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 June 2023

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. This research was conducted while Goodman and Whitten were employees at the U.S. Department of the Treasury, and the research was supported in part by the U.S. Department of Agriculture, Economic Research Service. 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 and should not be construed to represent any official USDA or U.S. Government determination or policy, nor do they necessarily represent those of the National Bureau of Economic Research. 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.

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 June 2023
JEL No. H0,H2,H24,I3

ABSTRACT

Each year Americans spend over 1.7 billion hours and $33 billion preparing individual tax returns, and these filing costs are regressive. To lower and redistribute the filing burden, researchers and policymakers have proposed having the IRS prepopulate tax returns for individuals. We evaluate this hypothetical policy using a large, nationally representative sample of returns filed for tax year 2019. Our baseline results indicate that between 66 and 75 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 $8.2 billion in refunds due to 11 million (20 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. The Internal Revenue Service (IRS) estimates that the average filer spends $210 and 11 hours to file their 2019 Form 1040 (Internal Revenue Service, 2020). This adds up to 1.7 billion hours and $33 billion in tax preparation fees, software costs, and filing fees. This burden is not distributed in proportion to income. On average, for tax year 2010, a taxpayer earning between $10,000 and $15,000 spent about 10.3 hours and $114 to file their return, while one earning between $100,000 and $200,000 spent about 14.5 hours and $328 (Marcuss et al., 2013).

Some researchers and policymakers, for example Goolsbee (2006) and Warren (2022), have proposed that the IRS pre-populate Form 1040 for some or all individuals. 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 additionally encourage 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 "item-based approach" because we define success based on the absence of specific failure situations. [1]

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 the "overall tax-liability approach", or "tax-liability approach" for short. The success rate under the overall tax-liability approach will tend to be understated if the tax authority can use a richer tax calculator.

Using our baseline tolerance thresholds, our overall tax liability and item-based approaches are successful for 66 and 75 million returns (42 and 48 percent of filers), respectively. In about two-thirds of the cases where the tax-liability approach is inaccurate, the pre-populated liability is higher than the reported liability, with a (rounded) median gap of $1,700. For both approaches, we find that success rates are decreasing in income, from around 60-80 percent in the bottom income decile to around 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 successfully complete their return. Among the 52 percent of taxpayers who fail the item-based approach,

[1] We use the words success (failure) interchangeably with match (mismatch).

the majority have only one failure situation. Such taxpayers would typically 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) in order to correct their pre-populated return.

Next, we consider refinements of the pre-population procedure in which the IRS pre-populates returns for 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. We show in supplemental analysis of a stratified sample of randomly audited tax returns that some, though not all, of this mismatch can be explained by non-compliance, whether intentional or unintentional.

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 47.1 million adults who received an information return, did not file a tax return, and did not appear to have a filing obligation based on income on their information returns.[3] Using our tax-liability approach, our pre-populated returns indicate that 16 percent (7.5 million) of these individuals are entitled to a refund, with an average potential refund of $390. We additionally identify 7.8 million non-filers who appear to have a filing obligation based on their information returns. Among this population, 54 percent (4.2 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.

Finally, we consider the welfare effects of a pre-population regime. In particular, we estimate the marginal value of public funds (MVPF) associated with the policy, which represents the ratio of the dollar-valued net benefits of the policy scaled by the fiscal costs of the policy to the government (Hendren and Sprung-Keyser, 2020). In general, a MVPF greater (less) than one implies that a given policy is more (less) efficient than a non-distortive lump sum transfer. In our context, the net benefits include the net reduction of individual compliance costs and any net reduction in income tax payments resulting from the policy, while the fiscal costs include the net reduction in tax payments plus the government's administrative costs. Each element in this formula requires assumptions about behavioral or technological parameters, and neither the data nor the existing literature provide much guidance in selecting those parameters. We provide estimates for a broad range of plausible parameters, showing they could lead to an MVPF anywhere between 0.4 and 18. While we are unable to precisely identify the MVPF for pre-population, our formula may help to illuminate the relative roles of the behavioral and technical parameters that affect the potential impact of pre-population on social welfare.

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.

[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.

Moreover, our data allow 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. The IRS would need the resources and time to process information returns and generate a pre-populated tax return, and taxpayers would need time to review the tax return, make any necessary changes, and send it back to the IRS. 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 mostly 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. The latest OECD tax administration data show that 33 of 38 OECD countries use third-party information to fully or partially pre-fill tax returns. While a few countries only pre-fill personal information or income types, 28 countries pre-fill amounts for at least some forms of income.[4]

The U.K. uses an exact-withholding system in which the majority of taxpayers do not need to file and the remaining taxpayers file a self-assessment in cases where there are income or deductions not known to His Majesty's Revenue and Customs. In 2022, 10.2 million tax units filed a tax assessment on a base of 30 million tax units.[5] In Germany, most taxpayers are required to file, direct e-filing is available, and pre-populated returns with e-filing have been available since 2014. Blaufus, Hechtner and Jarzembski (2019) find that compliance costs (time costs plus financial costs) dropped 33-38 percent from 2008 to 2016 though much of this decrease is attributable to changes in e-filing availability, a larger standard deduction, and exact withholding of capital gains tax; pre-population of returns was only introduced in 2014 and was still a modest percentage of tax returns at the time of the study. Spain allows taxpayers to request tax information from the tax authority, which provides a pre-filled return for the majority of taxpayers. For taxpayers with more complex tax situations, the authority provides available income information based on third-party information returns for taxpayers. This includes, for example, those with income sources from personal businesses or property rentals. Nearly all taxpayers receive one of these, with over 70 percent of taxpayers receiving a pre-filled return. Around half of these taxpayers accepted their pre-filled return with no alteration (Vaillancourt, 2011).

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 pilot, the ReadyReturn program, ran between 2003 and 2005. Each year roughly 50,000 people were mailed a pre-populated

[4] Data from the OECD Tax Administration Services database at `https://data.rafit.org/regular.aspx?key=63544710`

[5] See `https://www.gov.uk/government/news/more-than-102-million-filed-their-self-assessment-by-31-january` and `https://www.gov.uk/government/statistics/number-of-individual-income-taxpayers-by-marginal-rate-gender-and-age`.

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. The Colorado pilot, known as File-4-Me, was in effect from 1999 through 2002; at its peak, it prepared fewer than 15,000 returns per year (U.S. Department of the Treasury, 2003).

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.

Existing empirical work in the U.S. context uses 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 it to be 40% of taxpayers. 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).[6] 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.[7] 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

[6] 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.

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

random sample of 358,800 individual income tax returns for 2019 constructed by the Statistics of Income (SOI) Division of the IRS.[8] These data include information from nearly every line or box on Form 1040 and its schedules. The sampling rate increases with income, and SOI samples the highest-income taxpayers with certainty. In all specifications, we use sampling weights to make the sample representative of the full tax-filing population (157 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.[9]

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[10] and we use the 2018 Form 1040 to identify their 2018 dependents, filing status, and the identity of the 2018 primary and secondary filer.[11]

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." We refer to this method as the"item-based approach." The failure situations fall into three sets. The first set includes those situations where the tax unit composition 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.[12] 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.[13]

Second, we provide a more direct test: we compare the taxpayer's self-reported tax liability to their simulated pre-populated return liability calculated using NBER TAXSIM (Feenberg and Coutts, 1993).[14] 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 fea-

[8]The universe covers all those who filed a 2019 return in calendar year 2020 or 2021. It excludes those who filed a 2019 return in 2022 or later. As of June 8, 2023, we estimate that approximately 830,000 2019 tax returns were filed too late to make it into our sampling window. These returns are represented neither in our filer nor our non-filer samples.

[9]We deem a pre-populated return "successful" if and only if it matches the actually-filed return. This terminology is a simplification. Even an "unsuccessful" pre-populated return that requires adjustment may nonetheless benefit the taxpayer if it saves them time when filing. We return to this point below when discussing Table 2.

[10]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.

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

[12]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.

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

[14]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.

tures.[15] 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 upon its input data that simplify, and thus imperfectly reflect, the tax code. Therefore a richer calculator may accurately pre-populate additional returns. We refer to this exercise as the "tax-liability approach" to measuring the success rate of pre-population.

Overall the item-based approach predicts a higher success rate than does the tax-liability approach. Intuitively, this is because it is not possible for the item-based approach to fully enumerate every possible failure situation. However there are some subsets of taxpayers for whom the tax-liability approach predicts success despite the item-based approach finding one or more failures. 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. Additionally, a given return could have multiple failure situations with perfectly offsetting effects on liability.

III Results

Using our preferred tolerance levels, our baseline results are that 42 and 48 percent of pre-populated returns are successful under our tax-liability and item-based approaches, respectively, as shown in Table 1. For the tax liability 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 item-based 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 filing status errors, which improves the success rate of pre-population by around five percentage points (about 8 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 of our 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 item-based approach predicts fewer successes than the tax-liability approach 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.

[15] 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/`.

Next we examine the prevalence of the 22 failure situations that underlie our item-based approach in column 1 of Table 2. 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 17 percent of taxpayers. Itemized deductions are the next most common failure situation, affecting around 11 percent of taxpayers. Correcting these failures requires the taxpayer to file an additional schedule (C or A, respectively). In contrast, the third most common failure situation (affecting nine percent of taxpayers) is a mismatch between wages reported on Form 1040 and those appearing on Forms W-2. Correcting this failure may merely require editing one line of the pre-populated return.[16] In column 2, we see that five percent of taxpayers have a wage mismatch and *only* a wage mismatch. Overall, among the 52% of taxpayers with any failure situation, 57% ($\frac{.298}{.522}$) have exactly one failure situation. Though their pre-populated returns would be imperfect, these taxpayers may nonetheless benefit from a pre-population regime that accurately completes a substantial portion of their tax returns.

In Figure 2, we investigate the relationship between success rates and income. In particular, we plot the share of tax units with several common item-based 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 associated with EITC returns. The bunching of Schedule C filers in the EITC range is well documented starting with Saez (2010). The income patterns of W-2 wage mismatches relating to EITC income eligible ranges are consistent with those found in Mortenson and Whitten (2020). 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. 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-population is successful have lower average income. In the item-based approach, the successful group have average incomes of \$47,100 compared with \$103,100 in the unsuccessful group.

For the unsuccessful group under the tax-liability 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 have particularly high incomes, with an average of \$143,400. As we discuss in further detail in Section III.D, the extent to which taxpayers make adjustments to their pre-populated tax liability will determine the overall effects of a pre-population regime on tax revenue and the effective progressivity of the tax system.

The primary motivation for pre-populated tax returns is that they may reduce and redistribute the costs of tax filing. Table 3 hints at this in columns (1) and (3), as we see that 43 to 44 percent of taxpayers for whom pre-populated returns are accurate currently use paid preparers. Individuals may reduce their reliance on paid preparers if they decide to use the pre-populated return. 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.

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

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 mismatch rates for pre-populating specific lines of the 2019 Form 1040. Following our item-based 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 do not match the actually-filed returns. This represents the share of individuals who would need to make a correction to that line. 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 pre-populated wage line does not match the actually-filed return for 8% of taxpayers overall or 10% of tax returns with some wage income. Conditional mismatch rates are similar for taxable interest and qualified dividends, but the unconditional mismatch rates are lower due to the presence of many zeros. Conditional mismatch rates are somewhat higher for pensions (19%) and are much higher for capital gains (56%) and Schedule 1 income (76%), 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 differing mismatch 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.[17] Around one-fifth of 2019 filers satisfy these restrictions, and of these, we find that between 79 and 81 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 item-based approach, is a mismatch between Form W-2 and taxpayer-reported wages, which occurs in 69 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 73 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 item-based failure situation, 65 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,

[17] 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.

whose success rates are somewhat lower. In row 10, we allow all income types, increasing coverage by another 17 percentage points. However, as expected, the marginal success rates for this group are quite low: 23 and 19 percent, respectively, under our tax liability and item-based 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.[18]

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 implementable procedure would define taxpayer subsets based on the information 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 (item-based) and nine (tax-liability based) 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 item-based gap remains relatively stable through row 8 (adding capital gains), while the nine percentage point tax-liability based gap reduces to five to six 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 The Role of Non-compliance

In our analysis, we treat the filed return as the "truth". For example, if there is a mismatch between what a taxpayer reports on the wage line (Line 1) of Form 1040 and what is included on their Form W-2 – a situation that will immediately lead to an item-based failure and will likely lead to a tax-liability based failure – we assume that this mismatch reflects unique circumstances of the taxpayer that we do not observe. For instance, the taxpayer might have scholarship income or unreported tip income, which would not be reported on Form W-2 but would be included in Line 1 on Form 1040. Or the taxpayer may have corrected an error made by an employer on their Form W-2.

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 – there is 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). Additionally, taxpayers may make inadvertent mistakes, such as copying information incorrectly from their Form W-2 or including income on the wrong line. Therefore, some of the pre-populated returns that our analysis deems to be "failures" might actually measure true tax liability *more* accurately than the taxpayer-filed

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

return.

To quantify the role of non-compliance (intentional or unintentional) on our measured success rates, we would ideally compare our simulated pre-populated returns to taxpayer-filed returns that have been verified or corrected by hypothetical perfect audits. This is not feasible, both because (1) auditors in fact detect only a subset of non-compliance (Internal Revenue Service, 2019; Guyton et al., 2021) and, more practically, because (2) the tax returns in the SOI cross-sections – which serve as the source of the taxpayer-filed returns in this article – are not systematically or randomly subject to audit. While we cannot undertake this direct approach, we can nevertheless use a different set of randomly audited taxpayers to provide context for our results – though we remain subject to the limitation that auditors are not omniscient. In particular, we make use of data from the National Research Program (NRP), which is an IRS program that audits a stratified sample of taxpayers in certain years for the purpose of measuring existing compliance patterns. These data allow us to observe taxpayer-reported and audit-corrected amounts on many line items on Form 1040. We use data from tax year 2014, which is the most recent year available to us.

In Table 6, we report the share of taxpayers in the 2014 NRP that have an audit correction in excess of $500 (the same threshold used in Table 4) on each of several income lines. In column 1, we report the unconditional correction rate; in column 2, we restrict to those who report a non-zero amount on the line item in question. We find that 4.5% of all taxpayers had an audit correction on the wage line in 2014. Audit correction rates were much higher for income that was included on the 2019 version of Schedule 1, which includes (among other things) sole proprietor income and rental real estate income – consistent with existing research on the components of the tax gap (Internal Revenue Service, 2019). Interest and dividends rarely received audit corrections, while capital gains and retirement income were more likely to receive corrections.

We can compare these audit correction rates to the mismatch rates of Table 4. For wages, interest, dividends, and retirement income – all of which are subject to substantial information reporting – the mismatch rate is between 180 percent (wages) and 500 percent (interest) of the audit correction rate. On the one hand, this analysis suggests that most mismatches on many lines covered by information returns reflect legitimate differences between an information return entry and the corresponding tax return concept. On the other hand, this analysis also suggests that non-compliance is an important driver for the mismatches that we observe on the wage line, as the ratio of audit corrections to information return mismatches is largest on this line. This particular mismatch is one of the most consequential in implementing pre-population. As indicated in Table 2, the wage mismatch represents the unique failure situation for approximately 10% of item-based failures – and 61% if we restrict to the narrowest subset in Table 5, who might be the first eligible for a pre-population regime. Therefore, it is important to understand this particular mismatch.

To delve further into the sources for wage mismatches, we examine the prevalence of audit corrections among 2014 NRP taxpayers with wage mismatches between their Form W-2s and Form 1040s. Table 7 shows a cross-tabulation of the tax units in the 2014 NRP. Across the rows, we distinguish taxpayers by whether their W-2 income matches their Form 1040, Line 1 income (using the same method and $500 tolerance as the analysis underlying Table 4). Across the two columns, we separate taxpayers based on whether the tax unit had an audit correction on their Form 1040, Line 1. In this table, we drop tax units who have zero W-2 income and zero reported and audit-corrected wage income.

We find that the probability of receiving an audit correction is very low when Form W-2s and Form 1040 match. In the second row, we see that about 54 percent ($\frac{0.050}{0.050+0.043}$) of taxpayers with a W-

2/1040 mismatch receive an audit correction on the wage line of the Form 1040. Furthermore, the vast majority (approximately 87 percent) of such audit corrections simply replace the taxpayer-provided amount with the Form W-2 amount. These results suggest that our success rates with respect to true liability may be meaningfully underestimated, especially in the narrowest subset of taxpayers in Table 5. As a back-of-the-envelope calculation, if $0.54 \times 0.87 \approx 47$ percent of our wage mismatch errors were actually successes – and if this occurs uniformly throughout the population of those with a wage mismatch – then our item-based success rate in Table 5 in the first row would have been 86 percent instead of 81 percent.

III.C Non-filers

One potential benefit of pre-populated returns is that they 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.

We examine results for a subset of non-filers in Table 8. The table is based on a 0.1% random sample of the 54.8 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.[19]

We separate our analysis by those who do and do not appear to have a legal requirement to file a tax return.[20] Our sample, appropriately weighted, indicates that 47.1 million non-filers who receive information returns appear not to be required to file. On average, these non-filers had only $36 in taxes withheld. Sixteen percent (7.5 million) have a potential refund, and conditional on having a refund the average amount is $390. This finding is consistent with Hauck and Wallossek (2022), who document that, in Germany, many taxpayers without a filing obligation do not file even though a substantial portion are owed refunds. Among the 33 percent of German non-filers owed a refund, the average amount was €360. Other work suggests that time and out-of-pocket costs may exceed the benefits of filing for some taxpayers (Erard and Ho, 2001). Additionally, information frictions may increase the costs of filing. A pre-populated return could help reduce some of these barriers to claiming a refund.

In our sample of non-filers who appear not to have a filing obligation, one reason why potential refunds are not higher is that the vast majority (around 98 percent) would receive a pre-populated return without any dependents. Around eight percent would 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 among this population (Goldin et al., 2021). However, for the vast majority of non-filers, 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 0.7 percent of these non-filers (347,000 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

[19]We collected data on non-filers on June 8, 2023. We misclassify anyone who files after our data retrieval as a "non-filer". Information on non-filer age comes from Social Security Administration data. See Appendix A.1.2 for more detail on how we assign filing status. 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.

[20]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).

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 Table 8 row 7, we make use of this information by assigning dependents claimed via either of these non-filer tools to 2019 non-filers.[21] 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.

We also estimate that 7.8 million non-filers appear to be failing to meet a legal filing obligation. For the majority (54 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 \$3,982, with a (rounded) median of \$1,400. 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 file a late return after our data retrieval.

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).

III.D Welfare impacts

If the U.S. were to adopt a pre-population regime, its welfare impacts would depend on a number of factors, including the reduction in individuals' compliance costs, the increase in administrative cost to the tax authority, and the behavioral responses of taxpayers. Our analysis of pre-populated returns thus far has been mechanical, abstracting from behavioral responses and welfare considerations. Here we engage in a more speculative analysis of welfare effects.

To study the welfare effects of a pre-populated return program, we employ the marginal value of public funds (MVPF) approach of Hendren and Sprung-Keyser (2020). The MVPF is a parameter that measures the change in welfare resulting from a given policy change per dollar of government expenditure. It is equal to the net benefits accruing to policy recipients (more precisely, the aggregate amount that policy recipients would be willing to pay for the policy) divided by the net fiscal costs of the policy. Values above (below) one indicate that one dollar of government expenditure produces more (less) than one dollar of welfare gains.

In our context, we model the aggregate net benefits of pre-population as the sum of the net reduction in compliance costs ($-\Delta C$) and the net decrease in tax payments ($-\Delta T$) resulting from the adoption of pre-populated returns. We model the net fiscal costs of the policy as the sum of the aggregate increase in net administrative costs (ΔA) and the net decrease in tax payments. The MVPF is thus given by:

$$MVPF = \frac{-\Delta C - \Delta T}{\Delta A - \Delta T}. \quad (1)$$

To model the compliance cost reduction, we consider the change in aggregate compliance costs

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

separately for filers and non-filers: $\Delta C = \Delta C^{fil} + \Delta C^{non}$. For current filers, we model the reduction in compliance costs as a fraction of estimated compliance costs under current law: $\Delta C^{fil} = -\psi C^{fil}$, where C^{fil} represents baseline compliance costs, and $\psi \in (0, 1)$ represents the fraction of these costs reduced by pre-population. The true value of ψ is highly uncertain and depends on implementation details including the take-up rate of using the pre-populated return. We do not have reliable estimates of the reduction in compliance costs.Goolsbee (2006) suggested that compliance savings could reach \$6.3 billion per year (in 2019\$) after rolling the program out to 40 percent of the simplest returns, while a paper studying the Slovenian experience found that compliance costs fell over 70 percent although there were other tax policy changes occurring simultaneously (Klun, 2009).

We present results for a range of plausible values of ψ, the share of compliance costs reduced by pre-population: 0.05, 0.10, and 0.20. Baseline compliance costs for current filers, C^{fil}, are taken to be \$65.4 billion, which we derive from IRS estimates, given in the 2019 Form 1040 instructions, of average filing costs of \$210 and 11 hours.[22] In contrast, we expect compliance costs for current non-filers to rise, as some will be nudged into filing returns. We assume non-filers' compliance costs rise by $\psi(1-\psi)$\$149 on average, where \$149 is the minimum amount of average compliance costs (including monetized time costs) across income bins in Marcuss et al.'s (2013) study of filers, adjusted for inflation to 2019 dollars using the consumer price index. We multiply by $(1-\psi)$ to account for the general reduction in compliance costs for all filers relative to current law, and we multiply by ψ because we assume that more non-filers are nudged into filing when the compliance savings afforded by pre-populated returns (for filers) is greater.[23] Given our earlier estimate of 54.8 million non-filers receiving pre-populated returns, the aggregate increase in compliance costs for current non-filers is $\Delta C^{non} = 0.0548 \cdot \psi(1-\psi)$\$149= $\psi(1-\psi)$\$8.2 (measured in billions).

Next, we consider how tax payments may change under pre-population. In principle, ΔT could be equal to zero, as statutory liability is unchanged. However, the pre-populated return could function as a "nudge", encouraging some taxpayers to simply accept the pre-populated return, or make incomplete adjustments to it, causing payments to rise or fall. This could 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.

The evidence on overall tax liability changes after introducing pre-filled returns is sparse. In Finland, 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 (Kotakorpi and Laamanen, 2017). The authors found no appreciable average liability change, but the total effect on revenue remains unclear. Evidence from an online experiment in the U.K. suggests that, relative to receiving a blank return, taxpayers are more likely to overpay when liability is overestimated and to pay less when liability is underestimated (Fonseca and Grimshaw, 2017). 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.,

[22] We arrive at \$65.4 billion by summing the monetary costs, which amount to \$31.9 billion, and monetizing the 1.7 billion hours for the population at \$20 per hour. In 2019, average gross hourly earnings in the U.S. were around \$28 (U.S. Bureau of Labor Statistics, 2023). We use \$20 as a rough estimate of the average after-tax wage rate.

[23] Our functional form assumes that if pre-population saves 10% of filing costs for filers, then it will nudge 10% of the non-filer population into filing. If it saves 20% of filing costs, 20% of non-filers will start filing. This is simply a crude way of modeling our assumption that more useful pre-populated returns (in the sense of reducing compliance costs of filing) will lead to more take-up of filing among non-filers.

2015; Fochmann, Müller and Overesch, 2018; van Dijk et al., 2020).

We use two parameters to model the tax liability response, reflecting the extent to which taxpayers make positive and negative adjustments to their pre-populated tax liability. The change in tax liability ΔT is equal to the sum of the change in liabilities Δt_i for each taxpayer i. Let t_i^{pre} denote calculated liability on pre-populated returns and let t_i^{fil} denote tax payments under the current system.[24] The change in tax payments for a given individual is: $\Delta t_i = \phi_i \cdot (t_i^{pre} - t_i^{fil})$, where $\phi_i \in [0,1]$ takes on one of two values (ϕ_+ or ϕ_-) depending on whether the difference between liabilities on the pre-populated and actually-filed return, $t_i^{pre} - t_i^{fil}$, is positive or negative.[25] We use I_+ and I_- to denote the sets of individuals for whom this difference is positive and negative, respectively. In the extreme scenario where taxpayers merely accept their pre-populated returns, $\phi_+ = \phi_- = 1$. If instead taxpayers make all adjustments to arrive at the returns they file absent pre-population, $\phi_+ = \phi_- = 0$. In another extreme scenario, if taxpayers only make favorable changes to pre-populated returns that reduce tax payments, we might have $\phi_+ = 0$ and $\phi_- = 1$. In general, if taxpayers are more likely to make favorable changes, we have $\phi_- > \phi_+$. Also, to the extent that some fraction of taxpayers ignore their pre-populated returns, the values of ϕ_+ and ϕ_- will be lower. We present results for several possible pairs of (ϕ_+, ϕ_-): (0,0), (0.1,0.1), (0.1,0.2), and (0.1,0.3).

Finally, we turn our attention to the one remaining term in Equation (1): the net increase in administrative costs (ΔA). There are no up-to-date estimates of this term, and previous estimates varied widely (U.S. Department of the Treasury, 2003). Perhaps most relevantly, Vaillancourt (2011) estimates that the ongoing administrative cost of California's ReadyReturn was \$2.40 per taxpayer for tax year 2008. Scaling this value by the number of 2019 filers and non-filers and adjusting for inflation equals around \$607 million (2019\$). However, the ReadyReturn program differed significantly from the broader pre-population regime we model. Individuals opted in, and it was only available to a narrow subset of taxpayers. Computing pre-filled returns for all situations would be more costly, although per-capita administrative costs of a national program may benefit from larger scale. Importantly, Vaillancourt (2011) suggests that the individual reduction in compliance costs likely exceeded the increase in administrative costs.[26]

We model new administrative costs as a fraction of current aggregate taxpayer compliance costs: $\Delta A = \alpha C^{fil}$, with $\alpha \in (0,1)$. Intuitively, α captures the extent to which the centralized tax authority can complete pre-populated returns more efficiently than households.[27] Given the uncertainty surrounding this parameter, we present results for a range of plausible values: 0.01, 0.05, and 0.10. The smallest of these values (0.01) implies an annual cost of \$654 million, close to the naively extrapolated cost from the ReadyReturn program experience mentioned above. The largest of these values (0.10) implies that the IRS budget would have needed to increase by roughly 50 percent to pre-populate returns for tax year 2019.[28]

Expanding equation (1), our expression for the MVPF in billions of 2019\$ becomes:

[24] For non-filers, t_i^{fil} denotes observed withholdings.

[25] Since we use ϕ_+ and ϕ_- to describe both the behavior of current filers and current non-filers, they should be thought of as (weighted) averages of underlying parameters governing current filer and current non-filer behavior including any behavioral response in filing behavior.

[26] In our framework described below, this would imply that $\alpha < \psi$.

[27] The parameter α includes both the fixed and variable costs of implementing a pre-populated return regime. Implicitly, we have in mind the long-run costs, which are most relevant to the overall welfare impact of pre-population. There may be additional up-front costs in the set-up and early years of the program.

[28] In fiscal year 2020, when IRS would have hypothetically pre-populated 2019 returns, its budget was \$12.3 billion (Internal Revenue Service, 2022), and $0.10 \cdot \$65.4$ billion $= \$6.5$ billion.

$$
\begin{aligned}
MVPF &= \frac{-\Delta C - \Delta T}{\Delta A - \Delta T} \\
&= \frac{-\Delta C^{fil} - \Delta C^{non} - \sum_i \Delta t_i}{\Delta A - \sum_i \Delta t_i.} \\
&= \frac{\psi C^{fil} - \Delta C^{non} - \phi_+ \sum_{i \in I_+} \left[(t_i^{pre} - t_i^{fil}) \right] - \phi_- \sum_{i \in I_-} \left[(t_i^{pre} - t_i^{fil}) \right]}{\alpha C^{fil} - \phi_+ \sum_{i \in I_+} \left[(t_i^{pre} - t_i^{fil}) \right] - \phi_- \sum_{i \in I_-} \left[(t_i^{pre} - t_i^{fil}) \right]} \\
&= \frac{\psi \$65.4 - \psi(1-\psi)\$8.2 - \phi_+ \$255.7 + \phi_- \$271.3}{\alpha \$65.4 - \phi_+ \$255.7 + \phi_- \$271.3}, \qquad (2)
\end{aligned}
$$

where the amounts \$255.7 billion and \$271.3 billion come from our calculations comparing pre-populated and actually-filed liabilities.[29]

One benefit of our MVPF specification is that the reader can easily calculate it using their preferred values of α, ψ, ϕ_+, and ϕ_-. We explore several possibilities in Table 9, showing that the MVPF varies widely, ranging from 0.44 to around 18, depending on parameter values. Panel A covers cases in which the IRS is very efficient, implementing pre-populated returns at an administrative cost equal to one percent of the compliance burden currently borne by individual taxpayers. In these scenarios the MVPF is always above one and can be as high as 18 if taxpayers adjust to pay the same liability as they currently do ($\phi_+ = \phi_- = 0$) and the individual compliance burden declines by 20 percent ($\psi = 0.2$). Holding the implementation costs α constant, the compliance burden reduction to the taxpayer (ψ) has a large impact on the estimated MVPF. For example, in the top row of Panel A, increasing the burden reduction for taxpayers from one-twentieth to one-fifth moves the MVPF from 4.4 to 18, assuming no "nudge" effect on tax payments ($(\phi_+, \phi_) = (0, 0)$).

Comparing Panels A, B, and C illustrates the importance of implementation costs (α) on the MVPF. When compliance savings are smaller than administrative costs($\psi \leq \alpha$), MVPFs generally fall below one indicating that costs exceed benefits. Such a policy may nonetheless be desirable for distributional or other reasons outside of the scope of the MVPF calculation.

Finally, we see that symmetric changes in how taxpayers' liability adjustment (e.g. moving from $(\phi_+, \phi_) = (0, 0)$ to $(0.1, 0.1)$) have relatively little effect on the MVPF. This is because these nudge effects add (or subtract) an equal amount from the numerator and the denominator, reflecting the assumption that changes in tax liability are a pure transfer. When adjustments are symmetric the change in tax liability is small relative to other items in the formula because the mean pre-population error is relatively close to zero.[30] However, when we assume taxpayers make larger favorable adjustments than unfavorable adjustments ($\phi_- > \phi_+$) in rows 3 and 4 of each panel, this has the effect of increasing both the numerator and the denominator by the same meaningful amounts, bringing the MVPF closer to one. Intuitively, if $\phi_- > \phi_+$, the pre-population regime functions partially as tax cut which we assume is non-distortionary and therefore mechanically has an MVPF of one.

[29] For current non-filers we use withholdings. Non-filers whose pre-populated liability exceeds withholdings contribute \$16.8 billion to the \$255.7 figure. Non-filers whose pre-populated liability does not exceed withholdings contribute \$8.2 billion to the \$271.3 figure.

[30] In Panel A, the other item in the denominator (implementation cost) is very small, so even the small mean pre-population error is enough to change the MVPF non-trivially.

This discussion focuses on the MVPF for the full population of taxpayers, but any program that targets a subset of taxpayers may have very different MVPF values. Our population-wide assumed parameters may all differ by taxpayer characteristics, and the underlying compliance costs and differences in tax liability between filed returns and pre-populated returns may also differ systematically with taxpayer characteristics. However, it is not straightforward to predict whether the MVPF would be higher or lower in certain groups. For instance, if pre-population is restricted to those with the simplest tax situations, administrative costs would be smaller (increasing the MVPF), but the reduction in private compliance costs would also be smaller (reducing the MVPF). We leave further exploration of heterogeneity of these parameters to future research.

We further caution that our MVPF calculation offers an incomplete analysis of the welfare implications of pre-population. Some important omitted factors are (i) whether taxpayers will change labor supply, saving, or other real economic behavior in response to any changes in their tax payments, (ii) increased compliance costs associated with any acceleration of third-party information-return reporting to IRS, (iii) the reduction in the producer surplus of tax preparers, and (iv) issues concerning taxpayer morale, or trust in government more generally.

Additionally, the MVPF ignores distributional consequences (Finkelstein and Hendren, 2020). In the context of pre-populated tax returns, there are two broad distributional effects for policymakers to evaluate. First and most directly, there is a shift in the burden of preparing returns. The costs of administering pre-populated returns, like other government spending, will be borne broadly by taxpayers in proportion to tax liability. Because current filing costs are regressive (Marcuss et al., 2013), and because income tax liability is progressive, the net effect here is likely an increase in the progressivity of the filing burden.

Second, to the extent that pre-population alters tax payments (that is, to the extent that ϕ_+ or ϕ_- are positive), it could alter the progressivity of the tax system. We provide some evidence on the distribution of Δt_i among filers in Figure 3. Specifically, we plot the average difference between pre-populated and actual liability as a share of income by percentile of income (AGI as filed).[31] Average pre-populated liability exceeds true liability for every percentile except the top three percent; differences are especially large for those with income between \$10,000 and \$30,000 where the EITC may be especially relevant.[32] Pre-populated liability is smaller than true liability in the highest three percentiles, and especially so in the top one percent, likely due to receipt of income types that are subject to less information reporting. Together, these results mean that the tax liability on pre-populated returns is less progressive than existing liability. Under our modeling assumptions, this implies that (so long as ϕ_+ and ϕ_- are positive) pre-population could significantly reduce the effective progressivity of federal income taxes. However, this effect could be muted if the highest-income groups were ineligible for pre-population, or if ϕ_+ and ϕ_- were heterogeneous by income.

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-

[31]The figure excludes those with negative income and, for readability, excludes those in the lowest percentile of positive income, whose mean difference is approximately \$200 relative to mean income of \$270.

[32]Appendix Figure A3 provides additional detail on the distribution of pre-populated liability over- and understatements.

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 7.5 million non-filers who are not required to file yet are owed refunds averaging $390.

As discussed above, the experience of other OECD countries provides some context for our results. Twenty-eight of 38 OECD countries pre-fill taxpayer information and some forms of income.[33] Experiences with pre-filled returns have not always led to successful broad implementation. A comparison of experiences across various jurisdictions suggests that those who used the pre-filled returns were very satisfied, yet not all jurisdictions were able to expand pilots to cover the majority of taxpayers; some ended the projects or limited pre-filled returns to a small group of taxpayers (Vaillancourt, 2011). Take-up among eligible taxpayers also varies substantially across jurisdictions, which is an important consideration for the relative costs and benefits of such a program. In our analysis of the marginal value of public funds, variation in the reduction to compliance burden parameter (ψ) can move the MVPF from values below one to around two, even when IRS implementation costs are high (see Table 9 Panel C). The varied international experiences point to both the role of implementation and the high level of uncertainty about take-up and success associated with a new pre-population program.

While our analysis adds clarity to the estimated success rates of pre-population and for whom, it is also critical to understand the many parameters, costs, and benefits that we have not estimated or considered. For example, pre-population will likely require either a shift in when information returns (e.g. W-2s, 1099s) must be filed, or a delay in when taxpayers file individual returns using pre-populated data. Additionally, to the extent that pre-population acts as a nudge, it could alter the distribution of who pays the federal income tax, potentially in a regressive manner.

Implementation details are likely essential in determining the overall success of a pre-population program. Such implementation details include but are not limited to the following considerations. First, the program could approximate tax unit composition and dependent status (based on prior years' information, as in our baseline approach) or ask tax units to provide this information prior to or concurrent with pre-population. Second, in the case of taxpayer inaction, the IRS could treat the pre-populated return as accepted by the taxpayer or develop an alternative approach. Third, it may be beneficial for the IRS to balance the goals of making the program user-friendly while guaranteeing data security and taxpayer privacy. Finally, the IRS may consider what role – if any – the pre-populated return should play in post-filing enforcement activities.

In sum, we find that tax returns can be pre-populated accurately for nearly half 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*

[33] Data from the OECD Tax Administration Services database at `https://data.rafit.org/regular.aspx?key=63544710`

Notes, 107.

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

Blaufus, Kay, Frank Hechtner, and Janine K Jarzembski. 2019. "The income tax compliance costs of private households: empirical evidence from Germany." *Public Finance Review*, 47(5): 925–966.

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.

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

Finkelstein, Amy, and Nathaniel Hendren. 2020. "Welfare Analysis Meets Causal Inference." *Journal of Economic Perspectives*, 34: 146–67.

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.

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).

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2021. "Tax Evasion at the Top of the Income Distribution: Theory and Evidence." *NBER Working Paper*.

Hauck, Tobias, and Luisa Wallossek. 2022. "Optional (Non-)Filing and Effective Taxation." *Working Paper*.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A unified welfare analysis of government policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318.

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.

Internal Revenue Service. 2020. "IRS Tax Year 2019 Form 1040 and 1040-SR Instructions." *https://www.irs.gov/pub/irs-prior/i1040gi–2019.pdf*.

Internal Revenue Service. 2022. "Internal Revenue Service Data Book, 2021." Washington DC Publication 55-B.

Klun, Maja. 2009. "Pre-filled Income Tax Returns: Reducing Compliance Costs for Personal Income Taxpayers in Slovenia." *Financial Theory and Practice*, 33: 219–233.

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

Marcuss, Rosemary, George Contos, John Guyton, and Patrick Langetieg. 2013. "Income Taxes and Compliance Costs: How are they related?" *National Tax Journal*, 66: 833–854.

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. Bureau of Labor Statistics. 2023. "Average Hourly Earnings of All Employees, Total Private [CES0500000003]." *retrieved from FRED, Federal Reserve Bank of St. Louis*.

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.

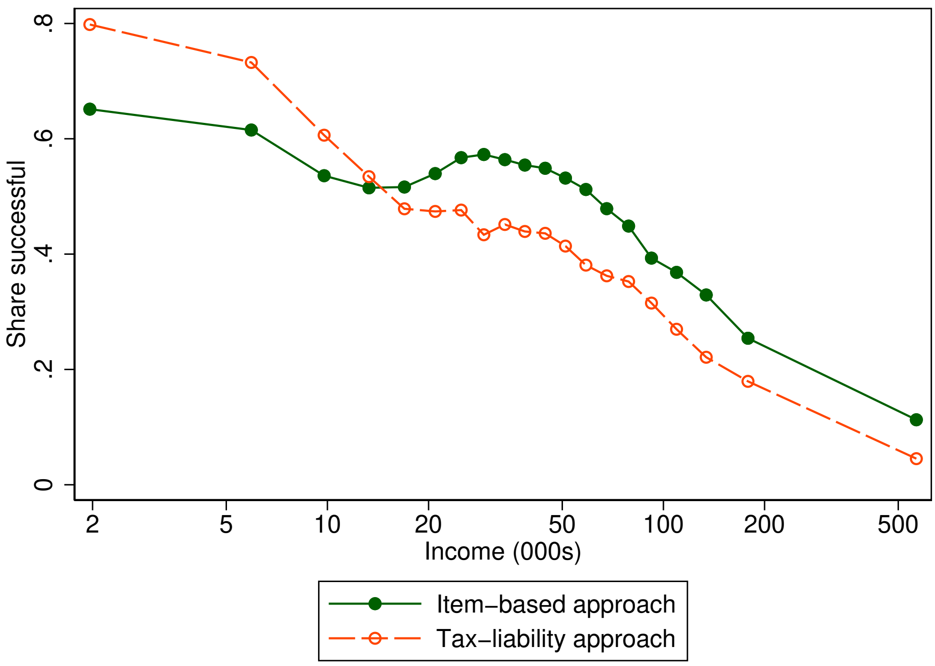
Vaillancourt, François. 2011. *Prefilled Personal Income Tax Returns: A Comparative Analysis of Australia, Belgium, California, Québec, and Spain.* Fraser Institute.

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.

Warren, Elizabeth. 2022. "Senator Warren Leads 22 Colleagues in Introducing the Tax Filing Simplification Act of 2022." Press Release.

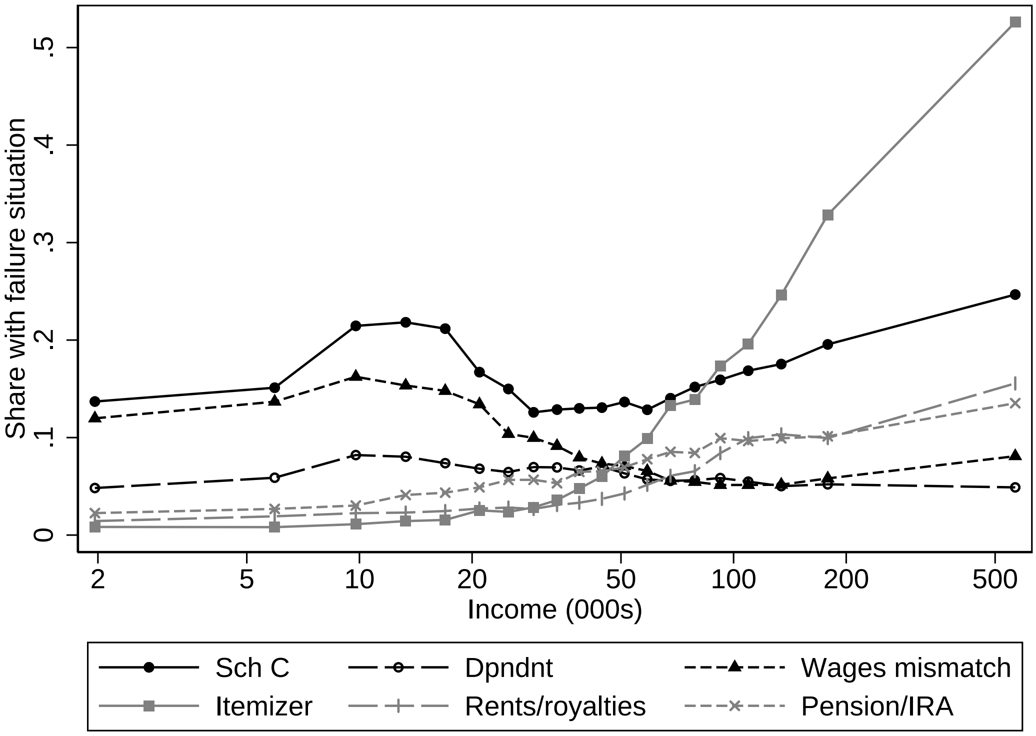
Tables and Figures

Figure 1: Pre-populated success rates by income



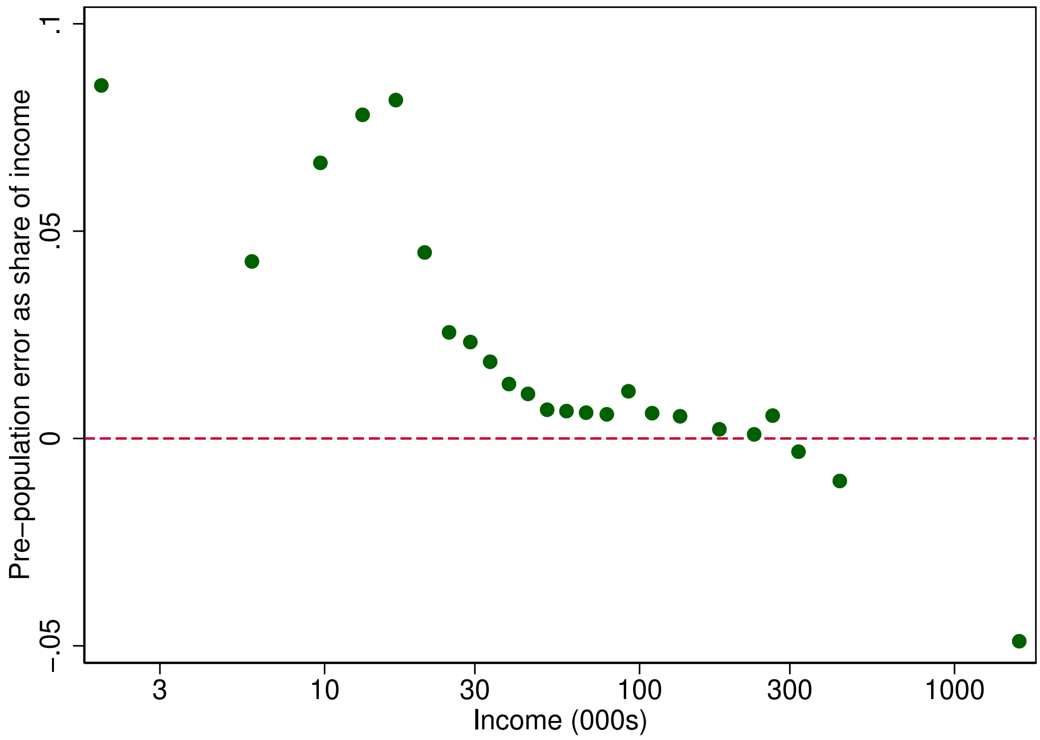
Notes: The figure plots the rate of accurately pre-populated tax returns for the 2019 tax filing population as a function of taxpayer-reported adjusted gross income under each of our two approaches. We use the preferred tolerance thresholds given in column 3 of Table 1. Data points are shown at the average 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: Mean pre-population error scaled by income, as a function of income



Notes: This figure plots the ratio of the unconditional mean pre-population error (that is, pre-populated liability minus liability as filed) to income, separately by percentile of income. Income is defined as adjusted gross income (AGI) as filed on Form 1040. Each bin is a percentile of positive income (filers with non-positive income are dropped). We omit the lowest percentile bin for readability; its value would be approximately 0.75. 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: Tax-liability approach					
Baseline	0.340	0.387	**0.422**	0.467	0.554
With known deps. & filing status	0.381	0.433	0.472	0.523	0.615
Tax difference tolerance	$10	$50	$100	$200	$500
Panel B: Item-based approach					
Baseline	0.429	0.445	**0.478**	0.507	0.596
With known deps. & filing status	0.473	0.491	0.528	0.559	0.660
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 tax-liability approach, so the relevant tolerance is the acceptable difference between simulated and actual tax liability. Panel B shows results for the item-based 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 item-based approach 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	0.166	0.077
Itemized deductions in 2019	0.109	0.040
Wages do not match W-2 wages	0.092	0.052
Taxable pension/IRA income does not match 1099-R	0.068	0.017
Dependent mismatch from 2018 to 2019	0.062	0.019
Sched E rents/royalties (except from K-1)	0.054	0.014
Child/dependent care credit	0.038	0.017
Capital gains income does not match 1099-B	0.035	0.008
Change in filing status from 2018 to 2019	0.035	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
S corp/partnership income does not match K-1	0.027	0.004
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	0.001
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	0.522	
Precisely one failure situation	0.298	

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**

	Item-based approach		Tax-liability approach		
	No failure situation?		Correct calculation?		
	✓	X	✓	high	low
	(1)	(2)	(3)	(4)	(5)
Mean					
Married	0.26	0.47	0.24	0.45	0.50
Has dependents	0.23	0.39	0.17	0.47	0.33
Uses paid preparer	0.44	0.58	0.43	0.55	0.62
Primary filer age	42	48	43	45	51
AGI (thousands)	47.1	103.1	39.1	82.0	143.4
Liability: calculated (thousands)	3.8	15.0	2.9	13.4	16.6
Liability: taxpayer-reported (thousands)	3.7	15.0	2.9	8.8	24.9
Liability: calculated less reported (thousands)	0.1	-0.4	0.0	4.5	-8.3
Failure situations	0.0	1.7	0.2	1.4	1.4
Median					
Primary filer age	39	47	38	43	50
AGI (thousands)	33.4	51.9	27.3	49.8	64.1
Liability: calculated (thousands)	1.5	2.6	0.9	3.7	2.1
Liability: taxpayer-reported (thousands)	1.4	2.3	0.9	1.7	4.4
Liability: calculated less reported (thousands)	0.0	0.3	0.0	1.6	-1.2
Failure situations	0	1	0	1	1
Count (millions)	74.9	81.7	66.1	58.8	31.7

Notes: The table provides mean and median values for 2019 tax unit characteristics conditional on either success or failure of our item-based and tax-liability 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 item-based approach is successful, while the second column describes those for whom it is unsuccessful. The third column describes taxpayers for whom the tax-liability approach is successful. The fourth and fifth columns describe taxpayers for whom the tax-liability approach is unsuccessful separated by whether the simulated tax is too high or too low, respectively, relative to reported liability. Medians of dollar-denominated variables are rounded to the nearest $100.

Table 4: **Mismatch rates for selected lines on Form 1040**

Line	Unconditional mismatch rate	Conditional mismatch rate
	(1)	(2)
Wages	0.084	0.099
Taxable interest	0.030	0.082
Qualified dividends	0.023	0.096
Income from Sch. 1	0.302	0.755
Taxable IRA and pensions	0.071	0.194
Capital gains	0.107	0.559
AGI	0.476	0.473
Taxable income	0.461	0.532
EITC	0.110	0.548
Child tax credit	0.145	0.375

Notes: This table reports the match rates for several lines on Form 1040. Column (1) reports the share of observations whose values derived from information returns matches the amounts reported on Form 1040. In each row of column (2), we restrict to observations that report non-zero values of that particular line item. 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 population	Cumulative success rate		Marginal success rate	
		Tax-liability approach	Item-based approach	Tax-liability approach	Item-based approach
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.20	0.79	0.81	0.79	0.81
2. Allow married	0.22	0.77	0.79	0.55	0.54
3. Allow dependents	0.30	0.68	0.73	0.41	0.58
4. Add interest/dividends	0.34	0.68	0.74	0.75	0.82
5. Add Social Security	0.36	0.69	0.74	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	0.67	0.73	0.60	0.60
9. Add high income	0.52	0.66	0.72	0.51	0.68
10. All income types	0.69	0.55	0.59	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: **2014 National Research Program: rate at which certain income items require audit correction**

Line	Unconditional correction rate	Conditional correction rate
	(1)	(2)
Wages	0.045	0.052
Taxable interest	0.006	0.011
Qualified dividends	0.005	0.017
Income from Sch. 1	0.171	0.387
Taxable IRA and pensions	0.024	0.045
Capital gains	0.029	0.118

Notes: This table reports the rate at which various income items from Form 1040 were corrected under audit. The sample is the set of randomized audits undertaken during the National Research Program for tax year 2014; the analysis uses sample weights to make the sample representative of all 2014 tax units. We use a tolerance of $500. Column (1) reports the unconditional correction rate. In each row of column (2), we restrict to observations that report non-zero values of that particular line item. “Schedule 1 income” refers to the set of income items that were included on Schedule 1 of the 2019 Form 1040.

Table 7: **Cross-tabulation of W-2 mismatches and wage audit corrections, 2014 National Research Program sample**

	Any wage audit correction?	
	No	Yes
W-2/1040 match	0.902	0.005
W-2/1040 mismatch	0.043	0.050

Notes: This table reports a cross-tabulation of (1) whether the Form 1040 wage amount matches the Form W-2 (box 1) wage amount within \$500 (rows) against (1) whether the taxpayer received an audit correction on the wage line in excess of \$500 (columns). The sample is restricted to those who have positive wages either on Form W-2, Form 1040 prior to audit, or Form 1040 after audit. The sample is the 2014 National Research Program sample; sample weights are used to make the sample representative of 2014 tax filers.

Table 8: **Pre-populated return characteristics for non-filers**

	No filing obligation		Has filing obligation	
	Mean	Median	Mean	Median
Age	56	59	47	46
Female	0.50	–	0.33	–
Filed a 2018 return	0.11	–	0.36	–
Pre-populated return with dependents	0.02	–	0.09	–
Calculated childless EITC >0	0.077	–	0.058	–
Calculated child EITC or CTC >0	0.007	–	0.089	–
...with revealed dependents	0.008	–	0.095	–
Calculated AGI	1,117	0	49,997	34,300
Calculated AGI (if potential refund)	3,825	2,800	38,399	27,200
Tax withholding	36	0	3,848	1,500
Tax withholding (if >0)	270	100	4,730	2,200
Has potential refund	0.16	–	0.44	–
Has potential taxes owed	0.00	–	0.54	–
Potential refund (if >0)	390	200	1,531	900
Potential taxes owed (if >0)	–	–	3,982	1,400
Count (millions)	47.1		7.8	

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.

Table 9: **Marginal value of public funds (MVPF) estimates**

Panel A: implementation costs $\alpha = 0.01$			
	compliance burden reduction		
adjustments to pre-populated liability	$\psi = 0.05$	$\psi = 0.10$	$\psi = 0.20$
$(\phi_+, \phi_-) = (0.0, 0.0)$	4.40	8.87	17.99
$(\phi_+, \phi_-) = (0.1, 0.1)$	2.01	3.33	6.02
$(\phi_+, \phi_-) = (0.1, 0.2)$	1.08	1.18	1.38
$(\phi_+, \phi_-) = (0.1, 0.3)$	1.04	1.09	1.20
Panel B: implementation costs $\alpha = 0.05$			
	compliance burden reduction		
adjustments to pre-populated liability	$\psi = 0.05$	$\psi = 0.10$	$\psi = 0.20$
$(\phi_+, \phi_-) = (0.0, 0.0)$	0.88	1.77	3.60
$(\phi_+, \phi_-) = (0.1, 0.1)$	0.92	1.52	2.76
$(\phi_+, \phi_-) = (0.1, 0.2)$	0.99	1.08	1.27
$(\phi_+, \phi_-) = (0.1, 0.3)$	0.99	1.04	1.14
Panel C: implementation costs $\alpha = 0.10$			
	compliance burden reduction		
adjustments to pre-populated liability	$\psi = 0.05$	$\psi = 0.10$	$\psi = 0.20$
$(\phi_+, \phi_-) = (0.0, 0.0)$	0.44	0.89	1.80
$(\phi_+, \phi_-) = (0.1, 0.1)$	0.55	0.91	1.65
$(\phi_+, \phi_-) = (0.1, 0.2)$	0.90	0.98	1.15
$(\phi_+, \phi_-) = (0.1, 0.3)$	0.94	0.99	1.08

Notes: The table presents estimates of the MVPF for pre-populating tax returns, under a range of parameter values. Parameter α is greater when pre-populated returns are more costly for the IRS to implement. Parameter ψ is greater when pre-populated returns are more beneficial for taxpayers, in terms of reducing compliance costs. Parameter ϕ_+ (ϕ_-) is greater when taxpayers are more likely to make unfavorable (favorable) adjustments to pre-populated returns, increasing (decreasing) tax payments relative to accepting the pre-populated return. See Section III.D for further discussion of these parameters and, more generally, the assumptions underlying our formulation of the MVPF.

A Additional details

In this appendix, we provide additional details on the assumptions we make in our item-based and tax-liability 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 who 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.[34]

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 item-based failure situation. In such cases, the IRS would have prepared a return that includes the wrong people. We do not attempt the tax liability computation in these cases. Rather, we code this as an immediate tax-liability 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.

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

A.1.3 The treatment of dependents

For the item-based 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 tax-liability 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 item-based 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.[35] 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 on information returns.

A.2 More detail on item-based failure situations

In this section, we define each of the item-based 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

[35]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 situation, and the majority of these taxpayers have only one.

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.[36]

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 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[37] – 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

[36]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.

[37]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.

have received a taxable scholarship. See Section III.B for further discussion.

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 157 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 for tax year 2019. Median AGI in the sample is $40,700, median taxable income is $23,700, and median tax liability is $1,700 (all rounded to the nearest hundred). All three measures are skewed in the sample with much larger means than medians.

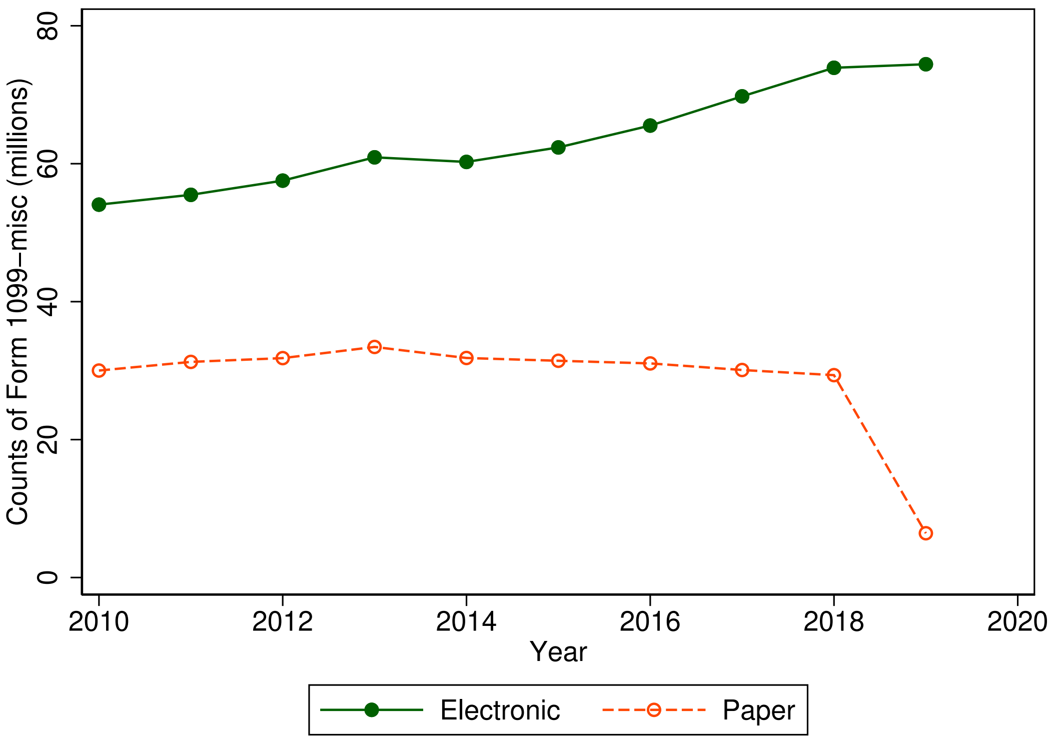
A.3.1 Success rates by age

Figure A2 shows how our item-based and tax-liability estimates of the share of pre-populated returns that are successful vary by age. The item-based success rate starts at just over 80 percent for 18-year-olds but falls steadily to around 40 percent for 36-year-olds. It stays relatively flat across the age distribution from there. The tax-liability success rate starts around 90 percent for 18-year-olds and falls even more sharply than the item-based approach, 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

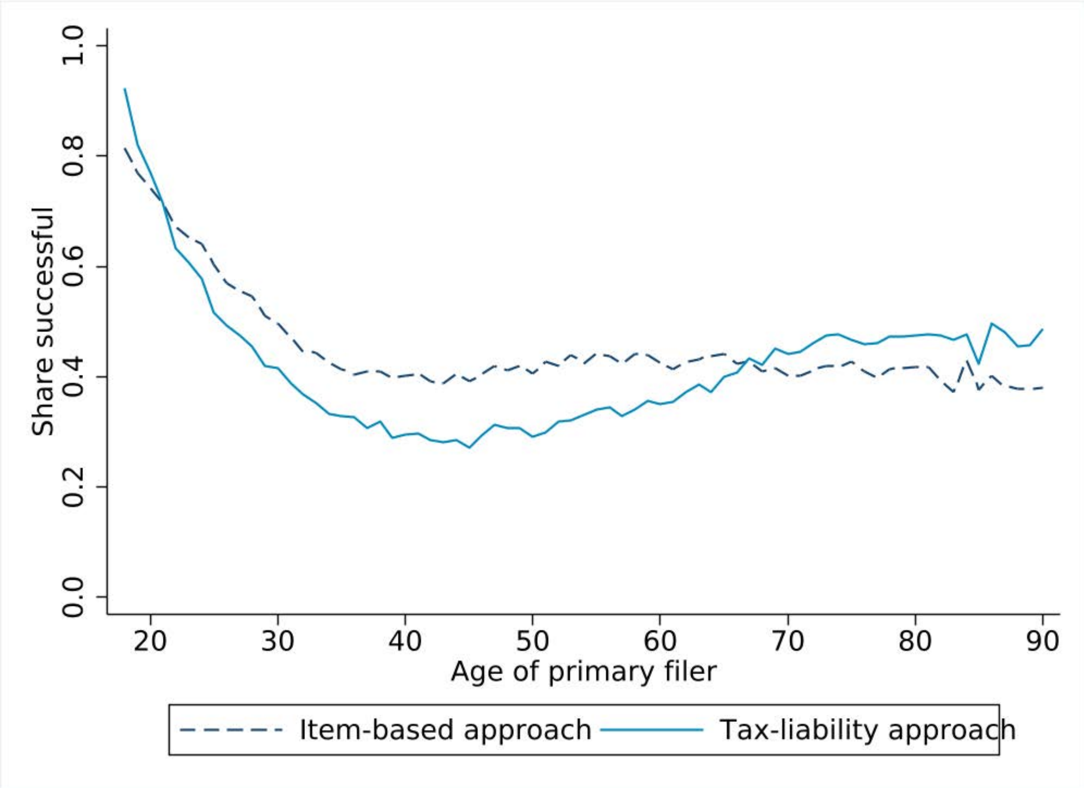
Figure A3 breaks out our tax-liability estimates by whether the pre-populated return under- or overstates liability. In panel (a) we see that, in all income ventiles except for the top five percent, pre-populated returns are more likely to overstate liability than to understate it. Panel (b) further illustrates that for most income bins, the average magnitude of the error is below $5,000 (conditional on there being an error in excess of $100 magnitude). Panel (c) shows these error magnitudes relative to income (omitting the bottom income ventile, which has extremely low income). We see that, for the bottom twenty-five percent of taxpayers, when liability is overestimated, it is overestimated by a sizeable share of income: well over twenty percent. But for most taxpayers, the average conditional errors in pre-populated liability are more modest. Panel (d) goes beyond averages to show various percentiles of overstated liability on the pre-populated return. In most income bins, when pre-populated liability is overstated, the median amount of overstatement is a few thousand dollars. However, the 75th percentile of overstatement is over $5,000 for many income bins. Panel (e) shows that the situation is much different for understatement of pre-populated liability. Only the top few ventiles see a 75th percentile of understatement greater than $5,000, conditional on understatement.

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



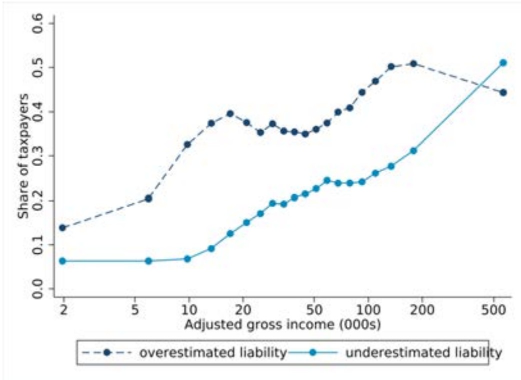
Notes: The figure plots counts of Form 1099-MISC in the underlying database of administrative records from which we draw our data, separately by electronically-filed and paper-filed forms.

Figure A2: Pre-population success rates by age

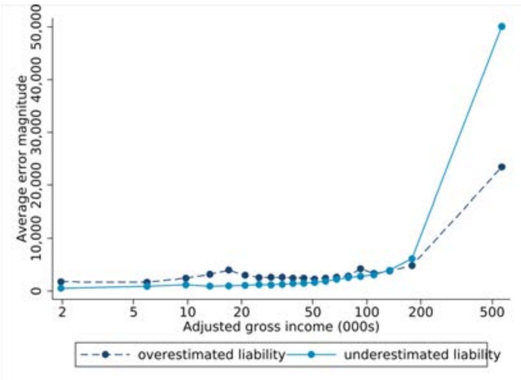


Notes: The figure plots the share of tax year 2019 pre-populated returns that would be within our preferred tolerance levels for accuracy under our item-based and tax-liability approaches, by primary filer age. Sample weights are used; thus the results are representative of the 2019 tax filing population.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

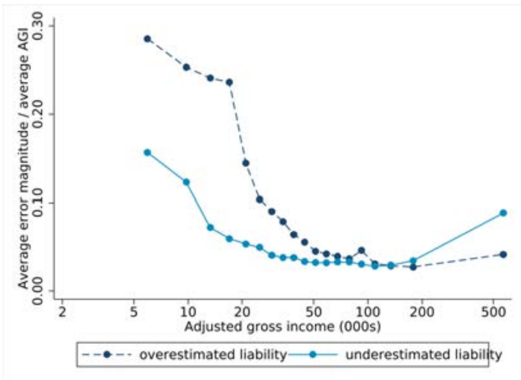
Figure A3: Lower bound direction and magnitude of liability errors



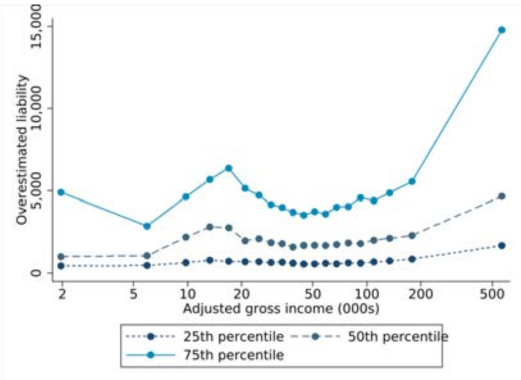
(a) Share with errors



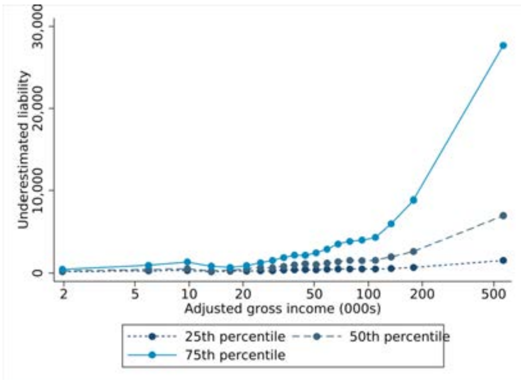
(b) Average error magnitude



(c) Average error magnitude relative to AGI



(d) Percentiles of overstated liability



(e) Percentiles of understated liability

Notes: The figure plots information regarding 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 taxpayers belonging to 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 average error magnitudes of panel (b) divided by average AGI within each AGI bin. Panels (d) and (e) plot the 25th, 50th, and 75th percentiles of error magnitude for the groups with overestimated and underestimated liability, respectively. In all panels, 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 errors relative to its small average AGI. The figure was created by the authors using IRS tax data.

Table A1: **Summary statistics**

	Mean	Median
Primary filer age	45	43
Married	0.37	–
Has dependents	0.32	–
Number of dependents	0.56	0
Uses paid preparer	0.51	–
Claims Earned Income Credit	0.17	–
Claims Child Tax Credit	0.30	–
Adjusted gross income	76,310	40,700
Taxable income	58,863	23,700
Tax liability	9,588	1,700
Count (millions)	156.6	

Notes: The table describes our sample of 2019 tax units from IRS Statistics of Income data. 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 population	Cumulative success rate		Marginal success rate	
		Tax-liability approach	Item-based approach	Tax-liability approach	Item-based approach
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.17	0.72	0.76	0.72	0.76
2. Allow married	0.27	0.68	0.70	0.62	0.59
3. Allow dependents	0.35	0.61	0.67	0.39	0.56
4. Add interest/dividends	0.38	0.62	0.67	0.65	0.74
5. Add Social Security	0.42	0.63	0.68	0.73	0.74
6. Add pension/IRA distributions	0.44	0.63	0.68	0.58	0.58
7. Add gambling, UI, state tax refunds	0.47	0.62	0.67	0.53	0.66
8. Add capital gains	0.48	0.62	0.67	0.59	0.58
9. Add high income	0.53	0.60	0.66	0.44	0.58
10. All income types	0.69	0.53	0.57	0.28	0.27
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.