NBER WORKING PAPER SERIES

AUTOMATIC TAX FILING:
SIMULATING A PRE-POPULATED FORM 1040

Lucas Goodman
Katherine Lim
Bruce Sacerdote
Andrew Whitten

Working Paper 30008
http://www.nber.org/papers/w30008

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2022, Revised October 2022

We thank participants at the Office of Tax Analysis Annual Research Conference, the Georgetown University Law Center Tax Law and Public Finance workshop, and the IRS-TPC Joint Research Conference on Tax Administration for helpful comments. We thank Maxwell Grozovsky for excellent research assistance. The views expressed in this paper do not necessarily represent the official position of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. This research was conducted while Goodman and Whitten 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. The authors declare that they have no relevant financial interests that relate to the research described in this paper. The authors declare that they have no relevant financial interests that relate to the research described in this paper.

Automatic Tax Filing: Simulating a Pre-Populated Form 1040
Lucas Goodman, Katherine Lim, Bruce Sacerdote, and Andrew Whitten
NBER Working Paper No. 30008
April 2022, Revised October 2022
JEL No. H0,H2,H24,I3

ABSTRACT

Each year Americans spend over two billion hours and $30 billion preparing individual tax returns, and these filing costs are regressive. To lower and redistribute the filing burden, some commentators have proposed having the IRS pre-populate tax returns for individuals. We evaluate this hypothetical policy using a large, nationally representative sample of returns filed for the tax year 2019. Our baseline results indicate that between 64 and 73 million returns (42 to 48 percent of all returns) could be accurately pre-populated using only current-year information returns and the prior-year return. Accuracy rates decline with income and are higher for taxpayers who have fewer dependents or are unmarried. We also examine 2019 non-filers, finding that pre-populated returns tentatively indicate $8.5 billion in refunds due to 12 million (21 percent) of them.

Lucas Goodman
Office of Tax Analysis
Department of Treasury
1500 Pennsylvania Avenue NW
Washington, DC 20220
Lucas.Goodman@treasury.gov

Katherine Lim
Minneapolis Federal Reserve Bank
90 Hennepin Avenue
Minneapolis, MN 55401
katie.lim@mpls.frb.org

Bruce Sacerdote
6106 Rockefeller Hall
Department of Economics
Dartmouth College
Hanover, NH 03755-3514
and NBER
Bruce.I.Sacerdote@dartmouth.edu

Andrew Whitten
Office of Tax Analysis
Department of the Treasury
Washington, DC 20220
andrew.whitten@treasury.gov

The direct costs of individual income tax compliance are large and growing. In the instructions for the 2019 Form 1040, the Internal Revenue Service (IRS) estimates that the average filer spends $210 and 11 hours to file. This adds up to 1.7 billion hours and $32 billion in tax preparation fees, software costs, and filing fees. To make the process more efficient, and to redistribute this burden more progressively, some commentators have proposed having the IRS pre-populate returns on behalf of individual taxpayers. After all, the IRS already receives information from third parties on many income sources reported on tax returns, including wages, unemployment compensation, interest, dividends, capital gains, non-employee compensation, and income from partnerships and S corporations. Moreover, the IRS could infer filing status and dependents based on prior year returns. Pre-population may have the additional benefit of encouraging non-filers to claim refunds (e.g., from unclaimed Earned Income or Child Tax Credits) or to meet their filing obligations.

In this paper, we explore the degree to which such pre-populated returns would have been "successful" for tax year 2019. To our knowledge, we are the first to simulate pre-populated returns in the U.S. context. We take two approaches to measure success. In our first approach, we check for tax situations that would result in an inaccurate pre-populated return, such as reporting a nontrivial amount of income, deduction, or credit on the 2019 filed tax return that the IRS does not observe on an information return. A return without any of these failure situations is deemed a success under this approach, which we refer to as the "upper-bound" approach because it will tend to overstate the success rate due to our inability to enumerate and measure *all* potential failure situations.

In our second approach, we simulate a pre-populated 2019 Form 1040 line by line, using only the taxpayer's 2019 information returns and their 2018 Form 1040 if they filed one. We calculate tax liability using NBER TAXSIM (Feenberg and Coutts, 1993), generally assuming that the taxpayer has the same filing status and dependents as the prior year. Under this approach, a pre-populated return is deemed successful if its calculated tax liability is approximately equal to the tax liability actually reported on the 2019 tax return (within a tolerance threshold of $100 in our baseline results). We refer to this as a "lower-bound" approach because it will tend to understate the success rate relative to a richer tax calculator.[1]

Using our baseline tolerance thresholds, our lower and upper bound procedures are successful for 64 and 73 million returns (42 and 48 percent of filers), respectively. In about two-thirds of the cases where the lower bound approach is inaccurate, the pre-populated liability is higher than the reported liability, with a mean gap of $4,200. For both the lower and upper bounds, we find that success rates are decreasing in income, from 60-80 percent in the bottom income decile to 10-30 percent in the top income decile. The higher rates of failure for higher-income taxpayers are largely driven by increasing rates of itemized deductions, which cause a divergence between pre-populated and actual returns. Pre-population is particularly successful for taxpayers who are single, young, and lack dependents. Suggestive of the potential benefits to pre-population, among those taxpayers who would have an accurately pre-populated return, 43 to 44 percent used a paid preparer when filing.

Pre-population may yield significant time savings even for taxpayers for whom the IRS cannot fully complete their return. Among the 52 to 58 percent of taxpayers with inaccurately pre-populated returns, the majority would need to make only one change or complete one additional schedule (e.g., reporting self-employment income and deductions on Schedule C -– and attaching any schedules required for that Schedule C -– which would be the most common edit required).

[1] We use the term "bound" loosely. Our approaches do not yield true mathematical upper or lower bounds, as explained further in Section II.

Next, we consider refinements of the pre-population procedure in which the IRS pre-populates returns for certain subsets of taxpayers. In the narrowest subset – single taxpayers with no dependents, with only wage income, and with only deductions or credits that are observed by the IRS[2] – we estimate that around 80 percent of these returns could be successfully pre-populated. The imperfect success rate for this group is partly due to the mismatch rate between wages reported on information returns vs. tax returns. Among wage earners, we accurately pre-populate this line of the Form 1040 only 90 percent of the time.

Beyond the time and cost savings to current filers, another potential benefit of pre-populated returns is that they could be created for current non-filers, potentially nudging them into claiming refunds or paying taxes due. For tax year 2019, we find 46.3 million adults who received an information return, did not file a tax return, and did not appear not to have a filing obligation based on income on their information returns.[3] Our pre-populated returns indicate that 17 percent (8.0 million) of these individuals are entitled to a refund, with an average potential refund of $386. We additionally identify 8.4 million non-filers who appear to have a filing obligation based on their information returns. Among this population, 55 percent (4.6 million) appear to have a balance due. Guyton et al. (2016) and Goldin et al. (2021) find that non-filers are more likely to file after receiving a reminder, and pre-populated returns may provide this reminder.

Our findings extend the existing work on pre-populated returns in the U.S. – which we describe in the next section – by using individual-level administrative data to calculate success rates along a number of taxpayer characteristics. We also update the literature by using data from after the 2017 tax reform, which (among other changes) reduced the prevalence of itemized deductions substantially. Moreover, our data allows us to focus on specific subgroups of taxpayers as well as non-filers. Our detailed results can inform the design of a pre-population program for some or all U.S. taxpayers.

To be sure, substantial challenges would arise in any implementation of pre-populated returns. For one, either providers of information returns – such as employers and financial institutions – would need to submit information returns earlier than they currently do, or the deadline for individual returns would have to be delayed. Additionally, a pre-populated return may function as a nudge toward accepting the unmodified pre-populated liability, likely changing the distribution of taxes paid. To the extent that this nudge has an effect and these pre-populated liabilities differ from "true" liabilities under the Internal Revenue Code, various equity and efficiency policy aims could be undermined. While we abstract from these issues in this paper, we stress that policymakers must take them into account when deciding whether, and how, to implement a pre-populated return program.

I Background

Pre-populated tax returns are not a new idea. Over 45 countries at least partially pre-populate their personal income tax returns (OECD, 2021). A number of these countries operate tax agency reconciliation systems, in which taxpayers may elect to have the tax authority complete their personal tax return and send it to them for review. In the U.S., more limited programs in California and Colorado have been attempted (U.S. Department of the Treasury, 2003; Fichtner, Gale and Trinca, 2019). The California pi-

[2]Unobserved items include itemized deductions; moving expenses; educator expenses; self-employment health insurance; and credits other than the Earned Income Tax Credit, Child Tax Credit, American Opportunity Tax Credit, or the credit for excess Social Security tax.

[3]For this analysis, we include all information returns with U.S. addresses, including but not limited to Form W-2, the Form 1099 series, and the Form 1095 series.

lot, the ReadyReturn program, ran between 2003-2005. Each year roughly 50,000 people were mailed a pre-populated return and cover letter. One fifth of recipients used the pre-populated ReadyReturn and customer satisfaction among the users was high (Bankman, 2005, 2008). Based on the success of the pilot, the program was expanded each year through 2013; however, it was then discontinued.

Informed by international and domestic experiences, academics and practitioners have analyzed the possibility of pre-populating tax returns in the U.S. (Internal Revenue Service, 1987; U.S. General Accounting Office, 1996; Gale and Holtzblatt, 1997; U.S. Department of the Treasury, 2003; Goolsbee, 2006; Cordes and Holen, 2010; Fichtner, Gale and Trinca, 2019). These reports note that, unlike the U.S, countries currently using tax agency reconciliation systems generally have individual-level taxation (i.e., tax schedules independent of family structure) where most taxpayers face the same marginal tax rate with few deductions, allowances, and credits.

Empirical work in the U.S. context has used tax-return data (without linking to information returns) to estimate the share of taxpayers that could be eligible for either pre-populated returns or exact withholding of tax liability (i.e., ensuring that no tax payments or refunds are due at the time of tax filing). Using a sample of 1992 administrative data, U.S. General Accounting Office (1996) finds that 45% of taxpayers claim the standard deduction and have only income types that are subject to third-party reporting. Goolsbee (2006) uses the IRS Individual Public Use File to calculate this figure for 2001 and finds 40%. Gale and Holtzblatt (1997) use a sample of administrative data and find that in 1994, a hypothetical improved withholding system could have accurately withheld liability for around 54% of taxpayers. U.S. Department of the Treasury (2003) conducts a similar exercise for 1999 and estimates success for 41% of taxpayers.

Our paper builds on existing work by providing estimates under the current tax law environment, and by making use of prior-year tax and information returns. As we show in Section III, our data reveal additional sources of error in pre-populated returns: mismatches between information returns and their analogous line items on the tax return, and changes in filing status and dependents from one year to the next. We also study the application of pre-populated returns to non-filers, which was not feasible in prior research.

II Analytical Approach and Data

Throughout this paper, we consider variations of the following simple procedure, where in our case $t = 2019$. The IRS begins by observing several items from a $t - 1$ tax return for filer i: the identity and age of i's dependents, i's filing status, and the identity of both the primary and secondary filer (if any).[4] The IRS then prepares a pre-populated return for year t assuming that filing status, dependent status, and the identity of the primary and secondary filers remain constant – except that we assume that a dependent who reaches age 24 in t is no longer a dependent in t.[5] In preparing the year t return, all income calculations are exclusively derived from year t information returns. Moreover, we make the strong assumption that all information returns are made available to the IRS prior to the construction of the pre-populated return.

To evaluate the success of this hypothetical production of pre-populated returns, we use a stratified random sample of 344,400 individual income tax returns for 2019 constructed by the Statistics of

[4]We also assume the IRS observes on the $t - 1$ tax return the amount of any capital loss carryforward and any amount of state income tax deducted, the latter of which is necessary to compute taxable refunds of state income tax.

[5]The rules for claiming dependents are substantially stricter for dependents above age 24.

Income Division of the IRS. These data include information from nearly every line or box on Form 1040 and its schedules. The sampling rate is increasing in income, with the highest-income taxpayers sampled with certainty. In all specifications, we use sampling weights to make the sample representative of the full tax-filing population (152 million tax units). We treat filed tax returns as the "truth". That is, we gauge the success of pre-populated returns by how closely they match these actually-filed returns. This approach is conservative, as we misclassify returns as "unsuccessful" when taxpayer-reported returns contain errors and the pre-populated return is, in fact, accurate.

We merge the sampled 2019 tax returns with their associated administrative data in 2018 and 2019. In particular, we measure income using 2019 information returns[6] and we use the 2018 Form 1040 to identify their 2018 dependents, filing status, and the identity of the 2018 primary and secondary filer.[7]

We use two methods to gauge the accuracy of pre-populated returns. First, we identify situations where the pre-populated return is almost certainly inaccurate – we refer to these as "failure situations". If a taxpayer does not have any of these failure situations, their return is deemed a "success." Because we cannot enumerate and measure all possible failure situations, we refer to this approach as an estimated "upper bound" on the success rate of pre-population. These failure situations fall into three sets. The first set includes those situations where the tax unit composition appears to have changed – e.g., a baby is born, or the taxpayer married or divorced. The second set includes situations where the taxpayer claims a tax credit or deduction that is not covered by information returns. This includes essentially everyone who itemizes their deductions, as there is no information return for state property taxes or charitable contributions, which (together) are nearly universal among itemizers.[8] The third set includes situations where the taxpayer has income that either is unreported on information returns or does not match (subject to a tolerance threshold) the income based on information returns. For example, a failure situation occurs when Schedule C (sole proprietorship) net income does not match non-employee compensation from Form 1099-MISC, which can occur if a taxpayer has other self-employment income or has expenses to deduct. In total, we identify 22 failure situations.[9]

Second, we provide a more direct test: we compare the taxpayer's self-reported tax liability to her simulated pre-populated return liability calculated using NBER TAXSIM (Feenberg and Coutts, 1993).[10] If the two amounts are within a tolerance threshold, then we deem the procedure successful. TAXSIM captures the most important features of the federal income tax system, including the EITC, the CTC, the alternative minimum tax, and the qualified business income deduction, among other features.[11] Additionally, we augment TAXSIM to incorporate the student loan interest deduction and the saver's credit. Nevertheless, the calculator has a limited set of input variables and therefore cannot accommodate some of the data derived from information returns, including the reconciliation of premium tax credits received under the Affordable Care Act. In addition, TAXSIM imposes some constraints

[6] We use income from the following information returns: Form W-2, Form W-2G, Form 1099-DIV, Form 1099-INT, Form 1099-G, Form 1099-R, Form 1099-SSA, Form 1099-B, Form 1099-MISC, Form 1099-K, and Forms 1065 and 1120S (Schedule K-1). We note that some 2019 Forms 1099-MISC appear to be missing from the database, but this should not substantially affect our results; see Appendix A.1.4.

[7] See Appendix A.1.3 for details on our assumptions regarding dependents.

[8] We calculate the earned income tax credit (EITC) and child tax credit (CTC) automatically, based on assumed dependent and eligibility status from 2018 and calculated income. The presence of EITC or CTC is not a failure situation.

[9] In Appendix A.2, we describe all of the failure situations that we consider.

[10] For this purpose, we conform the definition of "tax liability" to match the TAXSIM definition. Tax liability is defined as "total tax" (Line 16 of 2019 Form 1040) less the sum of self-employment tax and all refundable credits. For our purposes, "refundable credits" do not include direct payments (e.g., withholding or estimated tax payments) or excess Social Security tax withheld. This concept of tax liability does not include payroll (FICA) tax, though it does include the 0.9% Additional Medicare Tax.

[11] To protect taxpayer privacy, we use a local version of TAXSIM; however, the calculator used is identical to the public version available here: `http://www.nber.org/taxsim/`.

upon its input data that simplify, and thus imperfectly reflect, the tax code. Therefore a richer calculator may accurately pre-populate additional returns. For this reason, we refer to this exercise as a "lower bound" the success rate of pre-population.

We note that these "bounds" are not strict upper and lower bounds in the mathematical sense. For example, a tax unit may have a failure situation that affects its income but not its tax liability if its taxable income would have been zero regardless. This would result in a lower bound success but an upper bound failure. Additionally, a given return may have multiple failure situations that happen to have perfectly offsetting effects on liability, again creating a lower bound success but an upper bound failure. These situations are more common for certain groups of taxpayers, which can result in an estimated "upper bound" success rate that falls below the "lower bound", as indeed occurs for some income groups in Figure 1 (described later).

III Results

Using our preferred tolerance levels, our baseline results are that 42 and 48 percent of pre-populated returns are successful under our lower and upper bound approaches, respectively, as shown in Table 1. For the lower bound approach, we consider returns with differences between pre-populated tax liability and taxpayer-reported liability of \$100 or less to be a success. For the upper bound approach, we similarly allow differences between pre-populated and taxpayer-reported tax credits of up to \$100, while for income and deduction amounts we allow differences of up to \$500. Implicitly this assumes a 20 percent effective tax rate, such that a \$500 adjustment to income would change tax liability by \$100. The probability of success varies as we alter the error tolerances, but within a wide range of tolerances we find that between 33 and 60 percent of taxpayers have an accurate pre-populated return.

In Table 1, we also consider an alternative pre-population regime wherein the IRS first solicits accurate information about current year t filing status and dependents, then pre-populates a return using income from information returns. Perfectly implemented, this exercise eliminates dependent and filling status errors, which improves the success rate of pre-population by around five percentage points (7.6 million taxpayers) using our preferred tolerance thresholds. We note, though, that procuring this additional information may be costly for the IRS and taxpayers alike.

In Figure 1 we show that both our upper and lower bound approaches have higher rates of success for low- to moderate-income taxpayers, classified by the AGI reported on the tax return. In the graph, each data point represents an income ventile (conditional on positive AGI) and therefore each represents the same number of taxpayers. For taxpayers in the bottom three ventiles, with average incomes between \$0 and \$10,000, success rates are between 55 and 80 percent. Success rates remain high well into the middle part of the income distribution. For example, taxpayers earning around \$60,000 (the thirteenth ventile) have success rates between 38 and 51 percent. The graph illustrates the looseness of our "upper bound" and "lower bound" terminology, as the upper bound lies below the lower bound for the bottom four income ventiles. This is because many of these taxpayers have taxable income of \$0, meaning that the presence of a failure situation need not imply an incorrect tax liability calculation.

Next we examine the prevalence of each of the 22 failure situations that underlie our upper bound approach in column 1 of Table 2, which is sorted from most to least common. The most common failure situation is the presence of Schedule C income that doesn't match Form 1099-MISC non-employee compensation, which occurs for 16 percent of taxpayers. Next, around 11 percent of taxpayers itemize deductions. Taxpayers with these situations would need to file an additional schedule (C or A, respectively). In contrast, the next most common failure scenario is a nontrivial mismatch between Form W-2 and reported wage income. The nine percent of taxpayers with this issue may only need to edit one line of the tax return.[12] In column 2, we show the share of taxpayers who have a given failure situation and no other failure situations; they would be an upper bound success *but for* that particular failure. For the three most common failure situations, the share of taxpayers who only have that failure situation is 35-55% of the total share who have that failure situation. Overall, among the 52% of taxpayers with any failure situation, 56% ($\frac{.293}{.519}$) have exactly one failure situation.

In Figure 2, we investigate further the relationship between success rates and income. In particular, we plot the share of tax units with several common upper bound failure situations as a function of adjusted gross income. Itemization, the presence of rents/royalties, and pension/IRA mismatches are increasing in income. Wage and Schedule C mismatches exist across the income distributions, but are hump shaped, with noticeable increases in the range of income typically associated with EITC returns.[13] Dependent mismatches are comparatively flatter with respect to income.

Table 3 provides summary statistics on the taxpayers for whom pre-population is successful under the two approaches. On average, taxpayers for whom pre-population works are several years younger and 21 to 26 percentage points less likely to be married than taxpayers for whom the pre-population exercise is unsuccessful. As expected, taxpayers who have successful pre-populated returns are less likely to have dependents. Consistent with Figure 1, the group of taxpayers for whom pre-populated returns is successful have lower average income. In the upper bound approach, the successful group have average incomes of $47,200 versus $104,000 in the unsuccessful group. For the unsuccessful group under the lower bound approach, we separately consider those where pre-populated tax liability is too high (column 4) and those where it is too low (column 5). The group for whom the pre-populated liability is too low are particularly high income, with average incomes of $142,200.

The primary benefit of a pre-populated tax return is that it may reduce and redistribute the costs of tax filing. Table 3 illustrates this in columns (1) and (3), as we see that 43 percent of taxpayers for whom pre-populated returns are accurate currently use paid preparers. Moreover, among individuals for whom pre-populated returns are inaccurate, the median number of failure situations is one. That is, most individuals who would need to modify their pre-populated returns would only need to edit one line or schedule (e.g., filling out a Schedule C – and any other schedules required to be attached to that Schedule C – which would be the most common edit required, as we saw in Table 2). For this group the pre-populated return may substantially reduce the costs of tax filing as well.

III.A Success Rates by Types of Income and Tax Unit Composition

Next, in the spirit of Goolsbee (2006), we investigate the extent to which pre-population might be more successful for certain types of income or subsets of taxpayers with simple tax situations. These analyses can inform policymakers considering more limited approaches to implementing pre-population.

To inform this analysis, Table 4 reports the success rate for pre-populating specific lines of the 2019 Form 1040. Following our baseline upper bound approach, for income lines we use a tolerance of being within $500 and for credits we use a tolerance of $100. Column (1) shows the share of taxpayers where the pre-populated returns match the actually-filed returns, including cases where both are zero. This represents the share of individuals who would need to make no corrections to that line.

[12] In some circumstances, such as unreported tip income, additional forms are required.

[13] The bunching of Schedule C filers in the EITC range has been well known since at least Saez (2010). Our results on the income patterns of W-2 wage mismatches are consistent with those found in Mortenson and Whitten (2020).

In column (2), we restrict attention to those observations where the line on the actually-filed return is nonzero, which speaks more directly to the relative accuracy of our procedure for that line. The wage line is pre-populated accurately for 92% of taxpayers, including 90% of those who report wages on their tax return. Conditional match rates are similar for taxable interest and qualified dividends, but the unconditional match rates are higher due to the presence of many zeros. Conditional match rates are somewhat lower for pensions (81%) and are much lower for capital gains (44%) and Schedule 1 income (23%), a category that includes Schedule C (sole proprietor) income, Schedule E income (rents, royalties, S corporation income, and partnership income), unemployment compensation, and other less common income types.

Given the higher success rates for certain types of income, Table 5 analyzes the success of pre-population starting with a subset of taxpayers with simple tax situations covered by information returns and then gradually broadening the target population. Of course, income types are not the only dimension of simplicity; holding income types constant, being married, having dependents, having high income, or having certain deductions or credits can introduce complexity. Thus, for our narrowest set of taxpayers, we restrict to those who are single, with no dependents, whose only income source is wages, whose income is under $100,000, and who have no unobserved credits or deductions.[14] Around one-fifth of 2019 filers satisfy these restrictions, and of these, we find that between 80 and 82 percent of pre-populated returns would be successful (Table 5). It is noteworthy that success rates are substantially less than 100 percent even for such simple situations. The most common reason for failure, under the upper bound approach, is a mismatch between Form W-2 and taxpayer-reported wages, which occurs in 46 percent of failures within the narrowest subset.

When we expand the subset of taxpayers analyzed to include married couples with dependents, this increases the number of eligible tax units by 50 percent, but the success rates fall to a range of 68 to 74 percent, as indicated in columns (2) and (3). Columns (4) and (5) describe the success rates among the people *added* in each row. That is, among tax units who are included in row 2 but not row 1, we see a success rates of 54 to 55 percent. The relatively low success rate in this group is driven largely by transitions in filing status (including from non-filing). Among those included in row 2 but not row 1 with an upper bound failure situation, 35 percent have a mismatch between this year's filing status and last year's filing status.

When we allow pre-populated returns to have additional sources of income that are at least partially covered by information returns in rows 4 through 8, the share of eligible tax units increases by 17 percentage points without a meaningful effect on success rates. In row 9, we add higher-income taxpayers, whose success rates are somewhat lower. In row 10, we allow all income types, increasing coverage by 17 percentage points. However, as expected, the marginal success rates for this group are quite low: 23 and 19 percent, respectively, under our lower and upper bound approaches. Moving to the full set of taxpayers in row 11, we recover our baseline results, with the success rate ranging from 42 to 48 percent.[15]

One limitation of this exercise is that it does not inform a procedure that could be easily implemented because the IRS would not know which categories taxpayers fall under when pre-populating returns. A more readily implemented procedure would define taxpayer subsets based on the informa-

[14]This last restriction eliminates people who itemize deductions, who claim above-the-line deductions for moving expenses, educator expenses, or health insurance for self-employed persons, or who claim any credit other than the EITC, CTC, American Opportunity Tax Credit, or credit for excess Social Security tax.

[15]By construction, upper-bound marginal success rates in rows 10 and 11 are positive only when items are smaller than our tolerance thresholds.

tion from the prior year's (i.e., 2018) return. Appendix Table A2 considers this alternative. The success rates in Appendix Table A2 are universally lower than in Table 5 except in the final row, where they are mechanically equal. The success rates are five (upper bound) and nine (lower bound) percentage points lower in the narrowest subset. This largely reflects the additional source of error caused by year-to-year changes. For instance, some taxpayers with only wages in 2018 might have self-employment income in 2019.

The five percentage point upper-bound gap remains relatively stable through row 8 (adding capital gains), while the nine percentage point lower-bound gap reduces to five percentage points. Both gaps shrink to two percentage points when we allow all income types. In sum, requiring taxpayers to provide categorical information prior to a pre-populated return has a modest effect on accuracy rates. This result is similar to our result from Table 1 that shows that if taxpayers provided the IRS updated information about filing status and dependents prior to pre-population overall success rates would increase by around five percentage points. Policymakers contemplating these options would need to weigh the increases in accuracy from obtaining updated information against the additional burdens faced by taxpayers and administrators.

III.B Non-filers

One potential benefit of pre-populated returns is that it could encourage individuals who otherwise would not file to do so. This could help low-income taxpayers access refunds either through eligibility for the EITC, the CTC, or from tax withholding. It could also encourage compliance among non-filers with a filing obligation.

The IRS already creates pre-populated returns for certain non-filers who appear to owe significant liability, under the Automated Substitute for Return (ASFR) program (Treasury Inspector General for Tax Administration, 2017). These returns assist with IRS enforcement procedures and are sent out along with letters to delinquent taxpayers indicating proposed tax assessments. Liability is based on income from information returns, and no dependents appear on the returns. However, the ASFR program is limited in application. For example, it was temporarily suspended in parts of 2015, 2016, and 2017 due to IRS resource constraints, and in fiscal year 2019 only 380,349 cases were selected (National Taxpayer Advocate, 2019).

We examine results for a subset of non-filers in Table 6. The table is based on a 0.1% random sample of the 54.7 million 2019 non-filers who would have received a pre-populated return under our baseline policy simulation. These are individuals between the ages of 18 and 105 who (i) appear on a 2019 information return with an address in the 50 states or the District of Columbia, and (ii) did not file a 2019 tax return as a primary or secondary filer.[16]

We separate our analysis by those who do and do not appear to have a legal requirement to file a tax return.[17] Our sample, appropriately weighted, indicates that 46.3 million non-filers who receive information returns appear not to be required to file. On average, these non-filers had only $37 in taxes withheld. Seventeen percent (8.0 million) have a potential refund, and conditional on having a refund the average amount is $386. This is consistent with evidence that, for many non-filers, the time and

[16] We collected data on non-filers on June 8, 2022. We misclassify anyone who files after our data retrieval as a "non-filer". We exclude individuals who filed jointly in 2018 if their spouse filed a 2019 return. We stress that this is not the full population of non-filers, as it excludes individuals who do not receive any information return.

[17] Our proxy for the filing obligation is an indicator for whether the individual's pre-populated AGI calculation exceeds the standard deduction of $12,200 ($13,850 for those age 65 or older). Our estimate of non-filers with an obligation to file is consistent with Erard et al. (2014).

financial costs of filing exceed the benefits (Erard and Ho, 2001). However, a pre-populated return could help eliminate some of these barriers to claiming a refund.

One reason why potential refunds are low for these non-filers is that only two percent claimed dependents on the prior-year return. Thus for the vast majority, pre-populated returns would not include child tax benefits such as the CTC and any EITC amounts would be small under 2019 law. Nonetheless, for around one percent of these non-filers (337,000 taxpayers), pre-populated returns may increase child EITC and CTC take-up. One might expect that this number would be substantially higher under a pre-population regime wherein the IRS first solicits information about dependents and then pre-populates returns, but the evidence suggests otherwise. During the COVID-19 pandemic, the IRS solicited information on dependents from non-filers for the purpose of distributing Economic Impact Payments in 2020 and 2021 as well as the 2021 advanceable CTC. In row 7, we make use of this information by assigning dependents claimed via either of these non-filer tools to 2019 non-filers.[18] When we do so, the number of non-filers without a filing obligation who are owed child EITC or CTC increases by only 24,000. Many more of our sampled non-filers without a filing obligation – eight percent – appear eligible for the childless EITC, consistent with Guyton et al. (2016). If pre-populated returns provide a reminder to file or lower filing costs, they may increase EITC take-up (Goldin et al., 2021).

We also estimate that 8.4 million non-filers appear to be failing to meet a legal filing obligation. For the majority (55 percent), pre-populated returns indicate a balance due. Reflecting the general skew in income tax liability, the mean balance due among those who owe is $5,039, with a median of $1,500. We stress that this calculation uses information returns only; these taxpayers may have deductions that we do not observe that more than fully offset their income. Additionally, these taxpayers may have filed a late return after our data retrieval.

IV Discussion and Conclusion

Our results suggest that pre-populated returns would be accurate for a substantial share of U.S. taxpayers. With our preferred set of tolerance thresholds, we estimate that between 42 and 48 percent of taxpayers could be sent an accurate pre-populated return. Success rates are much higher for low- and moderate-income taxpayers and those who do not have unobserved deductions or self-employment income. Furthermore, most taxpayers could benefit from a pre-populated system by receiving a return with a large number of lines correctly pre-populated as most inaccurate returns have precisely one aspect (i.e., one line or one schedule) of the return that requires editing. Success rates for some lines on the Form 1040 are quite high. We calculate wages, taxable interest, and taxable dividends correctly over 90 percent of the time. Non-filers and taxpayers that fail to take up the EITC or CTC could see significant benefits (Plueger, 2009; Guyton et al., 2016). We estimate that there are eight million non-filers who are not required to file yet are owed refunds averaging $386.

However, other important considerations must be weighed when appraising whether, or how, a pre-population program should be offered. A detailed analysis of these issues is beyond the scope of this paper, but we briefly discuss some aspects here. First, any pre-populated return program would shift some of the filing burden to information reporters and tax administrators. It would require an earlier deadline for third-party information reports to avoid delays in refunds relative to the current filing deadline (April 15th), and the IRS would have to quickly process and match information returns to generate pre-populated returns. At present, there are no up-to-date estimates of the costs to the federal

[18] We additionally require that those dependents are not claimed by someone else in 2019.

government of moving to a tax agency reconciliation system, and previous estimates varied widely (U.S. Department of the Treasury, 2003). Updated technology and improved computing capabilities may lower these costs, and a reduction in taxpayer errors may save some current tax enforcement costs. The California experience suggests that administrative costs of pre-populated returns at the state level may be cost-effective (Bankman, 2005). Thus a well-designed federal program may achieve overall compliance cost reductions.

Second, pre-populated returns may change taxpayer behavior and overall revenue in unpredictable ways. The pre-populated return could function as a "nudge", encouraging taxpayers to accept the returns without adjustment – causing effective tax liability to differ from what it is under current law. This may occur due to the general tendency of individuals to accept default options (Sunstein, 2013) or due to a (perhaps mistaken) belief that accepting a pre-populated return would reduce one's audit likelihood (Martinez-Vazquez and Sanz-Arcega, 2020). The pre-populated return may also inform taxpayers about the set of information that the IRS has, which could in turn change reporting behavior.

When Finland introduced pre-populated returns, taxpayers reported less income from sources not subject to information reporting but were also less likely to claim deductions and expenses for which there were no information returns. The changes were concentrated among low dollar amounts and lower-income taxpayers, and there was no appreciable overall liability change (Kotakorpi and Laamanen, 2017). Experimental evidence is also consistent with the idea that pre-populated returns can nudge taxpayers to report different amounts. Studying 554 U.K. residents asked to assume the role of a fictitious taxpayer, Fonseca and Grimshaw (2017) find that when pre-populated returns overestimate liability, a larger share of participants pay more than they owe, relative to a regime with a blank tax return. Likewise, among those presented with an underestimate of liability, a larger share reported less liability. These "default" effects of pre-populated tax returns are consistent with other experimental studies, though some suggest that taxpayers are less likely to correct underestimated tax liability than overestimated liability (Bruner et al., 2015; Fochmann, Müller and Overesch, 2018; van Dijk et al., 2020). The introduction of new information returns covering currently unreported income or deductions could improve the accuracy of the pre-populated return, lessening this "nudge" effect (Gillitzer and Skov, 2018).

To the extent that taxpayers would accept a pre-populated return without adjustment, this would have disparate effects on tax liability across the income distribution. For the bottom 95 percent of the distribution of reported income, pre-populated tax liability is more often overestimated rather than underestimated, relative to taxpayer-reported liability on Form 1040; we show this in panel (a) of Figure 3. The magnitudes of these errors compound the effect. In panel (b) we see that for most income groups, the average overestimated liability is larger in magnitude than the average underestimated liability. Thus, for the bulk of the population, pre-populated liabilities exceed taxpayer-reported liabilities by a wide margin, in aggregate. Indeed, if all taxpayers outside the top five percent (by reported income) simply accepted their pre-populated returns without amendment, tax receipts would rise by $96 billion, around 19 percent of their reported federal individual income tax liability.[19]

This pattern reverses at the top of the income distribution. Among taxpayers in the top five percent of reported income, underestimated liability is slightly more common than overestimated liability. For these taxpayers, the magnitude of the average underestimate far exceeds the average overestimate, likely due to income types that are subject to less information reporting. These taxpayers are few in

[19]addition, $15 billion in tentative liability in excess of withholdings among 2019 non-filers potentially would have further increased receipts.

number but report 64 percent of overall federal income tax liability. Thus their behavioral response to pre-populated tax returns – as well as details of the proposed pre-populated return policy, such as income limits – would be crucial to determining the aggregate effect on tax revenue.[20]

Lastly, it is possible that pre-populated returns could improve compliance to the extent that they reduce misreporting on lines covered by information returns (and to the extent that such misreporting is not already recovered through post-return enforcement). In our analysis, we have treated the filed return as the "truth", with mismatches between information returns and the tax return reflecting the unique circumstances of the taxpayer which we do not observe. However, it is well-known that income misreporting on Form 1040 is substantial (Internal Revenue Service, 2019; U.S. Department of the Treasury, 2021). While most of this misreporting occurs on lines that are not covered by information returns – such as Schedule C income and rental income – others have found evidence consistent with strategic misreporting on lines covered by information returns as well (Mortenson and Whitten, 2020; Ramnath and Tong, 2017). Pre-populated returns could, in principle, improve compliance along this margin (Fochmann, Müller and Overesch, 2018).

This insight reveals that our estimated accuracy rates – relative to returns filed in the absence of misreporting – may be biased downwards. That is, some mismatches on information-return covered lines, such as wages, dividends, interest, and pensions, may be attributable to taxpayer error. We can bound the effect of such misreporting by repeating our upper bound analysis but removing from the set of failure conditions any mismatches on the wages, interest, dividends, or pension/IRA lines. This raises our upper bound success rate from 48 to 56 percent. This eight percentage point jump inevitably overstates the role of taxpayer misreporting, though, since some mismatches reflect errors on the part of the information return issuer or have other legitimate explanations.

In sum, we find that pre-populated returns have the potential to substantially reduce filing burdens for a large majority of taxpayers using existing information available to the IRS. While implementation details are challenging, we hope that this exploration will inform the discussion of innovative policies regarding pre-populated tax returns.

References

Bankman, Joseph. 2005. "Simple Filing for Average Citizens: The California ReadyReturn." *Tax Notes*, 107.

Bankman, Joseph. 2008. "Using Technology to Simplify Individual Tax Filing." *National Tax Journal*, 61: 773–789.

Bruner, David, Michael Jones, Michael McKee, and Christian Vossler. 2015. "Tax reporting behavior: Underreporting Opportunities and Pre-populated tax returns." *Department of Economics, Appalachian State University*, 15-11.

Cordes, Joseph, and Arlene Holen. 2010. "Should the Government Prepare Individual Income Tax Returns?" *SSRN Working Paper*.

Erard, Brian, and Chin-Chin Ho. 2001. "Searching for ghosts: who are the nonfilers and how much tax do they owe?" *Journal of Public Economics*, 81: 25–50.

Erard, Brian, Pat Langetieg, Mark Payne, and Alan Plumley. 2014. "Missing Returns vs. Missing Income." *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 107: 1–21.

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

[20]For these taxpayers, pre-populated returns show $110 billion lower tax liability than their actual returns, around 12% of their reported federal individual income tax liability.

Fichtner, Jason, William Gale, and Jeff Trinca. 2019. "Tax Administration: Compliance, Complexity, and Capacity." Bipartisam Policy Center.

Fochmann, Martin, Nadja Müller, and Michael Overesch. 2018. "Less cheating? The effects of prefilled forms on compliance behavior." *arqus Discussion Paper*, , (227).

Fonseca, Miguel, and Shaun Grimshaw. 2017. "Do Behavioral Nudges in Prepopulated Tax Forms Affect Compliance? Experimental Evidence with Real Taxpayers." *Journal of Public Policy & Marketing*, 36: 213–226.

Gale, William, and Janet Holtzblatt. 1997. "On the possibility of a no-return tax system." *National Tax Journal*, 50: 475–485.

Gillitzer, Christian, and Peer Skov. 2018. "The use of third-party information reporting for tax deductions: evidence and implications from charitable deductions in Denmark." *Oxford Economic Papers*, 70: 892–916.

Goldin, Jacob, Tatiana Homonoff, Rizwan Javaid, and Brenda Schafer. 2021. "Tax Filing and Take-Up: Experimental Evidence on Tax Preparation Outreach and EITC Participation." National Bureau of Economic Research.

Goolsbee, Austan. 2006. "The 'Simple Return': Reducing America's Tax Burden Through Return-Free Filing." *Hamilton Project Discussion Paper*.

Guyton, John, Dayanand Manoli, Brenda Schafer, and Michael Sebastiani. 2016. "Reminders & Recidivism: Evidence from Tax Filing & EITC Participation among Low-Income Nonfilers." *NBER Working Paper*, , (21904).

Internal Revenue Service. 1987. "Current Feasibility of a Return-Free Filing System."

Internal Revenue Service. 2019. "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013." IRS Research Applied Analytics and Statistics 1415.

Kotakorpi, Kaisa, and Jani-Petri Laamanen. 2017. "Prefilled income tax returns and tax compliance: Evidence from a natural experiment." *Working Paper*.

Martinez-Vazquez, Jorge, and Eduardo Sanz-Arcega. 2020. "Can prepopulated tax returns enhance tax compliance? The effects of the Spanish Renta Web initiative from a sociology of taxaption perspective." *International Center for Public Policy Georgia Statue University Working Paper*, 20-09.

Mortenson, Jacob A., and Andrew Whitten. 2020. "Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule." *American Economic Journal: Economic Policy*, 12(3): 402–432.

National Taxpayer Advocate. 2019. "Annual Report to Congress." Publication 2104. Catalog Number 23655L.

OECD. 2021. "Tax Administration 2021: Comparative Information on OECD and other Advanced and Emerging Economies." *OECD Publishing*.

Plueger, Dean. 2009. "Earned Income Tax Credit Participation Rate for Tax Year 2005." *IRS Research Bulletin*.

Ramnath, Shanthi P., and Patricia K. Tong. 2017. "The Persistent Reduction in Poverty from Filing a Tax Return." *American Economic Journal: Economic Policy*, 9(4): 367–394.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points." *American Economic Journal: Economic Policy*, 2(3): 180–212.

Sunstein, Cass R. 2013. "Deciding by Default." *University of Pennsylvania Law Revue*, 162(1): 1–58.

Treasury Inspector General for Tax Administration. 2017. "A Significantly Reduced Automated Substitute for Return Program Negatively Affected Collection and Filing Compliance." 2017-30-078. September 29, 2017.

Treasury Inspector General for Tax Administration. 2021. "Effects of the COVID-19 Pandemic on Business Tax Return Processing Operations." 2021-46-064. September 2, 2021.

U.S. Department of the Treasury. 2003. "Report to the Congress on Return-Free Tax Systems: Tax Simplification is a Prerequisite."

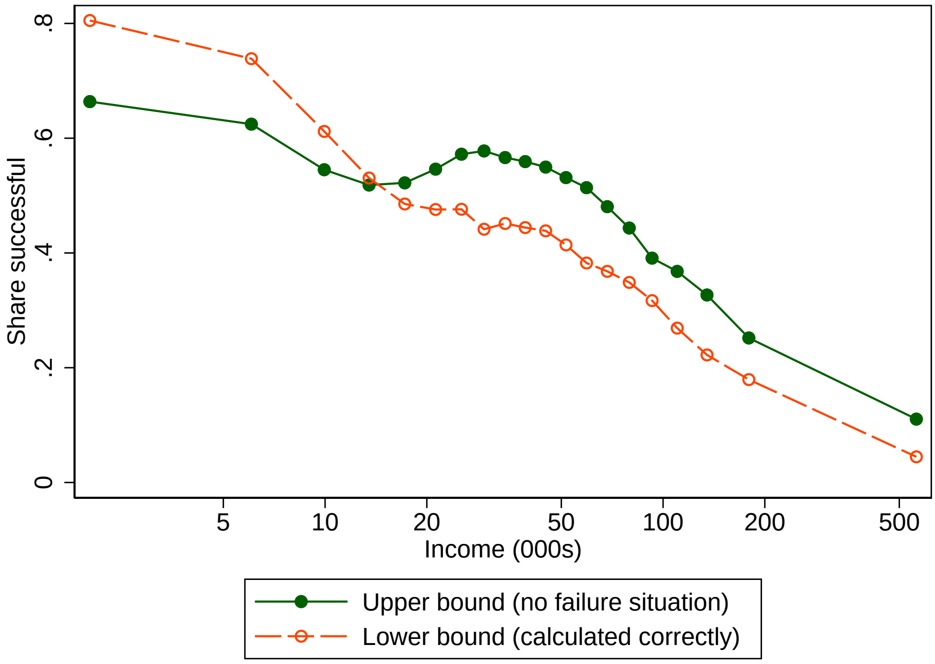
U.S. Department of the Treasury. 2021. "The American Families Plan Tax Compliance Agenda."

U.S. General Accounting Office. 1996. "Tax Administration: Alternative Filing Systems." GAO/GGD-97-6.

van Dijk, Wilco, Sjoerd Goslinga, Bart Terwel, and Eric van Dijk. 2020. "How choice architecture can promote and undermine tax compliance: Testing the effects of prepopulated tax returns and accuracy confimation." *Journal of Behavioral and Experimental Economics*, 87.

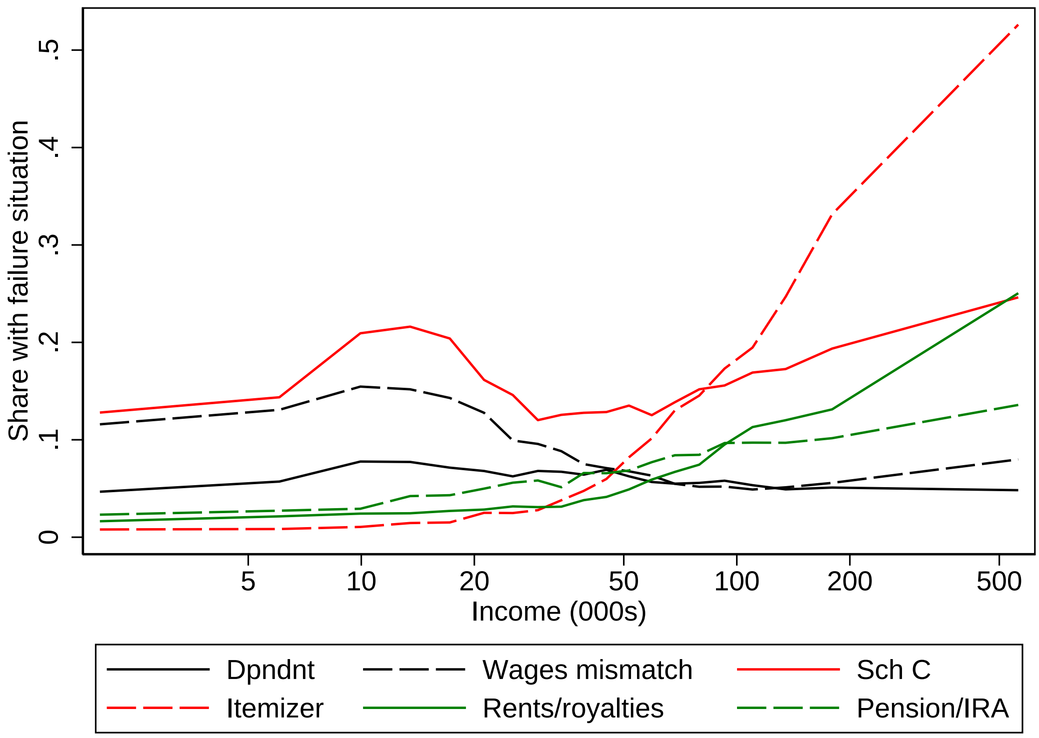
Tables and Figures

Figure 1: Pre-populated success rates by income



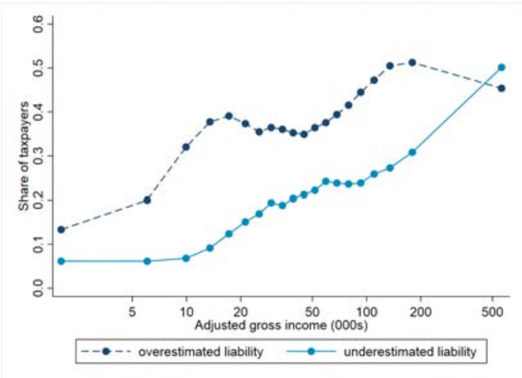
Notes: The figure plots the rate of accurately pre-populated tax returns for the 2019 tax filing population as a function of taxpayer-reported adjusted gross income under each of our two approaches. We use the preferred tolerance thresholds given in column 3 of Table 1. Data points are shown at the average AGI within each AGI bin. AGI bins are defined by ventiles among the subset with positive AGI; taxpayers with zero and negative AGI are excluded. The x axis has a log scale. The figure was created by the authors using IRS tax data.

Figure 2: Pre-population success by Income by Situation

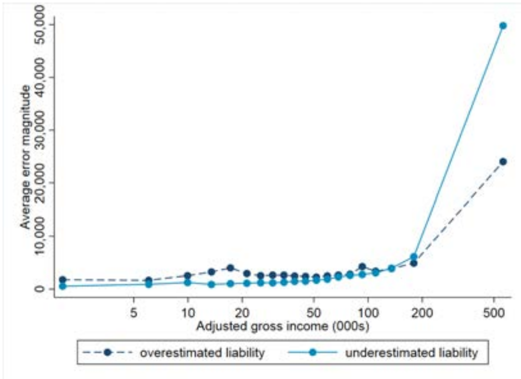


Notes: The figure plots the rate of certain common failure situations for the 2019 tax filing population as a function of taxpayer-reported adjusted gross income. We use the preferred tolerances, given in column 3 of Table 1. Data points are shown at the average AGI within each AGI bin. AGI bins are defined by ventiles among the subset with positive AGI; taxpayers with zero and negative AGI are excluded. The x axis has a log scale. Additional detail on certain failure situations is given in Appendix A.2. The figure was created by the authors using IRS tax data.

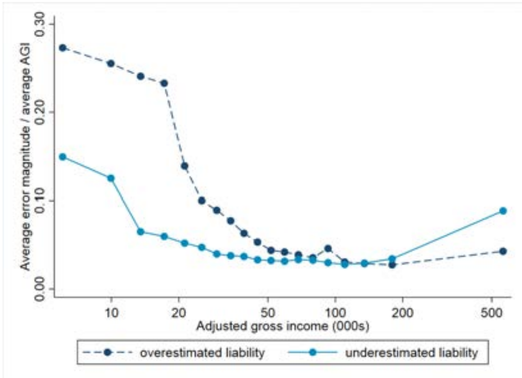
Figure 3: Lower bound direction and magnitude of liability errors



(a) Share with errors



(b) Magnitude of errors



(c) Magnitude of errors relative to AGI

Notes: The figure plots information for two subsets of the 2019 tax filing population: those for whom the difference between their simulated pre-populated tax liability and reported liability exceeds \$100, and those for whom the difference is negative and exceeds \$100 in magnitude. Panel (a) plots the share of each group by AGI bin. Panel (b) plots the average magnitude of the difference between the simulated pre-populated tax liability and the reported liability within AGI bin for each group. Panel (c) plots the averages error magnitudes of panel (b) divided by average AGI within each AGI bin. Sample weights are used; thus the results are representative of the 2019 tax filing population. Data points are shown at the average AGI within each AGI bin. AGI bins are defined by ventiles among the subset with positive AGI; taxpayers with zero and negative AGI are excluded. The x axis has a log scale. Panel (c) omits the bottom income ventile, which has large average errors relative to its small average AGI. The figure was created by the authors using IRS tax data.

Table 1: **Pre-population Success rates**

	Strictest (1)	(2)	Preferred (3)	(4)	Least strict (5)
Panel A: Lower bound					
Baseline	0.342	0.389	**0.424**	0.470	0.556
With known deps. & filing status	0.383	0.435	0.475	0.525	0.617
Tax difference tolerance	$10	$50	$100	$200	$500
Panel B: Upper bound					
Baseline	0.432	0.448	**0.481**	0.509	0.598
With known deps. & filing status	0.476	0.494	0.530	0.562	0.662
Credit tolerance	$10	$50	$100	$200	$500
Income/deduction tolerance	$100	$200	$500	$1000	$5000

Notes: This table reports the rates at which we are able to accurately pre-populate returns for 2019 tax filers. Across the columns, we relax our tolerance thresholds for defining an observation to have an "accurate" pre-populated return. Panel A shows results for the lower-bound approach, so the relevant tolerance is the acceptable difference between simulated and actual tax liability. Panel B shows results for the upper-bound approach, which requires two separate tolerances: one for failure situations related to credit amounts, and one for failure situations related to income/deduction amounts. Our preferred tolerances are given in the third column. Within each panel, the first row uses the baseline procedure. The second row considers a counterfactual where the IRS has full knowledge of filing status and dependency.

Table 2: **Share with each upper bound failure situation**

Situation	Share with this failure situation (1)	Share with *only* this failure situation (2)
Schedule C income does not match 1099-MISC NEC	0.162	0.075
Itemized deductions in 2019	0.109	0.040
Wages do not match W-2 wages	0.089	0.050
Taxable pension/IRA income does not match 1099-R	0.067	0.017
Sched E rents/royalties (except from K-1)	0.065	0.014
Dependent mismatch from 2018 to 2019	0.061	0.019
Child/dependent care credit	0.039	0.017
Capital gains income does not match 1099-B	0.036	0.008
S corp/partnership income does not match K-1	0.034	0.003
Change in filing status from 2018 to 2019	0.034	0.012
1099-R with taxable amount not determined	0.033	0.003
EITC dependent mismatch from 2018 to 2019	0.029	0.006
Interest does not match 1099-INT	0.028	0.006
Certain above-the-line deductions	0.023	0.001
Dividends do not match 1099-DIV	0.021	0.005
Section 199A deduction complications	0.020	0.000
Lifetime learning credit	0.014	0.008
Residential energy credits	0.011	0.004
Schedule F income	0.010	0.003
Noncapital gains (Form 4797)	0.009	0.000
Received or paid alimony	0.005	0.002
Early distribution from Roth IRA	0.005	0.000
Any failure situation	0.519	
Precisely one failure situation	0.293	

Notes: The table reports the share of tax units that have each failure situation in the 2019 tax filing population. We use the preferred tolerances, given in column 3 of Table 1, Panel B. Column (1) provides the share of tax units with each failure situation; the table is sorted in descending order by column (1). Column (2) provides the share of tax units who have the given failure situation and no other failure situation. The bottom row gives the share of tax units with at least one failure situation. This table includes the full set of failure situations we identify. Additional detail on certain failure situations is given in Appendix A.2.

Table 3: **Taxpayer characteristics conditional on success/failure**

	No failure situation? (upper bound)		Correct calculation? (lower bound)		
	✓	X	✓	high	low
	(1)	(2)	(3)	(4)	(5)
Mean					
Married	0.26	0.48	0.24	0.45	0.50
Has dependents	0.24	0.39	0.17	0.48	0.34
Uses paid preparer	0.43	0.58	0.43	0.55	0.61
Primary filer age	42	48	43	45	51
AGI (thousands)	47.2	104.0	39.3	83.9	142.2
Liability: calculated (thousands)	3.8	15.2	3.0	13.8	16.5
Liability: taxpayer-reported (thousands)	3.7	15.1	3.0	9.1	24.7
Liability: calculated less reported (thousands)	0.1	-0.3	0.0	4.6	-8.2
Failure situations	0.0	1.7	0.2	1.4	1.4
Median					
Primary filer age	39	47	38	44	51
AGI (thousands)	33.5	53.0	27.6	50.7	64.6
Liability: calculated (thousands)	1.5	2.7	0.9	3.7	2.2
Liability: taxpayer-reported (thousands)	1.4	2.4	0.9	1.8	4.4
Liability: calculated less reported (thousands)	0.0	0.4	0.0	1.7	-1.2
Failure situations	0	1	0	1	1
Count (millions)	73.1	78.9	64.5	57.1	30.4

Notes: The table provides mean and median values for 2019 tax unit characteristics conditional on either success or failure of our upper-bound and lower-bound approaches under our preferred tolerance thresholds. Sample weights are used; thus the results are representative of the 2019 tax filing population. The first column describes taxpayers for whom the upper bound approach is successful, while the second column describes those for whom it is unsuccessful. The third column describes taxpayers for whom the lower bound approach is successful. The fourth and fifth columns describe taxpayers for whom the lower bound approach is unsuccessful separated by whether the simulated tax is too high or too low, respectively, relative to reported liability.

Table 4: **Success rate for selected lines on Form 1040**

Line	Unconditional match rate	Conditional match rate
	(1)	(2)
Wages	0.918	0.903
Taxable interest	0.970	0.919
Qualified dividends	0.977	0.905
Income from Sch. 1	0.700	0.241
Taxable IRA and pensions	0.929	0.806
Capital gains	0.893	0.442
AGI	0.528	0.531
Taxable income	0.541	0.471
EITC	0.892	0.456
Child tax credit	0.856	0.628

Notes: This table reports the match rates for several lines on Form 1040. Column (1) reports the share of observations with a non-zero value on the relevant line. In column (2), among this subset, we calculate the share of observations whose values derived from information returns matches the amounts reported on Form 1040. In the rows marked "AGI" and "taxable income", we do not make the restriction to non-zero values. In all rows, we restrict to the set of tax units whose composition remained constant from 2018 to 2019. We use a tolerance margin of $500 in all rows, except that we use a tolerance margin of $100 in the EITC and the Child Tax Credit rows. For the purpose of this table, the Child Tax Credit includes both the refundable and non-refundable part aggregated together.

Table 5: **Pre-population success rates for subsets of taxpayers defined based on 2019 tax returns**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.20	0.80	0.82	0.80	0.82
2. Allow married	0.22	0.78	0.80	0.55	0.54
3. Allow dependents	0.30	0.68	0.74	0.42	0.59
4. Add interest/dividends	0.34	0.69	0.75	0.75	0.82
5. Add Social Security	0.36	0.69	0.75	0.76	0.78
6. Add pension/IRA distributions	0.43	0.68	0.74	0.64	0.70
7. Add gambling, UI, state tax refunds	0.44	0.68	0.74	0.54	0.71
8. Add capital gains	0.47	0.68	0.73	0.60	0.60
9. Add high income	0.52	0.66	0.73	0.51	0.68
10. All income types	0.69	0.55	0.60	0.23	0.19
11. **Broadest**: Eliminate deduction and credit restrictions	1.00	0.42	0.48	0.14	0.22

Notes: This table reports the rates at which we are able to accurately pre-populate returns for certain subsets of 2019 tax filers. We use our preferred tolerances in Table 1. Subsets are defined based on characteristics from the 2019 tax return. "Unobserved deductions" include all itemized deductions and above-the-line deductions for moving expenses, educator expenses, and health insurance for self-employed persons. "Unobserved credits" are all credits other than the EITC, Child Tax Credit, and American Opportunity Tax Credit. The cumulative success rate (columns 2 and 3) reflects the success rate for all individuals up to and including the row in question. The marginal success rate (columns 4 and 5) reflects the success rate for individuals included in a given row but not the prior row.

Table 6: **Pre-populated return characteristics for non-filers**

	No filing obligation		Has filing obligation	
	Mean	Median	Mean	Median
Age	55	59	47	46
Female	0.50	–	0.33	–
Filed a 2018 return	0.11	–	0.37	–
Pre-populated return with dependents	0.02	–	0.10	–
Calculated childless EITC >0	0.081	–	0.049	–
Calculated child EITC or CTC >0	0.007	–	0.096	–
...with revealed dependents	0.008	–	0.101	–
Calculated AGI	1,162	0	55,390	36,000
Calculated AGI (if potential refund)	3,873	2,900	39,895	28,100
Tax withholding	37	0	4,436	1,700
Tax withholding (if >0)	260	100	5,387	2,300
Has potential refund	0.17	–	0.44	–
Has potential taxes owed	0.00	–	0.55	–
Potential refund (if >0)	386	200	1,491	900
Potential taxes owed (if >0)	–	–	5,039	1,500
Count (millions)	46.3		8.4	

Notes: The table describes a 0.1% random sample of non-filers: individuals between the ages of 18 and 105 who are neither a primary nor secondary filer of a 2019 tax return and who have a 2019 information return with a U.S. address. We exclude individuals who filed jointly in 2018 if their spouse filed a 2019 return. Otherwise we construct tax units based on the 2018 filing, assuming single status if no 2018 return was filed. In the table, individuals are classified by whether there is an apparent filing obligation, i.e., whether simulated AGI is greater than the standard deduction for singles (conditional on age). All rows present results using our baseline specification except for the row showing the prevalence of potential EITC or CTC credits based on "revealed" dependents: those that are claimed by the individual either in 2018, 2020, or 2021 and not claimed by a 2019 filer. Sample weights are used; thus the results are nationally representative. Medians of dollar-denominated variables are rounded to the nearest $100.

A Additional details

In this appendix, we provide additional details on the assumptions we make in our upper and lower bound calculations. We also discuss additional results shown in the appendix tables and figures.

A.1 Further details on procedure

Throughout this paper, the baseline procedure that we consider is the following:

1. The IRS observes filing status, dependents, and tax unit individuals based on the prior year (2018, in our case) return. By "tax unit individuals", we mean the one or two individuals that are the primary and secondary filers.
2. The IRS prepares a current-year (2019, in our case) return for that tax unit based on the information returns received by the tax unit individuals in the current year. The IRS assumes that dependent status is unchanged from year to year, except for mechanical aging-out effects.[21]

A.1.1 Tax unit composition

In our empirical approach, we begin with the set of current-year tax units, as that is the sampling frame of the SOI file that we use throughout the paper. In the case of a tax unit where the 2019 tax unit individuals does not match the 2018 tax unit individuals – e.g., in the case of marriage, divorce, death, etc. – this creates an upper bound failure situation. In such cases, the IRS would have prepared a return that includes the wrong people. We do not attempt the lower bound direct computation in these cases. Rather, we code this as an immediate lower bound failure.

A.1.2 Non-filers

The universe of non-filers under consideration in our non-filer pre-populated return exercise includes the union of two sets:

1. "Single" non-filers: Individuals who (1) received any information return in 2019 with an address in the 50 states or the District of Columbia, (2) did not file in 2019 and (3) were either non-filers in 2018 or filed in 2018 using any status other than married filing jointly. If the individual was a non-filer in 2018, we assume that the IRS would prepare a single return for that individual in 2019. If the individual was a filer in 2018, the IRS would use the filing status on their 2018 return.
2. "Married" non-filers: These are married couples where (1) either member of the couple received any information return in 2019 with an address in the 50 states or the District of Columbia, (2) they filed jointly in 2018, and (3) neither member of the couple was a filer on any return in 2019. We assume that the IRS would prepare a married-filing-jointly return for these tax units.

We ignore the set of 2019 non-filers not included in either of these sets. One component of those excluded include individuals who do not receive information returns. Additionally, we exclude situations where a married couple filed jointly in 2018, but exactly one member of the couple filed a return in 2019. Among those that receive information returns, we estimate that there are approximately 400,000 such individuals that fit this second condition.

[21] See Section A.1.3 for further discussion of dependents.

A.1.3 The treatment of dependents

For the upper bound approach, we identify the set of individuals predicted to be a dependent of a given tax unit in 2019, as well as the set of individuals predicted to be an EITC dependent of a given tax unit in 2019. We refer to the former as an "apparent 2019 dependent" and the latter as an "apparent 2019 EITC dependent." Apparent 2019 dependents are those who were dependents in 2018 and are any age other than 23 in 2018. For those tax units that claimed the EITC in 2018, then apparent 2019 EITC dependents are dependents actually claimed in 2018 for EITC purposes. For those tax units that did not claim the EITC in 2018, then all 2018 dependents who are younger than 18 in 2019 are apparent 2019 EITC dependents.

A given tax unit will trigger the dependent mismatch failure situation if the merge between actual 2019 dependents and apparent 2019 dependents is less than perfect – i.e., if there is an actual 2019 dependent that is not an apparent 2019 dependent, or vice versa. A given tax unit will trigger the EITC dependent mismatch failure situation if (1) they claim the EITC in 2019 and (2) the merge between actual 2019 EITC dependents and apparent 2019 EITC dependents is less than perfect. Because it would require an amendment of the pre-populated return, we count it a failure situation if the identities of the dependents do not match, even if the total count of (EITC) dependents matches.

In contrast, for the lower bound approach, we calculate only the *number* of apparent 2019 dependents, as well as the number of dependents eligible for the CTC, and the number of dependents eligible for the EITC. The apparent 2019 dependent follows the same procedure as the upper bound approach. For tax credit eligibility purposes, we use a purely age-based threshold: a 2018 dependent is assumed to be a 2019 dependent unless he or she attains age 24 in 2019 and if an apparent dependent is under the age of 17 (18), we deem them to be eligible for the CTC (EITC).

A.1.4 Data problem: Form 1099-MISC

As we show in Figure A1, the number of Forms 1099-MISC in the database in 2019 is substantially lower than in 2018 and prior years; additionally, this apparent missing mass is fully explained by a missing mass of over 20 million paper-filed Forms 1099-MISC.[22] This data imperfection could, in principle, bias our success rates downwards. That is, there may be some taxpayers whose Schedule C income matches these missing 1099-MISCs perfectly and who have no other mismatches between their return and their simulated pre-populated return. In this case, we would have incorrectly coded such individuals as a failure instead of a success. However, this data imperfection likely has an immaterial effect on our results. Only 7.5% of taxpayers have a Schedule C mismatch and no other failure modes, per Table 2. Moreover, among those we identify to have 1099-MISC income, our success rates range from 2.6% to 5.5%. This likely reflects the fact that such taxpayers tend to deduct expenses, which are not observed to the IRS.

A.2 More detail on upper bound failure situations

In this section, we define each of the upper bound failure conditions that are not self-explanatory. The full set of failure situations is included in Table 2, which shows their likelihood of occurring. The most common failure situations are Schedule C mismatches (16 percent of taxpayers), followed by itemized

[22]It is possible that many of these missing forms may have been among the approximately 30 million information returns that the IRS destroyed due to pandemic-related operational prioritization (Treasury Inspector General for Tax Administration, 2021).

deductions (11 percent), wage mismatches (9 percent), taxable pension/IRA mismatches (7 percent), and Schedule E rents and royalty mismatches (7 percent). The least common failure situation is an early distribution from a Roth IRA (0.5 percent). Overall, just over half of taxpayers have any failure situations, and the majority of these taxpayers have only one.

Figure 2 shows how the prevalence of the most common six types of failure situations vary by income. Itemizing is not a prevalent source of failure until AGI reaches about $50,000, above which point the likelihood of itemizing rises rapidly. In contrast, the presence of Schedule C income is common at many income levels and is non-monotonic in income. Schedule C income is common for tax units earning $10,000-$20,000 and then falls in prevalence before rising again, reflecting the bunching of self-employed taxpayers near the first EITC kink (Saez, 2010; Mortenson and Whitten, 2020). Mismatches between reported wages and Form W-2 amounts are also non-monotonic, reaching a peak around $10,000 and a nadir around $100,000. Rents and royalties are less common than itemizing but otherwise follow a similar pattern; they are less common below $50,000 and their prevalence rises linearly with log income above $50,000.

Dependent mismatch and *EITC dependent mismatch*: see Section A.1.3, above.

Change in filing status: This failure situation is triggered in a number of cases. First, it is triggered when the identity of the tax unit (its primary and secondary filers, without regard to who is primary) is not constant from 2018 to 2019 – e.g., when the tax unit is a married couple in 2019 but two singles in 2018. Second, this failure situation is triggered when the 2018 filing status of the tax unit does not match the 2019 filing status (e.g., due to a change from single to Head of Household, or vice versa). An important exception is that a taxpayer who files as unmarried in 2019 and was a non-filer in 2018 does not trigger this failure situation.

Form 1099-R with taxable amount not determined: This situation is triggered when a taxpayer receives a Form 1099-R in 2019 (other than from an IRA) with box 2b checked, indicating that the taxpayer needs to perform additional calculations (not fully observable to the IRS) in order to calculate the taxable amount.

Certain above-the-line deductions: These refer to the deduction for moving expenses, self-employed health insurance, and educator expenses.

Section 199A deduction complications: Under section 199A, taxpayers are generally entitled to a deduction for 20% of qualified business income. However, there are several limits to this deduction that are not fully observable to the IRS, such as high-income taxpayers whose income is derived from a specified service trade or business (SSTB). This failure situation is triggered whenever a high-income individual's reported Section 199A deduction differs from our calculated value.[23]

Residential energy credits: This refers to the taxpayer claiming credits on line 5 of Form 1040, Schedule 3.

Early distribution from Roth IRA: This failure situation is triggered if the taxpayer receives a Form 1099-R with distribution code J in box 7. This is a failure mode because the taxable amount of such a distribution depends on the taxpayer's basis in the Roth IRA, which is not observed by the IRS.

Schedule C income does not match Form 1099-MISC NEC: This condition can arise for many reasons. First, the taxpayer may have small business income that is not included on a Form 1099-MISC. Second, the taxpayer may have expenses that are correctly deductible on Schedule C but of course are

[23] An additional limitation on the 199A deduction arises because the deduction is limited to 20% of ordinary taxable income, which is often binding for lower-income taxpayers. We do not consider mismatches arising from this limitation a 199A failure because any other failure situation resulting in the wrong ordinary taxable income would flow into this calculation.

not observed by the IRS. Third, the taxpayer may fail to report 1099-MISC income on Schedule C.

Itemized deductions in 2019: This is a failure situation because most itemized deductions – in particular, neither state/local taxes paid nor charitable contributions[24] – are not covered by information returns. Over 99.5% of itemizers claim at least one of these two deductions.

Wages do not match Form W-2 wages: This could occur for a variety of reasons; the following is a non-exhaustive list. First, the taxpayer may (intentionally or otherwise) misreport. Second, the Form W-2 itself could be erroneous – perhaps because we have identified the "wrong" W-2 in the case of duplicates. Third, the taxpayer could have unreported tip income. Fourth, the taxpayer could have received a disability pension distribution prior to their firm's retirement age. Fifth, the taxpayer could have received a taxable scholarship.

Taxable pension/IRA income does not match Form 1099-R: We assume that the IRS would pre-populate a return treating the entire distribution amount (with Box 7 codes 1, 2, or 7) as taxable. A mismatch between this amount and the taxpayer-reported amount could occur for a variety of reasons; the following is a non-exhaustive list. First, this failure condition is often an implication of the "taxable about not determined" failure mode – which often reflects the fact that some of the distribution represents basis recovery and thus not all of the distribution is taxable. Second, the distribution could comprise the distribution part of an indirect rollover (e.g., from a 401(k) to an IRA), which is correctly non-taxable. Third, the taxpayer may (intentionally or otherwise) misreport. Fourth, the Form 1099-R itself could be erroneous.

Lifetime Learning Credit: In general, post-secondary education expenses can give rise to one of two credits (but not both): the American Opportunity credit (AOC) or the lifetime learning credit (LLC). If a taxpayer is eligible for both, the AOC dominates the LLC. We assume that a pre-populated return would generate an AOC in the presence of education expenses. However, the AOC has additional restrictions relative to the LLC – most notably, the AOC can be taken only with respect to expenses for the first four years of post-secondary education. Additionally, a taxpayer may not claim the AOC if she has been convicted of a drug-related felony. These conditions are not directly observed by the IRS. Thus, if the taxpayer is observed to claim the LLC instead of the AOC (presumably due to one of the conditions mentioned above), this would cause a discrepancy between the pre-populated return and true liability. Thus, we treat it is as an upper-bound failure situation.

A.3 Additional Tables and Figures

The characteristics of the sample are shown in Table A1, weighted to represent the tax filing population of 152 million tax units. A little over a third of filers are married and nearly a third have dependents and claimed the CTC. Pre-populated returns may limit the need for using a paid preparer, which around half of taxpayers did in 2019. Median AGI in the sample is $41,200, median taxable income is $24,100, and median tax liability is $1,800. All three measures are skewed in the sample with much larger means than medians.

A.3.1 Success rates by age

Figure A2 shows how our upper and lower bound estimates for the share of pre-populated returns that are successful vary by age. The upper bound success rate starts at over 80 percent for 18-year-olds but

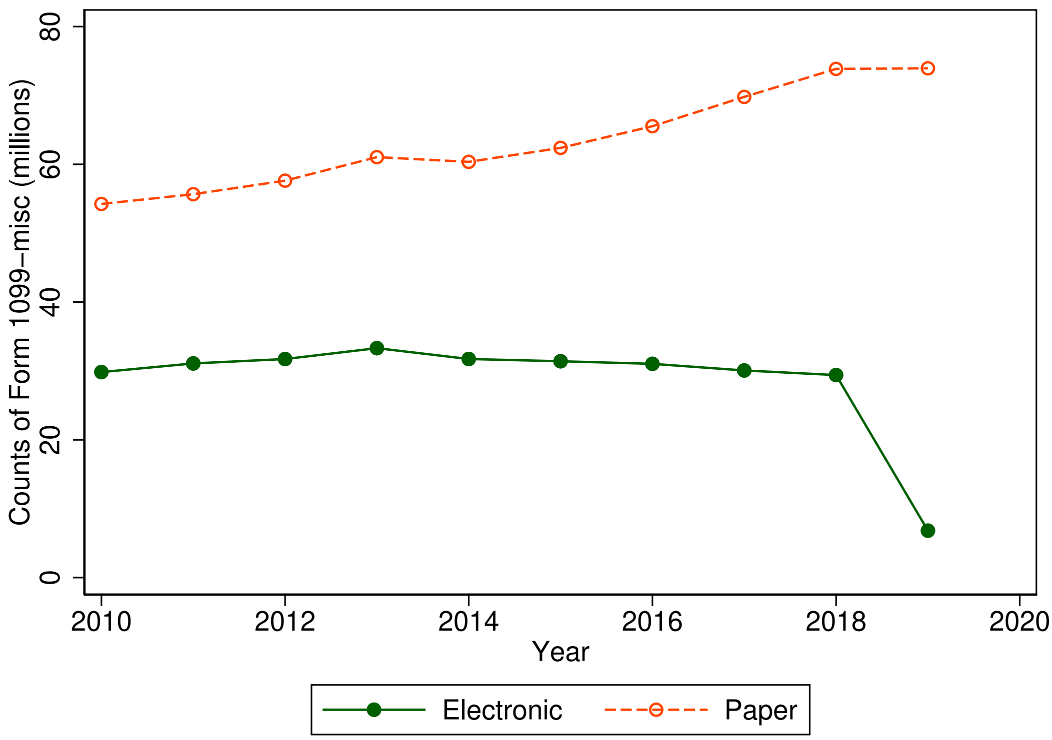
[24]The IRS observes state income tax withheld on certain information returns, but it does not observe property tax paid nor any amount of state tax paid directly.

falls steadily to around 40 percent for 36-year-olds. It stays relatively flat across the age distribution from there. The lower bound success rate starts around 90 percent for 18-year-olds and falls even more sharply than the upper bound, reaching around 30 percent for 36-year-olds. It stays low until about age 45, above which the estimate climbs steadily until plateauing at around 45 percent for ages 70 and above.

A.3.2 Over- vs. understated liability

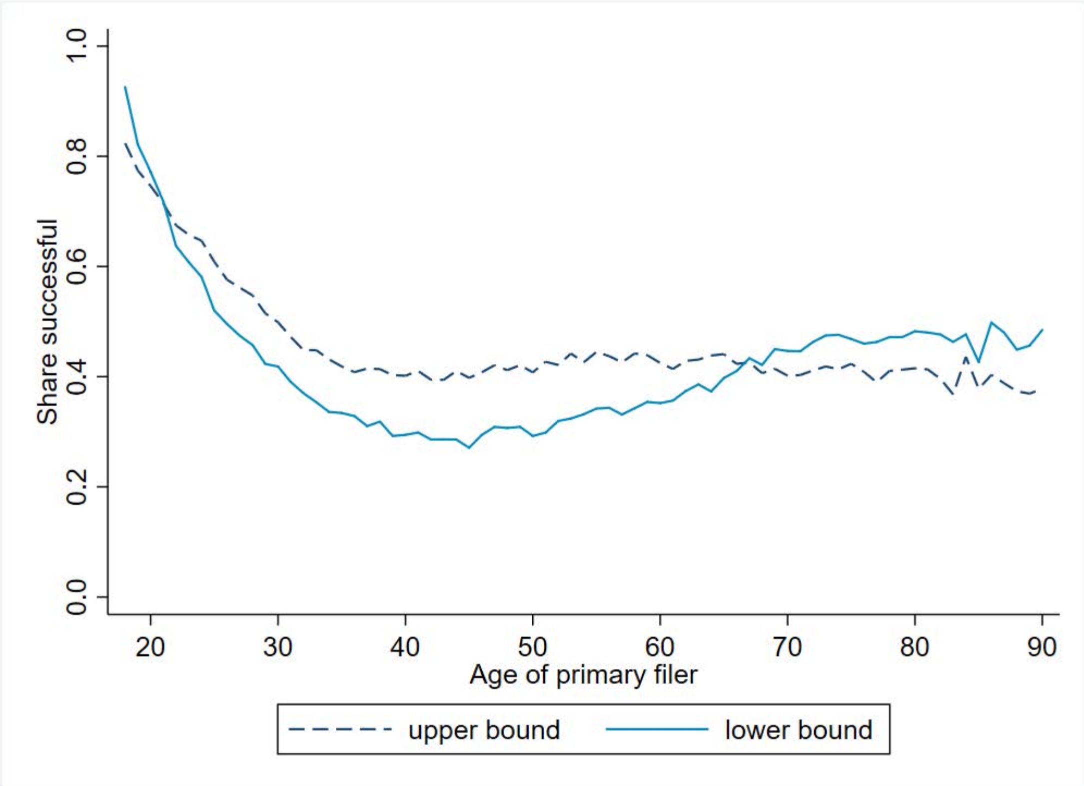
Figure 3 breaks out our lower bound estimates by whether the pre-populated return under- or overstates liability. In panel (a) we see that, in all income bins except for the top five percent, pre-populated returns are more likely to overstate liability than to understate it. Panel (b) further illustrates that for most income bins, the average magnitude of the error is fairly small: below $5,000. Panel (c) shows these error magnitudes relative to income. We see that, in the bottom income decile, liability is overestimated by a sizeable share of income: over twenty percent. But for most taxpayers, the average error in pre-populated liability is less than five percent of income.

Figure A1: Counts of Form 1099-MISC in the underlying database



Notes: The figure plots counts of Form 1099-MISC in the underlying database of administrative records from which we draw our data, separately by electronically-filed and paper-filed forms.

Figure A2: Pre-population success rates by age



Notes: The figure plots the share of tax year 2019 pre-populated returns that would be within our preferred tolerance levels for accuracy under our upper-bound and lower-bound approaches, by primary filer age. Sample weights are used; thus the results are representative of the 2019 tax filing population. Age is censored at 18 and 90.

Table A1: **Summary statistics**

	Mean	Median
Primary filer age	45	43
Married	0.37	–
Has dependents	0.32	–
Dependents	0.57	0
Uses paid preparer	0.51	–
Claims Earned Income Credit	0.17	–
Claims Child Tax Credit	0.30	–
Adjusted gross income	76,650	41,200
Taxable income	59,083	24,100
Tax liability	9,598	1,800
Count (millions)	152.0	

Notes: The table describes our sample of 2019 tax units from the Statistics of Income. Sample weights are used; thus the results are representative of the 2019 tax filing population. For dollar-denominated variables, medians are rounded to the nearest $100.

Table A2: **Pre-population success rates for subsets of taxpayers defined based on 2018 tax returns**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.17	0.72	0.77	0.72	0.77
2. Allow married	0.26	0.69	0.71	0.64	0.61
3. Allow dependents	0.35	0.62	0.68	0.39	0.57
4. Add interest/dividends	0.37	0.62	0.68	0.65	0.74
5. Add Social Security	0.42	0.63	0.69	0.73	0.75
6. Add pension/IRA distributions	0.44	0.63	0.68	0.58	0.58
7. Add gambling, UI, state tax refunds	0.47	0.63	0.68	0.54	0.66
8. Add capital gains	0.48	0.62	0.68	0.59	0.58
9. Add high income	0.53	0.61	0.67	0.44	0.58
10. All income types	0.69	0.53	0.58	0.28	0.26
11. **Broadest**: Eliminate deduction and credit restrictions	1.00	0.42	0.48	0.19	0.27

Notes: This table reports the rates at which we are able to accurately pre-populate returns for certain subsets of 2019 tax filers. We use our preferred tolerances in Table 1. Subsets are defined based on characteristics from the 2018 tax return. "Unobserved deductions" include all itemized deductions and above-the-line deductions for moving expenses, educator expenses, and health insurance for self-employed persons. "Unobserved credits" are all credits other than the EITC, Child Tax Credit, and American Opportunity Tax Credit.