American Economic Journal: Economic Policy 2020, 12(4): 38–57
https://doi.org/10.1257/pol.20180664

How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs[†]

By Youssef Benzarti*

This paper uses a quasi-experimental design to estimate the cost of filing taxes. Using US tax returns, I observe how taxpayers choose between itemizing deductions and claiming the standard deduction. Taxpayers forgo large 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. These estimates suggest substantial costs of filing federal income taxes, significantly larger than previously estimated using surveys. (*JEL* H24, H26, H31)

Income taxes represent the largest source of tax revenue in the United States. Today, 8.8 percent of GDP is transferred from individuals to the federal government through income taxes. While an extensive literature documents the efficiency costs of taxation, we know less about the costs 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.

* Economics Department, University of California, Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106 and NBER (email: benzarti@ucsb.edu). Dan Silverman was coeditor for this article. 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, Takeshi Murooka, Michaela Pagel, Colin Raymond, Alex Rees-Jones, Antonio Rosato, Jesse Rothstein, Josh Schwartzstein, Joel Slemrod, Charles Sprenger, Justin Sydnor, Alisa Tazhitdinova, Danny Yagan, and seminar participants at the Tax Systems Conference (Oxford University and University of Michigan), Paris II, CREST, Aarhus University, SITE Psychology and Economics, the US Department of Treasury, Princeton University, Harvard Kennedy School, UCLA, McMaster University, University of Massachusetts Amherst, and University of Toronto. Funding from the Robert D. Burch Center for Tax Policy and Public Finance is gratefully acknowledged.

[†] Go to https://doi.org/10.1257/pol.20180664 to visit the article page for additional materials and author disclosure statement or to comment in the online discussion forum.

[1] See, for example, Currie (2006); Bertrand, Mullainathan, and Shafir (2006); and, more recently, Bhargava and Manoli (2015).

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 nonzero, 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 tax data. To estimate the cost of itemizing, I proceed in the following way. If individuals forgo tax benefits because of compliance costs, then 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 density of itemizers exhibits a missing mass just above the standard deduction.

To confirm that this shape is due to taxpayers responding to the standard 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 postreform standard deduction threshold. The postreform density is systematically lower than the prereform density just above the postreform 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 approach compares two cross sections before and after a reform.[3]

I find that the cost of itemizing ranges from 0.6 percent to 0.8 percent of adjusted gross income (AGI); i.e., the disutility derived from itemizing is equivalent to working 10 to 15 hours, which is substantially larger than previous estimates. If individuals switch to the standard deduction because they value their time more than the benefits they can derive from itemizing, then richer households should forgo more tax benefits than poorer ones should. To test this hypothesis, I break down individuals by income deciles and use the procedure outlined above on each subgroup. The results show an increasing relationship between forgone benefits and income—while controlling for marginal tax rates—consistent with the hypothesis that tax filing imposes a higher cost on richer individuals because they have a higher marginal value of time.

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.

[2] My estimates are not affected by the alternative minimum tax, variation in marginal tax rates, or the phaseout of the personal interest deduction in 1987.

[3] My approach is related in spirit to a difference-in-difference approach where the treatment group includes taxpayers just above the standard deduction threshold and the control group is individuals far above the standard deduction threshold.

Behavioral responses to concave kink points predict that taxpayers will respond to variations in marginal tax rates but should not respond to variations in income while holding marginal tax rates fixed. The fact that forgone benefits increase with income—while controlling for marginal tax rates—supports the compliance costs explanation. A second alternative explanation 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. 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. And deductions tend to nominally increase over time, which implies that taxpayers should have a precise belief over the lower bound of the tax savings they can derive from itemizing, therefore contradicting the idea that uncertainty prevents them from itemizing.

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. Procrastination 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 lines of 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, it finds a smaller cost of itemizing of $107, which is equivalent to 0.12 percent of AGI and is 5 to 7 times smaller than my cost estimates. It uses 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. I discuss their approach and some of their assumptions in more detail in online Appendix Section A.[4] 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 disutility of filing taxes.[5]

Finally, this paper adds to a long tradition in public economics emphasizing the need to screen out applicants using ordeal mechanisms (Nichols and Zeckhauser 1982). If poorer individuals value their time less, then such policies can

[4] This paper is also related to a literature that estimates the effect of tax simplicity on individual and firm behavior. See, for example, Abeler and Jäger (2015) and more recently Tazhitdinova (2020); Harju, Matikka, and Rauhanen (2019); Aghion et al. (2017); and Tazhitdinova (2018).

[5] See online Appendix Table 1.7 for a list of papers using survey-based estimates.

successfully 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 costs are, such policies could be screening out too many individuals.

I. Data and Institutional Background

A. *Institutional Background*

Taxpayers can reduce their taxable income by choosing to itemize their deductions or claiming the standard deduction. The decision to itemize deductions requires comparing two numbers: the sum of itemized deductions and the standard deduction amount. Itemizing, however, is more costly: taxpayers need to keep a record of all the expenses they want to deduct and file Schedule A. Approximately two-thirds of the population claim the standard deduction. The standard deduction varies by filing status and by whether the person is blind or older than 65.

Taxpayers claim four main deductions: (i) state and local income taxes, (ii) mortgage interest, (iii) real estate taxes, and (iv) charitable donations. They represent 17 percent, 40 percent, 14 percent, and 12 percent of total deductions, respectively.

Schedule A is a one-page schedule and relatively easy to fill out, as it only requires copying numbers from receipts and statements and then summing them up. Record keeping is more time consuming since it requires archiving various records of expenses.

B. *Data*

I use the Statistics of Income (SOI) dataset. It consists of repeated annual cross sections of individual tax returns. 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 in this paper uses these weights to reflect population averages. No particular sample restrictions are made except for properly assigning individuals to filing types (single, joint, etc.) and marginal tax brackets (details in online Appendix Section C). In both cases, I rely on the information provided in tax returns. In addition, I use a panel of tax returns known as the University of Michigan tax return panel. The panel covers 1979 to 1990 and contains the same variables as the SOI files but has a smaller sample size.

II. Missing Mass

If some taxpayers are claiming the standard deduction when the sum of their itemized deductions is greater than the standard deduction, then there should be a missing mass just above the standard deduction threshold. I graph the density of

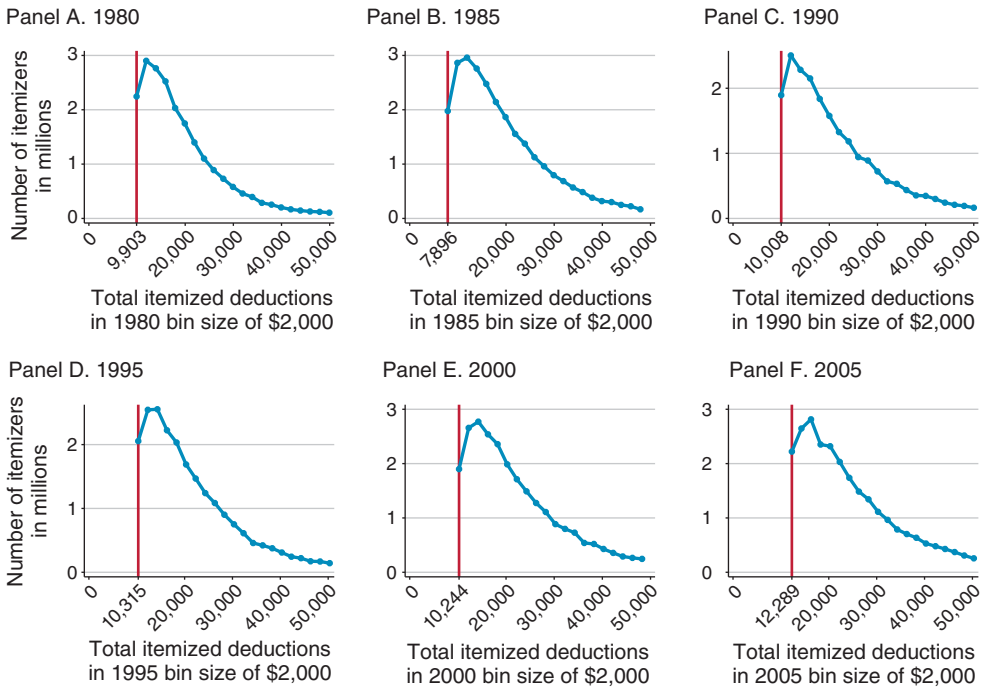

FIGURE 1. MISSING MASS JUST ABOVE THE STANDARD DEDUCTION

Notes: These figures 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 online Appendix Figures H.9, H.10, H.11, and H.12 and online Appendix Figure H.13 for single filers.

deductions for several years in Figure 1 (the remaining 27 years are shown in online Appendix Figures H.8–H.12).[6] The bin closest to the standard deduction includes only those itemizers whose deductions are strictly larger than the standard deduction. The density is systematically low just above the standard deduction and then increases and peaks two to three bins away. This is 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 is only suggestive.[7]

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. 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.[8] Although the magnitudes of the estimated costs for the 1971, 1975, and 2003

[6] All dollar amounts are in 2016 dollars throughout the rest of the paper.
[7] Online Appendix Figure H.22 shows different alternative scenarios that could create a missing mass.
[8] More details about the parallel system and other changes are provided in online Appendix Section E.

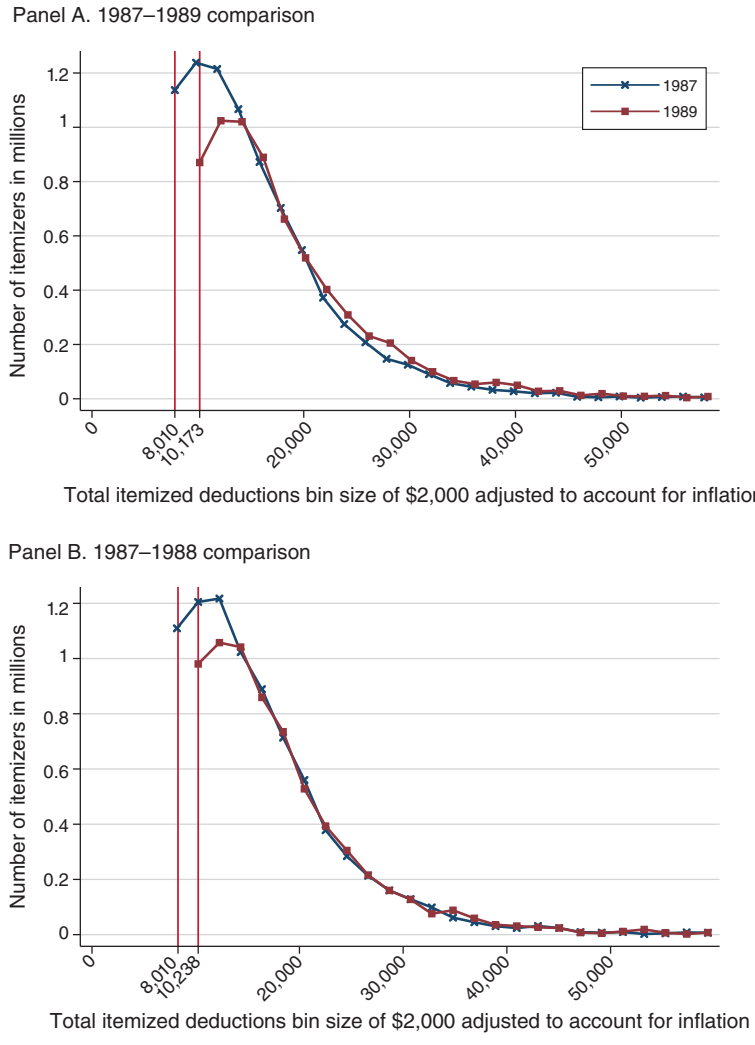

FIGURE 2. DENSITY OF DEDUCTIONS FOR ITEMIZERS FILING JOINTLY PRE- AND POSTREFORM

Note: 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.

reforms are inaccurate, they still provide reduced-form evidence of the *existence* of compliance costs.

I compare the pre-1988 reform year to the postreform year to account for lagged behavioral responses. Figure 2, panel A and panel B graph the density of deductions in pre- and postreform years for the 1988 reform.[9] The shape of the distribution in year $t + 1$ mirrors that of years t and $t - 1$, and the missing mass precisely follows

[9] Online Appendix Figure H.14 reports these densities for the 1971, 1975, and 2003 reforms and shows that the changes are qualitatively consistent with the 1988 reform.

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 were a feature of the distribution and not due to the standard deduction, then it would not track the standard deduction once it is increased.

III. Cost Estimation

A. *Cost Estimation Methodology*

To calculate the distribution of forgone benefits in the population, I need to reconstruct the counterfactual distribution of itemizers. Using the prereform year as the counterfactual distribution would lead to an underestimate of the cost because the prereform distribution is distorted by its proximity to the standard deduction, as shown in Figure 2, panel A. 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 except for assumptions A1 and A2 below. No assumptions about the drivers of the cost are needed in this section: the distribution of costs can be due to the costs of record keeping and filing or due to fear of audits and uncertainty. While alternative explanations—discussed in Section V—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 probability density function (pdf) of itemizers, assuming that there is no standard deduction and no cost of itemizing, as illustrated in Figure 3.[10] Denote by $g_S(\cdot)$ the observed probability density function of itemizers when the standard deduction is equal to S. Then, $g_0(\cdot)$ and $g_\delta(\cdot)$ correspond, respectively, to the pre- and postreform pdf of itemizers when the standard deduction increases from 0 to δ. The cumulative distribution function (cdf) of the cost of itemizing is denoted by $C_S(\cdot)$ and 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,

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

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

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

[10] An illustrative simulated example is provided in online Appendix Section B.B.

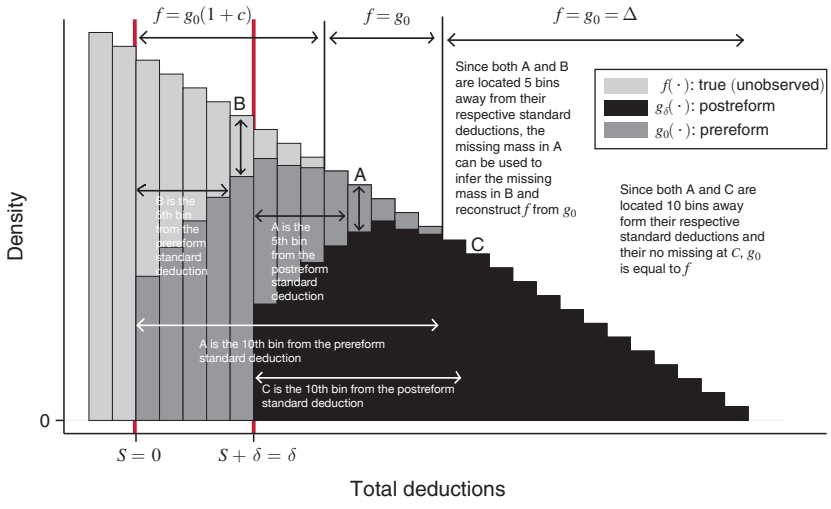

FIGURE 3. RECOVERING THE COUNTERFACTUAL DISTRIBUTION OF DEDUCTIONS

Note: This graph illustrates the method used in Section IIIA to reconstruct the counterfactual density of itemizers $f(\cdot)$ using the pre- and postreform densities $g_0(\cdot)$ and $g_\delta(\cdot)$.

In other words, the cost of itemizing is related to the missing mass $g_S(d)/f(d)$. 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: The cost is similar pre- and postreform.
- A2: The 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

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

which implies that the same proportion of individuals is missing d deductions above the prereform standard deduction and $d + \delta$ deductions above the postreform 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.[11] A failure of A2 introduces a bias: in online Appendix Section B.A, I provide an upper bound on the size of this bias and show that it is small. Intuitively, if A2 fails, then $g_\delta(\cdot)$ can be used instead of $f(\cdot)$. This will necessarily yield lower bounds, since $g_\delta(\cdot) < f(\cdot)$. For joint filers in the 28 percent bracket, for example, the estimated cost would lie between \$519 and \$591 if A2 fails, instead of

[11] See online Appendix Figure H.15.

\$591. In addition, if these two assumptions hold, then the missing masses in the pre- and postreform distributions should be proportional. Online Appendix Figure H.18 shows that this proportionality assumption indeed holds.

To estimate $C_0(\cdot)$ 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 postreform, 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 $C_0(d) = 1$ and $g_\delta(d) = g_0(d) = f(d)$; i.e., the pre- and postreform distributions of itemizers overlap for ranges of deductions exceeding the postreform 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 prereform 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 prereform taxpayers do not forgo any deductions, but the postreform ones do. This corresponds to the middle area in Figure 3. As a consequence, the prereform 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 postreform distribution is distorted, i.e., $g_\delta(d) < f(d)$. From equation (3), it follows that $\forall d \in [c_{\max}; \delta + c_{\max}]$:

$$C_0(d - \delta) = \frac{g_\delta(d)}{f(d)} = \frac{g_\delta(d)}{g_0(d)}, \tag{4}$$

which allows me to estimate $C_0(\cdot)$ over $[c_{\max} - \delta; c_{\max}]$.

Third, if $d \in [c_{\max} - \delta, c_{\max}]$, then both the prereform and postreform itemizers are forgoing deductions. This corresponds to the leftmost area in Figure 3. In this case, both the pre- and postreform 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 $C_0(\cdot)$ over $[c_{\max} - \delta; c_{\max}]$ from equation (4) to correct the prereform distribution by using the definition of $g_0(\cdot)$ from equation (2): $f(d) = g_0(d)/C_0(d)$. From equation (3), it follows that $\forall d \in [c_{\max} - \delta; c_{\max}]$,

$$C_0(d - \delta) = \frac{g_\delta(d)}{f(d)} = \frac{g_\delta(d)}{g_0(d)} C_0(d), \tag{5}$$

which allows me to estimate $C_0(\cdot)$ 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 $C_0(\cdot)$ and $f(\cdot)$ over $[0, c_{\max}]$.

B. *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 phaseout of the personal interest deduction, which I

control for (details in Section VD).[12] Each cost estimate is performed on individuals who have the same marginal tax rate and were not subject to the 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 postreform year because the reform occurred in 1988. If taxpayers learned about the increase in the standard deduction when filing their taxes, then we should observe the full response in 1989. Figure 2, panel B confirms that the effect was smaller during the reform year.

Table 1 shows the estimated costs for single, joint, and head-of-household filers in the 15 percent and 28 percent marginal tax brackets. Costs range from 0.57 percent to 0.85 percent of AGI. In dollar amounts, they vary from $175 for single filers in the 15 percent bracket to $591 for joint filers in the 28 percent 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 shown in Section IVA.

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 online Appendix Table I.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, with the exception of bins 10, 11, and 13, which are statistically significantly different at the 5 percent and 10 percent level, with differences of a very small magnitude (less than ten times that of the first or second bins).

IV. Anatomy of the Missing Mass

A. *Costs Increase with Income*

If rich taxpayers value their time more than poor ones do because their hourly wage is higher, then we should expect them to forgo more deductions. I test this using the income reported on tax returns. I break down the sample used above by deciles of income. This reduces the power, which I deal with in two ways. First, I use a moving average of income deciles. Second, I focus on joint filers in the 28 percent marginal tax bracket, as they represent by far the largest group of taxpayers.

I then fit a flexible polynomial through each deduction bin and 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 Section III.[13] The results are plotted in Figure 4, panel A: as income increases, taxpayers forgo more

[12] See online Appendix Section D for the TRA 86 reforms.

[13] I only report results for the first seven groups because deductions and income are positively correlated and very few high-income individuals are close to the standard deduction threshold. However, the increasing relationship is robust to including those bins; see online Appendix Figure H.19.

TABLE 1—COST ESTIMATES

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

Note: This table shows the cost of itemizing as estimated in Section III for different brackets and filing status.

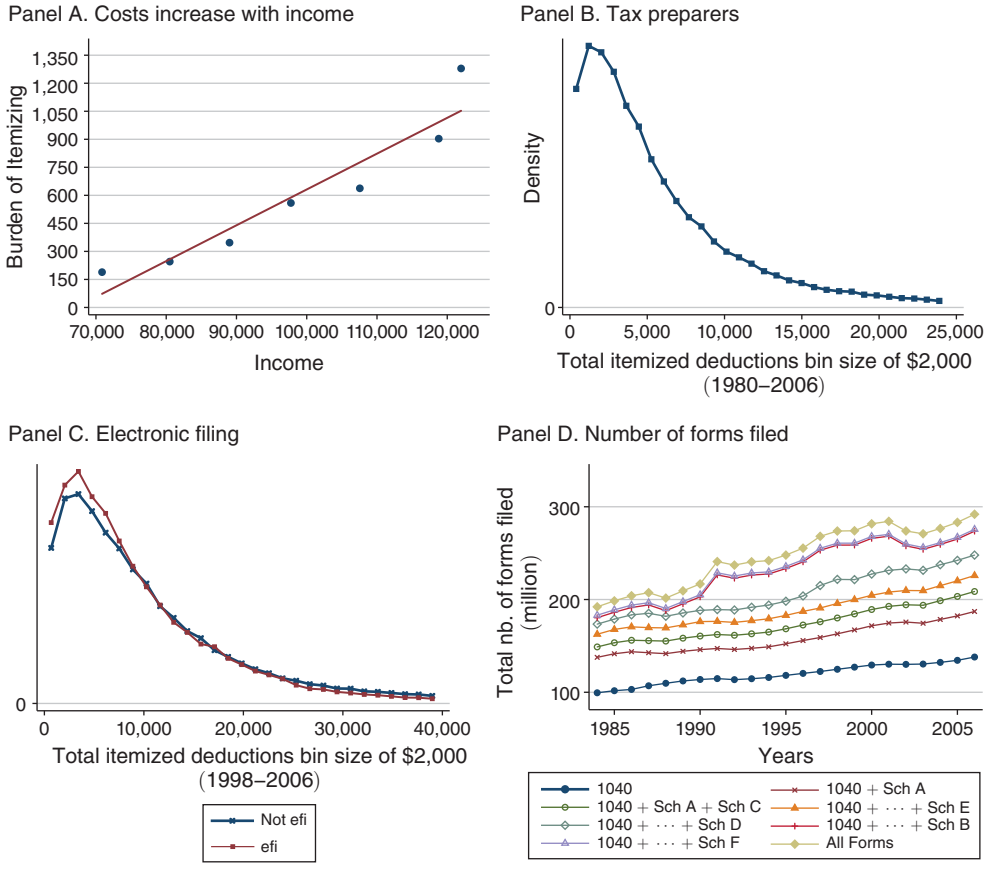

FIGURE 4. ANATOMY OF THE MISSING MASS

Notes: Panel A plots the relationship between income and the cost of itemizing. This relationship controls for the variation in MTR across the different income groups. Panels B and C show the distribution of itemized deductions for taxpayers who use tax preparers and e-filing. The *x*-axis is normalized such that 0 corresponds to the standard deduction threshold. Panel B pools years 1980 to 2006 and panel C years 1998 (start of e-filing) to 2006. Panel D shows the evolution of the number of forms filed over time.

deductions, consistent with the idea that they value their time more. Notice, first, that all taxpayers in Figure 4, panel A fall in the 28 percent 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. Second, even though itemized deductions increase with income, this is not what drives the increasing relationship between income and forgone benefits. Because I compare the same income groups before and after the reform, I am implicitly controlling for the relationship between income and deductions.

B. *Record Keeping*

Electronic filing (e-filing) and tax preparers may reduce the cost of filling out forms. However, they do not affect the cost of record keeping. Therefore, e-filing and tax preparers are unlikely to eliminate the observed missing mass if it is due to the hassle of filling out forms. Survey estimates of the cost of filing taxes have consistently documented that record keeping is the main driver of the cost of itemizing.[14]

To test for the effect of e-filing and tax preparers, I plot the density of itemizers who use a tax preparer or e-file in Figure 4, panel C and panel B. In both cases, the distribution of itemizers exhibits a clear missing mass, implying that e-filing or tax preparers do not eliminate the cost of itemizing.

Figure 4, panel C compares the density of taxpayers who use e-filing to those who do not. It shows a slightly smaller missing mass for taxpayers who e-file.[15] However, e-filing only 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, making a direct comparison of the missing mass impossible without additional parametric assumptions.[16]

Given that costs appear to be driven by record keeping rather than filling out Schedule A, it is likely that costs increase with the number of receipts one has to claim but not necessarily with the dollar amount claimed.[17] Consider the mortgage interest deduction or the state income tax deduction: taxpayers receive a statement, the 1098 for mortgage interest and W2 for state income taxes, close to the filing deadline, summing up all mortgage interest or state income taxes paid. For this reason, the record-keeping costs of claiming $1,000 or $10,000 worth of mortgage interest are likely to be similar. However, this is not true anymore for non-third-party reported items, in particular for charitable donations. In this case, since no statement

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

[15] The difference is statistically significant: bootstrapped standard errors are reported in online Appendix Table I.6

[16] One could also compare states with and without state income taxes. Since state income taxes are reported on W2s, they are unlikely to impose a large cost of record keeping. Online Appendix Figure H.20 shows a missing mass for states both with and without income taxes, suggesting that the state income tax deduction is not a large driver of costs.

[17] A direct test of this assumption is not possible with the data I use, because the number of receipts is not reported in the data.

is provided, each additional donation is likely to be costly in terms of record keeping. This suggests that the main driver of record-keeping costs is not necessarily the level of donations but instead the number of separate donations: a $1,000 donation imposes less record-keeping costs than 20 different $50 donations even though they add up to the same total amount. The fact that costs are likely to be variable for certain deductions is consistent with the findings from Rees-Jones (2018), which shows that taxpayers can manipulate the balance due on their taxes—presumably by adjusting their deductions—to precisely target a balance of $0. Overall, if record-keeping costs are partly variable, partly fixed, it is likely that the variable portion of these costs is pivotal to the decision to itemize. This would imply that the deductions that impose a variable cost on taxpayers, such as charitable donations, are likely to be the main driver of the missing mass.

V. Alternative Explanations to Compliance Costs

In this section, I consider alternative explanations for the missing mass. Note that information is unlikely to explain it; since I focus on taxpayers who switch from itemizing to claiming the standard deduction, they should be aware of the decision to itemize. In addition, taxpayers are reminded on the 1040 Form that they can make the choice between itemizing and claiming the standard deduction.

A. *Concave Kink Points*

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, possibly inducing some taxpayers to be indifferent between these two points. Depending on the curvature of the indifference curve, this could create a bimodal distribution with a missing mass both to the right and to the left of the standard deduction. This argument is illustrated in Figure 5.

However, in this case, the size of the missing mass should not respond to variations in income when controlling for marginal tax rates. The only reason that taxpayers would adjust their deductions in response to a concave kink point is variation in marginal tax rates, while changes in income should not matter. 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 marginal tax rates. Consistent with the compliance cost interpretation, Figure 4, panel A graphs the relationship between forgone benefits and income—*controlling for marginal tax rates*—and finds an increasing relationship.

In addition, behavioral responses to concave kink points lead individuals to locate away from the concave kink point. If behavioral responses to concave kink points led to the observed missing mass, then as the standard deduction increases, the bimodal distribution should track the new standard deduction threshold, and the pre- and postdistribution peaks should not overlap. The observed pre- and postdistributions in Figure 2, panel A and panel B contradict this prediction: the pre- and

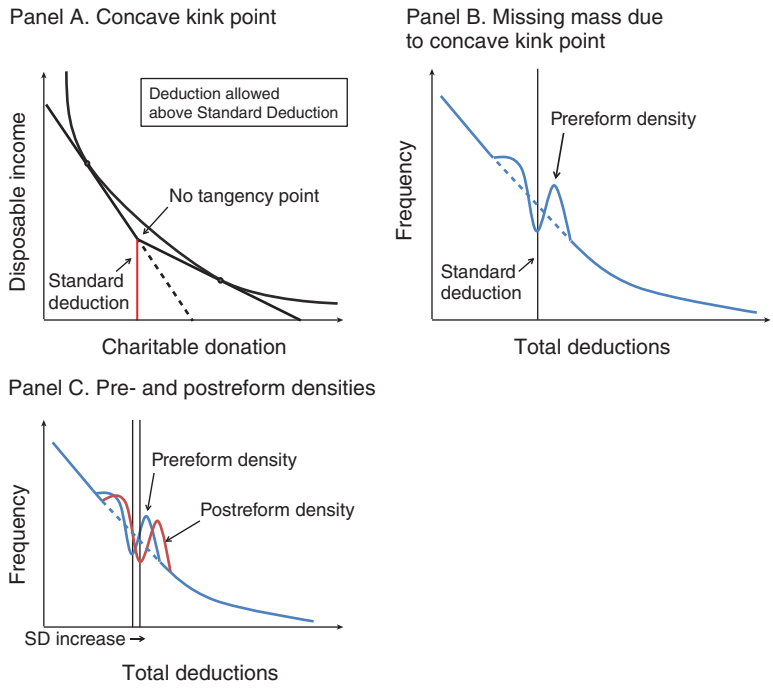

FIGURE 5. CONCAVE KINK POINT: POSTREFORM DENSITIES SHOULD NOT OVERLAP

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, then we should observe that the pre- and postreform densities are not overlapping just above the standard deduction. This is contradicted by Figure 2, panel A.

postdistribution peaks are overlapping, once again rejecting the hypothesis that the missing mass is caused by behavioral responses to concave kink points. A graphical illustration of this argument is provided in Figure 5.

Overall, both of these empirical tests rule out responses to concave kink points. This is consistent with the previous literature. Saez (2010), Kleven and Waseem (2013), and Tazhitdinova (2015) directly test the predictions of a behavioral response to both concave and convex kink points and find responses to convex kink points but not to concave kink points. Kleven (2016), in a survey of the bunching literature, confirms that there is no evidence of bunching at concave kink points.

B. *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 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.[18] The survey allows me to elicit their beliefs about both the audit probabilities and the perceived costs of audits.

Surveyed individuals have levels of income similar to those of joint filers in the 1988 28 percent marginal tax bracket. On average, they believe that audits occur with a probability of 8.72 percent, which is 7.9 times the true audit probability.[19] This accounts for, at most, 25 percent of the $591 estimated forgone benefits for joint filers in the 28 percent marginal tax bracket.[20]

C. *Rational Inattention*

Can uncertainty about the level of deductions lead taxpayers to switch to the standard deduction and explain the observed missing mass? Online Appendix Table I.4 shows the results of the calibration of a model illustrative of this type of behavior with varying levels of risk aversion.[21] Taxpayers would need an uncertainty range of at least $\pm$\$14,000 in order to forgo amounts of money as 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 percent of total deductions are mortgage interest, state taxes, and property taxes, which are relatively stable for a given person year after year. Indeed, online Appendix Figure H.21 plots the distribution of year-on-year percent change in the level of deductions by taxpayer, using the 1979 to 1990 panel. It shows that, on average, deductions tend to increase nominally. Therefore, it is likely that taxpayers who were previously itemizing will believe that their deductions will be nominally higher this year and therefore will hold relatively precise beliefs over the lower bound on the tax savings they can derive from itemizing.[22]

D. *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 III have already accounted for these adjustments. The fact that the pre- and postreform densities overlap away from the standard deduction threshold shows that the prereform density is a relevant counterfactual for the postreform

[18] Online Appendix Section F details the survey instrument.

[19] This is consistent with Bhargava and Manoli (2015), which finds that EITC filers believe that audit probabilities are eight times greater than the true ones.

[20] On average, expected audit costs are $147, with a 95 percent confidence interval of $[126, 169]$.

[21] The model is outlined in online Appendix Section G.

[22] Taxpayers may also hold biased beliefs over tax rates, leading them to misperceive the savings they can derive from itemizing. Kuziemko et al. (2015), for example, shows that taxpayers are inattentive to certain exemption thresholds, which could lead them to overestimate average tax rates. On the other hand, de Bartolome (1995) and Rees-Jones and Taubinsky (2019) show that taxpayers tend to "iron" tax schedules, leading to an underestimate of marginal tax rates. Abeler and Jäger (2015) shows that complexity plays a large role in these misperceptions. Since the 1987, 1988, and 1989 tax schedules were only comprised of two tax brackets, these biases, while possible, are unlikely to be meaningfully large.

density in Figure 2, panel A 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 percent of their personal interest in 1987, 40 percent in 1988, and 20 percent 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 percent 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 percent marginal tax rate bracket. If there had been a behavioral effect, then we should observe more deductions for individuals above the 28 percent 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 postreform densities. Away from the standard deduction, the two graphs overlap, implying that the postreform density is an appropriate counterfactual for the 1989 density.

E. *Contemporary Relevance*

There are three main reasons why the tax filing costs estimated in this paper are likely to be relevant for recent years. First, while electronic filing has been rising rapidly over the past two decades, it does not appear to have drastically reduced filing costs. Figure 6, panel A shows that the distribution of itemized deductions still exhibits a missing mass just above the standard deduction, even in the most recently available year (2009). Figure 6, panel B similarly shows the missing mass for electronic filers in 2009.[24] These figures imply that taxpayers are still failing to itemize in spite of the advent of electronic filing. This result is consistent with filing costs being mostly driven by record keeping.

In addition, it is likely that overall filing costs have increased over time, as we observe a steady increase in the *number* of schedules filed over time. Figure 4, panel D plots the number of forms filed by schedule over time. While one additional schedule was filed for each 1040 Form in the mid-1980s, this proportion is almost two to one in recent years, suggesting that costs may have been increasing over time, as taxpayers have to file more and more schedules.

[23] See online Appendix Figure H.23.

[24] Similar patterns are observed in 2009 for taxpayers who use a tax preparer (Figure 6, panel C) and self-file their returns (Figure 6, panel D).

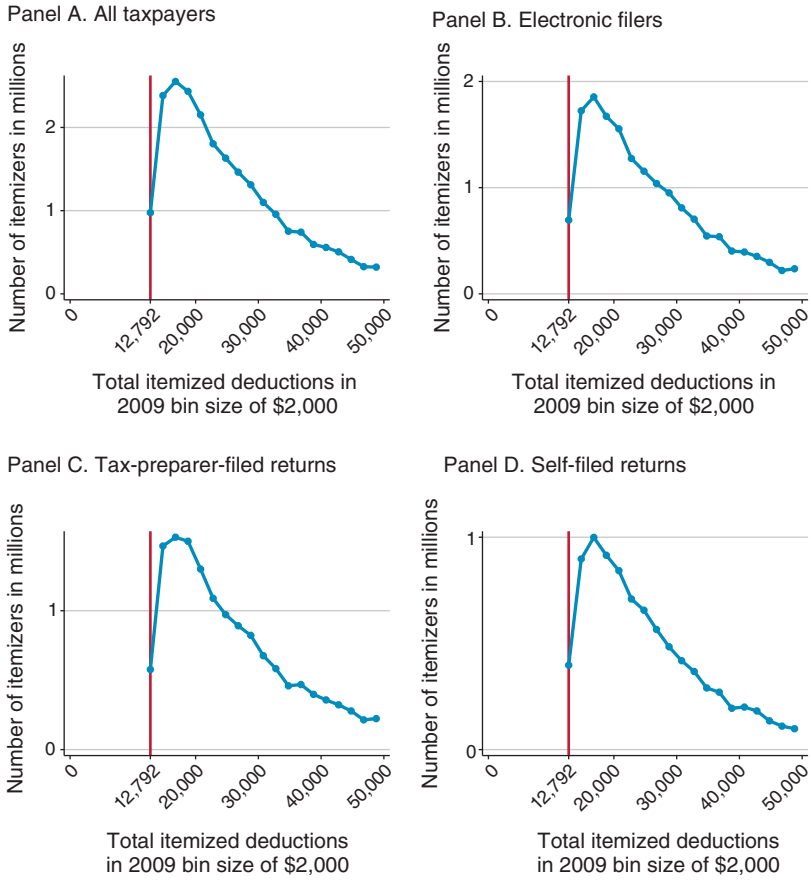

FIGURE 6. DISTRIBUTIONS IN THE MOST RECENTLY AVAILABLE YEAR (2009)

Notes: These figures plot the distribution of itemizers in 2009, which is the most recently available year in the SOI dataset. Panel A shows the distribution for all taxpayers, panel B for taxpayers who file electronically, panel C for taxpayers who use a tax preparer, and panel D for taxpayers who self-file.

VI. Compliance Costs or Behavioral Costs?

There is extensive evidence that individuals are time inconsistent and tend to procrastinate.[25] If taxpayers procrastinate on filing their taxes, then one should observe a large proportion of taxpayers filing on April 15th and procrastinators forgoing more deductions.[26]

First, consistent with individuals procrastinating on filing their taxes, I find that taxpayers bunch at the April 15th deadline. Figure 7, panel A graphs the volume of Google searches for the term "1040" by week, and Figure 7, panel B uses data from

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

[26] This argument is formalized in online Appendix Section H, which shows that procrastination can lead to high record-keeping costs, resulting in individuals failing to itemize.

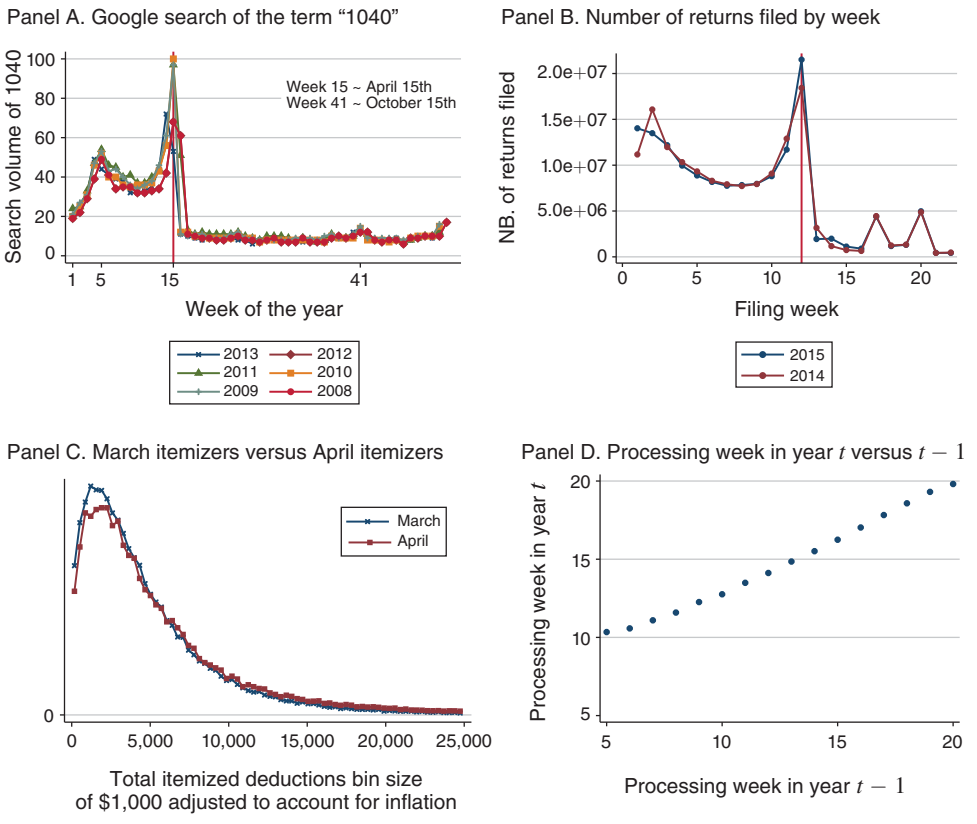

FIGURE 7. DEADLINE EFFECTS

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.

the IRS[27] 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, Reck, and Slemrod (2015), which shows that taxpayers search more actively on Google and Wikipedia for capital-gains-tax-related information close to April 15th.

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 7, panel C[28] shows that the missing mass for close-to-the-deadline filers (the 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: (i) by delaying filing, they forgo interest on their refunds, and (ii) they expose

[27] See https://www.irs.gov/uac/2016-and-prior-year-filing-season-statistics.
[28] Online Appendix Section C.E explains how the graph is constructed.

themselves to higher filing costs. Indeed, the sample I use to generate Figure 7, panel C 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.[29] 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.[30]

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 taxpayers who file late in year t are likely to file earlier in year $t+1$. To test for this, in Figure 7, panel D, 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, then the relationship should be constant, as we should observe mean reversion. If they are doing so because of a systematic bias, then the relationship should be increasing as the year t week of processing should predict the year $t+1$ week of processing. Figure 7, panel D shows a clear increasing relationship between the processing weeks in year t and year $t+1$, consistent with the explanation that late filing is due to a systematic bias.

VII. 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 tax benefits, suggesting large tax filing costs. 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. It can also be used when identifying responses from a censored distribution above or below a certain threshold.

REFERENCES

Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy* 7 (3): 1–28.

Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva. 2017. "Tax Simplicity and Heterogeneous Learning." NBER Working Paper 24049.

Benzarti, Youssef. 2020. "Replication Data for: How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E114723V1.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir. 2006. "Behavioral Economics and Marketing in Aid of Decision Making among the Poor." *Journal of Public Policy and Marketing* 25 (1): 8–23.

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

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

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." *American Economic Review* 105 (11): 3489–3529.

Blumenthal, Marsha, and Joel Slemrod. 1992. "The Compliance Cost of the U.S. Individual Income Tax System: A Second Look after Tax Reform." *National Tax Journal* 45 (2): 185–202.

Currie, Janet. 2006. "The Take-Up of Social Benefits." In *Public Policy and the Income Distribution*, edited by Alan J. Auerach, David Car and John M. Quigley, 80–148. New York: Russel Sage Foundation.

de Bartolome, Charles A.M. 1995. "Which Tax Rate Do People Use: Average or Marginal?" *Journal of Public Economics* 56 (1): 79–96.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.

Gronau, Reuben. 1973. "The Effect of Children on the Housewife's Value of Time." *Journal of Political Economy* 81 (2, Pt. 2): S168–99.

Guyton, John L., John F. O'Hare, Michael P. Stavrianos, and Eric J. Toder. 2003. "Estimating the Compliance Cost of the U.S. Individual Income Tax." *National Tax Journal* 673–88.

Harju, Jarkko, Tuomas Matikka, and Timo Rauhanen. 2019. "Compliance Costs vs. Tax Incentives: Why Do Entrepreneurs Respond to Size-Based Regulations?" *Journal of Public Economics* 173: 139–64.

Hoopes, Jeffrey L., Daniel H. Reck, and Joel Slemrod. 2015. "Taxpayer Search for Information: Implications for Rational Attention." *American Economic Journal: Economic Policy* 7 (3): 177–208.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics* 8: 435–64.

Kleven, Henrik J., and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics* 128 (2): 669–723.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *American Economic Review* 105 (4): 1478–1508.

Nelson, Forrest D. 1977. "Censored Regression Models with Unobserved, Stochastic Censoring Thresholds." *Journal of Econometrics* 6 (3): 309–27.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *American Economic Review* 72 (2): 372–77.

Pitt, Mark M., and Joel Slemrod. 1989. "The Compliance Cost of Itemizing Deductions: Evidence from Individual Tax Returns." *American Economic Review* 79 (5): 1224–32.

Redelmeier, Donald A., and Christopher J. Yarnell. 2012. "Road Crash Fatalities on US Income Tax Days." *Journal of the American Medical Association* 307 (14): 1486–88.

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–78.

Rees-Jones, Alex, and Dmitry Taubinsky. 2019. "Measuring 'Schmeduling'." NBER Working Paper 22884.

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

Slemrod, Joel, and Jon Bakija. 2008. *Taxing Ourselves: A Citizen's Guide to the Debate Over Taxes*. Cambridge: MIT Press.

Slemrod, Joel, Charles Christian, Rebecca London, and Jonathan A. Parker. 1997. "April 15 Syndrome." *Economic Inquiry* 35 (4): 695–709.

Slemrod, Joel, and Nikki Sorum. 1984. "The Compliance Cost of the U.S. Individual Income Tax System." *National Tax Journal* 37 (4): 461–74.

Tazhitdinova, Alisa. 2015. "Adjust Me if I Can't: The Effect of Firm Incentives on Labor Supply Responses to Taxes." https://mpra.ub.uni-muenchen.de/81611/1/MPRA_paper_81611.pdf.

Tazhitdinova, Alisa. 2018. "Reducing Evasion through Self-Reporting: Evidence from Charitable Contributions." *Journal of Public Economics* 165 (C): 31–47.

Tazhitdinova, Alisa. 2020. "Are Changes of Organizational Form Costly? Income Shifting and Business Entry Responses to Taxes." *Journal of Public Economics* 186 (C): 104–87.