effects are removed).[14] The latter estimate is statistically significant. Given that the majority of taxpayers report wage/salary income (see Table 2), this full sample result is intuitive. For Visibility Group 2, a one percentage point increase in audit rates decreases NMA by 3.0−5.2% (again statistically significant only when tax year fixed effects are omitted). This group includes taxpayers in a variety of situations, such as retirees with pensions/annuities income and taxpayers between jobs receiving unemployment income. For Visibility Group 3, a one percentage point increase in audit rates decreases NMA by 2.1−11.9% (significant without tax year fixed effects). This group includes partnership/S corporation income, capital gains, and alimony income—sources of income with some limited information reporting.

For Visibility Group 4, audit rates do not have a discernable effect on noncompliance (in either specification regarding tax year fixed effects). Income in this group is subject to very little information reporting (if any), so taxpayers may not respond to audit rates in the expected way. Further, as discussed in Section 5.1, NMA estimates for low visibility line items are likely to be understated.

Finally, audit rates have the expected effect on adjustments, deductions, exemptions, and credits (Visibility Groups 5 and 6). A one percentage point increase in audit rates decreases NMA on adjustments, deductions, and exemptions by 10.4% in the baseline specification and decreases NMA on refundable and nonrefundable credits by 14.9−21.8%.

[14] The result for Visibility Group 1 is consistent with a separate analysis we conducted using Automated Underreporter (AUR) data (results are not included here for conciseness) among a sample of tax returns taken from the entire population. AUR matches third-party information documents sent to the IRS with what taxpayers report on their tax returns. This screens for noncompliance on line items with substantial information reporting, such as wages and salaries. We construct a measure of NMA based on AUR-corrected line items. While NRP-adjusted NMA is available only for NRP audits, AUR-adjusted NMA is available for all taxpayers using third-party information documents. This approach allows us to evaluate a sample of taxpayers outside the standard NRP population for this analysis.

TABLE 3. Full Sample Baseline Regression Results

	Dependent variable: Log NMA						
	TARC	Visibility Group 1	Visibility Group 2	Visibility Group 3	Visibility Group 4	Visibility Group 5	Visibility Group 6
Audit Rate (2-Yr Lag)	0.049*	-0.022	-0.030	-0.021	0.003	-0.104***	-0.218***
	(0.028)	(0.019)	(0.033)	(0.059)	(0.033)	(0.037)	(0.028)
Corrected TARC	0.00001***						
	(0.00000)						
Correct Amount for Visibility Group		-0.00000	0.00000***	0.00000***	0.00000***	-0.00000***	-0.00004***
		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Total Exemptions 1	1.649***	-0.335***	0.262**	0.250*	1.712***	0.657***	0.619
	(0.143)	(0.057)	(0.114)	(0.147)	(0.136)	(0.089)	(0.401)
Total Exemptions 2	3.134***	-0.574***	0.478***	0.843***	1.603***	3.451***	1.611***
	(0.145)	(0.059)	(0.118)	(0.163)	(0.144)	(0.092)	(0.401)
Total Exemptions 3	3.609***	-0.553***	0.579***	1.008***	1.728***	3.817***	2.302***
	(0.147)	(0.061)	(0.121)	(0.173)	(0.149)	(0.096)	(0.402)
Total Exemptions 4	3.665***	-0.557***	0.493***	0.736***	1.692***	3.860***	2.532***
	(0.150)	(0.063)	(0.124)	(0.179)	(0.155)	(0.101)	(0.403)
Total Exemptions 5+	4.005***	-0.472***	0.540***	1.195***	1.884***	3.996***	2.973***
	(0.153)	(0.065)	(0.128)	(0.187)	(0.161)	(0.106)	(0.404)
Wage Income	-0.087**		-0.058*	-0.169***	-0.135***	0.116***	-0.103**
	(0.038)		(0.032)	(0.050)	(0.041)	(0.042)	(0.049)
Claimed child tax credit	-0.226***	-0.058***	-0.195***	-0.042	-0.224***	-0.574***	-0.205***
	(0.031)	(0.018)	(0.034)	(0.068)	(0.052)	(0.034)	(0.031)
Itemized	1.385***	-0.301***	0.056	-0.372***	-0.122**	4.823***	0.196***
	(0.042)	(0.027)	(0.035)	(0.055)	(0.058)	(0.043)	(0.054)
Deducted mortgage interest	-0.474***	-0.048*	0.012	0.386***	0.401***	-0.721***	-0.402***
	(0.041)	(0.026)	(0.034)	(0.056)	(0.058)	(0.044)	(0.053)
Over 65	-0.609***	-0.270***	0.879***	-0.171***	-0.463***	-0.301***	-0.507***
	(0.040)	(0.028)	(0.032)	(0.052)	(0.049)	(0.043)	(0.055)
Used paid preparer	0.038*	-0.045***	-0.153***	-0.109**	0.101***	-0.113***	0.012
	(0.023)	(0.013)	(0.022)	(0.043)	(0.037)	(0.024)	(0.026)
Filed electronically	-0.167***	0.049***	-0.088***	-0.069*	-0.292***	-0.084***	-0.019
	(0.025)	(0.015)	(0.023)	(0.040)	(0.035)	(0.027)	(0.029)
Married-Joint Status	-1.689***	0.083***	0.036	-0.419***	0.179***	-2.768***	-1.055***
	(0.033)	(0.019)	(0.036)	(0.080)	(0.055)	(0.036)	(0.035)
Constant	1.226***	1.263***	0.779***	2.854***	3.179***	0.179	1.506***
	(0.157)	(0.067)	(0.135)	(0.206)	(0.160)	(0.117)	(0.405)
Observations	88,521	72,938	63,795	40,112	58,919	73,990	48,083
Tax Year fixed effect	Y	Y	Y	Y	Y	Y	Y
Adjusted R2	0.133	0.023	0.034	0.085	0.235	0.342	0.154
F Statistic	400.675***	52.387***	66.589***	110.544***	533.957***	1,129.561***	258.305***
Degrees of Freedom	88,437	72,846	63,742	40,064	58,856	73,871	48,023

Notes: Standard errors displayed in parentheses. *p<0.10, **p<0.05, ***p<0.01 Corrected amounts (for TARC and by visibility group) are specified in unscaled dollar values. Although statistically significant, the estimated coefficients are small because the amount of noncompliance is small relative to the overall amount of income or offsets.

5.2.2 Subsample Analysis

There is reason to believe that taxpayers respond differently to audit rates depending on their tax situation (such as the amount and types of income they earn). Certain activity codes make up a disproportionate amount of income/expenses in each visibility group (see Figure 6). To the extent some line items are more responsive to audit rates, we would expect to see indirect effects vary for taxpayers of different activity codes as well. Prior literature also has identified disparate indirect effects based on the taxpayer's amount and visibility of income (e.g., Tauchen, Witte, and Beron (1993); Slemrod, Blumenthal, and Christian (2001)).

We estimate Equation (1) separately for each activity code segment of the population. Audit rate is specified as the audit rate for each activity code. Figure 7 displays the effects on TARC NMA by activity code. Audit rate has an unexpected positive effect on Activity Code 272 (taxpayers with income below $200,000 and no business income), which likely drives the full sample result in Column 1 of Table 3. Audit rates also have an unexpected positive effect on Activity Code 270 (EITC-claiming taxpayers with no business or business income below $25,000). Other subsample results are mixed and largely statistically insignificant. In sensitivity analysis, we find the expected negative effect of audit rate using a different lag and a more aggregate audit rate. This finding suggests that a two-year lag of audit rate may not be salient to all taxpayers.

FIGURE 7. Effect of Activity Code Audit Rate on TARC NMA

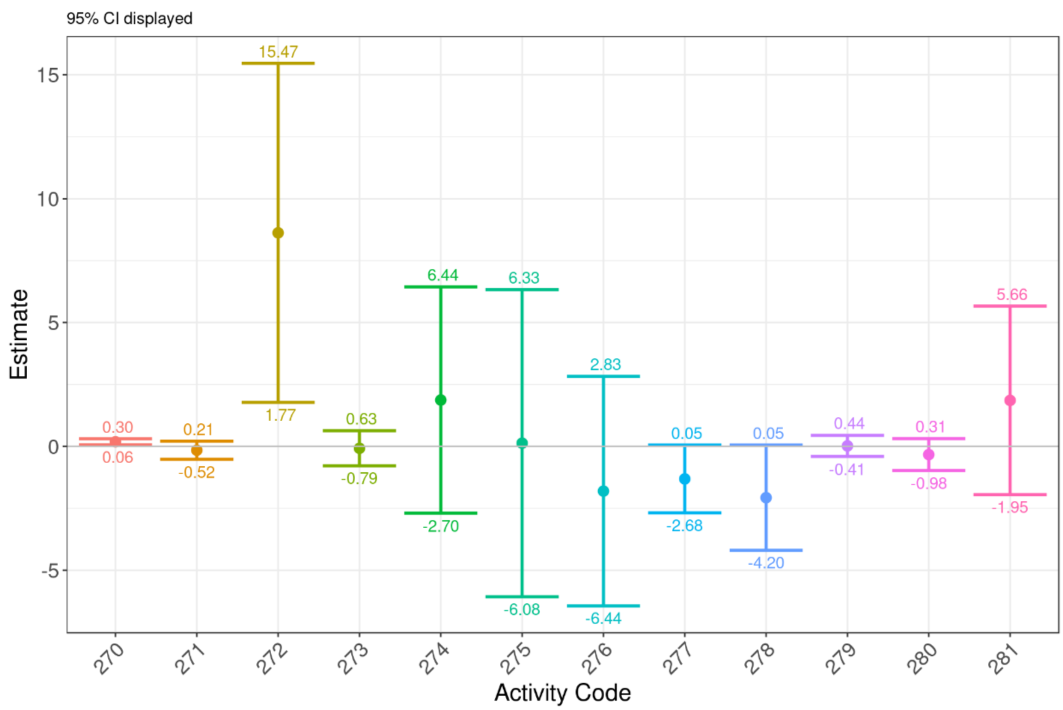


Figure 8 displays effects on Visibility Group 1 NMA by activity code. Like the full sample results in Table 3, subsample results are mostly statistically insignificant across the board. There is an unexpected positive effect of audit rate on Activity Code 281. Subsample results for Visibility Group 2 are largely statistically insignificant (Figure 9). The exceptions are an unexpected positive effect for Activity Code 272 and the expected negative effect for Activity Code 280.

FIGURE 8. Effect of AC Audit Rate on Visibility Group 1 NMA

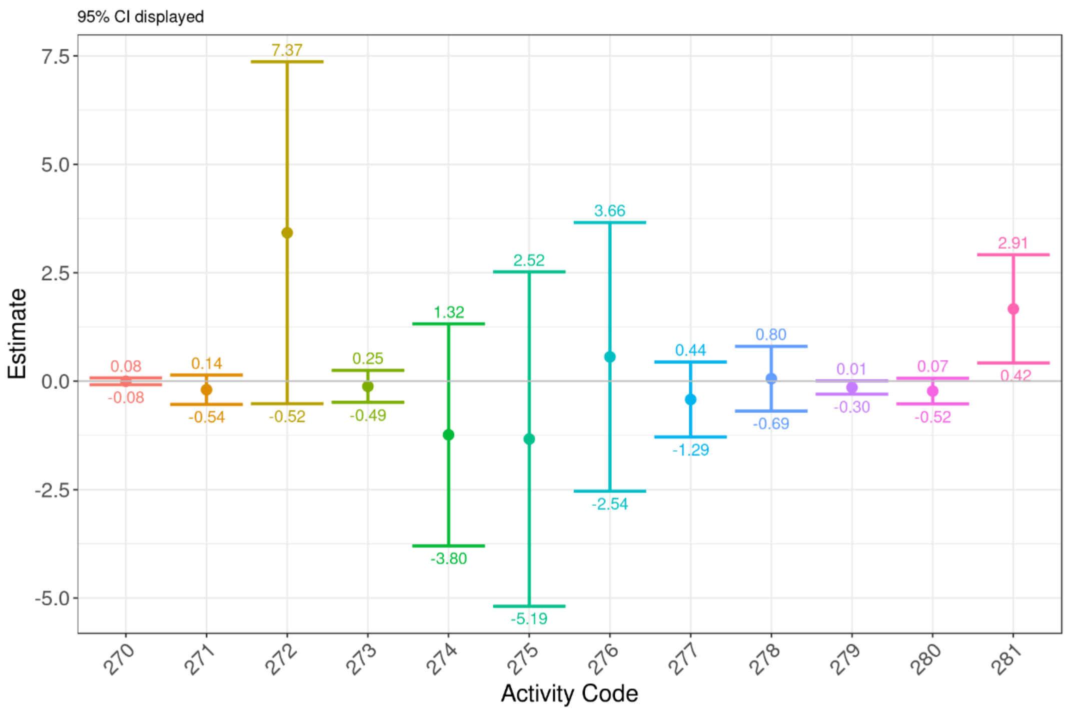


FIGURE 9. Effect of AC Audit Rate on Visibility Group 2 NMA

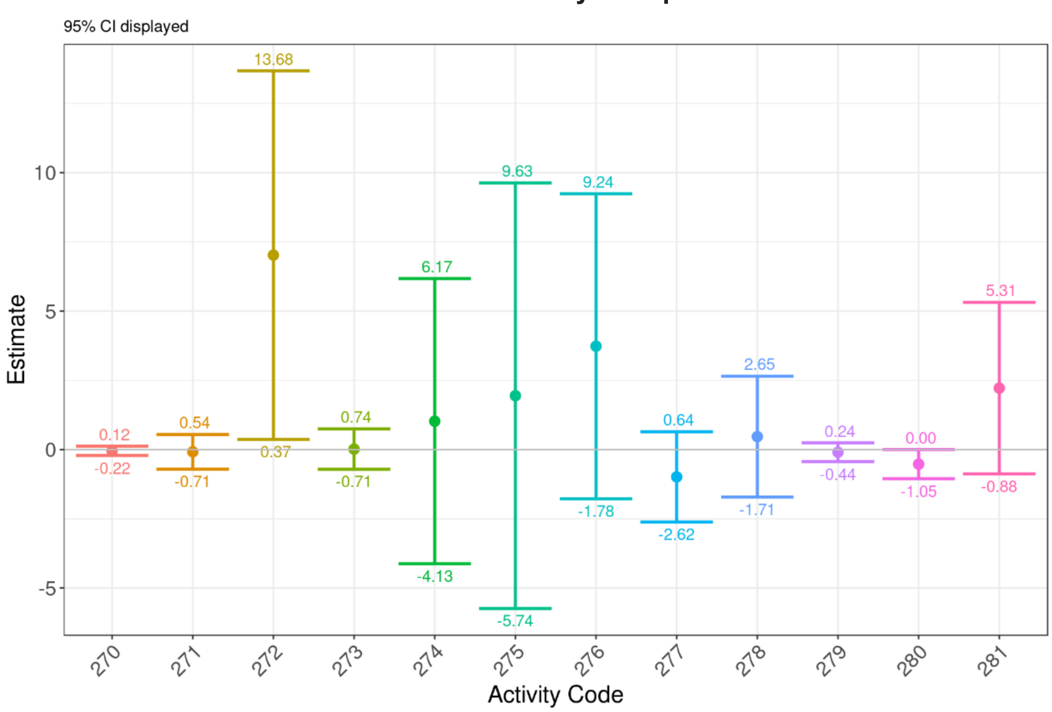


Figure 10 displays effects on Visibility Group 3 NMA by activity code. Audit rates have the expected negative effect for Activity Code 280. This activity code covers taxpayers earning between $200,000 and $1 million income without business income, who make up a disproportionate share of the NMA in this category (relative to their portion of the population).

FIGURE 10. Effect of AC Audit Rate on Visibility Group 3 NMA

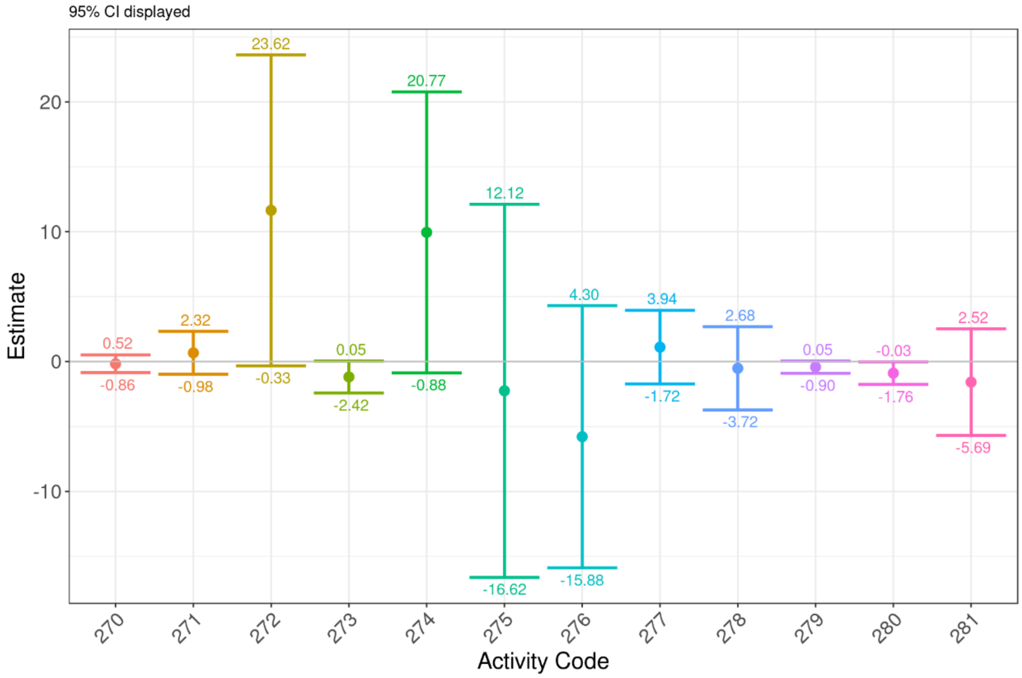


Figure 11 displays effects on Visibility Group 4 NMA by activity code. Audit rates have the expected negative effect (statistically significant) for Activity Codes 277 and 279 and an unexpected positive effect on Activity Code 270. Lastly, Figure 12 and Figure 13 display effects on Visibility Group 5 and 6 (respectively). When it comes to Visibility Group 5 NMA, audit rates have an unexpected positive effect on 272 and 274, although these effects do not hold in sensitivity analyses. For Visibility Group 6 NMA, audits have the expected negative effect on Activity Codes 270 and 276 and an unexpected positive effect on Activity Codes 275 and 280.

FIGURE 11. Effect of AC Audit Rate on Visibility Group 4 NMA

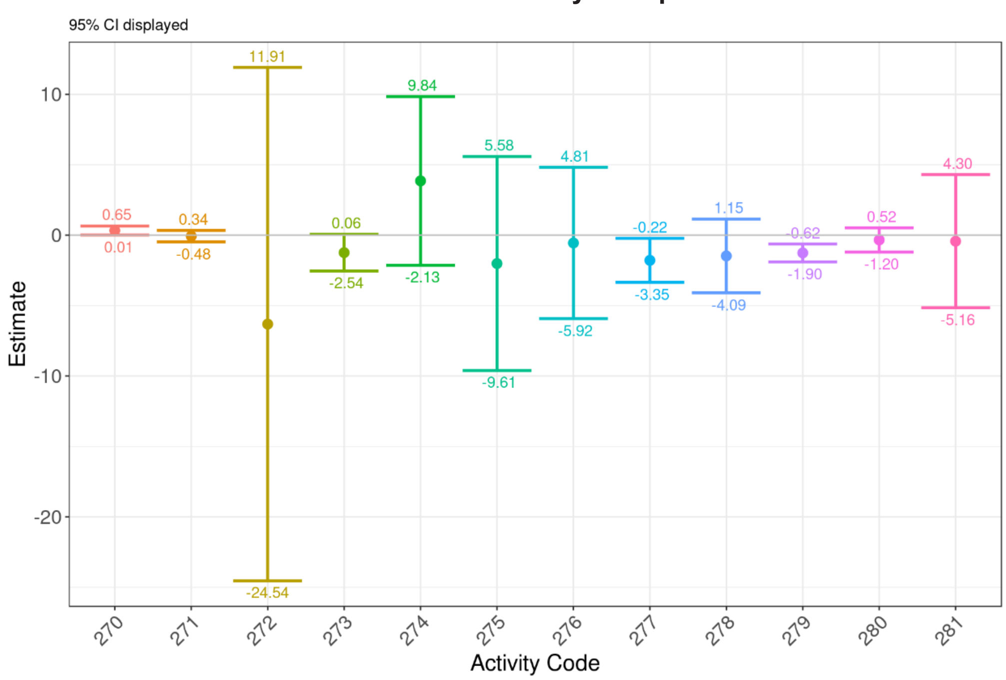


FIGURE 12. Effect of AC Audit Rate on Visibility Group 5 NMA

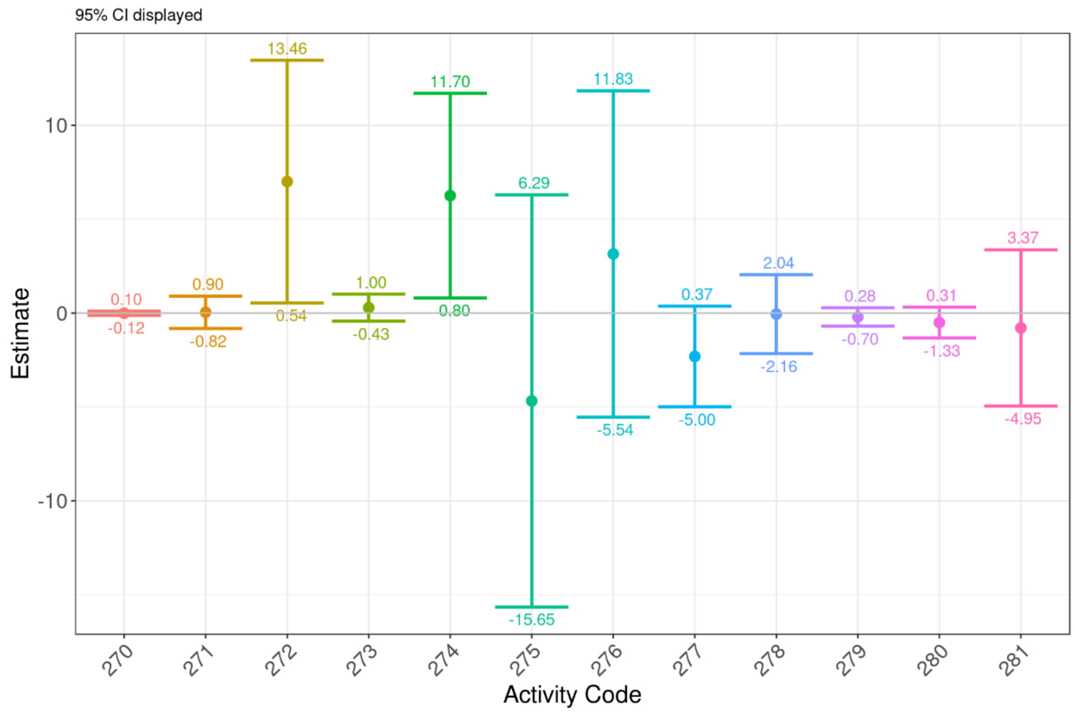
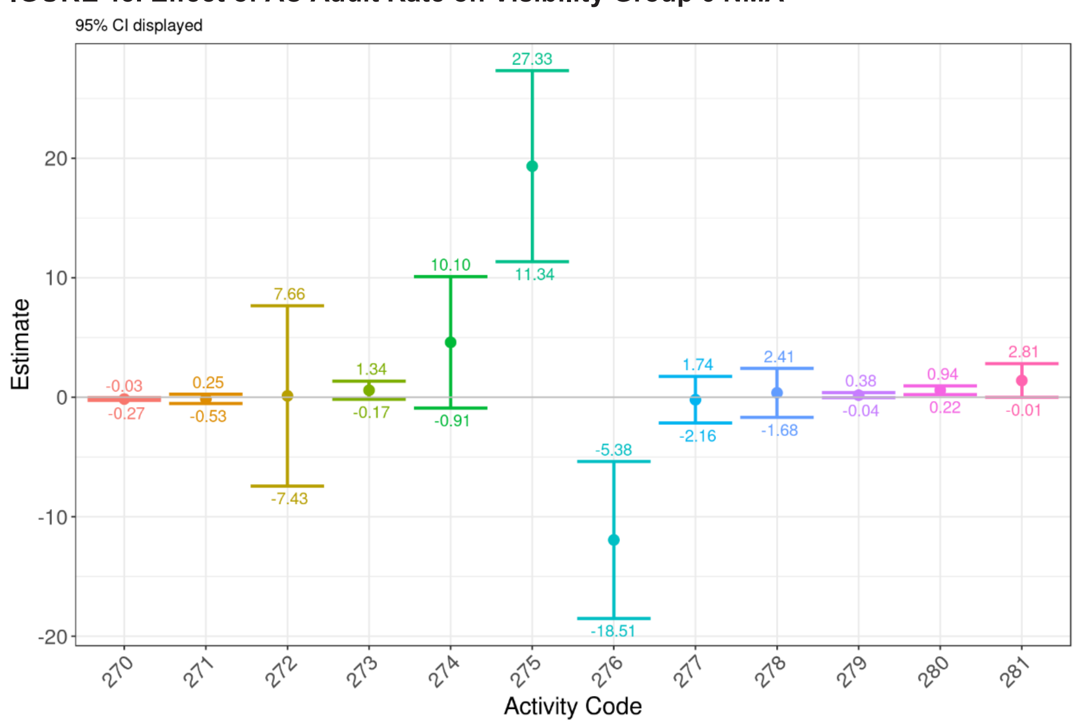


FIGURE 13. Effect of AC Audit Rate on Visibility Group 6 NMA



5.2.3 Sensitivity Analysis

While most general indirect effects papers use a treatment/control group study design, our approach in the vein of the comprehensive indirect effects literature relies on varying levels of treatment across the entire population. As such, results may be particularly sensitive to the specification of the treatment variable (audit rates). We add to the literature by using lagged audit rates in our baseline specification instead of contemporaneous ones and by testing for heterogenous effects across a variety of line items and taxpayer subpopulations. In this section, we evaluate other specifications of the audit rate that may be salient in understanding taxpayer response to IRS enforcement.

In addition to our baseline specification of a two-year lag of activity code audit rate, we evaluate alternative specifications: different lags of audit rate, audit rates grouped across some activity codes, the change in audit rate (rather than the level), and the average of audit rate over time.

Different lags of audit rate: We replace audit rate in Equation (1) with one- through five-year lags of the activity code audit rate. All five lags are included in each model concurrently. In the full sample models, audit rates under this specification are mostly statistically insignificant (across all seven NMA outcomes). In activity code subsamples, various lags have the expected negative effect on NMA (statistically significant for certain activity codes). Activity codes with statistically significant effects in the baseline specification of two-year lags (per Section 5.2.2) typically show similar results for a few other lags as well. In some cases, using other lags produces the expected negative effect when the baseline specification does not. This suggests that different lags of the audit rate are salient for different taxpayers.

Grouped audit rate: We replace audit rate in Equation (1) with audit rate defined at the level of the four activity code groupings in Table 5. Each return is assigned the two-year lag of the audit rate for the activity code group to which it belongs. Results under this specification tend to mirror the subsample results in Section 5.2.2., in some cases picking up a negative effect when the baseline specification does not. For example, the own-group audit rate has the expected negative effect on TARC NMA for certain activity codes (while the baseline regressions for TARC NMA did not yield a discernable effect).

Change in audit rate: We replace audit rate in Equation (1) with the change in the activity code audit rate over time. We specify one- through five-year changes, each in a separate regression. This specification produces the expected negative effect of audit rate on TARC NMA (statistically significant), while the baseline specification did not. This specification produces an unexpected positive effect on Visibility Groups 1 and 5 NMA, however. Results for other NMA measures are statistically insignificant.

Average audit rate: We replace audit rate in Equation (1) with the average activity code audit rate over time. We specify one- through five-year averages, each in a separate regression. This approach produces the expected negative effect of audit rate on Visibility Group 1 NMA and an unexpected positive effect on TARC and Visibility Group 4 NMA. Results for other NMA measures are statistically insignificant. The drawback of using this specification is that it dulls the variation in audit rates over time, thereby weakening our identification strategy.

6. Discussion

While most research on the impact that IRS enforcement efforts have on the compliance behavior of taxpayers in the general population evaluates specific taxpayer contexts and networks (or mechanisms), this paper contributes to a small literature on the comprehensive indirect effects of IRS enforcement on voluntary compliance. The estimation of comprehensive indirect effects aims to capture 1) the effects of all IRS enforcement activities, 2) the effects across the general taxpayer population, and 3) the effects propagating through various and multiple types of mechanisms. As such, these effects are relevant for IRS budget justification, which currently cites the ROI of enforcement on direct revenue and merely alludes to the existence of deterrence effects (IRS (2023)).

We advance understanding of the nature and magnitude of comprehensive indirect effects by implementing several novel or rarely used approaches. Ours is one of the few papers to use microdata in this area. This allows for more nuanced modeling of taxpayer behavior and the ability to control for return-level characteristics. Departing from prior papers, we use lagged audit rates to proxy for knowledge of IRS enforcement levels. While audit rates for the tax year at hand reflect the true aggregate probability of an audit, taxpayers (and their accountants) can plausibly know only past audit rates. Additionally, using lagged audit rates solves the reverse causality (endogeneity) problem; an earlier audit rate is not impacted by this year's compliance.

We do not find the expected effect of audit rates on bottom-line noncompliance (i.e., on TARC). However, we find that audit rates have the expected deterrence effect on certain groups of line items. Further, the effect is larger for items subject to less information reporting. Misreporting on high visibility income (wage and salaries) drops by 5.1% with a one percentage point increase in audit rates. The same change in audit rates induces a 5.2% drop in misreporting on income subject to substantial information reporting but not withholding (such as unemployment compensation), and a 11.9% drop in misreporting on income subject to only limited information reporting (such as partnership income and capital gains). The effect on misreporting income offsets is even larger—a one percentage point increase in audit rates decreases misreporting of adjustments, deductions, and exemptions by 10.4% and misreporting of refundable and nonrefundable credits by up to 21.8%.

These results are intuitive. High visibility line items such as wages and salaries are screened by automated underreporter programs, and misreporting on these line items may be less sensitive to audit rates per se. On the other hand, misreporting on income *not* validated automatically should be more responsive to enforcement actions such as audit rates.

Notably, we did not find a discernable effect of audit rates on misreporting of income subject to little or no information reporting. This result suggests that the deterrence effects of IRS enforcement depend on how well noncompliance can be detected and validated. Taxpayers improve reporting of visible line items in the face of rising audits rates because when 1) true tax obligations are visible to the IRS (taxpayer-reported income can be measured against income reported by third parties) and 2) the increased probability of audit increases the likelihood of the IRS validating these measures, there is an incentive to change behavior and increase compliance. Conversely, null or unexpected results can arise when the true amounts that should be reported on certain line items are difficult for the IRS to detect. In this case, taxpayers do not have deterrent incentives to increase compliance even in the face of an audit, so their behavioral response is unclear. Moreover, for these line items, researchers cannot rely on enforcement data—even that from the NRP—to be correct (due to measurement error).

6.1 Limitations and Future Research

As discussed, a primary limitation of this research is that NRP audits may not detect all noncompliance among taxpayers with high and unreported income. Prior research has attempted to shed light on previously undetected offshore accounts and passthrough income (Guyton *et al.* (2021)), but has not explored its relation to changes in compliance over time.

There are several near-term extensions we plan to address. Some visibility groups include a mixture of line items that reflect very different tax situations. For example, Visibility Group 2 includes retirees earning pension and Social Security income, as well as taxpayers between jobs receiving unemployment income. Further subdividing NMA within visibility group may yield better understanding of taxpayer behavior based on their circumstance. We also plan to evaluate different transformations of the NMA outcome variables, since negative NMAs are dropped in a log transformation.[15] We also plan to convert our estimates of the comprehensive indirect effect into dollar values (to directly align with ROI figures). However, our results herein provide strong evidence that the answer to our opening question (How much additional revenue would likely be generated if the IRS enforcement budget were increased by $X per year?) depends on how the additional funding is allocated across taxpayer groups and noncompliance opportunities or issues.

Finally, the ultimate goal of this research is to support IRS budget justification by estimating the ROI of all IRS activities. IRS service, outreach, education, and IT investments plausibly have an impact on compliance, as well. Most taxpayers desire and strive to properly report their income. These IRS services help taxpayers become more informed and better equipped to report and pay their taxes correctly the first time. To account for this, we hope to incorporate into future iterations of this work measures such as IRS website hits and level of service. Lastly, although we focus on individual taxpayers in this paper, prior research indicates that corporations track IRS enforcement activities in their accounting practices (Hoopes, Mescall, and Pitman (2012)). Estimating the deterrence effect of enforcement on corporate voluntary compliance is another area of future work.

[15] We tried an Inverse Hyperbolic Sine (IHS) transformation in lieu of the log transform, but results were highly sensitive to variable scaling. Alternatively, we plan to estimate regressions with an untransformed dependent variable (i.e., level of NMA).

References

Alstadsæter, Annette, Wojciech Kopczuk and Kjetil Telle. (2019). "Social Networks and Tax Avoidance: Evidence from a Well-Defined Norwegian Tax Shelter." *International tax and Public Finance* 26: 1291–1328. https://doi.org/10.1007/s10797-019-09568-3.

Boning, William C., John Guyton, Ronald Hodge, and Joel Slemrod. (2020). "Heard It Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment." *Journal of Public Economics* 190: 104261. https://doi.org/10.1016/j.jpubeco.2020.104261.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. (2013). "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103(7): 2683-2721. https://www.jstor.org/stable/pdf/42920668.pdf.

Department of the Treasury. (2019). "FY 2019 Budget in Brief: Internal Revenue Service, Program Summary by Budget Activity." https://home.treasury.gov/system/files/266/16.-IRS-FY-2019-BIB-FY2019.pdf.

Dubin, Jeffrey A., and Louis L. Wilde. (1988). "An Empirical Analysis of Federal Income Tax Auditing and Compliance." *National Tax Journal* 41(1): 61–74. https://www.jstor.org/stable/41788709.

Dubin, Jeffrey A., Michael J. Graetz, and Louis L. Wilde. (1990). "The Effect of Audit Rates on the Federal Individual Income Tax, 1977–1986." *National Tax Journal* 43(4): 395–409. https://www.jstor.org/stable/41788861.

Dubin, Jeffrey A. (2007). "Criminal Investigation Enforcement Activities and Taxpayer Noncompliance." *Public Finance Review* 35(4): 500–529. https://authors.library.caltech.edu/99369/1/sswp1200.pdf.

Hoopes, Jeffrey L., Devan Mescall, and Jeffrey A. Pittman. (2012). "Do IRS Audits Deter Corporate tax avoidance?" *The Accounting Review* 87(5): 1603–1639. https://www.jstor.org/stable/pdf/41721904.pdf.

Internal Revenue Service. (2022). "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2014-2016." *Publication 1415* (Rev. 08–2022). https://www.irs.gov/pub/irs-pdf/p1415.pdf.

Internal Revenue Service. (2023). "Fiscal Year 2024 Congressional Budget Justification & Annual Performance Report and Plan." *Publication 4450* (Rev. 2–2023). https://www.irs.gov/pub/irs-pdf/p4450.pdf.

Grana, Jess, Nelson Foster, Navya Kambalapally, India Lindsay, Betsy Lydon, Max McGill, Alexander McGlothlin, and Brian Rhindress. (2022). "Indirect Effects Research and Support." *MITRE Technical Report.*

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. (2021). "Tax Evasion at the Top of the Income Distribution: Theory and Evidence." *National Bureau of Economic Research*, No. w28542. https://www.nber.org/system/files/working_papers/w28542/w28542.pdf.

Lykke, Lucia, Max McGill, Leigh Nicholl, and Alan Plumley. (2020). "The Specific Indirect Effect of Correspondence Audits: Moving from Research to Operational Application." *The IRS Research Bulletin*. https://www.irs.gov/pub/irs-prior/p1500--2021.pdf.

Meiselman, Ben S. (2018). "Ghostbusting in Detroit: Evidence on Nonfilers from a Controlled Field Experiment." *Journal of Public Economics* 158: 180–193. https://doi.org/10.1016/j.jpubeco.2018.01.005.

Perez-Truglia, Ricardo, and Ugo Troiano. (2018). "Shaming Tax Delinquents." *Journal of Public Economics* 167: 120–137. https://doi.org/10.1016/j.jpubeco.2018.09.008.

Plumley, Alan. (1996). "The Determinants of Individual Income Tax Compliance." Department of the Treasury, Internal Revenue Service, *Publication 1916* (Rev. 11–96).

Pomeranz, Dina. (2015). "No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105(8): 2539–2569. https://www.jstor.org/stable/pdf/43821348.pdf.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. (2001). "Taxpayer Response to an Increased Probability of Audit: Evidence From a Controlled Experiment in Minnesota." *Journal of Public Economics* 79(3): 455–483. https://doi.org/10.1016/S0047-2727(99)00107-3.

Tauchen, Helen V., Ann Dryden Witte, and Kurt J. Beron. (1993). "Tax Compliance: An Investigation Using Individual Taxpayer Compliance Measurement Program (TCMP) Data." *Journal of Quantitative Criminology* 9: 177–202. https://www.jstor.org/stable/23365801.

Appendix

Data Summary

TABLE 5. Activity Code Definitions

Activity Code	Description	Percent of Population	Group
270	EITC present & TPI < $200,000 and Schedule C/F TGR < $25,000 or EITC w/o Sch C/F (As of TY 2008)	17.1%	EITC
271	EITC present & TPI < $200,000 and Sch C/F TGR > $24,999 (As of TY 2008)	1.2%	EITC
272	TPI < $200,000, no Sch C, E, F, or Form 2106 (As of TY 2008)	55.3%	Non-Business Mid-Income
273	TPI < $200,000 and Sch E or Form 2106, no Sch C or F (As of TY 2008)	10.8%	Non-Business Mid-Income
274	Non-Farm Business w/ Sch C/F TGR < $25,000 and TPI < $200,000 (As of TY 2008)	7.3%	Business
275	Non-Farm Business w/ Sch C/F TGR $25,000 - $99,999 and TPI < $200,000 (As of TY 2008)	2.1%	Business
276	Non-Farm Business w/ Sch C/F TGR $100,000 - $199,999 and TPI < $200,000 (As of TY 2008)	0.6%	Business
277	Non-Farm Business w/ Sch C/F TGR > $199,999 and TPI < $200,000 (As of TY 2008)	0.5%	Business
278	Farm Business Not Classified Elsewhere and TPI < $200,000 (As of TY 2008)	0.9%	Business
279	No Sch C or F and TPI > $199,999 and < $1,000,000 (As of TY 2008)	2.4%	Non-Business High-Income
280	Sch C or F present and TPI > $199,999 and < $1,000,000 (As of TY 2008)	1.0%	Business
281	TPI > $999,999 (As of TY 2008)	0.3%	Non-Business High-Income

FIGURE 14. Audit Rate vs. Aggregate NMP, by Activity Code (Trimmed Data)

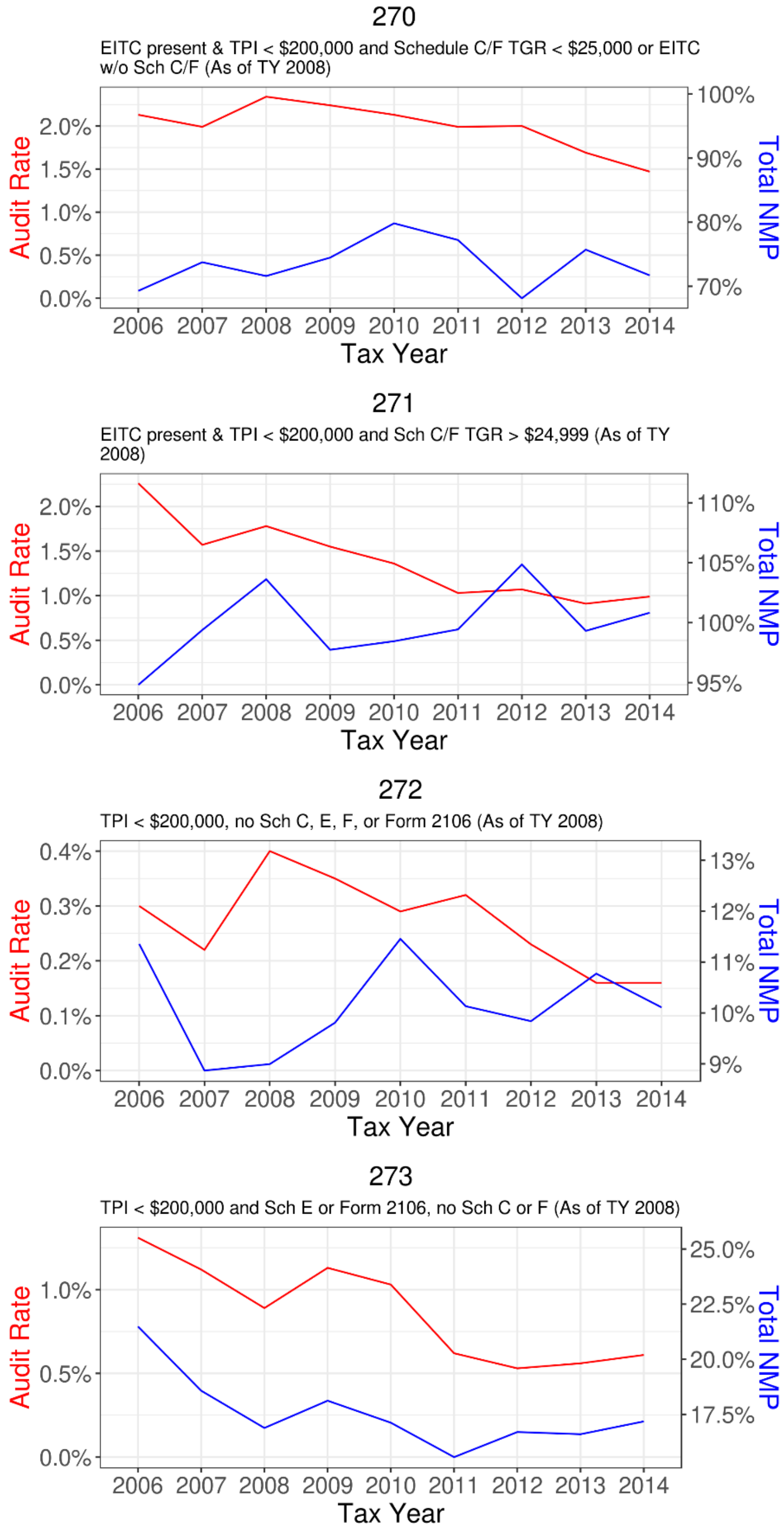


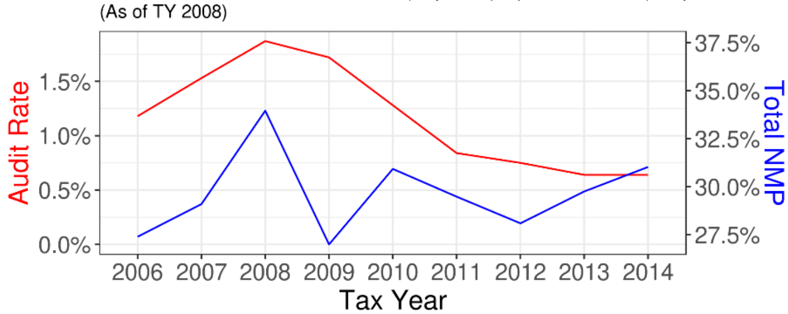
FIGURE 14. Audit Rate vs. Aggregate NMP, by Activity Code (Trimmed Data) (Continued)

274

Non-Farm Business w/ Sch C/F TGR < $25,000 and TPI < $200,000 (As of TY 2008)

275

Non-Farm Business w/ Sch C/F TGR $25,000 - $99,999 and TPI < $200,000 (As of TY 2008)



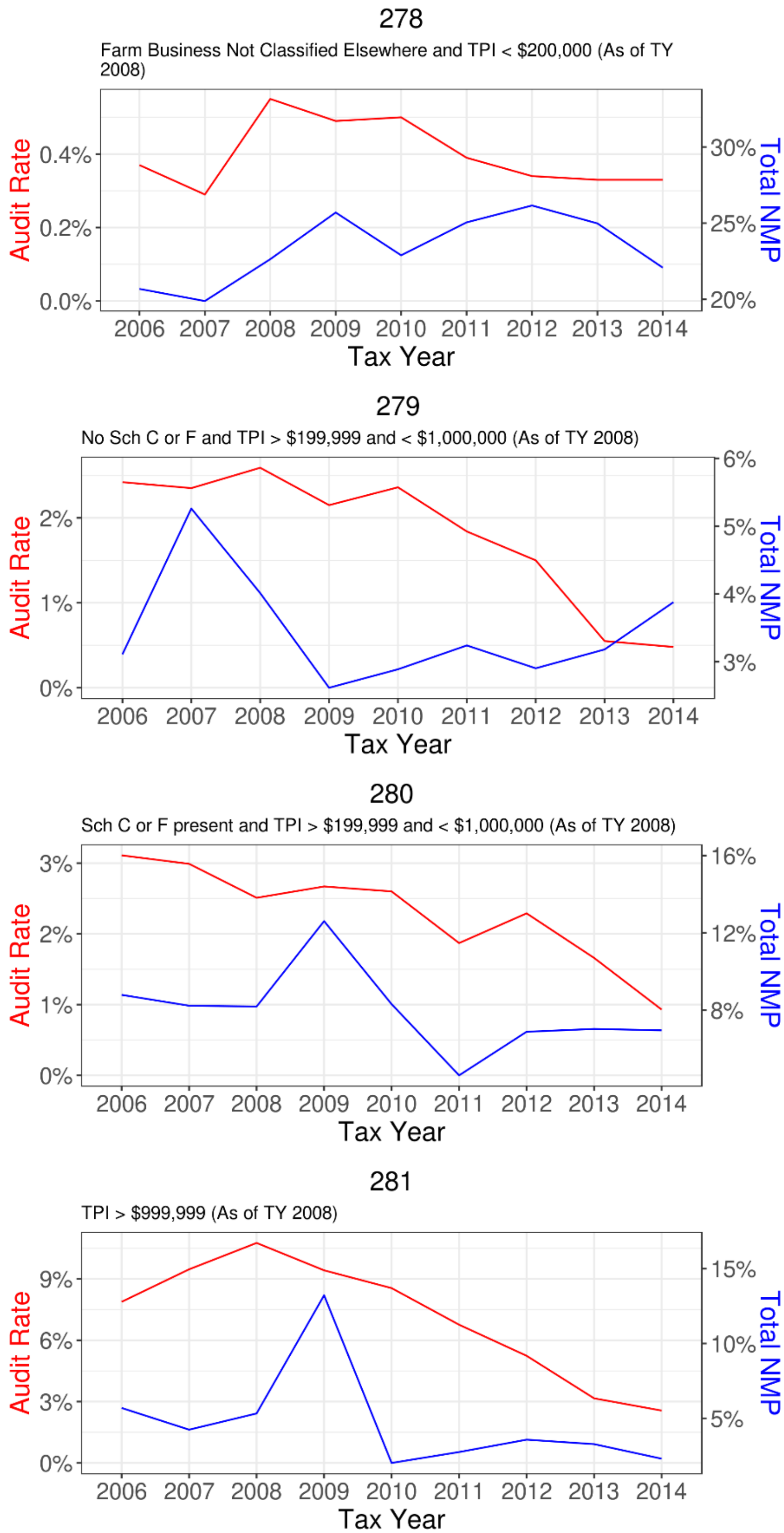
276

Non-Farm Business w/ Sch C/F TGR $100,000 - $199,999 and TPI < $200,000 (As of TY 2008)

277

Non-Farm Business w/ Sch C/F TGR > $199,999 and TPI < $200,000 (As of TY 2008)

FIGURE 14. Audit Rate vs. Aggregate NMP, by Activity Code (Trimmed Data) (Continued)



Supplementary Results

TABLE 6. Full Sample Baseline Results without Tax Year Fixed Effects

	Dependent variable: Log NMA						
	TARC	Visibility Group 1	Visibility Group 2	Visibility Group 3	Visibility Group 4	Visibility Group 5	Visibility Group 6
Audit Rate (2-Yr Lag)	0.099***	-0.051***	-0.052*	-0.119**	0.005	-0.018	-0.149***
	(0.025)	(0.017)	(0.030)	(0.051)	(0.030)	(0.032)	(0.024)
Corrected TARC	0.00001***						
	(0.00000)						
Correct Amount for Visibility Group		-0.00000	0.00000***	0.00000***	0.00000***	-0.00000***	-0.00004***
		(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Total Exemptions 1	1.638***	-0.338***	0.267**	0.246*	1.707***	0.658***	0.651
	(0.143)	(0.057)	(0.114)	(0.147)	(0.136)	(0.089)	(0.401)
Total Exemptions 2	3.123***	-0.577***	0.481***	0.838***	1.598***	3.450***	1.643***
	(0.145)	(0.059)	(0.118)	(0.163)	(0.144)	(0.092)	(0.401)
Total Exemptions 3	3.598***	-0.553***	0.581***	1.005***	1.726***	3.818***	2.326***
	(0.147)	(0.061)	(0.121)	(0.173)	(0.149)	(0.096)	(0.402)
Total Exemptions 4	3.652***	-0.558***	0.496***	0.735***	1.690***	3.860***	2.554***
	(0.150)	(0.063)	(0.124)	(0.179)	(0.155)	(0.101)	(0.403)
Total Exemptions 5+	3.995***	-0.473***	0.546***	1.188***	1.882***	3.996***	2.987***
	(0.153)	(0.065)	(0.128)	(0.187)	(0.161)	(0.106)	(0.404)
Wage Income	-0.092**		-0.061*	-0.168***	-0.135***	0.113***	-0.099**
	(0.038)		(0.032)	(0.050)	(0.041)	(0.042)	(0.049)
Claimed child tax credit	-0.229***	-0.057***	-0.193***	-0.038	-0.225***	-0.577***	-0.208***
	(0.031)	(0.018)	(0.034)	(0.068)	(0.052)	(0.034)	(0.031)
Itemized	1.386***	-0.300***	0.056	-0.376***	-0.122**	4.818***	0.198***
	(0.042)	(0.027)	(0.035)	(0.055)	(0.058)	(0.043)	(0.054)
Deducted mortgage interest	-0.477***	-0.048*	0.01	0.382***	0.398***	-0.723***	-0.400***
	(0.041)	(0.026)	(0.034)	(0.056)	(0.058)	(0.044)	(0.053)
Over 65	-0.612***	-0.274***	0.878***	-0.166***	-0.462***	-0.302***	-0.522***
	(0.040)	(0.028)	(0.032)	(0.052)	(0.049)	(0.043)	(0.055)
Used paid preparer	0.033	-0.044***	-0.154***	-0.110***	0.101***	-0.116***	0.018
	(0.023)	(0.013)	(0.022)	(0.043)	(0.037)	(0.024)	(0.026)
Filed electronically	-0.142***	0.037**	-0.090***	-0.079**	-0.286***	-0.065**	0.009
	(0.025)	(0.015)	(0.023)	(0.040)	(0.035)	(0.027)	(0.029)
Married-Joint Status	-1.689***	0.081***	0.038	-0.422***	0.176***	-2.763***	-1.046***
	(0.033)	(0.019)	(0.036)	(0.080)	(0.055)	(0.036)	(0.035)
Constant	1.264***	1.298***	0.818***	3.083***	3.235***	0.166	1.351***
	(0.157)	(0.067)	(0.135)	(0.206)	(0.160)	(0.117)	(0.405)
Observations	88,521	72,938	63,795	40,112	58,919	73,990	48,083
Tax Year Fixed effect	N	N	N	N	N	N	N
Adjusted R2	0.133	0.022	0.033	0.084	0.235	0.341	0.151
F Statistic	520.980***	67.154***	85.658***	142.824***	697.565***	1,473.283***	329.790***
Degrees of Freedom	88,445	72,854	63,750	40,072	58,864	73,879	48,031

Notes: Standard errors displayed in parentheses. *p<0.10, **p<0.05, ***p<0.01 Corrected amounts (for TARC and by visibility group) are specified in unscaled dollar values. Although statistically significant, the estimated coefficients are small because the amount of noncompliance is small relative to the overall amount of income or offsets.

3

Understanding Contemporary Taxpayers

Lin ♦ Samarakoon

Lopez-Luzuriaga ♦ Scartascini

Hoopes ♦ Menzer ♦ Wilde

Who Are Married-Filing-Separately Filers and Why Should We Care?

Emily Y. Lin and Navodhya Samarakoon (Office of Tax Analysis, U.S. Department of Treasury)[1]

1. Introduction

The U.S. federal income tax system recognizes families as an economic unit. Married couples file the income tax return jointly by pooling the spouses' incomes and deducting combined allowable expenses. Married couples can also elect to file separate returns, claiming the Married-Filing-Separately (MFS) filing status, but they are likely to face a higher tax liability as a result due to the unfavorable tax treatment of MFS status relative to Married-Filing-Jointly (MFJ) status. Of the 164.4 million federal individual income tax returns filed for Tax Year (TY) 2020, 55.3 million were filed as MFJ and 3.9 million were filed as MFS.[2] Counting each MFJ return as two married filers, for TY 2020, 96.6% of married filers who claimed either filing status filed jointly, leaving only 3.4% electing to file as MFS.

While it is well known that claiming the MFS status generally results in a higher federal income tax liability than claiming the MFJ status, it is little known to what extent married couples who file as MFS have a separate filing penalty where they face a higher federal income tax liability by filing separately.[3] Because there is no single condition or formula to apply, a married couple may not know which filing status leads to a lower tax liability until they run the calculation for both statuses. The Internal Revenue Service (IRS) Publication 501, Dependents, Standard Deduction, and Filing Information, informs married filers that the combined tax on separate returns is "generally" higher than the tax they would face on a joint return. The publication instructs married filers to "figure your tax both ways (on a joint return and on separate returns)" to be certain that they are "using the filing status that results in the lowest combined tax."[4] In addition to a possible lower tax bill, married couples may file separate returns for non-tax reasons. Numerous online articles, posted mostly by the media and tax preparation professionals or software companies, guide married taxpayers on how to choose the "better' filing status or when it makes sense to file separately.

In this paper, we summarize the situations in which married couples may prefer to file as MFS and situations in which filing a joint return may not be a choice for some married individuals. In addition, this paper fills the knowledge gap by providing novel statistics and data analysis about MFS claims. Possibly because of the assorted reasons why married couples file as MFS, no single profile can describe the small number of MFS filers. Our analysis shows that MFS filers consist of a diverse group of taxpayers across the income distribution and by how long they use this filing status. MFS filers are represented in all segments of the income distribution, with one-third having Adjusted Gross Income (AGI) below $30,000 and 14% having AGI above $100,000 in 2020. In addition, using administrative tax data for TYs 2013–2021, we find that over the 9-year period, more than half of the MFS filers claimed the status for only 1 year, and nearly 80% used it for 3 years or shorter. However, about 5% of those who ever claimed the status during this period did so for more than 7 years.

Examining the extent and the level of the separate filing penalty where a married couple pays more federal income tax by filing separate returns, our analysis shows that approximately 19–23% of MFS filers have federal income tax benefits by filing separately for an average amount of $1,513 (in 2021 dollars). Slightly fewer than a

---

[1] The authors would like to thank Adam Cole and the participants in the 2023 IRS/TPC Joint Research Conference for helpful feedback. The authors analyzing the tax data were employees at the U.S. Department of the Treasury. The findings, interpretations, and conclusions expressed in this paper 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. Any taxpayer data used in this research was kept in a secured Treasury or IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed.

[2] Refer to the Statistics of Income tax statistics posted on https://www.irs.gov/statistics/soi-tax-stats-individual-statistical-tables-by-filing-status.

[3] This paper examines only the federal income tax. Examination of the total federal and state liability is beyond the scope of this paper.

[4] See page 7 of IRS Publication 501, Dependents, Standard Deduction, and Filing Information

quarter of MFS files face the same federal income tax liability when filing a joint return, and about 53 to 59% of MFS filers have a federal income tax penalty by filing separately with an average penalty of $1,863 to $2,140. Taxpayer income and the presence of itemized deductions are positively associated with the bonus status. In addition, because MFS filers are generally ineligible for the Earned Income Tax Credit (EITC), the separate filing penalty is more prevalent among MFS filers who would claim the EITC when filing a joint return.

Although MFS status concerns only a small number of taxpayers, this filing status is associated with several policy and tax administration issues, including complexity, equity, and compliance. Complexity arises not only because married couples may have to calculate tax twice to decide on the filing status, but for taxpayers going through a separation or divorce, it can be confusing to determine the correct filing status. Separating individuals may have the living arrangements akin to those of unmarried individuals who file as single or head-of-household, but the tax rules governing marital status for the determination of filing status can be complicated to understand and interpret. On equity, MFS filers face an unfavorable tax treatment with respect to refundable tax credits. For low-income individuals who have difficulties in filing a joint return with their spouses, they would be denied the credits by filing as MFS. Lastly, the disparate tax liabilities across filing statuses, wherein filing as MFS leads to a higher tax liability than filing as unmarried filers, creates an incentive for separating individuals to misreport filing status. Because there is no third-party information about taxpayers' filing status, it is challenging for the IRS to detect this potential noncompliance absent an audit.

Our analysis shows that the regulatory and legislative efforts to allow vulnerable MFS filers to claim the Premium Tax Credit (PTC) and EITC in limited situations resulted in a very small fraction of MFS filers claiming each credit. Fewer than 2% of MFS filers claimed the PTC in recent years and, in 2021, the first year in which EITC was extended to MFS filers, 2% of MFS filers claimed the EITC. Because it is difficult to assess taxpayer eligibility under the specified rules that entitle MFS filers to these credits, further study is needed to determine whether the current claims are at their potential levels. Also, given the narrowly defined situations in which these credits are made available for MFS filers, MFS filers may need assistance in understanding their eligibility.

Finally, using audit results from a stratified random sample of individual income tax returns, we find that MFS status is susceptible to misreporting. During TYs 2006–2014, an average of 1.74% of returns were filed as MFS each year, but the audit results suggest that 2.68% of returns should have used this filing status. The vast majority of these misreporting cases involved erroneous claims of single or head-of-household status by audit-determined MFS filers. However, a small percentage of taxpayers incorrectly claimed the MFS status and as determined by the audit examiner, should pay a lower tax than the amount they reported on the tax return.

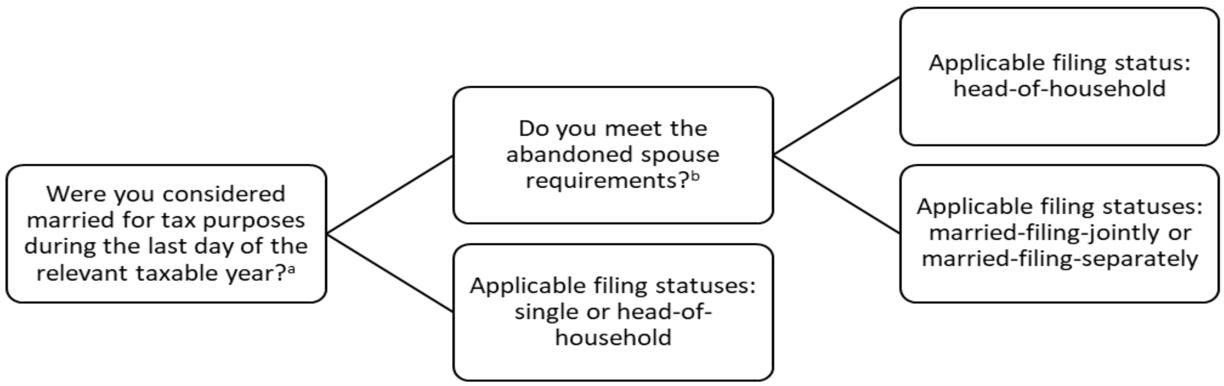
II. Tax Rules for Married-Filing-Separately

Prior to the enactment of joint taxation for married couples in 1948, the U.S. income tax system had only one filing status applied to all individuals regardless of marital status. With the creation of joint taxation in 1948, the joint status was used by married individuals, and the single status was used by unmarried individuals, as well as by married individuals who elected to file separate returns. At that time, the tax bracket schedules were designed in a way that all married couples paid no more federal income tax by filing jointly than they would if they filed separate returns using the single status. That is, there were federal income tax benefits, but no tax penalties, for married couples to file joint returns. Subsequent tax cuts were extended to unmarried individuals, which resulted in the marriage penalty and led to disparate tax schedules applied to different groups of non-joint filers. Specifically, head-of-household status was created in 1951 and expanded in 1954 to reduce the tax burden on unmarried individuals who had family responsibilities; the bracket widths for unmarried individuals who did not qualify as heads of households were broadened by the Tax Reform Act of 1969 (TRA69). In contrast, the tax schedule for married individuals who filed separate returns was maintained in these legislations.

As a result of the TRA69, married taxpayers who filed separate returns faced a different tax schedule from that for single taxpayers, effective in TY 1971. In addition, not only was MFS status less favorable than MFJ and head-of-household statuses, but it was also less favorable than single status. Despite frequent tax changes after 1969, the bracket disadvantage associated with the MFS status has never been eliminated. Table A-1 in the Appendix shows the tax brackets by filing status for TY 2022.

A taxpayer's marital status for tax filing purposes is determined by the taxpayer's status on the last day of the tax year.[5] A person is considered not married if the person is legally separated from the spouse, according to the state law, under a final decree of divorce or separate maintenance. A married couple can elect to file a joint return using the MFJ status or separate returns using the MFS status. In addition, certain married individuals who live apart from their spouses are considered as unmarried for filing status under the so-called abandoned spouse rules. The abandoned spouse rules are met if the taxpayer furnishes over half of the cost of maintaining the household that constitutes the principal place of abode of the taxpayer and a qualifying child for more than half of the tax year, and the taxpayer's spouse is not a member of the household during the last 6 months of the tax year. Figure 1 displays a flow chart on the determination of marital status for tax purposes and the filing status for married taxpayers.

FIGURE 1. Determination of Marital Status and Applicable Filing Status



[a] Per the Internal Revenue Code (IRC) Section 7703(a), an individual legally separated from the spouse, according to the state law, under a decree of divorce or of separate maintenance, is not considered married for tax purposes.

[b] Per IRC Section 7703(b), a married individual meets the abandoned spouse requirements, and thus is considered as unmarried for filing status, if (i) the taxpayer maintains as his home a household which constitutes for more than one-half of the tax year the principal place of abode of a qualifying child, (ii) the taxpayer furnishes over half of the cost of maintaining the household during the relevant taxable year, and (iii) the taxpayer's spouse is not a member of the household during the last six months of the taxable year.

As the tax system became more complex, an increasing number of provisions contributed to the unfavorable tax outcome for the MFS status relative to other filing statuses. MFS filers face limited eligibility for tax credits. They were not eligible for the EITC until TY 2021 when limited exceptions were allowed. They have very limited eligibility for the PTC and the Child and Dependent Care Tax Credit (CDCTC). They cannot take the education credits and the adoption tax credit at all, and are eligible for a reduced amount of the saver's credit. In addition to the unfavorable treatment with respect to tax credits, other rules lead to a higher tax liability for MFS status.[6] Specifically, MFS filers cannot take the exclusion for adoption expenses, the deduction for student loan interest, or the exclusion for interest income from qualified U.S. savings bonds used for higher education expenses. The maximum amount of the child and dependent care exclusion is half the size for MFS filers than for other taxpayers. If one spouse claims itemized deductions on the MFS return, the other spouse cannot take the standard deduction. In certain situations, MFS filers cannot claim the credit for the elderly or the disabled and must include in income a higher percentage of Social Security benefits. In addition, the income range of the phase-out schedule for the save's credit, the Child Tax Credit (CTC) and the credit for other dependents (ODC), as well as the exemption level for the Alternative Minimum Tax and the capital loss

[5] If a spouse dies during a tax year, the determination is made as of the time of the death.

[6] Refer to IRS Publication 501.

deduction limit are lower for MFS status than for MFJ status. These tax features could further increase the tax liability associated with MFS status for some married couples.

MFS filers are eligible to claim the PTC,[7] EITC,[8] and CDCTC[9] in limited circumstances. In March 2014, the IRS extended eligibility for the PTC to victims of domestic abuse and spousal abandonment, but a taxpayer cannot claim this relief for more than 3 consecutive tax years. To be eligible, the taxpayer must live apart from the spouse at the time of filing the tax return. A taxpayer is a victim of spousal abandonment if they cannot locate the spouse after a reasonably diligent effort is made (Mitchell (2016)). Beginning in 2021, separating couples who file as MFS may claim the EITC if they are separated under a legally binding, written separation agreement (but not a decree of divorce) and live apart from their spouses at the end of the tax year, in addition to meeting the same eligibility rules as the EITC. MFS filers may also be eligible to claim the EITC if they meet the abandoned spouse rules except for the household maintenance test. As for the CDCTC, MFS filers may be eligible to claim the credit if they meet the abandoned spouse rules except that the household they maintain is the home they reside in with a qualifying person for the CDCTC purposes (e.g., a disabled sibling) who is not a dependent child.[10] Table 1 summarizes the credit eligibility rules for MFS filers, most of which are similar to the criteria for being considered as unmarried for filing status.

For some married taxpayers, electing MFS status may result in a lower tax liability than filing jointly. If one spouse has low income and significant deductions subject to an AGI floor, it is possible that filing separately is advantageous. For example, medical expenses are deductible to the extent that expenses exceed 7.5% of a taxpayer's AGI. In the same manner, prior to 2018, business, investment, and certain miscellaneous expenses were deductible, subject to a 2% AGI floor. In a different scenario, if either spouse is a nonresident alien at any time during the tax year, the couple cannot file as MFJ and each spouse generally uses MFS status to report income subject to U.S. tax. However, U.S. persons married to a nonresident alien may elect to treat the nonresident alien spouse as a resident alien and file a joint federal income tax return. With this election, the worldwide income of both spouses is subject to U.S. income taxation, which can lead to undesirable tax consequences (Drumbl (2016)). Finally, based on the laws and regulations at the state level, the filing status choice on the federal return(s) may inform or determine the filing status the spouses may use on the state return(s). The filing status that minimizes the federal income tax liability may not be the same status that minimizes the total federal and state income tax liability. This paper focuses on the federal tax liability and leaves the choice of filing status when state tax is considered for future research.

[7] Refer to IRS Publication 974.

[8] Refer to IRS Publication 596.

[9] Refer to IRS Publication 503.

[10] If the qualifying person for the CDCTC is a qualifying child, then the filer will meet all the abandoned spouse rules and may file as head-of-household.

TABLE 1. Criteria for MFS Filers To Qualify for Tax Credits

Criteria or variations	Criterion for a married individual to file as head-of-household	When electing MFS status, is the criterion necessary to qualify for…			
		CDCTC	EITC: either (1) or (2)		PTC
			(1)	(2)	
Furnishes over half of the cost of maintaining the household during the relevant taxable year	Yes	Yes			
Maintains as the taxpayer's own home a household which constitutes the principal place of abode for a qualifying child for more than one-half of the tax year	Yes	Maintains as the taxpayer's own home a household which constitutes the principal place of abode for a qualifying CDCTC person for more than one-half of the tax year	Yes	Yes	
The taxpayer's spouse is not a member of the household during the last 6 months of the taxable year	Yes	Yes	Yes	Spouse does not live in the same household at the end of the taxable year	Spouse does not live in the same household at the time of tax filing
Taxpayer and spouse are separated under a legally binding written separation agreement or a decree of separate maintenance				Yes	
Is unable to file jointly because the taxpayer is a victim of domestic abuse or is unable to locate the spouse after reasonable diligence					Yes
Note	All of the above are met	All of the above and CDCTC eligibility are met	All of the above and EITC eligibility are met	All of the above and EITC eligibility are met	All of the above and PTC eligibility are met; cannot use the relief for more than 3 consecutive years

Data source: Author tabulation of tax rules and instructions in IRS publications.

III. Possible Non-Tax Reasons To Claim Married-Filing-Separately Status

Some married taxpayers may file as MFS involuntarily. Treasury regulations, as outlined above, allow domestic abuse and spousal abandonment exceptions for MFS filers to claim the PTC. In the instance of spousal abuse, the perpetrator may be noncooperative, refusing to furnish the necessary financial information needed to file jointly. It is also possible that the victim has left the home and does not wish to contact the abuser to file a joint return. For spousal abandonment, if the taxpayer does not have dependent children, does not meet the household maintenance test, or does not live apart from the spouse for the last six months of the tax year, the taxpayer is considered married and, because the spouse cannot be located, the taxpayer would have no choice but to file as MFS. With the PTC exception, the taxpayer may take the credit.

Married couples may file separately if each spouse would like to be responsible only for their own tax liability. In general, when a joint return is filed, both spouses are responsible for the tax and interest or penalties due on the return except for limited situations. Married taxpayers may file separately if they distrust that the spouse is accurately reporting the financial situation for tax purposes. In these cases, electing MFS status protects a taxpayer from IRS audits conducted on the spouse's tax return. Along a similar vein, choosing MFS status could protect a married person from being liable for the spouse's tax bill or from a refund offset that applies to the spouse. Estranged spouses who no longer live together or who do not have an emotionally co-dependent relationship may not share financial information to file a joint return. Couples in the process of getting a divorce may file separately to avoid the potential hassle of dealing with the IRS on a joint return after the divorce. It is not uncommon for couples in the process of a divorce to file separate returns. These individuals may use MFS status if they are not legally separated under a required court action or do not meet the specified rules such as living apart for the last 6 months of the year. Finally, married couples may file separately simply because the spouses want to stay financially independent.

One scenario in which it may be financially beneficial for married couples to file separate returns is if a taxpayer has large student loan expenses subject to an income-based repayment plan (Drumbl (2016)). When married taxpayers file jointly, the repayment amount will be based on the spouses' total incomes and therefore may be higher than if they file the tax separately. Lastly, married taxpayers may simply, and unfortunately, file separately because they lack access to necessary and accurate tax advice on what filing status would provide the most beneficial tax outcome.

IV. Data Analysis

A. *Shares of MFS Returns and Income Distribution*

Table 2 presents the shares of returns by filing status. Between TYs 2011 and 2020, the percentage of MFS returns grew steadily from 1.8% to 2.4% of total returns each year. MFS filers, along with single filers, have made up a rising share of the tax-filing population over the past decade. Conversely, the shares of MFJ and head-of-household filers have declined. Thus, MFS returns, despite constituting a small fraction of all returns, have increased in relative and absolute terms. Counting each MFJ return as two married filers, the share of married filers used MFS status increased from 2.4% to 3.4% during the period. Because some separate-filing spouses claimed head-of-household status, including these individuals resulted in 3.6% of married filers filing separate returns for TY 2020.

TABLE 2. Shares of Returns by Filing Status, 2011-2020

Tax Year	Number of All Returns	Married-Filing-Jointly	Married-Filing-Separately	Head-of-Household	Single	Share of Married Filers Using MFS
2011	145,370,240	36.7%	1.8%	15.2%	46.3%	2.4%
2012	144,928,471	37.1%	1.8%	15.1%	46.0%	2.4%
2013	147,351,298	36.6%	1.9%	14.9%	46.5%	2.5%
2014	148,606,578	36.3%	2.0%	14.9%	46.8%	2.7%
2015	150,493,262	36.1%	2.0%	14.7%	47.2%	2.7%
2016	150,272,156	36.0%	2.0%	14.4%	47.5%	2.7%
2017	152,903,232	35.8%	2.1%	14.3%	47.8%	2.9%
2018	153,774,296	35.7%	2.1%	14.2%	48.0%	2.9%
2019	157,796,805	34.7%	2.4%	13.7%	49.2%	3.3%
2020	164,358,794	33.7%	2.4%	13.1%	50.9%	3.4%

Data Source: Author calculations from data published in IRS SOI Publication 1304, Individual Income Tax Returns Complete Report (years 2011 through 2020).

By TY 2020, 3.9 million taxpayers filed as MFS. Table 3 breaks down the returns by AGI for each filing status in TY 2020, the most recent year for which the IRS's published statistics are available (IRS (2020)). MFS filers were represented in all income segments, but the fractions were higher than average for the income range of $30,000 to $100,000. Compared with other filing statuses, the income distribution suggests that MFS filers had higher income relative to single and head-of-household filers. In addition, when the spouses' incomes on separate returns were added to arrive at couple-level income, Table 3 shows that separate-filing couples were more likely to have income below $100,000, and less likely to have income of $100,000 or more, compared to joint-filing couples.

TABLE 3. Distribution of Income by Filing Status for TY 2020

Adjusted Gross Income (AGI, $)	All Returns	Married-Filing-Jointly	Married-Filing-Separately	Head-of-Household	Single	Couples Filing Separately[a]
≤ 0	3.2%	1.7%	4.0%	1.2%	4.7%	3.0%
0–15k	18.9%	5.2%	13.1%	14.8%	29.2%	6.8%
15k–30k	17.8%	7.6%	15.1%	29.2%	21.7%	8.2%
30k–50k	18.2%	11.5%	24.7%	27.9%	19.8%	13.4%
50k–75k	13.8%	14.7%	20.1%	14.4%	12.7%	18.0%
75k–100k	8.7%	14.7%	9.2%	6.1%	5.5%	15.0%
100k–200k	13.6%	30.2%	11.2%	5.2%	5.0%	26.0%
200k–500k	4.6%	11.5%	1.8%	1.0%	1.1%	8.2%
500k–1 million	0.8%	1.9%	0.4%	0.1%	0.2%	0.8%
≥1 million	0.4%	0.9%	0.3%	0.1%	0.1%	0.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Count	164,358,792	55,322,922	3,919,416	21,463,538	83,652,916	2,026,869

Data Source: Author calculation of the IRS Statistics of Income publications (IRS (2020)) and individual income tax returns filed for TY 2020.

[a] Of all MFS returns filed for TY 2020, 2.45 million returns can be paired, based on the taxpayer's and the spouse's identification numbers provided on the returns, as being filed by 1.22 married couples. Approximately 0.26 million MFS filers had a spouse who claimed the head-of-household status, and another 0.54 million MFS filers had a spouse who did not file a tax return, who we assumed had no income. The total includes 2.02 million married couples where at least one spouse filed as MFS. Couple-level income cannot be calculated for MFS filers who did not provide information about the spouse's taxpayer identification number. These filers are not included in this column.

B. Tax Penalty for Filing Separately

One way to assess the reasons why married couples file separately is to understand the prevalence and the level of tax penalty and bonus faced by MFS couples. The source of data we use for this evaluation is the population of approximately 31.8 million MFS returns filed for TYs 2013 to 2021. When a married couple files separately, the two MFS returns could be paired and the potential tax liability of the couple filed jointly can be calculated and compared with the combined liability on the separate returns. MFS filers are instructed to enter the spouse's name and either the Social Security Number (SSN) or Individual Taxpayer Identification Number (ITIN) on the tax return. We link an MFS return to the spouse's MFS return where the spouse's identification number matches the filer's identification number on another MFS return. This link results in 9.3 million married couples where both spouses claimed the MFS status for a total of 18.6 million MFS returns in the data.[11]

A large number, about 13.2 million, of MFS returns in our file are not linked to another MFS return for the spouse for various reasons. First, about 2.2 million MFS filers had a spouse who claimed the head-of-household status. For the analysis, we use the information reported on the spouses' head-of-household returns to calculate the couples' tax liability if filing jointly. Next, 2.1 million 1040-NR returns were filed by nonresident aliens who did not provide the spouse's Taxpayer Identification Number (TIN), either the SSN or ITIN. We

[11] In some cases, a spouse's MFS return is found but the match is not two-way. That is, return A has the spouse's Taxpayer Identification Number (TIN), either the Social Security Number or Individual Taxpayer Identification Number matched to the TIN of the filer on another MFS return B, but return B does not list a spouse corresponding to the filer's TIN on return A. Our analysis treats the two returns as filed by a married couple. See Table A-2 in the Appendix for details of the spouse-matching results.

assume the spouse was also a nonresident alien and exclude these 1040-NR MFS returns from analysis because these couples cannot use MFJ status. In addition, we exclude another 4.4 million MFS returns for which the spouse's TIN was missing or the spouse's TIN was found on a joint return for the same tax year because it is not straightforward as to how to determine these spouses' tax liability when filing as MFS.[12] For the rest of the 4.5 million unmatched returns, for which the spouse's identification number was available and did not appear on any tax return for the same tax year, the analysis assumes that these spouses were nonfilers and did not have income. We present results including and excluding these nonfiler cases. Also, Table A-2 in the Appendix shows the counts and percentages of the MFS returns with the various matching outcomes.

We use the TAXSIM model (Feenberg and Coutts (1993)) to simulate a couple's federal tax liability when the couple files jointly as well as the spouses' combined federal tax liability when filing separately.[13] We then take the difference between the two simulated liabilities (in 2021 dollars) to calculate the tax penalty (in negative values) and bonus (in positive values) facing the couple when the spouses file separate returns. We use a threshold of $5 to define the separate filing penalty or bonus. That is, a couple is considered as having a separate filing penalty (bonus) when the combined liability on the two separate returns is higher (lower) than the liability on the joint return by more than $5. The evaluation of the penalty or bonus is at the couple level, and the outcome for a couple applies to both spouses in the matched cases.

Table 4 presents the simulation results for all MFS returns as well as for MFS returns where both spouses filed a return. The average separate filing penalty is $646 for matched MFS returns and $987 for all MFS returns, including nonfiler cases. Most MFS filers face a separate filing penalty; about 53% of matched MFS filers and 59% of all MFS filers have a penalty, averaging $1,863 and $2,140, respectively. Slightly fewer than a quarter of MFS filers face the same liability filing separately or jointly, whereas about 19% to 23% have a bonus by filing separately, with a bonus amount of $1,513 on average. To put the amount of penalty and bonus in context, the total separate filing penalty represents about 12% (for matched returns) or 17% (for all returns) of the total joint liability for those who face a separate filing penalty. In comparison, the total separate filing bonus as a ratio of total joint liability is about 9%.

[12] Future research may explore ways to incorporate the spouse's tax information for the 3.2 million MFS returns with a missing spouse TIN through returns filed in different years.

[13] For this paper, we use TAXSIM version 32.

TABLE 4. Mean Variables of Tax Simulation Results for MFS Returns

Variable	ALL	Separate Filing Penalty	Neutral	Separate Filing Bonus
All MFS Returns, Including Couples Where One Spouse Did Not File				
Tax penalty ($)[a]	-987	-2,140	0	1,513
Fraction of all returns	100%	59.35%	21.93%	18.72%
Penalty as % of joint liability[b]	-7.45%	-17.34%	0%	9.15%
Adjusted gross income ($)	58,244	58,456	49,727	67,553
Age	46.96	47.51	45.99	46.37
Itemizer (0/1), self or spouse	0.3351	0.2902	0.1699	0.6713
Child tax credit (0/1)	0.1698	0.1786	0.1364	0.1811
EITC on joint return (0/1)	0.1041	0.1560	0.0054	0.0552
CDCTC on joint return (0/1)	0.0267	0.0236	0.0005	0.0673
Number of dependents	0.3251	0.3493	0.2454	0.3417
Any dependents (0/1)	0.2136	0.2235	0.1732	0.2295
Number of observations	24,761,774	14,695,734	5,431,231	4,634,809
Fraction with penalty or bonus by spouse filing status:				
Spouse MFS	100%	56.73%	26.78%	16.49%
Spouse head-of-household	100%	24.01%	2.22%	73.76%
Spouse nonfiler	100%	88.11%	11.89%	0%
Matched MFS Returns Only				
Tax penalty ($)[a]	-646	-1,863	0	1,513
Fraction of all returns	100%	53.16%	24.09%	22.75%
Penalty as percent of joint liability[b]	-4.13%	-11.77%	0%	9.15%
Adjusted gross income ($)	61,533	61,523	55,871	67,553
Age	46.62	47.44	45.03	46.37
Itemizer (0/1), self or spouse	0.3636	0.3180	0.1736	0.6713
Child tax credit (0/1)	0.1610	0.1602	0.1439	0.1811
EITC on joint return (0/1)	0.0630	0.0924	0.0054	0.0552
CDCTC on joint return (0/1)	0.0325	0.0320	0.0005	0.0673
Number of dependents	0.2969	0.3033	0.2406	0.3417
Any dependents (0/1)	0.2020	0.2025	0.1750	0.2295
Number of observations	20,376,065	10,831,696	4,909,576	4,634,793

Data source: Individual income tax returns filed for TYs 2013–2021.

Notes: The data contain 31,805,626 MFS returns. A total of 6,508,7554 returns are excluded from the analysis either because both spouses are nonresident aliens or the spouse's tax liability is not readily determinable. Another 99 observations are dropped in simulation due to missing variables. The table also excludes 543,998 returns (or 2.1% of returns in simulation) for which the difference between the simulated federal tax and the reported tax for either spouse is greater than $15,000, following an approach in Lin and Tong (2017). All money amounts are in 2021 dollars.

a The tax penalty evaluated at the couple level is applied to the spouse(s).

b This is the ratio of total penalty to total tax.

Table 4 shows that MFS filers in tax-neutral status have the lowest income of the three groups of MFS filers by penalty status. Low-income taxpayers may have no tax liability under either filing status and, thereby do not incur either the penalty or bonus. In addition, given the tax-related reasons for filing separately, as expected, claiming itemized deductions is positively related to having tax savings by filing separate returns. Nearly 70% of those with a bonus itemized their deductions, compared to the average rate of 35% for all MFS filers. Not reported in the paper, the bonus rate dropped by 40% in TY 2018 after the enactment of tax changes in the Tax Cuts and Jobs Act (TCJA) that significantly lowered the fraction of taxpayers who benefited from itemizing deductions. Specifically, for MFS filers matched to the spouses' returns, the share with the bonus declined from 28% before 2018 to 17% in 2018, and the share with the bonus for all MFS filers, including those with nonfiling spouses, declined from 23% to 14%. Another variable that is positively associated with the bonus status is taxpayer income, which may be related to the presence of itemized deductions.

In contrast, the separate filing penalty is more prevalent among individuals who would receive the EITC when they filed jointly. Because MFS filers generally cannot claim the credit, claiming it when filing a joint return would result in the separate filing penalty. About 16% (9%) of all (matched) MFS filers with the separate filing penalty would receive the EITC when filing jointly, exceeding the average rate of 10% (6%). For the same reason, the bonus rate was high, and the penalty rate was low, for MFS filers whose spouses filed as head-of-household because the head-of-household spouses could claim the EITC on their own returns. The bottom three rows of the top panel show the penalty percentages by the spouse's filing status. Only 24% of MFS filers with a head-of-household spouse incur the separate filing penalty, compared to 59% of all MFS filers.

C. Longitudinal Data on the Use of MFS Status

No data are available to accurately group MFS filers by reason for electing this status. We hypothesize that the duration for which an individual uses this filing status may inform the possible reason. In many of the scenarios as laid out above, the situation, such as if the couple is going through a divorce or if a separating couple misses the 6-month test in the first taxable year, but continues to live apart in future years, may be temporary. Hence, couples with shorter MFS election periods may be in transition from being married to single or have temporary difficulties in filing jointly. A long duration of MFS claims may indicate a prolonged separation with neither spouse meeting the abandoned spouse rules or any longer-term scenarios such as when spouses would like to keep tax or financial independence. Persistent MFS elections may also suggest a tax bonus from filing separate returns. Understanding the distribution of the length of MFS claims also helps to assess the extent to which a policy or tax law change affecting this filing status would have a short-term or long-term effect on taxpayers.

During the 9-year period from 2013–2021, about 13.4 million individuals with unique TINs claimed the MFS status for a total of 31.8 million tax returns. Calculating the number of years for which an individual used this filing status, we find that the majority of individuals who ever claimed the MFS status used it for a relatively short period a time. Table 5 shows that more than half of the individuals claimed the status for only 1 year, and nearly 80% used it for 3 years or shorter. However, about 5% of those who ever claimed the status did so for more than 7 years. On average, long-term users of MFS status are older and have a higher income compared to short-term users, with the average age and AGI increasing with the number of years for filing separately.

TABLE 5. Characteristics and the Percentages of Taxpayers by Duration of MFS Claims in 2013–2021

Number of Years with MFS Filing	Percentage of All MFS Filers	Accumulated Percentage	Mean Age in 2021	Mean Adjusted Gross Income (AGI) in 2021$
1	51.7%	51.7%	46	$55,290
2	18.3%	70.0%	48	$59,228
3	9.8%	79.8%	49	$64,292
4	6.0%	85.8%	51	$65,794
5	4.1%	89.9%	53	$67,757
6	2.9%	92.8%	54	$73,397
7	2.3%	95.1%	55	$74,360
8	1.9%	97.0%	57	$98,647
9	3.0%	100.0%	61	$118,026

Data source: Individual income tax returns filed as MFS for TYs 2013 to 2021.

Notes: Results in this table are calculated based on 13,370,930 individuals who ever filed as MFS in 2013–2021. For the mean AGI listed in the last column, we calculate the average AGI in 2021 dollars of each individual over the years when they filed as MFS, and then take the mean of the individual-level average across all individuals.

V. Complexity, Equity, and Compliance

Determining the filing status can be confusing for couples who are separated or in the process of getting a divorce. Taxpayers may not know if their separation agreement or living situations meet the standard of being considered as unmarried for filing status purposes. Under the tax code, separating couples are considered as unmarried only if they are separated under a court action recognized by state law as permanently severing the marriage relation. This definition excludes a non-final decree of divorce, a legally binding written separation agreement, or a court order of support. Further complicating the situation, due to differences in state law, the court action that meets the standard of legal separation varies across states (Ulven (1992)). As stated above, the tax code also provides an exception for married couples living apart to be considered as unmarried. The abandoned spouse rules are determined based on the couple's living arrangements, but they do not apply to separating individuals who do not have dependent children, do not live apart from their spouses for a required period, or do no furnish more than half of the cost of maintaining the household.

These restricted rules in defining marital status can disadvantage low-income taxpayers going through a separation or divorce. Low-income taxpayers may lack the resources to obtain tax advice that would help them minimize tax liability and determine the correct filing status. Consequently, they may be prone to paying a higher tax by filing separately or using an incorrect filing status inadvertently. In addition, a study finds that couples in prolonged separation tend to have low family income, have young children, and be racial and ethnic minorities (Taxpayer Advocate Service (2012)). Couples in extended separation may have no choice but to file separate returns from their estranged spouses because the spouse cannot be located or refuses to file jointly. However, they may remain married for a long time for tax filing purposes because they are not legally separated under a required court action and do not meet the abandoned spouse rules. For low-income taxpayers, the requirement to furnish more than half of the cost of maintaining the household can be particularly challenging as means tested public programs, such as food stamps and rental subsidies, count as outside support.

To a limited extent, this equity concern over vulnerable taxpayers was addressed by allowing certain MFS filers to claim a tax credit. Given the narrowly defined circumstances in which MFS filers may be eligible for these credits, as described in Section II, and the similar, but disparate, eligibility rules for each credit, it is important for low-income MFS filers to receive necessary assistance in understanding their eligibility. Table 6 shows the fractions of MFS filers who received the PTC, EITC and the CDCTC. The PTC was claimed by 0.6%−1.8% of MFS filers each year, with the fractions increasing gradually over the period. About 2% of MFS filers claimed the EITC in 2021, the first year in which the credit was extended to MFS filers. A persistent small fraction, less than 1%, of MFS filers claimed the CDCTC each year. With slightly fewer than 4 million taxpayers filing as MFS in recent years, only approximately 40,000 (1% of total) to 80,000 (2% of total) MFS filers claimed each of these credits. This result suggests that the special rules that relax the credit eligibility for MFS filers in sympathetic circumstances have created several very small groups of MFS filers claiming various tax credits. However, whether the current claims represent their potential levels is largely unknown because it is difficult to assess taxpayer eligibility under these special rules.

TABLE 6. Share of MFS Filers Claiming Certain Tax Credits, 2013–2021

Tax Year	PTC	EITC	CDCTC [b]
2013	X	X	0.6%
2014	0.6%	X	0.5%
2015	1.1%	X	0.4%
2016	1.3%	X	0.5%
2017	1.4%	X	0.5%
2018	1.5%	X	0.5%
2019	1.4%	X	0.5%
2020	--[a]	X	0.4%
2021	1.8%	2.0%	0.9%

Data source: Individual income tax returns filed as MFS for TYs 2013–2021.

Notes: The table is calculated based on a total of 31,805,626 MFS returns. “X” indicates that the credit was not available in the year (PTC) or not available for MFS status (EITC).

a We cannot use the same tax forms to calculate the share of MFS returns claiming the PTC for 2020 due to a temporary change in the requirement to repay excess advance payments of the PTC for that year.

b For TYs 2013–2020, the share is calculated based on the claims reported on Form 1040. We cannot use the same tax variable for 2021 due to the temporary expansion of the credit that resulted in changes to the tax form. The reported share for 2021 reflects the percentage of returns reporting the child and dependent care expenses on Form 2441.

According to the IRS, claiming an incorrect filing status is one of the common errors taxpayers make on their returns.[14] Not only do the complex rules increase the likelihood that taxpayers make inadvertent errors in filing status, but the lack of third-party information about individuals’ marital status and living arrangements, coupled with the disparate liabilities across filing statuses, also makes filing status susceptible to intentional errors. Absent an audit, IRS does not know whether a previously married person has the required court action for legal separation to be considered as unmarried. In addition, facts and circumstances of married couples living apart are similar to those of unmarried persons filing as single or head-of-household, especially because the abandoned spouse rules for married filers resemble the criteria for the head-of-household status for unmarried filers. Using data from random audits from TYs 2006–2008, Leibel (2014) finds that about 4% of all EITC claimants, including 2% of single taxpayers and 9% of head-of-household taxpayers claiming this credit, had MFS as the audit-determined filing status and thereby were ineligible for the credit. Hence, although the special rules that allow MFS filers to claim certain credits add to tax complexity, they may reduce the incentive for separating couples to misreport filing status as unmarried persons to claim the credits.

We investigate the extent of taxpayer reporting errors associated with MFS status using data from random audits. Table 7 shows the results from the audits of a stratified random sample of individual income tax returns conducted by the IRS’s National Research Program (NRP) for TYs 2006–2014. During this period, an average of 1.74% of tax returns claimed the MFS status each year. However, according to the filing status determined by the examiner, an average of 2.68% of tax returns should have claimed this filing status. The additional 0.94 percentage points come from 1.02% of returns that should have filed as MFS, but erroneously used either head-of-household (making up 79% of the erroneous claims) or single (the remaining 21% of the erroneous claims) status. This is net of 0.08% of returns that should have filed as MFJ, but erroneously used the MFS status. Overall, there is a large degree of misreporting associated with MFS status (1.10% of all returns) relative to the level that should be reported (2.68% of all returns).

[14] Refer to IRS tax tips for common errors on a return (https://www.irs.gov/newsroom/common-errors-on-a-tax-return-can-lead-to-longer-processing-times).

TABLE 7. Filing Status Errors Associated with MFS Status, 2006–2014

Reported Filing Status Is MFS (% of All Returns)	Corrected Filing Status Is MFS (% of All Returns): No	Yes	Total
No	97.24	1.02	98.26
Yes	0.08	1.65	1.74
Total	97.32	2.68	100.00

Data source: The NRP 1040 Study, 2006-2014.

Note: Results in this table are calculated based on 126,668 tax returns in the NRP 1040 Study for TYs 2006–2014.

For the total 1.10% of returns that made filing status errors associated with MFS status, we investigate the tax adjustments recommended by NRP examiners for these returns.[15] Table 8 shows the results by the type of errors. For returns where the filing status was changed from single or head-of-household to MFS by the examiner, the vast majority, or 96%, had a positive adjustment, meaning that the audit-corrected liability exceeded the liability reported on the return. Also, nearly 70% of these returns, as determined by the examiner, overclaimed the two child-related refundable credits. The audit-recommended increase in tax liability was $4,196 on average, including a recommended decrease of $2,318 in the two refundable credits. For returns that were corrected away from MFS status, 60% had a positive tax adjustment, whereas one-third had a negative adjustment. Refundable credit errors were not prevalent among these returns. The fact that some taxpayers overreported their liability by claiming the MFS status may indicate taxpayer confusion about their correct filing status.

TABLE 8. Recommended Tax Adjustments for Returns Associated with MFS Filing Status Errors

Adjustment Type	Reported Other Status, Corrected to MFS		Reported MFS, Corrected to Other Status	
	Mean	Std. dev.	Mean	Std. dev.
Adjustment for tax after credits ($)	4,196	4,791	2,204	8,046
Positive adjustment (0/1)	0.9619	0.1915	0.5992	0.4924
Negative adjustment (0/1)	0.0139	0.1171	0.3372	0.4750
Adjustment for EITC and additional CTC ($)	-2,318	2,497	-106	1,077
Negative adjustment (0/1)	0.6842	0.4651	0.0651	0.2480
Positive adjustment (0/1)	0.0118	0.1081	0.0463	0.2110
Number of observations		937		106

Data source: The NRP 1040 Study, 2006–2014.

Note: All money amounts are in 2021 dollars.

VI. Conclusion

Fewer than 4 million federal individual income tax returns were filed as MFS for TY 2020, representing only 2.4% of all returns or 3.4% of married filers. Because married couples generally face a higher federal income tax liability by filing separate returns, this paper examines the characteristics of MFS filers to understand why and how taxpayers use this filing status. We find that despite constituting a small share of taxpayers, MFS filers consist of a diverse group of individuals by income and by how long they use this filing status. MFS filers were represented in all segments of the income distribution and, while most MFS filers used the filing status for a

[15] A return can have multiple errors. The amount shown in the table does not distinguish the tax adjustment due to the filing status error from the adjustment due to other errors.

brief period, a small fraction used it for more than 7 years. This finding is by no means surprising because, as documented in the paper, married taxpayers in different circumstances may file as MFS for a variety of reasons.

Our analysis further shows that most MFS filers incur a separate filing penalty by paying more federal income tax than they would if they filed jointly with their spouses. Only 19%–23% of MFS filers enjoy a federal income tax bonus by filing separately, slightly fewer than a quarter of MFS filers face the same tax liability between filing separately and filing jointly, and 53–59% have a separate filing penalty by claiming the MFS status. The bonus status is positively associated with taxpayer income and the claim of itemized deductions. In contrast, the separate filing penalty is more prevalent among MFS filers who would receive the EITC if filing joint returns.

Finally, this paper considers complexity, tax administration, and compliance issues associated with the MFS status. Complexity arises because, for certain married taxpayers who are separated from the spouses, their living arrangements may be akin to those of unmarried individuals, but they are not considered as unmarried for tax filing unless very specific criteria are met. Taxpayers who have difficulties in filing joint returns but remain married must file as MFS, which makes them ineligible for various tax credits. Although the credit eligibility rules were relaxed for vulnerable MFS filers in limited circumstances, the percentage of MFS filers who claimed these credits was extremely low. Given the restrictive credit eligibility criteria, MFS filers may not know about their eligibility without IRS outreach and assistance. Finally, due to the tax incentive for separating persons to file as unmarried and a lack of third-party information for the IRS to verify filing status, compliance with MFS status is a concern. Our analysis shows a large percentage of filing status errors are associated with MFS status, some of which likely results from taxpayer misunderstanding about the correct filing status.

References

Drumbl, Michelle Lyon. (2016). "Joint Winners, Separate Losers: Proposals to Ease the Sting for Married Taxpayers Filing Separately." *Florida Tax Review*, 19(7): 399–464.

Feenberg, Daniel and Elisabeth Coutts. (1993). "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management*, 12(1), 189–194.

IRS. various years. *SOI Tax Stats—Individual Income Tax Returns Complete Report*. Publication 1304. https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-returns-complete-report-publication-1304.

Leibel, Kara. (2014). "Taxpayer Compliance and Sources of Error for the Earned Income Tax Credit Claimed on 2006–2008 Returns." *IRS Publication* 5161. https://www.irs.gov/pub/irs-soi/15rpeitctaxpayercompliancetechpaper.pdf.

Lin, Emily Y. and Patricia K. Tong. (2017). "Using Administrative Tax Data to Estimate Work Participation and Earnings Elasticities of Married Couples." *International Tax and Public Finance*, 24(6), 997–1025.

Mitchell, David S. (2016). "An Unhappy Union: Married Taxpayers Filing Separately and the Affordable Care Act's Premium Tax Credit." *Tax Lawyer*, 69(2): 453–476.

Taxpayer Advocate Service. (2012). "National Taxpayer Advocate 2012 Annual Report to Congress."

Ulven, Mark. (1992). "The Separation Penalty: Problems in Establishing Legal Separation for Filing Status." *Tax Lawyer*, 45(3): 903–913.

Appendix

TABLE A-1. Tax Brackets by Filing Status, TY 2022

2022 Individual Income Tax Table								
	Taxable Income							
	Married-Filing-Jointly		Married-Filing-Separately		Head-of-Household		Single	
Marginal Tax Rate	over	not over	over	not over	over	not over	over	not over
10%	$0	$20,550	$0	$10,275	$0	$14,650	$0	$10,275
12%	$20,550	$83,550	$10,275	$41,775	$14,650	$55,900	$10,275	$41,775
22%	$83,550	$178,150	$41,775	$89,075	$55,900	$89,050	$41,775	$89,075
24%	$178,150	$340,100	$89,075	$170,050	$89,050	$170,050	$89,075	$170,050
32%	$340,100	$431,900	$170,050	$215,950	$170,050	$215,950	$170,050	$215,950
35%	$431,900	$647,850	$215,950	$323,925	$215,950	$539,900	$215,950	$539,900
37%	$647,850	-	$323,925	-	$539,900	-	$539,900	-

Data source: IRS Revenue Procedure 2021-45.

TABLE A-2. Married-Filing-Separately (MFS) Returns by Spouse Matching Outcome

	Count	Percent of All (%)
All	31,805,626	100.0
1. Spouse's MFS return was found	18,585,142	58.4
2. Spouse claimed head-of-household status	2,245,219	7.1
3. MFS return was 1040-NR with a missing identification number for the spouse	2,135,598	6.7
4. Spouse's identification number was missing and the MFS return was not 1040-NR	3,206,243	10.1
5. Spouse filed a joint return	1,166,914	3.7
6. Spouse is a nonfiler	4,466,510	14.0

Data source: All MFS returns filed for TYs 2013–2021.
Note: MFS returns in groups (1), (2) and (6), or 79.5% of all MFS returns, are included in the simulation of the separate filing penalty.

Willing but Unable To Pay? The Role of Gender in Tax Compliance

Andrea López-Luzuriaga (Universidad del Rosario) and Carlos Scartascini (Inter-American Development Bank)[1]

1. Introduction

Do women and men behave differently when faced with tax obligations? Abundant evidence from field interventions (Wenzel (2006); Kleven *et al.* (2011); Alstadsaeter and Jacob (2013); Cabral *et al.* (2015); Advani *et al.* (2017)) and laboratory experiments (Fortin *et al.* (2007); Bazart and Pickhardt (2011); Eisenhauer *et al.* (2011); Castro and Rizzo (2014); Kogler *et al.* (2016); D'Attoma *et al.* (2017); D'Attoma *et al.* (2020)) shows that women are more likely to comply with their tax obligations than men. The main hypotheses for explaining the difference are that women are more risk-averse than men (Hibbert *et al.* (2013); Engstrom *et al.* (2015); Skatun (2017); Charness *et al.* (2018)), and women have higher levels of tax morale than men (Alm and Torgler (2006); Torgler (2005); Torgler and Valev (2010); Shafiq (2015); Cyan *et al.* (2016)).

If women are more likely to pay their taxes than men, does that imply they would respond more to a letter from the tax agency? There is no consensus on this matter. If women exhibit higher levels of tax morale or are more risk-averse, and noncompliance is driven by insufficient information or erroneous beliefs, an intervention could potentially be more successful in altering their behavior.[2] However, the intervention's impact cannot be disentangled from their initial compliance level (potential ceiling effect) or disposable income (potential corner solution).

In this article, we investigate whether women respond more to a message aimed at enhancing property tax compliance by evaluating the results from Castro and Scartascini (2015) across gender. Castro and Scartascini (2015) carried out a large field experiment exploring the determinants of property tax compliance in the municipality of Junín, Argentina, in 2011. The experiment included three treatment arms: one emphasizing penalty and detection probability (deterrence message), and two others conveying distinct tax morale messages (reciprocity and peer-effects messages).

The city government calculates the property tax based on basic indicators, such as the linear size of the lot fronting the street and the availability of public services in the neighborhood (serving as a low-accuracy proxy for housing values) and issues a tax bill bimonthly. Information asymmetries are absent, leaving taxpayers with a simple decision: to pay or not (no partial payments are accepted). The tax design, monitoring, availability of payment plans, or any other associated aspects do not factor in gender.

Our empirical findings reveal that women pay more than men, both at baseline and post-intervention. The data also suggest that women, following receipt of the deterrence message, tend to make earlier payments, hence increasing the likelihood of timely payment (intensive margin). However, overall compliance remains unchanged—those initially disinclined to pay remain unaffected by the intervention. In contrast, men in the treatment group exhibit an increased propensity to pay compared to their counterparts in the control group (extensive margin).

[1] We would like to thank the staff of the Municipality of Junín during Mayor Mario Meoni's tenure for providing the data, Lucio Castro for helping with the original data collection and intervention, and the Institutional Capacity Strengthening Fund (ICSF) of the Inter-American Development Bank, funded by the Government of the People's Republic of China, for its financial support for the original data collection. We have benefited from comments by many colleagues at numerous seminars and conferences. Our gratitude to all of them. The opinions presented herein are those of the authors and thus do not necessarily represent the official position of the institutions to which they belong. This paper also appears as IDB Working Paper 1330. Available at https://publications.iadb.org/en/willing-unable-pay-role-gender-tax-compliance.

[2] This would suggest an interior solution to the decision.

To understand these intriguing results, we perform multiple analyses. Firstly, we study the heterogeneous effects of the treatments and discover that the size of the tax liability impacts women's compliance (higher compliance at lower tax levels) but not men's, implying that women's decisions may be contingent on their financial situation.

Secondly, we employ survey data, targeted at the same population as the original experiment (though not the same sample), to explore the differences in motivations and resources between men and women. The survey data indicate that female-headed households are more likely to internalize enforcement probabilities (*i.e.*, they have a stronger belief in the city government's enforcement capabilities). However, they are also more likely to be poorer and perceive the tax as excessively high. These findings suggest that women are responsive to the messages but may be hindered by budget constraints.

The context of the field experiment and the design of the tax point towards potential liquidity constraints. Given that the property tax is independent of current income level, it may exceed a taxpayer's budget. In Argentina, mortgage financing is almost nonexistent, contributing to less than one percent of GDP, one of the lowest rates globally. Thus, the correlation between wealth stock and income flow is less significant than in other countries. This disconnect between taxation and current income is common in the developing world due to shallow credit markets, limited options for leveraging assets as collateral, and heavy reliance on indirect taxes—personal income taxes account for approximately 10% of total revenues in Latin America and the Caribbean, in contrast to around 25% in the OECD (Corbacho *et al.* (2013); Acosta-Ormaechea *et al.* (2022)).[3]

To gain insights into these results, we introduce a simple analytical model where the only decision taxpayers make is whether to pay the tax (with the government determining the size of the tax bill), mirroring the scenario with property taxes. The model predicts that individuals with higher levels of tax morale or risk aversion are more likely to enhance their compliance following an intervention that increases the perceived likelihood of detection. However, liquidity constraints could force a corner solution: if the tax exceeds current disposable income, individuals do not respond to the intervention.

Our results carry significant implications. Firstly, they highlight a gender disparity in compliance—women, given the same enforcement levels, comply more frequently than men. As a result, taxation could widen post-tax income inequality between genders in countries with low enforcement where a significant portion of the population evades taxes. This is compounded by the fact that women-led households typically have lower incomes. Therefore, they are disproportionately affected in developing countries where a substantial share of taxation is not income based. Secondly, reactions to the same messages vary across individuals, implying that tax authorities might need to tailor their interventions accordingly. Lastly, liquidity constraints could influence tax compliance when the tax base does not correlate highly with income.

2. Background and Data

The data for this analysis originate from a large-scale field experiment conducted by Castro and Scartascini (2015) to investigate the determinants of property tax compliance in Junín, Argentina, in 2011. The city government calculates the tax and sends the bills every two months. The property tax is levied on homes, farms, business premises, and most other real estate in the city of Junín. The tax is calculated based on the length of the street front of the property in meters (not on the size of the property nor its quality), the number of streetlights around the property, and the trash collection and street cleaning services provided to the area where the property is located. All these variables are known by the city government and cannot be influenced by the taxpayer.

The intervention introduced a message into the tax bill. Three distinct treatment messages were used: a deterrence message detailing the penalties for late payment, a reciprocity message describing the uses of the collected funds, and a peer-effect message providing information about the overall compliance rate. Each message's text can be found in Table 1. An example of the tax bill is available in Figure A1 in the Appendix.

[3] Moreover, in Argentina and other developing countries, a significant proportion of taxpayers owing income tax are part of a simplified tax regime. In these regimes, the tax owed remains constant within broad income brackets. For instance, in Argentina in 2021, individuals at the lower bound of the first bracket paid about 2% of their sales in income taxes, while those at the upper bound of the same bracket paid less than 1%.

The tax has two due dates. The initial due date typically falls in the second week of the month, with the secondary due date in the following week. While payment is expected by the initial due date, no late fees are levied if payment is made by the secondary due date. Any outstanding liabilities incur a monthly compound interest rate of 2%. We leverage this payment scheme to analyze compliance by gender at different times.[4]

The taxpayer database includes the names of each property owner and the individual responsible for paying the tax. From this information, we were able to infer the gender assigned at birth to the individual liable for the property tax. In Argentina, parents are permitted to select their children's names from a pre-approved list of approximately 10,000 female and male names.[5] Using this list, we constructed a gender variable for 92% of the sample, or about 21,500 taxpayers, 34% of whom were women.[6] There are only a few names that can be used by both women and men. The gender variable is balanced across treatments, control groups, and all other baseline observables (see Table A2 in the Appendix).

We use additional data from two external surveys to analyze the interplay between liquidity constraints and the impact of the intervention. These surveys target the same demographic as the original experiment but do not necessarily include the exact same individuals. The first survey, conducted by the city government following the intervention, targets the household member responsible for property tax payment and asks about their attitudes towards the tax. The second survey is the Urban Household Survey of 2011 (Encuesta Anual de Hogares Urbanos, EAHU), which we use to understand the characteristics of households led by women.[7]

3. Empirical Results

Does gender affect compliance? Using the baseline (pre-treatment) information, we find that women are more likely to pay than men (44% versus 39%), to pay on time (24% versus 21%), and to have paid at least once in the past (54% versus 49%). These results align well with the existing stylized facts in the literature. In addition, properties owned by female-headed households share some common characteristics. Their properties are smaller and receive more public services from the municipal government, which means that they are more centrally located. Men own more properties than women, on average. We control for all these characteristics (the log of the number of properties of each taxpayer, the log of the average linear font size of the properties, trash collection, and street lighting services) across our analysis. As we have mentioned, there is balance across treatment and control groups (characteristics of the tax and property by gender and balance test are in Table A1 and Table A2 in the Appendix).

Building upon Castro and Scartascini (2015)'s treatment assignment and property tax payment scheme, we assess three payment outcomes: payment by the first due date, payment by the second due date, and full payment within the 2 months billing cycle (paid). Castro and Scartascini (2015) reported that the deterrence message was the most successful on average for increasing compliance. Analyzing all individuals together, taxpayers who receive the deterrence letter are more likely to pay by the first and second due dates and more likely to pay overall.[8]

To explore gender disparities, we conduct two types of analysis. First, we introduce an interaction term with the gender variable in Castro and Scartascini (2015)'s baseline regressions to assess gender differences in treatment. Second, we examine the treatment effects within each gender sample (results are presented in Table 2).

The results show very little difference across genders. When examining the main variable of interest—the payment of the tax by the end of the period—it appears that men respond slightly more to the reciprocity

[4] More details about the intervention are available in Castro and Scartascini (2015).

[5] See https://data.buenosaires.gob.ar/dataset/nombres.

[6] Due to data availability, our analysis is limited to gender differences assigned at birth.

[7] The National Institute of Statistics and Censuses in Argentina ("Instituto Nacional de Estadíisticas y Censos") conducts the EAHU annually. While the survey represents the subregion level, it does not accurately represent the city level. Buenos Aires province is split into six subregions: Buenos Aires (city), Gran La Plata, Bahía Blanca, Partidos del GBA, Mar del Plata, and several smaller cities combined into one region. Junín is included in this final region.

[8] They are 2 percentage points more likely to pay by the first due date, 3 percentage points more likely to pay by the second due date, and 5 percentage points more likely to have paid the tax bill. Our results are slightly different from those presented in Castro and Scartascini (2015) because our sample is smaller–we could not infer the gender for all individuals.

message. However, this result seems to be driven more by a decrease in compliance for women rather than an increase for men, which aligns with the overall finding in Castro and Scartascini (2015). In their study, taxpayers receiving more public goods from the government (in this case, women) showed a negative response to the government's depiction of the utilization of the tax revenue. Therefore, the observed effect appears to be contingent on location rather than gender.

Given baseline differences across genders, what happens when we look within samples? Once we divide the population according to gender, more significant differences appear, particularly for the deterrence message, which has been shown to be the most relevant, on average. The deterrence message has two objectives: increase the perception of risk as well as the salience of the penalty. It reminds the taxpayer of the legal tools the city government has to collect unpaid taxes; this part of the message aims to increase the perceived probability that the tax authority will enforce the penalty. The message also explains the fines for not paying, illustrating how a compound interest rate works. This part of the message aims to make the fine more salient.

Looking first at paid (at the end of the period), we find that the deterrence letter did not significantly increase the overall payment among women. Still, it increased the timeliness of payment (paid by the first and second due date). For women who received the treatment letter, the probability of paying by the first due date and the second due date was 4 percentage points and 3 percentage points higher, respectively, than the women in the control group, both results significant at the 5% level. In contrast, it had a larger effect on payment behavior among men. Men who received the deterrence letter were more likely to pay overall than men in the control group by 2 percentage points. There is no difference in the payment by the first due date between men in the treatment and control groups. Men in the deterrence group are 1 percentage point more likely to pay by the second due date than men in the control group, but that difference is only significant at the 10% level. Figure 1 and Figure 2 summarize the results. These results are compatible with an analytical model with cash constraints, which we describe next.

4. A Gender-Based Compliance Analytical Framework

There is some evidence that women are better taxpayers than men. There are two possible explanations in the literature: women are more risk-averse and have higher levels of tax morale. Disposable income could be another potential source of systematic differences in tax compliance if women face more liquidity constraints. This mechanism would have the opposite effect by making women less likely to pay their tax liabilities. To disentangle the impact of these three channels and focus on the role of enforcement in tax payments, we build a simple model to understand compliance behavior, allowing for tax morale, risk aversion, and income differences. In our model, available as an IDB Working Paper,[9] the taxpayers maximize their expected utility of after-tax income. They can pay their government-assessed tax, T, or they can enter a lottery, where they would pay the tax and a fine, θ, with probability , or keep their full income with probability $(1 - p)$. Following Dwenger *et al.* (2016), we model the intrinsic motivation to pay taxes, S, as a positive monetary value that is added to the income after tax.

We find that, in equilibrium, there is a probability, p^*, that makes individuals indifferent between paying the tax or not. Suppose the taxpayer's perceived probability of enforcement is lower than this indifference probability. In that case, the taxpayer will decide not to pay the tax, but will pay the tax if the perceived probability is higher. This indifference probability decreases with respect to the intrinsic motivation parameter and the coefficient of absolute risk aversion. Those individuals with higher tax morale or risk aversion should react more to an intervention that increases the salience of the probability of being prosecuted for not paying the tax. Consequently, if women have higher tax morale and risk aversion levels than men, as identified in the broad literature, women will comply more than men and react more than men to an intervention.

While these predictions would hold for a tax proportional to income, predictions may be more nuanced if there are liquidity constraints. In many developing countries, where credit and housing markets are underdeveloped, the property tax is calculated based on some general characteristics of the house (such as lot and construction size) and not on the house's value. As such, the assessed tax may be disconnected from the asset's

[9] https://publications.iadb.org/en/willing-unable-pay-role-gender-tax-compliance

value. Also, because mortgages are rare and owners cannot convert the asset into income flows, the tax may be disconnected from current or liquid income. For instance, taxpayers were more likely to decrease their consumption after an increase in the property tax in Mexico City (Brockmeyer *et al.* (2021)). To account for this fact, we add a budget constraint given by a minimum required consumption level to the model. When the disposable income (income minus assessed tax) is lower than the minimum level of consumption needed, the taxpayer does not pay the tax (in the model and the actual world, partial payments of the property tax are not possible). These cash-constrained taxpayers do not react to the intervention (tax agency deterrence message) even if the message successfully alters their perceptions (*i.e.*, they find themselves in a corner solution).

Therefore, given the stylized facts about gender differences in the literature, the model predicts that if women have higher levels of tax morale or are more risk-averse, they will react more than men to an intervention that increases the salience of the probability of being prosecuted for not paying the tax. If current income of women is lower than that for men, then there is a higher probability that more of them will face a corner solution and be unable to react even in the context of an intervention that increases their perceived probability of detection.

5. Discussion

The analytical model shows that if women are financially constrained, then the empirical results where women pay more on average—but those who do not pay do not react to the treatment—are plausible. To evaluate the likelihood of this, we turn to survey data. First, we look at the data from a survey of taxpayers in the city of Junín. Responses to the survey indicate that women indeed perceive higher levels of enforcement; see Figure A2 in the Appendix. Women are also more likely to think the property tax is too high and say that they are unwilling to pay a higher tax—see Figure A3 in the Appendix. Second, looking at the urban household survey, we learn that female-headed households are poorer (male income is about 30% higher) and less likely to have a steady income than male-headed households (men have a 13-percentage point higher probability); see Table A3 and Figure A4 both in the Appendix.

Our findings, in addition to the suggestive evidence coming from the survey data, seem to indicate that women might be more willing to pay taxes for fear of enforcement, react more to a deterrence treatment as indicated by the model, but have lower resources to face a tax that is not highly correlated to income. We find additional support for this hypothesis in the heterogeneous analysis by looking at the treatment's impact according to tax size. The effect of the deterrence letter is positive and significant for women whose tax bill is lower (up to 10 percentage points). Yet the difference disappears as the tax liability increases—suggesting that the amount of the tax is essential for women in deciding whether to pay. For men, however, the effect does not change significantly as the tax liability increases; see Figure 3.

6. Conclusion

Our findings reveal that women generally exhibit higher compliance with tax obligations than men and may be more responsive to deterrence letters issued by city governments. In the treatment group, women who received these letters were more inclined to make timely payments compared to those in the control group. Notably, the deterrent effect of the letters on women's compliance was markedly pronounced when tax liability was low. However, this effect diminished as the tax liability increased, possibly due to the high illiquidity of the taxed asset. These outcomes align with an analytical model that considers budget constraints. Further analysis, using survey data, validates the model and empirical outcomes, indicating that women—more so than men—trust the government's tax enforcement ability, yet are more vulnerable to cash constraints. This susceptibility likely stems from their generally lower income, lesser likelihood of earning a fixed income, and greater tendency to perceive the tax as high.

Our research highlights that, in scenarios characterized by lax tax enforcement and significant evasion, tax policy and enforcement mechanisms could inadvertently widen the income gap between genders. Given that women typically earn lower salaries yet are more likely to comply with their tax obligations, this dynamic may exacerbate existing income disparities, particularly in developing countries where a small fraction of tax is proportional to income. As such, tax policy and enforcement initiatives should recognize and address these disparate impacts. Optimally, more robust enforcement under a given tax policy should strive to diminish, not amplify, inequality. Policy tools could potentially ameliorate this gender disparity without infringing upon the principle of horizontal equity in tax design. In the context of property taxes in illiquid markets, or taxes not proportional to wealth, a plausible solution might involve tying property tax indirectly to current income levels. For example, low-income households could receive a property tax discount or access differentiated payment plans based on income.

We hope our study encourages additional field experiments that explicitly incorporate gender into their design and explore a variety of enforcement strategies. Tax authorities ought to pursue enforcement methods that are, at the very least, gender neutral. Gaining a nuanced understanding of when and how such gender neutrality can be achieved remains a critical endeavor.

References

Acosta-Ormaechea, Santiago, Pienknagura, Samuel, and Pizzinelli, Carlo (2022). "Tax Policy for Inclusive Growth in Latin America and the Caribbean." In: IMF Working Paper 22.8.

Advani, Arun, Elming, William, and Shaw, Jonathan (Oct. 2017). "The Dynamic Effects of Tax Audits." Working Paper.

Alm, James, and Torgler, Benno (Apr. 2006). "Culture Differences and Tax Morale in the United States and in Europe." In: *Journal of Economic Psychology* 27.2, pp. 224–246.

Alstadsaeter, Annette, and Jacob, Martin (Apr. 2013). "Who Participates in Tax Avoidance?" SSRN Scholarly Paper ID 2261637. Rochester, NY: Social Science Research Network.

Bazart, Cecile, and Pickhardt, Michael (Jan. 2011). "Fighting Income Tax Evasion with Positive Rewards." In: *Public Finance Review* 39.1, pp. 124–149.

Brockmeyer, Anne, Estefan, Alejandro, Ramírez Arras, Karina, and Suárez Serrato, Juan Carlos (Apr. 2021). "Taxing Property in Developing Countries: Theory and Evidence from Mexico." Working Paper 28637. Series: Working Paper Series. *National Bureau of Economic Research*.

Cabral, Ana Cinta G., Myles, Gareth, and Kotsogiannis, Christos (May 2015). "Self-Employment Underreporting in Great Britain: Who and How Much?" In: 14th Journées Louis-André Gérard-Varet.

Castro, Lucio, and Scartascini, Carlos (Aug. 2015). "Tax Compliance and Enforcement in the Pampas Evidence from a Field Experiment." In: *Journal of Economic Behavior & Organization* 116, pp. 65–82.

Charness, Gary, Eckel, Catherine, Gneezy, Uri, and Kajackaite, Agne (Feb. 2018). "Complexity in Risk Elicitation May Affect the Conclusions: A Demonstration Using Gender Differences." In: J*ournal of Risk and Uncertainty* 56.1, pp. 1–17.

Corbacho, Ana, Fretes Cibil, Vicente, and Lora, Eduardo (2013). More Than Revenue: Taxation as a Development Tool. Inter- American Development Bank and Palgrave McMillan.

Cyan, Musharraf R., Koumpias, Antonios M., and Martinez-Vazquez, Jorge (Dec. 2016). "The Determinants of Tax Morale in Pakistan." In: *Journal of Asian Economics* 47, pp. 23–34.

Dwenger, Nadja, Kleven, Henrik, Rasul, Imran, and Rincke, Johannes (Aug. 2016). "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." In: *American Economic Journal: Economic Policy* 8.3, pp. 203–232.

D'Attoma, John, Volintiru, Clara, and Malézieux, Antoine (2020). "Gender, Social Value Orientation, and Tax Compliance." In: *CESifo Economic Studies* 66.3, pp. 265–284.

D'Attoma, John, Volintiru, Clara, and Steinmo, Sven (2017). "Willing to Share? Tax Compliance and Gender in Europe and America." In: *Research Politics* 4.2.

Eisenhauer, Joseph G., Geide-Stevenson, Doris, and Ferro, David L. (Mar. 2011). "Experimental Estimates of Taxpayer Ethics." In: *Review of Social Economy* 69.1, pp. 29–53.

Engstrom, Per, Nordblom, Katarina, Ohlsson, Henry, and Persson, Annika (Nov. 2015). "Tax Compliance and Loss Aversion." In: *American Economic Journal: Economic Policy* 7.4, pp. 132–164.

Finocchiaro Castro, Massimo, and Rizzo, Ilde (Aug. 2014). "Tax Compliance Under Horizontal and Vertical Equity Conditions: An Experimental Approach." In: *International Tax and Public Finance* 21.4, pp. 560–577.

Fortin, Bernard, Lacroix, Guy, and Villeval, Marie-Claire (Dec. 2007). "Tax Evasion and Social Interactions." In: *Journal of Public Economics* 91.11, pp. 2089–2112.

Hibbert, Ann Marie, Lawrence, Edward R., and Prakash, Arun J. (2013). "Does Knowledge of Finance Mitigate the Gender Difference in Financial Risk-Aversion?" In: *Global Finance Journal* 24.2, pp. 140–152.

Kleven, Henrik Jacobsen, Knudsen, Martin B., Kreiner, Claus Thustrup, Pedersen, Soren, and Saez, Emmanuel (May 2011). "Unwilling or Unable To Cheat? Evidence From a Tax Audit Experiment in Denmark." In: *Econometrica* 79.3, pp. 651–692.

Kogler, Christoph, Mittone, Luigi, and Kirchler, Erich (Apr. 2016). "Delayed Feedback on Tax Audits Affects Compliance and Fairness Perceptions." In: *Journal of Economic Behavior & Organization*. Taxation, Social Norms and Compliance 124, pp. 81–87.

Shafiq, M. Najeeb (Apr. 2015). "Aspects of Moral Change in India, 1990-2006: Evidence from Public Attitudes Toward Tax Evasion and Bribery." In: *World Development* 68, pp. 136–148.

Skatun, John Douglas (July 2017). "Taxation, Risk Aversion, and the Wage Gaps in Tournaments." In: *Oxford Economic Papers* 69.3, pp. 834–845.

Torgler, Benno (2005). "Tax Morale in Latin America." In: *Public Choice* 122.1/2, pp. 133– 157.

Torgler, Benno and Valev, Neven T. (Oct. 2010). "Gender and Public Attitudes Toward Corruption and Tax Evasion." In: *Contemporary Economic Policy* 28.4, pp. 554–568.

Wenzel, Michael (Sept. 1, 2006). "A Letter From the Tax Office: Compliance Effects of Informational and Interpersonal Justice." In: *Social Justice Research* 19.3, pp. 345–364.

Tables

TABLE 1. Messages included in the tax bill

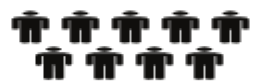
Message / Group	Text	Image
Deterrence	Did you know that if you do not pay the CVP on time for a debt of AR$ 1,000 you will have to disburse AR$ 268 in arrears at the end of the year and the Municipality can take administrative and legal action?	
Reciprocity	In the first 6 months of this year, CVP's collection contributed to placing 28 new streetlights, water connections in 29 streets and sewerage networks in 21 blocks	
Peer-effects	Did you know that only 30% of taxpayers do not pay the CVP? What about you?	
Control	No message	No image

TABLE 2. Average Treatment Effect (Sep/Oct)

	Paid 1st Due Date			Paid 2nd Due Date			Paid		
	Men	Women	Int.	Men	Women	Int.	Men	Women	Int.
Men			-0.009			-0.009			-0.008
			(0.009)			(0.007)			(0.007)
T1: Deterrence	0.006	0.036**	0.037**	0.013*	0.028**	0.029**	0.020**	0.011	0.012
	(0.008)	(0.016)	(0.016)	(0.007)	(0.012)	(0.012)	(0.008)	(0.010)	(0.010)
T1: Deterrence × Men			-0.031			-0.016			0.008
			(0.019)			(0.015)			(0.012)
T2: Reciprocity	0.006	-0.024*	-0.023*	0.005	-0.016	-0.014	0.006	-0.014	-0.013
	(0.011)	(0.012)	(0.012)	(0.009)	(0.011)	(0.011)	(0.009)	(0.011)	(0.011)
T2: Reciprocity × Men			0.030*			0.020			0.020*
			(0.017)			(0.012)			(0.011)
T3: Peer-Effect	0.005	0.000	0.000	0.006	0.001	0.002	0.001	-0.001	-0.001
	(0.006)	(0.014)	(0.014)	(0.007)	(0.011)	(0.011)	(0.009)	(0.011)	(0.011)
T3: Peer-Effect × Men			0.005			0.004			0.002
			(0.016)			(0.011)			(0.014)
N	14,003	7,301	21,304	14,003	7,301	21,304	13,995	7,292	21,287
Gender	Men	Women	Both	Men	Women	Both	Men	Women	Both
$F\left(\beta_{T1}=0, \beta_{gender}=0, \beta_{T1\times gender}=0\right)$			5.780			5.502			2.495
$p\left(\beta_{T1}=0, \beta_{gender}=0, \beta_{T1\times gender}=0\right)$			0.004			0.005			0.084
$F\left(\beta_{T2}=0, \beta_{gender}=0, \beta_{T2\times gender}=0\right)$			1.347			0.992			1.070
$p\left(\beta_{T2}=0, \beta_{gender}=0, \beta_{T2\times gender}=0\right)$			0.283			0.413			0.380
$F\left(\beta_{T3}=0, \beta_{gender}=0, \beta_{T3\times gender}=0\right)$			0.497			0.809			0.538
$p\left(\beta_{T3}=0, \beta_{gender}=0, \beta_{T3\times gender}=0\right)$			0.688			0.501			0.661
Pr. of paying for each category									
T1: Men	0.260		0.260	0.348		0.348	0.417		0.417
T1: Women		0.329	0.330		0.412	0.413		0.467	0.468
T2: Men	0.260		0.260	0.340		0.340	0.403		0.404
T2: Women		0.270	0.270		0.368	0.369		0.442	0.442
T3: Men	0.259		0.259	0.341		0.340	0.398		0.398
T3: Women		0.294	0.294		0.385	0.386		0.455	0.455
C : Men	0.254		0.254	0.335		0.335	0.397		0.397
C : Women		0.294	0.293		0.384	0.384		0.456	0.456

Notes: All regressions include as controls the lagged variable, fixed effects for blocks, variables for public service provision (trash collection and street lighting services during the period), the (log of the) number of properties that each taxpayer has, the (log of the) average linear front size of the properties, and a dummy that controls for those taxpayes who elected to pay monthly. Standard errors in parentheses are clustered by randomization blocks.

* p<0.10, ** p<0.05

Figures

FIGURE 1. Effects of the Deterrence Treatment

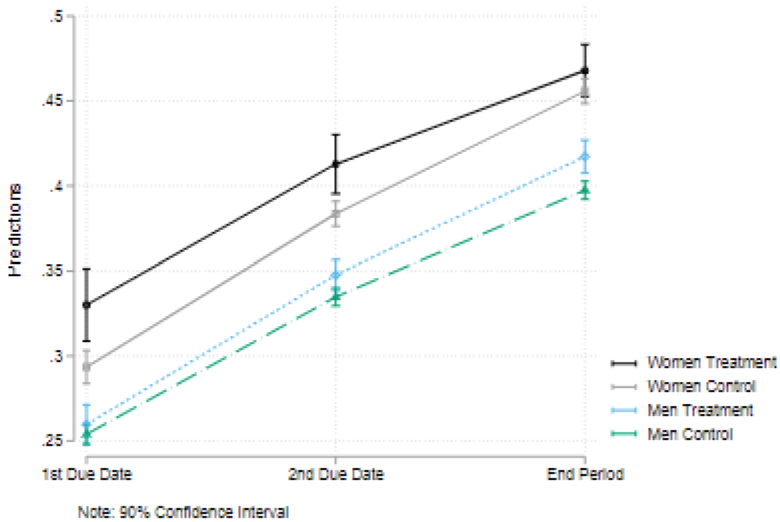
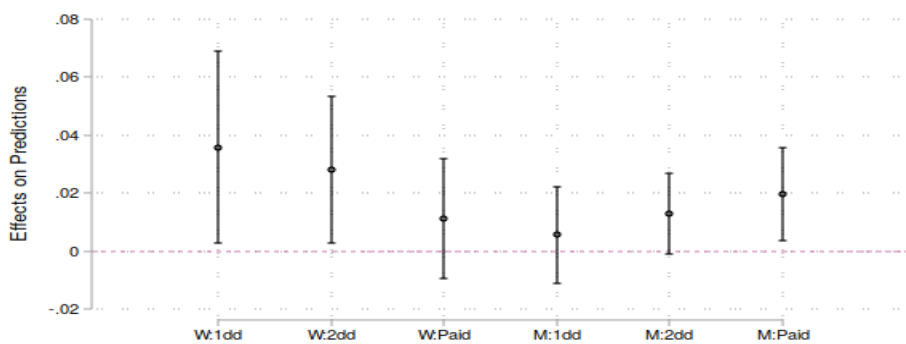


FIGURE 2. Marginal Effects of the Deterrence Treatment



Note: 95% confidence interval

W:1dd – Women 1st due date

W:2dd – Women 2nd due date

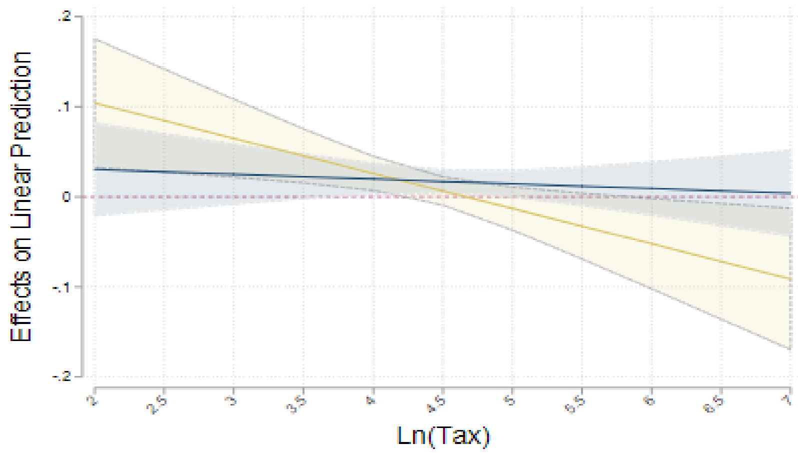
W:Paid – Women paid

M:1dd – Men 1st due date

M:2dd – Men 2nd due date

M:Paid – Men paid

FIGURE 3. Heterogeneous Effects of the Deterrence Treatment by Tax Liability and Gender



Note: Blue line corresponds to men and the yellow to women.

Appendix Tables

TABLE A1. Baseline Difference Between Women and Men (May/June)

	Diff.: Men		Women		N
Paid before 1st due date	-0.029	(0.006)	0.236	(0.005)	21,312
Paid before 2nd due date	-0.042	(0.007)	0.362	(0.006)	21,312
Paid	-0.055	(0.007)	0.441	(0.006)	21,310
Number street lights	-0.148	(0.017)	2.840	(0.014)	21,312
Paid monthly	-0.001	(0.002)	0.011	(0.001)	21,304
Log(number of properties)	0.046	(0.004)	0.783	(0.003)	21,312
Log(lineal front m)	0.031	(0.007)	2.568	(0.006)	21,312
Unrecoverable debtor	0.045	(0.007)	0.274	(0.005)	21,205
Paid all the liabilities	-0.055	(0.007)	0.441	(0.006)	21,310
At least one tax bill	0.055	(0.007)	0.541	(0.006)	21,341
Log(Mean property val.)	-0.099	(0.016)	9.151	(0.013)	21,466
Public services index	-0.142	(0.014)	0.456	(0.011)	21,312

Each row shows a regression of the pre-treatment variable in question on gender and a constant term. Observations are presented for the bimonthly period prior to treatment (May/June). The constant captures the value for the women. Unrecoverable debtors are taxpayers who have never paid their tax bill. Monetary amounts are in Argentine Pesos (AR$).
Robust standard errors are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A2. Baseline Difference Between Treatment Groups (May/June)

	Diff.: Deterrence-T1		Diff.: Reciprocity-T2		Diff.: Peer-effect-T3		Control		N
Paid before 1st due date	-0.0112	(0.0090)	0.0136	(0.0095)	-0.0034	(0.0083)	0.2126	(0.0234)	23,195
Paid before 2nd due date	-0.0142	(0.0103)	-0.0027	(0.0097)	-0.0076	(0.0104)	0.3316	(0.0350)	23,195
Paid	-0.0159	(0.0099)	0.0022	(0.0088)	-0.0039	(0.0107)	0.4037	(0.0401)	23,192
Number street lights	-0.0250	(0.0157)	-0.0116	(0.0216)	-0.0413	(0.0276)	2.7548	(0.1192)	23,195
Paid monthly	0.0020	(0.0026)	0.0055	(0.0029)	-0.0019	(0.0020)	0.0100	(0.0019)	23,186
Log(Number of Properties)	0.0052	(0.0037)	-0.0088	(0.0044)	-0.0036	(0.0057)	0.8101	(0.0117)	23,195
Log(lineal front m)	-0.0032	(0.0070)	-0.0034	(0.0109)	-0.0020	(0.0114)	2.5968	(0.0269)	23,195
Unrecoverable debtor	0.0039	(0.0092)	-0.0020	(0.0098)	-0.0085	(0.0095)	0.3041	(0.0384)	22,900
Paid all the liabilities	-0.0159	(0.0099)	0.0022	(0.0088)	-0.0039	(0.0107)	0.4037	(0.0401)	23,192
At least one tax bill	0.0121	(0.0093)	0.0059	(0.0094)	0.0017	(0.0109)	0.5763	(0.0404)	23,329
Log(mean property val.)	-0.0216	(0.0215)	-0.0024	(0.0174)	0.0101	(0.0215)	9.0990	(0.1677)	23,470
Public services index	-0.0268	(0.0132)	0.0002	(0.0166)	-0.0167	(0.0206)	0.3710	(0.1336)	23,195
Men	-0.0016	(0.0088)	0.0016	(0.0086)	-0.0194	(0.0140)	0.6589	(0.0104)	21,476

Notes: Each row shows a regression of the pretreatment variable in question on treatment and a constant term. Observations are presented for the bimonthly period prior to treatment (May/June). The constant captures the value for the control group. Unrecoverable debtors are taxpayers who have never paid their tax bill. Monetary amounts are in Argentine Pesos (AR$). Standard errors are clustered by the block level.

TABLE A3. Difference in Resources Between Male- and Female-Headed Households Urban Household Survey (EAHU–2011)

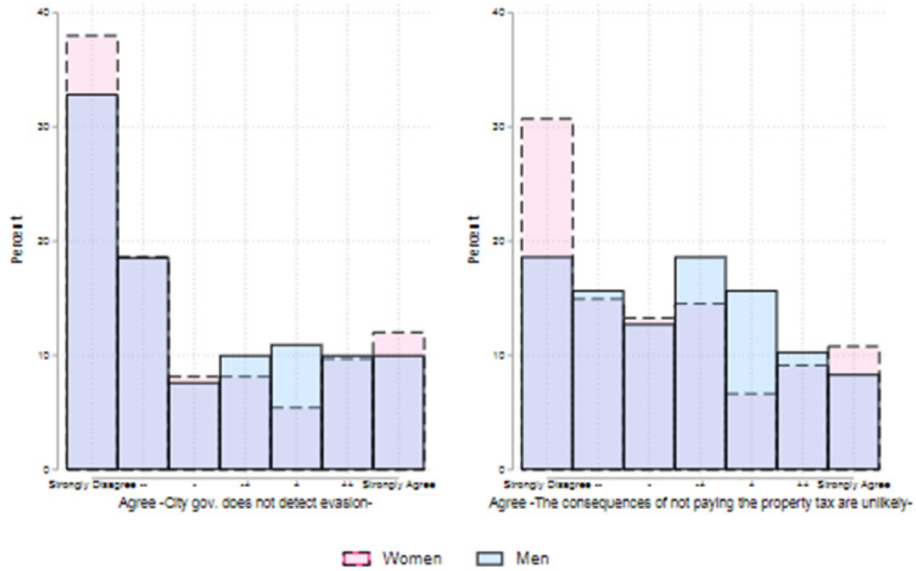
	Diff.: Men		Women		N
Num. individuals in the household	0.641***	(0.004)	2.447***	(0.004)	1,023,748
Num. members household younger 10 years old	0.149***	(0.002)	0.288***	(0.001)	1,023,748
Members household 10 years old or older	0.492***	(0.003)	2.159***	(0.003)	1,023,748
Monthly income	1,513.940***	(9.027)	4,092.824***	(7.016)	1,023,748
Monthly income per capita	215.264***	(4.468)	1,971.834***	(3.599)	1,023,748
Number of rooms in the house	0.089***	(0.003)	2.999***	(0.002)	1,023,748
Pr. receiving a wage	0.133***	(0.001)	0.673***	(0.001)	1,023,748
Num. years of education of the household head	0.274***	(0.028)	7.498***	(0.026)	990,495
Pr. spouse/parner is part of the household	0.619***	(0.001)	0.154***	(0.001)	1,023,748

Notes: Each row shows a regression of the pretreatment variable in question on gender and a constant term. The constant captures the value for the households where the head is female. Monetary amounts are in Argentine Pesos (AR$).
Robust standard errors are in parentheses.
*** p<0.01

APPENDIX Figures

FIGURE A1. Sample Tax Bills With Treatment Messages (in Spanish)

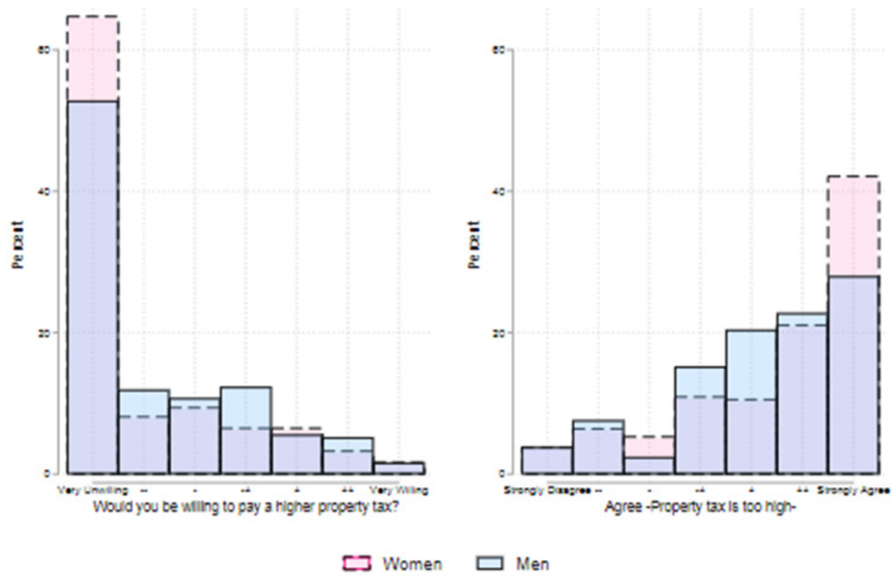
FIGURE A2. Perception of the Tax Enforcement Survey: City Government Junín



Question:
If you think of reasons people do not pay the property tax, use a scale of 1 to 7 where one means "strongly disagree" and seven means "strongly agree." How much do you agree with the following affirmations?
Left: The risk that the city government becomes aware of the property tax evasion is low.
Right: The consequences of being discovered not paying the property tax are unlikely.

The City government of Junín sponsored the survey. It targeted the individuals in charge of paying the property tax and was made on December 2015.

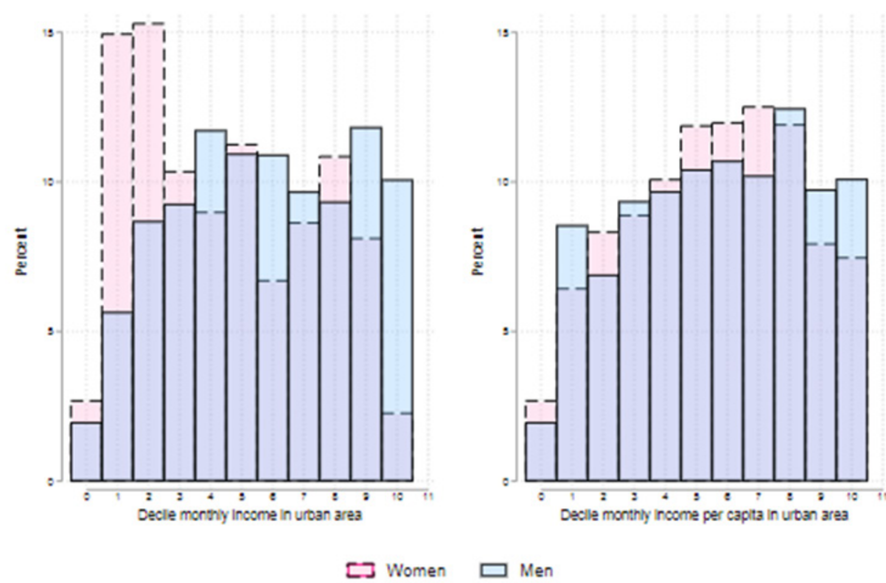
FIGURE A3. Perception of the Tax Burden Survey: City Government Junín



Question left:
Using a scale from 1 to 7, one means "very unwilling" and 7 "very willing." Would you be willing to pay higher property tax so that the services of the municipality of Junín would improve?
Question right:
If you think of reasons people do not pay the property tax, use a scale of 1 to 7 where one means "strongly disagree" and seven means "strongly agree." How much do you agree with the following affirmations?
People believe that the property tax is too high.

The City government of Junín sponsored the survey. It targeted the individuals in charge of paying the property tax and was made on December 2015.

FIGURE A4. Income Deciles Urban Household Survey (EAHU–2011)



Urban Household Survey - Third Quarter 2011
EAHU-2011

Who Sells Cryptocurrency?

Jeffrey L. Hoopes (University of North Carolina at Chapel Hill), Tyler S. Menzer (University of Iowa), and Jaron H. Wilde (University of Iowa)[1]

1. Introduction

On March 9, 2022, U.S. President Joe Biden issued Executive Order 14067, "Ensuring Responsible Development of Digital Assets" "outlining the first ever, whole-of-government approach to addressing the risks and harnessing the potential benefits of digital assets and their underlying technology" (White House (2022)). This wide-sweeping order directs or encourages digital-assets-related efforts from major U.S. agencies and regulators, including the Federal Reserve, the U.S. Treasury, the Financial Stability Oversight Council, and the Commerce Department. Notably, the Order also comes amid the recent "explosive growth" in digital assets. These assets now exceed $3 trillion (White House (2022)), and follow recent uncertainties about and disparities within and across various regulatory and policy bodies (e.g., Financial Accounting Standards Board (FASB), the U.S. Department of the Treasury, the IRS, the U.S. Securities and Exchange Commission (SEC)) regarding how to account for, tax, regulate, and oversee cryptocurrencies and cryptocurrency market places.[2] In this paper, we examine the population-level attributes of cryptocurrency reporters[3] who report their sales to the IRS.

Regulations governing financial products and activities often center on the attributes of the individuals involved. Indeed, U.S. law charges regulatory agencies to "seek the views of those who are likely to be affected, including...those who are potentially subject to such rulemaking" (Executive Order 13563, "Improving Regulation and Regulatory Review" (2011)). Thus, unsurprisingly, the nature of regulations across various financial areas often reflect the attributes of the individuals involved in particular financial activities. For example, U.S. law requires public, but not private, companies to provide audited U.S. Generally Accepted Accounting Principles (GAAP) financial statements and limits the investment vehicles that may be offered to accredited versus non-accredited investors.[4] In line with this intuition, we argue that understanding who uses cryptocurrency and how trends in cryptocurrency use are changing is essential to formulating an effective regulatory framework.

Protecting "Main Street" investors is at the heart of U.S. financial regulations. In fact, SEC Chairman Jay Clayton notes that "serving and protecting Main Street investors is my main priority at the SEC" (SEC (2018)). Current SEC regulation of products is contingent upon demographic attributes of the investors with

[1] All data work for this project involving administrative tax data was done on IRS computers, by authorized IRS personnel. In addition to being a PhD candidate at the University of Iowa, Tyler Menzer is an IRS employee under a Student Volunteer agreement through the Joint Statistical Research Program (JSRP). We thank John Guyton, Robert Hayden, and Anne Herlache of the IRS for help and guidance with this project and we thank Barry Johnson, Pat Langetieg, Alicia Miller, and Michael Weber for facilitating this project through the JSRP. We appreciate helpful comments from Andrew Belnap, Russ Hamilton, Patrick Hopkins, Stephen Lusch, Thomas Ruchti, Pradeep Sapkota, Casey Schwab, Cassie Mongold Thomas Omer, Scott Rane(discussant), Brian Williams, and workshop participants at the U.S. Treasury Office of Tax Analysis, the University of Iowa and the 2022 AAA Annual Meeting. The views expressed here are ours alone and do not reflect the views of the Internal Revenue Service.

[2] For example, the IRS issued Notice 2014-21 in 2014 detailing how cryptocurrency would be taxed as property by the IRS. The U.S. Treasury's Financial Crimes Enforcement Network (FinCEN) issued guidance treating cryptocurrencies as currency. Other market participants such as the FASB also have shown interest in updating rules for cryptocurrencies (Maurer 2022). The SEC has pursued regulatory action that presumes cryptocurrencies are securities. The picture for cryptocurrency regulation is even less clear internationally with some countries such as China, Egypt and others banning cryptocurrencies completely (Quiroz-Gutierrez 2022) while others have aimed to be cryptocurrency havens. Portugal, for instance, ruled that cryptocurrency traders are exempt from the country's 28 percent income tax in 2018 (Hall 2022).

[3] For this paper, 'cryptocurrency reporter' (uncapitalized) refers to our sample generally at the construct level; '*CRYPTOCURRENCY SELLER*' (all caps and italicized) refers to the actual taxpayers our variable captures; and 'cryptocurrency investor' is used when other sources discus individuals who invest in cryptocurrency.

[4] Prior research examines descriptive characteristics of the users of many other financial products, including the use of credit and credit cards, including age (Mathur and Moschis 1994; Limbu *et al.* 2012), student status (Limbu *et al.* 2012; Hayhoe *et al.* 2000), and race (Cohen-Cole 2011); the use of predatory lending services by race (Charron-Chénier 2020), disability (McGarity and Caplan 2019), gender (Nitani *et al.* 2020), and military status (Graves and Peterson 2005); stock market participation by gender (Almenberg and Dreber 2015), IQ (Grinblatt *et al.* 2011), geography (Brown *et al.* 2008), and age (Athreya *et al.* 2021); health insurance products by race (Monheit and Vistnes 2000; Hargraves and Hadley 2003); fintech products by gender (Chen *et al.* 2021), age (Singh *et al.* 2020; Carlin *et al.* 2017; Li *et al.* 2020), geography (Friedline and Chen 2021; Li *et al.* 2020), and race (Friedline and Chen 2021; Haupert 2022).

some products being allowed for investors perceived as more sophisticated, but not for others (e.g., 17 CFR § 230.144A–Private resales of securities to institutions).[5] Thus, understanding whether "Main Street investors" or higher income sophisticated investors predominantly use cryptocurrency is central to discussions surrounding cryptocurrency regulation. However, despite cryptocurrencies ostensible entrance into the mainstream and the major regulatory attention that now surrounds it (and other digital assets), little is known about the characteristics of individuals who own cryptocurrencies or how these characteristics have changed over time.

The inherent opacity surrounding publicly observable cryptocurrency activities accounts for the lack of evidence on who actually transacts in these digital assets.[6] While cryptocurrency publicly records the individual transactions and the unique identifiers (wallets) of the transacting parties, tying these transactions to individuals and their demographic characteristics has, ironically, proven elusive.[7] We overcome this challenge by using proprietary data from the IRS on reported sales of key cryptocurrency assets to provide the first population-level evidence of the characteristics of U.S. cryptocurrency reporters. We focus our analyses on cryptocurrency reporters who own cryptocurrency through cryptocurrency exchanges or directly on the blockchain. We analyzed that group rather than those who own cryptocurrency indirectly because cryptocurrency held indirectly, such as through public trusts or investment funds, are regulated the same as traditional securities.

Our primary objective is to contribute evidence on the characteristics of those who report cryptocurrency sales—the most common places they live, their income, age, marital and student status, and the industries in which they tend to work—and how these characteristics change over time. We find that the average income of cryptocurrency reporters has declined over time, suggesting the base of sellers has expanded in recent years, though the average cryptocurrency reporter reports higher income than taxpayers who are not associated with reported investment activity (e.g., who do not report cryptocurrency sales, sales of capital assets, or dividends). The average cryptocurrency reporter is just under 33 years old, which is considerably younger than the average non-crypto investor (about 56 years old). This age gap has grown from 2013 to 2020, even while the percentage of U.S. taxpayers reporting cryptocurrency sales has grown. Men, married individuals, college students, individuals with higher wages, individuals with more dividend income, and homeowners have become significantly more likely to sell cryptocurrency over time. We also found some evidence that workers in a broader range of professions own cryptocurrency and that cryptocurrency reporters have become more geographically diverse. Finally, we document a small but important subsection of cryptocurrency reporters who start with relatively low incomes, but within a short period of time, recognize more than $1 million in taxable gain. Interestingly in these analyses, cryptocurrency has a resorting effect—the investors who are the lowest income quartile individuals who realized large cryptocurrency gains not only recognized larger gains, on average, than those who started with more income, but after 2 years, they persist in having more taxable income for at least 2 more years.

Overall, our unique, broad-sample evidence about who sells cryptocurrency and how the attributes of cryptocurrency reporters are evolving over time provide timely, policy-relevant insights that can inform current regulatory efforts and policy deliberations.[8] Further, we contribute to a growing literature regarding many aspects of how cryptocurrencies, as an asset class, fit into our financial system by providing evidence on the dynamic characteristics of individuals investing in that asset class (Bourveau, De George, Ellahie and Macciocchi (2022); Gan *et al.* (2021); Arnosti and Weinberg (2022); Malik *et al.* (2022); Cheng *et al.* (2019); Makarov and

[5] One important decision for regulators is whether cryptocurrencies should be regulated as property or currency. Current IRS guidance states that cryptocurrency is treated as property (Notice 2014-21). However, some supporters of cryptocurrency argue it should be treated as a currency. Under the Internal Revenue Code (IRC), certain foreign currency transactions can be classified as personal transactions which eliminates gain recognition when the foreign currency gain would be less than $200. This issue has only become more important as businesses have begun to accept cryptocurrency for normal purchases. Our data shows that the median yearly gain for CRYPTOCURRENCY SELLERs is only $27, potentially providing initial evidence that the regulatory burden could be significantly reduced for taxpayers who use cryptocurrency for purchases, which may qualify for gain exclusion if cryptocurrency were treated as a currency instead of property.

[6] Due to the limitations of administrative data and tax reporting rules, we are able to observe only a subset of individuals who sell cryptocurrency and report those sales in an identifiable way through tax reporting. We discuss the specifics of this limitation in Section 3 and in the Online appendix.

[7] Even Chainalysis, a leader in tracking and identifying blockchain business users, does not provide individual level identification (https://blog.chainalysis.com/reports/service-level-data/).

[8] Apart from President Biden's Executive Order 14067 issued on March 9, 2022, the U.S. Federal Reserve released a report on a central bank digital currency in January 2022 and asked for comments from stakeholders on the proposal (Federal Reserve (2022)). The SEC has filed several lawsuits against cryptocurrency platforms such as Ripple and Blockfi (SEC (2020), SEC (2022)), and the U.S. Department of the Treasury is planning to issue new preliminary guidance (Versprille 2022).

Schoar (2020)). Finally, our analysis informs attempts to enforce taxation of cryptocurrency gains by providing evidence for profiles of the average cryptocurrency reporter.

2. Background

Bitcoin, the first cryptocurrency, is a decentralized, public, pseudo-anonymous payment network.[9] The backbone of Bitcoin is the blockchain, which serves as a public accounting ledger, maintaining the entire transaction history of Bitcoin. While a public ledger would appear to enable the linking of individuals to transactions, various issues cause this not to be the case. Prior studies have attempted to identify cryptocurrency reporters through different methods. Several papers use heuristics and machine learning to group individual wallet addresses together to identify "users" (Athey *et al.* (2016); Ron and Shamir (2013); Meiklejohn *et al.* (2013); Makarov and Schoar (2021)). These analyses are generally limited to single cryptocurrencies (such as Bitcoin) and identify only blockchain-based users, omitting users who buy or sell on certain exchanges. In addition, blockchain analyses make it difficult to distinguish between business users and individuals or to determine the geographic location of users. It also presents challenges when attempting to provide information at the user level—their income, gender, age, marital or student status, reported gambling activities, other investment sales, etc. These inherent limitations in the blockchain setting have led to dramatically different estimates of the number of unique Bitcoin users (Athey *et al.* (2016), Amiram *et al.* (2022), and Makarov and Shoar (2021)).

Other papers have taken a different approach to identifying cryptocurrency reporters. Hackethal *et al.* (2021) partnered with a German bank and received information on 100,000 investors, 872 of which invest indirectly in cryptocurrency or cryptocurrency-related investment products. They show cryptocurrency reporters were younger than non-cryptocurrency investors and had a greater level of wealth and income than non-cryptocurrency investors. Their point-in-time estimates find that cryptocurrency investors trade more frequently and hold a higher share of their investments in stocks. Similarly, Hasso *et al.* (2019), using a sample of 465,926 brokerage accounts from a U.K. brokerage, find that males between 35 and 44 years of age are most likely to trade cryptocurrency, but that females engage in less speculative trading and realize higher returns. In line with the UK brokerage's self-reported claim of being the market leader for contract-for-difference cryptocurrency trading, Hasso et al. note that 90 percent of active accounts trade cryptocurrency by 2017. They also note that investor trading patterns vary between asset classes, adding to the need for cryptocurrency specific research. While these studies provide some useful insights, it is unclear whether the results of these limited and unique samples generalize to larger populations of cryptocurrency reporters and how those characteristics have changed over time.

In addition to archival studies, several surveys highlight characteristics of cryptocurrency investors. For example, Benneton and Campiani (2021) describe three point-in-time surveys which report a wide variety of individual characteristics. The Survey of Consumer Payment Choice (SCPC) reports 1 percent of users claim cryptocurrency ownership, while the ING International Survey on Mobile Banking reports 8 percent among the subset of respondents who are familiar with cryptocurrency (65 percent). Another comprehensive survey by the Bank of Canada, the Bitcoin Omnibus Survey, done in 2016 and 2018 suggested that at the time of the survey, males were more likely to own Bitcoin (64 percent). These estimates are in line with surveys of U.S. respondents from other surveys such as the 2021 State of Crypto Literacy (64 percent) and the State of U.S. Crypto Report (74 percent). Survey evidence also suggest that cryptocurrency investors tend to be younger, consistent with the democratization of finance being a key tenant of the cryptocurrency space. However, the Bank of Canada survey suggests that both wealthier and more educated individuals own more cryptocurrency, consistent with our findings.

We add to the growing literature on owners of bitcoin by providing new evidence on how cryptocurrency reporters compare not only to other investors (similar to Hasso *et al.* (2019); Hackenthal *et al.* (2021)), but also how they compare to the non-investing public. Our unique data allows us to identify financial investment transactions and demographics. We also provide evidence on less identifiable characteristics such as wages and

[9] See Online Appendix for additional information on how Bitcoin transactions are recorded and processed.

other sources of income, geographic location, home ownership, family size, and employment characteristics. Since most individuals in the U.S. are required to file an annual tax return, our analysis also allows us to examine characteristics across the population of cryptocurrency reporters, especially in populations which may be less likely to answer survey questions, such as the very wealthy. Our large sample size also allows us to determine characteristics of cryptocurrency reporters when only a small proportion of the population engages with the technology. Finally, and importantly, the nature of our administrative data allows us to provide evidence for how the demographics of cryptocurrency reporters have trended over time.

3. Research Sample and Data

To examine demographic characteristics of those who sell cryptocurrencies, we access confidential taxpayer information from the IRS for tax returns filed between 2013 and 2020. Because tax reporting requirements do not discriminate between the different methods of holding cryptocurrency, the data should capture activity on cryptocurrency networks and trading cryptocurrency on exchanges. Although the IRS reporting would also cover cryptocurrency held indirectly by public firms, trusts, or other registered investment vehicles, those assets would already be subject to third-party reporting. We therefore do not include those assets in our calculation of cryptocurrency reporters, as they do not have control over the actual cryptocurrency. Similarly, consistent with our focus on individual taxpayers, we do not attempt to identify cryptocurrency held by businesses (Forms 1120, 1120S, and 1065) or trusts (Form 1041). To the extent that cryptocurrency transactions from flow-through entities affect individual tax returns, we will not identify those transactions.

While our sample is the largest time-series sample of cryptocurrency reporters to date, it also has its limitations. In particular, because the U.S. tax system relies upon the realization principle, the tax return reveals the most reliable information about cryptocurrency reporters only when they sell cryptocurrencies and report those transactions. While knowing everyone who buys and holds cryptocurrency would be useful, IRS data—potentially the best dataset available to answer these questions—is nonetheless imperfect. Using tax return data to study capital assets is especially problematic in equity markets, where investors must trade off their beliefs about expected returns and the value of tax deferral (Lei *et al.* (2020)). However, cryptocurrencies are unique in that this problem is partially mitigated by a key feature of the tax system. Investors sell and instantly repurchase cryptocurrencies to take advantage of tax losses (Cong *et al.* (2021)). In addition, for Tax Years 2019 and 2020, taxpayers must have checked a box on their tax returns stating whether they engaged in a variety of virtual currency transactions. Consequently, noting who sells should be a much better indicator of who owns cryptocurrencies than it would be for equities, especially given the volatility in the crypto market.

In addition, while there are severe financial penalties for tax noncompliance, some individuals invariably fail to report all their cryptocurrency transactions to the IRS. This problem is prevalent in much of the accounting literature, in which evidence of behavior is only observed contingent on the reporting or detection of such behavior (Cecchini *et al.* (2010); Hopkins *et al.* (2015)). In some cases described below, tax law requires third parties to report transaction level tax-related information to the IRS, enhancing the quality of the tax data. But to the extent that underreporting varies with the investor characteristics we study, our estimates may not reflect the true population of cryptocurrency reporters.

We obtain cryptocurrency sales data from two IRS forms: Form 8949 and Form 1099-B. Unless certain third-party reporting requirements are met, tax provisions require individuals to report individual stock transactions on Form 8949, including a description of the property, the date of purchase, date of sale, cost of property, sales price, and any adjustments.[10] We use a textual search to identify transactions that are likely to be cryptocurrency.[11] We focus on two types of cryptocurrencies, Bitcoin and Ethereum, which are, by far, the two most valuable, most widely held, and, well-known cryptocurrencies (2021 State of Crypto Report, Yougov

[10] Taxpayers are allowed to summarize transactions if gains are reported on Form 1099-B, with basis reported, and for which they have no adjustments.

[11] We note that while we use IRS data to perform our analyses, the process we use to identify them was designed and implemented by the authors of this research and does not represent the method the IRS may use to identify cryptocurrency transactions.

(2018)).[12] However, our textual analysis likely identifies other cryptocurrencies as well, especially cryptocurrencies with similar names such as Ethereum Classic or Bitcoin Cash.[13] After completing the textual search, we manually inspect a random sample of 3,000 cryptocurrency transactions in each of our sample years (i.e., 24,000 transactions) to assess the possibility of false positives. Overall, we find a false positive rate of 1.2 percent. This false positive rate was highest in 2013 (3.4 percent) and dropped over time to less than 1 percent for years after 2015.[14] To supplement our data from Form 8949, we also use the same method to search third-party reported descriptions filed on Form 1099-B.[15] Using form 1099-B filings allows us to identify some transactions that taxpayers may have summarized on their tax returns.

After identifying Bitcoin and Ethereum sales, we merge the Form 8949 data with individual taxpayer data from Form 1040 and its related schedules. We begin with 1,223,732,729 taxpayer-year records who have valid taxpayer identification numbers.[16,17] We restrict our analysis to electronically filed returns so we can capture all the fields we require for our analysis (reduction of 112,487,851 observations). This process yields a sample of 1,078,688,472 taxpayer-years (202,523,891 unique taxpayers), including 2,620,921 *CRYPTOCURRENCY SELLER* tax-year observations. We also merge in IRS data sourced from the Social Security Administration on the birth year and gender of taxpayers, as well as additional data from third-party reporting.

4. Results

4.1 Demographic Information on Crypto Sellers

We report general descriptive statistics for our sample in Table 1 and separate our sample into three groups, *NON-INVESTOR* (taxpayers with no capital asset sales, and no dividends), *NON-CRYPTO SELLING INVESTOR* (taxpayers with capital asset sales/dividends but no crypto sales), and *CRYPTOCURRENCY SELLER*. As in prior studies, we find that cryptocurrency reporters are younger with a mean age of 32.8 compared to 41.5 for non-investors and 56.3 for non-cryptocurrency investors. Consistent with the sentiment that cryptocurrency supports the "democratization of finance," we find that sellers have less income than other non-cryptocurrency investors, albeit more income than non-investors. We also find that they have less investment income (e.g., dividends, interest, and capital gains) and wages than non-crypto investors. Both the number of cryptocurrency transactions reported, and the yearly cryptocurrency gain is highly skewed, with the median reporting only one transaction while the average is 9.9.[18] The average reporter has a cryptocurrency gain of $12,484 per year, although the median is only $27. The average yearly cryptocurrency gain for the top 100 *CRYPTOCURRENCY SELLERs* ranges from almost $560 thousand to over $18 million, indicating that there are some taxpayers realizing and reporting very large cryptocurrency gains. We examine these taxpayers further in Section 4.2. Additionally, the average yearly cryptocurrency losses for the 100 *CRYPTOCURRENCY SELLERs* with the greatest losses ranges between losses of $115,000 and $5.7 million. In both cases, the largest

[12] The term "Ethereum" can refer to the cryptocurrency "Ether" and the blockchain platform on which Ether runs. In this paper, we use Ethereum to refer to the cryptocurrency rather than the blockchain network. We use this terminology for several reasons. First, major cryptocurrency platforms such as Coinbase, Kraken, and Binance all refer to the cryptocurrency Ether as Ethereum when listing it on their exchanges. Thus, many investors likely think of the term Ethereum as a cryptocurrency. Second, prior literature uses the term Ethereum to refer to the cryptocurrency (see Marakov and Shoar 2020 and Giffin and Shams 2020). Finally, popular news organizations, such as Coindesk, use both Ether and Ethereum to refer to the cryptocurrency interchangeably (e.g., https://cointelegraph.com/news/phishing-scammer-monkey-drainer-has-pilfered-as-much-as-1m-in-ethereum).

[13] As of March 21, 2022, Bitcoin is the largest cryptocurrency with a market cap of $889 billion while Ethereum has a market cap of $395 billion. The next largest currency is Tether, which has a market cap of $81 billion (https://coinmarketcap.com/historical/20220327/).

[14] We inspect individual transaction descriptions rather than tax returns. If a taxpayer reports multiple cryptocurrency transactions, we could still classify them correctly as a cryptocurrency seller even if one or more of the transactions that we identify are false positives.

[15] To identify cryptocurrency transactions reported on Form 1099-B, we follow the same process we use for Form 8949, with one exception. To avoid classifying cryptocurrency ETFs and related products as transactions related to direct interests in cryptocurrency, we remove transactions for which there is a valid CUSIP reported on the Form 1099-B. Notably, many Form 1099-B transactions report the CUSIP for the security being reported but cryptocurrencies are not regulated securities and thus do not have valid CUSIPs. .

[16] Each year contains between 124,222,137 (2013) and 148,493,792(2019) unique taxpayers.

[17] Because we are interested specifically in the reporting behavior of cryptocurrency owners in the reporting environment of the time, we restrict our sample to the first tax return filed by a taxpayer each year, and we remove tax returns filed more than 1 year after the close of the tax year (32,541,050). We also remove returns for which there are duplicate records filed at the same time (15,356).

[18] We caution interpretation of the number of transactions reported as taxpayers can and do often group transactions together or summarize them. To the extent cryptocurrency transactions are grouped together, it should bias the estimate downward.

gains and losses generally occur in the latter half of our sample. In addition, the median *CRYPTOCURRENCY SELLER* has no other non-gain investment income (*e.g.*, dividends or interest), also consistent with cryptocurrency investors being more like non-investors.

We graph several tax return characteristics in Figure 1. We find that *CRYPTOCURRENCY SELLERs* are much more likely to be enrolled in a university or college (*STUDENT)* than both other groups. In line with the lower income of *CRYPTOCURRENCY SELLERs*, we also find that they are more likely to claim the Earned Income Tax Credit (*EIC TAX CREDIT)* than other investors, but less likely to claim it compared with non-investors. We also look at taxpayer risk preferences by examining how likely *CRYPTOCURRENCY SELLERs* are to have reported gambling income (*GAMBLER)* and find that a similar proportion of *CRYPTOCURRENCY SELLERs* have gambling income compared with non-investors or other non-crypto investors (crypto-investors actually have slightly lower gambling income than non-crypto investors). Moreover, we consider the role of financial health and cryptocurrency sales by examining the percentage of *CRYPTOCURRENCY SELLERs* that receive cancellation of debt income (*CANCELLATION OF DEBT)* and find that *CRYPTOCURRENCY SELLERs* are similar in that respect compared with non-investors.

We graph the number of CRYPTOCURRENCY SELLERs over time in Figure 2, Panel A. The number of CRYPTOCURRENCY SELLERs is increasing dramatically over time, with less than 7,000 taxpayers reporting cryptocurrency per year between 2013 and 2016, and over 120,000 sellers in 2017. This large increase coincides with the price increase of Bitcoin in 2017 and the associated hype, broad media coverage, and surge in public interest. While Bitcoin started 2017 being valued around $1,000, it reached a high of nearly $20,000 before falling in 2018 (Higgins (2017)). Notably, we also observe a large increase in the number of CRYPTOCURRENCY SELLERs in each subsequent year. This pattern is consistent with survey evidence, which found that in 2020, 26 percent of users had acquired their cryptocurrency within the last year, 68 percent had acquired it within the last 2 years, and only 4 percent had acquired their cryptocurrency over 5 years ago (State of Crypto Report (2021)).[19]

We next examine how *CRYPTOCURRENCY SELLERs* have changed over time. In Figure 2, Panel B we directly compare *CRYPTOCURRENCY SELLERS* to *NON-CRYPTO SELLING INVESTORS.* The mean (standard deviation) age of *CRYPTOCURRENCY SELLERs* has markedly decreased over time, from 45.2 (18.6) in 2013 to 32.4 (10.9) in 2020. Over the same time, the average age of *NON-INVESTOR* (untabulated) [*NON-CRYPTO SELLING INVESTOR*] has remained relatively flat, from 41.1 (17.1) to 42.4 (17.8) [55.8 (19.6) to 55.4 (19.6)]. We also note that the average taxable income of *CRYPTOCURRENCY SELLERs* decreased over time, from an average of $299,217 in 2013 to only $76,147 in 2020. This change is particularly interesting because in the early part of our sample period (before 2018), *CRYPTOCURRENCY SELLERs* had more income than *NON-CRYTPO SELLING INVESTORs*, and this trend held even at the median. In 2017, the median taxable income for *CRYPTOCURRENCY SELLERs* was over $99,000, while the median *NON-CRYTPO SELLING INVESTOR* had a median income of only $70,000. By 2020, however, the median *NON-CRYPTO SELLING INVESTOR* had median income of $68,000, while the median *CRYPTOCURRENCY SELLER's* taxable income was only $32,000 (untabulated).

Noting the lower income of *CRYPTOCURRENCY SELLERS* over time, we examine the distribution of taxable income for these taxpayers in further detail in Figure 3. We produce a histogram of *TAXABLE INCOME* for *CRYPTOCURRENCY SELLERs* over our sample period, using $10,000 width bins. For ease of interpretation and due to the extreme skewness in *TAXABLE INCOME*, we limit the upper bound of the histogram to $270,000, which is approximately equal to the 95th percentile for cryptocurrency returns. Consistent with *CRYPTOCURRENCY SELLERs* reporting lower income, we find that 27.4 percent of these sellers have under $10,000 of taxable income, and over half of sellers report less than $40,000 in taxable income. The low income of these taxpayers is also unlikely due to excessive deductions, as only 13.1 percent of *CRYPTOCURRENCY SELLERs* file Schedule A for itemized deductions.

[19] The increase may also speak to increased compliance with tax laws. Regulatory factors such as the IRS John Doe Summons of a large cryptocurrency exchange and the resulting increase in third-party reporting may have resulted in increased regulatory scrutiny and compliance with tax reporting requirements. Consistent with this assumption, we find that the average number of CRYPTOCURRENCY SELLERS who receive a Form 1099-B for cryptocurrency increases from 6 percent in 2016 to 84 percent in 2020 (untabulated).

4.2 Cryptocurrency Millionaires

One area of interest related to cryptocurrencies is their ability to produce immense wealth as a result of the exponential growth in asset prices. This growth has created a rags to riches story for many early investors (Schlott (2022)). To examine this phenomenon further, we look specifically at individual taxpayers who report large cryptocurrency gains—the cryptocurrency millionaires. To begin, we partition taxpayers into five categories. Here, we first sum both *CAPITAL GAIN/LOSS*es and *CRYPTOCURRENCY GAIN*s by taxpayer for all years in our sample period. We label taxpayers who have a total *CRYPTOCURRENCY GAIN* greater than $1 million as *CRYPTOCURRENCY MILLIONAIRES*. Then, to calculate a taxpayer's total gain from traditional equities, we subtract a taxpayer's total cryptocurrency gain from the total capital gain reported on their tax return. We identify taxpayers who reported over $1 million of non-cryptocurrency capital gains as *EQUITY MILLIONAIRES*. We restrict *EQUITY MILLIONAIRES* to the group of individuals who report non-cryptocurrency *CAPITAL GAIN/LOSS* above $1 million, but *CRYPTOCURRENCY GAIN* less than $1 million. If a taxpayer's *CAPITAL GAIN/LOSS* and *CRYPTOCURRENCY GAIN* each exceed $1 million, we include them among the *CRYPTOCURRENCY MILLIONAIRES*. Finally, we treat all other taxpayers who do not fall into those two categories as we do in Table 1 (*e.g.*, *NON-INVESTOR*, *NON-CRYPTO INVESTOR*, *CRYPTOCURRENCY SELLER*), except that these categories now exclude observations relating to *CRYPTOCURRENCY MILLIONAIRES* and *EQUITY MILLIONAIRES*.

We report descriptive statistics for these groups of taxpayer-years in Table 2. When we compare *CRYPTOCURRENCY MILLIONAIRES* to *EQUITY MILLIONAIRES*, we note that both groups have higher incomes than the other groups, consistent with these groups being associated with higher wealth (income) than non-millionaire groups. *CRYPTOCURRENCY MILLIONAIRES* also report higher income, on average, than *EQUITY MILLIONAIRES*. Rather than the image of rags to riches, this pattern suggests that these individuals were already wealthy individuals. For example, the average wage income of *CRYPTOCURRENCY MILLIONAIRES* is $366,092 while the average wage income for *NON-CRYPTO SELLING INVESTORs* is only $86,215. In addition, it appears that *CRYPTOCURRENCY MILLIONAIRES* also report more cryptocurrency transactions than non-millionaire *CRYPTOCURRENCY SELLERs* (54.57 and 9.57, respectively). Taxpayers with at least $1 million of cryptocurrency gain are also less likely to receive a cryptocurrency Form 1099-B, which may indicate that these transactions were on-chain or private cryptocurrency transactions.

However, while *CRYPTOCURRENCY MILLIONAIRES* are on average wealthy to begin with, this masks large heterogeneity in initial incomes. To further examine the effects of large cryptocurrency gains, we examine a different set of taxpayers. We identify taxpayers who had a single tax year with a cryptocurrency gain of $1 million or more. Then, we determine the first year in which each of these taxpayers reported a cryptocurrency gain of at least $1 million and graph the taxable income of these investors in event time, with period being the first year the taxpayer had $1 million or more in cryptocurrency gains. We then divide taxpayers into quartiles based on their total taxable income in , allowing us to see the trend in income, conditioning on prior income, thus allowing us to explore whether these were, indeed, rags to riches stories, on average.

We report the average taxable income of each quartile over time in Figure 4. We see that the large cryptocurrency gain is a large shock to income for all quartiles. The highest-income quartile's taxable income appears to return to pre-cryptocurrency gain levels within 2 years, by $year_{(t+2)}$, with average income going from $7.19 million in to $year_{(t-2)}$ $6.69 million in $year_{(t+2)}$. However, for each of the bottom three quartiles of income, taxpayers in all three groups appear to report considerably higher income in $year_{(t+2)}$ than in $year_{(t-2)}$. The lowest quartile of income (as of $year_{(t-2)}$) exhibits the largest difference, with average taxable income starting at only $1,666 in $year_{(t-2)}$ and ending at $2,189,009 by in $year_{(t+2)}$. Although relatively few individuals have cryptocurrency gains over $1 million in any single year, this analysis provides evidence that at least some low-income taxpayers appear to experience potentially life-changing levels of income via cryptocurrency investments. Further, we note that individuals in the lowest quartile of income in *t-2* actually end up with the second highest income in *t+2*, suggesting that cryptocurrency gains do have, at least for a small section of the population, the ability to reorder income strata in meaningful ways.

4.3 Geographic Location of Cryptocurrency Reporters

We next turn our focus to the geographic location of *CRYPTOCURRENCY SELLER*s. In Figure 5, we map the ratio of *CRYPTOCURRENCY SELLER* tax returns to total number of tax returns by county for the continental U.S. for even numbered years. In the early sample years (2014 and 2016), we saw very few counties have *CRYPTOCURRENCY SELLER*s, with many counties having no *CRYPTOCURRENCY SELLER*s at all.[20] In 2018, we observed a much broader adoption across the U.S., suggesting that cryptocurrency was becoming more geographically widespread. Notably, some states appear to still have low cryptocurrency reporting rates even in 2020. West Virginia, which was rated 5th on a list of the "worst" states for cryptocurrency investors in 2022 (Newberry (2022)) and had the lowest search interest in Bitcoin in 2020 out of all 50 states (Google Trends analysis, untabulated), appears to have a relatively low incidence of *CRYPTOCURRENCY SELLER*s. New Hampshire also appears to have low cryptocurrency reporting and has below-average Google Trends search volume (rank 41) for 2020. However, somewhat puzzling is the relatively low cryptocurrency taxpayer reporting rates in Nevada, which had the highest Google Trend for Bitcoin out of all 50 states in 2020.

We next move to more granular data on location to examine cryptocurrency reporting at the city level. In Table 3, we report the top 10 cities with the highest *CRYPTOCURRENCY SELLER* ratio, as well as the 10 cities with the highest raw numbers of *CRYPTOCURRENCY SELLER*s. Overall, we see that California has some of the highest ratios of *CRYPTOCURRENCY SELLER*s throughout our sample period, with 5 out of 10 of the top cities being in California in 2014 and 8 out of 10 in 2020. This seems to indicate that these cryptocurrency "capitals" have maintained their positions throughout our sample period, and the West Coast continues to be the area with the highest concentrations of *CRYPTOCURRENCY SELLERS.* Examining the raw number of *CRYPTOCURRENCY SELLER*s without regard to population size is also insightful. We continue to find that more *CRYPTOCURRENCY SELLER*s live on either the West or East Coasts, with only four non-coastal cities in 2014 and only three non-coastal cities in 2020. Although there might be concern that population drives these results, we note that several of the top 10 largest cities in the U.S., such as Philadelphia, Phoenix, and San Antonio do not appear on the list. We conclude that although cryptocurrency has achieved a much wider adoption over the 8-year period of our sample, there is still significant geographic clustering of cryptocurrency reporters.

4.4 Cryptocurrency and Occupation

We next examine the occupations (industries of employment) of *CRYPTOCURRENCY SELLER*s using wages and Form W-2 information. We obtain the population of W-2 data for our sample years, which reports wage income and use the W-2 with the highest reported income each year.[21] Next, we identify the three-digit NAICS code based on the business tax return that filed the Form W-2. Similar to our geographic analysis, Table 4 reports the top 10 industry codes for 2014 and 2020 for both the ratio of sellers to total taxpayers and raw number of sellers.

In Panel A, we report the ratio of taxpayers who are *CRYPTOCURRENCY SELLERs* to the total number of taxpayers whose highest paid W-2 is in the given industry. We find that the highest ratio of *CRYPTOCURRENCY SELLER*s generally falls into more technology- or finance-related industries. Publishing- and news-related industries also make up a large portion of the top industries. We also see that even among the highest ratio industries, the ratio has increased over the sample period. For example, Other Information Services has increased from 0.04 percent of taxpayers in the industry reporting cryptocurrency sales in 2014 to 3.29 percent reporting sales in 2020. We also see that even the 10th highest ratio (Information, 1.65 percent) in 2020 is higher than all other industries in 2014, highlighting the growth in cryptocurrency adoption. We observe some changes in *CRYPTOCURRENCY SELLER* industry ranks over the sample period, with more retail industries (NAICS-3 454 and 443) in 2020 than in 2014, and two industries in the top 10 in 2014, Museums, Historical Sites, and Similar Institutions and Motion Picture and Sound Recording Industries, dropped off the list by

[20] Due to restrictions on IRS data and bias in small counties, we set any county with less than 10 cryptocurrency reporters or less than 1,000 tax returns to 0.

[21] For this test, if a tax return is filed as "Married Filing Joint" we identify the highest paid job for both the taxpayer and spouse for each year. If a "Married Filing Joint" tax return is a CRYPTOCURRENCY SELLER, we assume both spouses are CRYPTOCURRENCY SELLERs.

2020. We also present the information graphically in Figure 6, Panel A, along with data for the years 2016 and 2018. We observe that the general shift in the top industries happens between 2016 and 2018, which coincides with the large increase in the overall number of *CRYPTOCURRENCY SELLER*s.

When we analyze the raw number of *CRYPTOCURRENCY SELLER*s per industry in Panel B, we see more of a shift over time. In 2014, half of the top 10 highest ratio and highest raw counts are the same, such as Professional, Scientific and Technical Services, Other Information Services, Publishing Industries, and Computer and Electronic Manufacturing. However, toward the end of the sample period, we see the top industries with the most *CRYPTOCURRENCY SELLER*s are industries in which we would expect a large number of *CRYPTOCURRENCY SELLER*s simply because they are some of the largest industries (e.g., Food Services and Drinking Places, Educational Services, or Food and Beverage Stores). In fact, none of the industries with the largest ratios are included in the top 10 list by number of sellers by 2020. This change over time lends evidence to the broader adoption of cryptocurrency from a more niche investment to an asset with a much broader appeal and wider reach. Similar to the percentage rank, we also find that the majority of the change in the top industries happens between 2016 and 2018 as can be seen in Figure 6, Panel B.

4.5 Regression Analysis

We conclude our analysis with a model of the determinants of cryptocurrency reporting. To assess the determinants of being a *CRYPTOCURRENCY SELLERS* we estimate the following equation on a tax return-by-year basis:

$$\begin{aligned} &CRYPTOCURRENCY\ SELLERS_{it} = \alpha + \beta_1 AGE(Under\ 24)_{it} + \beta_2 AGE(25\text{-}44)_{it} + \beta_3 AGE(45\text{-}64)_{it} + \\ &\beta_4 LN\ WAGES_{it} + \beta_5 LN\ DIVIDENDS_{it} + \beta_6 MARRIED_{it} + \beta_7 SINGLE\ MALE_{it} + \beta_8 HOMEOWNER_{it} \\ &+ \beta_9 DEPENDENTS_{it} + \beta_{10} STUDENT_{it} + \delta_t + \varepsilon_{it} \qquad (1) \end{aligned}$$

We include 3 indicator variables for various age groups, with individuals greater than 65 being the base group. We include both the natural log of *WAGES* and natural log of *DIVIDENDS* to capture potentially different effects of labor income versus capital income.[22] We include the indicator variables *MARRIED* and *SINGLE MALE* to capture the effects of gender and tax reporting status. Because our observations are primarily at the tax return level, and not the taxpayer level, we do not attempt to allocate income or expenses between spouses, which is why for married couples we do not indicate a gender. We include *HOMEOWNER* to capture potentially differing asset or net worth values. We include *DEPENDENTS* to capture whether taxpayers with children have different investments. We include *STUDENT* to capture potentially differing socio-economic status and education level. Finally, δ_t reflects our year fixed effects to help control for the significant time trends in cryptocurrency reporting.

Due to the size of our data set (starting with well over a billion observations), we are unable to run a regression analysis on our full sample. To address this issue, we begin by taking a random sample of 10 million tax returns from the population of tax returns that have data available for the regressions.[23] We repeat each random sampling process 10 times and average the coefficients, standard errors, and adjusted R-squared from the models to report in Table 5. We also report the number of coefficients (out of 10) that are significant at the 1 percent level. Column 1 reports the results of estimating Eq. (1). As mentioned earlier, the overall probability of being a CRYPTOCURRENCY SELLERs is low, only a fraction of a percent. To aid in the interpretation of coefficient magnitude, we also report the overall probability of selling cryptocurrency for the full sample from which each random regression sample is chosen. Our objective with Model (1) and later models is to examine whether certain types of people are more likely to use cryptocurrency than other types of people, rather than attempt to develop a prediction model of cryptocurrency use. The explanatory power of our models is very low (with an R-squared generally below 1 percent). This pattern suggests that other factors not reflected in our tax

[22] We specifically avoid using TAXABLE INCOME due to the fact that cryptocurrency gains are a part of TAXABLE INCOME and we avoid CAPITAL GAIN/LOSS for the same reason. We note that if we were to try to remove cryptocurrency gains from income mechanically by subtracting them out, the variable would lose interpretability if there were other losses included on the return, as IRS rules do not allow TAXABLE INCOME to go below 0 or CAPITAL GAIN/LOSS to go below -3,000.

[23] In untabulated analysis described in the online appendix, we find that the random sampling process does a good job of maintaining the attributes of the full sample in the random sample.

return data, such as personal connections, investment advisors, technological aptitude, illegal behavior, or risk preferences, could better explain the variation in cryptocurrency use.

In line with prior survey evidence, we find that a significant predictor of cryptocurrency use is age, with *AGE(UNDER 24) (AGE(25-45))* tax returns having a probability 58.0 percent (82.3 percent)[24] above the percentage of cryptocurrency reporters in the entire population. In fact, all age groups under age 65+ are associated positively with selling cryptocurrency. We also find that a *SINGLE MALE* tax return has a 49 percent probability above the baseline. Being MARRIED is also positively associated with selling. While these associations are consistent with the univariate statistics discussed earlier, observing them in a regression framework allows us to understand these associations conditional on the other variables in the model.

Both measures of income are positively associated with owning cryptocurrency, but that the effect for capital income is larger, with a 1 percent increase in TAXABLE DIVIDENDS being 5.8 times the effect of a 1 percent increase in W*AGES*. We find that *HOMEOWNER* is positively associated with the probability of reporting cryptocurrency sales and the association holds even when conditioning on income and marital status, while the coefficient on the number of *DEPENDENTS* is negative, indicating that as taxpayers have more children, they are less likely to sell cryptocurrency. We also find that being enrolled in higher education (*STUDENT*) is also highly positively associated with being a *CRYPTOCURRENCY SELLER* with a probability 56.4 percent above the baseline, even after controlling for age.

We next examine how the associations between particular attributes and cryptocurrency sales have changed over time. To facilitate this comparison, we construct a trend variable, which equals 0 starting in 2013, 1 in 2014, and so forth (TREND). We then interact TREND with all variables from Eq. (1). Specifically, we estimate the following model:

$$\begin{aligned}&CRYPTOCURRENCY\ SELLER_{it} = \alpha + \beta_1 AGE(Under\ 24)_{it} + \beta_2 AGE(25\text{-}44)_{it} + \beta_3 AGE(45\text{-}64)_{it} + \\ &\beta_4 LN\ WAGES_{it} + \beta_5 LN\ DIVIDENDS_{it} + \beta_6 MARRIED_{it} + \beta_7 SINGLE\ MALE_{it} + \beta_8 HOMEOWNER_{it} + \\ &\beta_9 DEPENDENTS_{it} + \beta_{10} STUDENT_{it} + \beta_{11} TREND + \beta_{12} AGE(Under\ 24)_{it}{*}TREND + \beta_{13} GE(25\text{-}44)_{it}{*}TREND \\ &+ \beta_{14} AGE(45\text{-}64)_{it}{*}\ TREND + \beta_{15} LN\ WAGES_{it}{*}\ TREND + \beta_{14} LN\ DIVIDENDS_{it}{*}\ TREND + \beta_{16} MARRIED_{it}{*} \\ &TREND + \beta_{17} SINGLE\ MALE_{it}{*}\ TREND + \beta_{18} HOMEOWNER_{it}{*}\ TREND + \beta_{19} DEPENDENTS_{it}{*}\ TREND \\ &+ \beta_{20} STUDENT_{it}{*}\ TREND + \varepsilon_{it} \qquad (2)\end{aligned}$$

Table 5, Column 2 reports the result of this analysis. We find that the main effect for most variables is of the opposite sign as in column (1). We note that the coefficients on the interaction terms for both AGE(UNDER 24)*TREND and AGE(25-44)*TREND are positive, suggesting that CRYPTOCURRENCY SELLERs are indeed getting younger over time. The trend for both SINGLE MALE and MARRIED are also positive. Finally, we find that owning a home (having more dependents) is positively (negatively) associated with the probability of reporting a cryptocurrency sale over time. We also find a positive and large coefficient for STUDENT*TREND. Overall, we interpret the evidence to suggest that although the number of CRYPTOCURRENCY SELLERs has increased dramatically in recent years, such sellers continue to be different from the general population of taxpayers.[25]

We estimate our regressions on a sampling basis. Here, we describe some tests we performed to validate the sampling methodology we use in our regressions that are necessary because of the constraints in research computing power. In the full sample, 0.243 percent of the sample tax returns are CRYPTOCURRENCY SELLERs, while testing our random sample selection process results in between 0.242-0.246 percent (average 0.243 percent) of CRYPTOCURRENCY SELLERs. We find similar results when looking at the proportion of our sample that are NON-INVESTORS, 78.34 percent on average in our random samples, 78.35 percent in the full sample, and NON-CRYPTO SELLING INVESTORs, which make up 21.41 percent on average in our random samples and 21.41 percent in our full sample. On average each random sample contains approximately 21,900 CRYPTOCURRENCY SELLERs. These CRYPTOCURRENCY SELLERs also appear to be similar to the full population. For example, the average (median) wages reported by CRYPTOCURRENCY SELLERs in our ran-

[24] Calculated as (0.00384-0.00243)/0.00243 = .5804.

[25] In analysis described and tabulated in the online appendix, we estimate this regression during different parts of our sample period. While the direction of the results are similar to those reported here, some magnitudes do change.

dom samples is $77,039 ($46,036) while in our full sample the same statistics are $77,049 ($46,010). We note that the standard deviation of our random samples is typically smaller than the full population. For example, the standard deviation for WAGE INCOME for the full population is $355,060, while the average standard deviation for our random samples is only $209,624, with only one out of the 10 random samples having a standard deviation larger than the full sample. This pattern is likely due to the fact that the extreme observations on the right tail of the distribution have a very low probability of being selected for a given random sample. Thus, our models may not model the extreme end of the distributions very well. In untabulated robustness checks, we find that keeping the full sample of CRYPTOCURRENCY SELLERs observations and selecting a random control sample results in generally consistent inferences, although coefficient size does vary with the proportion of CRYPTOCURRENCY SELLERs to control observations.

4.6 Additional Analysis

We conduct two cross-sectional splits to further examine the attributes of CRYPTOCURRENCY SELLERs. First, we separately examine the two subsamples of our sample period based on tax year. Given the extreme discontinuous jump in CRYPTOCURRENCY SELLERs in 2017, when the number of reporters went from under 7,000 in 2016 to over 120,000 in 2017, we split our sample in half at this point. We report the results of these tests in Table 6, columns (1) and (2), where we partition our sample of 10 million tax returns into two subsamples, one for the 2013–2016 period and the other for the 2017–2020 period. Although the signs of the coefficients are consistent across all model variables in both regressions, there are several notable differences in magnitudes.

The coefficient on *AGE(25–44)* in column (1) is approximately the same magnitude as column (2) when scaled by the baseline probability (1.86 vs. 1.79). However, the youngest group of tax payers (*AGE (UNDER 24)*) are more likely to sell cryptocurrency in the later period (baseline adjusted of 1.06 to 1.57). The results suggest that this trend is reversed for *AGE(44-64) CRYPTOCURRENCY SELLER*s, who are more likely to sell cryptocurrency in the early period (0.80 to 0.39). We also find that investment income (*LN DIVIDENDS)* has a larger effect in the early period (0.53 to 0.13), while the coefficients on wage income (*LN WAGE)*, *HOMEOWNER*, and *STUDENT* are generally not significantly different from zero in the early period. The coefficients on the indicator variables for gender and marital status exhibit a stronger relation in the late period, but are large in both periods when scaled by the baseline percent. The effect for *DEPENDENTS* appears to weaken from the early to late period (-0.27 versus -0.12), suggesting that more taxpayers claiming dependents own cryptocurrency in later years.

We next examine how cryptocurrency reporters are different from NON-INVESTORS as opposed to NON-CRYPTO SELLING INVESTORs. We view this analysis as an important distinction of our study over prior work. Whereas prior studies compare cryptocurrency investors to other investors when using investment data (e.g., Hackethal, et al. (2021); Hasso, et al. (2019)), we compare cryptocurrency investors to a non-investing baseline separate from other investors. We present the results of these tests in Online Appendix Table 1, columns (3) and (4). When comparing CRYPTOCURRENCY SELLERs to other tax returns with investments, we find the largest predictors are AGE(UNDER 24) (8.49 times the baseline), STUDENT (8.45 times the baseline), AGE(25-44) (6.30 times the baseline), and SINGLE MALE (6.09 times the baseline). We also observe that CRYPTOCURRENCY SELLERs tend to have less dividend income than other investors and are less likely to own their home than other investors. We also examine CRYPTOCURRENCY SELLERs compared with the general NON-INVESTOR tax returns. While closer in age to non-investors, we find that CRYPTOCURRENCY SELLERs are still younger on average, more likely to be married, more likely to be male, and more likely to be a student. The large coefficient on LN DIVIDENDS is consistent with NON-INVESTORS having no other investment income, by definition. These results provide initial evidence that while CRYPTOCURRENCY SELLERs tend to be wealthier and have more income than the general population, some of that difference may actually reverse, depending on the comparison group. In addition, the fact that CRYPTOCURRENCY SELLERs report higher wage income than both comparison groups suggest they may be more sophisticated or wealthier, on average.

5. Conclusion

We offer the first broad-sample descriptive evidence on U.S. taxpayers selling cryptocurrency. As the number of cryptocurrency reporters increases and cryptocurrencies become a larger part of the financial ecosystem, it is imperative that regulators and rule-makers understand who sells cryptocurrencies. Our analyses suggest that despite increasingly widespread cryptocurrency investment, users are distinct from other U.S. investors and from non-investing taxpayers (e.g., certain geographic areas of the U.S. continue to be top cryptocurrency areas). Consistent with cryptocurrency gaining more mainstream appeal, we find that the number of counties in the U.S. with significant cryptocurrency reporting has increased dramatically. We also find evidence that the industries where cryptocurrency reporters work are becoming more diverse, moving from technology and finance related fields to areas such as restaurant workers. The association between certain personal attributes (e.g., gender, income, age, marital and student status) and cryptocurrency reporters are also changing over time, reinforcing the need for timely, broad-based evidence. Our study contributes to the growing literature on cryptocurrency and its users. We also provide timely evidence that can inform lawmakers and regulators as they seek to better target and construct legislation and rules.

References

Amiram, D., B. N. Jørgensen, D. Rabetti (2022). "Coins for Bombs: The Predictive Ability of On-Chain Transfers for Terrorist Attacks." *Journal of Accounting Research*, 60(2), 427–466.

Arnosti, N., S.M. Weinberg (2022). "Bitcoin: A Natural Oligopoly." *Management Science*.

Athey, S., I. Parashkevov, V. Sarukkai, & J. Xia (2016). "Bitcoin Pricing, Adoption, and Usage: Theory and Evidence." Stanford Graduate School of Business Working Paper.

Benetton, M., G. Compiani (2021). "Investors' Beliefs and Cryptocurrency Prices." The Cowles Foundation for Research in Economics, Yale University.

Bourveau, T., E.T. De George, A. Ellahie, D. Macciocchi (2022). "The Role of Disclosure and Information Intermediaries in an Unregulated Capital Market: Evidence from Initial Coin Offerings." *Journal of Accounting Research*, 60(1), 129–167.

Cecchini, M., H. Aytug, G.J. Koehler, P. Pathak (2010). Detecting management fraud in public companies. *Management Science*, 56(7), 1146–1160.

Cheng, S. F., G. De Franco, H. Jiang, P. Lin, P. (2019). "Riding the Blockchain Mania: Public Firms' Speculative 8-K Disclosures" *Management Science*, 65(12), 5901–5913.

Cong, Lin, Wayne R. Landsman, Edward L. Maydew, Daniel Rabetti (2021). "Tax-Loss Harvesting with Cryptocurrencies. Available at SSRN: https://ssrn.com/abstract=4033617 or http://dx.doi.org/10.2139/ssrn.4033617.

"Executive Order 13563, Improving Regulation and Regulatory Review." (Jan. 18, 2011) 76 Fed. Reg. 3821. https://obamawhitehouse.archives.gov/the-press-office/2011/01/18/executive-order-13563-improving-regulation-and-regulatory-review.

"Executive Order 14067, "Ensuring Responsible Development of Digital Assets" Fact Sheet re: President Biden to Sign Executive Order on Ensuring Responsible Development of Digital Assets. (March 9, 2022). Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/09/fact-sheet-president-biden-to-sign-executive-order-on-ensuring-responsible-innovation-in-digital-assets/.

Fabrichnaya, Elena, Alexander Marrow (January 21, 2022). "Russia Proposes Ban on Use and Mining of Cryptocurrencies" *Reuters* https://www.reuters.com/business/finance/russian-cbank-proposes-banning-cryptocurrencies-crypto-mining-2022-01-20/.

Federal Reserve. (January 20, 2022). "Federal Reserve Board Releases Discussion Paper That Examines Pros and Cons of a Potential U.S. Central Bank Digital Currency (CBDC)" Federalreserve.gov. https://www.federalreserve.gov/newsevents/pressreleases/other20220120a.htm.

Foley, S., J.R.Karlsen, T.J Putniņš (2019). "Sex, Drugs, and Bitcoin: How much Illegal Activity is Financed through Cryptocurrencies?" *The Review of Financial Studies*, 32(5), 1798–1853.

Gan, J., G. Tsoukalas, S. Netessine (2021). "Initial Coin Offerings, Speculation, and Asset Tokenization. *Management Science*, 67(2), 914–931.

Griffin, J. M., A. Shams (2020). "Is Bitcoin Really Untethered?" *The Journal of Finance*, 75(4), 1913–1964. http://dx.doi.org/10.1111/jofi.12903.

Hackethal, A., T. Hanspal, D.M. Lammer, K. Rink (2021). "The Characteristics and Portfolio Behavior of Bitcoin Investors: Evidence from Indirect Cryptocurrency Investments. *Review of Finance*.

Hall, J. (2022). "Tonga to Copy El Salvador's Bill Making Bitcoin Legal Tender, says Former MP." *Cointelegraph* (January 13, 2022). https://cointelegraph.com/news/tonga-to-copy-el-salvador-bill-making-bitcoin-legal-tender-says-former-mp.

Hall, Joseph. (Februart 16, 2022). "Portugal Slowly Becoming a 'Haven' for European Bitcoiners." *Cointelegraph* https://cointelegraph.com/news/portugal-slowly-becoming-a-haven-for-european-bitcoiners.

Hasso, T., M. Pelster, B. Breitmayer (2019). "Who Trades Cryptocurrencies, How Do They Trade It, and How Do They Perform? Evidence from Brokerage Accounts." *Journal of Behavioral and Experimental Finance*, 23, 64–74.

Higgins, Stan. "From $900 to $20,000: Bitcoin's Historic 2017 Price Run Revisited" (December 29, 2017). *CoinDesk*. https://www.coindesk.com/markets/2017/12/29/from-900-to-20000-bitcoins-historic-2017-price-run-revisited/.

Hoopes, J. L., P. Langetieg, S. Nagel, D. Reck, J. Slemrod, J., B.A. Stuart (2022). "Who Sells During A Crash? Evidence from Tax-Return Data on Daily Sales of Stock Short Title: Who Sells During A Crash?. *The Economic Journal*.

Hopkins, J. J., E.L. Maydew, M. Venkatachalam (2015). "Corporate General Counsel and Financial Reporting Quality." *Management Science*, 61(1), 129–145.

Kiernan, Paul. (August 3, 2021). "Crypto 'Wild West' Needs Stronger Investor Protection, SEC Chief Says." *Wall Street Journal* https://www.wsj.com/articles/sec-will-police-cryptocurrencies-to-maximum-possible-extent-chair-gary-gensler-says-11628007567?mod=article_inline.

Lei, Y., Y. Li, J. Xu (2020). "Two Birds, One Stone: Joint Timing of Returns and Capital Gains Taxes." *Management Science*, 66(2), 823–843.

Lopez, O. and E. Livni (2021). "In Global First, El Salvador Adopts Bitcoin as Currency." *The New York Times* (September 7, 2021). https://www.nytimes.com/2021/09/07/world/americas/el-salvador-bitcoin.html.

Makarov, I., A. Schoar (2020). "Trading and Arbitrage in Cryptocurrency Markets." *Journal of Financial Economics*, 135(2), 293–319.

Makarov, I., A. Schoar (2021). "Blockchain Analysis of the Bitcoin Market." *National Bureau of Economic Research*." (No. w29396).

Malik, N., M. Aseri, P.V. Singh, K. Srinivasan (2022). "Why Bitcoin Will Fail to Scale?. *Management Science*.

Maurer, Mark. (January 3, 2022). "FASB Targets New Long-Term Agenda, Rules on Expense Disclosure in 2022." *Wall Street Journal* https://www.wsj.com/articles/fasb-targets-new-long-term-agenda-rules-on-expense-disclosure-in-2022-11641205806.

Meiklejohn, S., M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G.M. Voelker, S. Savage. (October 2013). "A Fistful of Bitcoins: Characterizing Payments Among Men With No Names." *Proceedings of the 2013 Conference on Internet Measurement*, 127–140.

Nakamoto, Satoshi. (2008). "Bitcoin. A Peer-to-Peer Electronic Cash System." *Bitcoin.–URL:* https://bitcoin.org/bitcoin.pdf 4.

Newberry, Emma. (January 1. 2022). "Which U.S. States Are Worst for Crypto Investors?" *Motley Fool*. https://www.fool.com/the-ascent/cryptocurrency/articles/which-us-states-are-worst-for-crypto-investors/.

Quiroz-Gutierrez, Marco. (January 4. 2022)."Crypto Is Fully Banned in China and 8 Other Countries." *Fortune*. https://fortune.com/2022/01/04/crypto-banned-china-other-countries/.

Ron, Dorit, and Adi Shamir. (2013). "Quantitative Analysis of the Full Bitcoin Transaction Graph." *International Conference on Financial Cryptography and Data Security*. Springer, Berlin, Heidelberg.

Schlott, Rikki. (February 5, 2022). "Four Ordinary People Share How They Got Rich from Crypto." *New York Post*. https://nypost.com/2022/02/05/how-cryptocurrency-made-these-four-ordinary-people-rich/.

Securities and Exchange Commission. (December 22, 2020). "SEC Charges Ripple and Two Executives with Conducting $1.3 Billion Unregistered Securities Offering" SEC.gov. https://www.sec.gov/news/press-release/2020-338.

Securities and Exchange Commission. (February 14, 2022). "BlockFi Agrees to Pay $100 Million in Penalties and Pursue Registration of its Crypto Lending Product" SEC.gov. https://www.sec.gov/news/press-release/2022–26.

Singh, Manish. (February 15, 2022). “Cryptocurrency is Akin to ‘Ponzi scheme’ and Banning it is ‘Perhaps the Most Advisable Choice Says India’s Central Bank” *TechCrunch*. https://finance.yahoo.com/news/cryptocurrency-ponzi-scheme-banning-perhaps-163358692.html.

Sutton, Sam. (March 9, 2022). “Biden Orders First Sweeping Review of Federal Policy on Crypto.” *Politico* https://www.yahoo.com/now/biden-orders-first-sweeping-review-115213189.html.

Thompson, Mark. (January 19, 2022). “EU to Consider Banning Tax Haven Based Crypto Exchanges” *Law360*, https://www.law360.com/tax-authority/articles/1456696/eu-to-consider-banning-tax-haven-based-crypto-exchanges.

Versprille, Allyson. “Crypto Firms Brace For New Tax-Reporting Rules To IRS” (January 7, 2022). *Financial Adviser*, https://www.fa-mag.com/news/crypto-firms-brace-for-new-tax-reporting-rules-to-irs-65662.html?section=.

Warren, Elizabeth. Letter to Janet Yellen as head of the Financial Stability Oversight Council (FSOC). “FSOC Crypto Letter” (July 26, 2021). Available at: https://www.warren.senate.gov/imo/media/doc/FSOC percent20Crypto percent20Letter percent2007.26.2022.pdf.

Appendix A. Variable Descriptions

VARIABLE	DESCRIPTION
Variables of Interest	
CRYPTOCURRENCY SELLER	1 if either the description of a Form 8949 transaction is identified as cryptocurrency or a description from Form 1099-B is identified as cryptocurrency for tax returni in yeart. 0 otherwise. See online appendix A for a description of the textual analysis which identifies transactions as cryptocurrency.
NON-CRYPTO SELLING INVESTOR	1 if tax return in yeart reports either a non-zero amount for dividends or a non-zero amount for capital gain on Form 1040, and is not identified as a CRYPTOCURRENCY SELLER in yeart. 0 otherwise.
NON-INVESTOR	1 if a tax return is neither a CRYPTOCURRENCY SELLER nor a NON-CRYPTO SELLING INVESTOR, 0 otherwise.
CRYPTOCURRENCY GAIN*	Sum of the total gain or loss reported on Form 8949 for transactions identified as cryptocurrency for tax returni in yeart
NUM OF CRYPTO TRANSACTIONS*	Number of separate lines which are identified as cryptocurrency transactions on Form 8949 for tax returni in yeart
CRYPTOCURRENCY 1099B	An indicator equal to 1 if the primary or secondary taxpayer received any Form 1099-B which includes a transaction identified as cryptocurrency. See Online Appendix A. 0 Otherwise.
TREND	A year trend variable which takes the value of 0 in 2013 and increases in increments of 1.
CRYPTOCURRENCY MILLIONAIRE	1 for taxpayeri if ≥ $1,000,000
EQUITY MILLIONAIRE	1 for taxpayeri if ≥ $1,000,000 and CRYPTOCURRENCY MILLIONAIRE = 0
Continuous/Discrete Variables	
AGE	The year in which tax returnit was filed less the birth year for the primary taxpayer on tax returni
WAGES	Wages as reported on Form 1040 for tax returni in yeart.
TAXABLE INTEREST	Taxable Interest as reported on Form 1040 for tax returni in yeart.
TAXABLE DIVIDENDS	Taxable Dividends as reported on Form 1040 for tax returni in yeart.
CAPITAL GAIN/LOSS†	Capital Gain/Loss as reported in Form 1040 for tax returni in yeart.
TAXABLE INCOME	Taxable income after all deductions reported on Form 1040 for tax returni in yeart.
DEPENDENTS	Number of dependents reported on a taxpayer's return for yeart. This variable ranges from 0 to 4 dependents due to restrictions in IRS data.
Indicator Variables	
MARRIED	1 if tax returni in yeart reports both a primary taxpayer and a spouse, 0 otherwise.
SINGLE MALE	1 if tax returni in yeart does not report a spouse and census data lists the primary taxpayer as male. 0 if census data lists the primary taxpayer as female. Missing otherwise.

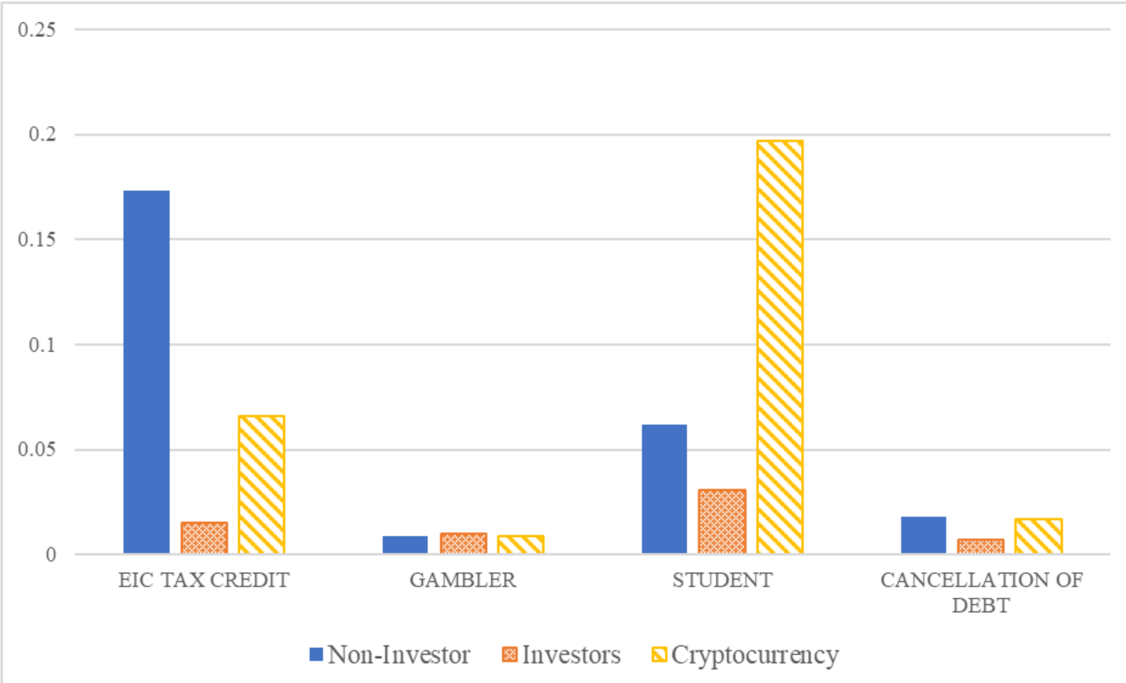
VARIABLE *Variables of Interest*	DESCRIPTION
SCH A‡	1 if tax returni in yeart had Schedule A for Itemized Deductions attached. 0 otherwise.
EIC TAX CREDIT‡	1 if tax returni in yeart included Schedule EIC for the Earned Income Tax Credit. 0 otherwise.
HOMEOWNER‡	1 if tax returni in yeart receives a Form 1098 for mortgage interest.
GAMBLER‡	1 if tax returni in yeart receives a W-2G for gambling winnings with reported amounts in Box 1 or Box 7
STUDENT‡	1 if tax returni in yeart receives a Form 1098-T for tuition and has reported amounts in Box 1 for Tuition and Fees in Box 1
CANCELLATION OF DEBT‡	1 if tax returni in yeart receives a 1099-C for the cancellation of debt and reports an amount in Box 2

* CRYPTOCURRENCY GAIN and NUM CRYPTO TRANSACTIONS are only non-zero for tax returns for which we identify cryptocurrency transactions. It is possible that some cryptocurrency transactions are summarized on these lines or are summarized on the Schedule D of Form 1040. Thus, they should be interpreted as lower bounds rather than absolute values.

† CAPITAL GAIN/LOSS is reported on Form 1040 after the capital loss limitation. The minimum value for this variable is -3,000. Losses in excess of -3,000 are carried forward and included in the next year's CAPITAL GAIN/LOSS amount.

‡The indicator variables for SCH A and EIC TAX CREDIT are indicators for the presence of their respective forms, Schedule A and Schedule EIC. Filing these forms is at the discretion of the taxpayer and does not mean that they claimed the credit or reduced the taxes due of the taxpayer. HOMEOWNER is an indicator variable for the presence of third-party reported information on mortgage interest. It therefore captures taxpayers who may or may not report the item on their individual tax returns, but it may not capture taxpayers who fall under the reporting thresholds. Such as taxpayers who pay less than $600 in Mortgage, interest.

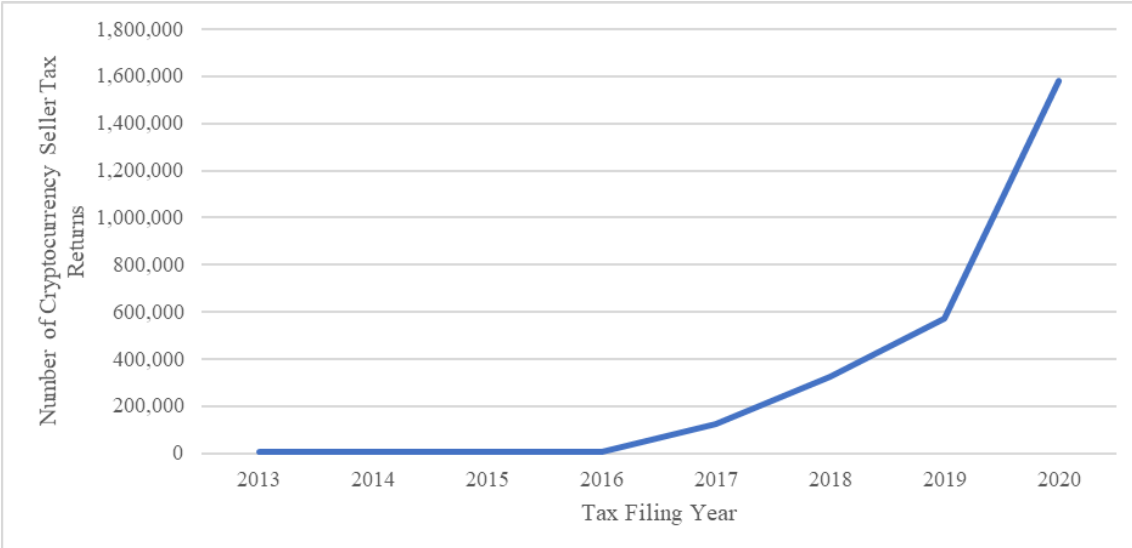
FIGURE 1. Characteristics by Taxpayer Type



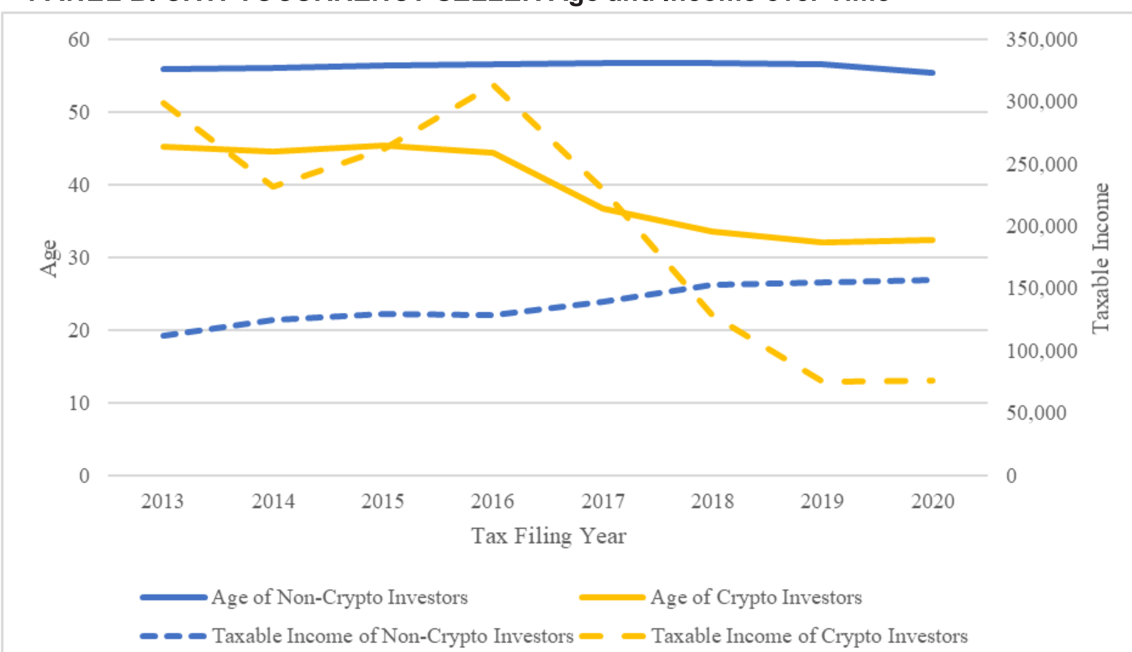
Note: Figure 1 shows the percentage of tax returns, split by taxpayer type, for various statistics. EIC is the percentage of tax returns which include the Earned Income Tax Credit, GAMBLER is the percentage of returns which receive a Form W-2G for gambling income, STUDENT is the percentage of returns which receive a Form 1098-T for tuition expense, and CANCELLATION OF DEBT is the percentage of returns which receive a Form 1099-C for cancellation of debt income.

FIGURE 2. Time Trends in Cryptocurrency

PANEL A. Number of *CRYPTOCURRENCY SELLER*s over time

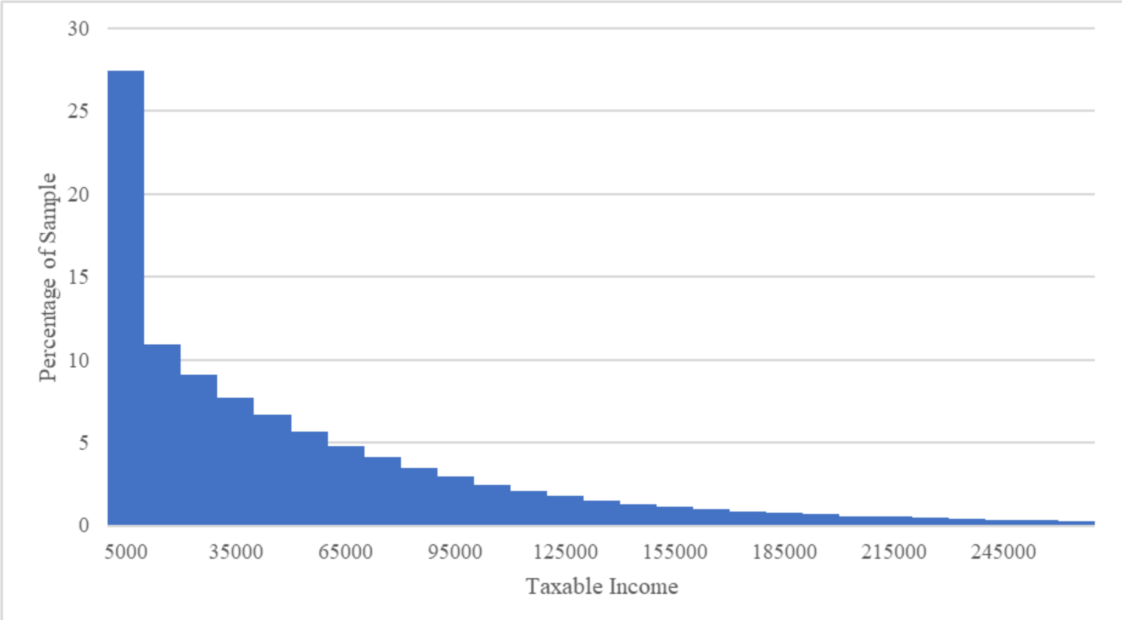


PANEL B. *CRYPTOCURRENCY SELLER* Age and Income over Time



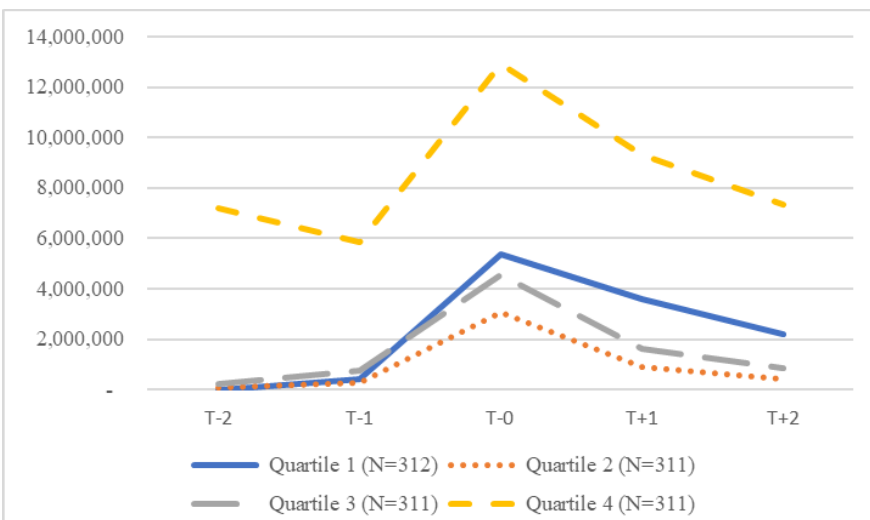
Note: Panel A reports the number of taxpayers who report cryptocurrency each year over our sample period. The number of reporters in trends upward from 4,344 (2013) to over 1.5 million (2020). Panel B splits the population into CRYPTOCURRENCY SELLERs or non-CRYPTOCURRENCY SELLERs and shows both the age (left-hand Y axis) and Taxable Income (right-hand Y axis) for both groups.

FIGURE 3. Histogram of Taxable Income for *CRYPTOCURRENCY SELLERs*



Note: Figure 2 shows the Histogram for Taxable Income for *CRYPTOCURRENCY SELLERS* across the sample period. We limit the Y axis to $270,000 of taxable income, which relates approximately to the 95th percentile. Bin width is $10,000, with midpoints listed on the x-axis. Tax returns which would have less than $0 of income due to losses or deductions are limited to $0 due to tax reporting rules.

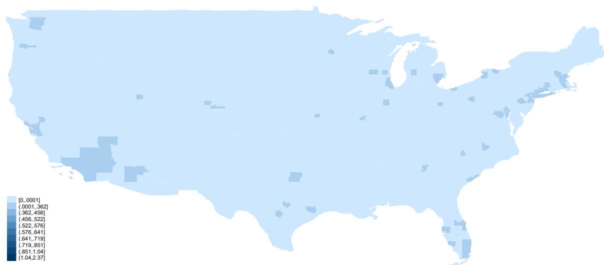
FIGURE 4. Mean Taxable Income of Taxpayers with >$1 Million Cryptocurrency Gains



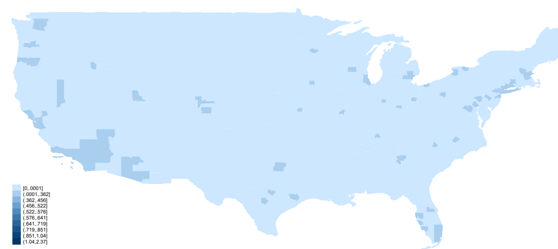
Note: Figure 4 graphs average taxable income over time of taxpayers who reported a Cryptocurrency capital gain of at least $1 million. Taxpayers are divided into quartiles based on their taxable income in T-2. If a taxpayer does not file a return in T-2, we assume that their taxable income is 0. In order to have data to complete quartiles, we only include gains beginning in 2015. The first cryptocurrency gain of at least $1 million is set as T-0.

FIGURE 5. Heat Map of CRYPTOCURRENCY SELLERs over time

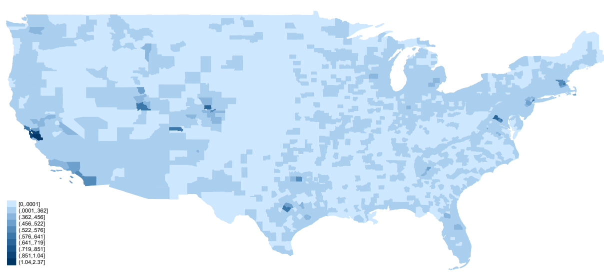
Panel A. 2014



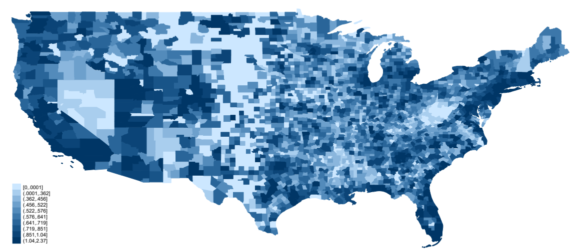
Panel B. 2016



Panel C. 2018



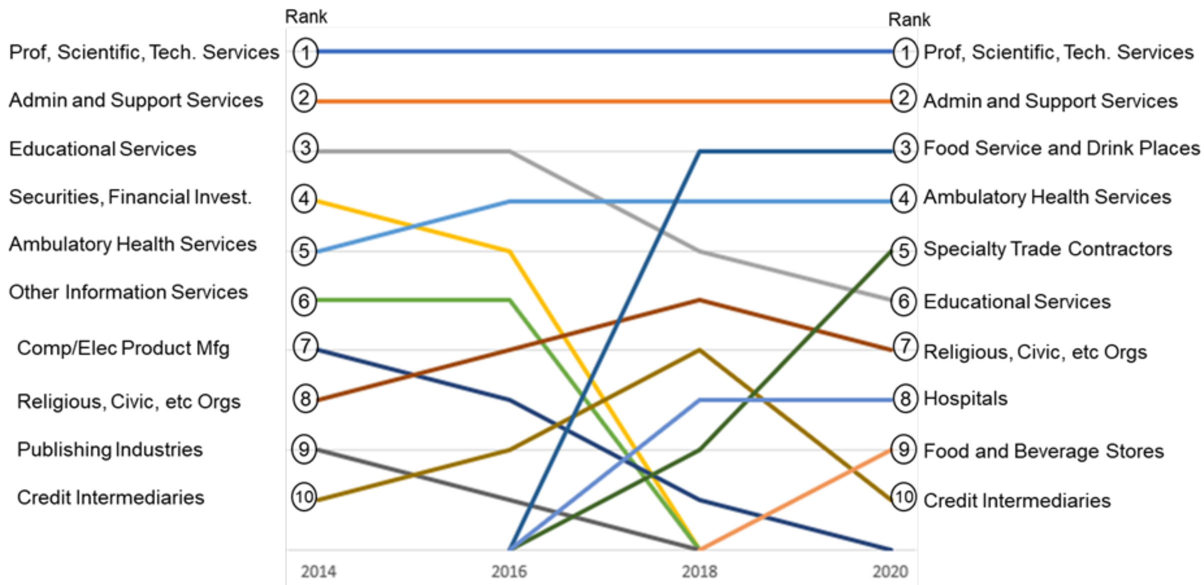
Panel D. 2020



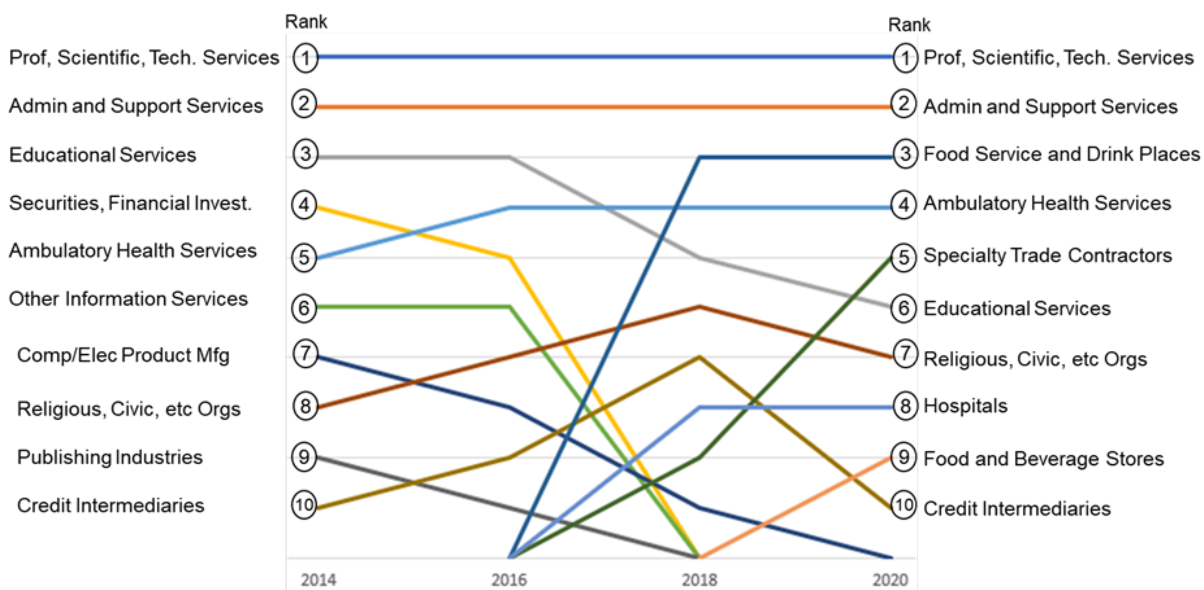
Note: Figure 5 displays heat maps of the percentage of cryptocurrency sellers in each county in the continental U.S. in 2014, 2016, 2018, and 2020. Breakpoints between colors are based on the decile rankings for 2020 to make colors comparable between graphs (Breakpoints: 0, >0 to 0.362, 0.362 to 0.456, 0.456 to 0.522, 0.522 to 0.576, 0.576 to 0.641, 0.641 to 0.719, 0.719 to 0.851, 0.851 to 1.04, and 1.04 to 2.37).

FIGURE 6. Industry of *CRYPTOCURRENCY SELLER*s over time

PANEL A: Top *CRYPTOCURRENCY SELLER* Job Industries Over Time by Percentage



PANEL B: Top CRYPTOCURRENCY SELLER Job Industries Over Time by Number



Note: Panel A presents the ratio of *CRYPTOCURRENCY SELLER*s in a particular business industry by year compared to all taxpayers in the given industry, Panel B presents the top business industries over time ranked by the number of *CRYPTOCURRENCY SELLER*s. To identify industry of a taxpayer, we obtain the population of W-2 data for our sample years, which reports wage income and use the W-2 with the highest reported income each year. Next, we identify the three-digit NAICS code based on the business tax return that filed the Form W-2. Since *CRYPTOCURRENCY SELLER* is calculated at the tax return level, if a joint tax return is filed, we assume both spouses are/are not holders of cryptocurrency. The denominator is the total taxpayers whose highest paid W-2 is in the given industry. Each taxpayer is assigned only a single industry. Data points are for each even numbered year between 2014 and 2020. A point at the bottom of each chart means that the specified industry was not in the top 10.

TABLE 1. Descriptive Statistics

Variables of Interest	NON-INVESTOR (N=845,102,236)			NON-CRYPTOCURRENCY INVESTOR (N=230,965,310)			CRYPTOCURRENCY SELLERS (N=2,620,926)		
	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
AGE	41.47	16.72	39	56.26	18.52	58	32.78	10.75	30
WAGES	39,506	257,012	26,604	86,318	392,742	37,413	77,049	355,060	46,010
TAXABLE INTEREST	98	49,955	0	2,757	152,022	47	1,733	1,263,304	0
TAXABLE DIVIDENDS	0	0	0	7,882	267,688	469	1,649	81,328	0
SCH A	0.168	0.374	0	0.442	0.497	0	0.131	0.338	0
MARRIED	0.316	0.465	0	0.584	0.493	0	0.378	0.485	0
MALE	0.314	0.464	0	0.182	0.385	0	0.541	0.498	0
STUDENT	0.062	0.242	0	0.031	0.174	0	0.197	0.398	0
LN WAGES	8.877	3.611	10.19	7.491	5.263	10.53	9.724	3.309	10.74
LN DIVIDENDS	-	0.000	0	5.663	3.194	6.15	1.686	2.690	0
Descriptive Variables									
CAPITAL GAIN/LOSS*	0	0	0	22,512	856,090	24	18,765	1,167,616	0
TAXABLE INCOME	34,346	88,971	19,747	138,353	1,079,166	67,115	91,421	1,086,150	36,372
CRYPTOCURRENCY GAIN	-			-			12,484	824,804	27
NUM OF CRYPTO TRANSACTIONS	-			-			9.90	100.40	1
EIC TAX CREDIT	0.173	0.378	0	0.015	0.123	0	0.066	0.249	0
GAMBLER	0.009	0.096	0	0.010	0.099	0	0.009	0.094	0
CANCELLATION OF DEBT	0.018	0.134	0	0.007	0.085	0	0.017	0.129	0
CRYPTOCURRENCY 1099B	-			-			0.784	0.412	1

Note: Table 1 reports descriptive statistics for the full sample of taxpayers (2013-2020) split out between NON-INVESTORs, NON-CRYPTOCURRENCY SELLING INVESTORs, and CRYPTOCURRENCY SELLERs. CRYPTOCURRENCY SELLERs are taxpayers who we identify as selling cryptocurrency for year t through textual analysis of Form 8949 Capital Gain descriptions, or who receive a Form 1099-B which we identify as relating to cryptocurrency through textual analysis of the description. NON-CRYPTOCURRENCY SELLING INVESTORs are taxpayers who are not identified as selling cryptocurrency but do report either Dividends or a Capital Gain or loss on their Form 1040 in year t. NON-INVESTOR Taxpayers are all other taxpayers. CAPITAL GAIN/LOSS is limited to the 3,000 capital loss limitation, however, CRYPTOCURRENCY GAIN is calculated on a transaction level basis and is not calculated with regard to the overall capital gain limitation. CRYPTOCURRENCY GAIN and NUM OF CRYPTO TRANSACTIONS are calculated only using information from Form 8949 and thus have limited non-missing observations (863,340 and 894,177 respectively). Some transactions reported on Form 1099-B may be summarized by taxpayers on their Form 1040 or Schedule D, and thus we would not be able to identify reported amounts for those transactions from the tax return. We avoid calculating reported amounts from Form 1099-B to avoid double counting transactions which are reported on both Form 8949 and Form 1099-B. SINGLE MALE and MARRIED are part of a categorical variable where the baseline is taxpayers who do not file a joint return and are female. Medians are calculated as the mean of the observations around the median observation per IRS disclosure guidelines. Due to missing values for gender and age in the Social Security Administration database, a small number of values for those amounts are missing. In order to comply with IRS data disclosure requirements, medians are calculated as a local average around the true median.

TABLE 2. CRYPTOCURRENCY MILLIONAIRE descriptive statistics

	NON-INVESTOR		NON-CRYPTOCURRENCY INVESTOR		CRYPTOCURRENCY SELLER		EQUITY MILLIONAIRE		CRYPTOCURRENCY MILLIONAIRE	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Continuous Variables										
AGE	41.47	16.72	56.26	18.52	32.69	10.67	46.76	13.37	40.53	11.49
WAGES	39,505	257,012	86,215	389,079	74,387	182,305	457,407	2,688,332	366,092	5,790,899
TAXABLE INTEREST	98	49,955	2,739	151,735	1,173	1,266,663	63,392	523,876	96,070	949,502
TAXABLE DIVIDENDS	-	-	7,853	267,254	797	24,845	102,780	513,208	165,186	2,225,614
CAPITAL GAIN/LOSS*	-	-	22,207	839,433	5,682	536,312	1,065,872	8,058,907	2,397,415	20,173,968
CRYPTOCURRENCY GAIN	-	-	-	-	3,402	33,204	-33,068	3,801,461	1,753,248	8,722,526
TAXABLE INCOME	34,345	88,960	137,908	1,068,095	74,386	355,173	1,699,424	8,233,654	2,833,299	16,598,034
NUM OF CRYPTO TRANSACTIONS	-	-	-	-	10	93	15	214	55	405
Indicator Variables										
TAXABLE LTCG	0.000	0.002	0.435	0.496	0.196	0.397	0.517	0.500	0.493	0.500
SCH A	0.168	0.374	0.442	0.497	0.127	0.333	0.805	0.396	0.569	0.495
EIC TAX CREDIT	0.173	0.378	0.015	0.123	0.067	0.250	0.002	0.049	0.006	0.078
MARRIED	0.316	0.465	0.584	0.493	0.375	0.484	0.747	0.435	0.559	0.497
MALE	0.314	0.464	0.182	0.385	0.543	0.498	0.223	0.416	0.414	0.493
CRYPTOCURRENCY 1099B	-	-	-	-	0.789	0.408	0.026	0.158	0.024	0.154
STUDENT	0.06	0.24	0.03	0.17	0.198	0.399	0.018	0.132	0.027	0.163
GAMBLER	0.01	0.10	0.01	0.10	0.009	0.094	0.012	0.108	0.015	0.122
CANCELLATION OF DEBT	0.02	0.13	0.01	0.09	0.017	0.129	0.005	0.069	0.006	0.080

Note: These statistics are for the full sample of taxpayers (2013-2020) split out between NON-INVESTORs, NON-CRYPTOCURRENCY SELLING INVESTORs, CRYPTOCURRENCY SELLERs, EQUITY MILLIONAIREs, and CRYPTOCURRENCY MILLIONAIREs. CRYPTOCURRENCY MILLIONAIRE is a time invariant indicator for taxpayers who recognize over $1 million of gain as a result of cryptocurrency transactions over the sample period. EQUITY MILLIONAIRE is a time invariant indicator for taxpayers who recognize over $1 million of equity capital gain and do not recognize over $1 million of cryptocurrency gain. CAPITAL GAIN/LOSS is limited to the $3,000 capital loss limitation; however, CRYPTOCURRENCY GAIN is calculated on a transaction level basis and is not calculated with regard to the overall capital gain limitation. CRYPTOCURRENCY GAIN and NUM OF CRYPTO TRANSACTIONS are calculated only using information from Form 8949 and thus have limited non-missing observations (863,340 and 894,177 respectively). Some transactions reported on Form 1099-B may be summarized by taxpayers on their Form 1040 or Schedule D, and thus we would not be able to identify reported amounts for those transactions from the tax return. We avoid calculating reported amounts from Form 1099-B in order to avoid double counting transactions which are reported on both Form 8949 and Form 1099-B. SINGLE MALE and MARRIED are part of a categorical variable where the baseline is taxpayers who do not file a joint return and are female. Medians are calculated as the mean of the observations around the median observation per IRS disclosure guidelines. Due to missing values for gender and age in the Social Security Administration database, a small number of values for those amounts are missing.

TABLE 3. Top 10 Cryptocurrency Cities for 2014 and 2020

PANEL A. Top Cryptocurrency Cities for 2014 and 2020 by Percentage of Taxpayer Returns

2014			2020		
Rank	City	Percentage of Sellers	Rank	City	Percentage of Sellers
1	Menlo Park, CA	0.0637%	1	Sunnyvale, CA	1.5402%
2	Mountain View, CA	0.0522%	2	Mountain View, CA	1.5245%
3	San Francisco, CA	0.0410%	3	Ross, CA	1.4648%
4	Palo Alto, CA	0.0384%	4	Milpitas, CA	1.4638%
5	Redmond, WA	0.0369%	5	Cupertino, CA	1.4595%
6	Cambridge, MA	0.0304%	6	Santa Clara, CA	1.4430%
7	New York, NY	0.0202%	7	Redmond, WA	1.4292%
8	Fremont, CA	0.0187%	8	Fremont, CA	1.4005%
9	Seattle, WA	0.0147%	9	Dublin, CA	1.3889%
10	Plano, TX	0.0137%	10	Secaucus, NJ	1.3819%

PANEL B. Top Cryptocurrency Cities for 2014 and 2020 by Number of Taxpayers

2014			2020		
Rank	City	Number of Sellers	Rank	City	Number of Sellers
1	San Francisco, CA	162	1	Brooklyn, NY	5,425
2	New York, NY	158	2	New York, NY	5,358
3	Seattle, WA	60	3	Los Angeles, CA	4,775
4	Brooklyn, NY	59	4	Chicago, IL	4,683
5	Austin, TX	47	5	San Francisco, CA	4,450
6	Los Angeles, CA	42	6	Houston, TX	3,912
7	Houston, TX	42	7	Austin, TX	3,880
8	Chicago, IL	41	8	Seattle, WA	3,643
9	San Jose, CA	30	9	San Jose, CA	3,593
10	Minneapolis, MN	30	10	San Diego, CA	3,546

Note: Panel A shows the top ten cities based on the percentage of CRYPTOCURRENCY SELLER tax returns filed in the given city over all tax returns filed in the given city. The percentage is not calculating the number of taxpayers as one tax return may relate to either one or two taxpayers given the filing status. Panel B shows the top ten cities based on the total number of CRYPTOCURRENCY SELLER tax returns filed in the given city. For both panels, we require any given city to have at least 1,000 tax returns filed in the year, and at least 10 CRYPTOCURRENCY SELLER returns filed in the year to reduce extreme percentages and due to IRS data restrictions. Taxpayer city is defined using taxpayer provided information on the Form 1040.

TABLE 4. Top Cryptocurrency Job Industries for 2014 and 2020

PANEL A. Top Cryptocurrency Job Industries for 2014 and 2020 by Percentage

	2014			2020		
Rank	**Industry**	**NAICS3**	**Percent**	**Industry**	**NAICS3**	**Percent**
1	Other Information Services	519	0.04%	Other Information Services	519	3.29%
2	Securities, Commodity Contracts, and Other Financial Investments and Related Activities	523	0.02%	Internet Publishing and Broadcasting	516	2.70%
3	Data Processing, Hosting, and Related Services	518	0.02%	Data Processing, Hosting, and Related Services	518	2.61%
4	Publishing Industries (except Internet)	511	0.01%	Publishing Industries (except Internet)	511	2.15%
5	Computer and Electronic Product Manufacturing	334	0.01%	Electronics and Appliance Stores	443	2.04%
6	Professional, Scientific, and Technical Services	541	0.01%	Professional, Scientific, and Technical Services	540	1.91%
7	Museums, Historical Sites, and Similar Institutions	712	0.01%	Computer and Electronic Product Manufacturing	334	1.76%
8	Motion Picture and Sound Recording Industries	512	0.01%	Nonstore Retailers	454	1.73%
9	Funds, Trusts, and Other Financial Vehicles	525	0.01%	Telecommunications	517	1.72%
10	Telecommunications	517	0.00%	Information	510	1.65%

PANEL B: Top Cryptocurrency Job Industries for 2014 and 2020 by Number

	2014			2020		
Rank	**Industry**	**NAICS3**	**Number**	**Industry**	**NAICS3**	**Number**
1	Professional, Scientific, and Technical Services	541	943	Professional, Scientific, and Technical Services	541	227,586
2	Administrative and Support Services	561	301	Administrative and Support Services	561	141,207
3	Educational Services	611	248	Food Services and Drinking Places	722	98,763
4	Securities, Commodity Contracts, and Other Financial Investments and Related Activities	523	212	Ambulatory Health Care Services	621	71,381
5	Ambulatory Health Care Services	621	191	Specialty Trade Contractors	238	56,926
6	Other Information Services	519	143	Educational Services	611	55,423
7	Computer and Electronic Product Manufacturing	334	124	Religious, Grantmaking, Civic, Professional, and Similar Organizations	813	41,559
8	Religious, Grantmaking, Civic, Professional, and Similar Organizations	813	121	Hospitals	622	39,738
9	Publishing Industries (except Internet)	511	108	Food and Beverage Stores	445	39,737
10	Credit Intermediation and Related Activities	522	107	Credit Intermediation and Related Activities	522	33,102

Note: Table 4, Panel A presents the ratio of Cryptocurrency Sellers in a particular business industry by year compared to all taxpayers in the given industry. To identify industry of a taxpayer, we obtain the population of W-2 data for our sample years, which reports wage income and use the W-2 with the highest reported income each year. Next, we identify the three-digit NAICS code based on the business tax return that filed the Form W-2.. Since *CRYPTOCURRENCY SELLER* is calculated at the tax return level, if a joint tax return is filed, we assume both spouses are/are not holders of cryptocurrency. The denominator is the total taxpayers whose highest paid W-2 is in the given industry. Each taxpayer is assigned only a single industry. Panel B presents the same information except that instead of the ratio of *CRYPTOCURRENCY SELLER* s total, industries are ranked by the raw number of *CRYPTOCURRENCY SELLER* s.

TABLE 5. Determinants of *CRYPTOCURRENCY SELLER*

Independent Variables	Dependent Variable: CRYPTOCURRENCY SELLER					
	Model 1			Model 2		
	Estimate	Std. Error	†	Estimate	Std. Error	†
AGE (UNDER 24)	0.00458	(0.000062)	10	-0.00395	(0.000068)	10
AGE (25-44)	0.00466	(0.000054)	10	-0.00372	(0.000058)	10
AGE (45-64)	0.00113	(0.000036)	10	-0.00073	(0.000039)	10
LN WAGES	0.00006	(0.000004)	10	-0.00004	(0.000005)	10
LN DIVIDENDS	0.00035	(0.000007)	10	-0.00021	(0.000007)	10
MARRIED	0.00247	(0.000035)	10	-0.00333	(0.000057)	10
SINGLE MALE	0.00353	(0.000041)	10	-0.00482	(0.000066)	10
HOMEOWNER	0.00029	(0.000036)	10	-0.00012	(0.00004)	6
DEPENDENTS	-0.00031	(0.000017)	10	0.00032	(0.000019)	10
STUDENT	0.00380	(0.000126)	10	-0.00487	(0.00016)	10
TREND				-0.00140	(0.000018)	10
AGE (UNDER 24) * TREND				0.00232	(0.000034)	10
AGE (25-44) * TREND				0.00227	(0.000028)	10
AGE (45-64) * TREND				0.00048	(0.000018)	10
LN WAGES * TREND				0.00002	(0.000002)	10
LN DIVIDENDS * TREND				0.00015	(0.000003)	10
MARRIED * TREND				0.00125	(0.000019)	10
SINGLE MALE * TREND				0.00179	(0.000022)	10
HOMEOWNER * TREND				0.00011	(0.00002)	10
DEPENDENTS * TREND				-0.00018	(0.00001)	10
STUDENT * TREND				0.00163	(0.000053)	10
Intercept	-0.00031	(0.000017)	10	0.00032	(0.000019)	10
Observations	10,000,000			10,000,000		
Year Fixed Effects	YES			NO		
Baseline Full Sample Probability of Crypto Seller	0.00243			0.00243		
Average Adjusted R2	0.002			0.002		

† The number of significant coefficients (out of 10) at the 1% level across the 10 random samples.

Note: Reported coefficient estimates, standard errors, and adjusted R^2 are average numbers over 10 iterations of random sampling. For each random sample, 10 million tax returns were selected at random from the full sample of tax returns (approx. 1.078 billion, from which the baseline full sample probability was computed). Numbers to the right of the coefficient are the number of coefficients that were significant at the 1% level over all iterations. Column (1) reports the results of model 1 on a random sample of all tax returns. Column (2) reports the results of model 2 on a random sample of all tax returns. Variables are defined in Online Appendix A. Robust Standard errors are reported in parentheses. To aid in the interpretation of coefficient magnitude, the baseline full sample probability of being a *CRYPTOCURRENCY SELLER* is reported at the bottom of each column.

TABLE 6. Cross Sectional Samples of Cryptocurrency Determinants

	Dependent Variable: CRYPTOCURRENCY SELLER							
	Sample (1)		Sample (2)		Sample (3)		Sample (4)	
	Early Sample 2013-2016		Late Sample 2017-2020		Only Crypto Sellers and Investors		Only Crypto Sellers and non-Investors	
Independent Variables	Estimate (Std. Error)	†	Estimate (Std. Error)	†	Estimate (Std. Error)	†	Estimate (Std. Error)	†
AGE (UNDER 24)	0.00004	8	0.00850	10	0.03255	10	0.00380	10
	(0.000012)		(0.000117)		(0.000574)		(0.000063)	
AGE (25-44)	0.00007	10	0.00856	10	0.02076	10	0.00348	10
	(0.000013)		(0.000098)		(0.000272)		(0.000053)	
AGE (45-64)	0.00004	7	0.00189	10	0.00090	10	0.00064	10
	(0.000012)		(0.000064)		(0.000135)		(0.000039)	
LN WAGES	0.00000	1	0.00010	10	0.00011	10	0.00004	10
	(0.000001)		(0.000008)		(0.000015)		(0.000005)	
LN DIVIDENDS	0.00002	10	0.00062	10	-0.00342	10	0.16578	10
	(0.000002)		(0.000012)		(0.000029)		(0.00085)	
MARRIED	0.00003	10	0.00469	10	0.00694	10	0.00256	10
	(0.000006)		(0.000067)		(0.000137)		(0.00004)	
SINGLE MALE	0.00004	10	0.00661	10	0.01865	10	0.00327	10
	(0.000006)		(0.000076)		(0.000257)		(0.000041)	
HOMEOWNER	0.00001	2	0.00051	10	-0.00788	10	0.00029	10
	(0.000007)		(0.000069)		(0.000155)		(0.000041)	
DEPENDENTS	0.00000	2	-0.00058	10	0.00006	0	-0.00030	10
	(0.000003)		(0.000033)		(0.000096)		(0.000017)	
STUDENT	0.00000	0	0.00413	10	0.02613	10	0.00363	10
	(0.000016)		(0.000163)		(0.000912)		(0.000126)	
Intercept	0.00000	2	-0.00058	10	0.00006	0	-0.00030	10
	(0.000003)		(0.000033)		(0.000096)		(0.000017)	
Observations	10,000,000				10,000,000			
Year Fixed Effects	YES		YES		YES		YES	
Baseline Full Sample Probability of Crypto Seller	0.00004		0.00459		0.00309		0.01122	
Average Adjusted R^2	0.000		0.005		0.011		0.003	

† The number of significant coefficients (out of 10) at the 1 percent level across the 10 random samples.

Note: Reported coefficient estimates, standard errors, and adjusted R^2 are average numbers over 10 iterations of random sampling. For each random sample, 10 million tax returns were selected at random from the full sample of tax returns (approx. 1.078 billion, from which the baseline full sample probability was computed). Numbers to the right of the coefficient are the number of coefficients that were significant at the 1 percent level over all iterations. Columns (1) and (2) report the results of model 1 run on the same random samples split by early sample period (2013-2016) and late sample period (2017-2020). Each column is therefore only a portion of the full 9 million tax return random sample. Columns (3) and (4) report the results of model 1 where the control sample consists of only NON-CRYPTO SELLING *INVESTORs* (3) or *NON-INVESTORs* (4). Each column is therefore only a portion of the full 10 million tax return random sample. Variables are defined in Online Appendix A. Robust standard errors are reported in parenthesis. To aid in the interpretation of coefficient magnitude, the baseline full sample probability of being a *CRYPTOCURRENCY SELLER* is reported at the bottom of each column. The number of observations in each random sample varies based on the cross-sectional split and random sample.

Online Appendix—Not for print publication

Overview of the Bitcoin Network and Transactions

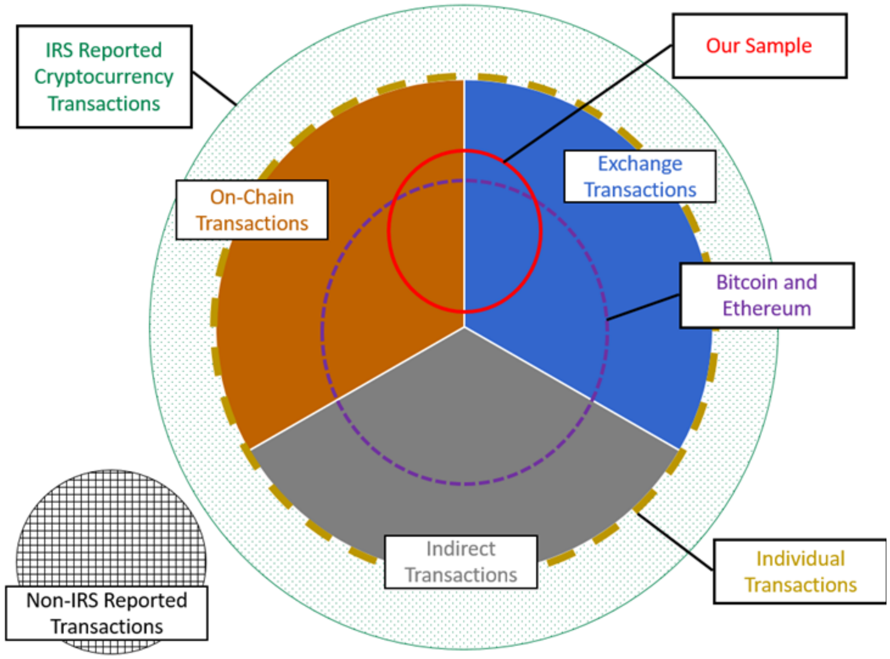
Bitcoin can refer both to the unit of account as well as the ledger which records transactions denominated in Bitcoin. Although we will discuss Bitcoin specifically, the general information applies to many other similar cryptocurrencies, and we attempt to note important differences. Bitcoin is a decentralized public ledger and can be thought of as serving a similar function to a bank. The Bitcoin ledger, commonly referred to as "the blockchain," contains and updates a list of transactions which can be used to identify how much Bitcoin is associated with each account. We next go over the key features of Bitcoin and similar cryptocurrencies.

Unlike a traditional bank, Bitcoin is decentralized. This means there is no central authority approving or processing transactions. Instead, when an individual wants to send Bitcoin, they broadcast the transaction to the entire Bitcoin network. Then, individuals or groups known as "Miners" observe those transactions and compete with each other for the right to confirm those transactions are legitimate and post them to the blockchain. This competition helps to ensure that no single entity has control over which transactions are or are not posted to the ledger. As a reward for the effort, miners are rewarded with both Bitcoin transaction fees paid by users, as well as a set Bitcoin reward which is created for each new batch of transactions that is confirmed. For Bitcoin, the competition for the right to post transactions is based on computing power, where miners with more computing power are more likely to win. Other cryptocurrencies use other mechanisms to determine which transactions are recorded on the blockchain.

Although the blockchain is a form accounting ledger, there are several differences which make it unique from other systems. First, the blockchain does not record running totals like a bank account. Instead, each account (also called a wallet) is the sum of all transactions that have taken place relating to that wallet. Therefore, in order to know the current balance of a wallet, one must examine the entire history of the blockchain, not simply the most recent transactions. The second difference is that Bitcoin accounts cannot be split. If a user has 100 Bitcoin in a wallet and want to spend 10 Bitcoin, then they must send 90 Bitcoin to another wallet owned by themselves and 10 Bitcoin to the external recipient. Wallets are reusable and can receive unlimited deposits. The third detail about Bitcoin is that it is a sender-based system. In order to send Bitcoin, all a sender needs is a Bitcoin account address, and Bitcoin can be sent without any action or even knowledge of the receiver. Taken together, this makes Bitcoin pseudonymous. The entire transaction history of each individual Bitcoin account can be observed, however, a single user can have an infinite number of accounts. In addition, because there is no central processing party, the identity of the owner of individual Bitcoin accounts is difficult to determine without additional information outside of the blockchain.

Several factors have led to innovation and changes within the cryptocurrency space. First, long transaction approval times (greater than 10 minutes for many Bitcoin transactions) and high transaction fees have led users to both transact with centralized Bitcoin market makers and develop competing cryptocurrencies with the aim of reducing the inefficiencies in Bitcoin. Second, although Bitcoin is pseudonymous, newer cryptocurrencies have been designed to increase privacy and security. Finally, new blockchains have been developed which allow users to increase the complexity of transactions. For example, a user could set up a transaction to send some value of cryptocurrency only if a specific set of identifiable outcomes is realized. Ultimately, the Bitcoin and cryptocurrency ecosystem continues to rapidly evolve and change over time, offering new opportunities but also challenges for investors, regulators, and researchers.

ONLINE Appendix Figure 1. Cryptocurrency Transaction Types (Not drawn to scale)



Notes: This figure is a representation of the various types of cryptocurrency transactions that taxpayers may engage in and how those definitions relate to our sample of identified cryptocurrency transactions. The figure is not intended to be definitive but is provided to help understand the relationship between the universe of transactions and those which we identify. We provide additional definitions below:

Non-IRS Reported Transactions: These are cryptocurrency transactions which are not reported to the IRS on an individual tax return on Form 8949 nor are reported to the IRS by third parties on Form 1099-B.

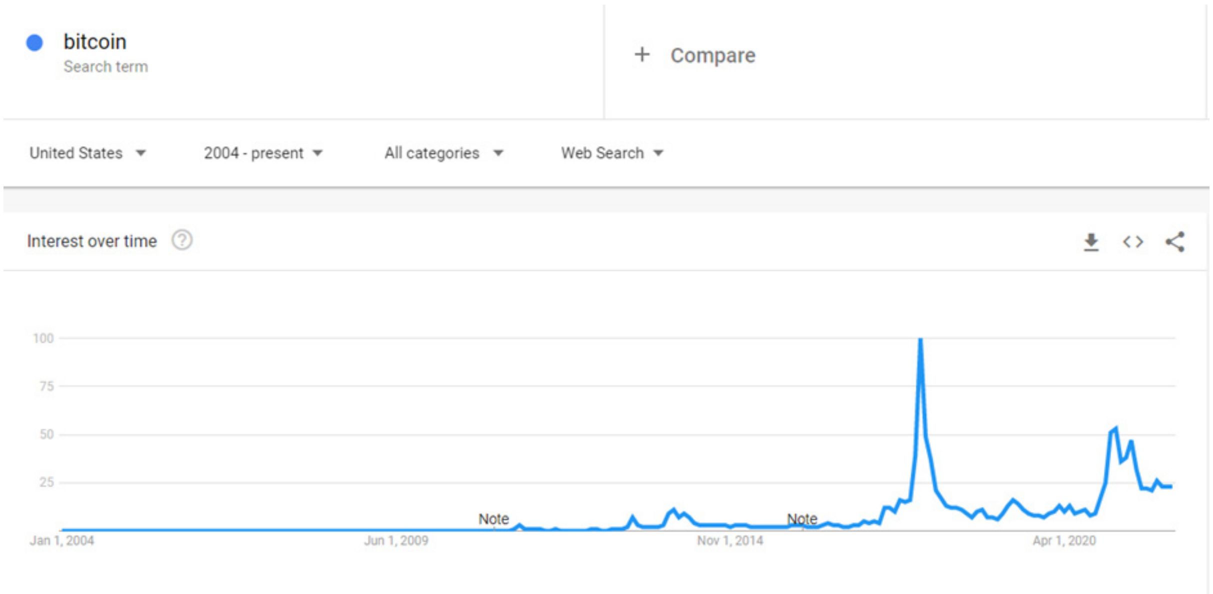
IRS Reported Cryptocurrency Transactions: These are transactions which are reported to the IRS. This could be through reporting on tax returns (Individual, Business, and Trust), or through reported to the IRS through various third-party reporting (e.g., Form 1099-B, Form 1099-MISC).

Individual Transactions: These transactions include only those transactions that are reported directly on Form 1040 for an individual taxpayer or are reported on Form 1099-B where the taxpayer has a Social Security Number. Individual transactions do not include transactions which are reported by businesses even if those transactions may eventually flow through to an individual return on Schedule K-1 and Schedule E.

On-Chain Transactions: On-Chain transactions refer to cryptocurrency transactions which are recorded permanently on a public blockchain. These transactions include sending or receiving cryptocurrency directly to individual wallets as well as sales of cryptocurrency made directly on the blockchain. The details of these transactions is generally publicly available but pseudo-anonymous. It is thus difficult to link the public blockchain data directly to taxpayers. On-Chain Transactions also are not generally subject to third-party reporting unless there is a centralized intermediary facilitating the transaction.

Exchange Transactions: Exchange transactions refer to cryptocurrency transactions done through a centralized third-party outside of the blockchain. These transactions are generally recorded on internal accounts or ledgers of the centralized party. Thus, individual transactions may not appear, or may only appear in aggregate on the blockchain.

ONLINE Appendix Figure 2. Google Trends Index for "bitcoin"



4

▽

Hidden Assets, Hidden Networks

Wind ◆ Bratt ◆ Graff ◆ Herlache

King ◆ Soto ◆ Yismaw ◆ Doyle ◆ Horvath ◆ Nowicki
Hess ◆ Gleason ◆ Sundstrom ◆ Brooks ◆ Mastrangelo
Stavrianos ◆ Hales

Following K-1s: Considering Foreign Accounts in Context

Tomas Wind, David Bratt, Alissa Graff, and Anne Herlache
(IRS: Research, Applied Analytics, and Statistics)[1]

I. Introduction

U.S. taxpayers' use of offshore accounts has been an area of focus for the IRS for some time. There are several reasons why this is the case. First, given the requirement that U.S. taxpayers pay taxes to the IRS regardless of their country of residence,[2] assets in overseas accounts held by taxpayers living outside of the United States may generate income that is legally subject to U.S. taxes. Second, and more significantly, foreign financial institutions' historically limited reporting of U.S. taxpayers' overseas assets to the IRS means that taxpayers who hold or move assets overseas may be underreporting their assets, thereby, not remitting the full amount of the U.S. taxes that they owe. For these and other reasons, the IRS has long implemented programs in its criminal investigatory and civil components to increase the number of U.S. taxpayers with overseas assets who come into and stay in compliance with the U.S. tax laws that pertain to their overseas assets.

Such compliance is a difficult task to ensure. In addition to the fact that some U.S. taxpayers who reside overseas may be unaware of their obligation to report their overseas assets to the IRS, recent immigrants to the U.S. who are obliged to pay taxes on their overseas accounts to the IRS may also be unaware of their need to do so. Moreover, the gaps in relevant reporting from overseas financial institutions described above makes it difficult for U.S. tax authorities to clearly understand the holdings of U.S. persons with overseas assets. And, finally, this murky picture is further clouded by the rise of the use of pass-through entities in a variety of tax scenarios.

In light of these challenges, the IRS's civil component has implemented several pathways over the past twentysomething years for those persons who owed U.S. taxes on their overseas holdings but had previously not reported these assets to the IRS to voluntarily come into compliance with U.S. tax law. While each of these programs (which are summarized below in Section II) had different criteria for participation, they generally allowed such persons to avoid the significant criminal penalties that they may have faced if their noncompliance had been discovered by IRS Criminal Investigations (CI). In doing so, they have made a substantial impact on the estimated size of the noncompliant population of U.S. persons with overseas assets.

While the particularities of U.S. tax law shaped the specifics of these initiatives, they were not rolled out in global isolation. Indeed, the Organisation for Economic Co-operation and Development, G20, and other international bodies made significant strides during this same time period to promote transparency in countries whose opaque financial industries masked potential noncompliance with tax laws on the part of citizens of a variety of countries. These efforts to combat tax evasion have been and continue to be an important backdrop to efforts undertaken by the IRS.[3]

Our paper examines one dynamic by which these initiatives may have had their ultimate effect on the population of noncompliant U.S. persons. In particular, we examine the influence of those who reported a foreign account during 2006–2020 on those other U.S. persons with whom they are linked through what we will refer to as a K-1 network. We show that a relationship exists between sharing a network with taxpayers that have reported a foreign account and reporting a foreign account.

[1] The findings, interpretations, and conclusions expressed in this paper 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.

[2] Filing requirements for U.S. citizens or resident aliens living or traveling outside the United States is determined by the amount of gross income from worldwide sources, filing status, and age.

[3] For an overview of these related international efforts, see "A Step Change in Tax Transparency" (OECD 2013).

II. Background on Foreign Account Reporting

Report of Foreign Bank and Financial Accounts

The Bank Secrecy Act (BSA) of 1970 stipulates that some U.S. persons must file a Report of Foreign Bank and Financial Account a.k.a. "FBAR" (FinCEN Form 114). Those persons include a citizen, resident, corporation, partnership, LLC, trust, or estate that have a financial interest in or authority over one or more overseas accounts. In keeping with other portions of the BSA, this reporting requirement only applies if the aggregate value of the foreign accounts of the U.S. person in question is greater than $10,000 in the calendar year in which it was reported.[4] U.S. persons who are required to file an FBAR must do so by submitting FinCEN Report 114 (which replaced TD Form 90-22.1 in 2013).

Foreign Account Tax Compliance Act

The Foreign Account Tax Compliance Act (FATCA) was passed in 2010 as part of the Hiring Incentives to Restore Employment Act. In requiring certain U.S. persons and entities to report their foreign account holdings, FATCA is broadly similar to FBAR. However, FATCA has a higher asset-reporting threshold $50,000 to $400,000 (depending on residence and marital status). It does not apply to U.S. persons in U.S. territories, and its definition of assets is broader than that covered by FBAR. In particular, FATCA requires the reporting of foreign stocks and securities, foreign financial instruments, contracts with non-U.S. persons, and other interests in foreign entities.[5]

Offshore Voluntary Disclosure Programs

We use "OVD" (Offshore Voluntary Disclosure) to refer to a series of four initiatives that the IRS undertook from 2009–2018.[6] These initiatives built off of CI's longstanding practice of taking voluntary disclosures under consideration when determining whether to recommend criminal prosecution. Across the four iterations of OVD, the IRS used an evolving set of incentives (in the form of reduced civil penalties and, in most cases, the waiver of criminal liability) to encourage taxpayers with overseas financial accounts and assets to come into compliance with US tax law.

The dates for each iteration of the OVD are as follows:

- 2009 Offshore Voluntary Disclosure Program (OVDP): March 23, 2009–October 8, 2009
- 2011 Offshore Voluntary Disclosure Initiative (OVDI): February 8, 2011–September 9, 2011
- 2012 OVDP: January 9, 2012–June 30, 2014
- 2014 OVDP: July 1, 2014–September 28, 2018 (*technically a continuation of the 2012 program, but with significant modifications)*

By assessing penalties on the accounts and assets covered by this program, the IRS thereby encouraged a subset of those who had been willfully noncompliant, U.S. taxpayers looking overseas who were unaware of their U.S. tax obligations, so-called "quiet disclosers," and others to come into compliance with U.S. tax law.[7] At the end of a successfully completed OVD filing, the taxpayer would enter into a Specific Matters Closing Agreement with the IRS.

[4] This overview is based on and, in some places, directly quotes the IRS's overview of Report of Foreign Bank and Financial Accounts (FBAR). There are several exceptions to this reporting requirement; see the linked page for further details. See also "Report of Foreign Bank & Financial Accounts (FBAR) Reference Guide," IRS Publication 5569 (3-2022).

[5] Summary of FATCA Reporting for U.S. Taxpayers | Internal Revenue Service (irs.gov), Foreign Account Tax Compliance Act (FATCA): Definition and Rules (investopedia.com).

[6] There was a related program in 2003 called the "Offshore Voluntary Compliance Initiative" that in some ways was a precedent for what we call the "OVD programs." However, its penalty structure was different enough, the reported disclosures of the program small enough, and the geopolitical context was different enough from the OVD programs from 2009 on that we exclude it from our analysis. For a comparison of the penalty structures of these programs including the 2003 OVCI), see GAO (2013).

[7] For a summary of the provisions and history of OVDP, see IRM 4.63.3.1.

Streamlined Filing Compliance Procedures

While OVD's target population included taxpayers whose noncompliance was at least in part willful, the Streamlined Filing Compliance Procedures (SFCP) were aimed at those whose noncompliance was completely *non-willful*. This program began on September 1, 2012, and is still available to those who believe that they may have been non-willfully noncompliant with U.S. tax law. SFCP was designed to complement OVD, as many taxpayers who were non-willfully noncompliant were entering OVD and then withdrawing from the program after determining that the program's penalty structure was not appropriate for their situation.[8] After SFCP was set up and while OVDP was still running, taxpayers had to make a mutually exclusive choice for one or the other paths to compliance.

In certifying that their noncompliance was not willful, taxpayers who hope to avail themselves of SFCP attested that their noncompliance was the result of their "negligence, inadvertence, or mistake or conduct that is the result of a good-faith misunderstanding of the requirements of the law."[9] While the streamlined filing procedure was initially available only to U.S. taxpayers residing abroad, the IRS subsequently opened this program up to U.S. taxpayers residing in the United States as well. The two different sets of procedures are known, respectively, as the "Streamlined Foreign Offshore Procedures" and the "Streamlined Domestic Offshore Procedures" (SDO). Aside from the difference in residency status, they differed primarily in that SDO imposes a five-percent so-called "miscellaneous offshore penalty."

III. Literature Review

Work on Pass-through Entities and Network Analysis

Pass-through entities (PTEs) have the dual character of being required to submit a tax return while not themselves being subject to federal income tax. This function of passing the income (and the attendant tax obligation) that comes into the PTE on to its constituent members makes this entity structure an attractive option in a variety of different tax planning scenarios. As such, PTEs and the resulting K-1 networks that they create have become more widely employed by U.S. taxpayers. (See Olson *et al.* (2022), p. 2-3 for a more detailed discussion of this topic.) They have also been the subject of a good deal of analysis by academics, tax authorities, and other interested parties.

Since the seminal work of Cooper *et al.* in their 2016 paper on pass-through entities, scholars have made innovative use of K-1 network data in their analysis of various aspects of tax administration in the United States (Cooper *et al.* (2016). In "Entity Structure and Taxes: An Analysis of Embedded Pass-Through Entities," for example, the authors demonstrate that the passthrough entities are more likely to appear in relatively complex corporate structures. They also show that the presence of PTEs in corporate structures is "significantly associated" with tax avoidance and uncertainty (Agarwal *et al.* (2021), p. 30). While the scope of their study is limited to C-corporations, the authors' analysis demonstrates the power of K-1 networks as a tool for analyzing such data and assisted us in developing a number of metrics used in this paper.

An earlier paper (Agarwal *et al.* (2015)) by a subset of the same authors and other collaborators within the IRS demonstrates another way to profitably analyze K-1 network data. In particular, the authors of this paper show how one can analyze taxpayer networks linked by K-1s through the lens of social network analysis. In indexing the types and numbers of entities within K-1 networks and the different types of linkages between them, this paper defines typical network structures and identifies anomalous pass-through arrangements that may be of interest to tax authorities. Recent work (Love (2021)) has given a more refined description of the specific entities, dynamics, and capital flows that appear within K-1 networks. After observing that Cooper *et al.* were able to identify the ultimate recipients of 74 percent of the income flowing through the universe of K-1 data reported to the IRS, Love uses an expanded set of data points and tailored algorithms to attribute most of the income that Cooper *et al.* had been unable to pin down. (Love estimates that he is able to account for 99

[8] https://www.irsvideos.gov/business/FilingPayingTaxes/StreamlinedFilingComplianceProceduresAComplianceOptionForSomeTaxpayers.

[9] Streamlined Filing Compliance Procedures | Internal Revenue Service (irs.gov).

percent of the total income flowing through this space.) Love shows that much of this newly revealed activity is 1) associated with the financial industry, 2) makes use of a so-called "blocker entity" that is domiciled overseas and which reduces and/or redirects its owners' tax liability, and 3) largely directs capital from the U.S. to foreign partners. Given the opacity of the operations of such blocker entities, the final beneficiaries of such flows are unclear, but Love suggests that the general flow of passive investment income from the U.S. to overseas partners has deleterious effects that must be weighed against the benefits of the inflows of foreign investment into the U.S. that are also facilitated by U.S. tax policy.

Black *et al.* (2023) further advances our understanding of the landscape of partnerships that report to the IRS. Using various forms of graphical analysis to visualize the structure of partnership arrangements and the capital flows within them, Black et al suggest a general division of partnerships into simple, single-owner structures on one hand and complex structures with multiple owners on the other. Suggesting that the two constitute, respectively, 85 percent and 15 percent of the total population of partnerships, Black *et al.* show that this smaller group of complex partnerships is characterized by circular ownership structures in a given partnership as well as multidirectional flows of capital therein. In terms of methodology, the authors demonstrate that random forest models outperform traditional linear regressions in terms of the accuracy of each predicted noncompliance.

Work on Foreign Accounts

In addition to the research on K-1 networks outlined above, there have also been several studies on the effect of the range of initiatives that the IRS has taken over the past twenty years to increase taxpayer compliance with reporting requirements for overseas accounts. One notable early study of the effect of the earlier iterations of the Offshore Voluntary Disclosure programs (hereafter referred collectively as "OVD"; see section II above for a more detailed description of the chronology of the various iterations of this program) was produced by the Government Accountability Office (GAO) in 2014 (GAO (2014)). The GAO report makes several points that were amplified by later academic research.

First, GAO notes that almost all of the participants in the 2009 iteration of OVD received the maximum penalty possible under the program. Additionally, the authors note that most of these accounts were high-value accounts that were located in Switzerland. Moreover, they note that the 2012 iteration of OVD had broader reporting requirements for participants; GAO suggests that this is because the data the IRS received from earlier iterations of OVD did not provide sufficient information to fully understand the landscape of overseas noncompliance. Finally, GAO notes that there is reason to suspect that there were many U.S. taxpayers with unreported foreign assets who amended their filings for previous years without participating in OVD. This population of so-called "quiet disclosers," GAO suggests, pose a specific risk for noncompliance.

Johannesen *et al.* (2019) dives deeper into this same set of issues and provides analysis that includes several additional years of data. The authors show that OVD did result in a significant increase in compliance, but that they argue that the relatively narrow scope of the program may mean that the more robust enforcement mechanisms of FATCA were, in fact, warranted. The authors make their case by analyzing the enforcement effect that is reflected in data on OVD, in particular, and compare it to the data on populations, such as the "quiet disclosers" that the Government Accountability Office identified that are not included in that group.[10] The authors confirm the GAO's findings that, among those who came into compliance with U.S. tax law on overseas assets during this period, OVD participants tended to have higher-value accounts.[11] They also validate the GAO's concern about the population of quiet disclosers, as they suggest that the total amount of additional tax remitted by quiet disclosers was much larger than the value of additional tax remitted by those who participated in OVD. They further estimate that the combined effect of these programs led to the disclosure of roughly 50,000 additional accounts and $100 billion in new wealth. By way of conclusion, Johannesen *et al.*

[10] Johannesen *et al.* (2019) report that the number of U.S. residents who filed a Report of Foreign Bank and Financial Accounts (FBARs) increased from roughly 40,000 U.S. residents every year from 2005 to 2008 up to 90,000. Given that a mere 30 percent of these new-FBAR-filers participated in the voluntary disclosure program described below, the authors conclude that the majority of these new filers tried to sneak back into compliance via quiet disclosure.

[11] They also confirm that many of the OVD accounts in the early 2009 and 2012 iterations of the program were located in Switzerland, Liechtenstein, and Luxembourg. They attribute this to the fact that financial institutions in these countries were subject to heightened scrutiny and reporting requirements beginning in the late 2000s. For a detailed overview of this background, see Johannesen et al. (2019) p. 1-9).

(2019) suggest that the relatively narrow scope of OVD and the noncompliance that OVD did not address may warrant the more robust reporting requirements of FATCA.

Johannesen *et al.* (2023) looks at the early data that has emerged from FATCA to show the additional light that this initiative shines on the details of U.S. taxpayer assets held overseas. By requiring that foreign financial institutions report on the owner(s), holdings within, and particular uses of the accounts of U.S. taxpayers overseas, FATCA greatly increases the amount of information available to the IRS about U.S. taxpayers' overseas wealth. Thus, despite issues with data quality and additional potential forms of noncompliance that FATCA does not address, the authors conclude that FATCA and other administrative data show that U.S. taxpayers hold roughly $4 trillion dollars overseas. They further note that this ownership is 1) highly concentrated at the very top (0.01 percent) of the income spectrum and that 2) roughly $2 trillion of these assets are held in traditional tax havens. They conclude by suggesting the need to better understand the effect that FATCA has on the levels of voluntary compliance among taxpayers holding assets overseas.

IV. Graph Construction

We use IRS administrative data, which contains de-identified taxpayer data extracted from filed tax returns, enforcement information, and narrative data to construct a graph. Specifically, we observe data extracted from Form 1065 Schedule K-1, Form 1120S Schedule K-1, and Form 1041 Schedule K-1 to identify taxpayer networks; and from Form 1040 to examine reported individual income and spousal relationships. Finally, to identify taxpayers with foreign accounts, we rely on FBAR data, Forms 8938 and 8966, and voluntary disclosure programs filing information.

Building Out K-1 Graphs

The primary structure of the graph relies on the relationship between K-1 recipients (payees) and K-1 issuers (payers). We began by developing a graph for every year between 2006 and 2020 using data obtained from Forms 1065 Schedule K-1, 1120S Schedule K-1, and 1041 Schedule K-1. Each graph contains two types of nodes that represent payees and payers, as well as edges connecting payees and payers. Each edge contains data on the total payment reported on the K-1, which is used to calculate a proxy for ownership of the issuing entity (i.e., partnership in the case of Form 1065 Schedule K-1, S-corporation in the case of Form 1120S Schedule K-1 and trust for Form 1041 Schedule K-1). Ownership is estimated by dividing the absolute value of the gains and losses reported to a given payee by the sum of the absolute value of all gains and losses issued by the payer. We created a subset graph to keep only edges that represent at least a 1 percent stake by the payee. We took this step to limit the size of the graph for computational purposes, as well as a means to restrict associations between taxpayers to those that are more likely to be significant.

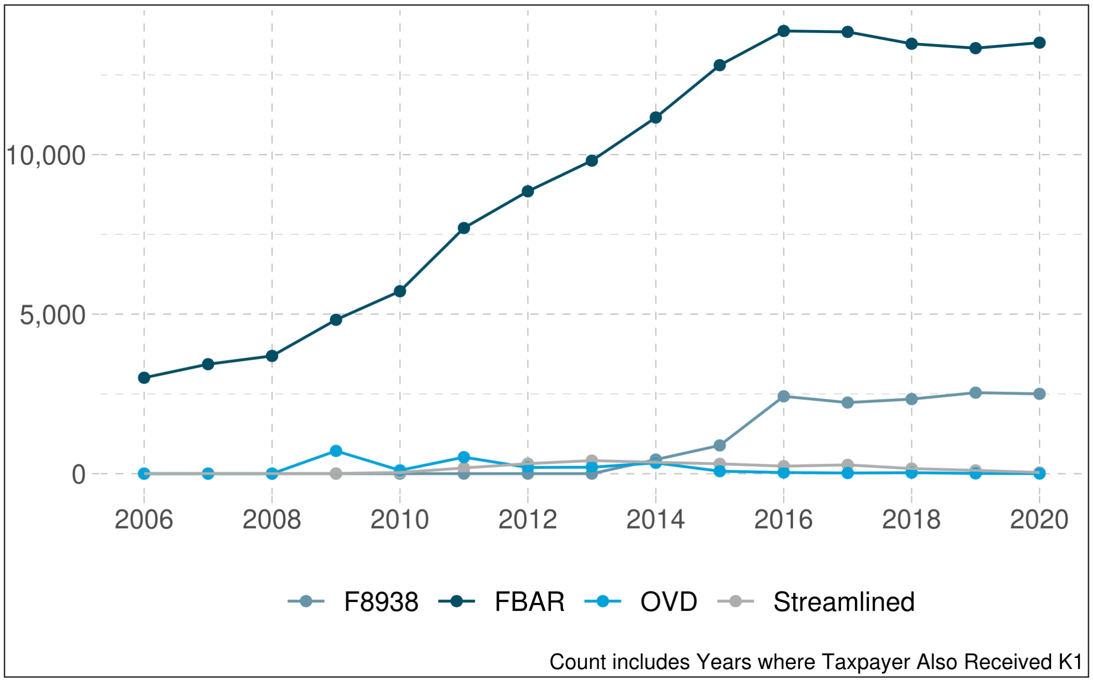
We then constructed a separate graph consisting of taxpayers with reported foreign accounts and their spouses. Taxpayers that reported holding foreign accounts in FBAR filings,[12] Form 8938, and in past OVD programs and streamlined voluntary disclosure programs are added to the graph. Spouse nodes and edges representing spousal relationships were then added to the graph using data from Form 1040. We then cross-referenced this foreign account taxpayer-spouse graph with the K-1 graphs and retained only individual taxpayers with foreign accounts that have received a K-1 at some point between 2006 and 2020. This represented the universe of taxpayers with reported foreign accounts.[13]

---

[12] Foreign account holders include account owners, joint owners, and taxpayers with signature authority, but no interest. In future versions of this work, we will likely take steps to distinguish between different types of account holders.

[13] We did not include Form 1040 Schedule B as a method to identify taxpayers with foreign accounts in an additional step to limit the size of the graph. Unlike FBAR and F8938 requirements, there is no threshold to report a foreign account on a Schedule B, therefore our focus was generally taxpayers with at least $10,000 in their accounts.

FIGURE 1. Count of Taxpayers by Type of Foreign Account Reporting



The final sample of taxpayers with foreign accounts was taken from individual taxpayers with a "significant stake" in at least one K-1 issuing entity from 2006 and 2020. We defined "significant stake" as directly receiving 30 percent[14] of the total absolute value reported by a payer. Figure 1 shows the count of sample taxpayers with foreign accounts and the years that they reported a foreign account and received a K-1. We then selected a sample of taxpayers for our comparison group. This group was made up of individual taxpayers that have never reported holding a foreign account, were never reported to have a foreign account on Form 8966, and were reported to hold a "significant stake" in at least one K-1 issuing entity between 2006 and 2020. Like the foreign account holding sample, we also included the spouses of the nonforeign account holding sample.

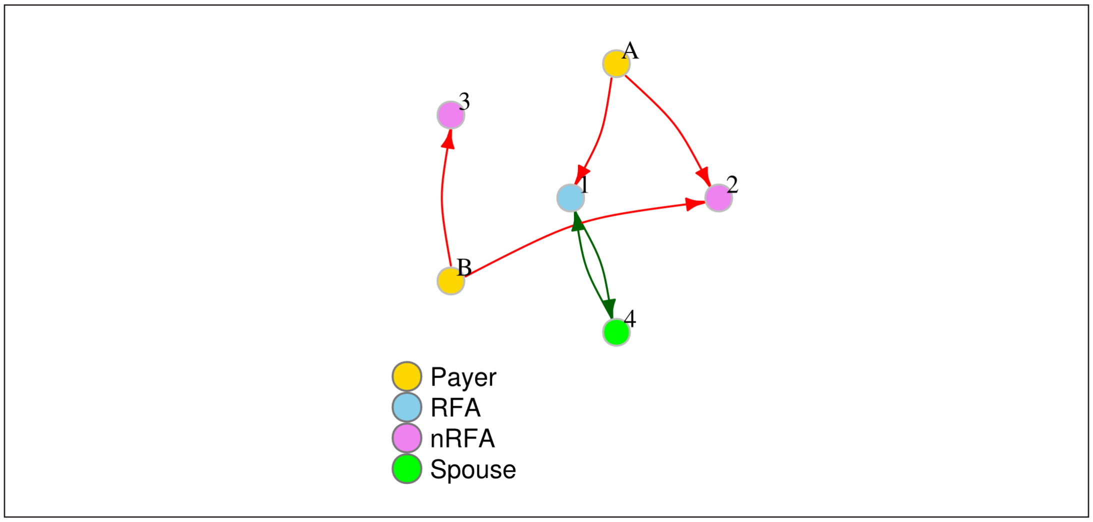
At this stage, we had two groups of taxpayers that will make up our study population-taxpayers with reported foreign accounts (RFA taxpayers) and taxpayers with no reported foreign accounts (non-RFA taxpayers). We then get their K-1 network for every year that a taxpayer has received a K-1. Specifically, we created nodes for all payers that issued K-1s to RFA and non-RFA taxpayers, other payees that received a K-1 from the same payer, and additional payers that issued K-1s to the taxpayer's neighbors. We repeated this process, for up to five levels from the initial taxpayer (see Figure 2). Finally, we took one more step to clean the graph by removing any nodes with over 500 edges. With the foundation of each yearly graph set, we then added metadata to each node from Form 1040, as well as create multiple measures and descriptors of each taxpayer's network.

Figure 2 depicts an example of a fictional network, of an RFA taxpayer we'll refer to as Node 1. Red edges depict K-1 relationships, while green edges show spousal relationships. In a given year, Node 1 received a K-1 from Node A, which also issued a K-1 to Node 2. In addition, Node 2 and Node 3 received a K-1 from Node B. Lastly, Node 1 and Node 4 were spouses in this year, but Node 4 did not receive any K-1s.[15]

[14] Note that the 30 percent significant stake we require for a taxpayer to be included in the sample is distinct from the one percent threshold we set when building the K-1 graph. We make the distinction because when building out a network we place a premium on having as complete a network as possible, taking into account computational restraints; while when we select sample taxpayers, we prioritize ensuring that the taxpayer is an important part of the network.

[15] More precisely, Node 4 did not receive any K-1s where it held at least a one percent stake.

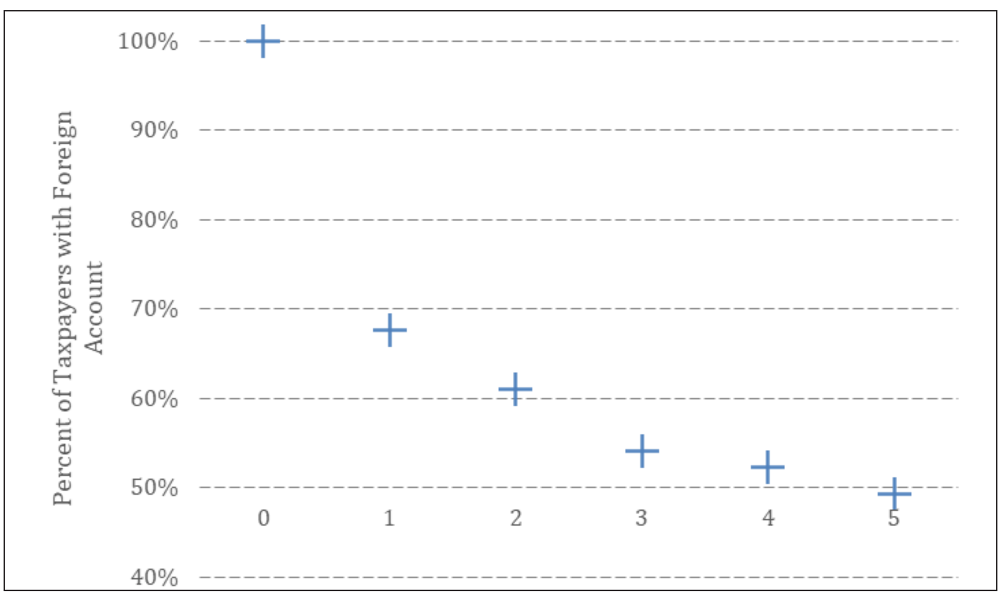
FIGURE 2. Example Network in Graph



V. Graph Content

The final sample contained 48,347 RFA taxpayers and 48,347 non-RFA taxpayers. Among the RFA taxpayers in the sample, 9,114 taxpayers (19 percent) reported a foreign account in only one year between 2006 and 2020, while 582 taxpayers (1 percent) reported a foreign account in all fifteen years. Taxpayers that report foreign accounts in multiple years are more likely to report a foreign account in the years directly following the first disclosure. As shown in Figure 3, among RFA taxpayers that first reported a foreign account between 2006 and 2015, 49 percent also reported a foreign account five years later.

FIGURE 3. Among RFA Taxpayers with First Foreign Account Between 2006-2015: Percent with RFA in Years Following First RFA



Taxpayers with reported foreign accounts generally had higher reported income on Form 1040. This was true across all income types we included. As noted in Table 1, the median adjusted gross income (AGI) for

RFA taxpayers in 2020 was around $240,000, while for non-RFA taxpayers it was around $107,000. The figures presented in Table 1 compare taxpayers who have ever reported a foreign account between 2006 and 2020 with taxpayers who did not, not just in years where they reported a foreign account. This suggests the type of taxpayers who reported foreign accounts and received K-1s are different, at least in terms of reported income, from K-1 recipients who never reported foreign accounts.

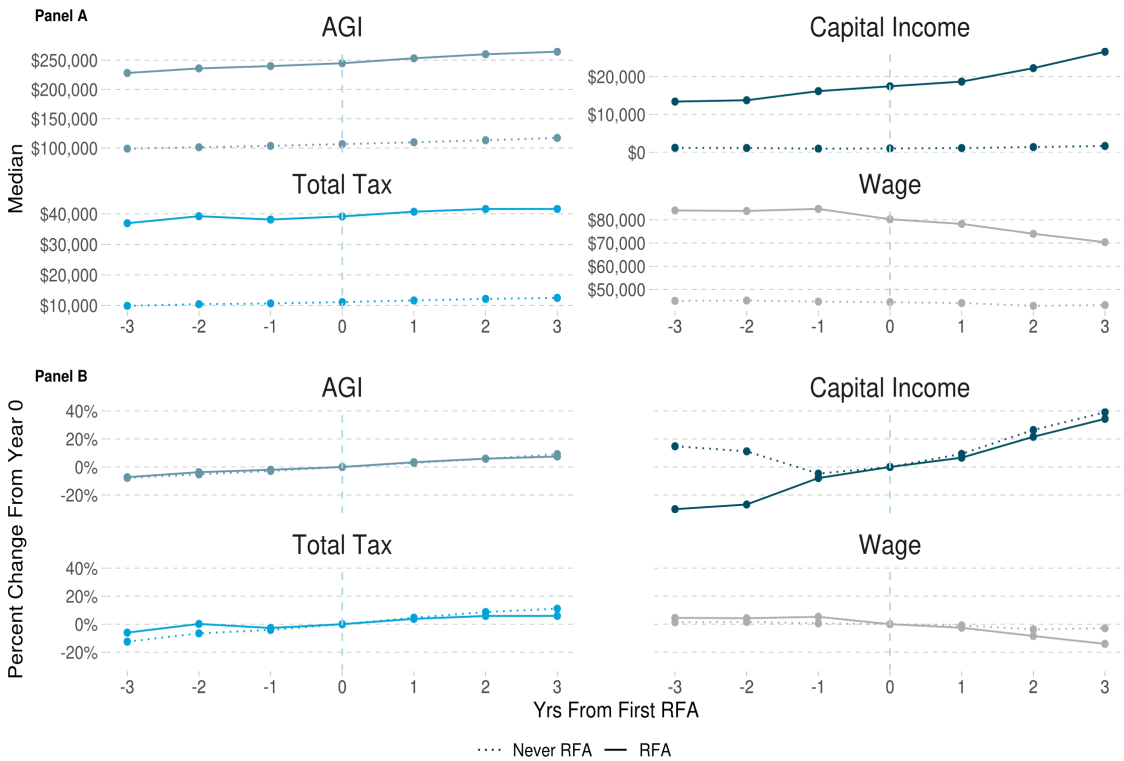
TABLE 1. Adjusted Gross Income by RFA Status, 2006–2020

Year	RFA Taxpayers			Non-RFA Taxpayers		
	25th Percentile	Median	75th Percentile	25th Percentile	Median	75th Percentile
2006	$86,000	$213,000	$650,000	$39,000	$85,000	$172,000
2007	$91,000	$224,000	$679,000	$38,000	$86,000	$175,000
2008	$73,000	$192,000	$545,000	$34,000	$81,000	$162,000
2009	$61,000	$168,000	$459,000	$30,000	$75,000	$149,000
2010	$65,000	$182,000	$508,000	$33,000	$79,000	$156,000
2011	$68,000	$187,000	$526,000	$34,000	$82,000	$163,000
2012	$77,000	$208,000	$602,000	$38,000	$88,000	$177,000
2013	$78,000	$208,000	$556,000	$40,000	$92,000	$179,000
2014	$86,000	$221,000	$600,000	$42,000	$96,000	$194,000
2015	$84,000	$225,000	$604,000	$43,000	$99,000	$199,000
2016	$82,000	$222,000	$572,000	$44,000	$100,000	$201,000
2017	$87,000	$233,000	$614,000	$45,000	$104,000	$210,000
2018	$91,000	$238,000	$624,000	$47,000	$107,000	$220,000
2019	$92,000	$247,000	$629,000	$48,000	$111,000	$228,000
2020	$84,000	$240,000	$638,000	$46,000	$107,000	$228,000

Note: Due to confidentiality concerns, all income amounts have been rounded to the nearest thousand dollars.

There does not appear to be much evidence of a change in reporting behavior among RFA taxpayers after first reporting a foreign account. In Figure 4, we look at RFA taxpayers who first received a foreign account between 2009 and 2017 and received a K-1 for the six years surrounding the first reported RFA to study whether reported income changes in the years after first reporting a foreign account relative to the years before. We then compared the results with non-RFA taxpayers to help ensure that any trends identified were not simply a result of the passage of time. We randomly assigned all non-RFA taxpayers a value between 2009 and 2017 and kept taxpayers with K-1s in all three years before and after that date. Panel A clearly shows the difference in median reported income and total tax among individuals who ever reported a foreign account and those that did not. Panel B shows the percent change in each category relative to year zero (defined as the first year with an RFA for RFA taxpayers and the randomly assigned value for non-RFA individuals).

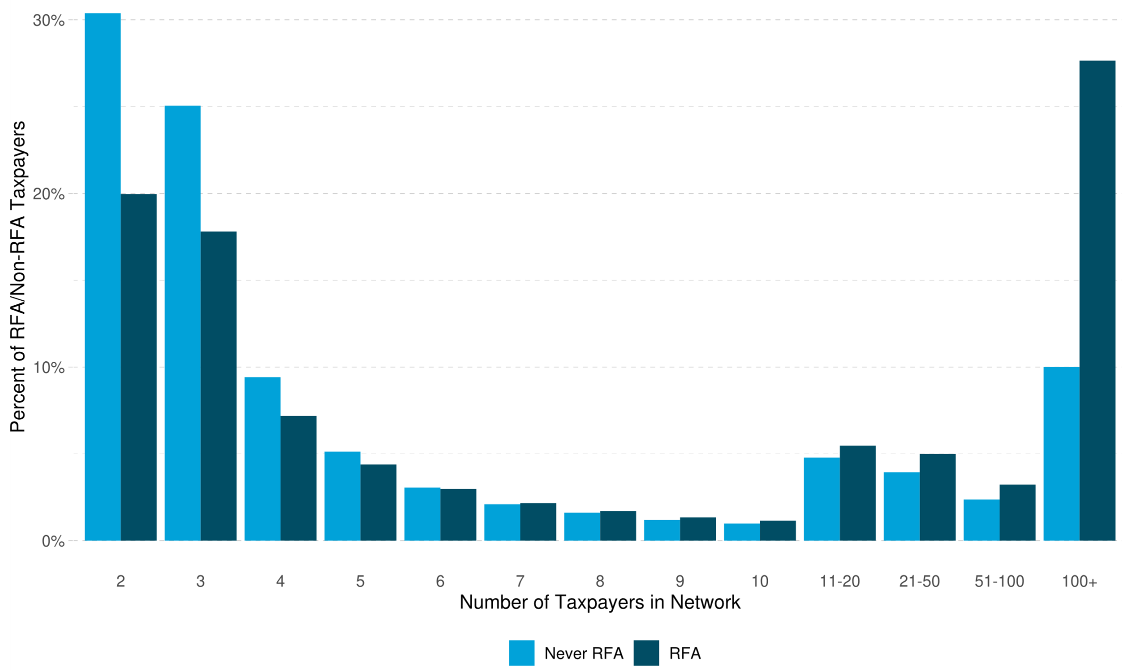
FIGURE 4. Reported Income for RFA and Non-RFA Taxpayers Over Time



In one final look at reported income surrounding first disclosing a foreign account, we focus in Figure 5 only on taxpayers who have at some point reported an offshore account. Here, we compared taxpayers who have reported a foreign account in a specific year with those who did not report one. Figure A1 in the Appendix shows once again that while there was some divergence between the two groups, it is not evident there was a change in reporting behavior in the years surrounding when a foreign account was reported. We hope to conduct a more rigorous statistical analysis to attempt to answer this question in future iterations of this work.

Similar to reported income, there does appear to be a difference in the network structures between RFA and non-RFA taxpayers. Networks of taxpayers who have ever reported a foreign account are larger both in the number of taxpayers and dollars. The median network among RFA taxpayers contains 6 taxpayers and has reported $268,000 flowing through the network, considerably larger than 3 and $49,000 for non-RFA taxpayers. As the median network size suggests, most K-1 networks contain only a handful of taxpayers. Around 55 percent of taxpayers who never reported foreign accounts contained fewer than four taxpayers. As is evident in Figure 5, while the network of a plurality of RFA taxpayers also contained three or fewer taxpayers, RFA taxpayers were far more likely to be in networks of over 100 entities.

FIGURE 5. Network Size by Whether Taxpayer Reported Foreign Account in a Given Year



VI. Modeling Approach and Results

While modeling directly from graphs is a growing area of research and would be worthy of exploration in this context in the future, we opted to capture elements of the graph in a flat file and. at that point, developed the models we used in this paper. Converting the graph into a flat file required creating numerous variables that adequately represented the relationships and other insights evident in the graph. The variables created from the graph and used for modelling were divided into three different categories: 1) network variables; 2) taxpayer Schedule K-1 variables; and 3) F1040 reported income variables. Network variables included descriptions of the taxpayer's K-1 network including the composition and size of the network (e.g., number of taxpayers, percent of network payers that are partnerships, multitiered pass-through entities as a percentage of network size) and foreign account disclosure by network members (e.g., whether the network contains any RFA owners or payers). Taxpayer K-1 variables directly described the relationship between the taxpayer and other members of its network (e.g., number of K-1s received, K-1s received from multitiered pass-through entities as well as from RFA taxpayers, the ratio of reported profits and losses on incoming Schedule K-1s). Lastly, F1040 variables captured amounts and types of income reported by the taxpayer on Form 1040. For F1040 values with negative values, we took the absolute value, as taxpayers with large reported losses often share characteristics with taxpayers reporting large gains. In addition, due to the large range of income variables, we divided each one into deciles of the absolute value of each income type. Deciles proved to be better predictors of reporting a foreign account than the amount itself. A complete list of network variables used for modelling is provided in Table A1 in the Appendix.

A primary objective of this work is to shed light on the relationship between a taxpayer's network and their likelihood to report a foreign account. To do this, we estimated a series of logistic regressions with individual and year fixed effects of the form:

$$y_{it}=\alpha+\beta x_{it}+\delta_t+\gamma_i+\mu_{it},$$

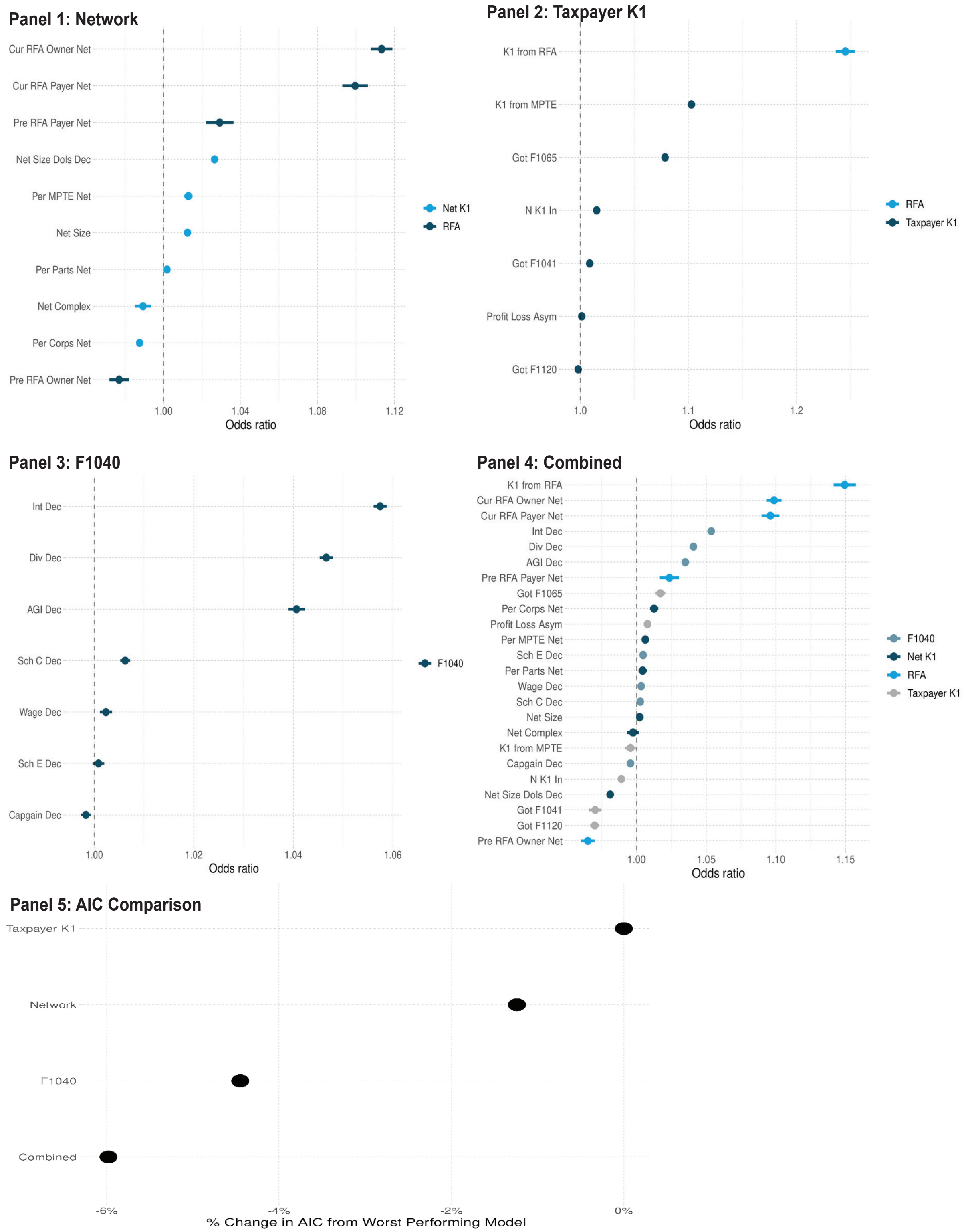
where $y_{i\,t}$ is the likelihood of whether taxpayer reported a foreign account in time *t*. We first estimated the specification separately for each of the groups of covariates described above. Therefore *X* is a vector of either network, taxpayer Schedule K-1 or F1040 variables for taxpayer *i* in time *t*, while *β* represents a series of coefficients for each individual covariates in *x*. The specification also includes year δ_t and individual γ_i fixed effects. All continuous variables were scaled to allow for comparison across different coefficients. Continuous variables were scaled by first subtracting the variable's mean from each observation and then dividing by the variable's standard deviation. Finally, we combined all the covariate groups and estimated an additional regression of the form:

$$y_{it}=\alpha+\beta 1net_{it}+\beta 2taxpayerk1_{it}+\beta 3f1040_{it}+\delta_t+\gamma_{it}+\mu_{it}$$

Panels 1−3 in Figure 6 show the results of the three regressions with the individual groups of variables. The coefficients have been transformed using the logistic function to represent the odds ratio for each variable. For example, having a current RFA payer in their network, holding all other values constant, increase the odds of reporting a foreign account in a given year by 1.1 compared to taxpayers without a current RFA payer in their network. In other words, holding all other variables at a fixed value, the odds of reporting a foreign account increases by 10 percent when the taxpayer's network contains a payer that reported a foreign account in the same year. The largest coefficient was in the taxpayer K-1 model and represented a 25 percent in the odds of reporting a foreign account when the taxpayer received a K-1 from an RFA taxpayer. All continuous variables have been scaled to facilitate comparing coefficients; however, it is important to keep in mind that a 0/1 change in an indicator variable may not be the same thing as a one standard deviation change in a scaled continuous variable.

We used the Akaike Information Criterion (AIC) to compare the goodness of fit of the models. AIC takes into account the likelihood of obtaining the observed data under the assumption that the model is correct, while also penalizing a model for containing more variables to account for overfitting. A lower AIC represents a better fitting model. Panel 4 in Figure 5 compares the AIC of the models specified. The model on the top row, the taxpayer K-1 model, was the worst performing model using this metric. Of the models containing just one group of variables, the F1040 models performed best, with an AIC 4.45 percent lower than the taxpayer K-1 model.

FIGURE 6. Grouped Variables Regression Results



The combined model, which included all sets of variables is the best performing. While it may appear intuitive that the specification that conveys the most information is the best performing, AIC takes into account the model complexity and penalizes additional covariates. Nevertheless, the combined model's AIC is 6 percent lower than the worst performing model. Table 3 in the Appendix presents the results of the combined regression. The analysis found that sharing a network with other taxpayers with reported foreign accounts is positively associated with reporting a foreign account. This confirms our hypotheses that considering not only a given taxpayer, but also the characteristics of their network is of informational value. This does not suggest that individual attributes should be ignored. We find that reporting higher income–(especially capital income) and receiving income from a partnership (as opposed to an S-corporation and trust) are also positively associated with reporting a foreign account in that year.

VII. Future Work and Conclusion

This work demonstrates the value in considering a taxpayer's K-1 network when assessing their likelihood for noncompliance through failing to report a foreign account. We merely scratched the surface of analyses that are possible by using a graph framework and other available taxpayer data in this space. For example, we included a relatively small amount of variables from a limited set of categories. We did not include additional taxpayer characteristics such as whether the taxpayer had a foreign address, whether they filed an amended return, country of foreign account, or their audit history. Moreover, other than a few notable exceptions (e.g., individual and year fixed effects, looking at whether related taxpayers had previously had foreign accounts) we did not fully exploit the time aspect of the data. On a related note, we took only a cursory look at how taxpayer's reporting behavior change after first reporting a foreign account; future work can dig deeper into this question.

We faced various challenges when compiling the data. We faced data challenges that are inherent with working with such a large and complex data set. To deal with some of these issues we trimmed the data by introducing thresholds at various stages (described in more detail above). Subsetting the graph has two main benefits: first, it allows us to focus on taxpayers that are more likely to be closely connected; and second, working with multiple years of full K-1 data is computationally expensive. However, this approach does have drawbacks such as arbitrarily flattening the networks by limiting the number of payees and levels. This resulted in less variation among the networks and may have biased our results towards representing network structure and neighbor characteristics as less significant than they actually are. Future work could experiment with constructing a graph that strikes a different balance between the trade-off between size and comprehensiveness.

We presented evidence that a relationship exists between a taxpayer's K-1 network and reporting a foreign account. Taxpayers with other payers and payees who report a foreign account are more likely to themselves report holding a foreign account in that same year. While network and K-1 variables are associated with reporting foreign accounts, the model using only F1040 variables is the best fitting regression specified using only a selection of variables. This, along with the results from the combined model, suggests that a combination of taxpayer and network characteristics is necessary to gain an understanding of reporting foreign accounts. In addition to exploring different taxpayer characteristics, future work can build on the framework we have laid out and develop predictive models using machine learning methods such as random forests or graph neural networks. In this work we lay a foundation for the work that is possible in this space and hope to continue to develop methods to identify possibly noncompliant taxpayers and further aid in tax administration.

Appendix

Figure A1 details income reporting behavior of reported foreign accounts (RFA) taxpayers that received a Schedule K-1 for six consecutive years surrounding the first reported RFA. We compared taxpayers who have reported a foreign account in a specific year with those who did not (but did in another year).

FIGURE A1. Reported Income Among RFA Taxpayers by Whether Reported Foreign Account in Specific Year

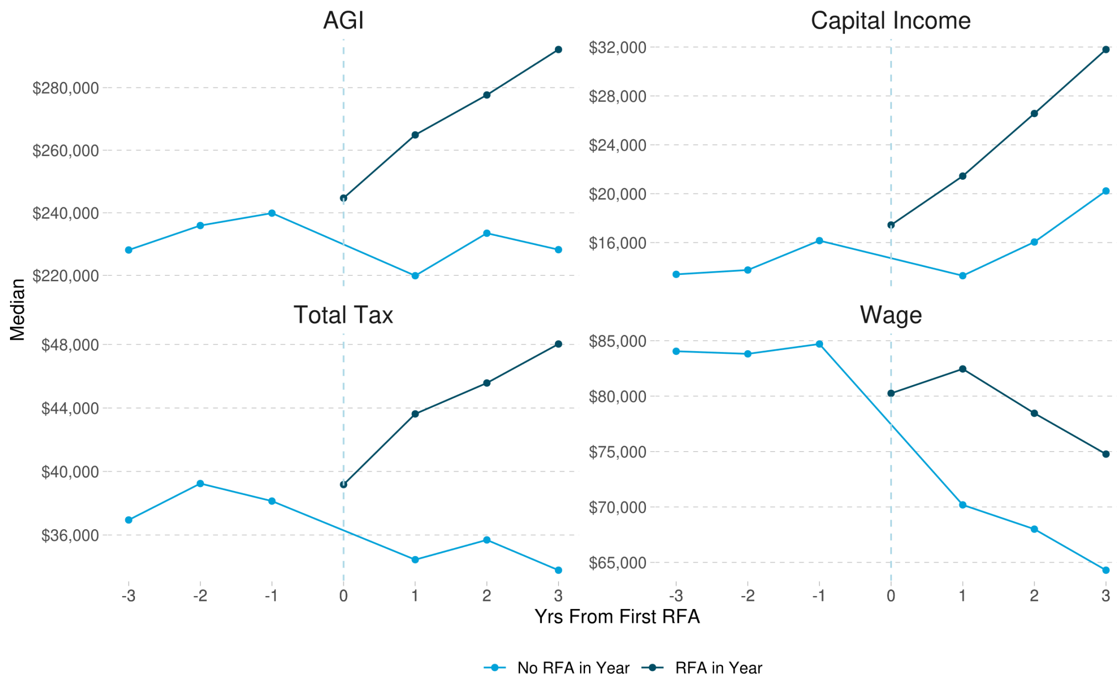


Table A1 details the variables used in the different models. To address extreme values, we winsorized the top .01 percent and (where applicable) the lowest .01 percent of F1040 reported income to deal with extreme values. In addition, we top-coded the absolute value of the K-1 dollar flows in the network to $1,000,000,000. In both cases, these steps were taken before dividing the data into deciles.

TABLE A1. Variable Descriptions

Variable	Description	Category
Div Dec	F1040	Decile of reported dividend income
Wage Dec	F1040	Decile of wage income
AGI Dec	F1040	Decile of the absolute value of adjusted gross income
Int Dec	F1040	Decile of the absolute value of reported interest income
Capgain Dec	F1040	Decile of the absolute value of capital gains income
Sch C Dec	F1040	Decile of the absolute value of Sch. C income
Sch E Dec	F1040	Decile of the absolute value of Sch. E income
Net Size	Net K-1	Total number of unique entities in network
Per Parts Net	Net K-1	Percent of network payers that issued 1065 Schedule K-1
Per Corps Net	Net K-1	Percent of network payers that issued 1120s Schedule K-1
Per MPTE Net	Net K-1	Pass-through entities as a percentage of the network size
Net Size Dols Dec	Net K-1	The decile of the absolute value of the K-1 flows in the network.
Net Complex	Net K-1	A network is complex if it contains a pass-through entity or a business entity as a payee. A network is simple if all payers issued K-1s directly to individuals.
Cur RFA Owner Net	RFA	Indicator whether there was an owner with a reported foreign account in the network in the current year, excluding the taxpayer and his or her spouse.
Pre RFA Owner Net	RFA	Indicator where there was an owner with a reported foreign account in a previous year, excluding the taxpayer and his or her spouse.
Cur RFA Payer Net	RFA	Indicator whether there was a payer with a reported foreign account in the network in the current year.
Pre RFA Payer Net	RFA	Indicator whether there was a payer with a reported foreign account in a previous year.
K-1 from RFA	RFA	Indicator whether the taxpayer received a K-1 from a payer with a foreign reported account in the current year.
Got F1065	Taxpayer K-1	Indicator whether the taxpayer received a 1065 Schedule K-1.
Got F1120	Taxpayer K-1	Indicator whether the taxpayer received a 1120s Schedule K-1.
Got F1041	Taxpayer K-1	Indicator whether the taxpayer received a 1041 Schedule K-1.
N K-1 In	Taxpayer K-1	Number of K-1s received
Profit Loss Asym	Taxpayer K-1	Sum of the absolute value of the difference between the percentage allocation of profits and percentage allocation of losses.
K-1 from MPTE	Taxpayer K-1	Indicator whether the taxpayer received a K-1 from a multitiered pass-through entity.

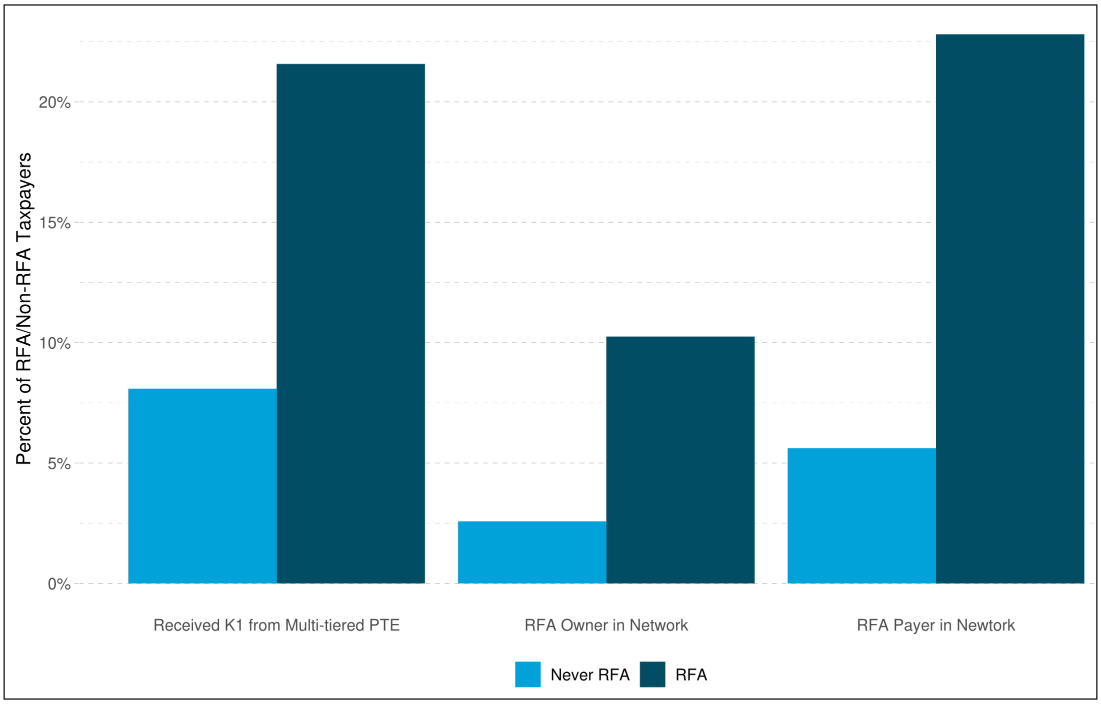
TABLE A2. Results of Fixed Effect Logit Models
DEPENDENT variable: RFA in Year

Variable	Network (1)	Taxpayer K-1 (2)	F1040 (3)	Combined (4)
Per MPTE Net	0.013*** (0.001)			0.006*** (0.001)
Net Size	0.012*** (0.001)			0.002*** (0.001)
Net Complex	-0.011*** (0.002)			-0.003 (0.002)
Per Parts Net	0.002** (0.001)			0.004*** (0.001)
Per Corps Net	-0.013*** (0.001)			0.012*** (0.002)
Net Size Dols Dec	0.026*** (0.001)			-0.019*** (0.001)
Cur RFA Owner Net	0.107*** (0.003)			0.094*** (0.003)
Cur RFA Payer Net	0.095*** (0.003)			0.092*** (0.003)
Pre RFA Owner Net	-0.023*** (0.003)			-0.036*** (0.003)
Pre RFA Payer Net	0.029*** (0.004)			0.023*** (0.003)
N K-1 In		0.015*** (0.001)		-0.011*** (0.001)
K-1 from MPTE		0.098*** (0.002)		-0.004** (0.002)
K-1 from RFA		0.219*** (0.004)		0.139*** (0.004)
Profit Loss Asym		0.001* (0.001)		0.008*** (0.001)
Got F1120		-0.002 (0.001)		-0.031*** (0.002)
Got F1041		0.009*** (0.002)		-0.030*** (0.002)
Got F1065		0.075*** (0.001)		0.017*** (0.002)
AGI Dec			0.040*** (0.001)	0.034*** (0.001)
Int Dec			0.056*** (0.001)	0.052*** (0.001)
Div Dec			0.046*** (0.001)	0.040*** (0.001)
Wage Dec			0.002*** (0.001)	0.003*** (0.001)
Sch C Dec			0.006*** (0.001)	0.003*** (0.001)
Sch E Dec			0.001 (0.001)	0.005*** (0.001)
Capgain Dec			-0.002*** (0.001)	-0.004*** (0.001)
AIC	783096.51	792926.51	757647.04	745516.38
Observations	704,857	704,857	704,857	704,857
R^2	0.050	0.036	0.084	0.099

Note: *p<0.1; **p<0.05; ***p<0.01.

Figure A3 shows the differences between the networks of RFA and non-RFA taxpayers across different dimensions. A higher percentage of taxpayers with foreign accounts received a K-1 from a pass-through entity in a multitiered network, had another RFA owner or an RFA payer in its network.

FIGURE A3. Differences Between RFA and non-RFA Taxpayer Networks



References

Agarwal, Ashish, Shannon Chen, and Lillian F Mills (2021). "Entity Structure and Taxes: An Analysis of Embedded Pass-Through Entities." The Accounting Review 96.6, 1–27.

Agarwal, Ashish, Shannon Chen, Ririko Horvath, Larry May, and Rahul Tikekar (2015). "Analysis of Flow-Through Entities Using Social Network Analysis Techniques." IRS TPC Research Conference, Washington D.C.

Black, Emily, Ryan Hess, Rebecca Lester, Jacob Goldin, Daniel E. Ho, Mansheej Paul, Annette Portz (2023). "The Spiderweb of Partnership Tax Structures." IRS Working Paper.

Cooper, Michael, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick (2016). "Business in the United States: Who Owns It, and How Much Tax Do They Pay?" Tax Policy and the Economy 30.1, 91–128.

Government Accountability Office (2014). "Offshore Tax Evasion: IRS Has Collected Billions of Dollars, but May be Missing Continued Evasion." GAO Publication No. 13-318.

Internal Revenue Service (2022). "Report of Foreign Bank and Financial Accounts (FBAR)."

https://www.irs.gov/businesses/small-businesses-self-employed/report-of-foreign-bank-and-financial-accounts-fbar.

Internal Revenue Service (2022). "Report of Foreign Bank & Financial Accounts (FBAR) Reference Guide." IRS Publication 5569.

Internal Revenue Service (n.d.). "Streamlined Filing Compliance Procedures–A Compliance Option for Some Taxpayers" https://www.irsvideos.gov/business/FilingPayingTaxes/StreamlinedFilingComplianceProceduresAComplianceOptionForSomeTaxpayers.

Internal Revenue Service (2022). "Summary of FATCA Reporting for U.S. Taxpayers." https://www.irs.gov/businesses/corporations/summary-of-fatca-reporting-for-us-taxpayers.

Internal Revenue Manual (IRM) § 4.63.3.1

Johannesen, Niels, Patrick Langetieg, Daniel Reck, Max Risch, Joel Slemrod (2019). "Taxing Hidden Wealth: The Consequences of U.S. Enforcement Initiatives on Evasive Foreign Accounts." NBER Working Paper No. w24366.

Johannesen, Niels, Daniel Reck, Max Risch, Joel Slemrod, John Guyton, and Patrick Langetieg (2023). "The Offshore World According to FATCA: New Evidence on the Foreign Wealth of U.S. Households." NBER Working Paper No. w31055.

Love, Michael (2021). "Where in the World Does Partnership Income Go? Evidence of a Growing Use of Tax Havens." Working Paper.

OECD (2013). "A Step Change in Tax Transparency." OECD Report for the G8 Summit.

Olson, Matt, Annette Portz, Mike Feldman, Devika Mahoney-Nair (2022). "Graph-based machine learning methods for case selection and population segmentation." IRS Working Paper.

Application of Network Analysis To Identify Likely Ghost Preparer Networks

Joshua W. King, Andrew J. Soto, Getaneh Yismaw, Izabel Doyle, Ririko Horvath, Ashley Nowicki, Chris Hess (IRS, Research, Applied Analytics & Statistics), Brandon Gleason (IRS, Criminal Investigations), Will Sundstrom, Jacob Brooks, Michael Mastrangelo, Mike Stavrianos, Daniel Hales (GCOM)

Introduction

More than half of individual taxpayers rely on paid tax return preparers to assist them in meeting their federal tax filing obligations. In Filing Year 2022, the IRS received 150,605,162 electronically filed individual returns, 85,885,109 (57 percent) of which were professionally prepared. Paid preparers are important IRS partners, as the Service depends on them to help taxpayers comply with tax laws. Identifying noncompliant preparers or groups of preparers who actively hide their identities, i.e., ghost preparers, is an essential component of the IRS's tax administration responsibilities. This paper describes ongoing efforts within the IRS to apply graph techniques and network analysis to: (1) identify clusters of suspected ghost prepared returns; (2) understand how ghost preparation unfolds during the filing season; and (3) study the impacts of ghost preparation on tax compliance.

Ghost Preparer Risk

Ghost preparers represent a risk to the integrity of the U.S. tax system. By not identifying themselves on the returns they prepare, they are not subject to oversight, and they are in violation of treasury rules and regulations. In addition, ghost preparers may engage in abusive tax practices which harm taxpayers and undermine the efficacy of the IRS. Compounding this concern, an individual ghost preparer may be responsible for tens if not hundreds or thousands of returns, giving any fraudulent or abusive behavior an outsized impact. According to the Treasury Inspector General for Tax Administration (TIGTA), ghost preparers disrupt and destabilize established IRS practices, legitimate preparers, and the taxpayer ecosystem (TIGTA (2018) and IRS (2009)).

Internal Revenue Manual (IRM) 25.20.2.1.4 defines a ghost preparer as a compensated tax return preparer who does not provide a Preparer Tax Identification Number (PTIN) on the returns they prepare as required by Internal Revenue Code (IRC) section 6109 and Treasury Regulation 1.6109-2. Ghost preparers may fail to include any identifying number on prepared returns, or they provide a number other than a PTIN in place of an appropriate PTIN, such as a series of random numbers, a Taxpayer Identification Number (TIN), a PTIN issued to another preparer, a made-up PTIN, etc. A PTIN must be obtained by all return preparers who are compensated for preparing or assisting with any U.S. federal tax return, refund claim, or other tax forms submitted to the IRS (unless specifically exempted).

Ghost preparers may intentionally or unintentionally hide their identity. For preparers who are unaware of the rules and requirements around PTINs, it is likely they are not qualified to provide advice and assistance to their clients. They may unknowingly put taxpayers at risk of not meeting their tax obligations and, in some cases, audit by the IRS.

For preparers who knowingly hide their identity, there are many motivations to do so. Ghost preparers may prey on taxpayers, stealing refunds or engaging in potentially illegal or fraudulent preparation strategies. Even when the taxpayer is collaborating with the ghost to falsify returns to maximize refunds or claim unearned credits, the ghost preparer enables that illegal behavior. Ghost preparers may lie about their activities to avoid their own tax liabilities or because they've already been identified as a problematic preparer and they are no longer allowed to prepare taxes. There are unknown risks as well; ghost preparers are likely engaged in schemes that have not yet been identified.

Innovation Lab

The Innovation Lab is an initiative at the IRS to encourage collaboration across the service on specific administrative or compliance challenges. In Fiscal Year 2021 the Innovation Lab sponsored research to apply network analysis to detect and identify ghost preparers.[1] The objectives of the Innovation Lab were to explore multiple network approaches to identify ghost prepared returns and to develop a tool for compliance staff to access and investigate these networks for treatment. Ghost preparers are inherently difficult to identify because they do not identify themselves on their clients returns. Many ghost preparers rely on do-it-yourself (DIY) software and not professional software that is more tightly regulated. Thus 100 returns completed by a ghost preparer may look like 100 individually filed returns. Network analysis offers the promise of identifying related returns from a ghost preparer by linking filings and revealing commonalities which point to a single preparer. Over the course of Innovation Lab-sponsored research, analysts delivered a dataset of individual income tax returns (Form 1040) networked across a range of filing and return characteristics for three filing years.

The lab delivered two network clustering approaches for grouping electronically filed self-prepared Form 1040 returns together into return clusters which may suggest ghost preparer involvement. Going forward, those results should be updated with more recent data, additional data sources to build out the context of the cluster networks, and for additional inputs to clustering approaches.

The lab produced a tool designed specifically to deliver ghost preparer results to compliance and enforcement staff. The tool includes features to prioritize and interrogate suspected ghost preparer return clusters to connect analytical outputs with field work.

Network Analysis

The concept of a network refers to "[an] object composed of elements and interactions or connections between these elements" (Brandes (2005)). Calculations in a network model are referred to as network analysis or graph analysis. Network analysis is useful in identifying interconnected groups of entities or clusters. Data stored in a network model can reveal patterns or relationships that were not previously apparent.

Key concepts of a network model are nodes and edges. Entities are referred to as nodes and can encompass wide range of things, depending on the application of that model. In the IRS context, nodes are generally derived from tax forms. Relationships are referred to as edges. Edges connect pairs of nodes and represent the relationship between them. Edges can contain information about the relationship between those two nodes and may also convey information about the directionality of the relationship.

A key concept of network analysis is the idea of clusters. Clusters are groups of nodes that are connected to each other within the network. Depending on the algorithm used or the structure of the network, clusters can represent meaningful sub-networks of the larger network dataset. The structure of the cluster (the size and distribution of nodes and edges) can convey information about which clusters are significant. In addition, it is possible to conduct calculations across the properties stored on nodes and edges to generate insights into clusters.

Ghost Preparer Workflow

The network analysis segment of the Ghost Preparer Project leveraged existing IRS data to reorient available tax return information into a network format to detect potentially ghost prepared returns. The modeling conducted thus far follows a general process with three avenues for altering and targeting clustering results: choosing data, establishing a network model, and applying clustering algorithms.

[1] Currently, ghost preparer compliance treatments involve ad-hoc referrals or cases picked up in related compliance efforts. This research is an attempt to systematically identify likely ghost preparer networks.

The process begins with identifying the datasets; the IRS maintains tax return information in a variety of formats across a range of systems. In addition to return information, the network model can incorporate additional datasets which the IRS maintains or has access to. Decisions regarding the data included in the model have significant effects on the modeling outcomes.

Once the data have been selected, the next step is to choose which elements to include in the network model. Here decisions are made regarding what should be a node and what should represent relationships between returns and what should be stored as a property. In addition, this provides an opportunity to limit elements added to the network model based on number of connections or other characteristics. It also provides the opportunity for additional data manipulation, such as normalizing data elements.

The final step and final avenue for targeting results in the general workflow is the application of the clustering model. There are several models available using a range of computing tools. Decisions about which modeling tool to use can have significant implications for the results generated, including whether results are deterministic or probabilistic, as well as the average size of clustering results. Current analysis relies heavily on the connected component algorithm, which is a deterministic clustering approach, and can be run against a given network and returns all distinct subnetworks.

Benefits of Network Analysis

Ghost preparation is characterized by complex networks of relationships between individuals involved; a network model can capture these relationships. By analyzing the structure of the network, we can identify patterns which suggest ghost preparation. In addition, data in a network format lends itself to seeing second and third order connections between returns which facilitates connecting potential ghost preparers to the clusters of returns identified.

Clustering Approaches

Approach 1. Risk-based

The Risk-based clustering approach scores returns and relationships to refine community detection algorithms to return actionable clusters of returns with limited false positives. The Risk-based approach provides analysts the ability to tune the analysis to focus on specific noncompliant behavior, suspicious behavior, and known schemes undertaken by ghost preparers. By extension, it also controls for returns which would be addressed using compliance programs outside ghost preparation, namely identity theft (IDT). In addition to targeting, the scoring removes spurious connections, which limits false positives or interconnected groupings of noncompliant preparers.

Background

The Risk-based approach evolved from IDT detection efforts that grouped returns by submission characteristics, which together are unique for most filers. Once grouped, the returns were assigned risk scores for a set of IDT indicators, which allowed for filtering returns to a set of suspicious and inter-connected returns. From this limited set of returns, additional related returns could be identified by considering a wider set of linking factors. This approach effectively finds groups of returns that indicated IDT behavior and, unexpectedly, also identified returns associated with ghost preparer behavior.

The Risk-based approach aims to address a series of challenges in interacting with networked returns to detect ghost preparers. A pure networking approach would allow for links to be drawn from all specified data fields within a tax return, but can create large unmanageable clusters, called super clusters. Super clusters (e.g., clusters exceeding 10k returns) can be formed when spurious factors link several small clusters together, or when return level data errors are present. Traditional networking algorithms alone can effectively identify clusters within graphs, but the presence of a cluster does not necessarily imply ghost preparation. While common factors can be used to identify ghost preparers, they can also generate false positives.

In the Risk-based approach, risk factors were identified from workstreams in IDT and from stakeholders. Many of the risk factors are centered around the use of duplicate information from the tax return submitted by the taxpayer/ghost preparer. In a group of properly self-prepared returns, a reasonably high degree of variability is expected, and it is suspicious for a large group of returns to share information. The more occurrences of these shared factors, the riskier the return/group of returns is/are.

The Risk-based approach evolved from IDT detection efforts used during the Economic Impact Payments stimulus program. The idea was to take returns and group them. Once grouped, risk scores could be assigned for predetermined factors indicative of IDT or anomalous behavior. Once risk scores were assigned, the data could be filtered to select the groups of returns with the highest scores. From there, a set of related returns could be built. This yielded manageable cluster sizes and a manageable number of clusters that were indicative of IDT behavior. An unexpected result of this approach was that it also showed returns with ghost preparer behavior.

Mechanics

First Iteration: The approach began by selecting electronically filed DIY returns and running them through various steps to create a set of indicator variables. Indicator variables were binary and represented if the factor was present or not present. Some factors were used to calculate an individual-level score, and that was used in subsequent steps. Once indicator variables were created, they were grouped and scoring factors were calculated. Most scoring factors were assigned scores based on the percent of that factor within a given group of returns. Scoring was comprised of a score from 0-4, with the higher the score the more present a factor is in each grouping (Table 1).

TABLE 1. Scoring Factors for Risk-Based Approach

Score	Criteria Percent
0	Not present/not calculable
1	<26%
2	≥ 26% and < 51%
3	≥51% and < 76%
4	≥76%

Once scores were assigned, groups of returns were selected using predetermined criteria. From there, the groups were transformed back into return-level data and duplicates were removed. Post-transformation data represented returns that were a part of groups that were considered suspicious based on return-level and group-level characteristics. This data was the starting point for building networks of related returns.

Having selected an initial set of suspicious returns, the approached then iteratively added additional returns with shared factors found in the suspicious set. This process was five times to build a full set of returns considered for the approach. With the full set of returns data was cleaned for formatting issues and known data exclusions and then reformatted into a networked graph. From those networked results, distinct subnetworks or clusters were identified using a connected components algorithm. Linking factors were determined based on the potential for noise to be introduced into the results. More specific linking factors were chosen to eliminate noise in the results and make for more manageable data and results representative of a ghost preparer.

Second Iteration: A second iteration of the Risk-based approach was created to improve community/cluster scoring and better address super clusters. The first step of the improved approach remains the same. Through a series of steps, indicator variables were created that would later become part of the risk score and were almost entirely the same as the indicators created for the first iteration of the Risk-based approach with additional focus on normalizing linking factors.

The updated approach builds a network using the normalized linking factors and generated clusters of results using the connected components algorithm. Initial clusters of returns were assigned a cluster name, and scores were calculated on risk factors using the same scoring method shown in Table 1. Returns were then filtered using a predetermined set of thresholds. Communities of returns that met the scoring criteria flagged and recorded (first run).

Returns from the first run were queried and clusters that had a return count of more than set threshold were separated. These large groupings were restaged as a network and linking factors which were either highly connected or had limited usage were removed. A second round of the community detection algorithm was run against this limited network generating a final set of clusters, with a limited number of super clusters.

Results

Results from the first iteration of the Risk-based approach identified 1,680 clusters with a total of 255,892 returns. The average cluster size of the returns from the first iteration of the Risk-based approach was 127.52 when excluding the single super cluster, and 152.32 when including the super cluster.

Results from the second iteration of the Risk-based approach showed there were 8,188 clusters with a total of 1,003,470 returns. The average cluster size of the returns from the second iteration of the Risk-based approach was 122.55.

Approach 2. Top-down with Degree Limits

The Top-down approach is a network first clustering approach. It relies on a connected components algorithm to understand connections between tax returns. The IRS has used this strategy effectively in several tools and contexts to find groupings of tax returns. The approach is straightforward to apply and gives simply communicated information about likely groups of tax returns. This method enables the visualization and analysis of the activity of ghost preparers by focusing on the direct correlations identified in the tax filing information. Nevertheless, there are advantages and disadvantages to using connected components to look for connections in tax filings, just like with any analytical tool. One advantage of this strategy over the Risk-based approach is it does not assume ghost preparers engage in clearly risky behaviors. This allows analysts to discover ghost preparers new and emerging schemes. It will also help identify preparers who create accurate returns but do not sign them.

Background

The IRS has a long-established process for identifying ghost preparers and identity thieves by grouping returns from tax return data. These connections show patterns of actions or choices that a ghost preparer took when preparing tax returns. This was primarily applied to smaller groupings of returns, generally limited by a geographic area or certain risk characteristics. The results were especially useful in identifying communities of tax returns filed by a specific ghost preparer and enabled analysts to identify key points of connection and target interventions based on the possibility of fraud within the returns.

When this approach was applied to a less restricted set of returns, false positives and overlapping communities became an issue, as well as massive unmanageable clusters or super clusters. Early attempts to find meaningful clusters using connected components dealt with this super cluster problem in a variety of ways. One constructed the links but made no attempt to identify communities inside the super clusters. This enabled analysts to manually visualize the relationships and detect clusters. The team decided that this was a starting point for building clusters and learning about the connections between them.

Mechanics

This approach considers electronically filed self-prepared tax returns. That data was reformatted to build a network with normalized linking factors. Following the creation of the network, the connected components technique was used to identify clusters based on a pre-determined set of linking factors deemed strong and meaningful.

Once the network was created, linking factors were removed based on their degree count. This network analysis technique removes nodes from a network based on the number of connections they have; this limits overly connected or infrequently used nodes.

Results

The connected components technique yielded a super cluster containing more than 14.8 million tax returns. It also discovered 5,011 clusters of 50 or more tax returns, containing a total of 499,258 tax returns. The number of tax returns within the clusters ranged from 50 to 1,413. After excluding the solitary super cluster, the average cluster size was 99.6 tax returns. The 5,011 clusters must be examined further to identify false positives and overlapping communities.

Approach 3. Label Propagation Algorithm

The previous two clustering approaches can both produce large super clusters of interconnected returns that are not useful for analysis. While those approaches both try to avoid these super clusters, the binary nature of edges in a graph (either present or not) means that they are inherently vulnerable to over connecting when using noisy data. Label Propagation (LPA) is a network clustering algorithm that identifies nodes which are closely related. We use LPA on the ghost preparer project to break up those super clusters into smaller, more meaningful, components.

Background

The over connection that leads to the formation of super clusters happens because we are using data elements that contain noisy data. These issues can aggregate together to create very large and connected components. We can break up these large clusters into smaller components and remove some of the spurious connections by using graph community detection algorithms, like LPA.

Since the over connected clusters can be the result of spurious connections, label propagation can remove connections that are not supported by other nearby connections. The ability to break up over connected clusters is important for two reasons. First, the networks in this process are generated in such a way that densely connected areas are considered suspicious by default. This means that the large, over connected clusters often contain a disproportionate amount of ghost prepared returns compared to the rest of the clusters, so it is even more important to be able to break them up. Second, the ability to break up these large clusters provides flexibility creating networks.

With community detection algorithms, the only way that we can try to reduce the size and frequency of these over connected clusters is to be more conservative when creating edges. This requires leaving out information, or implementing complicated rules for creating edges will become hard to manage. Label propagation gives us another avenue to address this problem so that we do not have to leave as much potentially valuable information on the table.

Mechanics

Label Propagation works as follows:

1. Each node is given a unique label.
2. Each node changes its label to the most common label among its neighbors, with ties decided randomly.
3. Repeat Step 2 until no nodes change their label or you reach a predetermined number of iterations.

Label propagation has the effect of finding communities in the larger graph that have a high density of internal connections. The core idea is that once densely connected community settles on one label, there are not enough connections to nodes outside of the community to change the labels of all the community members. In the early stages, since all nodes are initialized with unique labels, there will be many tie votes that are decided randomly. As the algorithm progresses and communities start to form, there are fewer and fewer ties, and so there are fewer randomly decided winners.

Results

Label Propagation was used to break up the super cluster of almost 15 million returns that was created during the Top-down clustering process. The initial cluster of 14,796,946 returns was broken down into 632,797 smaller clusters, ranging in size from 1 return to 5,649 returns. The mean cluster size was 23.4 returns, and the 25^{th}, 50^{th}, and 75^{th} percentiles for cluster size were 5, 8, and 17 returns, respectively.

Other Topics

Ghost Preparer Tool

From the initial planning phases of the Innovation Lab, the team emphasized the importance of operationalizing any ghost preparer analysis results. To facilitate this, the planning team requested and received funding to develop a ghost preparer specific tool. In January 2022, the innovation lab released a beta of the Ghost Preparer Tool (GPT). The tool is designed to identify potential cases and investigate networks of returns for ghost preparers. From a data perspective the tool consists of two main components: a multi-year dataset of e-file Form 1040 returns stored in a network format and clustering results. The tool was designed to take advantage of existing graph tools within the IRS and to be as flexible as possible to accommodate new data sources and clustering techniques as they become available.

Work Streams

The tool is designed to support two workstreams, the first of which is case discovery. The aim is to allow users to review clusters of self-prepared returns to detect previously unidentified schemes or cases. A key piece of this work stream is standard cluster metrics generated for all clusters that have been added to the tool. The tool allows users to specify thresholds and filter and order lists of clusters based on cluster metrics irrespective of clustering technique. Users can download lists of suspicious clusters or use the tool's interface to explore the clusters of returns they've selected. The tool also includes a feature set where users can add notes to clusters allowing for deconfliction and collaboration between users.

The second workstream is the investigation of suspicious groupings of returns. There are two features in the tool which support this workstream: a full text search engine and graph visualizations. Full text search enables users to quickly search a point of connection across the entire reference dataset to if they have already been identified as being part of a cluster and to quickly see all returns related. The second component of the tool which supports the investigative work stream is the graph visualization of clusters. Here, users can graphicly explore the connections which exist between returns as well as reports generated for clusters. Users also have the ability run a set of predefined and user defined graph traversals to reveal potential leads.

Graph Neural Networks

Background

Graph Neural Networks (GNNs) are a variation of standard neural networks that use connections between data to perform machine learning tasks. This connection data can be used in addition to standard tabular data, but it could also be used on its own, if the focus of the machine learning task is to learn about the graph structure. There are different types of GNN models, but one of the common types of GNN, convolutional neural networks (CNNs), takes inspiration from approaches that were developed for computer vision tasks.

CNNs use a convolution layer that has the effect of sweeping a small window across an image. These smaller, windowed image segments are then what the neural network is trained to recognize, and this approach provides several benefits. By moving a small window across the image, the model is trained to pick out smaller scale details like the borders between objects or facial features, ignoring their placement in the overall image. In other words, this sliding window focuses the learning efforts of the model on the local structure of the image as opposed to the global structure. This approach also makes the model better at generalizing to images of different sizes and shapes.

Graph Convolutional Networks (GCNs) are GNNs that take inspiration from CNNs. GCNs focus learning on small scale graph structures instead of whole graphs and are better able to handle graphs of different scales. One of the most interesting things about GCNs is that they are generally very efficient to implement, both in terms of data preprocessing and in terms of actual computation. Standard CNNs include extra hyperparameters and choices to be made about the convolution process. GCNs, on the other hand, can leverage the deep connections between graph theory and linear algebra so that they are relatively simple to create and computationally efficient.

Application to Ghost Preparers

GNNs can help when we are in a situation where the data must be evaluated in context. The returns prepared by ghosts are usually prepared with the cooperation of the filer, and so in isolation there may be few or no indications that the person who prepared this return was not the person who filed it. Ghost preparers are usually discovered because a group of returns that have been signed by different DIY filers all have some indication that they may actually have been prepared by one actor.

GCNs address this problem because they can take the values of features of neighboring nodes into account when creating vector representations (embeddings) of either nodes, edges or entire graphs. These embeddings can then be used for different downstream tasks, like classifying nodes, classifying graphs, predicting links, or detecting anomalies.

Current Plans

The ghost preparer team is currently working on a prototype GCN model. This initial model will be a link prediction model, with the goal of learning how preparers are linked to returns that they prepare so that we them predict links between ghost prepared clusters and who may be preparing those returns. One of the benefits of this research is that once there is a process in place to create meaningful embeddings of the data with the GCN, it is relatively simple to then apply those embeddings to different downstream tasks. Depending on how our link prediction experiment goes, we can explore unsupervised anomaly detection for finding potential ghost prep clusters and can also test out supervised models once we accumulate enough data on confirmed ghost preparers to use as labels.

Future Analysis

Cluster Timelines

Using network-based approaches to detect ghost preparers requires the suspect network to be mature enough that we can detect signs of ghost preparing behavior. Networks are considered fully formed at the end of the tax season and returns that would be present in any given cluster are present in the network. However, the end of the tax season may be post-refund and more cumbersome to deal with from a compliance standpoint. A cluster timeline analysis was completed to look at various snapshots throughout the tax season to see if Document Locator Numbers (DLNs) identified through the Risk-based approach would appear at the end of the tax season in a suspicious cluster of 50 or more returns.

Approach

The Risk-based approach was run using all the data available for filing season 2021. The dataset containing the selections of possible ghost preparers served as the comparison group and is referred to as the final set. The Risk-based approach was run again at three different times during filing season 2021: February (Group Timepoint 1 (T1)), March (Group Timepoint 2 (T2)), and April (Group Timepoint 3 (T3)). The approach was run in its entirety, and the approach remained the same for all the time points as it did for the final set, except that the cluster size minimums for groups T1 to T3 were dropped down from 50 returns to 5 returns. Results from groups T1 to T3 were transformed into cluster-level data. Returns were grouped by cluster size and the number of clusters that have that given cluster size were recorded (e.g., in cluster size 8, there were 3,687 clusters that had that given cluster size; see Table 2).

Additionally, the DLN count was computed for each cluster group size (e.g., there were 29,496 returns that were a part of a cluster group size 8). From there, four data points were computed for each given cluster group size: the number of suspicions DLNs (returns) that appeared in the final set; the percentage of DLNs that appeared in the final set; the percentage of the clusters where 100% of the returns appear in the final set (Percent of 100% Clusters in Final Set); and the number of clusters, where all of the returns appear in the final set (Number of 100% Clusters in Final Set).

TABLE 2. Timepoint Snapshot Example, February 2021

Group	Group Size	Total Clusters	Percent of Clusters	Total DLNs	Suspicious DLNs	% of DLNs in Final	# of 100% Clusters in Final Set
T1	6	10,435	7.06	62,610	4,493	7.18	737
T1	7	5,688	9.53	39,816	3,893	9.78	542
T1	8	3,687	13.48	29,496	4,063	13.77	497
T1	9	2,459	17.97	22,131	4,105	18.55	442
T1	10	1,900	21.53	19,000	4,254	22.39	409

Cluster-level data were further transformed into cumulative data for each group (T1 to T3), and the cluster group sizes were collapsed into groups of 10 plus, 20 plus, 30 plus, 40 plus, and 50 plus (e.g., 10 plus represented cluster group sizes of 10 or greater). For each transformed group, suspicious returns (DLNs), total returns (DLNs), DLN percentage in final set, and cluster percent in final set were calculated. The idea was to create groups of returns with realistic cutoff criteria but with no specific grouping size selected. Instead, it is more likely and more logical to say that cluster group sizes of 30 plus have more suspicious returns than cluster group sizes that have exactly 25 returns.

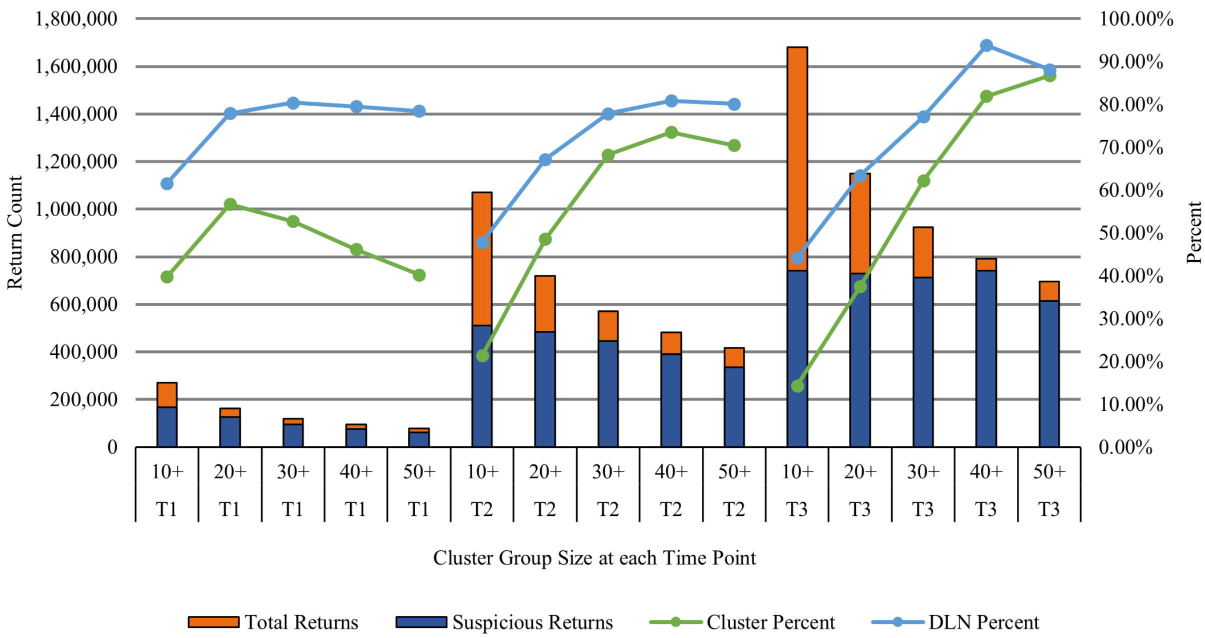
Results

Timepoint 1. Results showed that the Risk-based approach identified suspicious DLNs in T1 (February snapshot) among all groups of cumulative cluster sizes better than chance (50 percent). In T1 61.56 percent of DLNs showed up in the final set for cluster sizes of 10 or greater (see Figure 1). This percentage increased to 77.90 percent for clusters of 20 or greater and 80.37 percent for clusters 30 or greater (see Figure 1). Percentage decreased for cluster sizes of 40 or greater (79.52 percent) and 50 or greater (78.44 percent).

Timepoint 2. Results showed a similar pattern of increasing percentage as cumulative group size increased in T2 (March snapshot), but it began at a lower percentage than timepoint 1 (T1). For cluster group sizes of 10 plus, 47.83 percent of DLNs appeared in the final set (see Figure 1). Percentages increased in cluster group sizes of 20 plus (67.20 percent), cluster group sizes of 30 plus (77.82 percent), and cluster group sizes of 40 plus (80.86 percent). For cluster group sizes of 50 plus the percentage decreased to 80.11 percent.

Timepoint 3. A similar pattern of increasing percentage was seen in T3 (April snapshot), and like T2 the starting percentage began lower than 50 percent. For cluster group sizes of 10 plus, 44.16 percent of DLNs appeared in the final set and increased to 63.46 percent for cluster group sizes of 20 plus (see Figure 1). Percentages continue to increase for cluster group sizes of 30 plus (77.10 percent), and cluster group sizes of 40 plus (93.80 percent). Lastly, percentages decreased down to 88.12 percent for cluster group sizes of 50 plus.

FIGURE 1. Detection of Suspicious Ghost Preparer Clusters for Three Time Points



Results from the Cluster Timeline Analysis supported the idea that the Risk-based approach can detect suspicious returns at various timepoints in the tax season, and those same suspicious returns are likely to show up at the end of the tax season in a suspicious cluster of 50 or more returns. However, the Risk-based approach uses a predetermined set of scores to identify clusters of returns as suspicious, and it is possible that over time with various schemes those markers would evolve or become extinct. Additionally, the Risk-based approach uses degree filtering to break up large clusters, and it is possible that in earlier snapshots returns were in small cluster and identified as suspicious. In later snapshots, those returns could have fallen into a super cluster and have been broken into a smaller cluster that was not suspicious. Lastly, the approach does not track returns over time, as it is not looking at the same cluster and how it changes. Instead, the analysis represents three distinct snapshots and shows the percentage of DLNs that show up in the final set. The simplicity of the Cluster Timeline Analysis allows it to be implemented for any other networking approaches being used to identify ghost preparer networks.

Impact Analysis

The aim of this analysis is to leverage the clustering results generated during the Innovation Lab analysis and subsequent modeling to provide insight into the risks ghost preparers may pose to the fair and effective implementation of the U.S. tax code. To measure that risk, this analysis focuses on how returns changed from tax year to tax year for individuals who appear to transition to filing with a ghost preparer. This work can help us to begin to understand the potential impact ghost preparers have on a return-by-return basis and, in the future, can be extrapolated to provide a picture of the impact ghost preparers have in the aggregate. This analysis compares changes across return characteristics and established risk metrics.

Anecdotal evidence backed by analysis of the clusters identified during the Ghost Preparer Innovation lab suggests that ghost preparers do not work with a group of taxpayers that is representative of the entire taxpaying population. Ghost prepared clusters tend to be lower income filers, but there are indications that there may be other communities which disproportionately see ghost preparer involvement. At this point in our understanding of ghost preparers, keeping the analysis limited to individuals who appear to have worked with a ghost preparer limits the risks of attributing the filing behavior of a specific community or demographic to ghost preparers.

Approach

We considered clustering results generated using the risk-based clustering for Tax Years (TYs) 2019, 2020, and 2021. From those clustering results we selected a sample of 5,000 primary filers and considered their returns for the three tax years used in clustering. From that dataset we identified two groups of returns, those where the filer transitioned into a cluster and those where a filer stayed in a cluster. Conceptually we considered these two groups—joined ghost preparer and stayed with a ghost preparer. For the two groups, we compared year-over-year changes in their returns to attempt to understand the effect joining a suspicious cluster has on their filings.

When a primary filer filed a return the previous tax year that wasn't identified with a cluster and then files a return in the current year that is identified with one, we determined that the filer joined a ghost preparer. When a primary filer filed returns identified with a cluster in consecutive tax years, we determined that the filer stayed with a ghost preparer. Cases when the primary filer did not file in the previous year, transitioned out of a cluster, or remained outside of a cluster in consecutive years were excluded from consideration.

For TYs 2020 and 2021 of the 5,000 primary filers we considered, we found 1,934 returns where the filer joined a ghost preparer and 1,956 where the filer stayed with a ghost preparer. For these groups of returns, we measured year-over-year changes as the difference in values.

Limitations and Assumptions

There are several assumptions and limitations in the analysis. As stated elsewhere, we do not have labeled data, so we assume that the clusters of interconnected self-prepared returns we detect represent individual ghost preparers and that all returns in each cluster are ghost prepared. We recognize that there may be false associations, returns incorrectly identified as being ghost prepared, as well as returns where we do not detect the involvement of a sophisticated ghost preparer. We do not currently have a measure of the extent to which we misidentify ghost preparers.

This analysis hinges on the intuition that ghost preparers engage in consistent behavior year over year and that they treat new and returning clients to their illegitimate practice the same. An extension of that assumption is that where a ghost preparer adopts a new scheme or preparing practice, we assume it is generally employed across all returns they prepare. It is worth recognizing if a ghost preparer treats new patrons differently or if they prepare randomly, it could add a confounding variable to our analysis.

The second key assumption is that the changes in the returns of individuals we've identified as transitioning to ghost preparers are due to the ghost preparer rather than the motivation for the individual to seek out a ghost preparer. It is conceivable that individuals choosing to prepare with a ghost preparer may do so due to a change in their tax situation. It is difficult to statistically disambiguate victims of a ghost preparer and taxpayers working in collaboration with their preparer.

A final important consideration for this analysis is that it spanned the COVID-19 pandemic. This is a period of change in employment and earnings for many Americans, which had the potential to impact the results.

Return Distributions

When considering year-over-year changes, initial findings supported the assertion that ghost preparers do influence their clients' returns. Individuals joining a suspected ghost cluster in comparison to individuals remaining in a ghost cluster are more likely to see changes in their return characteristics from tax year to tax year. In addition, preliminary results showed that first year clients and returning clients are comparable across income and credits claimed suggesting that we are comparing a similar population of filers.

The most notable change from the perspective of tax administration was refunds. Risk-based cluster returns that transitioned to a ghost preparer saw a $733 increase in average refunds for TYs 2021 and 2022 combined, compared to returning ghost preparer cluster filers. Average annual increases in refunds were paired with increases in average reported incomes and withholding, which, while notable, did not provide a clear picture of noncompliance.

TABLE 3. Comparative Annual Changes in Reported Income and Refunds

Return Element	Tax Period	Average Value on F1040		Annual Change		Annual % Change	
		Joined GPC	Stayed GPC	Joined GPC	Stayed GPC	Joined GPC	Stayed GPC
Total Income	TY 2020	$43,154	$39,721	$6,853	$1,268	16%	3%
	TY 2021	$39,061	$42,941	$2,572	$745	7%	2%
	Total	$40,937	$41,337	$4,538	$1,005	11%	2%
Adjusted Gross Income	TY 2020	$42,594	$39,122	$6,880	$1,232	16%	3%
	TY 2021	$38,395	$42,413	$2,370	$851	6%	2%
	Total	$40,320	$40,774	$4,442	$1,040	11%	3%
W2 Wages	TY 2020	$39,331	$40,244	$2,220	-$388	6%	-1%
	TY 2021	$41,900	$45,075	$3,140	$3,530	7%	8%
	Total	$40,706	$42,648	$2,717	$1,579	7%	4%
Total Tax Amount	TY 2020	$3,786	$2,210	$1,125	$24	30%	1%
	TY 2021	$2,922	$3,138	$555	$353	19%	11%
	Total	$3,301	$2,687	$817	$189	25%	7%
Withholding Amount	TY 2020	$3,932	$3,891	$599	-$17	15%	0%
	TY 2021	$3,876	$4,384	$169	$411	4%	9%
	Total	$3,902	$4,138	$367	$198	9%	5%
Refund Amount	TY 2020	$3,744	$4,194	$508	-$43	14%	-1%
	TY 2021	$4,755	$4,394	$1,042	$170	22%	4%
	Total	$4,291	$4,294	$797	$64	19%	1%
Earned Income Credit	TY 2020	$2,843	$2,994	-$49	-$188	-2%	-6%
	TY 2021	$2,621	$2,775	$185	$29	7%	1%
	Total	$2,712	$2,884	$78	-$79	3%	-3%

An additional finding of note is that there was no major year-over-year change in Earned Income Tax Credit (EITC); neither the rates at which it is claimed, nor the average value of the credit. This is significant because it showed changes in refunds do not appear to be driven by the EIC and that, for this clustering approach, transitioning to a ghost preparer doesn't appear to have a major change in EITC behavior of taxpayers.

TABLE 4. Year-Over-Year Change in EITC Claims

	Dropped	No Change	Added	Net Change	Percent Change
Tax Year 2020					
Joined GPC	- 38	807	+ 43	5	1%
Stayed GPC	- 53	889	+ 32	-21	-2%
Tax Year 2021					
Joined GPC	- 32	965	+ 48	16	2%
Stayed GPC	- 48	907	+ 27	-21	-2%
Totals					
Joined GPC	- 70	1772	+ 91	21	1%
Stayed GPC	- 101	1796	+ 59	-42	-2%

We do observe filers joining suspicious clusters show higher year-over-year changes in Schedule C usage. This could indicate that ghost prepares are fabricating business income and losses to maximize refunds for their clients however this analysis doesn't provide evidence of noncompliance.

TABLE 5. Year-Over-Year Change in Schedule C Usage

	Dropped	No Change	Added	Net Change	Percent Change
Tax Year 2020					
Joined GPC	- 53	740	+ 95	42	5%
Stayed GPC	- 67	856	+ 51	-16	-2%
Tax Year 2021					
Joined GPC	- 37	858	+ 150	113	11%
Stayed GPC	- 40	879	+ 63	23	2%
Totals					
Joined GPC	- 90	1,598	+ 245	155	8%
Stayed GPC	- 107	1,735	+ 114	7	0%

Discriminant Function Score Distributions of Cluster Returns

In looking for ghost preparer effects we consider an existing IRS risk metric, the discriminant function (DIF) score, to provide insight into the compliance risks posed by clusters of interconnected 1040 self-prepared returns. The DIF scoring algorithm is a technique that has been used since 1969 to predict how likely a tax return is to have a significant adjustment. Individual and small corporation income tax returns, and S corporation, and partnership tax returns receive a DIF score during processing. The score is calculated and stored in the administrative data system, then is used in downstream systems during examination selection.

DIF encompasses a series of models that are specific to mutually exclusive tax classes known as activity codes. Returns are assigned an activity code based on total positive income, total gross receipts, and EITC. These classes help in guaranteeing fairness by providing balanced coverage for all tax return types. By developing a model for each activity code, similar returns can be compared to one another, enabling more accurate predictions about the population, and allowing for the most noncompliant returns to be selected. A high DIF score indicates that the return has a high likelihood for significant tax change overall and that auditing that return will lead to a tax change. DIF score distributions are not consistent across activity codes or processing years, meaning models cannot be compared to one another.

To allow for comparison year to year across all filing types we consider returns which fall within the top 5 percent of DIF scores, indicating they're the riskiest returns irrespective of processing year or activity code.

Results. For the two tax years considered, 2020 and 2021, we found that on net, 3 percent of returns where the filer joined a ghost cluster moved into to the 95th percentile of the DIF distribution from the previous tax year compared to 1 percent of returns where the primary filer stayed with a ghost preparer. These results are not conclusive, but they do indicate that ghost preparers do not lessen the audit risk to the taxpayer or improve compliance on average.

Overall, however, we found all returns in the population considered to be significantly more likely to fall in the top 5 percent of the DIF distribution. Returns in suspected clusters, both first time and repeat filers, fell within the top 5 percent of the DIF distribution at rates of 16 percent and 17 percent respectively across TYs 2020 and 2021, indicating they are more than 3 times as risky as returns overall when considered from a DIF perspective. For TY 2020, our sample includes 1,045 filers who self-prepared and would go on to be identified in a suspected ghost cluster in TY 2021, for that group 138 of the returns, or 13 percent fell within the top 5 percent of the DIF distribution. While not as high as the returns identified in ghost clusters for the same year, (15 percent) it still is much higher than we would anticipate.

TABLE 6. Returns in the 95th Percentile of the DIF Distribution

	Total Returns	Returns in 95% of DIF Distribution*	Left 95%	No Change	Joined 95%	Net Change
Tax Year 2020						
Joined GPC	888	14%	-52	772	+64	+12
Stayed GPC	974	16%	-61	849	+64	+3
Tax Year 2021						
Joined GPC	1,045	18%	-59	878	+108	+49
Stayed GPC	982	18%	-67	832	+83	+16
Totals						
Joined GPC	1,933	16%	-111	1,651	+172	+61
Stayed GPC	1,956	17%	-128	1,681	+147	+19

* This value is 5 percent in the population overall.

Next Steps

Validate

An important next step for the larger project and for this analysis is to measure the effectiveness of the various clustering approaches. Validation generates feedback which can improve existing processes and provides important context to the analysis results. There may be options to use existing IRS data or processes for this purpose, but each has its own challenges and limitations. One possibility would be to use compliance and enforcement data to check if previously identified ghost preparers would have been identified using network analysis. Another approach might be to use IDT identification detection processes to identify overlap with ghost preparation results. There may be additional options as well.

Validation can help to identify potential biases or gaps in clustering results which is critical to this effort. False positives could have serious consequences for taxpayers who legitimately self-prepare their returns as well as individuals falsely identified as ghost preparing returns. In addition to limiting risk to the taxpayer, verifying results can highlight failures of the model to identify known ghost prepared returns which may represent gaps in enforcement. Ghost preparer patterns and approaches likely evolve overtime, generating labeled data is key to improving models and staying ahead of new schemes.

When a ghost preparer is detected, the IRS does undertake outreach and compliance actions to help that preparer meet their legal obligation. In some cases, the IRS may pursue a criminal investigation of the preparer. As a result, the IRS does have information about known ghost preparers. One approach for validating clustering results would be to check if they detect known ghosts.

There are challenges in doing this as well as some limitations to how generalizable we could expect the results. Compliance datasets are primarily oriented around the ghost preparer, while the cluster analysis is oriented around returns. The complicates analysis because it requires dealing both with the uncertainty around the ghost and their clientele, for a known ghost we may not have an exhaustive picture of the returns they prepared and for the suspected returns we may not have a full picture of the ghost preparer. Creating a cross walk between the two will be a challenge.

A limitation, which may be abating, is the delay between detection and action. Investigations and prosecutions may take years, so many definitively identified ghosts were not active during the processing years for which the clustering results are available. This means that known ghosts, either prosecuted or treated and the returns identified using network analysis have limited overlap. Finally, ghosts the IRS has detected and treated may not be representative of ghost preparers overall, so looking at these datasets may not provide a true picture of the effectiveness of the clustering.

The IRS does commit resources to detecting IDT in real time to limit risk of individuals being unable to file returns and reduce the harm to the government. Some of these processes may also identify ghost preparers in real time. One option would be to collaborate with the IDT detection teams to look for overlaps. The major limitation of this approach is that while it may corroborate the current approaches, it is not definitive.

Ghost Preparer Compliance Study

By leveraging important insights learned thus far using the GPT network analysis techniques that identify suspected ghost preparers, a compliance study program could be established to enable the Service to study ghost preparers' compliance behaviors and effects on tax administration. A potential ghost preparer compliance study might involve examining a portion of tax returns they prepared to assess compliance changes. The study should consider elements such as income underreporting and credit overclaims which would require a two-step sampling design.

A formal compliance study, if undertaken, should involve various stakeholders to set priority goals and determine the size of the study needed based on available resources and additional data to be captured, among many factors. Since there is no ghost preparer audit data, what follows is a possible starting point.

Objectives of Ghost Preparer Compliance Study

1. Estimate ghost preparer population and related characteristics,
2. Estimate impact of ghost preparers on tax compliance and tax-administration, and
3. Use result of the study to enhance ghost preparer compliance strategy.

First-Stage Sample. Select a random sample of clusters (networks) from a population of suspected ghost preparer networks[2] identified by the two clustering approaches: Risk-based and Top-down. Provisionally, this can be done by selecting equal number of first-stage samples of suspected clusters from each clustering approach. Since suspected ghost preparer clusters range by sizes of returns, stratification of the first-stage sample by network size will be an appropriate approach to make sure some of the large volume clusters are included.

Second-Stage Sample. For each of the first-stage samples of suspected ghost preparer networks selected, select a random sample of returns (customers of suspected ghost preparers) with an explicit objective of positively identifying the suspected ghost preparer[3] and assess the nature and level of noncompliance at the return level. The second-stage sample size for each first-stage sample cluster will depend on resources, but a starting point could be proportional-to-size, with the minimum number of returns (customers) necessary to positively identify the ghost preparer, as determined by subject matter experts.[4]

While a formal ghost preparer study can be implemented as part of existing compliance programs, there will be extra efforts and data capturing needs that will require additional resources. However, the long-term benefit of the outcome data of the study will far outweigh the initial costs. In addition to being able to have a reasonable impact estimate with a Ghost Preparer Compliance outcome data, a more tailored compliance strategy and treatment option(s) can be established by developing supervised predictive machine learning algorithms, such as GNN, that can better identify potential ghost preparers. Without outcome data, the best modeling effort that can be done at this point is some sort of unsupervised anomaly detection method whose effectiveness and reliability cannot be as easily assessed during modeling. Furthermore, with outcome data, the effectiveness of the networking approaches can be valuated and improved, iteratively.

[2] The population of suspected ghost preparer networks refers to networks identified by each of the clustering approaches. However, these networks don't necessarily identify all ghost preparer networks. Hence, by the nature of the problem, we don't have the possible universe of the ghost preparer frame. Accordingly, though a big step up from current practices, the level of inference that can be made using this compliance data will be limited.

[3] For each first stage sample cluster, we would need a second stage sample size of returns audited to positively identify the ghost preparer and to produce compliance level at a network level.

[4] If results from the Ghost Preparer pilot treatment options are available, a formal model-based sampling design can be constructed.

References

Brandes, Ulrik (2005). "*Network Analysis: Methodological Foundations.*" Volume 3418. Germany: Springer Science & Business Media.

Internal Revenue Service (IRS) (2009*). "Publication 4832*, *Return Preparer Review,* Rev. Dec. 2009

TIGTA (July 25, 2018). "The Internal Revenue Service Lacks a Coordinated Strategy to Address Unregulated Return Preparer Misconduct." *Treasury Inspector General for Tax Administration Ref. No. 2018-30-042*

5

∇

Appendix

Conference Program

13th Annual IRS-TPC Joint Research Conference on Tax Administration
June 22, 2023

Program

9:00–9:30 Opening

Wendy Edelberg (Director of the Hamilton Project, Brookings Institution)

Eric Toder (Co-Director, Urban-Brookings Tax Policy Center) and

Barry Johnson (Deputy Chief Data and Analytics Officer, Research, Applied Analytics and Statistics (IRS)

9:30–11:00 Session 1: Service is Our Surname

Moderator: *Deena Ackerman (U.S. Department of The Treasury)*

» Looking Beyond Level of Service: Using Behavioral Insights to Improve Taxpayer Experience
Jan Millard (IRS, RAAS); Sarah Smolenski, Jonah Flateman, Jamil Mirabito, Omar Faruqi, Lauren Szczerbinski, Michael Stavrianos (ASR Analytics)

» The Balance Due Taxpayer: How Do We Reduce IRS Cost and Taxpayer Burden for Resolving Balance Due Accounts?
Howard Rasey, Shannon Murphy, Frank Greco, Javier Framinan (IRS, W&I); Angela Colona, Javier Alvarez (IRS, Taxpayer Experience Office)

» Understanding Yearly Changes in Family Structure and Income and Their Impact on Tax Credits: Can Tax Credits Be Advanced?
Elaine Maag, Nikhita Airi, Lillian Hunter (Urban-Brookings Tax Policy Center)

» Racial Disparities in Audit Rates
Thomas Hertz (IRS, RAAS)

Discussant: *Janet Holtzblatt* (Urban-Brookings Tax Policy Center)*
Emily Y. Lin (U.S. Department of the Treasury)*

10:40–10:55 p.m. – Break

10:55–12:25 p.m. – Session 2: Estimating Audit Aftershocks

Moderator: *John Guyton (IRS, RAAS)*

» Changes to Voluntary Compliance Following Random Taxpayer Audits
Allan Partington, Murat Besnek (Australian Taxation Office)

» The Long-Term Impact of Audits on Nonfiling Taxpayers
India Lindsay, Jess Grana (MITRE); Alan Plumley (IRS, RAAS)

» Silver Lining: Estimating the Compliance Response to Declining Audit Coverage
Alan Plumley, Daniel Rodriguez (IRS, RAAS); Jess Grana, Alexander McGlothlin (MITRE)

Discussant: *William Boning (U.S. Department of the Treasury)*

12:25–1:25 p.m. – Keynote Speaker/Lunch

Catherine Rampell (Washington Post)

1:25–2:55 p.m. – Session 3: Understanding Contemporary Taxpayers

Moderator: *Russell James (IRS, RAAS)*

» Who are Married-Filing-Separately Filers and Why Should We Care?
Emily Y. Lin, Navodhya Samarakoon (U.S. Department of the Treasury)

» Willing but Unable to Pay? The Role of Gender in Tax Compliance
Andrea Lopez-Luzuriaga (Universidad del Rosario); Carlos Scartascini (Inter-American Development Bank)*

» Who Sells Cryptocurrency?
Jeffrey L. Hoopes (University of North Carolina at Chapel Hill); Tyler S. Menzer, Jaron H.Wilde (University of Iowa)

Discussant: *Yan Sun (IRS, RAAS)*

2:55 p.m. - 3:10 p.m. – Break

3:10–4:40 p.m. – Session 4: Hidden Assets, Hidden Networks

Moderator: *Robert McClelland (Tax Policy Center)*

» Following K-1s: Considering Foreign Accounts in Context
Tomas Wind, David Bratt, Alissa Graff, Anne Herlache (IRS, RAAS)*

» Application of Network Analysis to Identify Likely Ghost Preparer Networks
Chris Hess, Joshua King, Ashley Nowicki, Andrew Soto, Getaneh Yismaw, Ririko Horvath (IRS, RAAS); Brandon Gleason (IRS, Criminal Investigation); Jacob Brooks, Daniel Hales, Michael Stavrianos, Will Sundstrom (ASR Analytics)

» The Offshore World According to FATCA: New Evidence on the Foreign Wealth of U.S. Household
Niels Johannesen (University of Copenhagen); Daniel Reck (University of Maryland); Max Risch (Carnegie Mellon University); Joel Slemrod (University of Michigan); John Guyton, Patrick Langetieg (IRS, RAAS)

Discussant: *Paul Organ (U.S. Department of the Treasury)*

4:40–4:45 p.m. – Wrap-up

Barry Johnson *(Deputy Chief Data and Analytics Officer, Research, Applied Analytics, and Statistics (IRS))*