Estimating the Costs of Filing Tax Returns and the Potential Savings from Policies Aimed at Reducing these Costs*

Youssef Benzarti (UCSB and NBER)

September 3, 2020

1 Executive Summary

The compliance costs of taxation, which encompass the money, time, hassle and effort spent filing tax returns every year, have received renewed interest recently with the passing of the Tax Cuts and Jobs Act in 2017 which promised to reduce these costs by shrinking the 1040 form to the size of a postcard. Attempts at reducing tax filing costs date back to at least the Tax Reform Act of 1986, that aimed at simplifying the tax code, in part by reducing the number of tax brackets and increasing the standard deduction. However, in spite of the policy interest surrounding this issue, there has been limited research aimed at putting a dollar value on these costs.

This paper presents a simple method of estimating tax filing costs. Importantly, this method moves beyond survey evidence, which has only accounted for the time spent filing taxes. Instead, the approach taken in this paper is able to capture all monetary costs, effort costs and disutility costs taxpayers experience by estimating their willingness to pay to avoid filing taxes. The intuition behind this approach is simple: individuals will only perform a

*Youssef Benzarti, UCSB and NBER (benzarti@ucsb.edu). This paper draws on some of the results and methods from Benzarti (2020): sections 2, 3.1 and 3.2 are mostly based on Benzarti (2020), while the rest is mostly new. I am very grateful to Robert Moffitt and Alisa Tazhitdinova.

costly action if the benefit of doing so exceeds its cost. Therefore, one can estimate the cost by measuring the forgone benefits associated with the action. I apply this intuition to the decision to itemize deductions. Given that itemizing is more costly than claiming the standard deduction, it only makes sense to itemize if the associated costs are smaller than the tax benefits of itemizing. I test this prediction empirically and show that some taxpayers claim the standard deduction even if their itemized deductions exceed the standard deduction, resulting in a missing mass in the distribution of itemized deductions just above the standard threshold. The magnitude of this missing mass, i.e. the proportion of itemizers who forgo deductions, determines the distribution of the cost of itemizing.

I develop a method to measure the missing mass in order to estimate how much taxpayers are willing to pay in order to avoid the hassle of itemizing their deductions. This number ranges from $175 to $591 depending on the type of taxpayer. I also show that this heterogeneity in willingness to pay mostly stems from income differences, this in turn generates heterogeneity in the opportunity cost of time of each taxpayer. I also show that, while tax preparers and electronic filing have little bearing on the magnitude of this cost, the timing of filing matters substantially. Taxpayers who file their taxes close to the April 15th deadline, tend to forgo more deductions that those who file further away.

Using this estimate, survey evidence from the IRS on the time it takes to file given schedules and a new methodology, I extrapolate these costs in order to estimate compliance costs for different types of individuals. I then extrapolate itemization cost to other types of schedules by relying on survey evidence from the IRS. Accounting for all taxpayers, I find that tax filing in the U.S. imposes a yearly cost in excess of 1% of GDP and that this cost has been steadily increasing since the 1980's. While this increase is in part driven by the fact that the number of tax filers increased, it is also due to the fact that taxpayers have to file an increasing number of schedules. For example, Schedule B must be filed when the taxable interest income or dividend income exceed $400. Because the $400 threshold has been nominally fixed for decades, the number of individuals who must file Schedule B has been steadily increasing.

A number of policies have been been proposed to reduce compliance costs. Using the methods and estimates derived above, I can provide an estimate of the efficacy of these policies at reducing compliance costs. The first such policy is to pre-populate tax forms with information that the IRS currently possesses (so no additional collection efforts are needed). Overall, such a policy could cut compliance costs by approximately 43%. Similarly, the government could offer a free electronic filing option to all taxpayers. Currently, only certain taxpayers have access to free electronic filing options, but this could be expanded, presumably at low costs. My estimates suggest that a free e-filing option could lead taxpayers to save $5.1 billion annually in fees paid to purchase tax software or use the services of tax preparers. Implementing one (or both) of these policies would bring the US tax system in line with most tax systems in OECD countries.

2 Estimating the Cost of Filing Tax Returns

2.1 Data

The empirical analysis in this paper relies primarily on the Statistics of Income (SOI) dataset which consists of a random sample of tax returns. These tax returns are sampled every year and represent approximately 0.1% of the US population. The randomization is designed to create a representative sample of the US population and weights are provided in the SOI dataset in order to make the sample representative. Note that the dataset oversamples high-income individuals allowing more precise estimates at the top of the income distribution. This dataset has been used extensively in public finance, including, for example, when analyzing the evolution of top-income earners as in Piketty and Saez (2003).

While this dataset is well suited for the current analysis, there are other administrative datasets that could be used with the proposed approach. Naturally, the full sample of US tax returns would allow more in depth analysis, since it not only contains more information (such as location, education level, family information, etc.) but it is also longitudinal and provides information on third-party reported items such as 1098 forms, W2 forms, etc.

This type of information can further shed light on the specific mechanics that are driving compliance costs and could help further validate the empirical approach adopted in this paper. For example, having access to the W2 and 1098 forms would allow identification of a portion of the missing deductions without relying on any particular assumptions: one would simply count the number of individuals who do not itemize and yet have mortgage interest and state tax amounts in excess of the standard deduction. Similarly, knowing what level of education a given taxpayer has reached could help explain why so much money is left on the table. As has been documented in the personal finance literature (e.g. Van Rooij et al. (2011)), financial literacy is one important driver of money left on the table, as is the case, for example, for the failure to save for retirement. Finally, the longitudinal dimension of the dataset would help track taxpayers over time and help us understand whether they are prone to learning and how they respond to different events. Longitudinal data could help us understand how long it takes for a taxpayer to start itemizing once their total deductions cross the standard deduction threshold, and whether it takes longer for less financially literate taxpayers or faster for taxpayers who use a tax preparer.

2.2 Approach

The basic approach used to identify and estimate the cost of filing tax returns relies on a simple and easy to export approach that requires minimal assumptions. There are two important data ingredients: repeated cross sectional data (longitudinal data would be even better but is not required) and a change in the amount of the filing threshold (an increase in the standard deduction amount in this case). With these two ingredients, a researcher can apply the methodology below to another setting and get the cost estimate that is the basis for estimating costs for the rest of the population and other forms.

This simple approach relies on the observation that, in the case of any filing requirement where a threshold is involved above which a taxpayer has to file a separate return or perform a costly action (submit a receipt, implement more calculations, consult with an accountant, etc.), we should observe a missing mass in the distribution of tax preparers just above the threshold. For example, if the cost of itemizing relative to claiming the standard deduction is positive, some taxpayers will choose the standard deduction even when the sum of their itemized deductions is greater than the standard deduction. This implies that the distribution of itemized deductions will have a missing mass just above the standard deduction threshold. The magnitude of the missing mass, i.e., the proportion of itemizers who forgo deductions, determines the distribution of the cost of itemizing.

This prediction can be easily tested by plotting the distribution of itemizers in the neighborhood of the standard deduction threshold. Using the SOI dataset, these distributions can be plotted for all available years and all filing statuses. Figure 1 shows the distributions for a number of years using the SOI data.[1] As can be easily seen, each distribution exhibits a missing mass right above the standard deduction threshold: the distribution of itemizers is systematically lower close to the standard deduction, increases steadily and then peaks two to three bins away from the standard deduction before decreasing again.

An important point worth noting is that this evidence is only suggestive. This is because we cannot observe the distribution of itemizers below the standard deduction threshold in any available datasets, even ones that are only available to the IRS. Indeed, while the mortgage interest and state and local taxes are third-party reported to the IRS, a number of deductions are not and are therefore not recorded in any data. These deductions include medical expenses and charitable donations, which constitute approximately 30% of an average taxpayer's total deductions. If we do not know what shape the distribution of deductions has below the standard deduction threshold, we cannot make any definitive statements as to whether the standard deduction threshold is *causing* the missing mass or whether the missing mass is merely a natural feature of the distribution. This issue is illustrated in Fig-

[1] Note that all remaining years can be found in Benzarti (2020). Note also that all dollar amounts discussed in the paper are in 2016 dollars.

ure 2, which plots hypothetical distributions below the standard deduction threshold and connects them to the observed distribution of itemizers above the standard deduction. Case (a), which assumes that the distribution of itemizers below the threshold is simply smoothly increasing throughout its support and then peaks just above the standard deduction, is impossible, simply because it would fail to account for most of the population of taxpayers who choose the standard deduction. Case (b), however, although unlikely, is possible: it assumes a distribution that is double peaked. Case (c), which is the one that is consistent with the standard deduction causing a missing mass in the distribution of itemizers just above the standard deduction is the most likely, but cannot be confirmed with only one cross section of data.

I now turn to a quasi-experimental design in order to show that the standard deduction threshold is indeed causing the missing mass in the distribution of itemizers by prompting some taxpayers to forgo tax benefits from itemizing and claim the standard deduction to avoid the costs of filing schedule A. My approach uses a change in the standard deduction amount and a pre- and post-change cross sections of the dataset to estimate the counterfactual distribution of itemized deductions and the implied compliance costs. The standard deduction changed four times during the studied period – in 1971, 1975, 1988 and 2003. I focus on the 1988 change for two reasons. First, it is the largest change thus providing more precise estimates. And second, the 1971, 1975 and 2003 changes coincided with other changes to the itemization rules.[2]

Figures 3a and 3b show the 1987-1989 and 1987-1988 distributions of itemized deductions. Three main patterns emerge. First, the missing mass clearly tracks the standard deduction threshold: as the standard deduction increases, the missing mass moves with the standard deduction. Second, when comparing the pre- and post-reform distributions away from the standard deduction it is clear that they are overlapping, i.e. only the distribution in the neighborhood of the standard deduction is affected by the standard deduction, the

[2] The graphical evidence for the 1971, 1975 and 2003 standard deduction changes can be found in Benzarti (2020).

remaining portion of it remains the same, further reinforcing that the standard deduction is indeed causing the missing mass. Third, the missing mass appears larger for the 1987-1989 distributions relative to the 1987-1988 distributions consistent with the idea that there is likely to be a lagged response. A likely explanation is that some taxpayers may only learn about the standard deduction increase once they file their return in 1988 and would, in that case, still itemize even if it would only save them a few dollars, simply because they have already done most of the record-keeping and have already worked through Schedule A. I account for such lagged responses by estimating the compliance costs using the 1987 pre-reform and 1989 post-reform distributions of itemizers.

Ideally, had the distribution of itemized deductions been observable, we could simply estimate compliance costs by comparing the difference in total itemized deductions available versus itemized deductions claimed, which would provide an accurate estimate of forgone deductions. If instead, we use the pre-reform distribution as a proxy for the distribution of itemized deductions, we would estimate a lower bound on the amount of forgone deductions, simply because the pre-reform distribution also exhibits a missing mass. However, the pre-reform distribution can be adjusted, using the post-reform distribution, in order to reflect the true counterfactual distribution of itemized deductions. Here I explain this procedure using a graphical and intuitive representation of the approach and a hypothetical example borrowed from Benzarti (2020). For a more formal treatment of the approach, interested readers can refer to Benzarti (2020).

Assume that $f(\cdot)$ is the counterfactual distribution of itemizers where all taxpayers with total deductions in excess of the standard deduction itemize (as if it was costless to itemize). This distribution corresponds to the lightest distribution in Figure 4. Further, assume that each bin size is equal to \$100 of deductions and that the cost distribution in the population is given by the following, where the cost is expressed in deductions rather than dollars and $C(\cdot)$ represents the cumulative distribution function (CDF) of costs:[3]

[3] In order to convert deductions into dollars, one simply needs to multiply the deduction amount by the marginal tax rate.

- 40% have a cost lower than 100, i.e, $C(100) = 40\%$
- 50% have a cost lower than 200, i.e, $C(200) = 50\%$
- 60% have a cost lower than 300, i.e, $C(300) = 60\%$
- 70% have a cost lower than 400, i.e, $C(400) = 70\%$
- 80% have a cost lower than 500, i.e, $C(500) = 80\%$
- 90% have a cost lower than 600, i.e, $C(600) = 90\%$
- 92% have a cost lower than 700, i.e, $C(700) = 92\%$
- 96% have a cost lower than 800, i.e, $C(800) = 96\%$
- 100% have a cost lower than 900, i.e, $C(900) = 100\%$

Assume that the standard deduction amount corresponds to the second bin in Figure 4. Consequently, the mass at every bin that comes after the standard deduction should be affected by the cost defined above. Denote by $g_0(\cdot)$ the pre-reform distribution, which we can estimate using the SOI dataset directly and represents the distributions plotted in Figure 1. $g_0(\cdot)$ corresponds to the second lightest distribution In Figure 4. Estimating the cost would have been easy had $f(\cdot)$ been observable: we would simply estimate the percentage difference between $f(\cdot)$ and $g_0(\cdot)$. But since $f(\cdot)$ is not observable we have to rely on comparing the pre- and post-reform distributions $g_0(\cdot)$ and $g_\delta(\cdot)$, where $g_\delta(\cdot)$ is the post-reform distribution of itemizers when the standard deduction increases by δ. For the purpose of illustration, assume that the standard deduction amount increases by \$500 (5 bins). Denote by d_i the distortion introduced by the standard deduction in bin i. Since 40% of the population has a cost that is smaller than 100, as assumed above, 60% of the population will claim the standard deduction in the first bin, which will be distorted by 60% i.e. $d_1 = 60\%$. Similarly, $d_2 = 50\%$, $d_3 = 40\%$, $d_4 = 30\%$, $d_5 = 20\%$, $d_6 = 10\%$, $d_7 = 8\%$, $d_8 = 4\%$ and $d_i = 0$ for any $i > 8$. Then we can recover the true density of deductions $f(\cdot)$ as follows.

Denote by b_i a bin of size 100 that starts at i dollars, i.e, $b_{200} = [200, 300]$ and $f(b_{200})$ is the number of itemizers with total deductions in the 200 to 300 dollar range. Since $f(\cdot)$ is the undistorted counterfactual distribution of itemizers, $f(b_{200})$ corresponds to the counterfactual number of itemizers with deductions ranging from 200 to 300 dollars had there been no distortion imposed by the cost of itemizing, i.e., every person, irrespective of their cost would itemize if their deductions exceed the standard deduction. Similar notation applies to $g_0(\cdot)$ and $g_\delta(\cdot)$: $g_0(b_{200})$ is the number of itemizers with deductions 200 to 300 in excess of the pre-reform standard deduction; and $g_\delta(b_{200})$ is the number of itemizers with deductions 200 to 300 in excess of the post-reform standard deduction once it increases by δ.

Using this notation, we can introduce the adjustment process. Start by considering $g_\delta(b_{900})/g_0(b_{1400})$, which represents the ratio of the darkest and second darkest distributions at the bin that is 14 bins away from the pre-reform distribution (and nine bins away from the post-reform distribution), in Figure 4. Here, $g_0(.)$ and $g_\delta(.)$ are overlapping, which implies that $g_\delta(b_{900})/g_0(b_{1400}) = 1$.

Since $g_\delta b_{900}/g_0 b_{1400} = 1$, no taxpayer has a cost in excess of 900. This has two implications: (1) costs do not exceed \$900, and relatedly, (2) both the pre- and post-reform distributions are undistorted when considering taxpayers with deductions \$900 in excess of the standard deduction. This means that for any deductions \$900 in excess of the standard deduction, $f(\cdot) = g_0(\cdot)$, i.e., $g_0(\cdot)$ is the true counterfactual distribution and $C(x) = 1$ for $x \geq 900$. Similarly, consider $g_\delta(b_{800})/g_0(b_{1300})$. As established above, costs do not exceed \$900 and therefore $g_0(b_{1300})$ must be undistorted and hence $g_\delta(b_{800})/g_0(b_{1300}) = g_\delta(b_{800})/f(b_{1300}) = 96\%$.

The same procedure is applied to recover $C(700) = 92\%$, $C(600) = 90\%$, $C(500) = 80\%$ and $C(400) = 70\%$. These are all the bins in Figure 4 where $f(\cdot) = g_0(\cdot)$, and where $g_0(\cdot)$ can be used as the true counterfactual.

Next, notice that $f(\cdot)$ and $g_0(\cdot)$ do not overlap anymore three bins away from the standard deduction. Therefore, we need to re-construct $f(\cdot)$ from $g_0(\cdot)$ and $g_\delta(\cdot)$. Notice that $f(\cdot)$ and $g_0(\cdot)$ diverge 8 bins away from the pre-reform standard deduction and that we know, from the calculations above, that 4% of individuals will not itemize when their deductions are lower than \$800 (since $g_\delta(b_{800})/f(b_{1300}) = 96\%$). This implies that the distortion imposed by the cost of itemizing on $g_0(\cdot)$ is 4%. Therefore, $f(800) = g_0(800)/96\%$.

Now that we know $f(800)$, it follows that $C(300) = f(800) - g_\delta(300) = g_0(800)/C(300) - g_\delta(300)$, which is equal to 60%. Similarly, we can calculate $C(200) = f(700) - g_\delta(200) = g_0(700)/C(200) - g_\delta(200)$, which is equal to 50% and $C(100) = f(600) - g_\delta(100) = g_0(600)/C(100) - g_\delta(100)$, which is equal to 40% and therefore recover the entire cost distribution.

3 Cost of Filing Tax Returns

3.1 Estimating the Cost of Itemizing Deductions (Schedule A)

In this section I estimate the cost of itemizing deductions by applying the approach outlined in the previous section to the 1988 increase in the standard deduction, which increased it from \$2,540 to \$3,000 for single filers, from \$3,760 to \$5,000 for joint filers, and from \$2,540 to \$4,400 for head-of-household filers (in 1988 dollars).[4] The cost estimation is performed separately for each one of these filing categories, generating three different estimates. Note that standard errors are estimated using bootstrap.

Table 1 reports the results of this estimation for single, joint and head-of-households filers in the 15% and 28% marginal tax brackets. The estimated costs vary from 0.57% to 0.85% of adjusted gross income (AGI), or from \$175 for single filers in the 15% bracket to \$591 for joint filers in the 28% bracket. Hence, costs expressed in dollars are systematically lower for individuals in lower tax brackets but are substantially more similar when expressed as a percent of AGI. This suggests that income matters in determining the cost, which we discuss next.

3.2 Main drivers of the cost of itemizing

Costs Increase With Income If compliance costs are indeed preventing taxpayers from itemizing even though they could benefit from it, taxpayers with a higher value of time

[4]The allowable amount for the personal interest deduction was decreased in 1988, I adjust for this in my calculations (more details available in Benzarti (2020)).

should be forgoing even more benefits. This can be tested by breaking down the sample of itemizers by income and re-estimating the cost of itemizing for each income group.

To do so, I construct ten subsamples of the population of itemizers by deciles of income and estimate the cost separately for each one of these deciles. The results of this estimation are shown in Figure 6: as income increases, taxpayers forgo more deductions, consistent with the idea that they value their time more.

Record Keeping Compliance cost surveys have systematically reported that record keeping costs are the main driver of compliance costs, as shown, for example, in Guyton et al. (2003), Slemrod and Sorum (1984), Slemrod and Bakija (2008) and Blumenthal and Slemrod (1992). If this is true, the observed missing mass should be present even for individuals who use electronic filing (e-filing) and tax preparers since these only reduce the cost of filling out forms but do not affect the cost of record keeping.

This prediction can be easily tested by plotting the density of itemizers who use a tax preparer or e-file. Figures 7 shows these distributions: both exhibit a clear missing mass, implying that e-filing and use of tax preparers do not eliminate the cost of itemizing. This is consistent with record keeping costs being the main driver behind the large cost of itemizing deductions.

Procrastination Anecdotal evidence suggests that some taxpayers systematically wait until the April 15th deadline to file their returns. While, in principle, waiting to file one's return does not necessarily mean that it will be more time consuming or costly to file it, waiting until the last moment can prove out to be costly if one misplaces records or forms or gets a surprise shock specifically on April 15th.

This prediction can be tested with the SOI data, as well as with public data on when returns are filed – available on the IRS website or using Google Search data. If taxpayers are indeed waiting until April 15th to file their taxes, graphically, this should result in an excess mass in the distribution of filing just before April 15th. And if April 15th filing results in more deductions being forgone, we should observe a larger missing mass in the distribution of itemizers just above the standard deduction threshold for those taxpayers who file just before the April 15th deadline compared to those that file their returns in February or March.

This prediction clearly appears in the data no matter what dataset is being used: a substantial portion of taxpayers tend to systematically bunch at the April 15th deadline. Figure 9a graphs the volume of Google searches for the term *1040* by week, and Figure 9b uses data from irs.gov and plots the number of tax returns filed by week: in both cases, there is a clear spike in the weeks that include April 15th.[5]

More importantly, Figure 9c shows that taxpayers who file close to the deadline tend to forgo more deductions: the missing mass for those that file close to the April 15th deadline is larger than for March filers, consistent with procrastination accounting for a portion of the estimated forgone deductions.

Tax Evasion Could it be that taxpayers believe that itemizers are more likely to be audited than individuals claiming the standard deduction which would then cause them to forgo deductions so as to avoid higher odds of being audited? The probabilities of audit for this share of the population are lower than 1% and are virtually the same for itemizers in the neighborhood of the standard deduction and individuals who claim the standard deduction.[6] In order to shed light on this possibility, consider a taxpayer with a Constant Relative Risk Aversion (CRRA) utility function $U(x) = \frac{1}{1-\theta}x^{1-\theta}$. Denote by p the probability of audit, S the after tax benefit of the standard deduction, T the after tax benefit of itemized deductions, and k the cost imposed by an audit on the taxpayer, which includes both a fixed cost of being audited (collecting receipts and dealing with the IRS) and the penalty that the taxpayer may have to pay. Consider the extreme case in which all charitable deductions are false.[7] Denote by C, the proportion of charitable donations to total deductions. On average, charitable

[5] Link for the IRS data: https://www.irs.gov/uac/2016-and-prior-year-filing-season-statistics.

[6] See Miller et al. (2012) and Slemrod and Gillitzer (2013).

[7] It is very hard for taxpayers to evade other major deductions such as the mortgage interest, state tax and property tax deductions since these are third-party reported.

deductions are equal to 13% of total deductions. Taxpayers evade taxes by reporting $C \times T$ fake deductions. Therefore, if a taxpayer is audited, her deduction level will be brought back to $T(1-C)$ from T and she will incur a cost k of being audited.

The taxpayer will itemize deductions if the expected benefit of itemizing given a probability p of facing an audit is greater than the benefit of claiming the standard deduction:

$$p\left[\frac{1}{1-\theta}(T(1-0.13)-k)^{1-\theta}\right]+(1-p)\left[\frac{1}{1-\theta}(T)^{1-\theta}\right] \geq \frac{1}{1-\theta}S^{1-\theta}. \quad (1)$$

In addition, the taxpayer will switch to the standard deduction if her total deductions reduced by the amount of charitable deductions is smaller than the standard deductions i.e. $(1-C)\times T < S$. Otherwise – if she is afraid of being audited – she can still itemize and only claim her true deductions, she should still be above the standard deduction threshold. This implies that any taxpayer with total deductions $T > \frac{S}{C}$ would not switch to the standard deduction.

This can be calibrated in order to estimate how much of the forgone benefits evasion could explain. The first term of equation 1 is the benefit derived if the taxpayer is audited: she can only deduct the standard deduction $(T-C)$ and incur the cost of itemizing (c) and the cost of evasion (k). It is multiplied by the probability of audit p. The second term is the benefit derived from itemizing: it is equal to the level of deductions T and is multiplied by the probability of not being audited $(1-p)$. Overall, the sum of these two terms is equal to the expected benefit of itemizing. The right hand side of the inequality is the benefit from claiming the standard deduction. To perform the calibration, I solve equation 1 for T, which determines the level of total deductions of a taxpayer who would stop itemizing because of evasion. Assuming a risk aversion coefficient of 1, that an audit would cost half a day of work and that taxpayers correctly perceive audit probabilities, taxpayers would reduce their deductions by 5 to 6 dollars to avoid an audit. If they misperceive audit probabilities and believe they are 20 times what they truly are, they would forgo 102 to 114 dollars to avoid an audit. Overall, this is not consistent with the average compliance cost I estimate.

In addition, if evasion was driving the result we should observe that taxpayers who itemize - even when they are close to the standard deduction threshold – have a low proportion of deductions that are easy to evade. Mortgage interest, state and local tax deductions are hard to evade because they are third-party reported to the IRS. Charitable donations, however, are easy to evade because they are not third-party reported.[8] Figure 10 shows the proportion of charitable donations for taxpayers who are close to the standard deduction threshold and rejects the assumption that taxpayers switch to the standard deduction by reducing their deductions because of a fear of audit.

3.3 Estimating the Costs of Filing Other Schedules and Tax Forms Over Time

Extrapolation Approach The cost I estimate is specific to the taxpayers who are just above the standard deduction. In order to get a cost estimate that is more representative of the rest of the population of taxpayers, I need to estimate the cost for individuals away from the standard deduction threshold. These individuals differ in their demographics, in particular, they are more likely to have higher incomes since deductions and income tend to be positively correlated and we know, from above, that higher income individuals tend to forgo more deductions. For this reason, I need to estimate the effect that these demographics have on cost. I do so for four variables: income, dependents, use of tax preparers and electronic filing.[9]

To estimate the effect of dependents and the use of tax preparers on the cost of itemizing, I apply the procedure outlined in Section 2.2 for the 1988 reform on subsamples of joint filers. I use joint filers to perform this subsample analysis because they represent more than 50% of the population of filers, which is essential to getting enough power when breaking down the

[8] Kleven et al. (2011) show that taxpayers understand that third-party reported deductions are harder to evade and behave accordingly.

[9] Unfortunately, the SOI dataset does not contain information on education or occupation.

main sample into multiple groups. This means that my approach relies on the assumption that these demographics affect filing costs in the same way for different types of filers and for different years. To estimate the effect of tax preparers, I compare the cost for joint filers who use a tax preparer to the the cost for those who do not. I use a similar approach for taxpayers with and without dependents. To estimate the effect of income on the cost of filing, I use the estimates from Section 3.2. Because electronic filing did not exist in 1988, I cannot use the procedure from Section 2.2. Instead, I pool all cross-sections in years in which electronic filing was commonly used – 1998 to 2006; fit a polynomial through the bins that are away from the standard deduction to extrapolate the counterfactual distribution close to the standard deduction; and compare the distribution of electronic filers and paper filers to this counterfactual to assess the size of the missing mass for each group. Formally, I assume that the filing cost is given by the following equation:

$$C = \beta * \{[(\alpha_{efi}\mathbb{1}_{efi} + \alpha_{\overline{efi}}(1 - \mathbb{1}_{efi})] + [\alpha_{prep}\mathbb{1}_{prep} + \alpha_{\overline{prep}}(1 - \mathbb{1}_{prep})] + [\alpha_{dep}\mathbb{1}_{dep} + \alpha_{\overline{dep}}(1 - \mathbb{1}_{dep})] + \sum_{i=0}^{9} \alpha_i \mathbb{1}_i, \quad (2)$$

where β is the baseline cost of itemizing, as estimated in Section 2.2; α_{efi} is the effect of electronic filing on the cost of itemizing; and $\alpha_{\overline{efi}}$ is the effect on the cost of not filing electronically. The remaining variables are defined similarly, with *prep* corresponding to the use of a tax preparer and *dep* having at least one dependent child. Each α_i coefficient corresponds to the effect of income on the cost. These are derived in Section 3.2: each α_i is equal to the ratio of the cost of itemizing for income group i divided by the average cost of itemizing for all groups.

The resulting coefficients are reported in Table 2. Note that the α coefficients provide estimates for both demographic and non-demographic characteristics of taxpayers. On the one hand, the α coefficients for the effect of income on costs, denoted by α_i where i is the income decile, tend to increase monotonically with i, which is consistent with a value of time explanation of the forgone benefits as discussed above. On the other hand, the α coefficients for using a tax preparer or electronic filing are not demographic characteristics but instead a choice made by the taxpayer. Note that the effect on cost of using either electronic filing or a tax preparer are very limited. The main reason for this limited effect is the fact that compliance costs appear to be mostly driven by record keeping rather than the cost of filing forms and returns. And neither electronic filing nor tax preparers can affect record keeping effectively since records have to be kept prior to using electronic filing or visiting a tax preparer. This is what likely explains that the estimated coefficients have so little effect on cost and is consistent with the evidence and discussion from Section 3.2.

Extrapolation to Other Schedules and Tax Forms To infer the cost of filing other schedules, I assume that, holding constant the number of hours spent working on a given tax schedule, taxpayers derive the same disutility from each tax schedule. In other words, they do not dislike filing any particular schedule more than others, as long as each requires the same number of hours. I also assume that the demographics estimated in Section 3.3 affect the cost of other schedules in the same way. The IRS provides estimates of the number of hours required to file each tax schedule based on surveys of taxpayers at the time of filing.[10] I use these survey estimates to scale the cost estimates of other schedules. For example, filing schedule B requires 1 hour and 19 minutes, which is 28% of the total time required to file schedule A; therefore, I assign a baseline cost of filing schedule B of 28% of that of schedule A. The filing cost for each taxpayer is given by an equation similar to equation (2), with a subscript x that corresponds to each tax schedule:

$$C_x = \beta_x * \{[(\alpha_{efi}\mathbb{1}_{efi} + \alpha_{\overline{efi}}(1-\mathbb{1}_{efi})] + [\alpha_{prep}\mathbb{1}_{prep} + \alpha_{\overline{prep}}(1-\mathbb{1}_{prep})] + [\alpha_{dep}\mathbb{1}_{dep} + \alpha_{\overline{dep}}(1-\mathbb{1}_{dep})] + \sum_{i=0}^{9}\alpha_i\mathbb{1}_i\}, \quad (3)$$

[10] According to the IRS survey estimates, the 1040 form requires 9.4 hours, Schedule A 4.5 hours, Sch. B 1.3 hours, Sch. C 9.6 hours, Sch. D 3.8 hours, Sch. E 5.8 hours, Sch. F 16.1 hours and Sch. SE 1.1 hours.

where C_x is the cost of schedule $x = 1040, A, B, C, D, E, F, SE$, and β_x is the baseline cost estimate of schedule x. The coefficients are reported in Table 2.

Filing Costs Have Been Increasing Since the 1980s Using equation (3), I estimate the total cost of filing all federal income tax schedules for every year from 1984 to 2006.[11] Figure 8a shows that costs have been increasing steadily, from \$150bn in 1984 to \$200bn in 2006 (both in 2016 dollars). Part of this increase is mechanically driven by an increase in the number of tax filers. But it is also driven by a steady increase in the number of taxpayers who have to file other schedules in addition to the 1040 form. Some of these schedules require a substantial amount of time to be filed. Figure 8b plots the number of forms filed by schedule over time. The proportion of non-1040 forms filed increased by 15% from 1984 to 2009. While it is often believed that filing costs have decreased over time since the 1980s because of the rapid increase in electronic filing, Figure 8b shows two countervailing forces to electronic filing that drive total costs upwards: the number of individuals who file taxes and the number of scheduled filed by each taxpayer. My estimates suggest the upward pressure on filing costs exerted by these two forces outweigh the cost savings of electronic filing. The number of non-1040 forms filed can be easily reduced by increasing the filing thresholds for their corresponding schedules. The cost savings from increasing these thresholds would need to be weighed against the effect they would have on increasing evasion. Tazhitdinova (2018) investigates these trade-offs in the setting of non-cash charitable donations in the U.S., providing evidence of the effectiveness of reporting requirements against evasion and the costliness of these requirements to taxpayers.

Note that, in order to construct Figures 8a and 8b, I simply use the number of forms filled by each taxpayer, which is information included in the SOI dataset. And as mentioned above, part of the increase in costs is mechanically driven by more and more taxpayers having to file more and more forms and schedules. This, in part, is due to taxpayers having income from more varied sources, not only wage income but also capital gain income, dividends,

[11] I start in 1984 because prior years are missing information on Schedule SE.

etc. but part of this is also due to the thresholds for having to file separate schedules being fixed in nominal terms. For example, Schedule B must be filed when taxable interest income or dividend income exceeds $400 and this limit has been fixed nominally for several decades, mechanically increasing the number of taxpayers having to file Schedule B. While differentiating between the two explanations is difficult with the current dataset, it could be possible with more detailed data.

4 Assessing Policy Proposals to Reduce the Cost of Filing Tax Returns

This section considers four different policy proposals to reduce tax filing costs and assesses their likely effects.

4.1 Pre-Populated Forms

One policy proposal that is believed to be one of the more cost-effective ways of reducing compliance costs is to pre-populate tax forms with information that tax authorities already have access to from third-party information reports sent to the IRS. Third-party reporting is the norm for wage earners, for example, whose earnings are reported by employers to the employee and to the IRS on the W2 form. This form also contains other pieces of information that can be used to pre-populate other tax return fields: such as the amount of state taxes paid, which can pre-populate part of Schedule A. Similarly, mortgage interest is also reported by lenders to borrowers and to the IRS and can be pre-populated. Finally, banks and brokers report some of their clients' transactions to the IRS via the 1099 and 1098 forms. There are also proposals to have counties report property taxes paid by home owners to the IRS.

There are two versions of this policy proposal. The first version proposes to only pre-populate forms with information the IRS already has access to using the W2, 1098 forms, etc. The second proposal expands informational reporting and requires more institutions to disclose information to the IRS. While expanding third-party reporting could result in large time savings for taxpayers, and could also reduce evasion substantially, it has the disadvantage of imposing a potentially heavy administrative burden on reporting institutions. And while this burden is likely to be small for larger firms, that already have a well-established informational system, it can be prohibitive for smaller firms. Harju et al. (2019), for example, show that some firms would reduce their activity just to avoid compliance costs. For this reason, we restrict our evaluation to the first policy proposal, but it should be noted that large savings, potentially at a low cost, can be achieved by requiring entities with well established informational systems to report information to the IRS. These entities may include counties, large charitable organizations, and banks.

There is one possible downside from pre-populating forms with information that is already available to the IRS, which has often been used to advocate against such a policy and is a similar argument used by Milton Friedman against tax withholding. According to Friedman and others, one would ideally want to make taxes as salient as possible and force taxpayers to work through all forms in detail in order for them to clearly understand the burden they are bearing and also so that they can reduce this burden as much as possible. While theoretically sound, empirically there is little basis for such argument. In fact Gillitzer and Skov (2018) explores this tradeoff in case of the introduction of third-party reporting of charitable donations in Denmark. They find that third-party reporting *reduces* the tax burden of Danish taxpayers. As forms are pre-populated, taxpayers end up with *more* charitable deductions claimed, simply because prior to forms being pre-populated, there were several charitable deductions they did not bother with. And while it is true that taxpayers are less likely to pay attention to their tax returns if they are already pre-populated, some may argue that the current forms are so complicated that few taxpayers engage with them and instead outsource them to tax preparers or use tax softwares.

In order to estimate the potential savings from pre-populating tax forms with the information that the IRS currently has, I classified every field in the 1040 form and also in every tax schedule into whether it could currently be pre-populated or not. I then used the proportion of fields that can be pre-populated to deflate the current compliance costs for each of the forms and schedules by that proportion. For example, I estimate that 70% of the fields in Schedule A can be pre-populated, which would reduce the compliance costs of filing Schedule A from $24 to $7 billion, if those fields are pre-populated. Of course, this simple approach relies on potentially strong assumptions. The first one is that pre-populating these forms is cost-free for the IRS. While this is likely to be true at the margin, especially for electronically filed forms, it will certainly involve substantial one-time fixed costs needed to implement a system that would automatically pre-populate forms. Second, it is likely that the reduction in record keeping costs is heterogenous: here I assume a one-to-one relationship between the reduction in fields needed to be filled and record-keeping costs. In reality, some fields might require higher record-keeping costs than others especially the fields that are not third-party reported. For example, since all mortgage interest is reported on one single form, which also has the advantage of being sent very close to the filing deadline, the record-keeping costs for this field is likely to be low. Whereas charitable donations appear in several different receipts, one for each donation, and are likely to be harder to keep track of and require larger record-keeping costs. While this is the best available estimate of pre-populating forms, conducting surveys aimed specifically at estimating record-keeping costs for each category could help refine this estimate further.

Overall, and according to these estimates, pre-populating forms reduce compliance costs by 43%. This reduction mostly comes from large time savings from the time required to file the 1040 form for individuals who cannot file the 1040A or 1040EZ forms but also from Schedule A and to a lesser extent from Schedule C.

Finally, it is worth emphasizing that this policy comes with a "double dividend", which is not accounted for in the estimations above: it allows compliance costs savings for taxpayers, but can also substantially reduce tax evasion. Pomeranz (2015) and Kleven et al. (2011) have shown that some of the most effective ways to reduce evasion are to require third-party reporting. Therefore one desirable consequence of such policies would be a reduction in the tax gap.

4.2 Free Electronic Filing Options for All Taxpayers

Another policy proposal, which is likely to be less costly than pre-populating forms but also less beneficial in terms of compliance costs reduction is to offer a free electronic filing option for taxpayers. According to the latest statistics published by the IRS for the 2020 filing season, 98% of taxpayers (114 million of them) either file their returns electronically or use a tax preparer. While there are free electronic filing options for taxpayers who have simple enough returns and low enough earnings, which represents approximately 37% of the population of taxpayers, these options are not available for the remaining of the population of taxpayers.

Here, I assess how much money US taxpayers would save by having access to a free electronic filing option. Similarly to the previous policy proposal, I do not assess how much it would cost the IRS to offer such a free service. However, similarly to pre-populating forms, this is likely to be a one time fixed expense with some small future maintenance costs. Instead, I simply estimate how beneficial the free electronic filing would be to taxpayers. And similarly to the previous analysis, the estimates presented here rely on several important assumptions which I lay out below.

In contrast to the analysis of the pre-populated forms, these estimates rely on using 2020 filing statistics from the IRS, which are publicly available on the IRS website. The SOI dataset is not needed because estimating the cost of a free electronic filing option only requires to know how many of each schedule and form are filed by taxpayers which does not require the underlying micro data.[12] All is left, in order to estimate how much money would be saved from offering a free electronic filing option, is to know how much do the electronic filing and tax preparer services cost. Unfortunately, the IRS filing statistics do not distinguish

[12] Link to the IRS filing statistics: https://www.irs.gov/statistics/filing-season-statistics

between taxpayers who electronically file because they use tax software versus taxpayers who are counted as electronically filing because they use the services of a tax preparer who in turn files electronically using their own tax software. For this reason, and in order to be more conservative, since tax preparer services are often substantially more expensive than using a tax software, I assume that all taxpayers who are counted as filing electronically use a tax software.

I use the prices charged by Intuit's TurboTax in order to estimate how much it costs taxpayers to use their software. TurboTax offers four pricing categories: (1) a free option available for taxpayers who either file a 1040EZ or a 1040A form and have W2 earnings that are less than $100,000, do not file any Schedule, do not have 1099 income and are not self employed; (2) a $60 option for taxpayers who either have W2 income in excess of $100,000 and/or itemize their deductions; (3) an $80 option for taxpayers who have substantial investment income and/or investment rental properties; and (4) a $120 option for self-employed taxpayers. TurboTax is the most commonly used tax software but other software packages offer similar prices.

Using these prices and the number of taxpayers who belong to each of the corresponding categories, I estimate that $5.1 billion have been spent in 2020 on tax preparation. It is worth emphasizing that this figure ignores any time or disutility cost that a taxpayer might have to bear in order to file these taxes and only accounts for filling out the required forms assuming all forms and receipts have been collected prior to it, and hence also ignores any record keeping costs.

4.3 Shortening the 1040 form

A recent policy proposal has been to shorten the 1040 form, which was implemented in the Tax Cuts and Jobs Act reform of 2017. This policy shortened the 1040 form to the size of a "postcard" and eliminated the 1040A and 1040EZ forms. Practically, since 98% of taxpayers either rely on tax software or tax preparers, there is no reason to expect that any of these taxpayers would save time or money because of the "postcard" size of the 1040 form. For the remaining 2% of taxpayers who file their returns on paper, it is unclear how their costs have changed: on the one hand, there are fewer lines on the 1040 "postcard" form but some of these lines were either going to be removed because of some of the changes imposed by the TCJA 2017 and would have been dropped even in the old 1040 forms or some of these lines were moved to other schedules therefore requiring taxpayers to spend more time working these schedules. Therefore, this policy, at best kept filing costs constant and at worst may have increased them for the 2% of the population that files returns by hand.

5 Conclusion and Discussions

In this paper, I show that compliance costs of taxation are large and have been increasing steadily over time. This paper discusses two policy proposals, namely pre populating forms and offering a free electronic filing option to all taxpayers, that can help alleviate these costs or even eliminate them. While we do not have a precise estimate of how costly these solutions are, we know that most OECD countries have adopted them. Even some US States, such as California, offer a free electronic filing to most taxpayers. Given their prevalence it is reasonable to assume that they are likely low cost enough to warrant implementing them in order to save taxpayers money, time and effort that could be used in more productive ways.

References

Benzarti, Y. (2020). How Taxing is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. *American Economic Journal: Economic Policy*.

Blumenthal, M. and J. Slemrod (1992). The compliance cost of the us individual income tax system. *National Tax Journal*, 185–202.

Gillitzer, C. and P. E. Skov (2018). The use of third-party information reporting for tax deductions: evidence and implications from charitable deductions in denmark. *Oxford Economic Papers 70*(3), 892–916.

Guyton, J., J. O'Hare, M. Stavrianos, and E. Toder (2003). Estimating the compliance cost of the US individual income tax. *National Tax Journal*, 673 –688.

Harju, J., T. Matikka, and T. Rauhanen (2019). Compliance costs vs. tax incentives: Why do entrepreneurs respond to size-based regulations? *Journal of Public Economics 173*, 139–164.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 651–692.

Miller, S., R. Marcuss, S. Boehmer, B. Johnson, and W. Kei (2012). Internal revenue service data book. Annual report.

Piketty, T. and E. Saez (2003). Income inequality in the united states, 1913–1998. *The Quarterly journal of economics 118*(1), 1–41.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review 105*(8), 2539–69.

Slemrod, J. and J. Bakija (2008). Taxing Ourselves: a Citizen's Guide to the Debate Over Taxes. *MIT Press Books*.

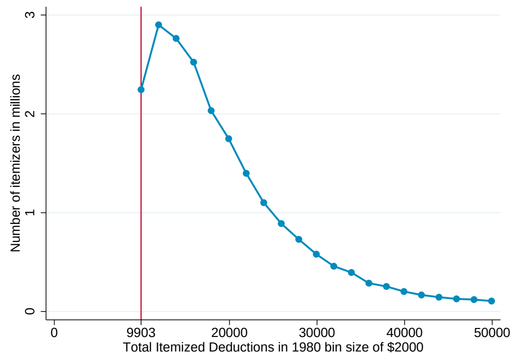
Slemrod, J. and C. Gillitzer (2013). *Tax Systems*. MIT Press.

Slemrod, J. and N. Sorum (1984). The Compliance Cost of the US Individual Income Tax System. *National Tax Journal*, 461 –474.

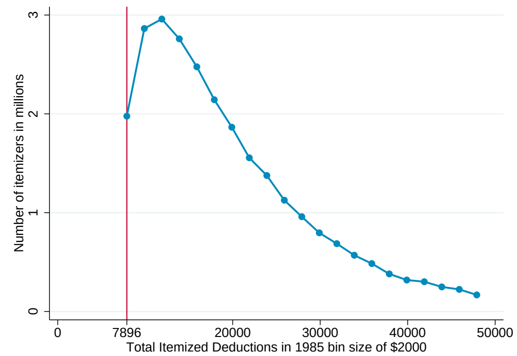
Tazhitdinova, A. (2018). Reducing evasion through self-reporting: Evidence from charitable contributions. *Journal of Public Economics 165*, 31–47.

Van Rooij, M., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics 101*(2), 449–472.

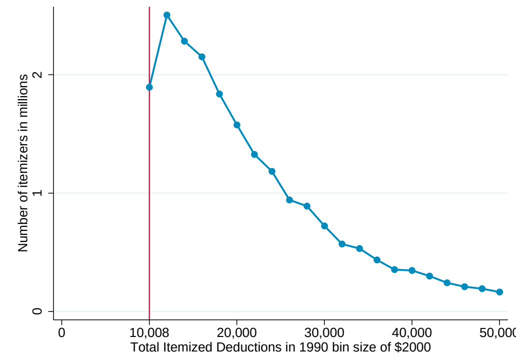
Figure 1: Missing Mass Just Above the Standard Deduction



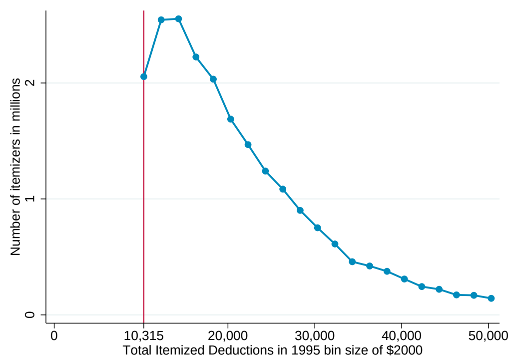
(a) 1980



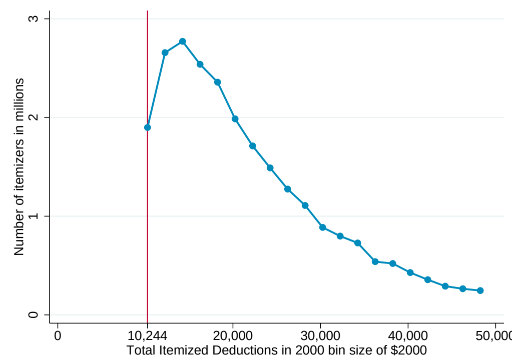
(b) 1985



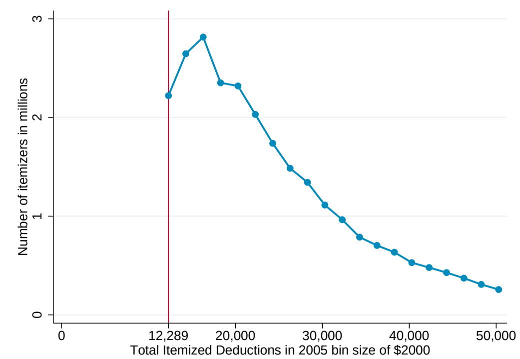
(c) 1990



(d) 1995



(e) 2000

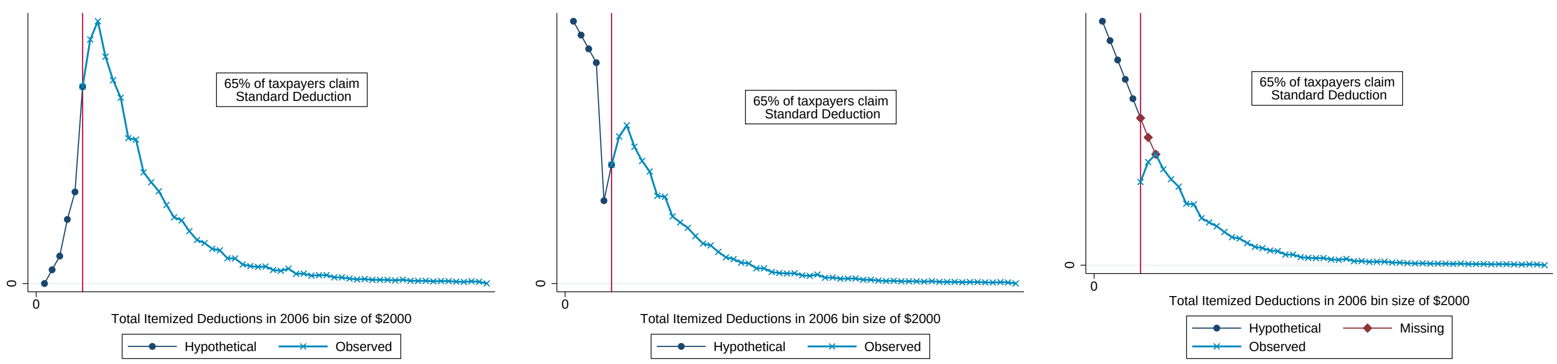


(f) 2005

Notes: The figures above plot the density of deductions for itemizers filing jointly. The bin size is $2,000 and the vertical line represents the standard deduction threshold for each year. Notice the missing mass just above the standard deduction threshold. Additional years can be found in Benzarti (2020).

Figure 2: Different Scenarios Below the Standard Deduction

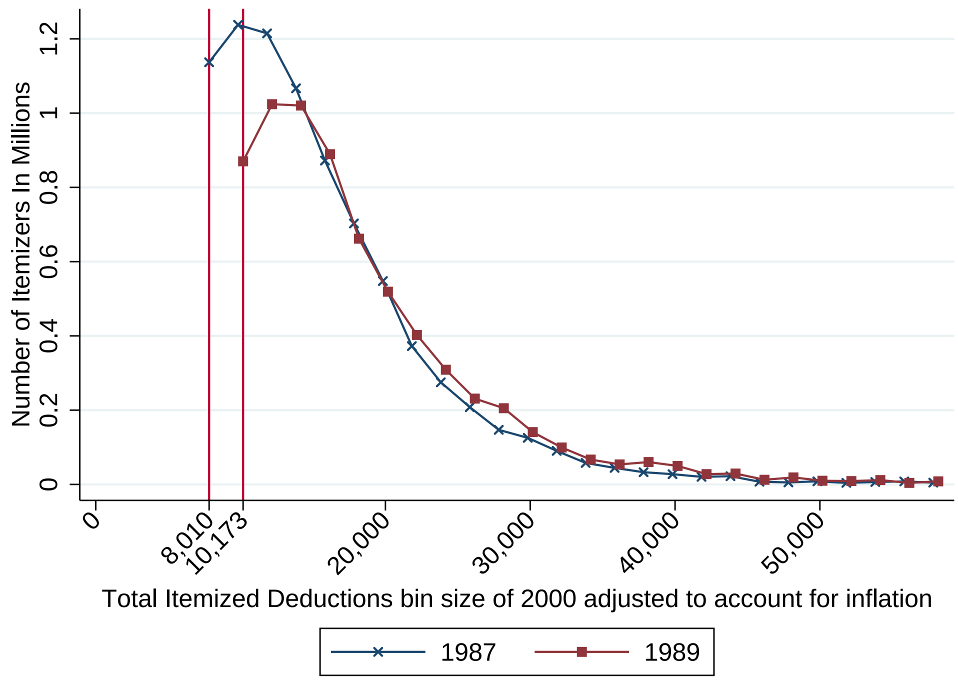
(a) Increasing: Impossible **(b) Double Peaked: Unlikely** **(c) Missing Mass**



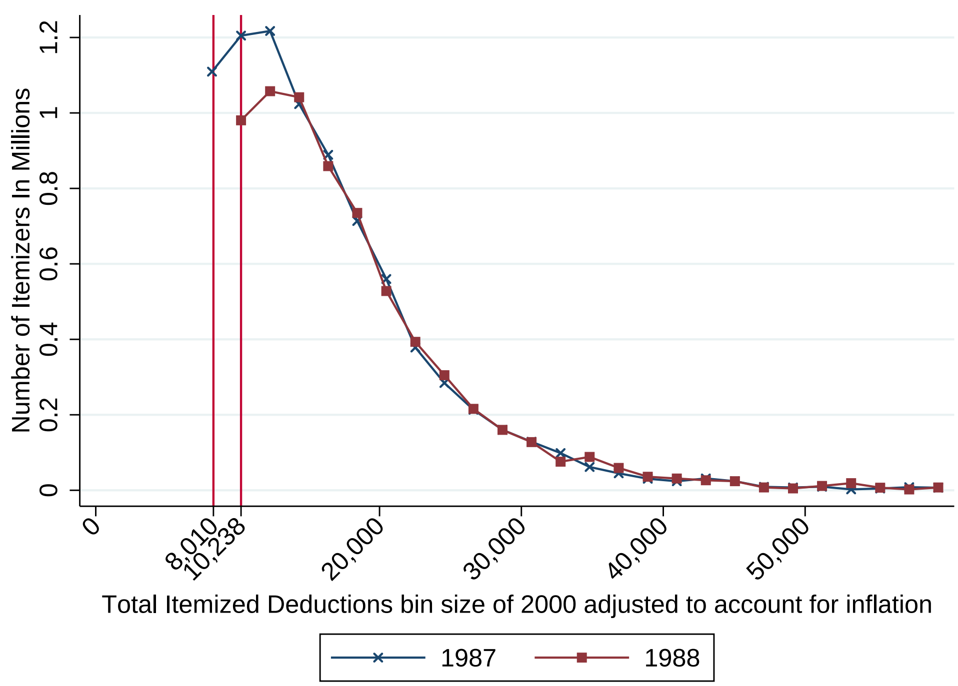
Notes: The graphs above plot the different scenarios that could be happening below the standard deduction. Graph (a) assumes that the density is strictly increasing, which is impossible given that 65% of taxpayers claim the standard deduction. This scenario would fail to account for most of the population of taxpayers. Graph (b) accounts for most of the population and is continuous at the standard deduction but the density is double peaked. This is possible but unlikely given that densities are usually single peaked. This however does not rule out densities that are double-peaked *because of the standard deduction*. Graph (c) assumes that there is a discontinuity at the standard deduction threshold because of compliance costs creating a missing mass.

Figure 3: Density of Deductions for Itemizers Filing Jointly Pre and Post Reform

(a) 1987-1989 Comparison

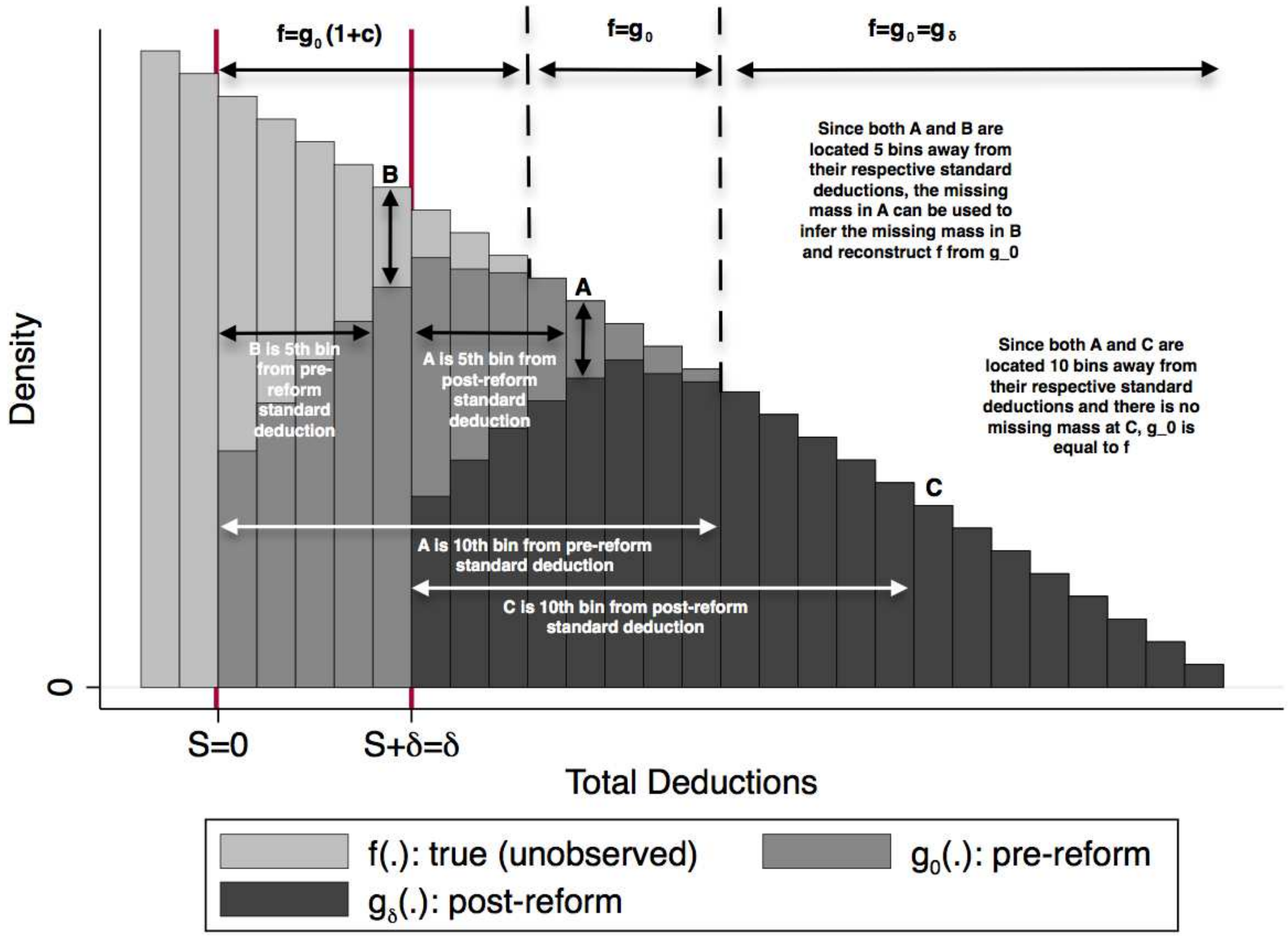


(b) 1987-1988 Comparison



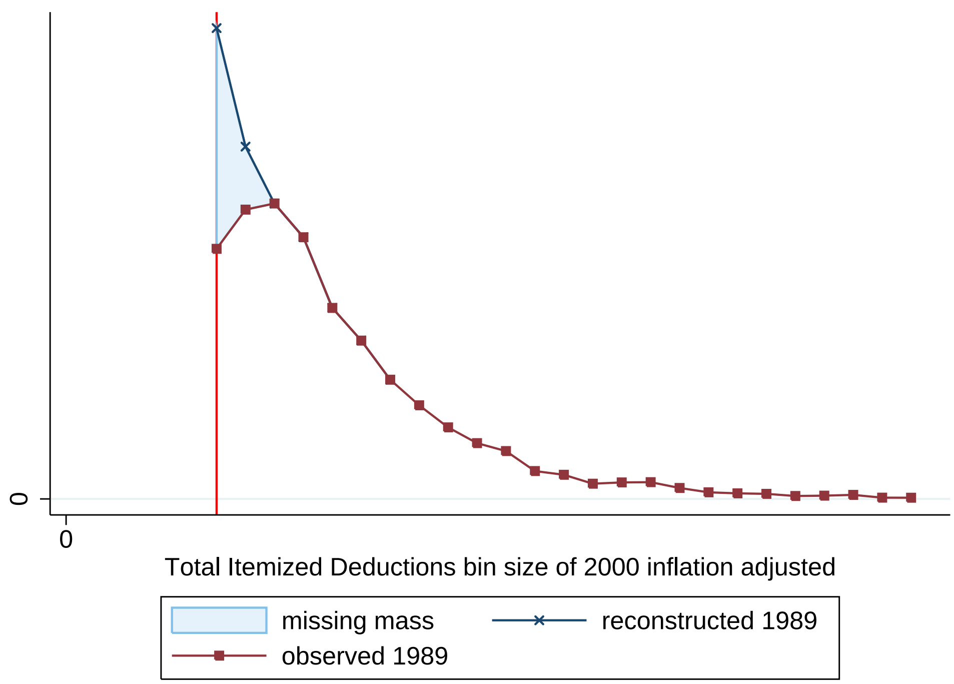
Notes: The first graph plots the density of itemizers one year before and one year after the standard deduction reform while the second one plots these densities one year before and during the reform.

Figure 4: Reconstruction of the Counterfactual Density



Notes: This graph illustrates the method used in section 2.2 to reconstruct the counterfactual density of itemizers $f(\cdot)$ using the pre- and post-reform densities $g_0(\cdot)$ and $g_\delta(\cdot)$.

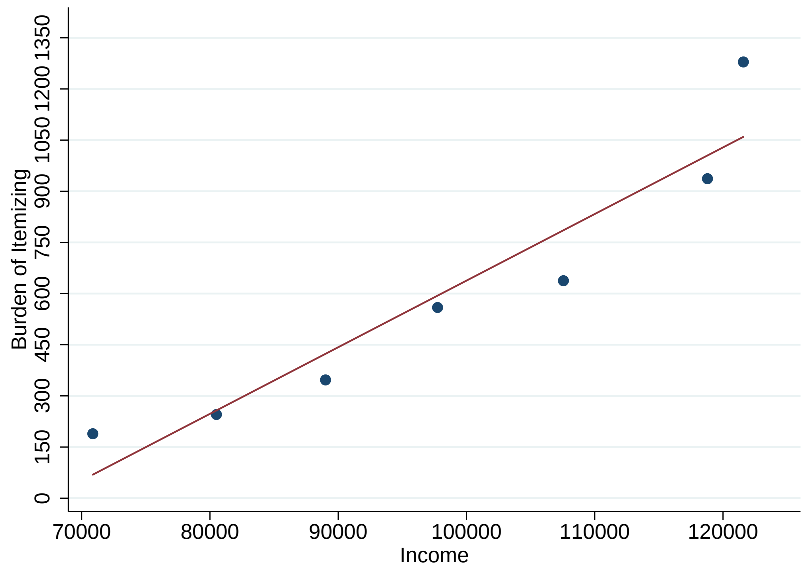
Figure 5: Reconstructed Counterfactual



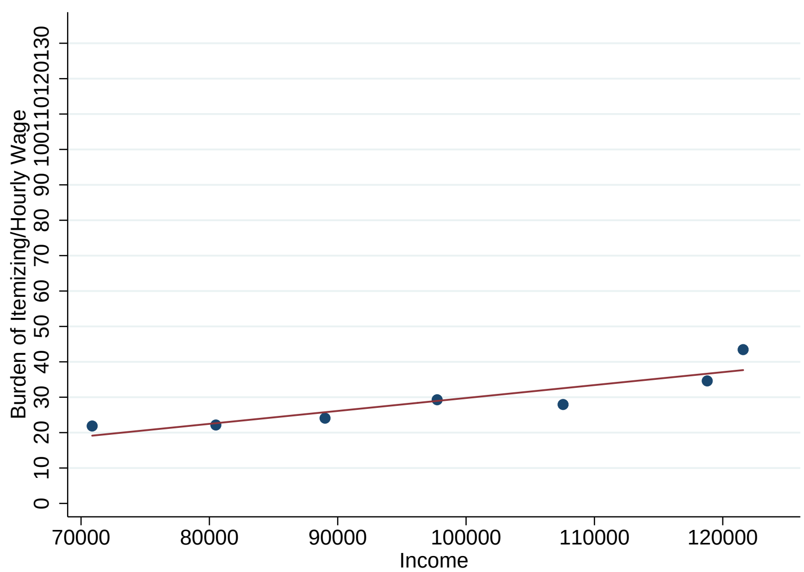
Notes: This graph plots the reconstructed counterfactual in 1989 using the method outlined in section 2.2 and the observed density for 1989. The missing mass used to estimate the burden of itemizing is given by the area lying between the two curves.

Figure 6: Relationship Between Income and the Burden of Itemizing Deductions

(a) Burden of Itemizing and Income



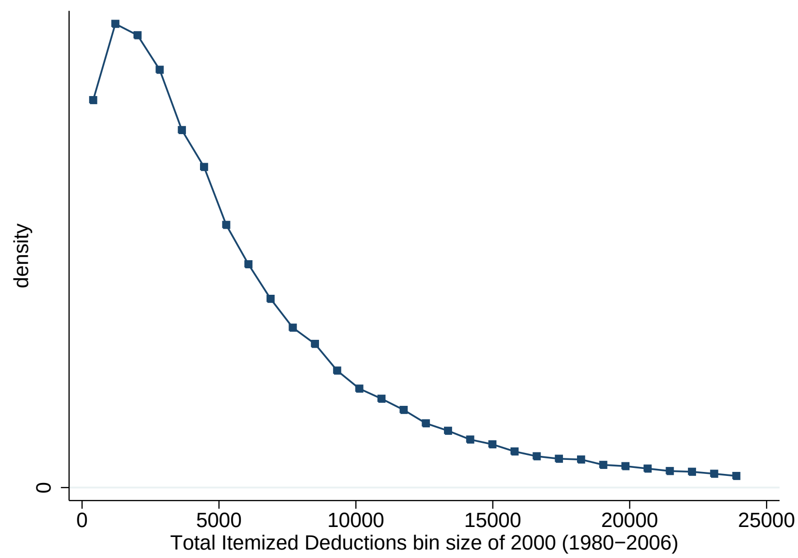
(b) Hours Spent Itemizing and Income



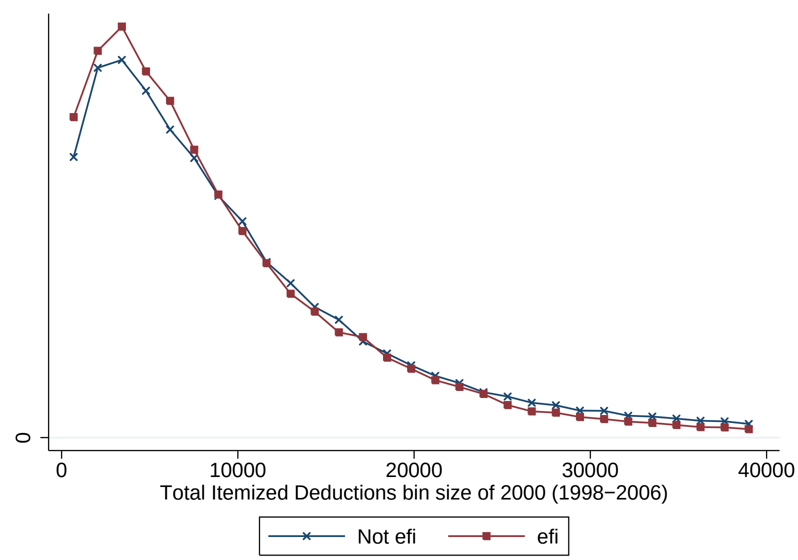
Notes: (a) The first graph shows the increasing relationship between income and the burden of itemizing: richer households are more likely to forgo deductions. This relationship controls for the variation in MTR across the different income groups. (b) The second graph divides the burden of itemizing by the hourly wage and shows the implied hours spent itemizing by each income group. The y-axes of each graphs have scales of the same magnitude.

Figure 7: Use of Tax Preparer and Electronic Filing

(a) Tax Preparer



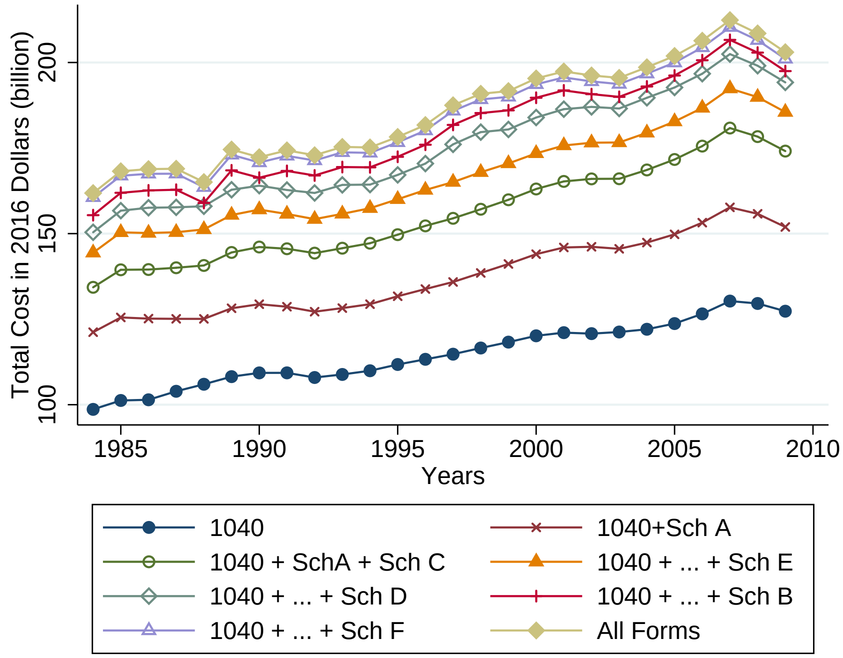
(b) Electronic Filing



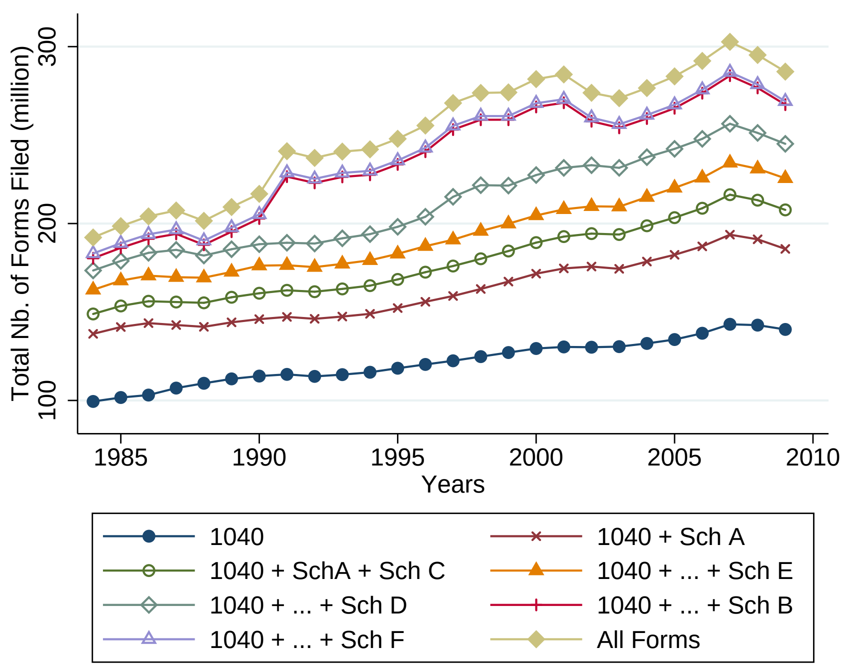
Notes: The x-axis is normalized such that 0 corresponds to the standard deduction threshold. Graph (a) plots the density of total deductions for taxpayers who use tax preparers from 1980 to 2006 by bin size of $2000. Graph (b) plots the density of total deductions for taxpayers who file returns electronically from 1998 to 2006 by bin size of $2000 and compares it to the density of taxpayers who do not file returns electronically. Both graphs exhibit a missing mass close to the standard deduction implying that neither tax preparers nor electronic filing eliminate the burden of itemizing. The use of electronic filing slightly reduces the missing mass consistent with compliance costs being the driver of the missing mass and record-keeping being the largest portion of the cost of itemizing.

Figure 8: Cost Trends

(a) Total Costs



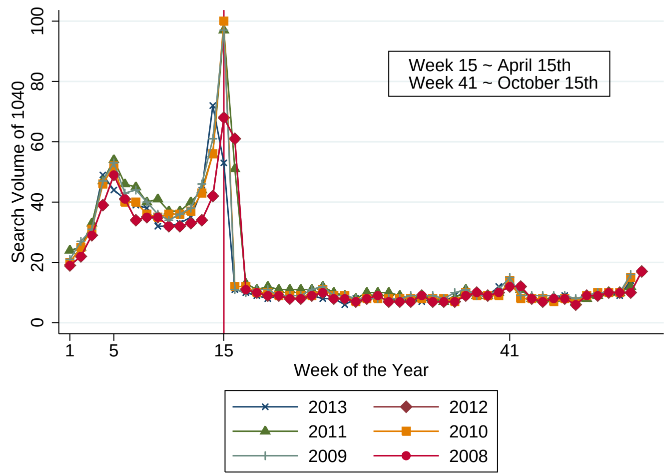
(b) Number of Forms Filed



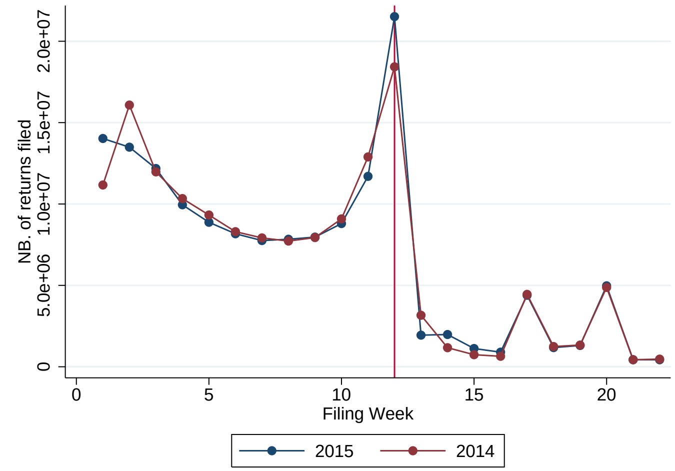
Notes: The first Figure plots the cost of filing each schedule for the total US population over time as estimated in equation (3). The second figure plots the total number of forms filed over time. Each curve is cumulative: it incrementally adds each schedule to the previous curve.

Figure 9: Deadline Effects

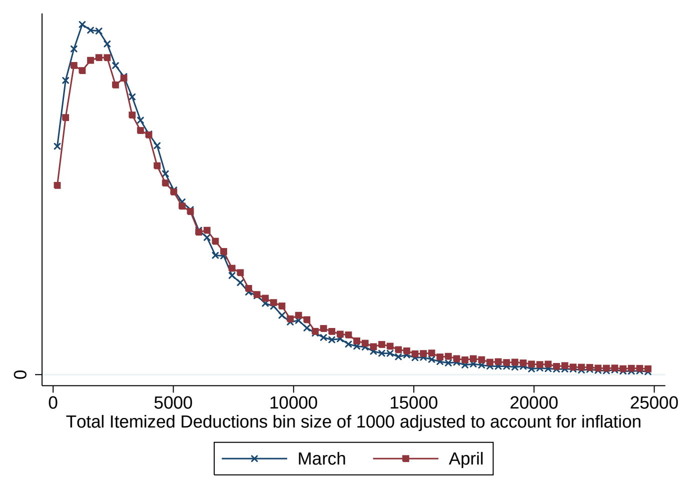
(a) Google Search of the Term 1040



(b) Number of Returns Filed by Week

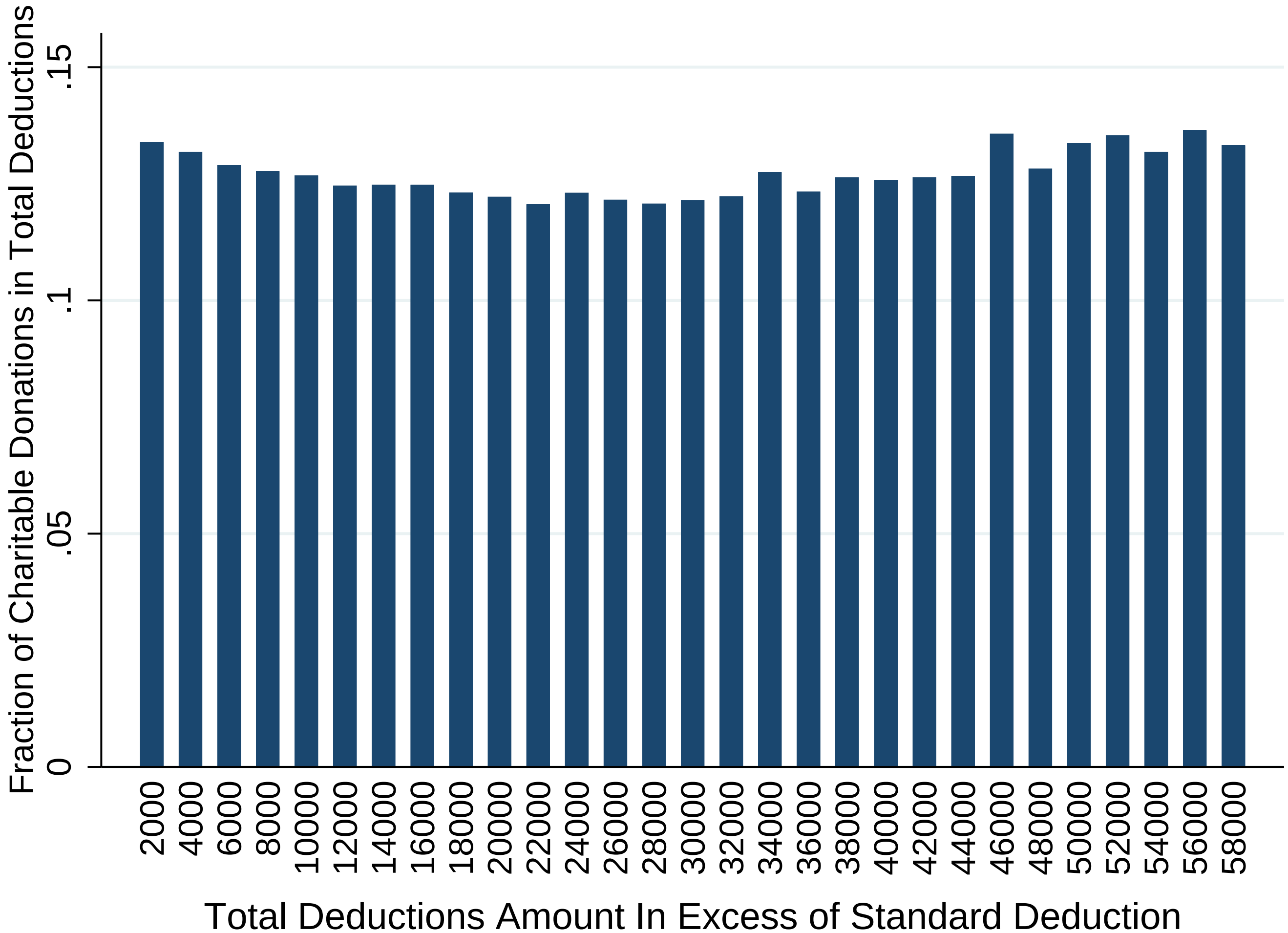


(c) March Itemizers v.s. April Itemizers



Notes: Panel (a) plots the volume of search of the term "*1040*" in Google and panel (b) plots the volume of tax returns filed by week in 2014 and 2015. The red vertical line corresponds to the week of April 15. Panel (c) plots the density of itemizers who file in March versus in April, the x-axis is normalized such that 0 corresponds to the standard deduction. Panel (d) plots the average week in which a return is processed in year t on the y-axis and the average week in which a return is processed in year $t-1$ on the x-axis.

Figure 10: Fraction of Charitable Donations in Itemized Deductions by Size of Total Deductions



Notes: This graph shows the proportion of deductions that are charitable donations for itemizers pooling all years from 1980 to 2006 by their distance to the standard deduction. Deductions are adjusted for inflation and the standard deduction amount is subtracted from them to calculate the distance to the standard deduction. The proportion of charitable donations does not change close to the standard deduction threshold implying that taxpayers do not respond to the change in the standard deduction by reducing their charitable donations. This rules out the explanations of the missing mass based on the behavioral response to a concave kink point and evasion.

Table 1: Cost Estimates

Filing Status	MTR	Cost as % of AGI	Cost in $
Single	15%	0.74	$175
Single	28%	0.85	$369
Joint	15%	0.57	$242
Joint	28%	0.74	$591
Head	15%	0.76	$270
Head	28%	0.72	$458

This Table shows the cost of itemizing as estimated in section 2.2 for different brackets and filing status.

Table 2: Cost Variables

Variable	Cost
β_A, Single, τ=15%	$175
β_A, Single, τ=28%	$369
β_A, Joint, τ=15%	$195
β_A, Joint, τ=28%	$591
β_A, Head, τ=15%	$270
β_A, Head, τ=28%	$458
β_{1040}	$2.08\beta_A$
β_B	$0.28\beta_A$
β_C	$2.13\beta_A$
β_D	$0.83\beta_A$
β_E	$1.29\beta_A$
β_F	$3.55\beta_A$
β_{SE}	$0.25\beta_A$
$\alpha_{\overline{dep}}$	0.99
α_{dep}	1.01
$\alpha_{\overline{efi}}$	1.07
α_{efi}	0.93
$\alpha_{\overline{prep}}$	0.99
α_{prep}	1.01
α_1	0.21
α_2	0.21
α_3	0.27
α_4	0.38
α_5	0.61
α_6	0.99
α_7	1.4
α_8	1.74
α_9	2.7

Notes: This table shows the estimates used in equations (2) and (3). β_A is estimated in section 2.2. β_{1040}, β_B, β_C, β_D, β_E, β_F and β_{SE} are estimated in section 3.3. Section 3.3 explains how α_{dep}, α_{efi} and α_{prep} are estimated. Section 3.2 explains how α_i, $i = 1, ..., 10$ are estimated.