How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs.

Youssef Benzarti*
UCLA and NBER

Abstract

This paper uses a quasi-experimental design to estimate the cost of filing taxes. First, using US tax returns, I observe how taxpayers choose between itemizing deductions and claiming the standard deduction. Taxpayers forgo tax savings to avoid compliance costs, which provides a revealed preference estimate of such costs. I show that costs increase with income, consistent with an opportunity cost of time explanation. Second, I estimate the cost of filing federal income taxes. I find that this cost is substantially larger than previously estimated, has been increasing since the 1980s and has reached 1.2% of GDP in recent years.
JEL Classification: H24, H31, H83
Keywords: Public Finance, Taxes, Compliance Costs.

*UCLA and NBER (benzarti@econ.ucla.edu). I thank Alan Auerbach, Stefano DellaVigna, Matthew Rabin and Emmanuel Saez for invaluable advice, guidance, and encouragement throughout this project. I benefited from discussions with Miguel Almunia, Pierre Bachas, Dan Benjamin, Kimberly Clausing, Erik Eyster, Alex Gelber, Daniel Gross, Ben Handel, Nathaniel Hendren, Hilary Hoynes, Emiliano Huet-Vaughn, Damon Jones, Louis Kaplow, Marc Kaufmann, Henrik Kleven, Wojciech Kopczuk, Laurence Kotlikoff, David Laibson, Etienne Lehmann, Attila Lindner, Adriana Lleras-Muney, Takeshi Murooka, Michaela Pagel, Colin Raymond, Alex Rees-Jones, Antonio Rosato, Jesse Rothstein, Josh Schwartzstein, Dan Silverman, Joel Slemrod, Charles Sprenger, Justin Sydnor, Alisa Tazhitdinova, Danny Yagan and seminar participants at the Tax Systems Conference (Oxford and U. of Michigan), Paris II, CREST, Aarhus University, SITE Psychology and Economics, the US Department of Treasury, Princeton, Harvard KSG, UCLA, McMaster, U. Mass Amherst and U. of Toronto. Funding from the Robert D. Burch Center for Tax Policy and Public Finance is gratefully acknowledged.

Income taxes represent the largest source of tax revenue in the United States. Today, 8.8% of GDP is transferred from individuals to the Federal Government through income taxes. While an extensive literature documents the efficiency cost of taxation, we know less about the cost of collecting taxes. But every year, 140 million taxpayers spend numerous hours gathering receipts and statements, filling out various tax schedules and forms and submitting them to the Internal Revenue Service (IRS). A large literature documents that individuals frequently leave "money on the table" in other domains because of transactional costs, which suggests that the compliance costs of taxation are likely to be very large.[1]

This paper provides the first estimate of this cost using quasi-experimental methods. I exploit the fact that taxpayers can choose between itemizing their tax deductions or claiming the standard deduction. Itemizing deductions requires some effort cost but can provide significant tax savings. Claiming the standard deduction saves time and effort but results in more taxes due.

With compliance costs, itemizing is beneficial only if it reduces the tax bill by more than the cost of itemizing. This implies that if compliance costs are non-zero, some taxpayers will claim the standard deduction, even though the sum of their deductions exceeds the amount of the standard deduction. The main identification challenge is to differentiate individuals who choose not to itemize because of compliance costs from those who claim the standard deduction because their total deductions are smaller than the standard deduction. This is particularly difficult because taxpayers who claim the standard deduction are not required to report their deductions, implying that their true level of deductions is not observable in the tax data. To identify the cost of itemizing, I proceed in the following way. If individuals forgo tax benefits because of compliance costs, there should be a missing mass in the density of deductions just above the standard deduction threshold. I test this hypothesis by graphing the density of deductions for the years 1980 to 2006 using a stratified random sample of US tax returns, weighted to be representative of the population of itemizers. The shape of the density function suggests the missing mass just above the standard deduction.

To confirm that this shape is due to taxpayers responding to the standard

[1]See, for example, Currie (2006), Bertrand et al. (2006) and, more recently, Bhargava and Manoli (2015).

deduction, I turn to a quasi-experimental design. Following a large increase in the standard deduction, I observe a drop in the mass of itemizers just above the post-reform standard deduction threshold. The post-reform density is systematically lower than the pre-reform density just above the post-reform standard deduction threshold, and the two densities overlap further away from the standard deduction. I ensure that no other reforms are affecting the densities of itemized deductions.[2] I use the missing mass to construct the distribution of forgone benefits. While related to bunching estimators, my approach is different: bunching estimators rely on one cross-section of data, while my approaches compares two cross-sections before and after a reform.[3]

I find that the cost of itemizing ranges from 0.6% to 0.8% of annual income, i.e., the disutility derived from itemizing is equivalent to working 12 to 16 hours, which is substantially larger than previous estimates. I find significant heterogeneity among taxpayers. Some taxpayers still itemize even when savings are modest, and some forgo large tax benefits, resulting in a large average cost of itemizing.

If individuals switch to the standard deduction because they value their time more than the benefits they can derive from itemizing, richer households should forgo more tax benefits than poorer ones. To test this hypothesis, I break down individuals by income deciles and repeat the estimation using the same identification strategy outlined above. The results show an increasing relationship between forgone tax benefits and income - while controlling for the marginal tax rate - consistent with the hypothesis that tax filing imposes a higher cost on richer individuals because they have a higher marginal time value.

The missing mass just above the standard deduction is consistent with taxpayers forgoing benefits to avoid the cost of itemizing. However, there are three alternative explanations for the missing mass. The first is that the standard deduction acts as a concave kink point, effectively changing the price of a deduction. Behavioral responses to concave kink points predict that taxpayers will respond to variations in marginal tax rates but should not respond to variations in in-

[2]My estimates are not affected by the Alternative Minimum Tax, variation in marginal tax rates and the phase out of the personal interest deduction in 1987.

[3]My approach is more related in spirit to a difference-in-differences approach where treated individuals are just above the standard deduction threshold while the control group are those individuals far above the standard deduction threshold.

come while holding the marginal tax rate fixed. The fact that forgone benefits increase with income - while controlling for the marginal tax rate - supports the compliance costs explanation. A second alternative explanation for the missing mass is that some taxpayers mistakenly believe that IRS audits are more likely when itemizing and, thus, switch to the standard deduction to avoid the expected cost of an audit. To assess this explanation, I conduct a survey of taxpayers to elicit their beliefs about audit probabilities and audit costs. I find that the perceived expected cost of audits would explain, at most, one fifth of the cost. A third alternative explanation is that the uncertainty that taxpayers face over the amount of deductions they can claim drives them to not itemize. The cost that I estimate is based on taxpayers who itemized the year before the reform; their deductions are stable over time, which implies that taxpayers should have a small uncertainty range over their level of total deductions. I show that for this theory to explain the result, the uncertainty range would need to be extremely large ($\pm\$14,000$).

I use the cost of itemizing to estimate the total cost of filing federal income taxes.[4] I find that total filing costs are significantly larger than previously estimated from surveys. This cost has been steadily increasing since the 1980s and has reached more than \$200 billion ($\simeq 1.2\%$ of GDP) in recent years.[5] This increase has occurred in spite of the rise in the number of electronic filers over time. It can be explained by growth in the total number of filers – which has been faster than the growth of the US population – and an increase in the number of additional schedules that taxpayers have to file as a supplement to their 1040 form. Along with the ever increasing complexity of the tax code – growing from 26,300 pages in 1984 to 74,608 in 2016 – the increase in compliance costs call for a much needed simplification of the tax code.

While the large magnitude of the costs could be explained by high levels of aversion to filing taxes, I gather empirical evidence suggesting that taxpayers procrastinate on filing their taxes, which leads them to incur high costs. Pro-

[4]Filing costs include both the cost of filling out forms, record keeping, learning about the law and sending documents to the IRS. I estimate the cost of filing the 1040 form, Schedule A, Schedule B, Schedule C, Schedule D, Schedule E, Schedule F and Schedule SE. I do not have information on Schedule R.

[5]See Slemrod (1989) for an example of a survey-based estimate of compliance costs. Other similar references are listed in Table I.9.

crastination provides two testable predictions: first, procrastinators will delay filing until the deadline; and, second, taxpayers who file close to the deadline will forgo more deductions. I provide empirical evidence consistent with both predictions and show that late filing is a persistent behavior, confirming that it is a systematic bias.

This paper is related to several prior work. It is the first and only paper to provide estimates of the cost of filing taxes using a quasi-experimental design. The most closely related paper is Pitt and Slemrod (1989): estimating the cost of itemizing using a censored model with unobserved censoring thresholds using maximum likelihood, they find a smaller cost of itemizing of $107, which is equivalent to 0.12% of AGI and is 5 to 7 times smaller than my cost estimates. They use estimators from Gronau (1973) and Nelson (1977) to address the fact that the distribution below the standard deduction is unobservable. While our approaches are related, my method is able to provide a reduced-form demonstration of the existence of compliance costs without relying on a structural model. Although both papers require assumptions to estimate the magnitude of the compliance costs, my paper requires fewer. Understanding why their structural estimation yields significantly smaller estimates is an important question left for future research. I discuss their approach and some of their assumptions in more detail in Appendix section A. There is also a literature that uses survey evidence to estimate compliance costs. Although informative of the time spent filing taxes, it does not capture the preferences of taxpayers and, in particular, any aversion to filing taxes or any behavioral biases. That literature also suffers from the usual biases of surveys, including high attrition rates and measurement errors.[6] This paper is also related to a literature that estimates the effect of tax simplicity on individual and firm behavior.[7]

Finally, this paper adds to a long tradition in public economics emphasizing the need to screen out applicants for welfare benefits using ordeal mechanisms such as waiting in line, filling out forms, etc.[8] If poorer individuals value their time less – possibly because they are unemployed – then such policies can suc-

[6]Slemrod and Sorum (1984) and Slemrod (1989) report an attrition rate of 71.3%.

[7]See for example Abeler and Jäger (2015) and more recently Harju et al. (2017) and Aghion et al. (2017).

[8]Nichols et al. (1971) and Duclos (1995).

cessfully target them by screening out richer individuals. My results lend support to this assumption because richer individuals tend to forgo more benefits than poorer ones. However, given how large hassle costs are, such policies could be screening out too many individuals. In addition, time inconsistency could lead to unwanted distortions, such as screening out procrastinators versus non-procrastinators rather than rich versus poor individuals.

1 Data and Institutional Background

1.1 The Decision to Itemize Deductions

Taxpayers can reduce their taxable income by claiming deductions. Consider, for example, a single person with an income of $150,000 in the 28% marginal tax bracket. If she spends a total of $10,000 on deductible expenses, her tax liability is reduced by $2,800. If, instead, she decides to claim the standard deduction – which in 1989 was $3,100 – her tax liability is reduced by only $868.

The decision to itemize deductions requires comparing two numbers: the sum of itemized deductions and the standard deduction amount. Itemizing, however, is administratively costly, as it requires collecting several documents and working through a separate tax form.

Approximately two thirds of the population claim the standard deduction. The standard deduction varies by filing status (single, joint, married filing separately and head of household) and by whether the person is blind or older than 65.

1.2 The Cost of Itemizing

Itemizing deductions is a two-step process. First, the taxpayer has to keep a record of all the expenses she wants to deduct during the year for which she is filing taxes, year t. Second, she has to file a separate form when itemizing: Schedule A.

The majority of taxpayers itemize four types of deductions:

- State and local income taxes: these are taxes paid in year t to the state or local government. They are reported on form W2 received in January of year $t+1$. On average, they represent 17% of total deductions.
- Mortgage interest: this is the interest paid to finance the main or second home of the taxpayer. It is reported on form 1098, which the taxpayer receives in January of year $t+1$. On average, they represent 40% of total

deductions.

- Real estate taxes: these are taxes paid on real estate owned by the taxpayer. They can be found in financial records or by calling the county tax assessor. On average, they represent 14% of total deductions.
- Charitable donations: any payment made for charitable purposes, including to religious institutions. Except for vehicle donations in excess of $500, these payments are not subject to third-party reporting. There are, however, documentation requirements for large donations.[9] In addition, taxpayers need to keep records of their own receipts. On average, charitable donations represent 12% of total deductions.

In addition, some taxpayers can also deduct other taxes (sales taxes in some years), other interest expenses (credit-card interest in some years), casualty or theft losses, medical and dental expenses and miscellaneous deductions.

Schedule A is relatively easy to fill out, especially if the taxpayer itemizes only the most common deductions outlined above. All she has to do is copy numbers from form 1098, form W2 or charitable contribution receipts, sum them up and copy the sum on the 1040 form. There are no complicated tax schedules or intricate tax operations. Record-keeping is more time consuming since one has to archive the various records of expenses to be able to recover them when tax season arrives. It is, however, easier to keep track of deductions that are third-party reported given that taxpayers receive forms W2 and 1098 in January of year $t + 1$.

1.3 Data

The dataset that I use to perform this analysis consists of annual cross-sections of individual tax returns. Constructed by the IRS, it is called the Individual Public Use Tax Files – commonly referred to as the Statistics of Income (SOI) files. The data are available annually for the periods that I analyze. The number of observation per year ranges from 80,000 to 200,000. The repeated cross-sections are stratified random samples in which the randomization occurs over the Social Security number. The data oversamples high-income taxpayers and taxpayers with business income, but the IRS provides weights that are used in this analysis to produce estimates representative of the total US population. All of the analysis

[9] Property donations in excess of $500,000 require an appraisal.

in this paper uses these weights to reflect population averages. In addition, I use a panel of tax returns known as the University of Michigan tax panel. The panel covers 1979 to 1990 and contains the same variables as the SOI files but has a smaller sample size (fewer than 40,000 observations per year). Sample restrictions for each figure and table are detailed in Appendix section C.

2 Missing Mass

If some taxpayers are claiming the standard deduction when the sum of their itemized deductions is greater than the standard deduction, there should be a missing mass just above the standard deduction threshold. I graph the density of deductions for 1980, 1985, 1990, 1995, 2000 and 2005 by bin sizes of $2,000 in Figure 1.[10] The bin closest to the standard deduction includes only those itemizers whose deductions are strictly larger than the standard deduction. Notice that the density is systematically low just above the standard deduction and then increases and peaks two to three bins away. This holds true across all years from 1980 to 2006 and for all filing statuses. Since I cannot observe the distribution of itemizers below the standard deduction, this cross-sectional evidence does not prove that the missing mass is caused by the standard deduction, and one could argue that it is merely a naturally occurring feature of the distribution.[11]

To prove that the missing mass is a distortion due to the standard deduction, I turn to a quasi-experimental design. There were four large increases in the standard deduction amounts since 1960. These changes occurred in 1971, 1975, 1988 and 2003. I use the 1988 reform to estimate the cost of itemizing because other changes occurred at the same time as the 1971, 1975 and 2003 reforms. In particular, the 2003 reform is likely to provide a *lower bound* on the cost of itemizing since there were changes in marginal tax rates and deduction rules that made it more attractive to itemize. In 1971 and 1975, there were changes to the parallel standard deduction system.[12] Although the magnitudes of the estimated costs for the 1971, 1975 and 2003 reforms are inaccurate, they still

[10] All remaining years from 1980 to 2006 are plotted in Appendix Figures H.10, H.11, H.12, H.13 and H.14. All dollar amounts are in 2016 dollars in the rest of the paper.

[11] Appendix Figure H.22 shows different alternative scenarios that could create a missing mass.

[12] More details about the parallel system of standard deduction and other changes are provided in Appendix section E.

provide reduced-form evidence of the *existence* of compliance costs.

I compare the pre-1988 reform year to the post-reform year to account for lagged behavioral responses. Figures 2a and 2b graph the density of deductions in pre- and post-reform years for the 1988 reform.[13] Notice that the shape of the distribution in year t+1 mirrors that of years t and t-1 and that the missing mass precisely follows the new standard deduction threshold. This shows that some itemizers switch to the standard deduction once it is increased, even though their deductions are larger than the standard deduction.

The fact that the missing mass closely follows the standard deduction establishes that there is a discontinuity in the distribution *caused* by the standard deduction. If this missing mass is a feature of the distribution and not due to the standard deduction, it should not track the standard deduction once it is increased.

3 Cost Estimation

3.1 Cost Estimation Methodology

To calculate the distribution of forgone benefits in the population, I need to reconstruct the counterfactual distribution of itemizers. Using the pre-reform year as the counterfactual distribution would lead to an underestimate of the cost because the pre-reform distribution is distorted by its proximity to the standard deduction, as Figure 2a shows. This section explains how I reconstruct the counterfactual distribution. Importantly, this estimation method is *model-free*: the estimated distribution of forgone benefits does not require nor depend on any assumptions made over the determinants of the forgone benefits. No assumptions about the drivers of the cost are needed in this section: $C_S(\cdot)$ can be due to the sum of the costs of record keeping and filing or can also include fear of audits and uncertainty, which I discuss in Section 5. While alternative explanations could change the interpretation of the estimated dollar amount of forgone benefits, they would not change the dollar amount itself.

Denote by $f(\cdot)$ the unobserved p.d.f. of itemizers, assuming that there is no standard deduction and no cost of itemizing, as illustrated in Figure 3. Denote

[13] Appendix Figure H.16 reports these densities for the 1971, 1975 and 2003 reforms and shows that the changes are qualitatively consistent with the 1988 reform.

by $g_S(\cdot)$ the observed probability density function (p.d.f.) of itemizers when the standard deduction is equal to S. Then, $g_0(\cdot)$ and $g_\delta(\cdot)$ correspond, respectively, to the pre- and post-reform p.d.f. of itemizers when the standard deduction increases from 0 to δ. The cumulative distribution function (c.d.f.) of the cost of itemizing is denoted by $C_S(\cdot)$ and is defined over $[0, c_{max}]$, where c_{max} denotes the largest cost an individual can have. Individuals whose total deductions exceed the standard deduction by less than the cost of itemizing choose the standard deduction. Formally,

$$\forall S = \{0; \delta\} : \quad g_S(d) = \begin{cases} 0, & \text{if } d \leq S \\ f(d)(1 - C_S(d - S)), & \text{if } S < d \leq c_{max} + S \\ f(d), & \text{if } d > c_{max} + S. \end{cases} \tag{1}$$

By rearranging (1) over $d \in [0, c_{max}]$:

$$C_S(d - S) = \frac{f(d) - g_S(d)}{f(d)}. \tag{2}$$

In other words, the cost of itemizing is related to the missing mass $f(d) - g_S(d)$, which is shown in Figures 1, 2a and 2b. However, because $f(\cdot)$ cannot be observed directly, it needs to be reconstructed using $g_0(\cdot)$ and $g_\delta(\cdot)$. Two assumptions are necessary:

- **A1**: Cost is constant across years.
- **A2**: Cost is independent of the level of deductions.

Assumptions A1 and A2 imply that $C_0(\cdot) = C_\delta(\cdot)$, and from equation (2), it follows that:

$$C_0(d) = \frac{f(d) - g_0(d)}{f(d)} = \frac{f(d + \delta) - g_\delta(d + \delta)}{f(d + \delta)} = C_\delta(d), \tag{3}$$

which implies that the same proportion of individuals is missing d deductions above the pre-reform standard deduction and $d + \delta$ deductions above the post-reform standard deduction.

Assumption A1 can be verified by graphing two densities in years with no reforms and ensuring that they are overlapping. This assumption is verified on

all years from 1980 to 2006.[14] A failure of A2 introduces a relatively small bias in the cost estimate: in Appendix section B, I provide an upper bound on the size of this bias and show that it is small.[15]

To estimate $C(\cdot)$ and reconstruct $f(\cdot)$, I proceed in three steps. First, if $d \in [\delta + c_{max}; +\infty]$, then the benefit of itemizing is greater than its cost both pre- and post-reform, and taxpayers will not forgo deductions by claiming the standard deduction. This corresponds to the rightmost area in Figure 3. Formally, if $d \in [\delta + c_{max}; +\infty]$, then $C(d) = 1$ and $g_\delta(d) = g_0(d) = f(d)$, i.e., the pre- and post-reform distributions of itemizers overlap for ranges of deductions exceeding the post-reform standard deduction δ by more than the largest possible cost c_{max}. And for any $d \in [\delta + c_{max}; +\infty]$, $f(d) = g_0(d)$, i.e., the pre-reform observed distribution of itemizers $g_0(\cdot)$ corresponds to the undistorted distribution $f(\cdot)$.

Second, if $d \in [c_{max}; \delta + c_{max}]$, then over this range, the pre-reform taxpayers do not forgo any deductions, but the post-reform ones do. This corresponds to the middle area in Figure 3. As a consequence, the pre-reform distribution is not affected by its proximity to the standard deduction and is equal to the undistorted distribution, i.e., $g_0(d) = f(d)$, but the post-reform distribution is distorted, i.e., $g_\delta(d) < f(d)$. From equation (3), it follows that $\forall d \in [c_{max}; \delta + c_{max}]$:

$$C(d - \delta) = \frac{f(d) - g_\delta(d)}{f(d)} = \frac{g_0(d) - g_\delta(d)}{f(d)}, \tag{4}$$

which allows me to estimate $C(\cdot)$ over $[c_{max} - \delta; c_{max}]$.

Third, if $d \in [c_{max} - \delta, c_{max}]$, then both the pre-reform and post-reform itemizers are forgoing deductions. This corresponds to the leftmost area in Figure 3. In this case, both the pre- and post-reform distributions are distorted by their proximity to the standard deduction and $g_0(\cdot)$ is now different from $f(\cdot)$. To reconstruct $f(\cdot)$, I use the estimate of $C(\cdot)$ over $[c_{max} - \delta; c_{max}]$ from equation 4 to correct the pre-reform distribution by using the definition of $g_0(\cdot)$:

[14] See Appendix Figure H.17.

[15] For joint filers in the 28% bracket, for example, the estimated cost would lie between \$572 and \$644 if A2 fails, instead of \$644.

$f(d) = g_0(d)(1 + C(d))$. From equation 3, it follows that $\forall d \in [c_{max} - \delta; c_{max}]$

$$C(d - \delta) = \frac{f(d) - g_\delta(d)}{f(d)} = \frac{g_0(d)(1 + C(d)) - g_\delta(d)}{f(d)}, \tag{5}$$

which allows me to estimate $C(\cdot)$ over $[c_{max} - 2\delta; c_{max} - \delta]$. By repeating this procedure over $[c_{max} - 3\delta; c_{max} - 2\delta]$, $[c_{max} - 4\delta; c_{max} - 3\delta]$, etc., I can recover $C(\cdot)$ and $f(\cdot)$ over $[0, c_{max}]$.

3.2 Cost Estimates

I apply the methodology outlined above to the 1988 reform, which increased the standard deduction 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-households filers. Besides the standard deduction reform, the only other 1988 reform that could have affected the amount of deductions was the phase-out of the personal interest deduction, which I control for (details in Section 5.5). There were no other reforms affecting deductions in 1988 or 1989, and the reforms affecting the 1987 distribution did not have lagged effects.[16] Each cost estimate is performed on individuals with the same marginal tax rate and who were not subject to the Alternative Minimum Tax (AMT). There was a marginal tax rate decrease for married filing jointly with income above \$45,000 (in 1987 dollars) in 1988. I address this by estimating the cost separately for individuals above and below this cutoff.

I use 1989 rather than 1988 as the post-reform year because the reform occurred in 1988. If taxpayers learned about the increase in the standard deduction when filing their taxes, we should observe the full response in 1989. Figure 2b confirms that the effect was smaller during the reform year.

Table 1 shows the estimated costs for single, joint and head-of-households filers in the 15% and 28% marginal tax brackets. Costs range from 0.57% to 0.85% of AGI. In dollar amounts, they vary from 175\$ for single filers in the 15% bracket to \$591 for joint filers in the 28% bracket. Costs expressed in dollars are systematically lower for individuals in lower tax brackets. They are, however, more homogenous when expressed as a percent of AGI. This suggests that income matters in determining the cost, as I show in the next section.

[16] See Section 5.5 for the full list of reforms and Appendix D for the TRA'86 reforms.

To calculate the standard errors of the difference between the bins in the 1987 and 1989 densities, I use a bootstrap procedure. The results are reported in Table 3. The difference between the first and second bins is statistically significant with large z statistics (6.55 and 3.47). The rest of the bins are all overlapping, with differences that are not significant, even at the 10% level, with the exception of bins 10, 11 and 13, which are statistically significantly different at the 5% and 10% level, with differences of a very small magnitude (less than ten times that of the first or second bins).

4 Anatomy of the Missing Mass

4.1 Costs Increase With Income

If rich taxpayers value their time more than poor ones because their hourly wage is higher, we should expect them to forgo more deductions. I can test this using the income reported on tax returns. To do so, I break down the sample used above by deciles of income. This raises power issues that I deal with in two ways. First, I use a moving average of the income deciles. For example, the lower-income group consists of every individual with income below the second decile threshold. And the second group consists of taxpayers with income above the first decile and below the third decile, etc. Second, I focus on joint filers in the 28% marginal tax bracket, as they represent by far the largest group of taxpayers.

Once the groups are constructed, I fit a polynomial of degree 3 through each deduction bin. I then calculate the difference in density for each bin. When this difference is not statistically significant, I consider that the bins are overlapping, and, therefore, no deductions are forgone in that specific bin. Using the predicted bins from this polynomial, I calculate the forgone benefits for each group by repeating the procedure developed in the previous section: I compare the distribution in 1987 to that in 1989, reconstruct the counterfactual distribution of itemized deductions and calculate the distribution of the cost of itemizing by comparing the counterfactual distribution to the true one. I report results for only the first seven groups because deductions and income are positively correlated. This implies that very few high-income individuals are close to the standard deduc-

tion threshold, making the estimates for these bins uninformative.[17] The results are plotted in Figure 4(a): as income increases, taxpayers forgo more benefits, consistent with the idea that they value their time relatively more. Two pieces of information are worth emphasizing. First, all taxpayers in Figure 4(a) fall in the 28% marginal tax bracket, implying that the positive relationship between income and forgone benefits is not due to marginal tax rate variation but, rather, to income per-se. Second, even though itemized deductions increase with income, this is not what drives the increasing relationship between income and forgone benefits. Because I use a quasi-experimental design and compare the same income groups before and after the reform, I am implicitly controlling for the relationship between income and deductions.

4.2 Tax Preparers and Electronic Filing

Electronic filing and the use of tax preparers may reduce the cost of filling out forms, as one need not file schedule A. However, it does not affect the cost of record keeping. Therefore, if record-keeping costs are non-zero, electronic filing or the use of tax prepapers will not reduce costs to zero, and one will still observe a missing mass close to the standard deduction threshold. Survey estimates of the cost of filing taxes have consistently documented that record keeping is the main driver of the cost of itemizing.[18]

To test for whether electronic filing or using a tax preparer eliminates the cost of itemizing, I graph the density of itemizers who use a tax preparer and those who use electronic filing in Figure 5 and look for whether there is still a missing mass close to the standard deduction threshold. The missing mass still exists, implying that tax preparers or electronic filing do not eliminate the cost of itemizing.

Figure 5(b) compares the density of taxpayers who use electronic filing to those who do not. It shows a slightly smaller missing mass for taxpayers who file electronically than for those who do not. The difference is statistically significant.[19] This is consistent with the missing mass being driven by taxpayers who claim the standard deduction to avoid the cost of itemizing. However, electronic filing only

[17] The remaining two income bins are reported in Appendix Figure H.21.

[18] See, for example, Guyton et al. (2003), Slemrod and Sorum (1984), Slemrod and Bakija (2008) and Blumenthal and Slemrod (1992).

[19] Bootstrapped standard errors are reported in Table I.7

slightly reduces the cost of itemizing and does not eliminate the missing mass, which is consistent with record keeping being the main driver of compliance costs.

I cannot perform a similar test for taxpayers who use tax preparers, as the two densities do not overlap away from the standard deduction – possibly because the population that uses tax preparers is intrinsically different from the population that does not – making a comparison of the missing mass impossible. Figure 5(a) shows, however, that the use of tax preparers does not eliminate the cost of itemizing.

5 Alternative Explanations to Compliance Costs

5.1 Information

Information or cognitive abilities are unlikely to explain the forgone deductions. Since the taxpayers on whom I focus switch from itemizing to claiming the standard deduction, they should be well aware of the decision to itemize and have the cognitive abilities to do so. In addition, taxpayers are reminded on the 1040 form that they can make the choice between itemizing and claiming the standard deduction.

5.2 Concave Kink Points

When claiming the standard deduction, taxpayers are paying the full cost of items such as charitable donations or mortgage interest that they may have incurred below the standard deduction. However, when they itemize, they pay only a portion of the cost because deductions are subsidized by 1 minus the marginal tax rate. The standard deduction acts as a *concave* kink point: the price of charitable donations is lower when itemizing than when claiming the standard deduction. The indifference curve of a given taxpayer can be tangent at two points of the concave kinked budget set (see Figure 6(a)), possibly inducing some taxpayers to be indifferent between two points, one above the standard deduction and one below it. Depending on the curvature of the indifference curve, this could create a bi-modal distribution with a missing mass both to the right and to the left of the standard deduction (see Figure 6(b)).

However, according to the assumption that taxpayers respond to concave kink points, the size of the missing mass should not respond to variations in income when controlling for the marginal tax rate. The only reason for taxpayers to

adjust their deductions in response to a concave kink point is the marginal tax rate. Income, per se, should not matter in this case. On the other hand, a behavioral response due to compliance costs predicts that richer taxpayers will forgo more money because they have a higher opportunity cost of time, even controlling for the marginal tax rate. Figure 4 graphs the relationship between forgone benefits and income - *controlling for the marginal tax rate* - and finds an increasing relationship, rejecting the hypothesis that taxpayers are responding to concave kink points in this setting.

In addition, behavioral responses to concave kink points lead individuals to locate away from the concave kink point. This mechanism is illustrated in Figure 6. If behavioral responses to concave kink points led to the observed missing mass, as the standard deduction increases, the bi-modal distribution should track the new standard deduction threshold, as illustrated in Figure 6(c) and the pre- and post-distribution peaks should not overlap. The observed pre and post distributions in Figures 2a and 2b contradict the prediction of Figure 6(c): the pre- and post-distribution peaks are overlapping, once again rejecting the hypothesis that the missing mass is caused by behavioral responses to concave kink points.

Overall, both the fact that forgone deductions increase with income and the shape of the post-reform distribution of deductions rule out responses to concave kink points. The absence of behavioral responses to concave kink points is consistent with the empirical public finance literature that documents behavioral responses to tax systems and does not find any evidence of responses to concave kink points. Saez (2010), Kleven and Waseem (2013) and Tazhitdinova (2015) directly test the predictions of a behavioral response to both concave and convex kink points, finding responses to convex kink points but no responses to concave kink points. For example, Saez (2010) finds evidence of bunching at the threshold of the first income tax bracket where tax liability starts and at the first kink of the EITC – which are both convex kinks; however, he finds no evidence of bunching at the second EITC kink – which is concave. Kleven (2016), in a survey of the bunching literature confirms that there is no evidence of bunching at concave kink points.

5.3 Evasion

An alternative explanation for the missing mass is that taxpayers are concerned with being audited by the IRS. They mistakenly believe that audit probabilities are higher when itemizing. Their beliefs about audit probabilities could lead them to switch to the standard deduction once it increases, in order to avoid the expected cost of an audit.

However, since audit probabilities are very low, for this behavior to explain the missing mass, taxpayers would need to mistakenly believe that audit probabilities are high or that audit costs are large. To address this, I conduct a survey of 195 individuals in a wealthy neighborhood of the city of Los Angeles, California to capture as many itemizers as possible. Appendix section F details the survey instrument. The survey allows me to elicit their beliefs about both the audit probabilities for itemizers and the perceived costs of undergoing an audit. Figure 8 reports the results of the survey.

Surveyed individuals have levels of income similar to those of joint filers in the 1988 28% marginal tax bracket. On average, they believe that audits occur with a probability of 8.72%, which is 7.9 times the true audit probability.[20] This accounts for, at most, 25% of the $591 estimated forgone benefits for joint filers in the 28% marginal tax bracket.[21]

5.4 Rational Inattention

Can uncertainty about the level of deductions lead a taxpayer to switch to the standard deduction and explain the observed missing mass? Table 4 shows the results of the calibration of a model illustrative of this type of behavior with varying levels of risk aversion.[22] Taxpayers would need an uncertainty range of at least $\pm\$14,000$ in order to forgo amounts of money like those found in this paper when their true deductions are $10,000. This uncertainty range is large and unlikely for two reasons. First, I focus on taxpayers who were itemizing in the previous year. Second, total deductions are highly serially correlated across years for a given individual since 71% of total deductions are mortgage interest,

[20] This is consistent with Bhargava and Manoli (2015) who find that EITC filers believe that audit probabilities are eight times greater than the true ones.

[21] On average, their willingness to pay to avoid an audit is $1,748, which implies that the expected cost of an audit is $147, with a 95% confidence interval of $[126, 169]$.

[22] The model is outlined in Appendix section G.

state taxes and real estate taxes, which are relatively stable for a given person year after year.

5.5 Other Reforms Affecting Deductions?

Other changes took place in 1988. In this section, I describe these changes and explain how I adjust for those that are likely to affect my estimates. The estimates derived in Section 3 have already accounted for these adjustments. The fact that the pre- and post-reform densities overlap away from the standard deduction threshold shows that the pre-reform density is a relevant counterfactual for the post-reform density in Figure 2a and that – after adjusting for these changes – the missing mass estimates are not affected by these changes.

The personal interest deduction was phased out starting in 1986. Taxpayers could deduct only 65% of their personal interest in 1987, 40% in 1988 and 20% in 1989. This is likely to affect the distribution of deductions from 1987 to 1989. To control for this effect, I adjust the 1987 distribution - which is the counterfactual for 1989 - by recalculating the personal interest deduction as if only 20% of it could be deducted. This leads some taxpayers to have deductions below the standard deduction, and I drop them from my sample. To ensure that there is no behavioral effect associated with the phasing out of the personal interest deduction, I compare the distribution of deductions for individuals above and below the 28% marginal tax rate bracket. If there had been a behavioral effect, we should observe more deductions for individuals above the 28% marginal tax bracket. I find no significant behavioral response of personal interest deductions.[23] This is consistent with the fact that the majority of the personal interest deduction is claimed for interest on student loans, which are hard to adjust once they are contracted. In addition, after making this correction, I can compare the overlap between the pre- and post-reform densities. Away from the standard deduction, the two graphs overlap, implying that the post-reform density is an appropriate counterfactual for the 1989 density.

[23] See Appendix figure H.23.

6 Total Filing Costs Increase Over Time

6.1 Average Treatment Effect

The cost I estimate is a *local* average treatment effect representative of filers who are just above the standard deduction. To get an average treatment effect, 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 different income levels since deductions and income tend to be positively correlated. 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.

To estimate the effect of dependents and the use of tax preparers on the cost of itemizing, I apply the procedure outlined in Section 3.1 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 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 4.1. Because electronic filing did not exist in 1988, I cannot use the procedure from Section 3.1. 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.[24] Formally, I assume that the filing cost is given by the following

[24]The three distributions are shown in Figure H.20.

equation:

$$C = \beta * \{[(\alpha_{efi}\mathbb{1}_{efi} + \alpha_{\overline{efi}}(1 - \mathbb{1}_{efi})] + [\alpha_{prep}\mathbb{1}_{prep} + \alpha_{\overline{prep}}(1 - \mathbb{1}_{prep})] + [\alpha_{dep}\mathbb{1}_{dep} + \alpha_{\overline{dep}}(1 - \mathbb{1}_{dep})] + \sum_{i=0}^{9} \alpha_i \mathbb{1}_i, \quad (6)$$

where β is the baseline cost of itemizing, as estimated in Section 3.1; 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 4.1: 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 coefficients are reported in Table 2.

6.2 Other Schedules

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 6.1 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.[25] 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 (6), with a subscript x

[25] According to 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. See Appendix Table I.11 for a breakdown of the cost.

that corresponds to each tax schedule:

$$C_x = \beta_x * \{[(\alpha_{efi}\mathbb{1}_{efi} + \alpha_{\overline{efi}}(1 - \mathbb{1}_{efi})] + [\alpha_{prep}\mathbb{1}_{prep} + \alpha_{\overline{prep}}(1 - \mathbb{1}_{prep})] + [\alpha_{dep}\mathbb{1}_{dep} + \alpha_{\overline{dep}}(1 - \mathbb{1}_{dep})] + \sum_{i=0}^{9} \alpha_i \mathbb{1}_i\}, \quad (7)$$

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.

6.3 Costs Have Been Increasing Since the 1980s

Using equation (7), I estimate the total cost of filing all federal income tax schedules for every year from 1984 to 2006.[26] Figure 7a 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 7b 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 7b 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.[27]

7 Compliance Costs or Behavioral Costs?

There is extensive evidence that individuals are time-inconsistent and tend to procrastinate.[28] If taxpayers procrastinate on filing their taxes, one should

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

[27] Tazhitdinova (2014) explores this tradeoff in the case of charitable donations.

[28] See DellaVigna (2009) for a survey of the literature.

observe a large proportion of taxpayers filing on April 15th and procrastinators forgoing more deductions.[29]

First, consistent with individuals procrastinating on filing their taxes, I find that taxpayers 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[30] and graphs the number of tax returns filed by week. Both exhibit a clear spike in the weeks that include April 15th. This is consistent with Hoopes et al. (2015), who show that that more calls are made to the IRS close to April 15th and that taxpayers search more actively on Google and Wikipedia for capital-gains-tax-related information.

Second, I also find that taxpayers who file close to the deadline tend to forgo more deductions, consistent with procrastination accounting for a portion of the estimated forgone deductions. Figure 9c[31] shows that the missing mass for close-to-the-deadline filers (first two weeks of April) is larger than for March filers.

Note that rational taxpayers should not file close to the deadline for two reasons: 1)by delaying filing, they forgo interest on their refunds; and 2)they expose themselves to higher filing costs. Indeed, the sample I use to generate Figure 9c includes only those taxpayers who are owed a refund by the IRS and, therefore, have an incentive to file as early as possible to save on interest.[32] Second, filing costs are substantially higher closer to the deadline because lines at the post office are longer; appointments with tax preparers are scarcer; and it is harder to get tax help from the IRS because their phone lines are busier than usual.[33]

Note also, that late filing is hard to reconcile with the option value of waiting for low cost realizations. One could argue that taxpayers who bunch at the deadline are rational taxpayers who wait for a low cost realization and face a series of idiosyncratic shocks that force them to file hastily at the very last moment and lead them to forgo benefits. If that is the case, then we should observe that

[29] I formalize this argument in Appendix section H and show that procrastination can lead to high record-keeping costs, resulting in individuals failing to itemize.

[30] https://www.irs.gov/uac/2016-and-prior-year-filing-season-statistics

[31] Appendix section C.6 explains how the graph is constructed.

[32] Slemrod et al. (1997) estimate that taxpayers forgo $46 million in interest by not claiming their refund as soon as possible.

[33] Redelmeier and Yarnell (2012), for example, report that there are more road fatalities on April 15th and argue that this is due to taxes.

taxpayers who file late in year t are likely to file earlier in year $t+1$. To test for this, in Figure 9d, I graph the average week in which returns are processed in year $t+1$ by week of processing in year t. If taxpayers who bunch at the deadline are doing so for rational reasons, the relationship should be constant, as we should observe mean reversion. If they are doing so because of a systematic bias, the relationship should be increasing as the year t week of processing should predict the year $t+1$ week of processing. Figure 9d shows an increasing relationship – with a t-statistic equal to 49 – between the processing weeks in year t and year $t+1$, consistent with the explanation that late filing is due to a systematic bias.

8 Conclusion

Using a quasi-experimental design and a novel method to recover the counterfactual density of deductions, I find that taxpayers forgo large amounts of deductions. This implies tax filing costs of a much larger magnitude than previously estimated using survey methods. The aggregate cost of filing federal taxes has been steadily increasing since the 1980s in part because of population growth, but also because of the increase in the number of forms that each taxpayer files.

Should the cost of tax filing be reduced? The IRS faces a tradeoff between requiring fewer forms and receipts (and, therefore, reducing filing costs) versus reducing evasion. The large magnitude of my estimates suggests that reducing reporting might be welfare-improving even if it leads to higher evasion costs.

Filing costs can also be reduced without reducing reporting. This is especially true if taxpayers tend to procrastinate. The IRS can ensure that the deadline for filing taxes falls on a day when people are likely to be less busy, such as a weekend day. In addition, the IRS could shift the cost of filing taxes to firms since they are less likely to procrastinate. This can be achieved through informational reporting, as is the case with the mortgage deduction.

However, some compliance cost can be efficient when designing a tax system and can be used as a policy instrument. This is especially true when the social gains of deductions or loopholes are small, and political economy concerns prevent the government from repealing these deductions. One way of ensuring that taxpayers do not claim them is to impose large compliance costs.[34]

[34]Discussed in Kaplow (1998).

Finally, the identification strategy used in this paper can be exported to estimate other compliance costs when individuals have a choice between a low-cost/low-benefit option versus a high-cost/high-benefit one, as is the case in, for example, deciding whether to file a lawsuit in small claims court versus regular court. It can also be used when identifying responses from a censored distribution above or below a certain threshold.

References

Abeler, J. and S. Jäger (2015). Complex tax incentives. *American Economic Journal: Economic Policy 7*(3), 1–28.

Aghion, P., U. Akcigit, M. Lequien, and S. Stantcheva (2017). Tax simplicity and heterogeneous learning.

Bertrand, M., S. Mullainathan, and E. Shafir (2006). Behavioral economics and marketing in aid of decision making among the poor. *Journal of Public Policy & Marketing 25*(1), 8–23.

Bhargava, S. and D. Manoli (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *The American Economic Review 105*(11), 3489–3529.

Blank, R. and D. Card (1991). Recent trends in insured and uninsured unemployment: Is there an explanation? *Quarterly Journal of Economics 106*(4), 1157–1189.

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

Currie, J. (2006). The take-up of social benefits. *Public Policy and the Income Distribution*, 80.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature 47*(2), 315 –372.

Duclos, J.-Y. (1995). Modelling the take-up of state support. *Journal of Public Economics 58*(3), 391–415.

Gronau, R. (1973). The effect of children on the housewife's value of time. *The Journal of Political Economy*, S168–S199.

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.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *The American Economic Review 103*(7), 2643–2682.

Harju, J., T. Matikka, and T. Rauhanen (2017). Compliance costs vs. tax incentives: why small firms respond to size-based regulations?

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Hoopes, J. L., D. H. Reck, and J. Slemrod (2015). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy 7*(3), 177–208.

Kaplow, L. (1998). Accuracy, complexity, and the income tax. *Journal of Law, Economics, and Organization 14*(1), 61–83.

Keys, B. J., D. G. Pope, and J. C. Pope (2014). Failure to refinance. *Working Paper*.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*(1).

Kleven, H. J. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The Quarterly Journal of Economics 129*, 669–723.

Little, A. D. (1988). Development of methodology for estimating the taxpayer paperwork burden.

Madrian, B. and D. Shea (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics 116*(4), 1149–1187.

Nelson, F. D. (1977). Censored regression models with unobserved, stochastic censoring thresholds. *Journal of Econometrics 6*(3), 309–327.

Nichols, D., E. Smolensky, and T. N. Tideman (1971). Discrimination by waiting time in merit goods. *The American Economic Review*, 312 –323.

Pitt, M. and J. Slemrod (1989). The compliance cost of itemizing deductions: Evidence from individual tax returns. *The American Economic Review*, 1224–1232.

Redelmeier, D. A. and C. J. Yarnell (2012). Road crash fatalities on us income tax days. *The Journal of the American Medical Association 307*(14), 1486–1488.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 180–212.

Slemrod, J. (1989). The return to tax simplification: An econometric analysis. *Public Finance Review 17*(1), 3–27.

Slemrod, J. and J. Bakija (2008). Taxing ourselves: a citizen's guide to the debate over taxes. *MIT Press Books*.

Slemrod, J., C. Christian, R. London, and J. A. Parker (1997). April 15 syndrome. *Economic Inquiry 35*(4), 695 –709.

Slemrod, J. and N. Sorum (1984). The compliance cost of the US individual income tax system. *National Tax Journal*, 461 –474.

Steuerle, E., R. McHugh, and E. M. Sunley (1978). Who benefits from income averaging? *National Tax Journal*, 19–32.

Sydnor, J. (2010). (over) insuring modest risks. *American Economic Journal: Applied Economics 2*(4), 177–199.

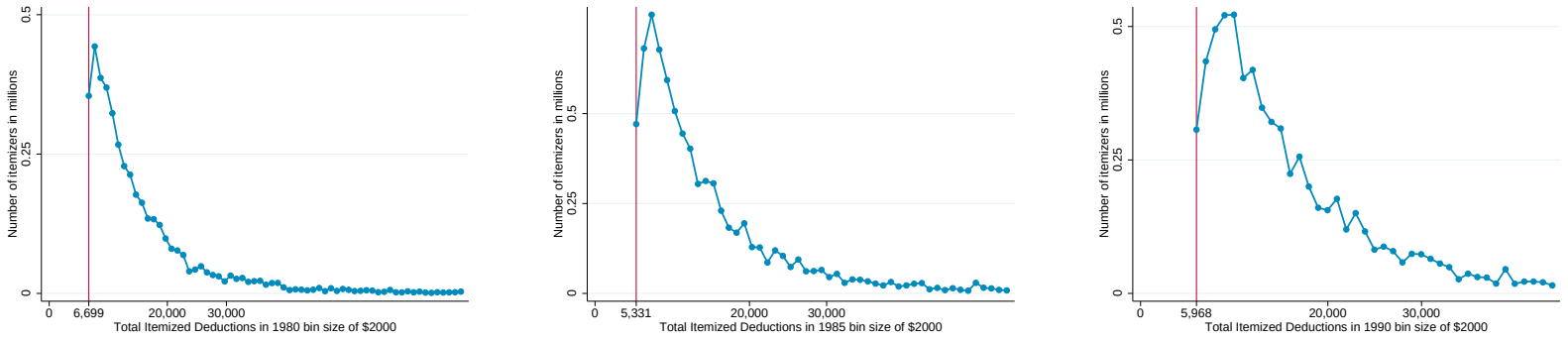
Tazhitdinova, A. (2014). Optimal reporting threshold: Theory and evidence from charitable contributions. Manuscript, UC Berkeley.

Tazhitdinova, A. (2015). Tax breaks for low earners: Who benefits from them? evidence from mini-jobs in germany. *Working Paper*.

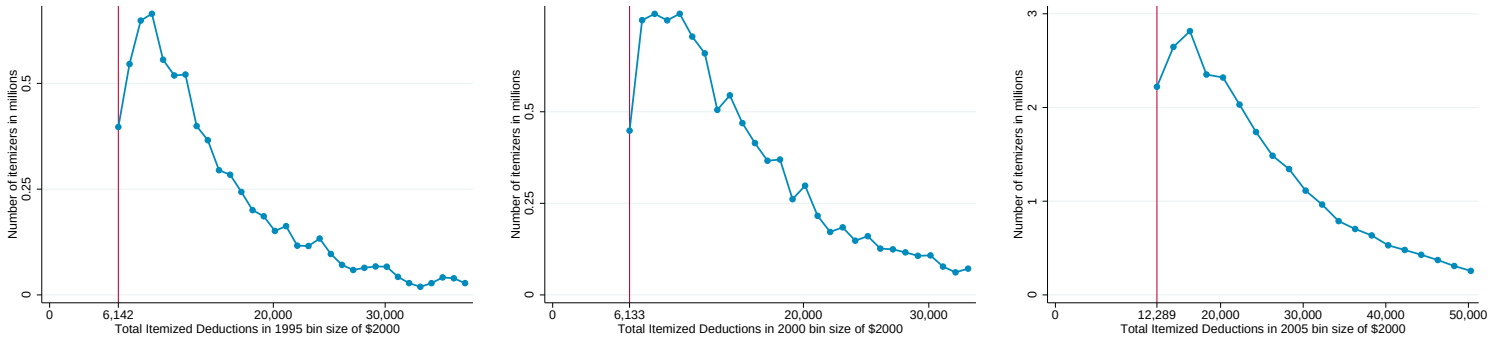
Wicks, J. H. (1965). Taxpayer compliance costs from the montana personal income tax. *Montana Business Quarterly 3*, 36–42.

Wicks, J. H. and M. N. Killworth (1967). Administrative and compliance costs of state and local taxes. *National Tax Journal*, 309–315.

Figure 1: Missing Mass Just Above the Standard Deduction



(a) 1980 **(b) 1985** **(c) 1990**

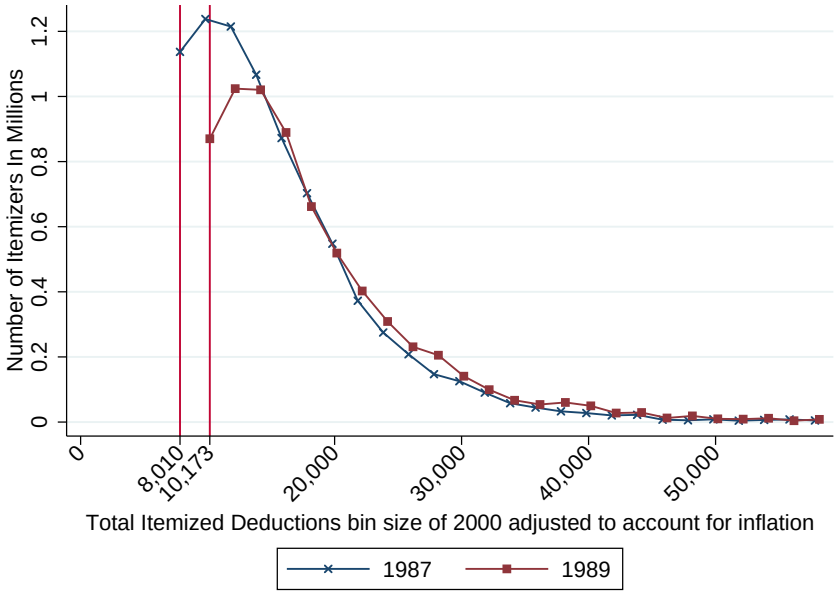


(d) 1995 **(e) 2000** **(f) 2005**

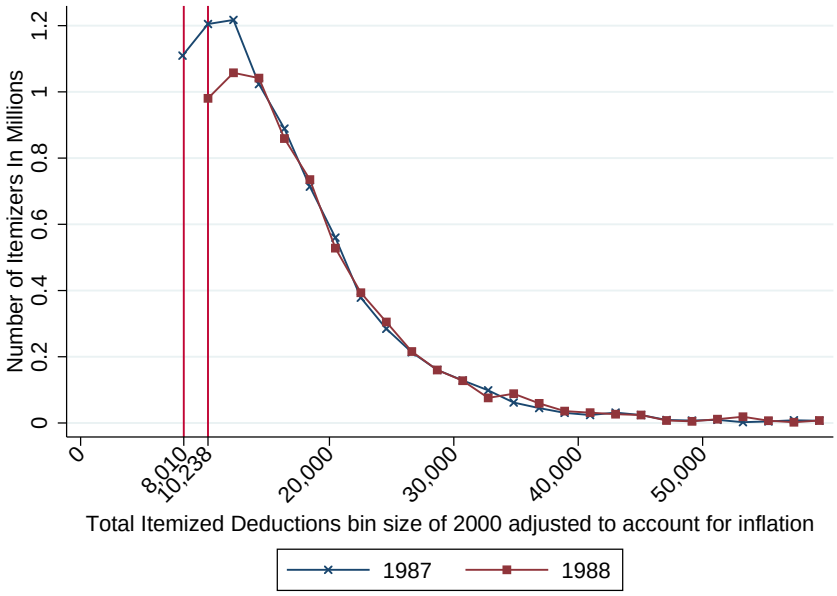
Notes: The figures above plot the density of deductions for itemizers filing jointly. The bin size is $2,000 and the vertical line represents the standard deduction threshold for each year. Additional years are reported in appendix figures H.11, H.12, H.13 and H.14 and figure H.15 for single filers.

Figure 2: Density of Deductions for Itemizers Filing Jointly Pre and Post Reform

(a) 1987-1989 Comparison



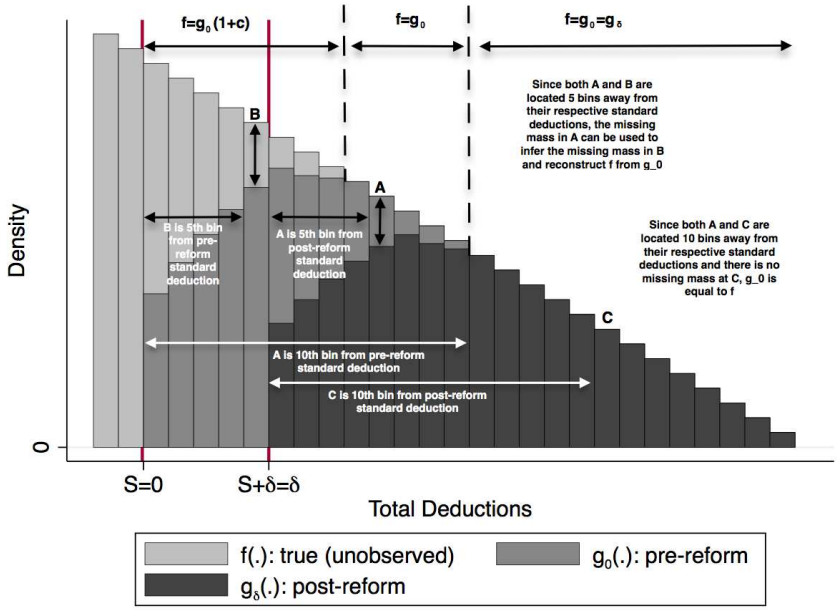
(b) 1987-1988 Comparison



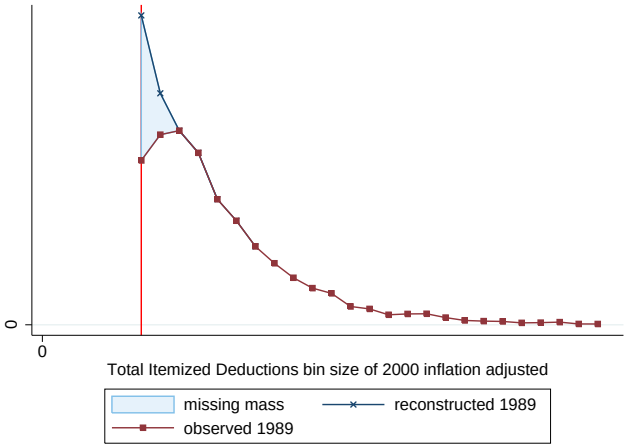
Notes: The first graph plots the density of itemizers one year before and one year after the standard deduction reform while the second one plots these densities one year before and during the reform.

Figure 3: Reconstruction of the Counterfactual Density

(a) Procedure Illustration

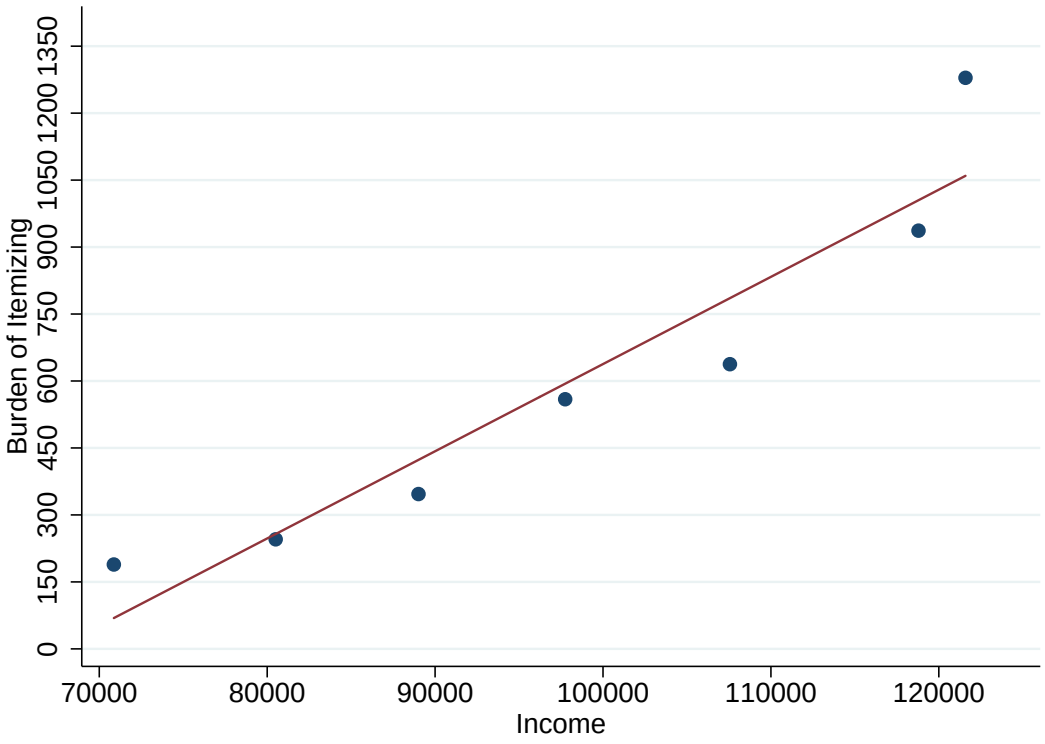


(b) Reconstructed Counterfactual



Notes: The first graph illustrates the method used in section 3.1 to reconstruct the counterfactual density of itemizers $f(.)$ using the pre- and post-reform densities $g_0(.)$ and $g_\delta(.)$. The second graph plots the reconstructed counterfactual in 1989 using the method outlined in section 3 and the observed density for 1989. The missing mass used to estimate the cost of itemizing is given by the area lying between the two curves.

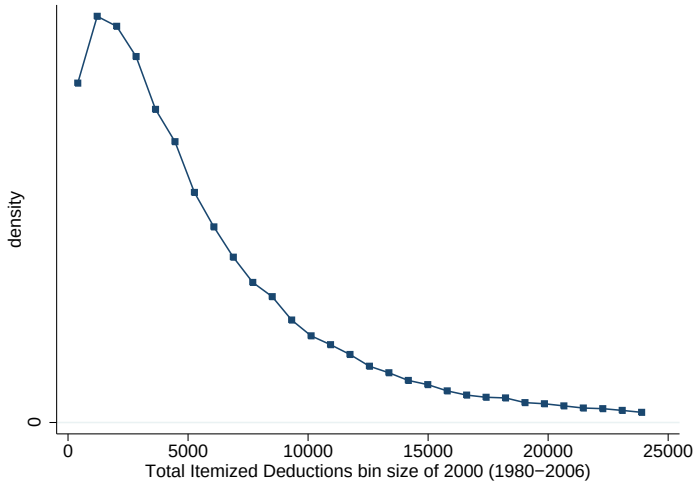
Figure 4: Relationship Between Income and the Cost of Itemizing Deductions



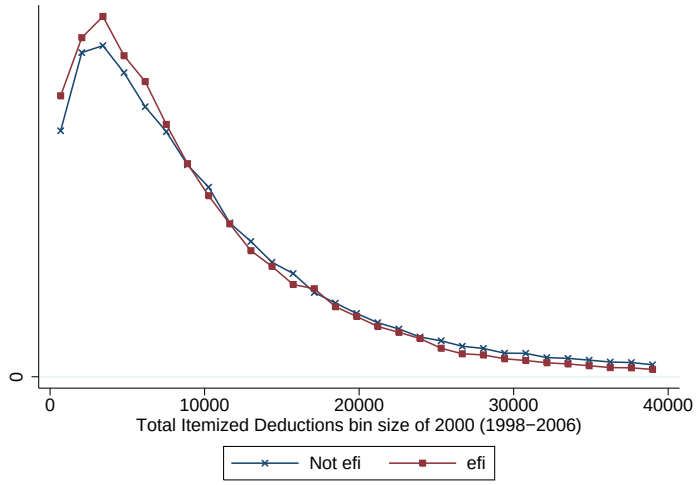
Notes: This graph shows the increasing relationship between income and the cost of itemizing: richer households are more likely to forgo deductions. This relationship controls for the variation in MTR across the different income groups.

Figure 5: Use of Tax Preparer and Electronic Filing

(a) Tax Preparer

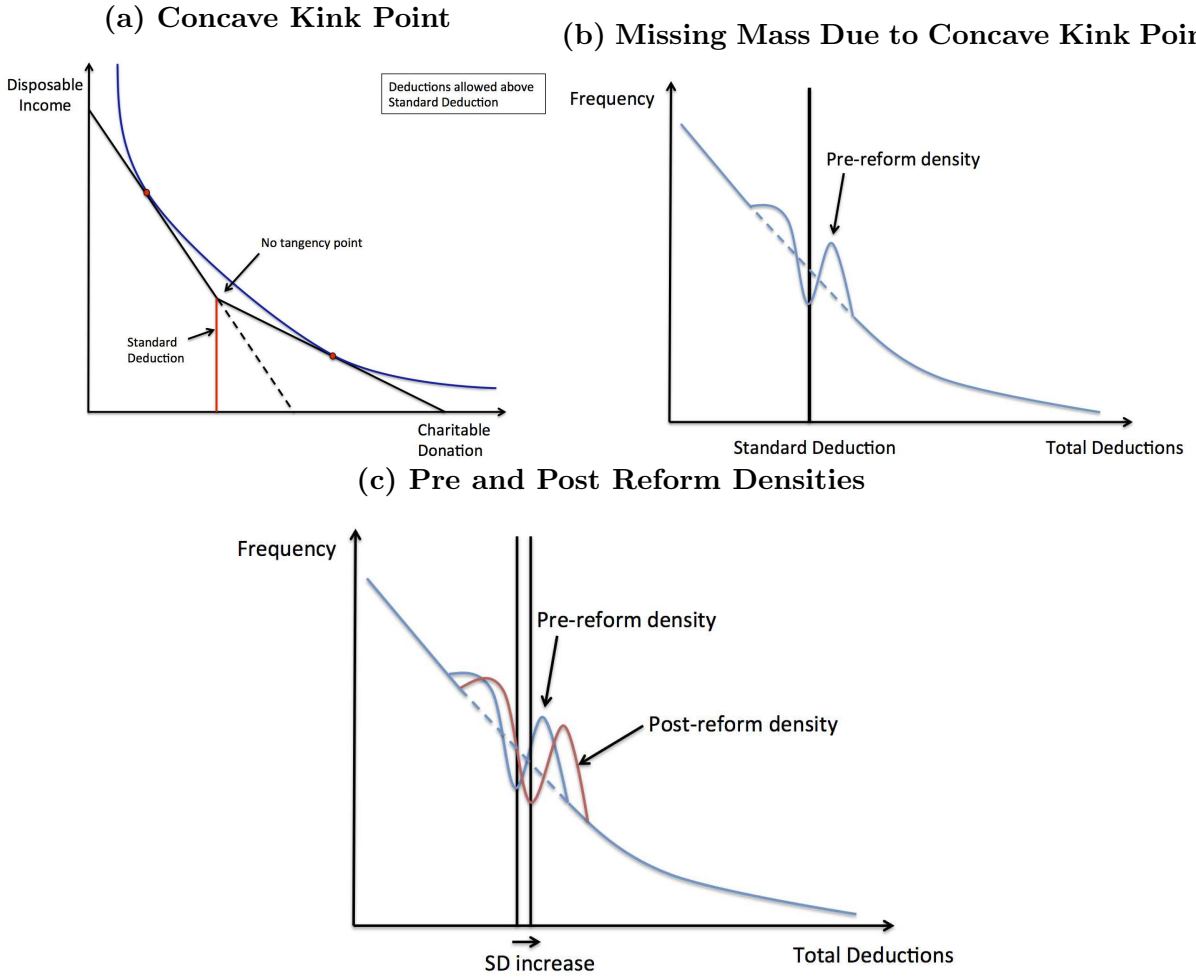


(b) Electronic Filing



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 cost 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.

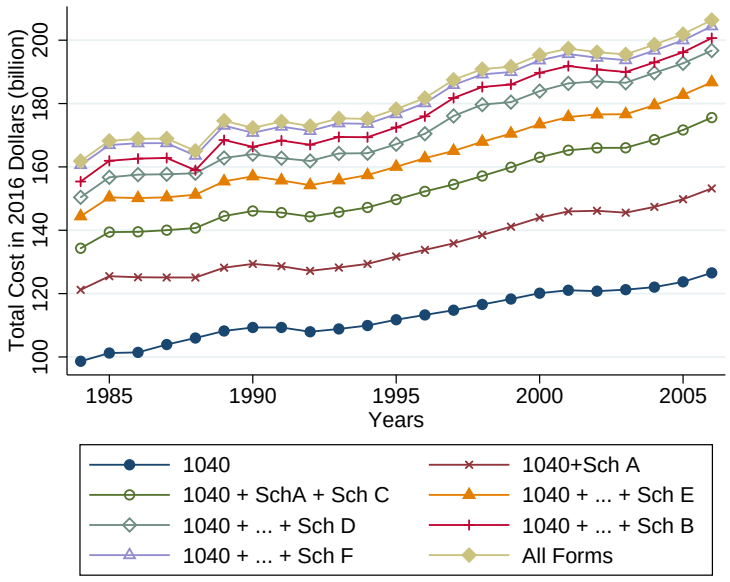
Figure 6: Concave Kink Point: Densities Following Reform Should Not Overlap



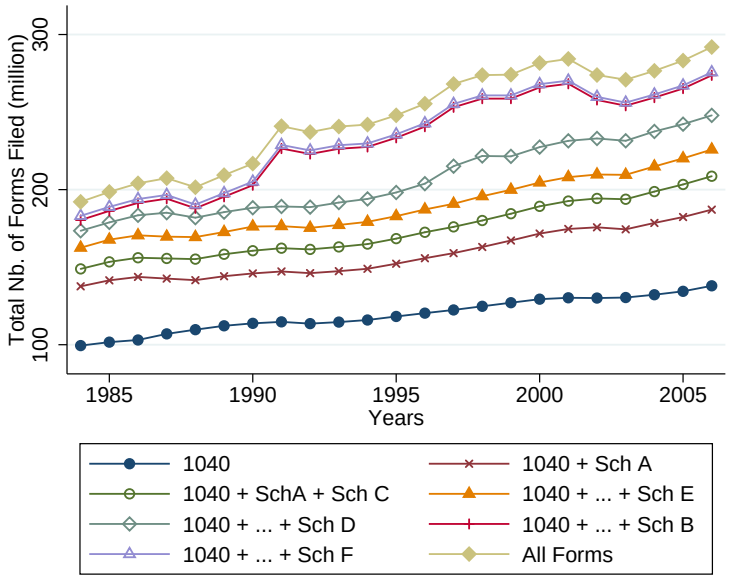
Notes: Panel (a) displays a budget set with a concave kink point. Panel (b) shows the effect that a concave kink point could in theory have on the density of itemizers. Panel (c) shows that if itemizers were responding to the concave kink point, we should observe that the pre and post reform densities are not overlapping just above the standard deduction. This is contradicted by figure 2a.

Figure 7: Cost Trends

(a) Total Costs

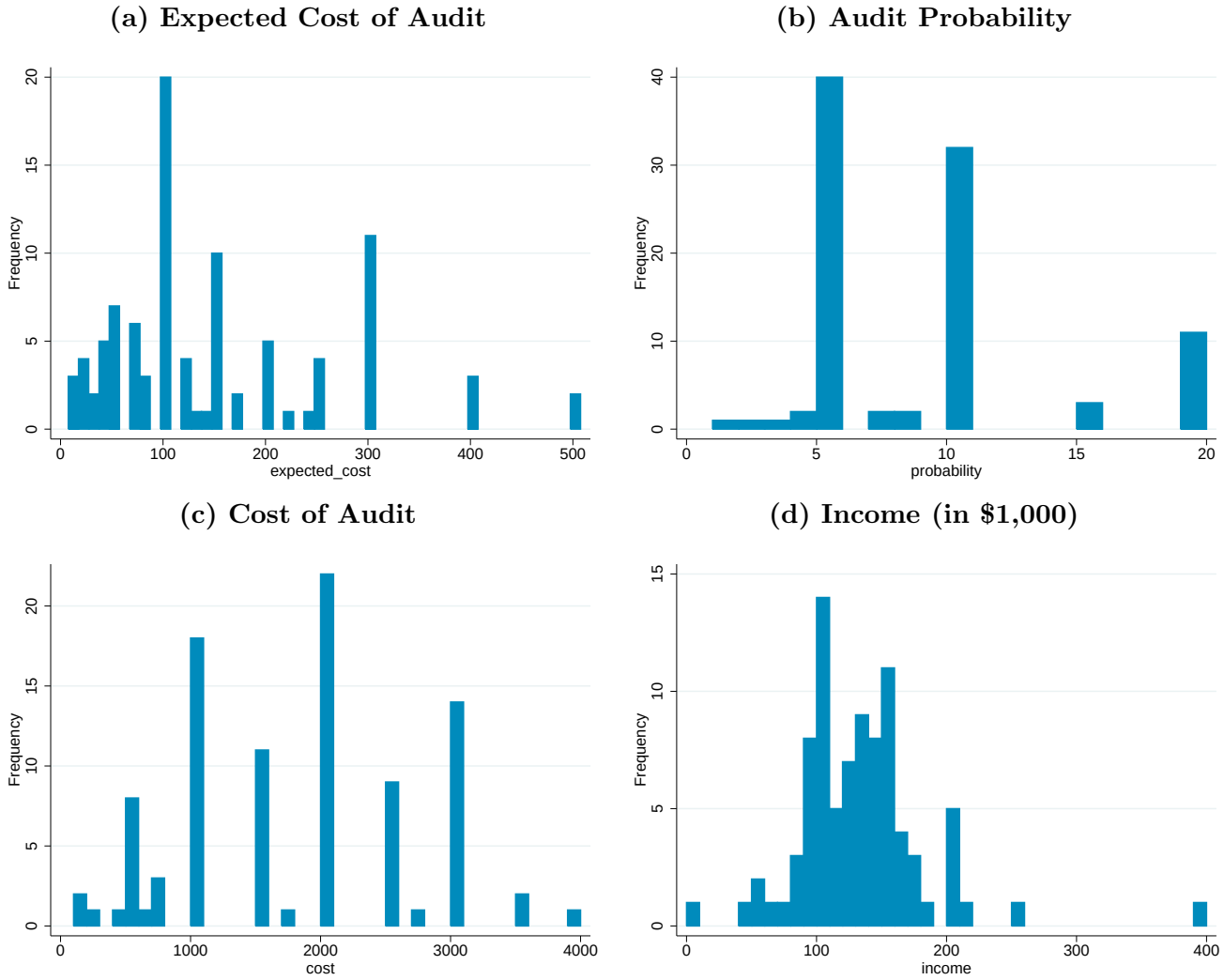


(b) Number of Forms Filed



Notes: The first figure plots the cost of filing each schedule for the total US population over time as estimated in equation (7). 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.

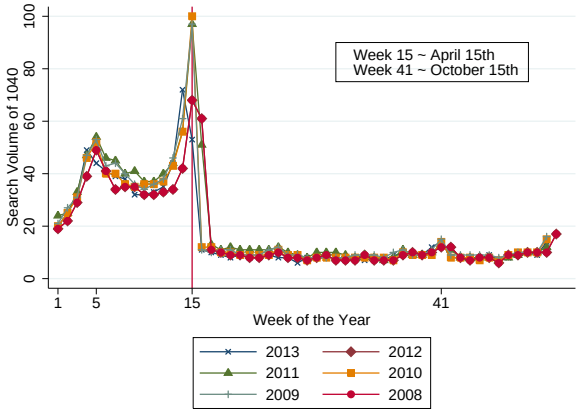
Figure 8: Audit Survey



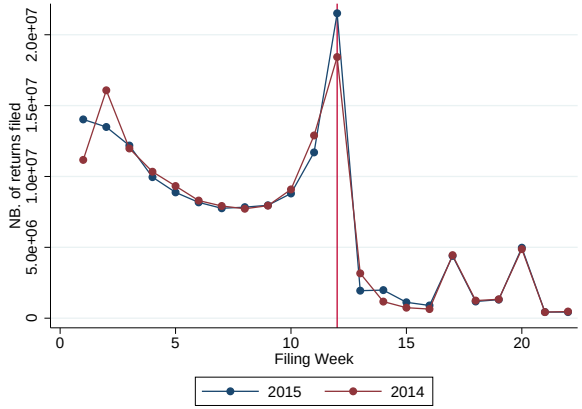
Notes: The number of observations for each panel is 95 individuals. Panel (a) is the distribution of expected cost of audit and is equal to the product of audit probabilities by cost of audit. Panel (b) is the distribution of perceived audit probabilities. Panel (c) is the distribution of cost of audit. Panel (d) is household income in brackets of $1,000. 195 individuals were surveyed, of which 95 files their taxes themselves and itemize deductions.

Figure 9: Deadline Effects

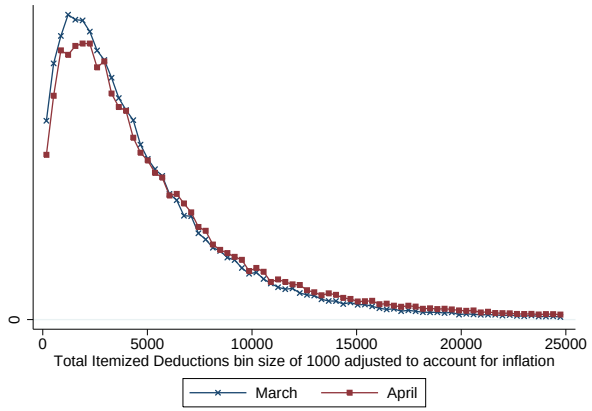
(a) Google Search of the Term 1040



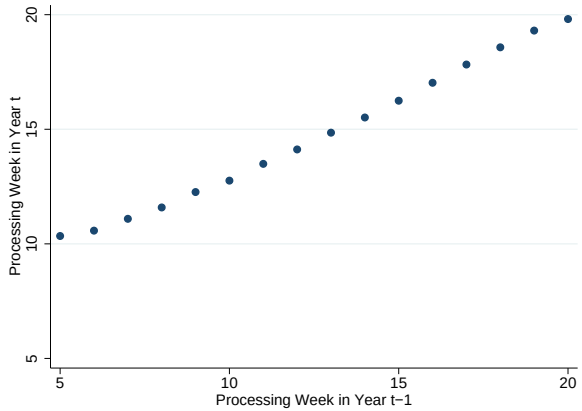
(b) Number of Returns Filed by Week



(c) March Itemizers v.s. April Itemizers



(d) Processing Week in Year t v.s. $t-1$



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.

Table 1: Cost Estimates

Filing Status	MTR	Cost as % of AGI	Cost in $
Single	15%	0.83	$175
Single	28%	0.85	$369
Joint	15%	0.57	$243
Joint	28%	0.74	$591
Head	15%	0.76	$270
Head	28%	0.72	$458

Notes: This table shows the estimates derived using the 1988 reform as explained in section 3.

Table 2: Other Tax Schedules and Demographics

Variable	Cost Coefficient
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 (6) and (7). β_{1040}, β_B, β_C, β_D, β_E, β_F and β_{SE} are estimated in section 6.2. Section 6.1 explains how α_{dep}, α_{efi} and α_{prep} are estimated. Section 4.1 explains how α_i, $i = 1, ..., 10$ are estimated.

Table 3: Standard Errors of the Difference Between the 1987 and 1989 Densities (figure 2a)

Bin	Deduction Range	Difference	Standard Errors	z-stat
1	[9991, 11991]	0.00311***	0.00047	6.55
2	(11991, 13991]	0.00190***	0.00044	3.47
3	(13991, 15991]	0.00000	0.00040	0.02
4	(15991, 17991]	-0.00047	0.00041	-1.13
5	(17991, 19991]	0.00022	0.00038	0.59
6	(19991, 21991]	-0.00010	0.00033	-0.31
7	(21991, 23991]	-0.00041	0.00028	-1.45
8	(23991, 25991]	-0.00042	0.00025	-1.67
9	(25991, 27991]	-0.00032	0.00020	-1.60
10	(27991, 29991]	-0.00042**	0.00018	-2.24

Notes: This table shows the bootstrapped standard errors for the difference between bins in figure 2a * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. I use 100 replications for the bootstrap estimation.

Table 4: Calibration of Rational Inattention Model

	Precision of Beliefs About Level of Savings (σ)							
	10	50	100	200	500	1000	2000	3000
CRRA coefficient								
0.1	0	0	0	1	5	28	70	152
0.25	0	0	0	2	11	44	167	203
0.5	0	0	1	4	22	86	64	625
0.8	0	0	1	5	35	134	462	880
1	0	1	2	7	44	164	547	1015
1.1	0	1	2	8	48	179	586	1074
1.25	0	1	2	9	54	200	640	1153
1.5	0	1	3	11	64	233	718	1262
1.8	0	1	3	13	76	270	799	1364
2	0	1	4	14	84	293	844	1417

Notes: This table shows the results of a calibration of the rational inattention model derived in section 5.4.

WEB APPENDIX FOR ONLINE PUBLICATION

A Pitt and Slemrod (1989)

Pitt and Slemrod (1989) very elegantly apply the methods of Gronau (1973) and Nelson (1977) to assess the compliance cost of itemizing deductions by estimating a censored model with unobserved censoring thresholds using maximum likelihood.

To do so they estimate a cost and benefit function of itemizing deductions. The benefit of itemizing is given by $TS_i = X_i\beta + u_i$ where X_i are exogenous and observed characteristics, β is a vector of parameters and u_i an error term. Similarly, the cost of itemizing is assumed to be $C_i = Z_i\gamma + v_i$, where Z_i are exogenous and observed characteristics, γ a vector of parameters and v_i an error term. A person will itemize if $TS_i \geq C_i$. TS_i is only observed when $TS_i \geq C_i$ but C_i is never observed. Gronau (1973) and Nelson (1977) show that if u_i and v_i are uncorrelated or if there are some characteristics present in X_i but not in Z_i then the model is identified and a likelihood function can be maximized to estimate both TS_i and C_i. Pitt and Slemrod (1989) acknowledge that there is no reason to assume that the errors are uncorrelated but that there are some characteristics that are likely to be present in X_i but not in Z_i, therefore arguing that identification should be valid.

The set of exogenous and observable characteristics they consider to estimate both β and γ are whether a person is married, her AGI, the square of AGI, whether a person owns a farming business, the number of age exemptions a person claims and the number of exemptions claimed. The set of exogenous characteristics specific to β are positive investment income, the average state income and sales taxes for an income of \$40,000, the average property tax rate in a given state and an index of medical costs in a given state.

If the assumptions from Gronau (1973) and Nelson (1977) hold and given these exogenous and observed characteristics, they can estimate the cost and benefit function. They find that the average cost of itemizing is \$107 (in 2016 dollars), i.e., 6 times lower than the cost I estimate.

Since Pitt and Slemrod (1989) acknowledge that u_i and v_i are likely to be correlated, for the Gronau (1973) and Nelson (1977) estimators to be consistent, the exclusion restriction imposed on X_i and Z_i becomes necessary for identification.

B Assumption A2

Assumption A2 states that the cost should not increase with the level of deductions. It makes sense to assume that the cost of deducting \$10,000 worth of mortgage interest is the same as deducting \$100,000 because total mortgage interest is reported on form 1098. However, it is also reasonable to assume that an individual who donates \$100,000 to charity is more likely to donate to more charities than an individual who donates \$10,000.

Assumption A2 is important for equation 3. Intuitively, it allows me to infer the distortion imposed by the standard deduction on the pre-reform distribution in bin j from bin $j+m$ when the pre- and post-reform standard deduction thresholds are m bins away. A2 can fail if the cost of itemizing decreases with the size of total deductions which would bias my cost estimate downwards. But more importantly it can fail if the cost of itemizing increases with the size of total deductions, which would overestimate the cost. There is an easy way to provide an upper bound for the bias introduced by a failure of A2: by using the pre-reform distribution $g_\delta(d)$ as the true counterfactual instead of $f(d)$. This is a generous upper bound because it assumes that the pre-reform distribution is undistorted just above the standard deduction in spite of figure 2a showing a clear distortion. In this case, the estimated cost would be \$519 instead of \$591. Therefore if A2 fails, the cost of itemizing would lie between \$519 and \$591.

C Sample Restrictions

C.1 Figure 1

The sample used for figure 1 are joint filers who itemize deductions. I focus on joint filers because they represent more than 50% of the population and the standard deduction is specific to the filing status. This means that I cannot show every tax filing status on the same graph because they would have different standard deductions. Figure H.15 shows the same patterns for single taxpayers.

C.2 Figures 2a, 2b

In figure 2a and 2b, I focus on taxpayers who are married filing jointly for the reasons outlined in section C.1. In addition, in 1988 and 1989 there were two tax brackets (15% and 28%) and a tax rate "bubble" (33%). Most taxpayers who itemize deductions fall in the 28% marginal tax bracket. Therefore, to control for the change in marginal tax rates, I only consider taxpayers who fall in the 28% marginal tax rate bracket. This allows me to precisely calculate the amount of after tax forgone benefit.

C.3 Figure 4

In figure 4, I use the same sample restrictions as in figure 2a and 2b and break down the sample into deciles of income.

C.4 Figure 5

To generate figure 5, I consider joint filers as explained in section C.1. In figure (a), I consider all years from 1980 to 2006 but exclude 1985 and 1990 because the tax preparer variable is missing in those years. In figure (b), I consider all years from 1998 to 2006 because few taxpayers used electronic filing prior to 2006.

C.5 Figure 9c

The variable indicating the week in which a return is processed by the IRS is only present in the SOI files in year 1980 to 1999. Thus, to generate figure 9c, I restrict attention to those years. I use the same sample restrictions as in figure C.1 in addition to dropping taxpayers who have a balance due to the IRS. If taxpayers owe money to the IRS, it is rational to wait as much as possible so as to save on interest.

C.6 Week of Filing Variable

The SOI files contain a variable that indicates the week in which a return is processed by the IRS. Slemrod et al. (1997) have access to the internal IRS files that record the filing date and compare it to the processing date from the SOI files. They find that the order in which returns are processed matches the order in which they are filed. Knowing the order is sufficient for my purposes because what I am interested in is comparing taxpayers who file close to the deadline to those who file earlier. I can therefore use the processing time variable to identify late filers and verify the predictions of the naive present bias model. The IRS

promises that returns are processed within 6 weeks. This constraint is likely to be binding for returns that are filed close to the deadline given that a lot of returns are processed at the time. Therefore, I assume that the processing time has a lag of 6 weeks.

I restrict the sample used to generate this graph to taxpayers who are owed refunds by the IRS and who do not have to file any other schedule but Schedule A. This allows me to rule out taxpayers who rationally delay filing to save on interest on the amount they owe to the IRS and taxpayers who cannot file early because others schedules sometimes require additional paperwork that only becomes available later in the year.

C.7 Taxpayers Who Have To Claim the Standard Deduction

In rare cases, taxpayers have to claim the standard deduction even when their itemized deductions exceed the standard deduction. These individuals are dropped from my sample. This happens in the following four cases:

1. A married taxpayer whose spouse files separately and itemizes deduction.
2. In some states, a taxpayer who wants to itemize on her state tax return has to itemize on her federal tax return as well.
3. A taxpayer who is neither a citizen nor a permanent resident of the United States.
4. A taxpayer who can benefit from itemizing for alternative minimum tax purposes even though the standard deduction is greater than the sum of her itemized deductions.

D Tax Reform Act of 1986 and Lagged Responses

Could there be any other exogenous variation altering the distribution of itemized deductions in 1989 affecting my main identification strategy? The majority of tax reforms happened following the TRA'86 and were enacted in 1987. Among those, there were some deduction reforms. Because I am comparing 1987 to 1989, I am implicitly controlling for the Tax Reform Act of 1986 (TRA'86) reforms. But there might be slow adjustments and lagged responses in 1988 or 1989. To rule these out, I consider all the reforms enacted by TRA'86 that could affect the level of deductions and show that it is reasonable to assume that the adjust-

ment is immediate. Because all of the reforms reduced the amount of eligible deductions, they have no lagged response. To see this consider a hypothetical example: assume the charitable donation deduction is capped at $10,000. A taxpayer who was donating $15,000 will now only be able to deduct $10,000. Will the taxpayer reduce her donations? She might reduce them up to $10,000 but there is no reason to expect that she will reduce them any further. What does this imply for the level of deductions? We should observe a drop in deductions to $10,000 in 1987 and then *no further* drop in 1988 or 1989, ruling out any lagged responses. Since I am comparing 1987 to 1989, any reform that caps the amount of deductions should not affect my estimates. The deduction reforms enacted in 1987 are the following (source: IRS):

- Prior to 1987, medical deductions in excess of 5% of the AGI are deductible. In 1987, this threshold is increased to 7.5% of AGI, further limiting the allowable amount of medical deductions. There is no reason to assume that there will be a slow adjustment that spills over into 1988 or 1989 in this case.
- Sales taxes are not deductible anymore. For similar reasons, one should observe a drop in the total deductions in 1987 as sales taxes were a large portion of it but there should be no lagged effect.
- The home mortgage interest deduction is subject to a new limit. The home mortgage interest deductions for a given year are capped at the value of one's house (plus renovations). Anything in excess of the value of the house have to be deducted as personal interest for which only 65% of the total value can be deducted. First, the IRS estimated that very few taxpayers were affected by this reform since it is very rare that one's home mortgage interest in one given year exceeds the total value of one's house. Second, there is no reason to expect a drop in levels in the subsequent years. If a person is affected by this reform, in 1987 she will be forced to claim less deduction than she was previously claiming.
- Any interest for home mortgages in excess of 1 million dollars is not deductible anymore. Again, there is no reason to expect any lagged effects due to this reform because it caps the amount of deductions.

There are no other reforms affecting directly or indirectly the amount of item-

ized deductions an individual can qualify for.

E The 1971, 1975 and 2003 reforms

E.1 The 1971 and 1975 reforms

In 1970 and 1975 taxpayers could claim as a standard deduction the smaller of the standard deduction or 10% of their income. In 1971, both thresholds were increased to \$8,809 or 13% of income if income is greater than \$46,983, and the larger of \$6,166 or 13% of income for taxpayers with income smaller than \$46,983. In 1975, a similar two tiered standard deduction existed with an AGI limit of 16% and a dollar limit of \$74,431.

If I were to only look at the density of itemizers above \$6,130 in 1970 and compare it to the density of itemizers above \$8,809 in 1971, my estimates would be biased because some taxpayers who have deductions greater than \$8,809 in 1971 are likely to stop itemizing – not because of compliance costs – but only because their deductions are now smaller than 13% of their income. This is why using 1971 and 1975 will not yield accurate estimate of compliance costs (they tend to over-estimate them).

E.2 The 2003 reform

Two main changes occurred in 2003 that affect the post-reform standard deduction. The first one is that tax rates were reduced 2 to 3 percentage points (depending on the bracket), reducing the incentive to itemize. The second one is that electronic filing was rapidly expanding in the early 2000's complicating the comparison between the pre and post-reform standard deduction.

F Audit Survey

The survey was carried outside a health food supermarket in Santa Monica, California. The location was chosen to attract as many wealthy individuals as possible to increase the proportion of itemizers. 195 individuals were surveyed of which 114 file their taxes themselves. Of those 95 itemize deductions, which constitutes the final sample. They were asked the following questions:

1. Do you file taxes yourself?

2. Do you itemize deductions or claim the standard deduction?

3. Per year, what do you think the chances of being audited are?

4. Assume the IRS wants to audit you. What is the highest amount you would pay a lawyer that would deal directly with the IRS and prevent you from being audited?

5. What is the annual income of your household? (Brackets of $1,000)

G Rational Inattention

Could taxpayers forgo large amounts of deductions because they are uncertain of whether their total deductions are larger than the standard deductions threshold?

Most of the deductions are relatively stable from year to year as they mostly consist of items that vary very little such as mortgage payments, real estate taxes or state income taxes. This means that taxpayers should have an accurate signal of their true deductions. In addition, the expenses associated with deductions are an active decision: if deductions increase or decrease by a large percentage, taxpayers are likely to be aware of this change because they caused it.

Therefore, for rational inattention to explain the magnitude of the estimated compliance costs, one would need to assume that taxpayers receive a very noisy signal which is unlikely given that deductions vary little from year to year. I formalize this argument in what follows:

Assume that the taxpayer has a Constant Relative Risk Aversion (CRRA) utility function given by $U(x) = \frac{1}{1-\theta}x^{1-\theta}$ if $\theta \neq 1$ and $U(x) = \log(x)$ if $\theta = 1$.

Denote by τ the after tax amount of deductions the taxpayer can claim (deduction multiplied by marginal tax rate) and by S the after tax amount of the standard deduction. Assume that the taxpayer has beliefs over τ that follow a normal distribution with mean μ and standard deviation σ. Denote by c the cost incurred by the taxpayer to calculate the total amount of deductions τ. The cost is only incurred when she itemizes, not when she claims the standard deduction.

The taxpayer will decide to itemize if the expected benefit from itemizing given her beliefs over τ exceeds the cost of figuring out the level of τ i.e. c. This occurs

when the following equation is satisfied:

$$\mathbb{E}\left[\frac{1}{1-\theta}(\tau - c)^{1-\theta}\right] \geq \frac{1}{1-\theta}S^{1-\theta}. \tag{8}$$

This equation does not have a closed form solution, so I use a Taylor expansion of second degree around the mean of $\tau - c$, as follows:

$$\frac{1}{1-\theta}(\mu - c)^{1-\theta} - \frac{1}{2}\theta(\mu - c)^{-1-\theta}\sigma^2 \geq \frac{1}{1-\theta}S^{1-\theta}. \tag{9}$$

And for $\theta = 1$, it is equal to:

$$\log(\mu - c) - \frac{\sigma^2}{2(\mu - c)^2} \geq \log(S). \tag{10}$$

The first term in equation 10 is the expected benefit that the taxpayer derives from itemizing. The second term is a correction for the risk aversion of the taxpayer: she will itemize deductions if the benefit of itemizing corrected for her risk aversion is greater than the benefit she derives from itemizing. Holt and Laury (2002) find a θ that ranges between -0.95 and 1.37. I assume here that $\theta = 1$ but also consider $0 < \theta \leq 2$[35] in table 4. I fix the standard deduction at \$10,000 for joint filers. The cost estimated by the IRS of the time required to itemize deductions is $c = 149$. I can calculate a lower bound on the standard deviation of the taxpayer's beliefs over τ (σ). Using these parameters, I find that for rational inattention to explain the magnitude of the forgone benefits, the standard deviation of after tax deductions σ has to be greater than \$1,814 (which corresponds to \$6,479 worth of deductions with a 28% marginal tax rate). This means that the taxpayer has a range of uncertainty of deductions of more than \$6,479. This implies very high uncertainty in the beliefs of the benefits that the taxpayer can save from itemizing which is unlikely given that deductions are relatively stable from year to year as they are mostly constituted of mortgage payments and state taxes and are the results of active decisions. If a taxpayer's total deductions were to increase or decrease dramatically, she would most likely know about it because it would be due to for example to large income variations, the take up

[35]Negative values of θ are not considered because they imply risk lovingness and would trivially reject rational inattention.

of a mortgage etc. which are salient.

If I assume a standard deviation of $\sigma = 200$ – which corresponds to a standard deviation of deductions of \$714 – then rational inattention with $\theta = 1$ predicts that taxpayers would claim the standard deduction up to total deductions of \$10,557 and forgo an average of \$557 worth of deductions, i.e., \$156 of after tax dollars given a cost c=\$149. With reasonable parameters, rational inattention predicts that taxpayers will forgo an additional \$7 in excess of the cost of \$149.

H Time Inconsistency: Model

I assume that the cost of record keeping continuously increases for every day that the receipt is not archived as soon as it is received. When the taxpayer is issued a receipt for a charitable donation and fails to archive it, the cost of keeping track of this receipt increases continuously because it is more likely to be lost or it could take more time to look for it. The rational taxpayer archives the receipt as soon it is issued. The naive present-biased taxpayer plans on archiving the receipt but fails to do so, leading to high record keeping costs.

Assume for simplicity that the taxpayer only needs to itemize one deduction for example for a charitable contribution she made. The taxpayer is facing two distinct costs when considering the decision to itemize deductions. The first one is that of record keeping, denoted here by c. The second one is filling out Schedule A itself which is denoted by k.

If the taxpayer succeeds in performing the two tasks she receives a one time benefit b in the subsequent period. Once the taxpayer gets the receipt for her charitable contribution, she can decide to archive it immediately by incurring a cost c or archive it later and incur a larger cost $c(1+r)$ next period where r is the rate at which the cost of record keeping grows if the receipt is not archived.

δ is the time-discount factor, β the present-bias parameter, t the period in which the record keeping is performed and Schedule A is filled out and $(t+1)$ the period in benefit b is received.

In what follows, I use two definitions:

Definition 1: For given β, δ, c, k, $(1+r)$ and t a task is said to be β-worthwhile if $-c(1+r)^t - k + \beta b > 0$.

Similarly:

Definition 2 For given δ, c, k, $(1+r)$, and t a task is said to be δ-worthwhile

if $-c(1+r)^t - k + \delta b > 0$.

The rational taxpayer has a standard utility function where per-period utility is discounted by δ in the future.

The decision to itemize or claim the standard deduction for the rational taxpayer can be written as follows:

$$\max_t \delta^t(-c(1+r)^t - k + \delta b),$$

conditional on itemizing being δ-worthwhile.

Cost c is incurred as soon as the taxpayer starts the record keeping. If she waits an additional t periods before archiving the receipt, the cost of record keeping is multiplied by $(1+r)$ for every additional period i.e. $(1+r)^t$ overall. Therefore, to minimize the cost of record keeping, the rational taxpayer will choose $t = 0$, this means that she will archive the receipt as soon as it is received and will incur a record keeping cost of c rather than $c(1+r)^t$.

The taxpayer is left with choosing t such that:

$$\max_t \delta^t(-c(1+r)^t - k + \delta b)$$

Assume the taxpayer is contemplating the decision to perform the record keeping task in the first period yielding utility: $-c - k + \delta b$. She will only perform it if $-c - k + \delta b > 0$. And if she waits an additional period she will receive $\delta(-c(1+r) - k + \delta b)$, which is smaller than the utility she would have enjoyed if the task had been performed in the first period. This means that the rational taxpayer will either archive the receipt immediately or never archive it because she does not plan on itemizing her deductions.

The naive present biased taxpayer can perform the record keeping in period t or can wait and perform it in period $t+1$. She will prefer performing it in period $t+1$ if the following inequality is satisfied:

$$-c(1+r)^t - k + \beta b < \beta[-c(1+r)^{t+1} - k + b].$$

This inequality simplifies to:

$$-c(1+r)^t - k < \beta[-c(1+r)^{t+1} - k]. \tag{11}$$

A sufficient condition for equation 11 to hold is:

$$(1+r)\beta < 1. \tag{12}$$

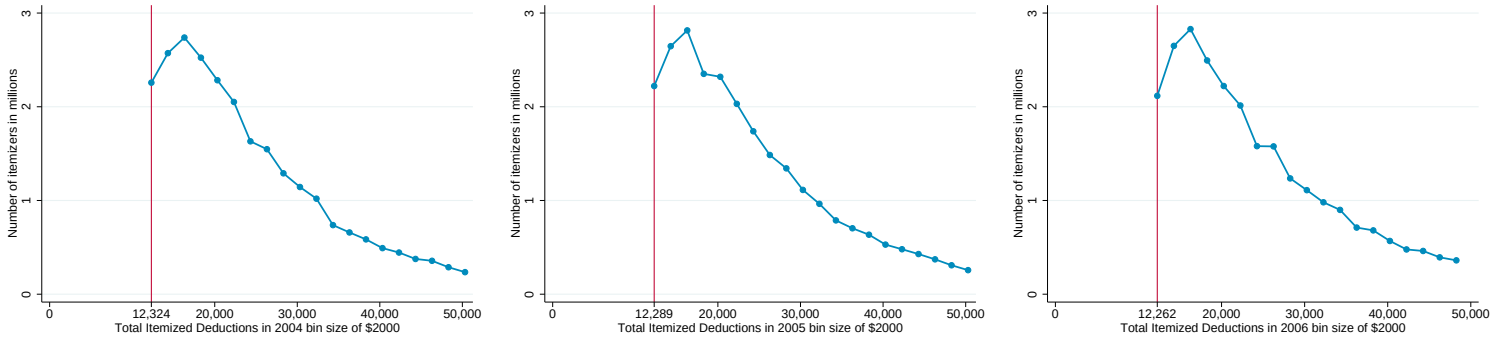
Intuitively, for the naive present-biased taxpayer to procrastinate on archiving her receipt, it is sufficient that the rate at which the record keeping cost increases be smaller than the rate at which she discounts the future.

Provided that condition 11 holds in period $t = 0$, it will also hold in any subsequent period $t > 0$ i.e. if itemizing is worthwhile but not performed in the very first period, the taxpayer will procrastinate until she reaches the deadline.

Testable Prediction 1: Naive present-biased taxpayers will file their returns at the deadline of April 15th when condition 11 holds.

Testable Prediction 2: The cost of record keeping for naive present-biased taxpayers is greater than for rational ones. This predicts that taxpayers who file close to the deadline are likely to forgo more deductions.

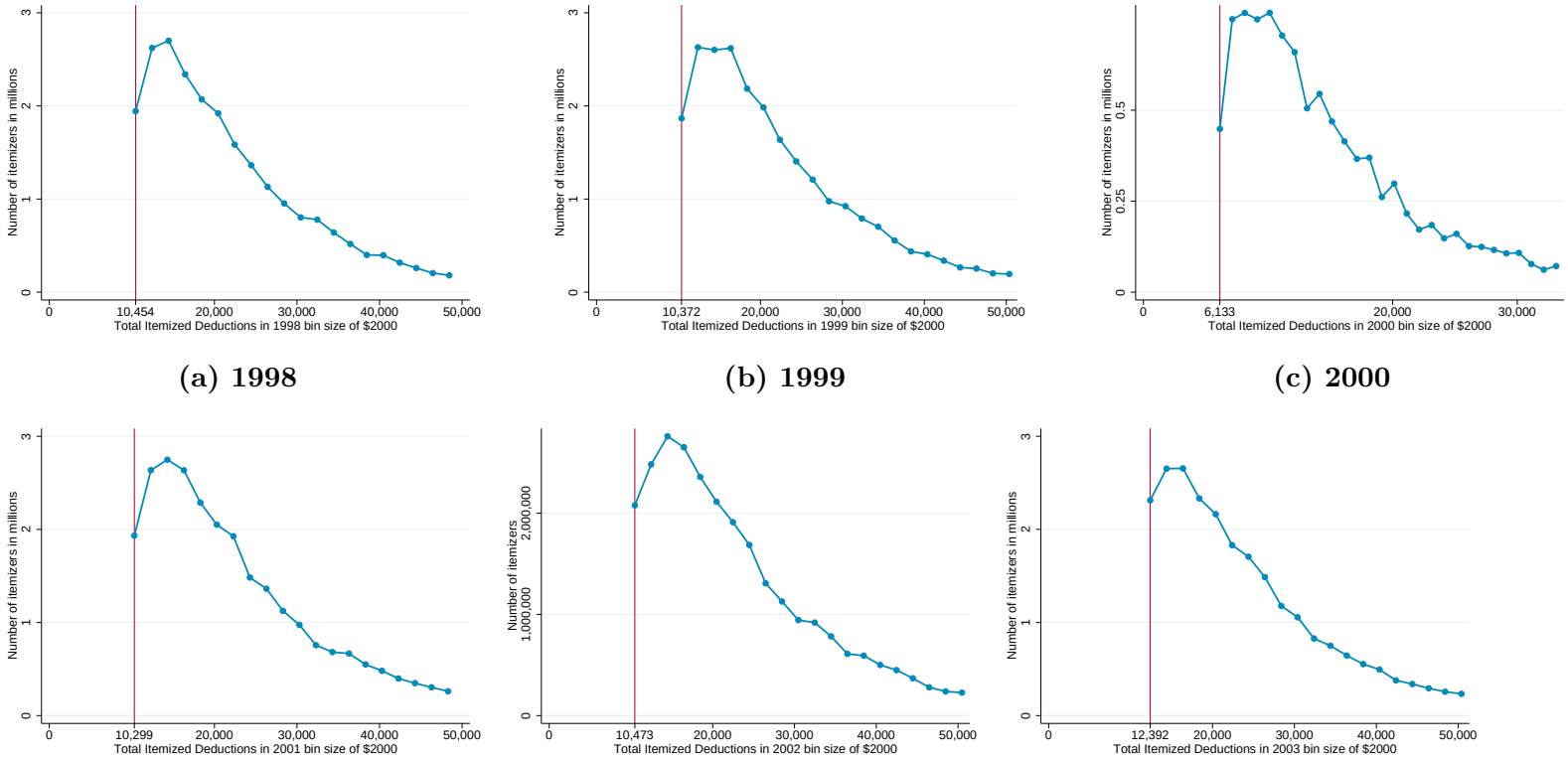
Figure H.10: Missing Mass Just Above the Standard Deduction 2004-2006 (Joint Filers)



(a) 2004 **(b) 2005** **(c) 2006**

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.

Figure H.11: Missing Mass Just Above the Standard Deduction 1998-2003 (Joint Filers)

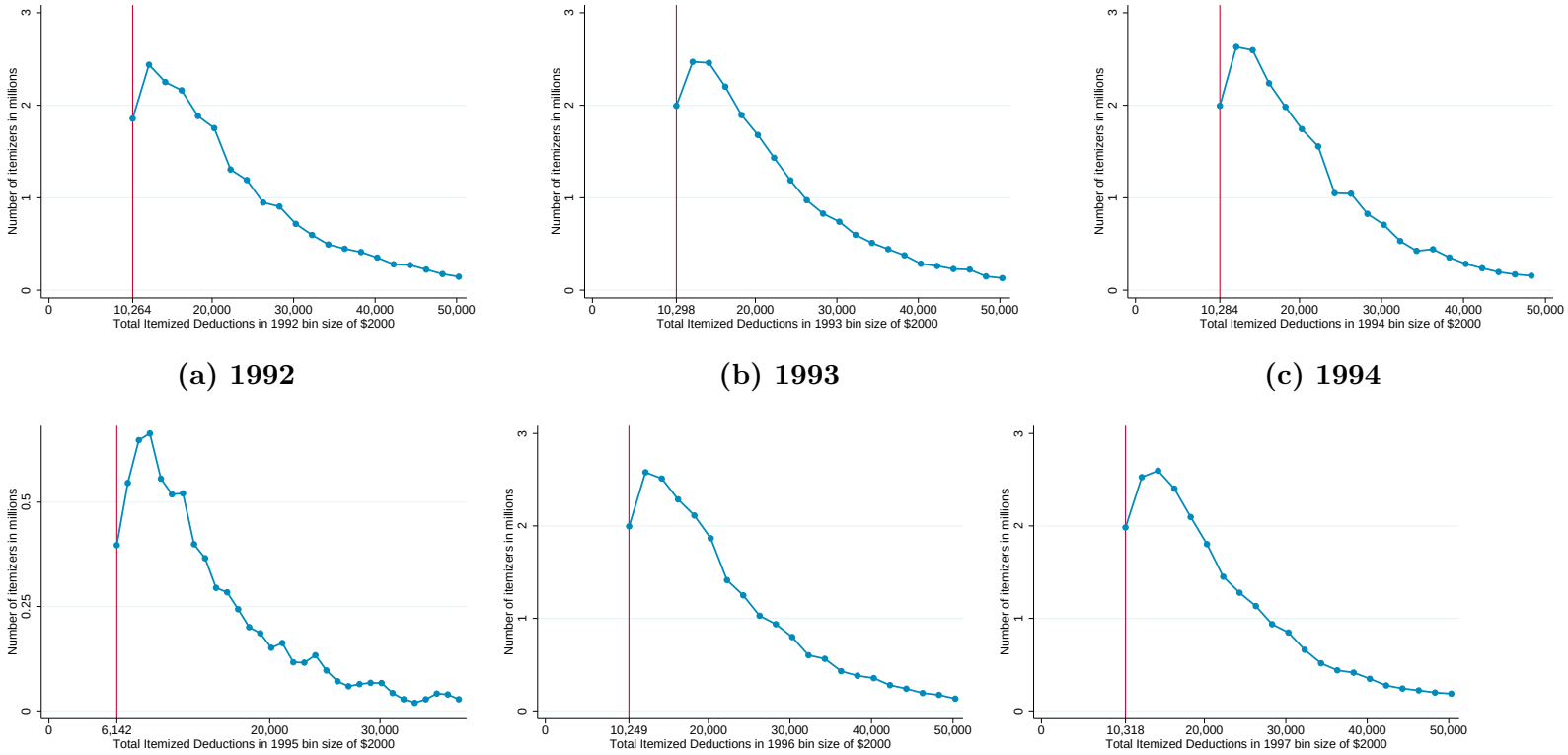


(a) 1998 **(b) 1999** **(c) 2000**

(d) 2001 **(e) 2002** **(f) 2003**

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.

Figure H.12: Missing Mass Just Above the Standard Deduction 1992-1997 (Joint Filers)

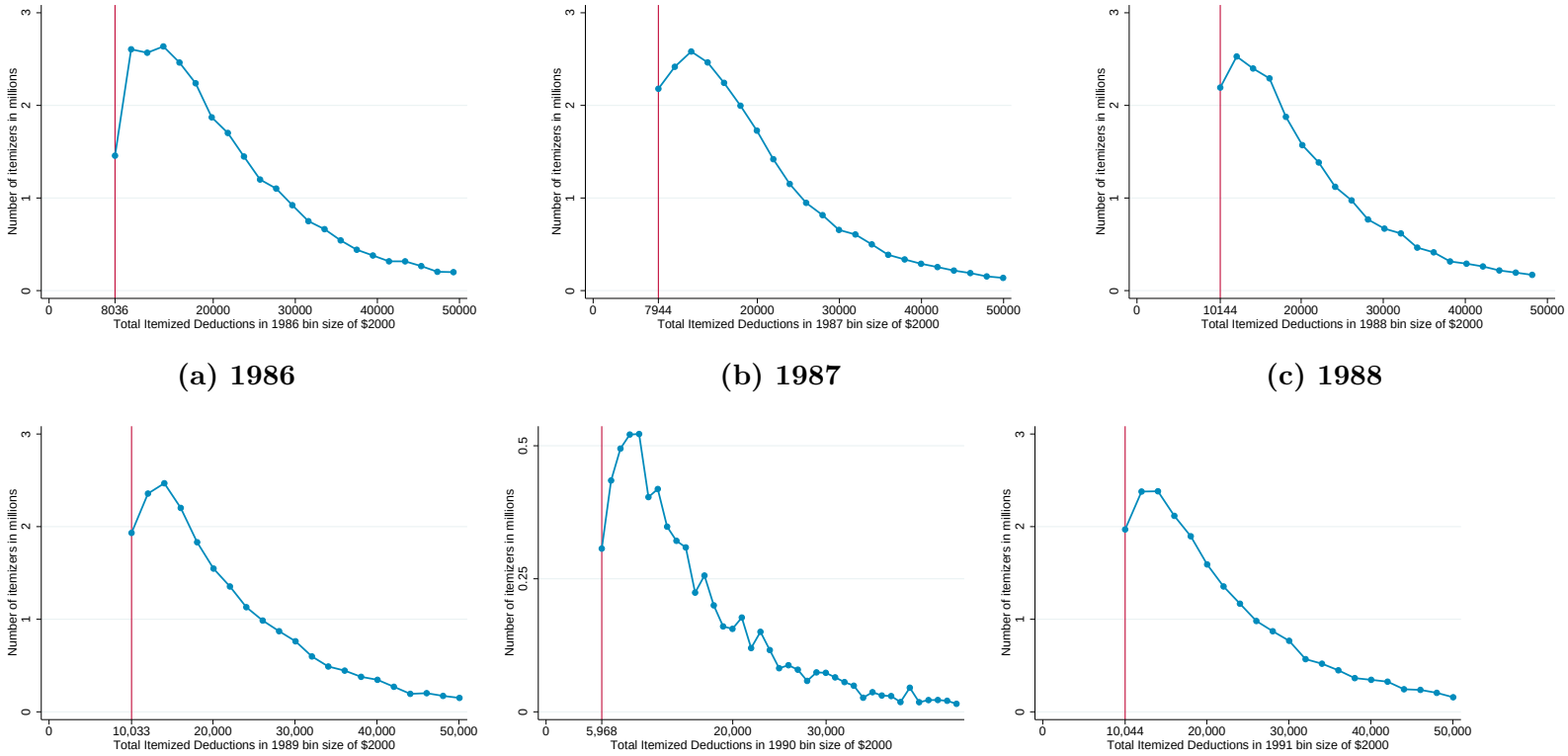


(a) 1992 **(b) 1993** **(c) 1994**

(d) 1995 **(e) 1996** **(f) 1997**

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.

Figure H.13: Missing Mass Just Above the Standard Deduction 1986-1991 (Joint Filers)

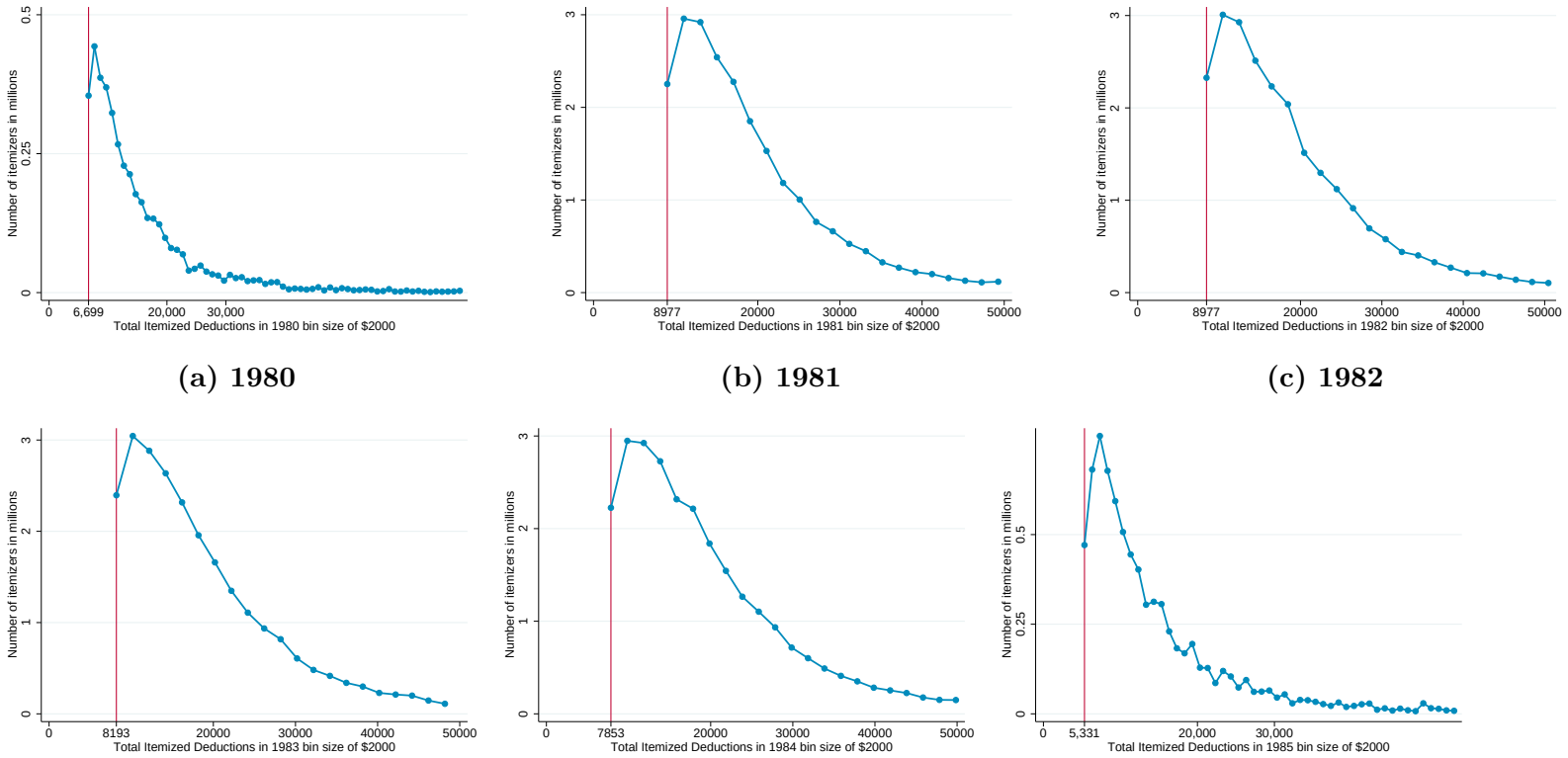


(a) 1986 **(b) 1987** **(c) 1988**

(d) 1989 **(e) 1990** **(f) 1991**

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.

Figure H.14: Missing Mass Just Above the Standard Deduction 1980-1985 (Joint Filers)

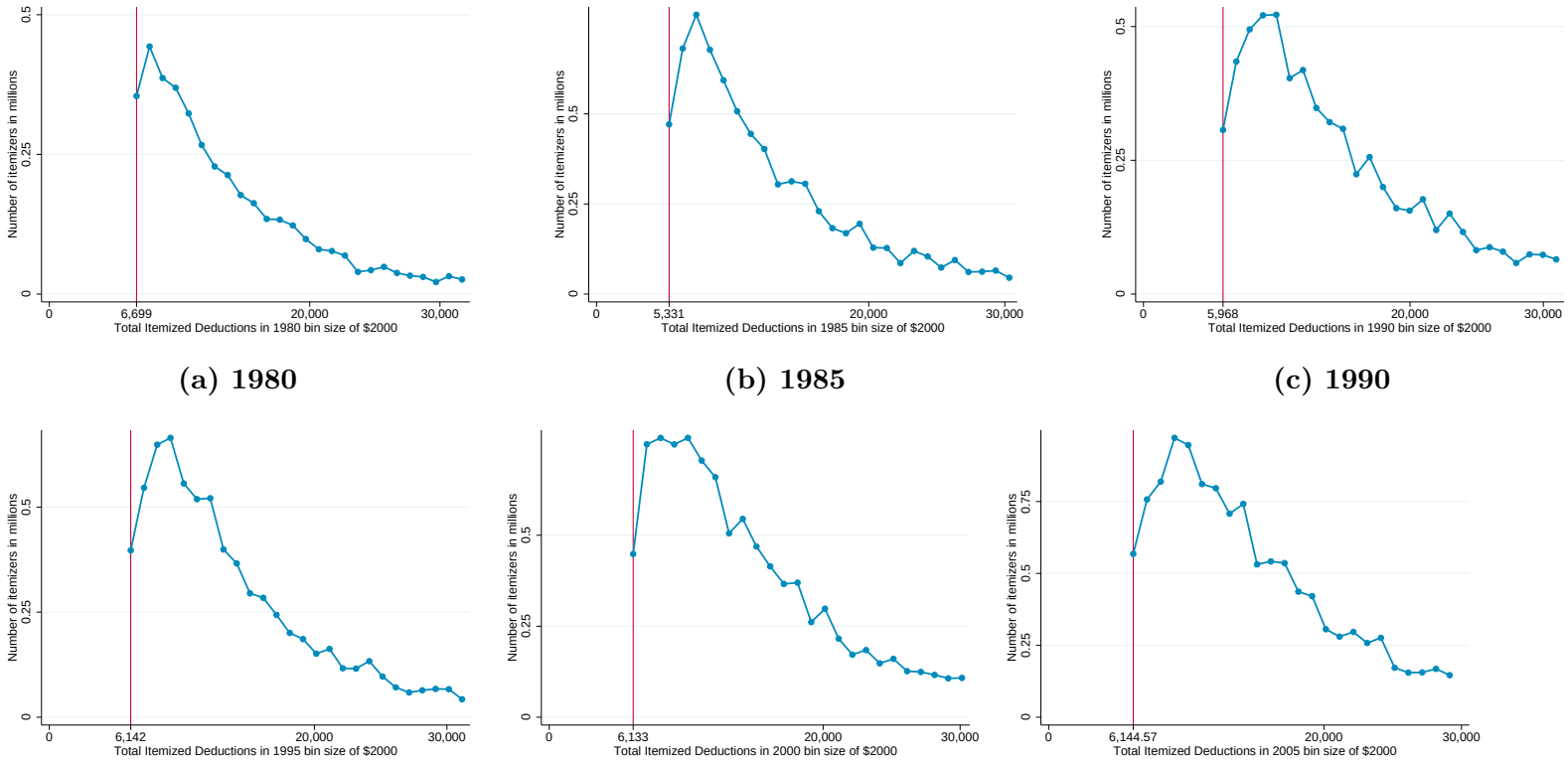


(a) 1980 **(b) 1981** **(c) 1982**

(d) 1983 **(e) 1984** **(f) 1985**

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.

Figure H.15: Missing Mass Just Above the Standard Deduction (Single Filers)



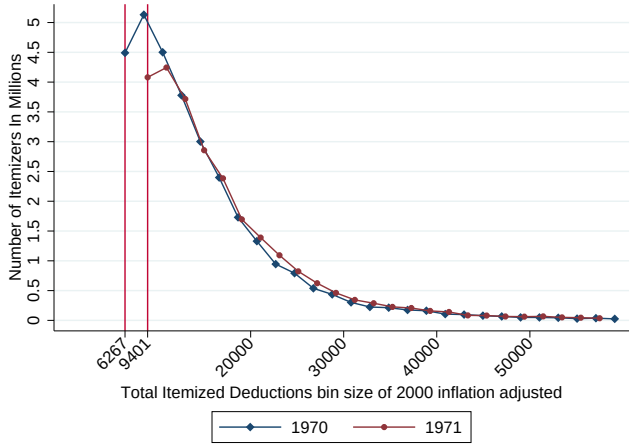
(a) 1980 **(b) 1985** **(c) 1990**

(d) 1995 **(e) 2000** **(f) 2005**

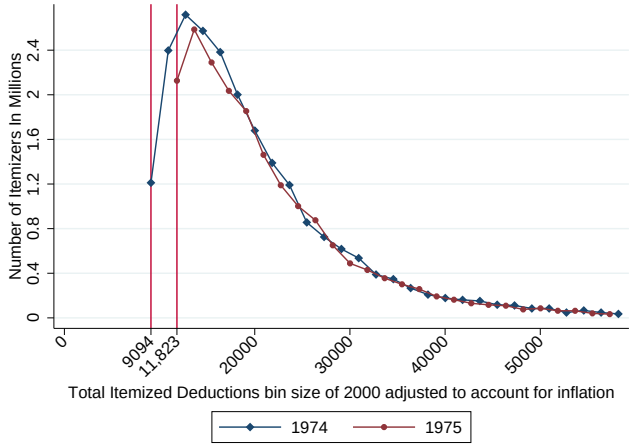
Notes: The figures above plot the density of deductions for single filers who itemize deductions. The bin size is $2,000 and the vertical line represents the standard deduction threshold for each year.

Figure H.16: Reduced Form Evidence of the Existence of Compliance Costs

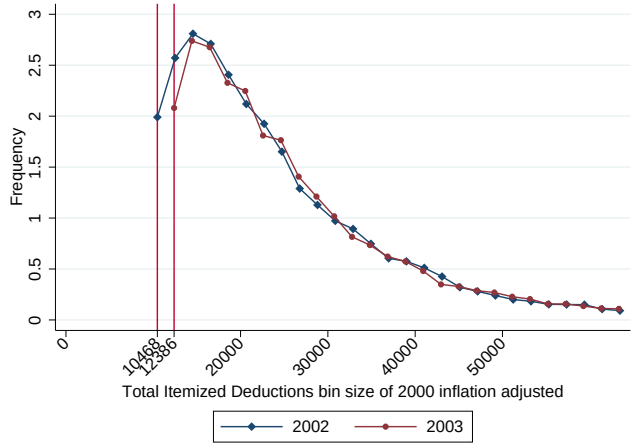
(a) 1970-1971 Comparison



(b) 1974-1975 Comparison

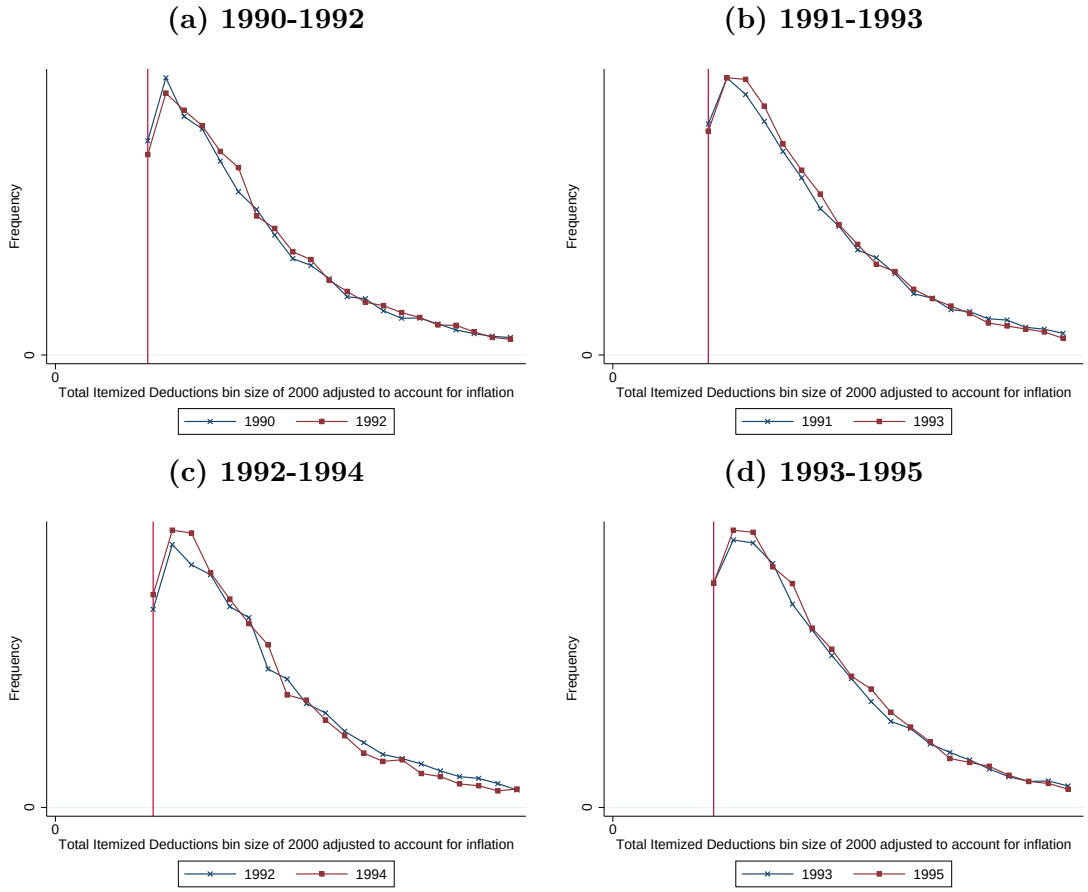


(c) 2002-2003 Comparison



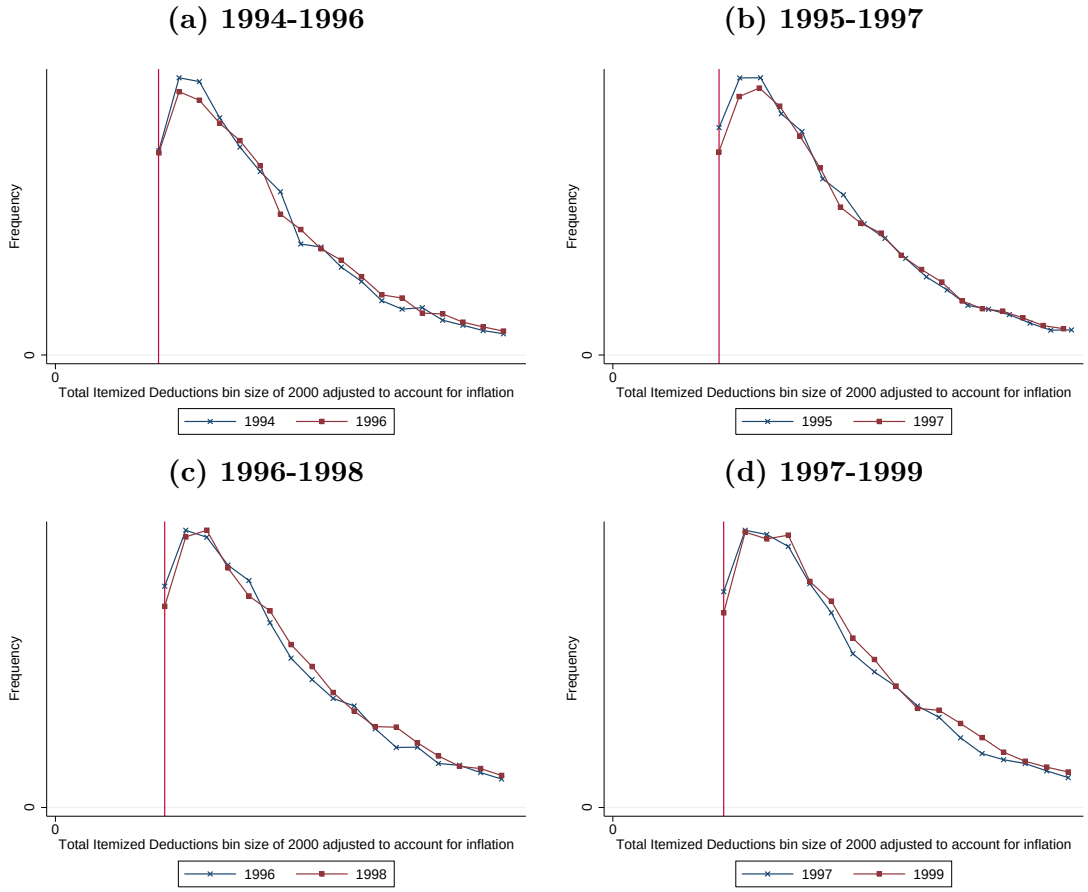
Notes: These graph plot the density of deductions before and at the time of the 1971, 1975 and 2003 reforms. While these show reduced form evidence of the existence of compliance costs, they do not provide accurate estimates of these compliance costs because other changes occurred at the same time.

Figure H.17: Placebo Test: Overlapping Densities In Years With No Reforms



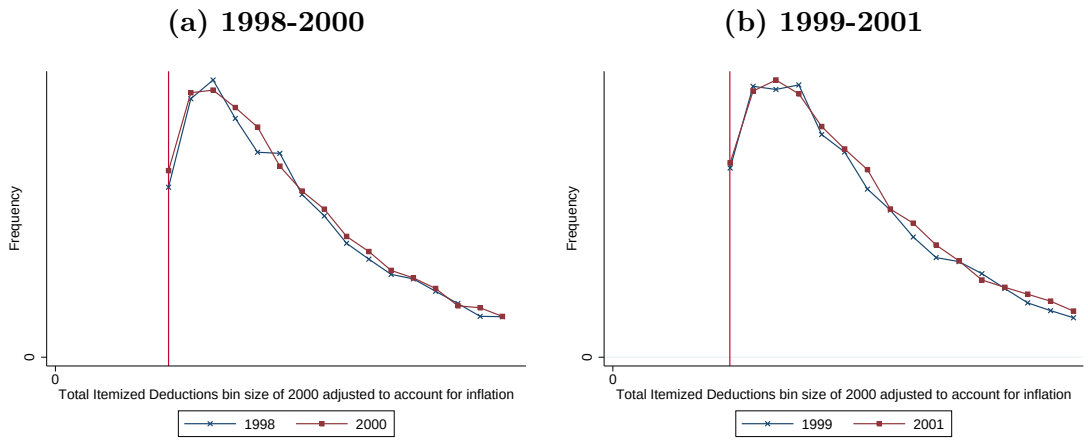
Notes: The figures above test assumption A1 which states the cost of itemizing does not vary from year to year.

Figure H.18: Placebo Test: Overlapping Densities In Years With No Reforms



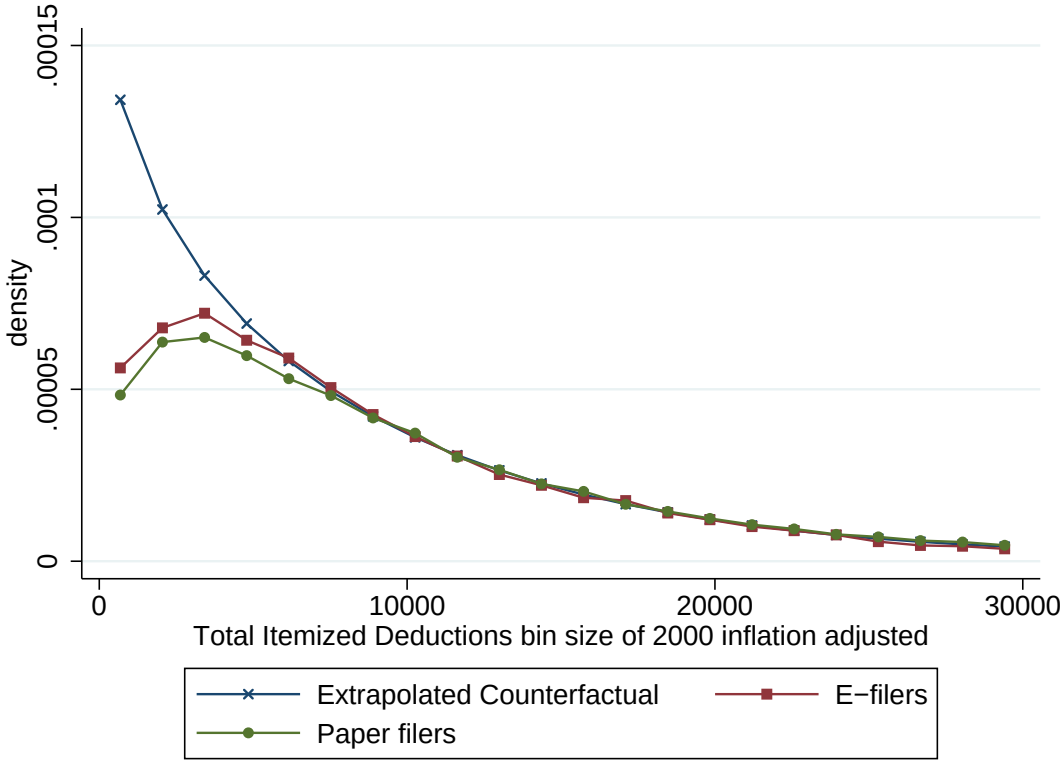
Notes: The figures above test assumption A1 which states the cost of itemizing does not vary from year to year.

Figure H.19: Placebo Test: Overlapping Densities In Years With No Reforms



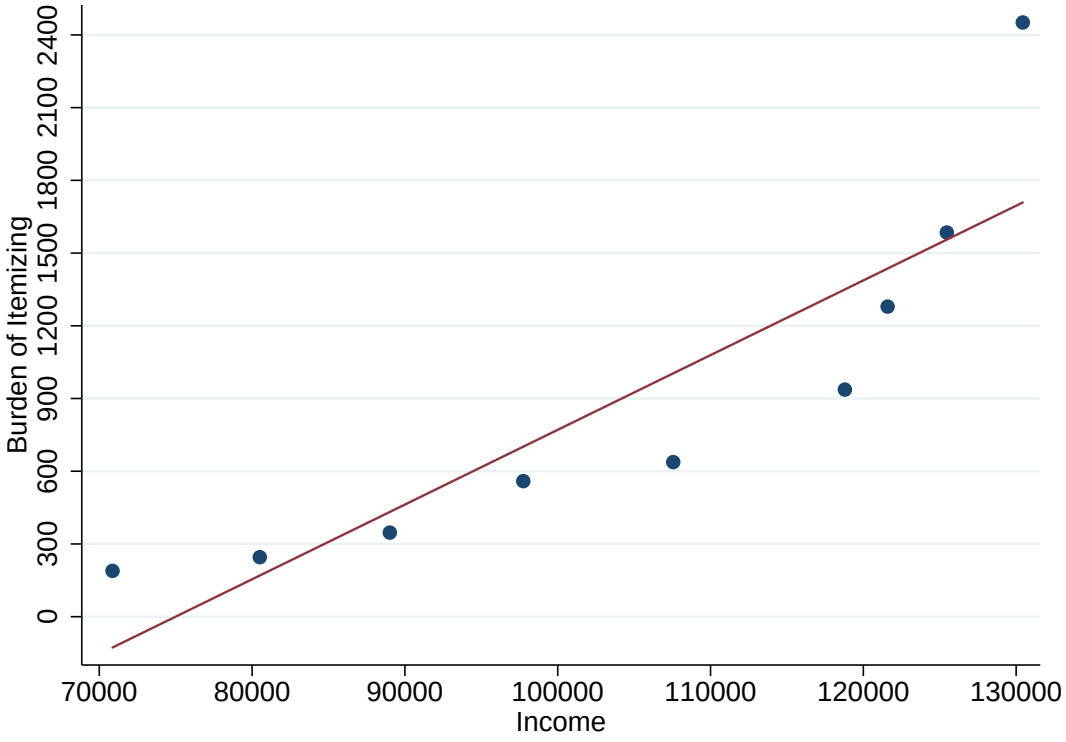
Notes: The figures above test assumption A1 which states the cost of itemizing does not vary from year to year.

Figure H.20: Effect of Electronic Filing on Cost



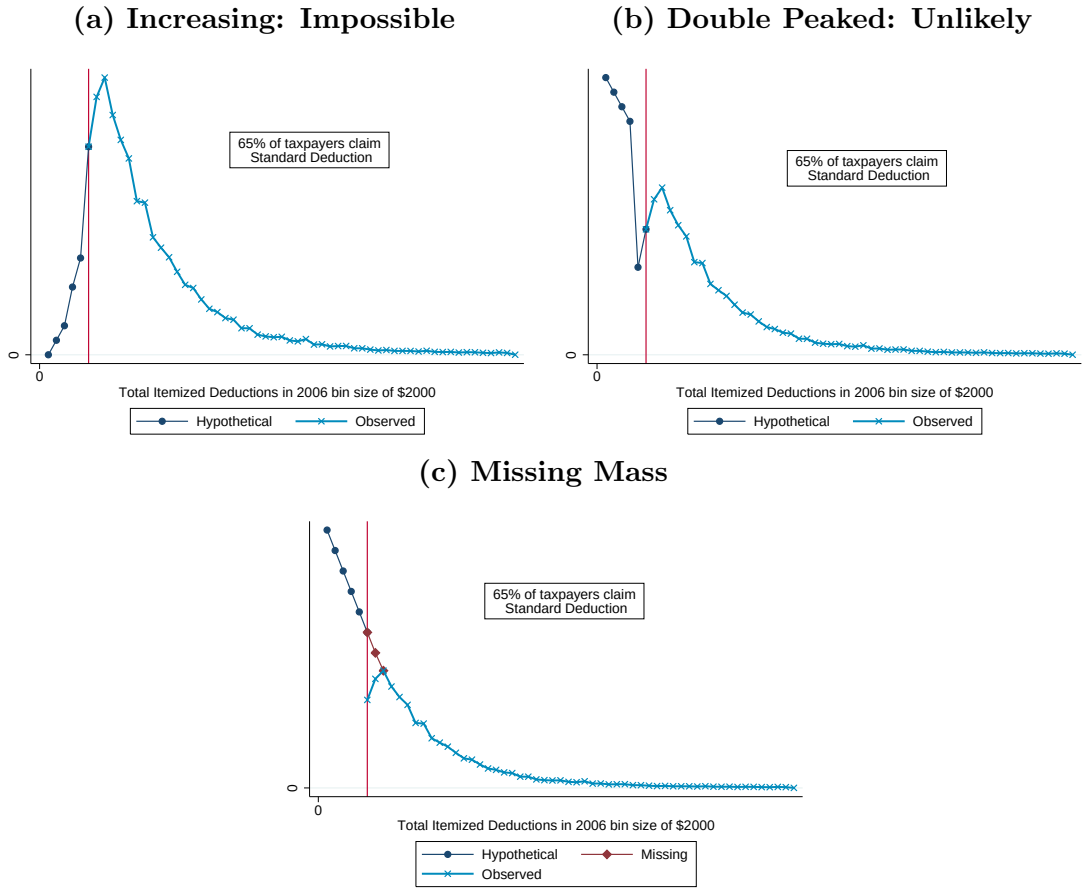
Notes: This graph pools all cross sections from 1998 to 2006 for joint filers and plots the distribution of itemizers using electronic filing and paper filing. It uses the area away from the standard deduction to extrapolate the shape of the counterfactual distribution of itemizers just above the standard deduction.

Figure H.21: Forgone Benefits Increase With Income



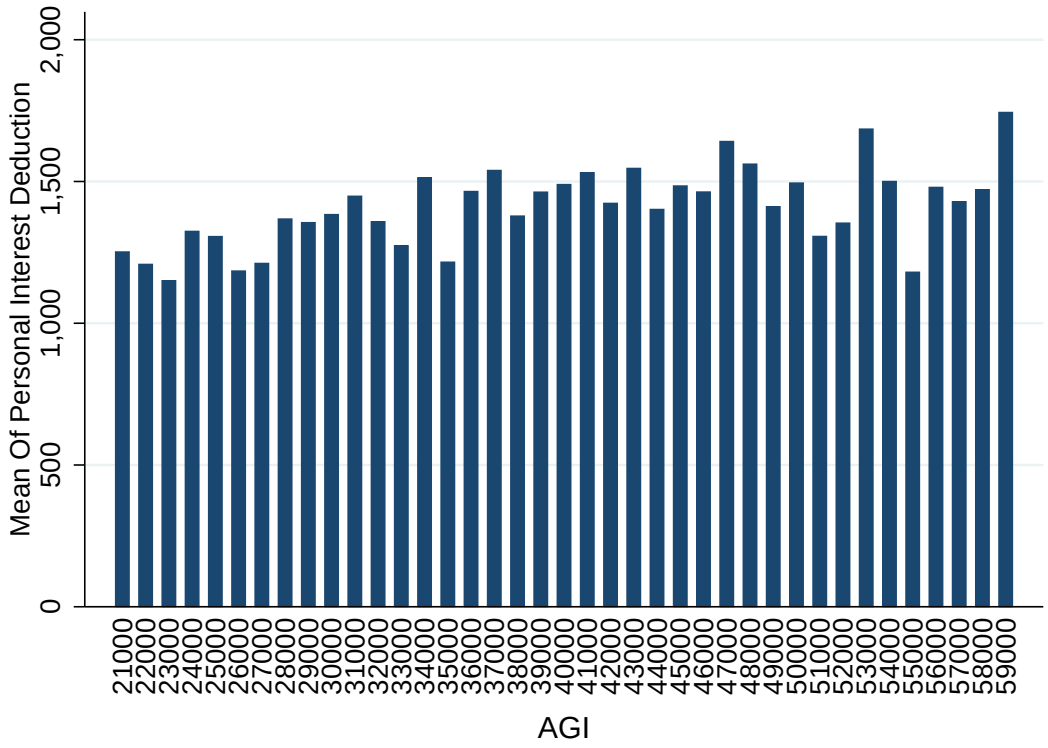
Notes: This graph plots the relationship between forgone benefits and income for all income deciles.

Figure H.22: Different Scenarios Below the Standard Deduction



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 H.23: No Behavioral Response For Personal Interest Deduction



Notes: This figure plots the average personal interest deduction claimed by income bins of $1000 in 1989. Below $30,950, the marginal tax rate is 15% for married filing jointly and above it is equal to 28%. If taxpayers were responding to tax incentives when claiming the personal interest deduction, one would observe a discontinuity at the MTR threshold. None is observed here.

I APPENDIX TABLES

Table I.5: Standard Deduction By Year For Joint Filers

Year	Standard deduction	S.D. in 2014 $	Growth Rate	Year	Standard deduction	S.D. in 2014 $	Growth Rate
1961	1000	7968	0.00%	1984	3400	7796	0.00%
1962	1000	7889	0.00%	1985	3540	7838	4.12%
1963	1000	7786	0.00%	1986	3670	7978	3.67%
1964	1000	7686	0.00%	1987	3760	7886	2.45%
1965	1000	7564	0.00%	**1988**	**5000**	**10070**	**32.98%**
1966	1000	7353	0.00%	1989	5200	9991	4.00%
1967	1000	7133	0.00%	1990	5450	9935	4.81%
1968	1000	6846	0.00%	1991	5700	9971	4.59%
1969	1000	6492	0.00%	1992	6000	10189	5.26%
1970	1000	6140	0.00%	1993	6200	10223	3.33%
1971	1500	8824	50.00%	1994	6350	10208	2.42%
1972	2000	11400	33.33%	1995	6550	10240	3.15%
1973	2000	10732	0.00%	1996	6700	10174	2.29%
1974	2000	9665	0.00%	1997	6900	10243	2.99%
1975	2600	11514	30.00%	1998	7100	10378	2.90%
1976	2800	11724	0.08%	1999	7200	10293	1.41%
1977	3200	12580	0.14%	2000	7350	10169	2.08%
1978	3200	11693	0.00%	2001	7600	10515	3.40%
1979	3400	11158	0.06%	2002	7850	10560	3.29%
1980	3400	9831	0.00%	2003	9500	12301	21.02%
1981	3400	8911	0.00%	2004	9700	12234	2.11%
1982	3400	8394	0.00%	2005	10000	12199	3.09%
1983	3400	8133	0.00%	2006	10300	12173	3.00%

Notes: The table shows the standard deduction amounts from 1961 to 2006 for joint filers and its growth rate. The years that I use to identify the cost of itemizing deductions are in bold.

Table I.6: Standard Errors of the Difference Between the 1970 and 1971 Densities (figure H.16a)

Bin	Deduction Range	Difference	Standard Errors	z-stat
1	[6140, 9140]	0.00373***	0.00102	3.64
2	(9140, 12140]	0.00288***	0.00090	3.20
3	(12140, 15140]	0.00307***	0.00074	4.11
4	(15140, 18140]	0.00083*	0.00046	1.81
5	(18140, 21140]	0.00019	0.00037	0.54
6	(21140, 24140]	0.00039	0.00027	1.45
7	(24140, 27140]	-0.00025	0.00018	-1.41
8	(27140, 30140]	-0.00001	0.00015	-0.09
9	(30140, 33140]	-0.00007	0.00011	-0.63
10	(33140, 36140]	-0.00010	0.00010	-0.94

Notes: This table shows the bootstrapped standard errors for the difference between bins in 1970 and 1971 for taxpayers with deductions below $30,000. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. I use 100 replications for the bootstrap estimation.

Table I.7: Standard Errors of the Difference Between the Density of Electronic Filers v.s. Paper Filers (Figure 5(b))

Bin	Deduction Range	Difference	Standard Errors	z-stat
1	[0, 2000)	7.08e-06***	1.44e-06	4.92
2	[2000, 4000)	3.02e-06*	1.55e-06	1.95
3	[4000, 6000)	5.91e-06***	1.39e-06	4.25
4	[6000, 8000)	3.44e-06**	1.54e-06	2.23
5	[8000, 10000)	5.10e-06***	1.49e-06	3.42
6	[10000, 12000)	1.47e-06	1.41e-06	1.04
7	[12000, 14000)	2.37e-07	1.42e-06	0.17
8	[14000, 16000)	-1.73e-06	1.18e-06	-1.47
9	[16000, 18000)	-1.93e-07	1.04e-06	-0.19
10	[20000, 22000)	-1.88e-06*	1.03e-06	-1.82

Table I.8: Standard Errors of the Difference Between the Density of March vs. April Filers (Figure 9c)

Bin	Deduction Range	Difference	Standard Errors	z-stat
1	[0, 2000)	-.000012***	2.87e-06	-4.17
2	[2000, 4000)	-.0000114***	2.83e-06	-4.02
3	[4000, 6000)	-5.08e-06	3.09e-06	-1.64
4	[6000, 8000)	-.0000141***	3.39e-06	-4.17
5	[8000, 10000)	-9.22e-06***	3.41e-06	-2.71
6	[10000, 12000)	-8.12e-06**	3.42e-06	-2.51
7	[12000, 14000)	-4.21e-06	1.42e-06	-1.23
8	[14000, 16000)	-5.94e-06	3.05e-06	-1.94
9	[16000, 18000)	-4.86e-07*	3.36e-06	-0.14
10	[20000, 22000)	-5.57e-06**	2.81e-06	-1.98

Notes: These two tables show the bootstrapped standard errors for the difference between the density of itemizers who use electronic filing versus paper filing in the first table and the difference between April and March Filers in the second table.
* denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.

Table I.9: Survey Based Estimates of the Compliance Costs of Taxation in the US

Article	Methodology	Cost of Itemizing Deductions	Aggregate Costs of Filing Taxes
Wicks (1965) and Wicks and Killworth (1967)	Survey of Montana residents	Not reported	32% of state and 11.5% of federal tax revenue
Slemrod and Sorum (1984)	Survey of 2000 Minnesota residents	Not reported	5% to 7% of total tax revenue
Little (1988), Commissionned by IRS	Two separate surveys of 750 and 6200 taxpayers	Not reported	1.59 billion hours
Slemrod (1989)	Estimate structural model based on survey of 2000 Minnesota residents	3.2 to 3.5 hours	Not reported
Blumenthal and Slemrod (1992)	Survey of 2000 Minnesota households in 1990	9 hours	85 billion dollars
Guyton et al. (2003)	Survey and ITBM* simulations	9.9 hours	18.7 billion hours dollars

Notes: This table reports the results of several research article documenting the cost of tax filing using survey evidence. *ITBM stands for the Individual Tax Burden Model.

Table I.10: Articles Documenting Low Take-Up Rates/Large Forgone Benefits

Article	Setting	Forgone Benefits
Steuerle et al. (1978)	Tax Benefits/Income Averaging	$666
Blank and Card (1991)	Unemployment Insurance Benefits	Take-up rate of less than 30% of eligible unemployed individuals
Madrian and Shea (2001)	Retirement Savings	50% match of retirement savings up to 6% of contributions
Sydnor (2010)	Home Insurance	Five times the insurance premium
Bhargava and Manoli (2015)	Taxes	Earned Income Tax Credit Benefits
Handel (2013)	Health Insurance	$2,032 per year
Keys et al. (2014)	Mortgage Refinancing	Present discounted cost of $11,500

Table I.11: IRS Hourly Cost Estimates

Form	Recordkeeping	Learning about the law or the form	Preparing the form	Copying, assembling and sending the form to the IRS	Total
1040	3 hrs., 7 min.	2 hrs., 32 min.	3 hrs., 10 min.	35 min.	9hrs., 24 min.
Sch. A	2 hrs., 47 min.	26 min.	1 hr., 1 min.	20 min.	4 hrs., 34 min.
Sch. B	33 min.	8 min.	16 min.	20 min.	1 hr., 17 min.
Sch. C	6 hrs., 13 min.	1 hr., 4 min.	1 hr., 56 min.	25 min.	9 hrs., 38 min.
Sch. D	1 hr., 2 min.	1 hr.	1 hr., 8 min.	35 min.	3 hrs., 45 min.
Sch. D-1	13 min.	1 min.	13 min.	35 min.	1 hr., 2 min.
Sch. E	2 hr., 52 min.	1 hr., 7 min.	1 hr., 16 min.	35 min.	5 hrs., 50 min.
Sch. F	9 hr., 41 min.	1 hr., 59 min.	3 hr., 52 min.	35 min.	16 hrs., 7 min.
Sch. R	20 min.	15 min.	22 min.	35 min.	1 hr., 32 min.
Sch. SE short	20 min.	11 min.	13 min.	14 min.	58 min.
Sch. SE long	26 min.	22 min.	37 min.	20 min.	1 hr., 45 min.

Notes: Each cell of this table is an estimate of the time it takes to perform each task associated with each tax schedule. They are based on IRS surveys of taxpayers at the time of filing and are reported in the 1040 instructions (on page 3 in 1989). There is no information on Sch. R in the SOI public use files so its cost is not estimated in this paper.