Abandoned Retirement Savings*

Lucas Goodman Anita Mukherjee Shanthi Ramnath

2021-01-28

Preliminary: Please do not cite or circulate.

Abstract

The shift towards defined contributions systems has resulted in individual retirement savings often being spread across multiple accounts. In parallel, the increasing use of default enrollment has led to non-salient savings. These forces raise the concern that people may forget or lose track of these retirement accounts. This paper makes three key empirical contributions on the study of abandoned retirement savings. First, we use individual tax data on Individual Retirement Accounts (IRAs) to study abandonment prevalence and correlates. We establish that in 2017, 2.7% of 72.5 year-olds abandoned $790 million; conditional on abandonment, the median account held $5,400. Second, we use a regression discontinuity design to show that abandonment is ten times higher in automatic rollover accounts. Third, we use state unclaimed property data to show that these primary portals are currently limited in their ability to reduce abandonment. We nest our findings in a model of retirement savings featuring forgetting to derive policy implications for passive and active savers.

JEL Codes: H24, H31, J14, J63
Keywords: retirement savings, defaults, memory, escheatment

*Goodman: U.S. Department of the Treasury, lucas.goodman@treasury.gov. Mukherjee: University of Wisconsin-Madison, anita.mukherjee@wisc.edu. Ramnath: Federal Reserve Bank of Chicago, sramnath@frbchi.org. We thank J. Michael Collins, Maria Fitzpatrick, Bill Gale, Damon Jones, Olivia Mitchell, Jim Poterba, John Shoven, Justin Sydnor, and seminar participants at the 2020 Stanford Working Longer and Retirement Conference, 2019 National Tax Association conference, 2018 and 2019 Social Security Administration conferences, and the University of Wisconsin-Madison for helpful comments. Corina Mommaerts provided extraordinary guidance on the state unclaimed property analysis. Lois Miller, Arpita Patnaik, and Alan Yang provided excellent research assistance. Erin Egan provided helpful information about state unclaimed property databases. We are grateful to the U.S. Social Security Administration and Wisconsin Alumni Research Fund for funding. All views are our own and not of the funding agencies. The views expressed in this paper do not necessarily represent the official position of the U.S. Department of the Treasury, The Federal Reserve Bank of Chicago, or the Federal Reserve System.

1 Introduction

A substantial share of private retirement savings in the U.S. is accumulated in tax-preferred retirement savings plans. Decisions regarding such plans, including funding, investment, and withdrawal decisions, rest almost entirely on individual account holders and failing to adequately manage these accounts can be costly. In the worst case, if an account is forgotten (i.e., abandoned or unclaimed), then resources available to retirees could be greatly reduced and threaten their overall retirement adequacy. Numerous policies exist to reduce frictions individuals may face when accessing their own retirement savings.[1] Nonetheless, relatively little is known about the prevalence of abandoned retirement accounts.[2]

The paper fills a gap in our understanding of the scope of abandoned retirement accounts in the U.S. We analyze multiple datasets derived from IRS administrative records and state unclaimed property databases to study the prevalence of abandoned accounts in the aging population. We find that 2.7% of 72.5-year-old IRA-holders in 2017 owned $790 million ($730 million in 2012 dollars, the unit of our analysis) in abandoned accounts. Conditional on abandonment, the median balance was $5,351.[3] The abandoned amounts are also meaningful as a share of income – the median account was 24% of income in 2003-06 and 12% in 2014-17. These correspond to estimated annuity values of 2.0% and 1.1% of income, respectively. To our knowledge, our paper is among the first to provide comprehensive estimates on the share of retirement accounts abandoned within the aging population.

We explore various factors that are correlated with abandonment. Abandonment decreases in account value, but remains substantial, even for higher value accounts – e.g., it is 3% for accounts worth $10,000 in 2017. Holding account value fixed, we find that measures of financial sophistication observable in the tax data – filing a return, earning capital income, and paying estimated tax – are correlated negatively with abandonment, as expected. Moreover, we find a strong relationship between abandonment and race, as proxied by the share non-Hispanic white within a person's zip code. All else equal, those living in zip codes that are 50% white are more than twice as likely to abandon an account than those living in zip codes that are 100% white. This correlation is diminished only modestly when adding further controls, such as zip-code level education, income, and population density.

[1] For example, states encourage plans to send abandoned accounts to unclaimed property offices. Additionally, the bipartisan Retirement Savings Lost and Found Act of 2020 was proposed to create a portal to locate retirement accounts.

[2] For example, Office (2019) states that "little is known about the amount of unclaimed retirement savings ultimately transferred to states as unclaimed property".

[3] To protect taxpayer privacy, all quantiles in this paper derived from tax data are quasi-percentiles, equal to the mean of the 30 individuals closest to the reported percentile.

Next, we find that among abandoned accounts, roughly 30% of those valued $100 or less and 60% of those valued at $3,000 or more are eventually claimed within 10 years.[4] We estimate IRAs worth $1.79 billion were abandoned cumulatively between 2003 and 2008. Of these, 60.7%, or $1.09 billion, were reclaimed within ten years (or would have been reclaimed if the individual remained alive for that period). Still, roughly 40% of high-value IRAs remain unclaimed even after 10 years.

Turning to unclaimed property data, we study retirement accounts that plans have turned over to state governments (i.e. escheated accounts). We find that 36% of individuals have account balances less than $100, suggesting that plans turn over accounts to state government escheat accounts that have management costs exceeding returns. There are almost no unclaimed accounts worth more than $10,000, though there is substantial density in this range in the tax data. In total, the unclaimed property data contain $4.4 million in abandoned IRAs compared to the $170 million in the tax data.

Next, we look at two specific states with different reclaiming mechanisms for unclaimed property to measure the amount of reclaiming that occurs among escheated accounts. We choose Massachusetts and Wisconsin as they represent two extremes measured in terms of the effort a state expends in finding account owners, where Wisconsin is a "high" effort state and Massachusetts is a "low" effort state. In Massachusetts only 3.4% of unclaimed retirement accounts reported in 2016 were claimed within two years, whereas in Wisconsin, 67% of accounts were claimed in that same time frame. Overall, these findings suggest that even when taking into consideration lost accounts that have been reunited with owners, there is still a substantial amount of money that is left unclaimed. Furthermore, the pattern of reclaiming suggests some level of rational abandonment, but a substantial level of forgetting.

Finally, we study whether individuals responding to a specific default policy – forced-transfer IRAs – appear more likely to forget an account. The policy allows employers to create a default option for employees leaving their jobs: if the separating employee takes no action, an account with a balance between $1,000 and $5,000 is allowed to be transferred automatically into an IRA.[5] This policy creates two empirical discontinuities, which we use to estimate the abandonment behavior among those induced into a forced-transfer IRA ("passive savers"). Under the admittedly strong assumption that any difference in behavior

[4]We measure "claiming" as the presence of a distribution taken, which is observed directly in the tax data.

[5]Employees who leave their jobs always have the option to make an active choice to take a cash distribution or to roll their account over into an IRA of their choosing. If the account has a balance less than $1,000, this default option is allowed to be a full distribution in cash. These "force-out" policies are meant to allow firms to rid themselves of small accounts, which can have relatively high administrative costs. If the account balance is greater than $5,000 then the employer must offer the option to maintain the account.

(conditional on amount of the rollover) is due to abandonment, we estimate abandonment rates that are much higher for defaulted savings than in the general population – ranging from 23% to 45% depending on the specification and threshold used.

This analysis of forced-transfer IRAs serves two purposes. First, because those affected by this default are primarily working-age, it suggests that abandonment could increase in importance as the current working-age population reaches retirement. Second, it informs welfare implications of default policies in general. In the final section of our paper, we show the conditions under which the presence of forgetting can reverse the welfare benefits of default policies, taking into account *both* myopia and forgetting. Forgetting in our model could also represent other mechanisms that lead to abandonment such as hassle costs, information frictions, or rollover difficulties. We show that the conceptual implications of abandonment differ for passive and active savers: specifically, the latter group may under-save out of fear of failing to manage and remember the account.

This study relates to several strands of literature within economics. The first connection is to passive behavior and inertia in retirement saving, which connects to potential abandonment later in the lifecycle. Passive saving, namely through default enrollment in retirement plans, has been shown to substantially increase participation (Thaler and Benartzi, 2004; Chetty et al., 2014; Madrian and Shea, 2001; Benartzi and Thaler, 2007). The benefits of auto-enrollment are complemented by research demonstrating inertia in retirement plan choices (Kim et al., 2016) and other settings (e.g., in tax withholding as studied in Jones 2012). Most of this prior work studies consumer choice and policy design, which an eye toward shifting behavior towards what is deemed more optimal. There is continued debate about default enrollment as a feature in plan design, however (Bubb and Warren, 2020; Scott et al., 2020; Bernheim and Gastell, 2020). We contribute to this literature by emphasizing a less recognized drawback of default policies such as auto-enrollment: the possibility that such accounts, due to being less salient (Ekerdt and Hackney, 2002), are more vulnerable to becoming abandoned. We also note prior research demonstrating for example that auto-enrollment increases savings among the financially less literate, creating concerns about management of those funds over the lifecycle (Goda et al., 2020; Carroll et al., 2009).

Our study is also related to the literature within economics on forgetting, hassle costs, and generally "leaving money on the table". Compelling evidence of forgetting has been documented in settings such as payment choices for a task (Ericson, 2011) and lapse-based insurance (Gottlieb and Smetters, 2014), with the main takeaway being that people exhibit overconfidence in their prospective memory. Ericson (2017) shows theoretically that ignoring limited memory can lead us to overestimate present bias – that paper also considers the

effects of procrastination, which is likely present in behaviors such as retirement account consolidation. Yet, prior work also shows a relative unwillingness to invest in reminder policies that could help curb such forgetting (Rogers and Milkman, 2016). Our paper contributes to this literature by incorporating forgetting in the retirement savings context, which has implications for various policies including defaults, reminders, and deadlines to rollover funds. We note that hassle costs could also explain at least some abandoned accounts. These costs have shown to be important in tax filing (Benzarti, 2020), health insurance (Baicker et al., 2012), and unemployment insurance (Ebenstein and Stange, 2010). Hassle costs in the present context are exacerbated by information frictions such as those faced by retirement account holders attempting to locate and manage their savings.

Finally, our research is tied more generally to retirement savings adequacy (Poterba, 2014) and consumption smoothing in retirement (Banks et al., 1998). Prior studies have examined whether households are saving enough for retirement (Scholz et al., 2006; Skinner, 2007), while others have studied the role of specific factors such as financial literacy (Lusardi and Mitchell 2014, Clark et al. 2006), information (Mastrobuoni, 2011), and "leakages" (i.e., cash-outs) at job separation (Armour et al., 2016; Clark et al., 2014; Munnell and Webb, 2015) or more generally prior to retirement (Goodman et al., 2019).

The rest of the paper is organized as follows. In the next section, we provide background on unclaimed retirement accounts. Section 3 describes the individual level tax data and the state unclaimed property data. Section 4 contains analysis related to the prevalence, trends, and correlates of abandoned accounts. Section 5 presents analysis on a specific type of default retirement saving account, the forced transfer IRA, to show how abandonment changes with default enrollment. Section 6 provides a policy discussion and Section 7 concludes.

2 Institutional Background

Unclaimed retirement savings are "savings that individuals are entitled to receive, but have not claimed because employers or other entities that maintain their retirement accounts cannot locate the individuals or because the individuals have forgotten about the savings" (Office, 2019).[6] However, whether or not an individual "claims" a retirement account is not a well-defined concept on its own. In particular, retirement accounts are long-term savings vehicles that are designed to go untouched for decades. In fact, distributions that take place before an individual turns 59.5 are generally penalized, though some exceptions

[6]We add to this definition the case when individuals may not have forgotten the savings, but cannot locate it or otherwise face high barriers to claiming it.

apply. On the other hand, withdrawals from accounts *must* begin at age 70.5 or the year the individual retires, and, the minimum amount an individual must withdraw each year is called a "Required Minimum Distribution" (RMD).

States have used RMD rules to help determine whether a retirement account should be considered an unclaimed asset. In particular, states specify a dormancy period, typically three to five years, after the date for when an RMD should haven taken place. If no distribution is taken at any point before that extended period, then states require that the participant's assets be remitted to the state, i.e. plans must escheat unclaimed funds. Each state then hosts an unclaimed property database aimed to link individuals with their unclaimed assets. The state unclaimed property databases generate meaningful activity, with owners having claimed $25 million in retirement savings in 2016 according to data from 15 states (2019).[7] These unclaimed accounts had an average value of $601 from 401(k) plans and $5,817 from traditional IRAs.

Reducing the number of unclaimed retirement accounts is a key policy concern in the U.S.. Congress is currently debating The Retirement Savings Lost and Found Act of 2018, which is focused entirely on this issue. The Act would expand an existing online database of pension and 401(k) account owners to help facilitate matching owners with their lost plans.[8] The proposed legislation would also clarify the rules for categorizing account owners as "missing" and place a greater burden on employers and plan managers to find these individuals.[9]

Our paper adds to this policy discussion by providing novel estimates of the extent to which unclaimed retirement accounts are prevalent in the U.S. Using a combination of administrative data sources, we are able to identify abandoned accounts that have not yet been escheated to states as well as determine the efficacy of escheatment for reunited owners with lost accounts.

3 Data Description

For our analysis, we make use of two administrative data sources: (1) federal tax and information return data and (2) state unclaimed property data. Each dataset comes with its own set of challenges, which we will describe in more detail below. Using both datasets

[7]Note that this report was limited to a description of aggregated state-level data, not account-level data as we study in this paper.

[8]This proposal goes far beyond the current PBGC registry, which contains pension plan information for only terminating plans and just recently started collecting information on DC plans.

[9]The full text of this Act is available here: `https://www.govtrack.us/congress/bills/115/s2474`.

in tandem, however, will allow for a more comprehensive analysis of unclaimed retirement accounts in the U.S.

3.1 Administrative Tax Data

We use data drawn from a database containing the near-universe of tax returns and information returns. In general, our analysis spans tax years 1999 through 2018, though we restrict to subsets of years for certain purposes. Tax returns from the Form 1040 are used to identify marital status and income. We are also able to link both primary and secondary filers to a given tax return.

Information returns are separate from tax returns and include forms sent by a third-party to the IRS and, in some cases, also to the relevant individual. We make heavy use of two information returns in particular. First is the Form 1099-R, which reports information on the distributions made from pensions, IRAs, and other similar accounts. This form is used to identify the amount of a distribution and whether it came from an IRA (versus another account). The custodian also reports (in Box 7) up to two codes which give further information about the nature of the distribution. In particular, we use the presence of Box 7 codes "G" or "H" to infer that the distribution was part of a direct rollover. The Form 1099-R also includes an identifier for the custodian of retirement funds, which we can link across IRS datasets. We use Form 1099-R to identify whether an individual received a distribution from a given IRA; we also use this form to determine the running variable in our analysis of forced transfer IRAs.

The second information return we use is Form 5498, which reports information about IRAs held by an individual. Importantly, this form identifies the type of IRA ("traditional", Roth, or certain specialized types of small-business IRAs), the value of the IRA as of the end of the year, and any contributions made (rollover or otherwise) during the year. Form 5498 is sent to the IRS every year in which the IRA maintains a positive value, though the account holder typically receives a copy only when a contribution is made. As with Form 1099-R, Form 5498 also includes an identifier for the custodian of the account.

For our analysis, we will focus on IRAs largely because of data availability. We observe IRA account-holders on Form 5498, even for individuals who are not interacting with the account. By contrast, we do not have analogous information on holders of DC plans held through employers. Fortunately, as estimated by Goodman et al. (2019), nearly 90 percent of distributions from the joint DC-IRA system to those age 65 or greater in 2015 were from

IRAs, not DC accounts, suggesting that most individuals have rolled their DC plans over into an IRA prior to reaching retirement-age.

Finally, we make use of other components of the IRS database for our analyses. For example, our sample construction makes use the Death Master file, which records each individual's date of birth, date of death, and sex. Additionally, Some of our analyses require controls for Social Security income, which we obtain from Form 1099-SSA (an information return).

3.2 Defining Unclaimed Accounts in Tax Data

Unlike in the state unclaimed property data, where accounts are by definition unclaimed, in the tax data we must infer which accounts have been abandoned. We begin with the universe of traditional IRA accounts held by individuals at age 72.5 (whose date of death, if applicable, is later than the year in question), which are indexed by the combination of individual i and custodian j.[10] We define an account ij to be abandoned if it satisfies each of the three following conditions:

1. The account ij has existed, with positive value, for at least four years (i.e., the 69.5 year, the 70.5 year, and the 71.5 year, in addition to the 72.5 year).
2. There have been no distributions from that account ij (as measured on Form 1099-R) over those four years.
3. The individual i has not satisfied his or her RMD in any year (i.e., the 70.5 year, the 71.5 year, and the 72.5 year).[11]

Using this criterion, in Section 4 we estimate the prevalence of unclaimed accounts in the U.S. as observed in the tax data.

3.3 State Unclaimed Property Data

Our data on escheated retirement accounts comes from state unclaimed property (SUP) databases. These data contain account-level information on each unclaimed property, and

[10] That is, if an individual receives multiple Forms 5498 from the same custodian, we aggregate to the individual-custodian level, after dropping Forms 5498 that appear to be duplicates on important variables.

[11] If an individual held only one account, then condition (2) would imply condition (3). However, the RMD is calculated with respect to *all* IRA distributions and *all* IRA assets held by the taxpayer; the taxpayer can remain in compliance with RMDs by taking a larger distribution from one IRA (j) and taking no distribution from another (j'). Our algorithm conservatively would *not* consider j' to be abandoned. Furthermore, for this purpose, we measure compliance with RMDs using the larger of 1099-R distributions and taxable IRA distributions reported on Form 1040.

include details such as the type of account (IRA or pension, for example), account value, and names and addresses of the account owner (e.g., an individual) and account holder (e.g., a firm or plan manager). The name and geographic information of the account owner enable inference of related account characteristics such as the owner's gender and neighborhood covariates. The data are collected separately from each state's division of unclaimed property (or related agency). An advantage of the SUP data is that all states use a standard reporting format following guidance from the National Association of Unclaimed Property Administrators (NAUPA). Our sample includes as many states as possible (up to 13) for different analyses.[12]

We use property codes associated with each unclaimed property to identify retirement accounts. The SUP data contain numerous property types – NAUPA lists 123 categories – including uncashed checks, securities, insurance property, mineral proceeds, and trusts. We isolate retirement accounts according to the codes listed in Table B.4, which include pension checks, Individual Retirement Accounts (IRAs), pension and profit-sharing plans, and annuities.[13] We supplement these data with information from the American Community Survey, the Census, and Social Security names databases to further ascertain the characteristics of these unclaimed account owners. We also analyze information on claims from unclaimed property from two states, Massachusetts and Wisconsin, to better understand frictions in these processes.

While much of our primary analysis will focus on results derived from tax data, an advantage of the state unclaimed property data over the tax data is that we can look at all types of retirement plans, not just IRAs. In particular, we will use more detailed data from Massachusetts and Wisconsin to help shed light on factors that influence accounts that are eventually reunited with owners. The additional data from these two states includes information on properties that were reclaimed after being sent to state unclaimed property. The Massachusetts's data are available from 1999 to 2018 and the Wisconsin data are available from 2010 to 2018.

Table 4 shows summary statistics on claiming for the two states. In Massachusetts, there were 3,320 retirement accounts reported as unclaimed to the state in 2016, of which

[12]States have different policies for sharing data with researchers. The 13 states that responded to our requests for data containing account-level property type and account value of unclaimed properties include: AK, CA, CT, FL, HI, LA, MA, MN, MO, NC, ND, NV, NY, OH, PA, TX, and WI.

[13]The number of retirement account codes has expanded over time (for example, the "IR" codes were introduced in 2010 and gradually adopted by states since that year). Two categories in Table B.4 merit further description based on our understanding: the first is "pension checks", which represents uncashed checks sent by plans to encourage required minimum distributions. The second is "pension and profit-sharing plans", which despite the name, consists of defined contribution plans (Office, 2019).

only 3.4% were claimed within two years. The average account value of claimed accounts is much higher at $2,110 than unclaimed accounts ($581), which is consistent with some lower-value accounts being abandoned or forgotten as owners need to initiate their own claims in the state. The Wisconsin claim data are available from 2016 to 2018, limiting our ability to study long-term claiming behavior. We observe that in 2016, there were 815 unclaimed retirement accounts reported to the state, of which two-thirds were claimed within two years. The average account value is higher than the national average at $980, and the ones that remain unclaimed are of higher value ($1,315 versus $812). As noted, the state uses taxpayer Social Security numbers to match account owners with escheated funds whenever possible; this method appears successful, as 54%[14] of accounts are reunited with their owners via the match. Of the accounts that are claimed, 80% are done so via the DOR Auto Match; an additional 12% of account owners initiate claims on via the state's website, and 6.5% of owners are connected to their lost accounts via an online locator service.

.

4 Abandoned Account Prevalence and Correlates

The left panel of Figure 1 plots the abandonment share over time using administrative tax data. The solid line plots the share of abandoned accounts, beginning from 1.0% in 2003 and more than doubling to 2.2% in 2017. The dashed line aggregates the accounts to the individual level and plots the share of individuals who owned at least one abandoned account. This share was 1.2% in 2003, and more than doubled to 2.7% in 2017. The right panel of Figure 1 plots the time trend in the dollar value of abandoned accounts (in 2012 dollars). This value increases from $285 million in 2003 to $730 million in 2017. In each of these two years, the aggregate amount abandoned represents approximately 0.35% of the total value of IRAs held by age 72.5 individuals. Note that the figures contain dashed vertical lines separating out the period 2009 through 2011. The reason is that the look-back period for these years (which is used to determine whether the individual satisfied an RMD or took a distribution) includes 2009, when RMDs were suspended by legislation. Thus, for these years, failing to take the RMD for two years would cause an account to be deemed as abandoned; in all other years, this categorization would require failing to take the RMD for three years. Therefore, in the analysis that follows, we drop observations for cohorts that were aged 72.5 in 2009, 2010, or 2011.

[14]This equals the proportion claimed, 0.67, multiplied by the proportion claimed via DOR Auto Match, 0.80, from Table 4.

We find clear evidence that a non-trivial share of retirement accounts become abandoned by the time an account owner turns age 72.5. If abandonment is systematically higher among low-value accounts, we could infer that individuals rationally abandon some accounts if the perceived costs of managing that account exceed its value. Abandoned low-value accounts could also reflect forgetting default savings from short duration jobs, including rollover IRAs. Alternatively, if there is no relationship between account value and the probability an account goes unclaimed, we could infer that abandonment is random and due fully to forgetting. Figure 2 plots the abandonment share by value of that account at age 72.5, separately for the early period of our sample (2003-2006) and the later part of our sample (2014-2017). The relationship appears linear on the log-log scale, with abandonment decreasing in account size. Nonetheless, abandonment is substantial for accounts that are large in value: the abandonment share for accounts worth $10,000 is about 1.5% in the early sample and 3% in the later sample.

We provide individual-level context for abandoned account values in Table 1. Column (1) presents the 25th, 50th, 75th, and 90th percentile of the value of abandoned accounts; column (4) presents the same quantiles for accounts that are not abandoned. Panel A aggregates from 2003 through 2006, while Panel B aggregates from 2014-2017. The table shows that, in both periods, the median abandoned account was much smaller than the median non-abandoned account. Furthermore, except at the 90th percentile, each quantile of abandoned accounts shrank in value from the early part of our sample (2003-2006) to the late part of our sample (2014-2017). Columns (2) and (5) report various quantiles of the ratio of account value to income for abandoned accounts and non-abandoned accounts, respectively.[15] As an alternative measure, in columns (3) and (6), we convert the gross IRA amount to an annuity value, using the Social Security life expectancy tables and a (conservative) assumption of a 1% interest rate. The median abandoned account was equal to 24.1% of income in the 2003-2006 period and 12.7% of income in the 2014-2017 period; these correspond to annuity values of 2.0% and 1.1% of income, respectively. We also observe that a non-trivial minority of abandoned accounts are much larger. At the 75^{th} percentile, the annuity value of an abandoned IRA represents 6.5% (2003-2006) or 4.8% (2014-2017) of income; at the 90^{th} percentile, these values are 16.7% and 14.5%.

Finally, we estimate the prevalence of abandonment using the state unclaimed property data.[16] In our sample of 13 states, about 36,500 retirement accounts totaling $18.3 million

[15] Our measure of income is adjusted gross income, plus non-taxable Social Security, minus taxable IRA distributions. The measure is meant to capture disposable income and therefore includes income sources that are not taxed but still available for consumption.

[16] The analysis is in Table B.1.

were escheated in 2016. The summary statistics reveal that abandonment is a problem that extends beyond IRAs, as expected – only 40% of the escheated accounts are of this type. Next, only for the purposes of producing a national estimate, we assume that these accounts belong to 74 year-olds and extrapolate the account-level data to the entire US using a log-linear model. In this exercise, we estimate that nearly 70,000 retirement accounts totaling $38 million were escheated in 2016. This amounts to about 3.3% of 74 year-olds having an abandoned retirement account in 2016 with an average value of $547. Note that this 3.3% estimate is not directly comparable to the tax-data based analysis because (1) low-value accounts are much more likely to be escheated due to federal guidance, as we will show, and (2) the escheated accounts are not limited to IRAs.

4.1 Decomposing the Sources of Abandonment

As shown in the left panel of Figure 1, we find that roughly 3% of all IRA account owners subject to an RMD had an unclaimed account in 2017. In this section, we examine individual-level factors that are correlated with this abandonment. Table 2 shows estimates of a regression of abandonment on various covariates, all conditional on year fixed effects interacted with fine bins of account value, meaning that these results can be interpreted as holding account value fixed. The first column includes four proxies for financial sophistication: filing a tax return, paying estimated tax during the year, having non-zero capital gains or dividend income, and having interest income.[17] The results show correlations in the expected direction: each of these proxies predicts abandonment. Additionally, the magnitude of these coefficients stays relatively similar across the columns, as additional controls are added.

The second column adds two demographic variables: sex and race. We observe sex for each individual, but we do not observe race. Instead, we use the zip code level race shares as a proxy. We find that males are slightly (0.1 percentage points) more likely to abandon accounts. More substantially, we find that the white share of a zip code strongly predicts abandonment. Moving from a zip code that is 50% white to a zip code that is 100% white (as measured by the ACS) reduces abandonment by 1.3 percentage points – a substantial effect relative to the mean abandonment of 1.55 percentage points. Column 3 adds a covariate for the population density of the zip code, which reduces the coefficient on the white share by about 15%. Column 4 adds additional zip code level demographics, including education and poverty status. This has very little effect on any of the previous coefficients. Lastly, column 5 adds fixed effects for year interacted with the custodian (j); this specification reveals the

[17] For this exercise, we use the sample from 2012 and on. Certain key variables, most notably the amount of estimated tax paid, are missing in earlier years.

extent to which the other covariates are working through the channel of different individuals selecting into different financial service firms. This specification does have a modest effect reducing the partial correlation between abandonment and the presence of capital income, reducing the magnitude of the coefficients on having dividends/capital gains or interest income. Additionally, this set of fixed effects reduces the coefficient on the share white by an additional 18%, suggesting that a modest share of the component of the race correlation is driven by selection into firms. That is, firms that serve more minority neighborhoods appear to be associated with greater abandonment, all else equal.

To best understand the correlates of abandoned accounts, we also analyze the state unclaimed property data. As with the tax data, we supplement the dataset with county-level demographic characteristics from the American Community Survey. Since we are able to use the first and last names of each account holder in this dataset, we also predict gender and race using this micro-level information. We find that individuals with unclaimed retirement assets are 10 percentage points less likely to be female, 12 percentage points more likely to be of Hispanic origin, and more likely to have a last known address in larger, more urban counties. Other county-level characteristics, such as the age makeup, educational attainment, migration, and median household income are similar between individuals with an unclaimed retirement asset and the population-weighted average county characteristics.[18] The demographic patterns suggest that the escheated accounts were more likely to be held by large plans in urban areas. As mentioned before, we highlight that the state unclaimed property data and the tax data are not comparable due to the known differences in account type, account value, and plan selection in the escheatment decision.

Lastly, one of the most striking patterns in Figure 1 is the more-than-doubling of abandonment of IRAs between 2003 and 2017. One possible explanation for this increase is that there has been a proliferation of low-balance accounts. However, Figure 2 shows that this cannot explain the entirety of the increase. In particular, the figure showed that the probability of abandonment conditional on IRA value increased through (almost) the entire distribution of IRA values. That is, holding IRA value fixed, the probability of abandonment still increased from 2003 to 2017. Additionally, we have performed a Oaxaca-Blinder style decomposition, decomposing the increase into the amount explained by changes in observable characteristics, such as account value and financial sophistication proxies, and the residual (unexplained) increase. We find that virtually 100 percent of the increase is unexplained by

[18]The analysis is in Table B.2.

compositional changes.[19] Further explanations for the increase in abandonment over time is a fruitful area for further research.

4.2 Reclaiming Abandoned Accounts

Next, we estimate the "reclaiming rate", i.e., rate at which abandoned accounts are subsequently claimed by their owners, in the tax data. The reclaiming rate is important for a number of reasons.[20] First, a reclaiming rate substantially less than 100% offers validity to our measurement of abandonment. Second, the welfare consequences of abandonment depend on when–and whether – the accounts are reclaimed. If the accounts are never reclaimed, then the savings within the IRA can be considered lost. On the other hand, if the accounts are reclaimed, then the individual may regain that consumption that would have been otherwise lost, though she may have made suboptimal consumption decisions while the account went unclaimed. We proxy for "reclaiming" by creating an indicator for whether the owner of an abandoned account receives some distribution from the account in question. For each year relative to the abandonment year (age 72.5), we compute the hazard of reclaiming as equal to the share of accounts with a distribution in that year, conditional on (1) not yet having been claimed and (2) the individual being alive for the entire year. This definition is subject to some measurement error in both directions. One important caveat to this definition is that the act of escheatment might lead to a distribution, which we would erroneously categorize as "reclaiming".[21] The left panel of Figure 3 plots the cumulative hazard of reclaiming over a ten-year period. The one-year hazard is just under 20%, meaning that over 80% of abandoned accounts are not claimed within one year. This cumulative hazard increases steadily to approximately 58% over 10 years. Thus, a slight majority of accounts are claimed within ten years (conditional on not dying prior to that point); however, a substantial minority are not claimed. Additionally, many more accounts remain unclaimed by those who die within ten years of the 72.5 year.

The right panel of Figure 3 plots the 10-year cumulative reclaiming hazard separately by 25 bins of account value. The lowest-value abandoned IRAs, with account balances under $100, have a reclaiming rate of approximately 30%; this increases to 60% for accounts with

[19] We present this analysis in Figure B.1.

[20] A key trigger in our definition of abandonment is whether the account owner has failed to take an RMD by age 72.5. Individuals may miss these early deadlines and take their RMDs later on, however. The cost to individuals of such behavior is the heavy penalties associated with missing RMDs, though in practice these penalties have a low probability of being levied by the IRS. The cost to the government of such behavior is approximately equal to the foregone tax revenue on the inside build-up of assets within the account.

[21] In 2018, the IRS issued Revenue Ruling 2018-17, which clarified that escheatment events should trigger a Form 1099-R issued in the individual's name. However, this ruling occurred after our sample period. In addition, we might fail to identify true reclaiming if the custodian of the account changes EIN.

balances around $3,000, after which point it remains approximately constant. This suggests that one component of failure to reclaim is rational inattention, but this cannot explain the 40% of high-value IRAs that remain unclaimed after 10 years. In sum, IRAs worth $1.79 billion were abandoned cumulatively between 2003 and 2008. Of these, we estimate that 60.7%, or $1.09 billion, were reclaimed within ten years (or would have been reclaimed if the individual remained alive for that period).

4.3 Reclaiming Escheated Accounts

Once accounts are escheated, we can no longer observe them in the tax data. Therefore, we turn to data from unclaimed property databases to better understand what happens to abandoned accounts that are passed on to states. Another advantage of the SUP data is that it includes all escheated accounts, as opposed to the tax data that is limited to IRAs. A natural question is to what extent do plans escheat abandoned retirement assets. While state policies prescribe escheatment after a dormancy period of two to five years, federal guidance makes escheatment optional. Plans may choose to do so when accounts are too costly to manage, or when they experience a structural change such as a merger or termination.

In Figure 7, we plot the number of individuals with abandoned IRAs that appear in the escheatment and tax datasets. Note that these samples are not subsets of each other; the tax data are for individuals aged 72.5, which is prior to when escheatment would typically occur.[22] Age of account holders for the escheatment data are not available. An immediate observation is that 36% of individuals in the escheatment data have account balances less than $100, suggesting that plans escheat accounts that have management costs exceeding returns. There are almost no escheated accounts worth more than $10,000, though there is substantial density in this range in the tax data. In total, the escheatment data in Figure 7 contain $4.4 million in abandoned IRAs compared to the $170 million in the tax data. A key takeaway from this chart is that state unclaimed property databases are not currently the most effective means of uniting abandoned accounts holders with their lost funds. In part this is because a large portion of the escheated funds are likely to be rationally abandoned, as the utility costs of claiming exceed the small value. For example, most state unclaimed property departments require account holders to provide proof of ownership via pay stubs or bank account statements to claim funds.

Next, we consider variation in reclaiming by state policy. In general, each state employs

[22]Note that we use six states in this analysis: FL, LA, MA, MN, OH, and TX. The reduced sample size is primarily due to keeping only account holders who also appear in the tax data. WI and HI are omitted to preserve account holder confidentiality; they each have less than 30 escheated accounts in the year studied.

a different procedure for reuniting escheated accounts with their owners. We focus on Massachusetts and Wisconsin because they help illustrate the spectrum of such policies. In Massachusetts, like most states, unclaimed account owners must initiate claims to their funds through the state. By contrast, in Wisconsin, the state uses Social Security number information to match unclaimed funds with their owners. This process started in June 2015 and is rare among states; it is known as the "Wisconsin model". These procedures result in stark differences: in Massachusetts only 3.4% of unclaimed retirement accounts reported in 2016 were claimed within two years, whereas in Wisconsin, 67% of accounts were claimed in that same time frame.

As we saw in Figure 3, accounts of higher value are more likely to be reclaimed in a ten year horizon. We investigate whether this same relationship holds true for accounts that have been escheated. Table 4 suggests this to be the case, however we provide a formal test below. The Massachusetts data are more representative of the country as most states are unable to use Social Security numbers to match unclaimed funds with their owners (i.e. the Wisconsin Model). As such, we continue with analysis of the Massachusetts data to examine whether account value predicts claiming behavior.[23] We estimate the following regression for each unclaimed account i separately for cutoffs of 1, 2,..., 13 years since the property was reported as unclaimed:

$$\text{Claimed}_i = \beta \ln(\text{Account Value}_i) + \eta_p + \gamma_y, \tag{1}$$

where $Claimed_i$ is whether the property was claimed within 1, 2,..., 13 years (13 separate regressions, with decreasing sample size), and $AccountValue$ is in units \$10,000, though the units do not matter as it is logged.[24] The fixed effects η_p and γ_y are for property codes and year reported as unclaimed, respectively. Figure 8 (with corresponding regression results in

[23]The relationship between account value and claiming in the Wisconsin data requires care in interpretation because the majority of claims are done automatically via Social Security number matches and not by owner initiation. We observe that the auto-matched claims are of smallest average value, which makes sense as the state claims these on behalf of the account owner as long as there is a Social Security number to match. Accounts reclaimed by the account owner initiating the process have higher value.

[24]We also estimate an accelerated failure time (AFT) model of the following form for each unclaimed account: $\log(t_i) = \beta\text{Account Value}_i + \eta_p + \gamma_y$, where t is the years until claim if the account is claimed, or years until 2018, the final year of observation, if the claim is not observed. *Account Value* is measured in units of \$10,000, and η and γ are the same as in equation (1). The AFT model takes into account the censored data (i.e., we do not observe all claims); the results are reported in Table B.3. An advantage of the AFT model is that we can interpret the coefficient of interest β as the impact on the time to an event, in our case a claim; the coefficient of -0.0893 in column (3) means that the time to claim decreases by 8.93% for each \$10,000 in unclaimed account value. These results appear monotonic; Figure B.2 plots the Kaplan-Meier hazard rates of claiming by account value, and we observe that the probability of claim at every year since reported as unclaimed increases with account value.

Table B.5) shows that the relationship between account value and claiming increases with time since the property is reported as unclaimed. In other words, higher value accounts are more likely than lower value accounts to be claimed many years after being reported as unclaimed. This pattern of abandoning lower-value accounts suggests rational abandonment due to hassle costs and other frictions in rollover or reclaiming.

5 Defaults and Account Abandonment

The previous section reported our best estimates of abandonment of IRAs among RMD-aged individuals. As the U.S. retirement system moves to a participant-driven model, the proliferation of default policies to encourage individual savings may become an important factor driving account abandonment. In particular, passive savers, whose behavior is impacted by such policies as this group may be at high risk of abandoning accounts.In this section, we test that hypothesis directly. We analyze a specific default policy where employers are allowed to force out accounts of employees who leave their jobs with DC balances under $5,000. For balances under $1,000, this force-out distribution is allowed to be in cash. For balances between $1,000 and $5,000, this force-out distribution must take the form of a forced transfer to an IRA established for this purpose. This policy variation is illustrated in Figure 4. We focus on the subpopulation of mostly working-age individuals who are induced into transferring their savings into an IRA due to their passive behavior. While, we cannot observe whether any given individual is a passive saver, the empirical patterns will allow us to recover the average level of account abandonment among passive savers at each of the $1,000 and $5,000 thresholds.

5.1 Empirical methodology

We recover the probability of abandonment among passive savers using the following procedure. We begin by tabulating all individuals who (1) make a rollover distribution between 2005 and 2015 and who (2) receive a rollover IRA contribution within 20% of the amount given in (1).[25]. The resulting dataset includes all individuals who made a rollover contribution to an IRA, whether induced by the forced transfer rules or not. Figure 5 plots the counts of observations in $10 bins. There are two approximate discontinuities in these counts. The counts jump immediately to the right of $1,000, reflecting some share of individuals that are induced into an IRA (rather than cashing out) by the forced transfer

[25]We then drop distributions that are even multiples of $500; such distributions predominantly represent partial distributions (i.e., distributions of less than the full account balance), which by construction are not induced by the forced transfer policy. We discuss this point further in Appendix 7

rules. Likewise, the counts fall immediately to the right of \$5,000, reflecting some individuals who would have been enrolled in a forced-transfer IRA if their accounts were worth slightly less.[26]

For the sake of exposition, we discuss the procedure in the context of the \$1,000 threshold. Those just to the left of \$1,000 include active savers who would enroll in an IRA even without the default policy–i.e. "always-takers" (AT) in the treatment effects language. Those just to the right represent a mix of both passive savers–i.e. "compliers" (C) in the treatment effects language–and active savers (AT). Thus, the mean outcome for active savers y_{AT}, is equal to the mean outcome of for those to the left of the \$1,000 threshold: y_L. The mean outcome to the right of \$1,000, y_R comprises a mix of passive and active savers so that:

$$y_R = \alpha y_C + (1 - \alpha) y_{AT},$$

where α is the share of passive savers. We can recover y_R and $y_L = y_{AT}$ using local linear regression. I.e., let x_i denote the rollover amount and let T_i denote an indicator that $x_i \geq 1000$.

$$y_i = \gamma^y + \beta^y(x_i - 1000) + \pi^y T_i + \eta^y(x_i - 1000) * T_i + \epsilon_i^y \tag{2}$$

y_L is equal to $\hat{\gamma}^y$ and y_R is equal to $\hat{\gamma}^y + \hat{\pi}^y$.[27]

Likewise, we can also recover α using regression discontinuity methods. We aggregate to the level of single dollar of rollovers to compute C_x, the count of rollover distributions equal to x.[28] We can run the following local linear regression, where each observation is a single dollar amount:

$$C_x = \gamma^{FS} + \beta^{FS}(x - 1000) + \pi^{FS} T_x + \eta^{FS}(x - 1000) * T_x + \epsilon_x^{FS} \tag{3}$$

Our estimate for α is equal to $\frac{\hat{\pi}^{FS}}{\hat{\pi}^{FS} + \hat{\gamma}^{FS}}$.

Using our estimate for α, y_L, and y_R, we can back out the mean outcome of the passive savers. We proceed with two potential outcomes. First, we look at the probability of having any interaction with the account over the next five years (a future distribution from or future

[26] We note that the discontinuity is not perfect. One possible explanation for this is that there is some (short) lag between when accounts are valued for determining whether they are eligible for a forced transfer and when the distributions are actually made, during which time asset prices may change. As discussed in Section 3.1, we do not observe the balance in employer plans, which would be the more ideal running variable.

[27] At the \$1,000 threshold, we use a triangular kernel with bandwidth \$400 and drop observations within a donut hole, $x_i \in [875, 1125]$.

[28] I.e., x takes on only integer values in our data

contribution into the IRA). Second, we look at the probability of ever updating the address with the IRA custodian within the next five years, conditional on moving at some point during that time. Both outcomes – failing to update one's address and not interacting with the account over the next several years – are likely to have a causal effect on abandonment. By failing to update the mailing address, the probability increases substantially that the custodian will be unable to communicate with the participant in the future. Likewise, future interactions with the account will tend to reduce the probability of forgetting in the future. Thus, even if these differences in conditional means did not reflect short-term abandonment, they may have causal effects on abandonment in the future.

Finally, under strong assumptions, we can produce an estimate for the share of accounts abandoned by the passive savers. Let A denote a dummy for abandonment. Let Y denote either of our two outcomes measuring interaction with accounts. We define A such that $Pr(Y = 1|A = 1) = 0$; that is, abandonment precludes any further interaction with the account. We also make the simplifying assumption that $A = 0$ for active savers. We can write y_C–the outcome of passive savers–as follows:

$$y_C = E(Y|C) = Pr(A = 1|C)E(Y|A = 1, C) + (1 - Pr(A = 1|C))E(Y|A = 0, C). \quad (4)$$

The left-hand-side of this equation is estimated directly, as discussed above. Furthermore, by assumption, $E(Y|A = 1, C)$ is zero. Given an estimate for $E(Y|A = 0, C)$, we can back out $Pr(A = 1|C)$–our object of interest. The term $E(Y|A = 0, C)$ represents the probability of Y for compliers, conditional on not abandoning the account. A natural, but strong, assumption would be that $E(Y|A = 0, C) = E(Y|AT)$. This assumption would be violated to the extent that passive and active savers are different – even conditional on non-abandonment – in ways that drive Y.

To be clear, the assumption that $E(Y|A = 0, C) = E(Y|AT)$ is strong for both of our measures for account interaction. Taking a distribution from an IRA in the near future is likely to be correlated with a myriad of unobservable characteristics, which differ between compliers and always-takers. Likewise, it may be the case that compliers are less likely to update their address even if they do not abandon the account. To the extent that $E(Y|A = 0, C) < E(Y|AT)$, our estimated abandonment share will be too high.

5.2 Results

Our estimates from equation 3 imply the share of compliers (α) at the \$1,000 threshold among the full population is 0.555. This can be read off the graph as the counts just to

the right of \$1,000 are approximately twice the counts just to the left of \$1,000. At the \$5,000 threshold, we estimate an α of 0.377. Additionally, we perform analogous estimates on the the restricted sample of people who change addresses (as indicated by Form 1040) is 0.567 (0.355), as the sample when using the outcome of updating one's address with the IRA custodian.[29] These estimates are given in Table 3.

Next we back out the probability of abandonment among passive savers at the two thresholds. The top panels of Figure 6 plot the mean probability of having any interaction with the account over the next five years. The left panel plots the data around the \$1,000 threshold and the right panel plots the data around the \$5,000 threshold. This illustrates the clear visual discontinuity in the share of individuals who subsequently interact with their account, implying that y_C is substantially less than y_{AT}, which further implies a fairly large estimated abandonment share. The bottom panels repeat these plots for the outcome of updating the IRA address (conditional on moving). These two panels give similar results as the top two panels: there is a clear visual discontinuity, implying a substantial gap between y_C and y_{AT}, and a fairly large estimated abandonment share.

The corresponding estimates for Figures 5 and 6 are presented in Table 3. The first two columns use any interaction with the account as the outcome. For passive savers, the estimated probability of interaction is .37 at the \$1,000 threshold and .36 at the \$5,000 threshold, relative to 0.60 and 0.65 for active savers, respectively. This leads to very large implied abandonment shares for passive savers: 0.38 at the \$1,000 threshold and 0.45 at the \$5,000 threshold. The second two columns use the address-based approach. In this sample (i.e., restricted to movers), the means for active savers are 0.70 and 0.75, respectively, relative to 0.45 and 0.58 for the passive savers. This leads to estimated abandonment shares of 0.36 and 0.23 at the two thresholds.

Compared to the population of older account holders in Section 4 where account abandonment was around 3%, these estimates are quite large. The magnitude of these estimates, however, may be driven to some extent by a violation of the assumption that $E(Y|C, A = 0) = E(Y|AT)$. Because passive and active savers differ along a variety of dimensions, it is plausible that passive savers would be less likely to take a distribution from the account or update their address, conditional on not abandoning. However, these differences in primitives would need to be very large in order to explain the differences in conditional means in the absence of abandonment. In Appendix Figure B.3, we show how these outcomes vary for active savers (at the \$1,000 threshold) as a function of income

[29] In particular, we require the zip code, as indicated by Form 1040, to change from one valid zip to another valid zip at some point within five years of the rollover.

and age; i.e., we plot $E(Y|AT, X)$, where X contains age or income.[30] For example, the probability of updating one's address varies from 60% to 80% as income increases. This provides context to the 25 percentage-point difference between $E(Y|C)$ and $E(Y|AT)$, as reported in column 1 of Table 3. In the absence of abandonment, this difference could be explained only if the difference between $E(Y|C, A = 0)$ and $E(Y|AT)$ were larger than the difference between $E(Y|AT, X)$ for the lowest and highest incomes in this sample. Likewise, the probability of interacting with the account within five years naturally increases substantially near retirement age, from just under 60% for those age 50 to over 80% for those age 65 and above. In the absence of abandonment, the 23 percentage point gap between active and passive savers (as reported in column 1 of Table 3) would be approximately equivalent to the gap between working-age and retirement-age individuals. While we cannot rule out the gap between $E(Y|C, A = 0)$ and $E(Y|AT)$ is indeed this large, we find this context suggestive that abandonment must be playing a role, even if the estimated abandonment shares in Table 3 are overestimates.

6 Policy Implications

The empirical analysis in Section 4 established that roughly 3% of all accounts among those who are legally required to draw down their assets are abandoned. Furthermore, we showed in Section 5, that abandonment appears to be substantially more prevalent in savings accounts created by a default policy that forced account transfers to IRAs. Under the additional strong assumption that the abandonment rates we find in Section 5 translate to abandonment rates of accounts induced by other default policies – e.g., automatic enrollment – then our findings raise the possibility that default policies could reduce welfare, unless myopia is very severe.

These results inform a broad set of possible policies. Tax policy providing preferential treatment for IRAs may require modification as many do not withdraw their savings and pay taxes at the expected ages. Defaults into retirement saving plans may require safeguards to prevent forgetting these accounts. State policy for connecting unclaimed property with their owners could feature greater automation to improve these efforts. We do not take a stand on any one of these policies, but believe that policymakers should pay attention to interventions that could reduce abandoned retirement savings. Here, we continue our discussion by providing a conceptual framework for thinking through the ambiguous welfare

[30] Appendix Figure B.4 plots the analogous series at the $5,000 threshold.

effects of default policies and how our findings on abandonment, both among the older population and among passive savers, can provide further insight.

We propose a two-period model of retirement savings and consumption that features active saving, passive saving (by default), and forgetting. Forgetting can be rational or boundedly rational. We begin by showing that incorporating forgetting alone into a consumption-savings model can have different implications for optimal policy design meant to increase retirement savings.

Each individual lives for two periods. In period 1, the individual earns W and saves S. In period 2, the individual consumes S and a Social Security benefit g. Thus, consumption in each period is given by:

$$c_1 = W - S, \qquad c_2 = S + g. \tag{5}$$

Also, suppose individuals are myopic regarding the retirement period and thus undersave as in Laibson (1997). If the individual undervalues retirement utility by a factor of $0 < \beta \leq 1$ at the time of the savings decision, total decision utility is given by:

$$U(S) = u(c_1) + \beta u(c_2) = u(W - S) + \beta u(S + g). \tag{6}$$

Note that β is a *myopia* parameter and not a discount factor, which we omit for simplicity. The optimal (interior) retirement savings for such an individual is $\hat{S}$, which satisfies:

$$\beta = \frac{u'(W - \hat{S})}{u'(\hat{S} + g)}. \tag{7}$$

Given equation (7), we present $\hat{S}$ as a function of β. Thus, $\hat{S}(\beta) < \hat{S}(1),\ \forall \beta < 1$, i.e., individuals undersave because of myopia.

Default savings in our empirical analysis resulted from automatic rollovers, but conceptually we should consider default savings to be from any type of passive behavior. We use d to represent these savings. The individual remembers to consume active and passive savings with different probabilities at retirement. Specifically, the individual remembers fraction $0 < \gamma_a \leq 1$ of active savings (S) and fraction $0 < \gamma_p \leq 1$ of passive savings (d), thereby allowing for some savings to be forgotten. (A more general model could have γ represent probability of being unclaimed in retirement, with this possibility emerging from forgetting or other factors such as hassle costs.) The individual does not forget g as these checks are generally sent to the account holder without any action necessary. Thus, the lifetime utility

optimization problem with default savings d is given by:

$$U(S, d) = u(W - S - d) + \beta u(\gamma_a S + \gamma_p d + g). \tag{8}$$

Policies to reduce retirement account abandonment have different implications for passive and active savers. As in Section 5, we import this terminology from Chetty et al. (2014), where active savers (estimated to be 15% of the population in that paper) respond to incentives and passive savers (the remaining 85%) default to automatic contributions. In what follows, we derive and discuss insights for those two types of savers. We also discuss a conceptual implication on the estimation of myopia in the presence of forgetting.

6.1 Implications for Passive Savers

Default enrollment in retirement savings is a key policy aimed toward increasing financial security for passive savers. Yet, the non-salience of these savings may lead to greater probability of forgetting about them in retirement. Here, we examine the conditions under which default contributions increase experienced utility. To simplify the model, we assume that all of the active savings are remembered, i.e., $\gamma_a = 1$, and that default savings are not salient (as they are intended). Thus, the individual continues to optimize S without considering d.

The reduction in lifetime utility when default savings can be forgotten ($\gamma_p < 1$) is relevant in considering the trade-offs of "nudge" policies.[31] Tension arises because default savings help savers correct for myopic beliefs about retirement needs but also introduce a risk of forgetting these savings. The agent's experienced utility (as a function of d) is given by:

$$V(d) = u(W - \hat{S}(\beta) - d) + u(\hat{S}(\beta) + \gamma d + g). \tag{9}$$

The derivative of this experienced utility with respect to d, evaluated at $d = 0$, is:

$$V'(0) = -u'(W - \hat{S}(\beta)) + \gamma u'(\hat{S}(\beta) + g). \tag{10}$$

Combining (10) with (7) (which defines $\hat{S}(\beta)$) reveals that $V'(0) > 0$ if and only if $\gamma > \beta$. Intuitively, the first dollar of default contributions increases experienced utility when the myopia correction $(1 - \beta)$ is greater than the probability of forgetting $(1 - \gamma)$. As such,

[31] If $\gamma_p = 1$, i.e., all passive savings are remembered, then the optimal d can fully fix the myopic saving problem by setting $\hat{d} = \hat{S}(1) - \hat{S}(\beta)$.

default savings could prove more beneficial when complemented by efforts in helping account owners remember and locate their savings (at least in retirement).

6.2 Implications for Active Savers

We now turn our attention to active savings. Again, we use (8) and to simplify the model, we consider the case with no default saving, i.e., $d = 0$. The optimal active saving, S^* satisfies:

$$\beta\gamma_a = \frac{u'(W - S^*)}{u'(\gamma_a S^* + g)}. \tag{11}$$

We denoted S^* as a function of β and γ_a. As expected, forgetting decreases optimal saving, i.e., $\hat{S}(\beta) > S^*(\beta, \gamma_a),\ \forall\gamma_a < 1$. This is an important consideration for active savers, as they may fail to take advantage of saving vehicles that arise throughout their working lives out of fear of forgetting them in retirement. (Another empirical signature of such fears could be cash-outs of valuable savings at job transitions.) The expected response from active savers suggests that policies helping individuals to consolidate and keep track of their retirement accounts could increase total savings.

Our model of retirement savings with forgetting raises a new question about the estimation of myopia among active savers. Myopia is an important parameter in setting the optimal default rate, as we have shown. It is also a key parameter in lifeycle models of retirement savings, for example in Laibson (1997) and Ericson (2017). If individuals reduce saving because of fear of forgetting as opposed to myopia, however, policymakers will overestimate the degree of myopia among savers and mistakenly increase default savings. The problem arises because the amount of saving in our model with forgetting is equal to one in a parallel economy with with a higher discounting rate, so a traditional analysis will conflate these two worlds. The implication is important because addressing myopia versus forgetting require different policies: while myopia can be aided by default enrollment, for example, forgetting requires reminders or other nudges (such as auto-rollover of funds) to keep funds with their account owners. Under our model and a simple calibration, we determine that the discount factor is overestimated by one percent for the level of abandonment we estimate in the overall sample ($\beta = 0.71$ if $\gamma_a = 1$, versus $\beta = 0.72$ if $\gamma_a = 0.97$).[32]

[32]These estimates assume a log utility function and $W = 1$, $g = 0.2$, $S = 0.3$.

7 Conclusion

The retirement savings landscape in the U.S. places a great deal of responsibility on individuals: how much to save, which saving vehicles to use, how to allocate investment funds, and more. Additionally, individuals must keep track of numerous savings accounts accumulated over their working lives. Failing to do so can result in these accounts being abandoned, either due to forgetting or hassle costs that likely increase with the age of these accounts. While there is considerable literature on various aspects of retirement saving, there is a marked gap in the study of abandoned accounts.

The present paper fills this gap by providing the first set of stylized facts regarding the extent of unclaimed retirement assets. We do this by analyzing individual level tax and information returns, along with account-level data from state unclaimed property. The descriptive analysis contains estimates and correlates of unclaimed accounts, including the extent to which they are "reclaimed" after missing their legal withdrawal minimums or being escheated to state unclaimed property. We also do a deep-dive into forced transfer IRAs, a type of automated rollover account, where we exploit sharp policy thresholds that enable a regression discontinuity analysis. Our contribution also includes an analysis of a theoretical model of retirement saving featuring empirically plausible values of forgetting among passive and active savers.

We find that nearly 2.7% of retirees own an abandoned retirement account, and this percentage is increasing over time. The amounts abandoned in IRAs are substantial at about $800 million per year, with over 40% remaining abandoned even after a decade. We also find that policies which promote retirement saving, such as auto-enrollment, may unintentionally encourage the accumulation of smaller balance accounts that appear to be about 10 times more likely to be abandoned over the lifecycle. Current policy to mitigate abandonment appears limited to the use of escheatment to unclaimed property, yet plan participation is mostly voluntary and most accounts are neither escheated nor reclaimed upon escheatment.

Taken together, our results demonstrate that retirement account abandonment is a timely topic needing policy attention. Forgetting and hassle costs are particularly important as retirees live longer and become vulnerable to cognitive decline (Chandra et al., 2020). We demonstrate the conceptual importance of our results for more general studies of lifecycle saving and consumption, as abandonment can result in suboptimal retirement consumption or decreased pre-retirement saving. We hope that future research on saving will incorporate frictions in retirement account consolidation and memory so that policymakers can consider the tradeoffs that these frictions create.

References

Armour, P., M. D. Hurd, and S. Rohwedder (2016). Trends in pension cash-out at job change and the effects on long-term outcomes. In *Insights in the Economics of Aging*, pp. 15–39. University of Chicago Press.

Baicker, K., W. J. Congdon, and S. Mullainathan (2012). Health insurance coverage and take-up: Lessons from behavioral economics. *The Milbank Quarterly 90*(1), 107–134.

Banks, J., R. Blundell, and S. Tanner (1998). Is there a retirement-savings puzzle? *American Economic Review*, 769–788.

Benartzi, S. and R. Thaler (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives 21*(3), 81–104.

Benzarti, Y. (2020). How taxing is tax filing? using revealed preferences to estimate compliance costs. *American Economic Journal: Economic Policy 12*(4), 38–57.

Bernheim, B. D. and J. M. Gastell (2020). Optimal default options: The case for opt-out minimization. Technical report, National Bureau of Economic Research.

Bubb, R. and P. L. Warren (2020). An equilibrium theory of retirement plan design. *American Economic Journal: Economic Policy 12*(2), 22–45.

Carroll, G. D., J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick (2009). Optimal defaults and active decisions. *Quarterly Journal of Economics 124*(4), 1639–1674.

Chandra, A., C. Coile, and C. Mommaerts (2020). What can economics say about Alzheimer's disease? *NBER Working Paper No. w27760*.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from denmark. *Quarterly Journal of Economics 129*(3), 1141–1219.

Clark, R. L., M. B. d'Ambrosio, A. A. McDermed, and K. Sawant (2006). Retirement plans and saving decisions: The role of information and education. *Journal of Pension Economics and Finance 5*(1), 45–67.

Clark, R. L., M. S. Morrill, and D. Vanderweide (2014). Defined benefit pension plan distribution decisions by public sector employees. *Journal of Public Economics 116*, 73–88.

Ebenstein, A. and K. Stange (2010). Does inconvenience explain low take-up? evidence from unemployment insurance. *Journal of Policy Analysis and Management 29*(1), 111–136.

Ekerdt, D. J. and J. K. Hackney (2002). Workers' ignorance of retirement benefits. *The Gerontologist 42*(4), 543–551.

Ericson, K. M. (2011). Forgetting we forget: Overconfidence and memory. *Journal of the European Economic Association 9*(1), 43–60.

Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association 15*(3), 692–719.

Goda, G. S., M. R. Levy, C. F. Manchester, A. Sojourner, and J. Tasoff (2020). Who is a passive saver under opt-in and auto-enrollment? *Journal of Economic Behavior & Organization 173*, 301–321.

Goodman, L., K. Mackie, J. Mortenson, and H. Schramm (2019). The evolution of leakage and retirement asset flows in the us. *Available at SSRN 3450866*.

Gottlieb, D. and K. Smetters (2014). Lapse-based insurance. *Working paper*.

Jones, D. (2012). Inertia and overwithholding: explaining the prevalence of income tax refunds. *American Economic Journal: Economic Policy 4*(1), 158–85.

Kim, H. H., R. Maurer, and O. S. Mitchell (2016). Time is money: Rational life cycle inertia and the delegation of investment management. *Journal of Financial Economics 121*(2), 427–447.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics 112*(2), 443–478.

Lusardi, A. and O. S. Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature 52*(1), 5–44.

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

Mastrobuoni, G. (2011). The role of information for retirement behavior: Evidence based on the stepwise introduction of the social security statement. *Journal of Public Economics 95*(7-8), 913–925.

Munnell, A. and A. Webb (2015). The impact of leakages from 401(k)s and iras. *Center for Retirement Research at Boston College* (15-2).

Office, G. A. (2019). Federal action needed to clarify tax treatment of unclaimed 401(k) plan savings transferred to states. Technical report.

Poterba, J. M. (2014). Retirement security in an aging population. *American Economic Review 104*(5), 1–30.

Rogers, T. and K. L. Milkman (2016). Reminders through association. *Psychological Science 27*(7), 973–986.

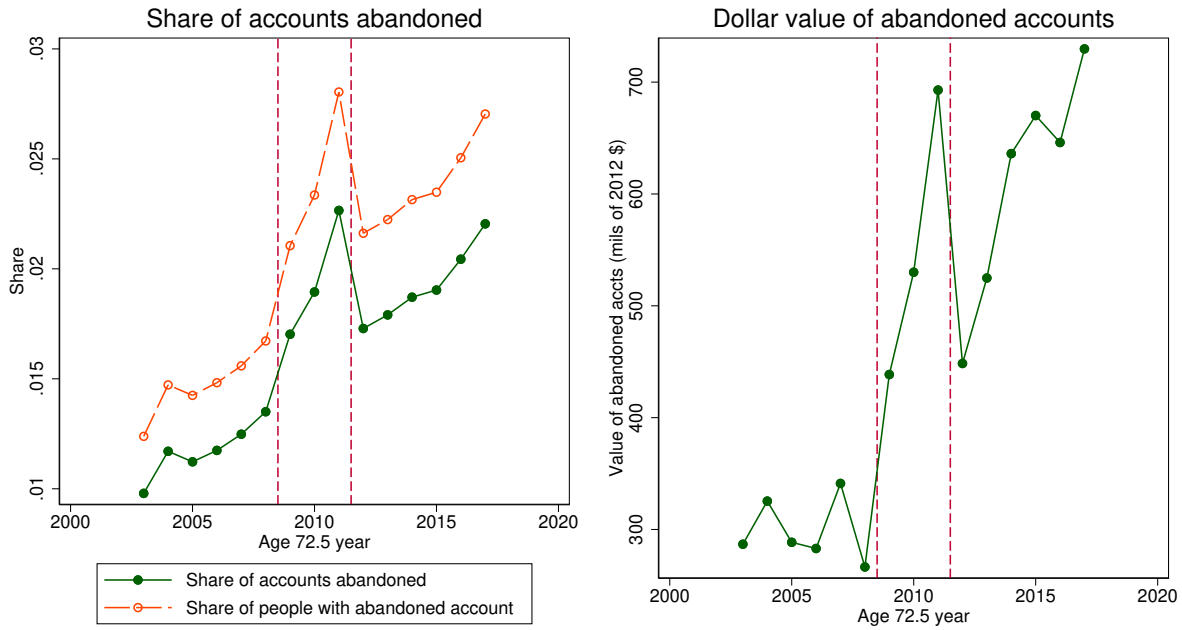
Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006). Are americans saving "optimally" for retirement? *Journal of Political Economy 114*(4), 607–643.

Scott, J. S., J. B. Shoven, S. N. Slavov, and J. G. Watson (2020). Can low retirement savings be rationalized? *NBER Working Paper No. w26784*.

Skinner, J. (2007). Are you sure you're saving enough for retirement? *Journal of Economic Perspectives 21*(3), 59–80.

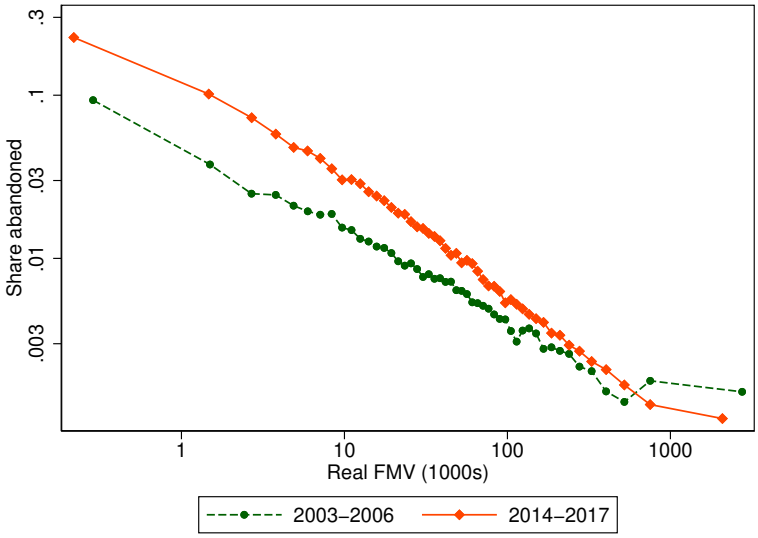
Thaler, R. H. and S. Benartzi (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy 112*(S1), S164–S187.

Figure 1: Time series of abandoned IRAs at old age over time, in tax data



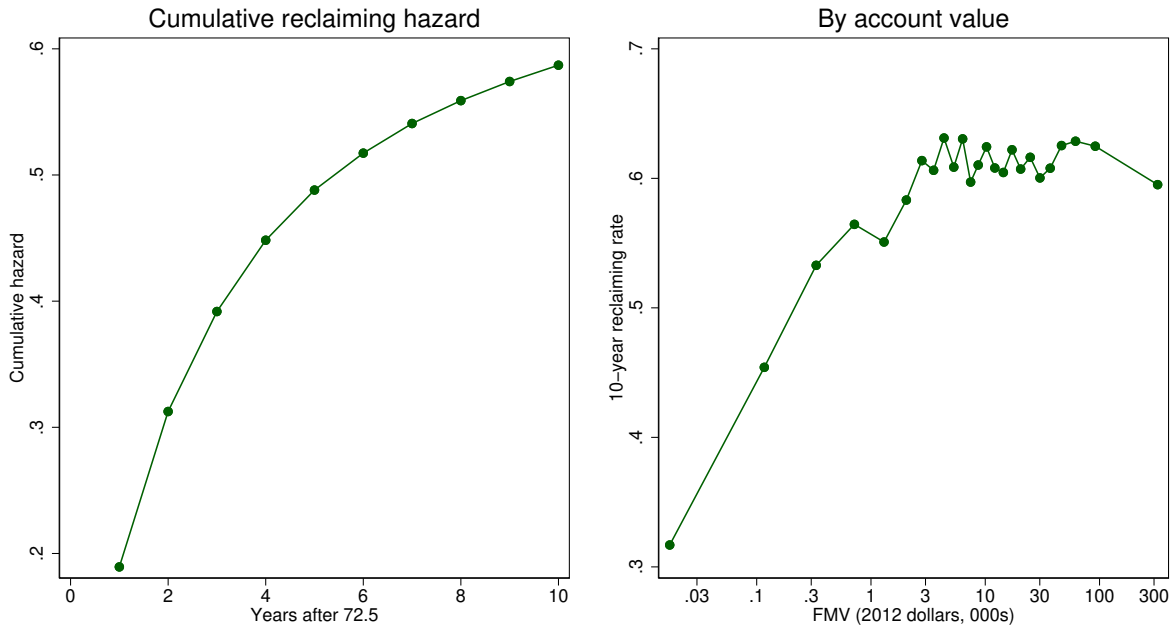
Notes: Left panel: This solid series plots the share of traditional IRAs of 72.5-year-olds that we estimate to be abandoned in each year, where an account is defined as the combination of individual and custodian. The dashed series plots the share of IRA-owners (age 72.5) with at least one abandoned IRA. Right panel: This series plots the aggregate dollar value (measured at the end of the age 72.5 year) of accounts that we estimate to be abandoned at age 72.5. Dollars are adjusted to 2012 dollars via the PCE deflator. These estimates are computed by the authors via analysis of Forms 1099-R, 5498, and other tax return data. See text for a detailed discussion of how we determine an account to be abandoned. Dotted lines bracket the years 2009 through 2011; for each of these three years, the measurement of abandonment is artificially affected by the 2009 suspension of required minimum distributions.

Figure 2: Probability of abandonment conditional on IRA size



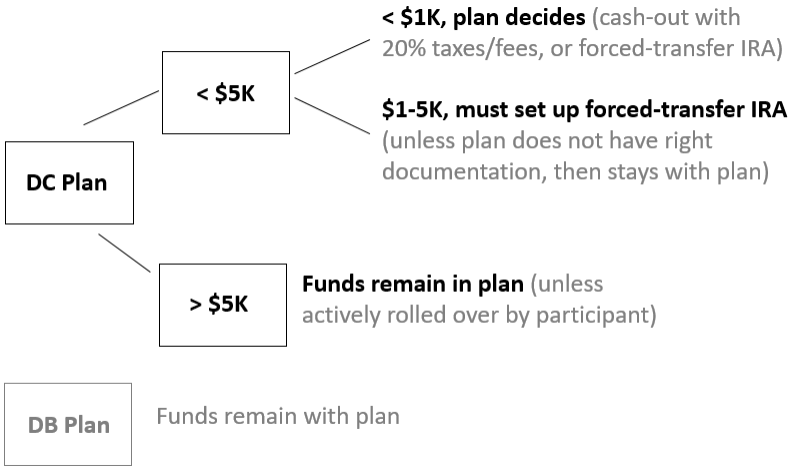
Notes: This figure plots the non-parametrically estimated probability of abandonment at age 72.5 as a function of the account value at age 72.5, measured in 2012 dollars. These series are plotted separately for the period (i.e., cohorts) 2003-2006 and 2014-2017. The estimate is computed by the authors via analysis of Forms 1099-R, 5498, and other tax return data. See text for a detailed discussion of how we determine an account to be abandoned.

Figure 3: Reclaiming of abandoned accounts in tax data



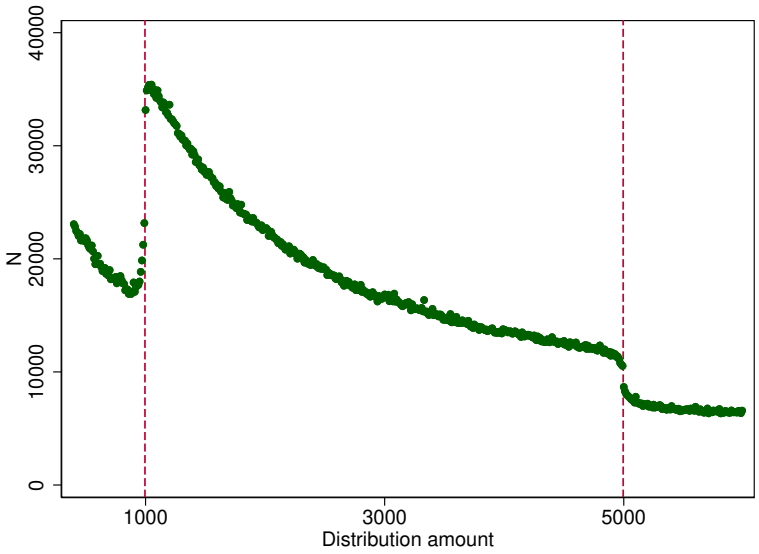
Notes: Left panel: This figure plots the cumulative probability of reclaiming an account in subsequent years, conditional on being estimated to be abandoned at age 72.5 and remaining alive through the year in question. Right panel: This figure plots the probability of reclaiming an account in ten years as a function of the value of the account at age 72.5. Reclaiming is deemed to occur when the individual takes any IRA distribution from the account, as indicated on Form 1099-R. This cumulative probability is computed using year-by-year hazard rates for each year t, which are estimated on the subset of the sample that remains alive through the end of t. This figure includes those who were age 72.5 between 2003 and 2008. See text for a detailed discussion of how we determine an account to be abandoned.

Figure 4: Default Path of Retirement Funds at Job Separation



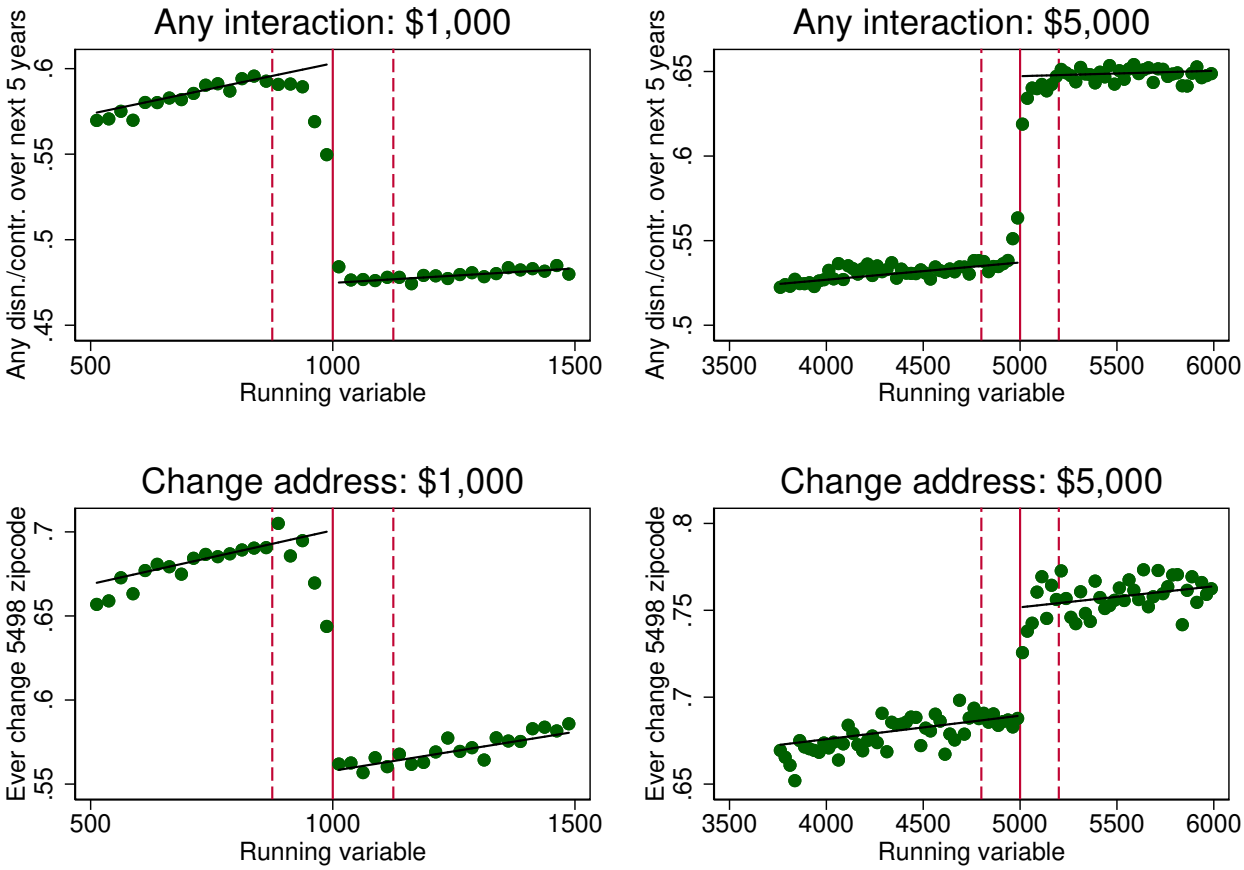
Notes: Figure shows the path of retirement funds at job separation for non-terminating plans. DC indicates defined contribution; DB indicates defined benefit.

Figure 5: Observation counts in $10 distribution bins



Notes: This figure plots raw counts of observations in the forced transfer sample with direct rollovers (Form 1099-R) as a function of the nominal distribution amount, in $10 bins. Distributions that are exact multiples of $500 are dropped. This figure uses distribution data from 2005 through 2015. Dotted lines indicate the policy thresholds at $1,000 and $5,000.

Figure 6: Ingredients for estimating abandonment of forced transfer compliers



Notes: Each panel plots the expected mean of some outcome as a function of the nominal distribution amount. Each panel also plots the linear fit on each side of the threshold, weighted using a triangular kernel, dropping observations within the donut hole indicated by the dashed vertical red lines. The left two panels use the $1,000 threshold while the right two panels use the $5,000 threshold. In the top two panels, the outcome is a dummy for having any interaction with the account – distributions or contributions – within five years of the initial distribution. In the bottom two panels, the outcome is a dummy for updating the address with the IRA custodian within five years, as measured using the address reported by the custodian on Form 5498; in these panels, we condition on the individual moving during this period, measured using the addresses reported on Form 1040 and/or Form 1099-SSA.

Figure 7: Individuals with abandoned IRAs in escheatment and tax databases

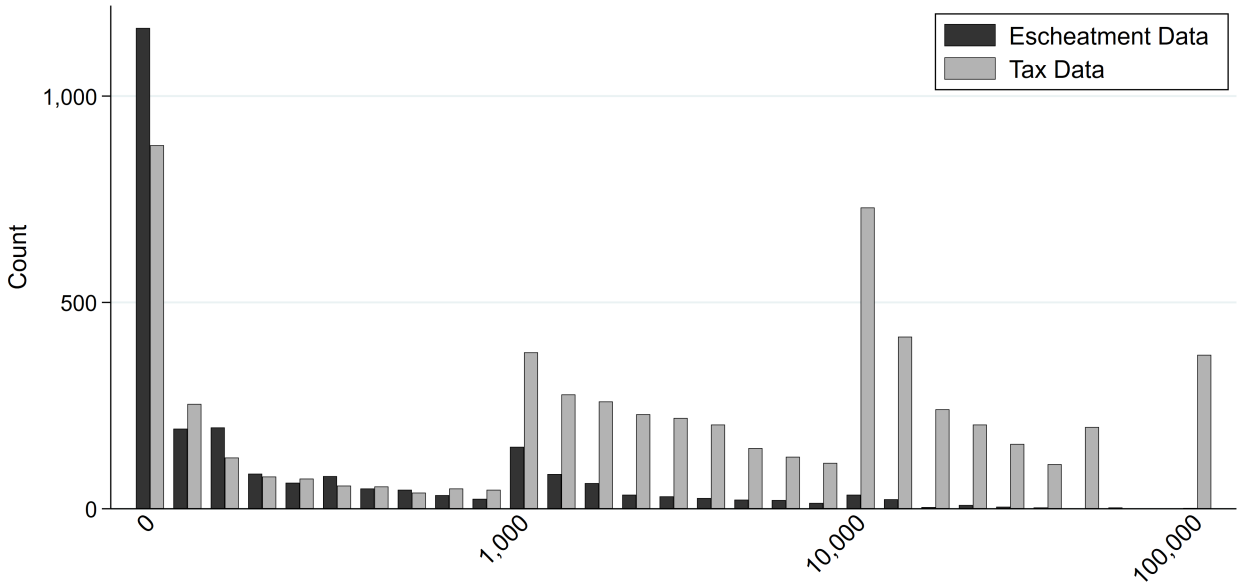
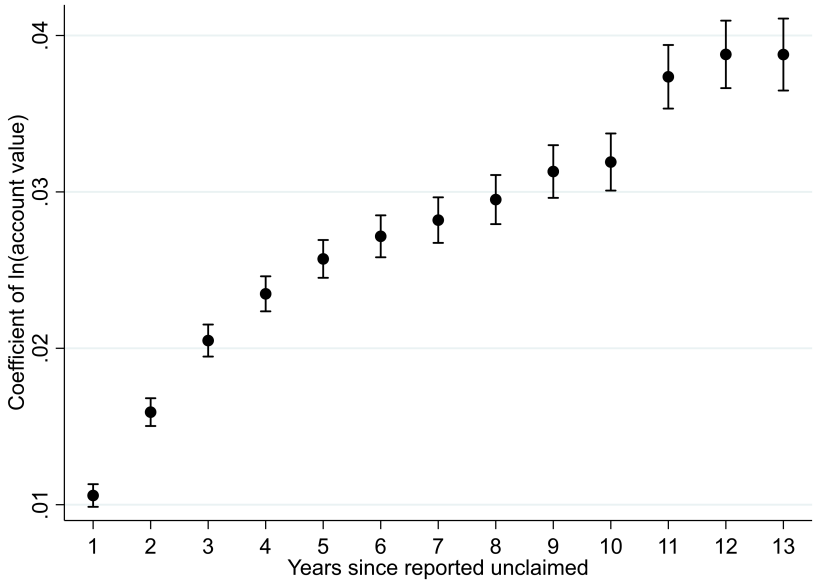


Figure shows the number of people with abandoned IRA accounts in the escheatment and tax databases in 2016. Note that the horizontal axis is not linear. The chart includes the six states (L, LA, MA, MN, OH, and TX) that report a sufficient number of names and addresses of the unclaimed property holders to enable match to the tax data while preserving account holder confidentiality.

Figure 8: Coefficient plot of the impact of account value on claiming



Notes: Figure shows the coefficients, each from a separate regression, on ln(Account Value) as the window of years since reported unclaimed increases from 1 to 13. The coefficients are from regressions of the following form: Claimed = $\beta \ln(\text{Account Value}) + \eta + \gamma$, where *Claimed* is whether the property was claimed within 1, 2,...,13 years (each separate regressions), *Account Value* is measured in units of \$10,000, η represents property type fixed effects, and γ represents year reported as unclaimed fixed effects. The full regression results corresponding to these coefficients are shown in Table B.5. *Data source*: Massachusetts unclaimed and claimed property data, 1998-2018.

Table 1: Abandoned account value as a share of income

	Abandoned accounts			Non-abandoned accounts		
	Value (2012 dollars)	Total share of income	Annuity share of income	Value (2012 dollars)	Total share of income	Annuity share of income
Panel A: 2003-2006						
25th percentile	$ 2,609	0.057	0.005	$ 18,125	0.447	0.037
50th percentile	$ 9,015	0.241	0.020	$ 45,148	1.245	0.104
75th percentile	$ 27,131	0.774	0.065	$111,772	3.241	0.272
90th percentile	$ 61,803	1.988	0.167	$276,117	7.184	0.603
Panel B: 2014-2017						
25th percentile	$ 677	0.012	0.001	$ 34,300	0.598	0.050
50th percentile	$ 6,277	0.127	0.011	$ 97,065	1.797	0.151
75th percentile	$ 25,640	0.571	0.048	$250,216	4.604	0.386
90th percentile	$ 73,182	1.727	0.145	$562,565	9.856	0.827

Notes: Column 1 of this table reports quantiles of the values of abandoned IRAs. Column 2 reports quantiles abandoned IRAs to income. Income is defined as adjusted gross income, plus non-taxable Social Security, minus taxable IRA distributions. Columns 3 reports quantiles of the ratio of the value of the abandoned IRA – converted to an annuity stream – to income. The annuity calculation assumes an interest rate of 1% and uses estimated mortality rates from the Social Security Administration, separately by gender. Columns 4 through 6 repeat the same calculation for accounts that are not abandoned. Panel A uses the early part of our sample (2003-2006), while Panel B uses the late part of our sample (2014-2017). To protect taxpayer privacy, all quantiles reported as psuedo-quantiles, equal to the 30 observations nearest the true quantile.

Table 2: Correlates of abandonment

	(1)	(2)	(3)	(4)	(5)
Financial sophistication metrics:					
Files tax return	-0.0412	-0.0410	-0.0412	-0.0413	-0.0411
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Pays estimated tax	-0.0014	-0.0013	-0.0013	-0.0018	-0.0019
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Has dividends or capital gains	-0.0044	-0.0037	-0.0039	-0.0042	-0.0040
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Has interest	-0.0023	-0.0021	-0.0021	-0.0021	-0.0017
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Demographics:					
Male		0.0014	0.0016	0.0016	0.0011
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
Zip share white		-0.0267	-0.0227	-0.0227	-0.0183
		(0.0003)	(0.0003)	(0.0004)	(0.0004)
Observations	6,760,000	6,757,000	6,757,000	6,757,000	6,747,000
Baseline mean	0.0187	0.0187	0.0187	0.0187	0.0187
Year-by-value FE	X	X	X	X	X
Control for zip density			X	X	X
Control for zip educ. and poverty				X	X
Year-by-payer FE					X

Notes: This table reports regression estimates for a regression of a dummy for abandonment (at the account-level) on various outcomes using data from 2012 through 2017, restricted to those observations with valid zip codes from Form 1040 and/or Form 1099-SSA. Each column corresponds to a different regression. Each regression includes fixed effects for year interacted with 500 bins of real IRA value. All coefficients are statistically significant ($p < 0.001$).

Table 3: Estimated abandonment probabilities in forced transfer sample

	Any interaction over next five years		Update address over next five years	
	(1)	(2)	(3)	(4)
	\$1,000 threshold	\$5,000 threshold	\$1,000 threshold	\$5,000 threshold
Share compliers	0.555	0.377	0.567	0.355
	[0.547, 0.562]	[0.367, 0.386]	[0.548, 0.588]	[0.333, 0.377]
$E(Y\|AT)$	0.603	0.647	0.701	0.752
	[0.597, 0.609]	[0.641, 0.653]	[0.687, 0.716]	[0.738, 0.765]
$E(Y\|C)$	0.372	0.355	0.449	0.576
	[0.361, 0.382]	[0.338, 0.372]	[0.423, 0.472]	[0.533, 0.611]
Prob. abandoned	0.383	0.452	0.360	0.234
	[0.362, 0.404]	[0.423, 0.479]	[0.316, 0.403]	[0.178, 0.299]
Observations	1,399,000	1,170,000	219,000	161,000

Notes: This table reports estimated parameters from the calculation of abandonment in the forced transfer sample. In columns 1 and 2, we use as the outcome a dummy for having any interaction with the account – distributions or contributions – within five years of the initial distribution. In columns 3 and 4, we use the outcome of updating the address with the IRA custodian within five years, as measured using the address reported by the custodian on Form 5498; in these columns, we condition on the individual moving during this period, measured using the addresses reported on Form 1040 and/or Form 1099-SSA. Columns 1 and 3 exploit the \$1,000 threshold, while columns 2 and 4 exploit the \$5,000 threshold. In the first row, we report the estimated share of compliers immediately to the right of the \$1,000 threshold or left of the \$5,000 threshold in the sample in question. The second row reports the estimated mean outcome for always-takers – those who enroll in an IRA voluntarily. The third row reports the estimated mean outcome for compliers – those who are induced to enroll in an IRA due to the forced transfer policy. The fourth row reports the implied share of compliers that abandon, under the assumptions maintained in the text. See text for further discussion about how these objects are calculated.

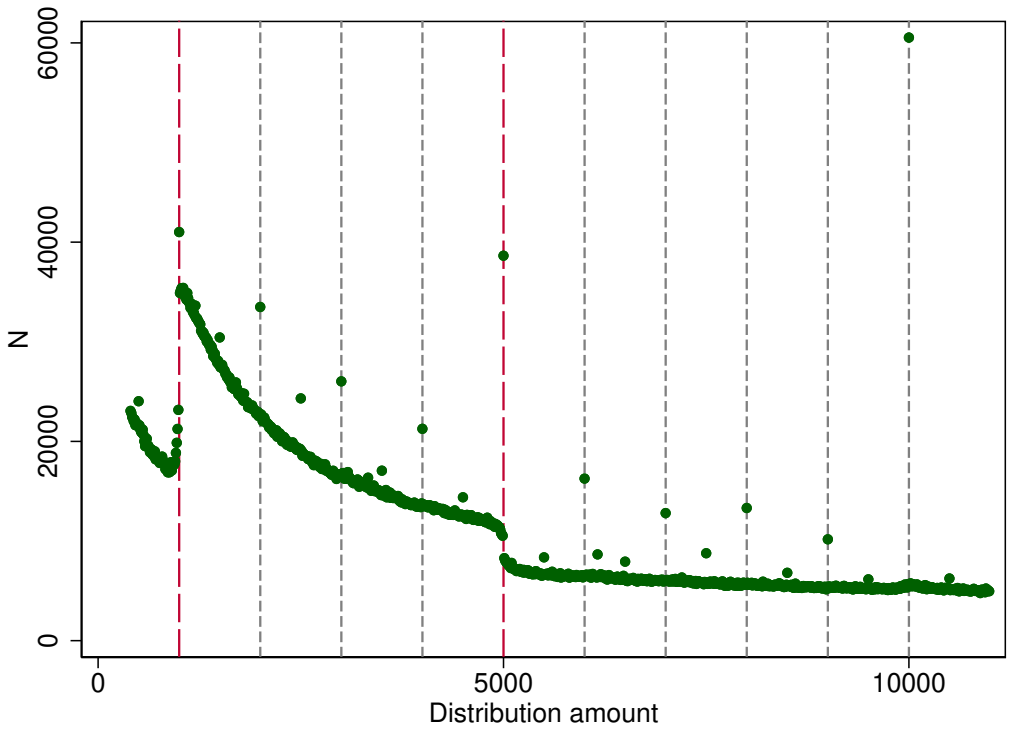
Table 4: Summary statistics on unclaimed and claimed retirement accounts

	MA	WI
Summary		
# unclaimed retirement accounts in 2016	3,320	815
Proportion claimed within 2 years	0.034	0.666
Avg account value	$633	$980
...of claimed accounts	$2,110	$812
...of accounts remaining unclaimed	$581	$1,315
Proportion of claims by initiation type		
DOR Auto Match	—	0.801
Online (own)	—	0.122
Online (locator service)	—	0.065
Other	—	0.013

Notes: Table shows summary statistics on unclaimed and claimed retirement accounts in Massachusetts and Wisconsin. Statistics are for accounts reported as unclaimed in 2016 for ease of comparison. The claim initiation data are not available for Massachusetts. *Data Sources:* Massachusetts Unclaimed and Claimed Property Data, 2016 to 2018; Wisconsin Unclaimed and Claim Property Data, 2016 to 2018.

Appendix A: Round-number bunching in rollover IRAs

Figure A.1: Counts of rollover distributions without dropping round numbers



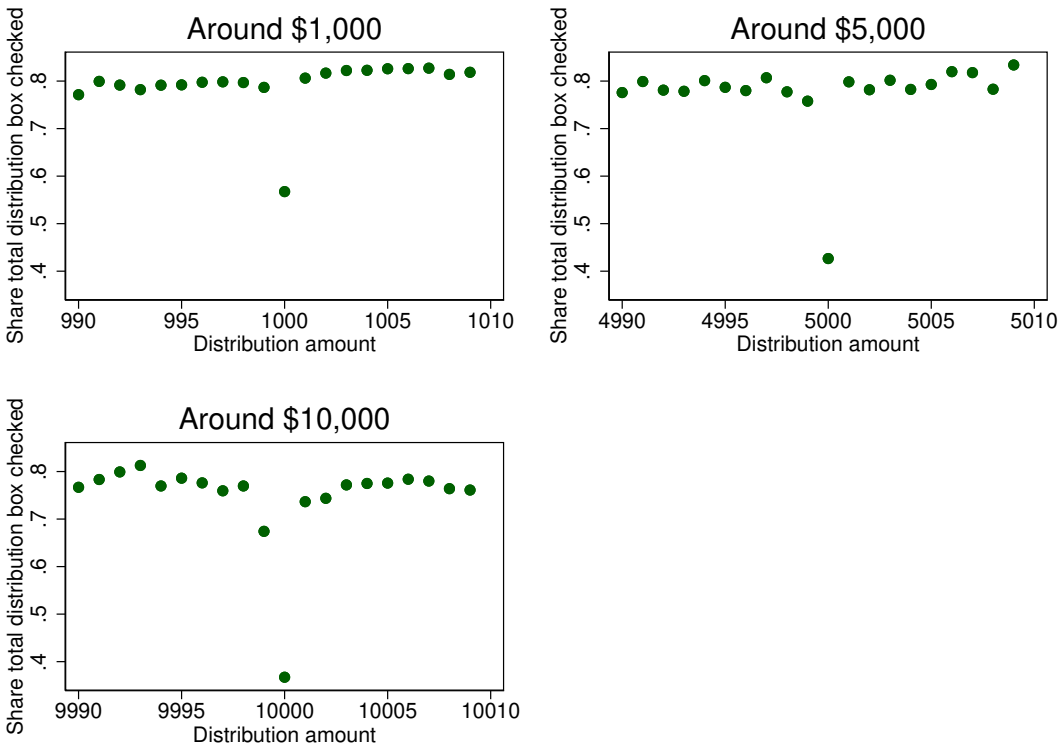
Notes: This figure plots raw counts of observations in the forced transfer sample with direct rollovers (Form 1099-R) as a function of the nominal distribution amount, in $10 bins, without dropping round numbers. Dotted lines indicate even multiples of $1,000.

Figure A.1 plots raw counts 1099-R rollover distributions as a function of distribution amount, in $10 bins, for distributions between $400 and $11,000, between 2005 and 2015. This figure shows the clear round number bunching at multiples of $1,000 and (to a lesser extent) multiples of $500. The bunching at $5,000 is very large; however, the similarly large bunching at $10,000 is reassuring that the magnitude of the $5,000 spike is related to the "roundness" of $5,000 rather than the policy threshold.

The spike of distributions exactly at $1,000 and $5,000 (among other places) would be highly problematic in the empirical approaches that we pursue, especially because round number bunchers are likely to be quite different than those with distribution amounts nearby. In particular, round number bunchers are plausibly much more likely to be rolling over only a portion of the account balance, rather than the entire balance. Such distributions could very plausibly bunch at round numbers. And, furthermore, such distributions will generally

be unaffected by our policy variation, while differing from those taking full distributions in important ways.

Figure A.2: Share of rollover distributions with total distribution box checked

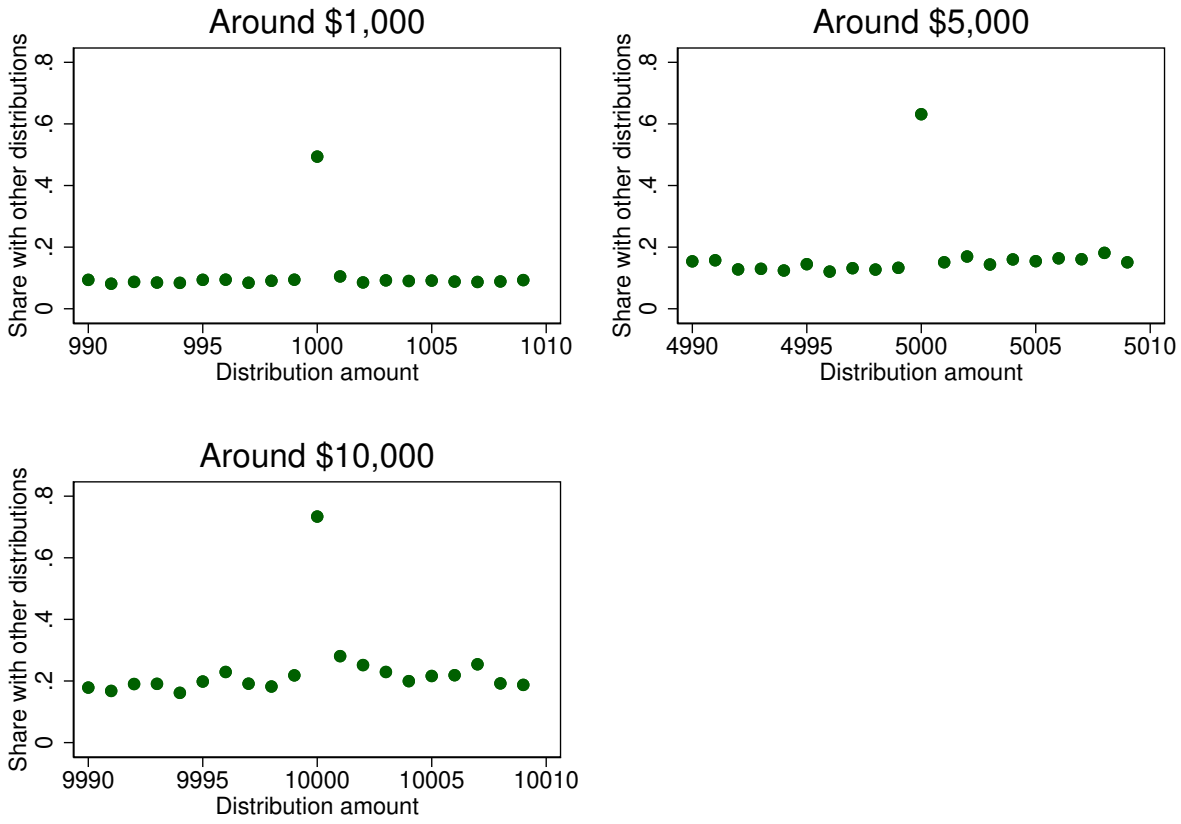


Notes: This figure plots the share of distributions in the forced transfer sample with the "total distribution" check-box checked, by single dollar of distributions. The three panels zoom in on thresholds of $1,000, $5,000, and $10,000 respectively.

To explore this, we plot in Figure A.2 the share of distributions in our data that have the "total distributions" checkbox checked on Form 1099-R, as a function of the distribution amount, by exact dollar of distributions around $5,000 and $10,000. Indeed, the share of distributions with the total distribution box checked declines substantially at round numbers, including round numbers (such as $10,000) that are outside of our policy variation.

One strategy to proceed would be to immediately restrict attention to 1099-R's with the total distribution checkbox checked, since the policy variation affects only total distributions. Unfortunately, there appears to be too much measurement error in the checkbox variable: some portion of distributions with the checkbox checked would in fact not be a total distribution. One common example of mismeasurement is the case when an individual makes two distributions to close out an account; e.g., suppose an individual at separation has a balance of $7,000, chooses to roll over $5,000 and take a cash distribution of $2,000. This

Figure A.3: Share of individuals with same-year regular distribution from same payer, conditional on total distribution box checked

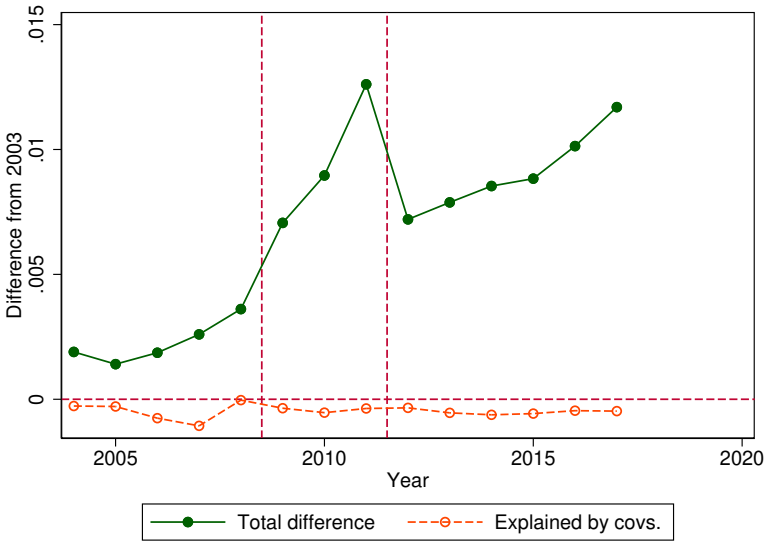


Notes: This figure analyzes the forced transfer sample, restricted to those with the total distributions checkbox checked. The figure plots the share of such observations that have some other distribution (with a Box 7 code of 1, 2, or 7, indicating a non-rollover distribution) from the same payer in the same year, by single dollar of distributions. The three panels zoom in on thresholds of $1,000, $5,000, and $10,000 respectively.

would generate two different Forms 1099-R, which sometimes would both have the total distribution checkbox checked. Figure A.3 explores this further. Among distributions with the total distribution box checked, we determine whether the individual received a different Form 1099-R from the same payer with a distribution code indicating a "normal" distribution (with codes 1, 2, or 7). We indeed find a large spike at both the $5,000 and $10,000 threshold. This suggests that this type of mismeasurement would cause the problem of round number bunching to remain if we used this restriction. For this reason, we take a simpler approach: we drop those with round number distributions (that is, at even multiples of $500) from our data. This conservative approach allows us to focus primarily on those rolling over the full account balance.

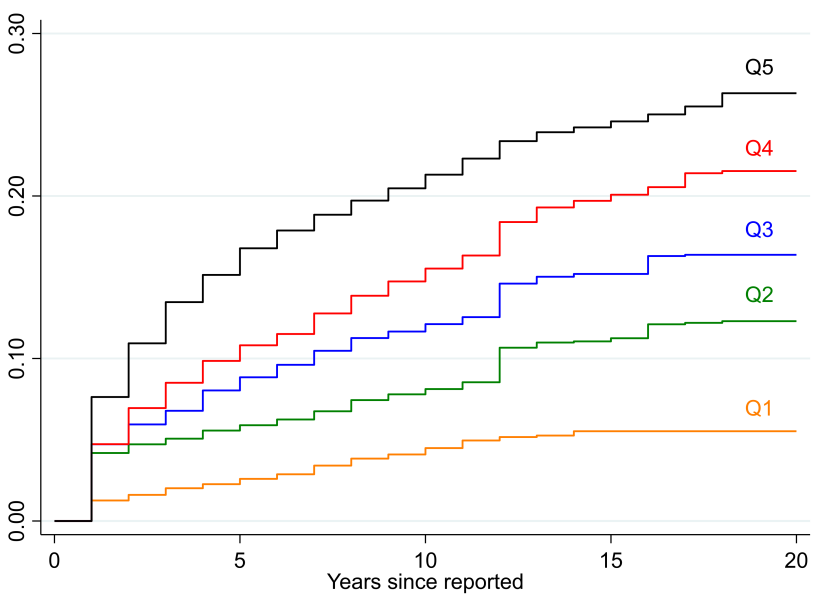
Appendix B: Additional Figures and Tables

Figure B.1: Oaxaca-Blinder decomposition of increase



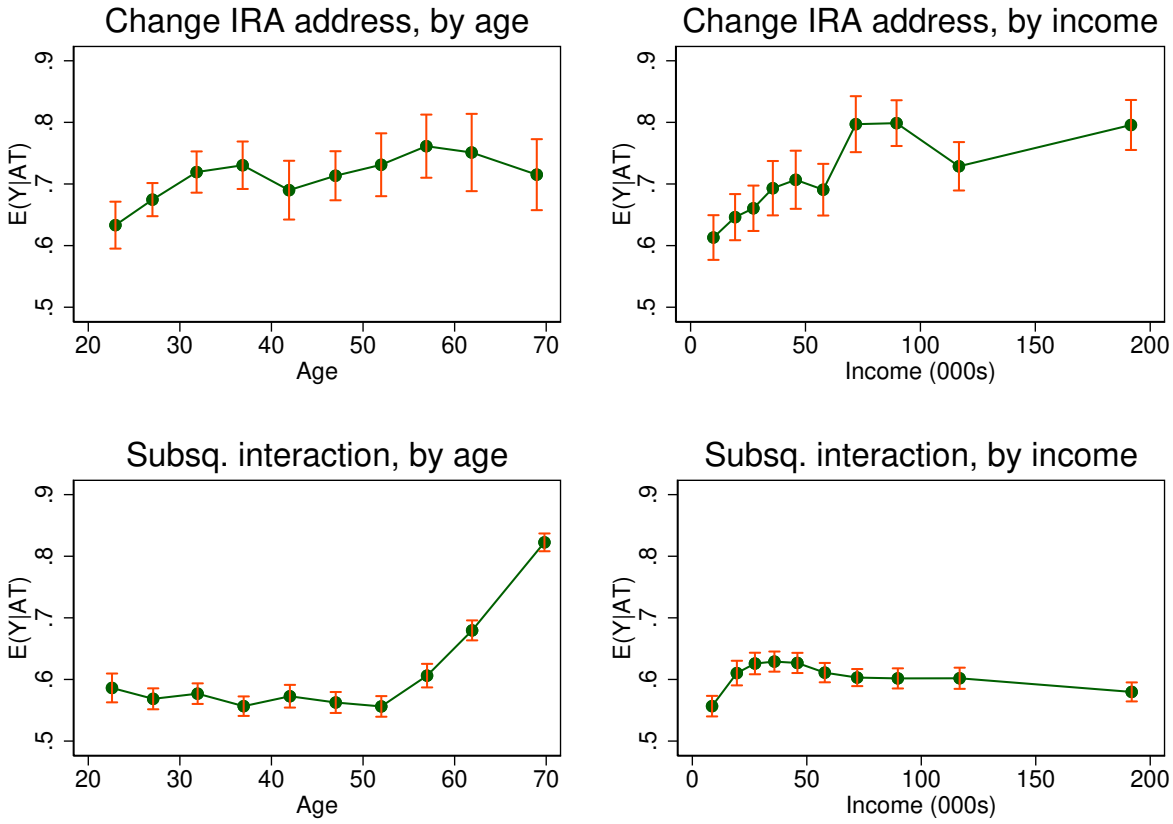
Notes: The solid series plots the share of IRAs abandoned in year t minus the share abandoned in 2003. The dashed series plots the amount of this difference that can be explained by covariates, $(X_t - X_{2003})'\beta_{2003}$. X includes a fully-interacted set of fixed effects for (1) real value of the IRA (in 50 bins), (2) 10 bins of Social Security income relative to the national distribution, (3) 10 bins of the white share of the zip code, (4) a dummy for being male, (5) a dummy for having any interest income, and (6) a dummy for having non-zero capital gains or dividend income. The coefficient β_{2003} is the coefficient from a regression of abandonment on X using observations in 2003.

Figure B.2: Kaplan-Meier claiming estimates by account value



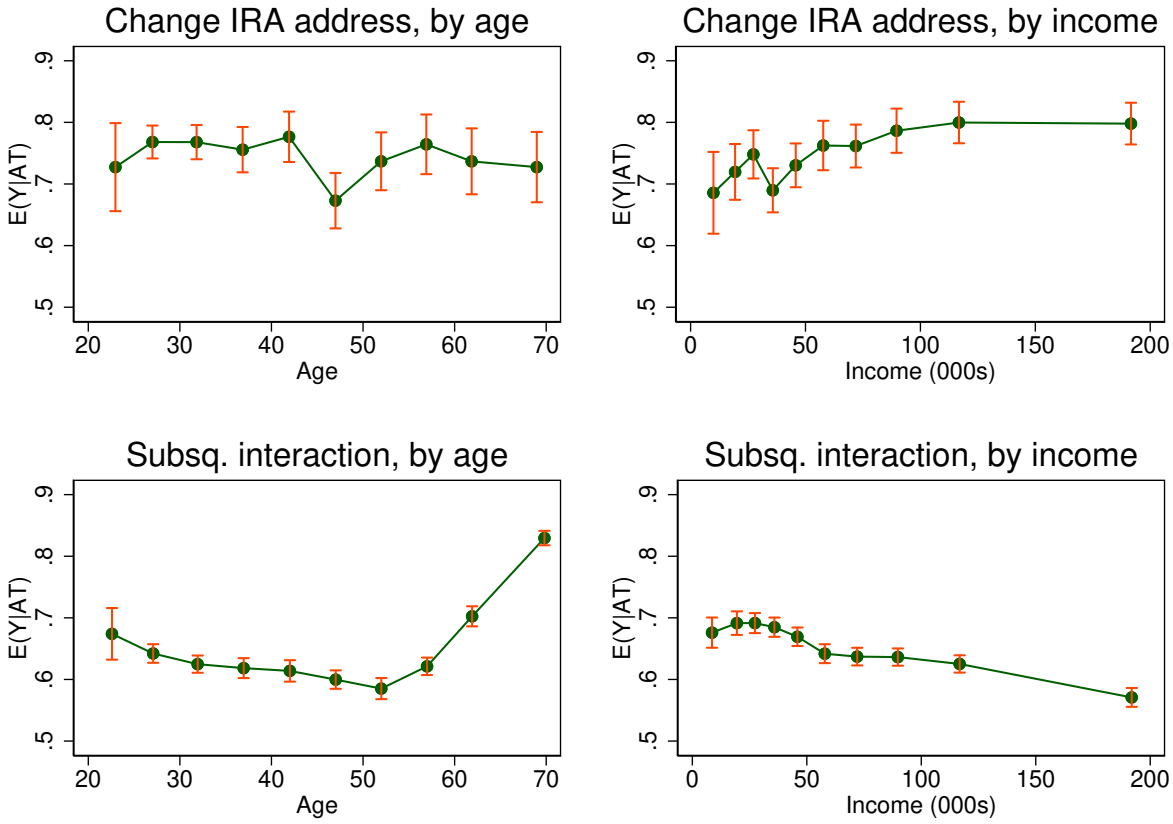
Notes: Figure shows Kaplan-Meier estimates (where the failure is a claim) for whether an account was claimed after being in the state unclaimed property database by quintiles of account value (Q5 is the highest value quintile). *Data Source:* Massachusetts Unclaimed and Claimed Property Data, 1998 to 2018.

Figure B.3: Variation in always-taker means as a function of age and income: $1,000 threshold



Notes: This figure plots the mean outcomes for always-takers at the $1,000 threshold as a function of age or income. These objects are computed as the constant of separate regressions of the outcome on distribution amount, using observations only to the left of $1,000 within a certain bin of age or income, weighted by a triangular kernel. The outcome in the top panel is a dummy for changing the IRA address within 5 years; this analysis is restricted to those who move, based on address reported on Form 1040 and/or Form 1099-SSA. The outcome in the bottom two panels is a dummy for having any further interaction (distribution or contribution) with the account within 5 years of the initial distribution.

Figure B.4: Variation in always-taker means as a function of age and income: $5,000 threshold



Notes: This figure plots the mean outcomes for always-takers at the $5,000 threshold as a function of age or income. These objects are computed as the constant of separate regressions of the outcome on distribution amount, using observations only to the right of $5,000 within a certain bin of age or income, weighted by a triangular kernel. The outcome in the top panel is a dummy for changing the IRA address within 5 years; this analysis is restricted to those who move, based on address reported on Form 1040 and/or Form 1099-SSA. The outcome in the bottom two panels is a dummy for having any further interaction (distribution or contribution) with the account within 5 years of the initial distribution.

Table B.1: Unclaimed retirement account summary and extrapolation, 2016

Sample	# accts	Total funds ($)	Mean amt ($)	Pop. age 74	Accts per 74	% US 74 pop
Data	36,529	18,347,524	633	898,227	.041	.43
National (extrap.)	69,507	38,012,684	547	2,094,035	.033	1.00

Notes: Table reports aggregate retirement-related unclaimed properties by our sample of 13 states and the extrapolated national sample. Data sources include state unclaimed property data and the U.S. Census Bureau Estimated State Population by Characteristics for 2016.

Table B.2: Characteristics of unclaimed retirement account owners

	Unclaimed retirement account owners	Overall 74 year old population
Name analysis		
Percent female	0.45	0.54
Percent Hispanic	0.24	0.12
County-level analysis		
Average county population	2,116,785	158,137
Percent age 65+	0.13	0.14
Percent white	0.69	0.72
Percent married	0.48	0.48
Percent bachelor degree	0.31	0.30
Percent born in state	0.57	0.57
Percent moved across state in past year	0.02	0.02
Percent own home	0.63	0.63
Median household income ($)	60,167	58,881
Percent of families below poverty line	0.11	0.11
Unemployment rate	8.60	7.60

Notes: Table reports average characteristics of individuals with unclaimed retirement assets (column 1) and average characteristics of 74 year olds (column 2). Data come from the sample of 13 states in the state unclaimed property data, Census population data, names databases derived from 2000 Census data and Social Security Administration data, and ACS county-level data.

Table B.3: Impact of account value on claiming

	(1)	(2)	(3)
Account Value ($0000)	-0.0810*** (0.0074)	-0.0830*** (0.0076)	-0.0893*** (0.0076)
Year FEs	N	Y	Y
Property Type FEs	N	N	Y
Observations	51,858	51,858	51,858

Notes: Table shows coefficients from accelerated failure time (AFT) models for the number of years until an account was claimed (if observed) after being in the state unclaimed property database by account value. *Data source*: Massachusetts unclaimed and claimed property data, 1998-2018.

Table B.4: NAUPA codes categorized as retirement accounts

Code	Description
CK11	PENSION CHECKS
IR01	TRADITIONAL IRA - CASH
IR02	TRAD IRA - MUTUAL FUNDS
IR03	TRAD IRA - SECURITIES
IR04	RESERVED FOR TRADITIONAL IRA
IR05	ROTH IRA - CASH
IR06	ROTH IRA - MUTUAL FUNDS
IR07	ROTH IRA - SECURITIES
IR08	RESERVED FOR ROTH IRA
IR09	IRA OTHER - RESERVED 1
IR10	IRA OTHER - RESERVED 2
MS14	PENSION & PROFIT SHARING PLANS
05	IRA's-Securities
55	Annuities
71	IRAs
78	Pensions, retirement funds

Notes: NAUPA (National Association of Unclaimed Property Administrators) Codes used to categorize unclaimed property. Codes starting with "IR" were introduced in 2010 and gradually adopted by states. Codes without any alphabetic characters are exclusive to California.

Table B.5: Impact of account value on claiming an unclaimed retirement account

	Claimed?						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(Account Value)	0.0106***	0.0159***	0.0205***	0.0235***	0.0257***	0.0272***	0.0282***
	(0.0004)	(0.0005)	(0.0005)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
R-squared	0.172	0.154	0.145	0.139	0.136	0.138	0.133
Observations	50,063	46,524	43,204	41,436	39,709	36,025	34,356
Years Unclaimed	1	2	3	4	5	6	7
	(8)	(9)	(10)	(11)	(12)	(13)	
ln(Account Value)	0.0295***	0.0313***	0.0319***	0.0374***	0.0388***	0.0388***	
	(0.0008)	(0.0009)	(0.0009)	(0.0010)	(0.0011)	(0.0012)	
R-squared	0.130	0.132	0.135	0.139	0.129	0.129	
Observations	32,760	30,777	28,063	26,362	25,709	24,376	
Years Unclaimed	8	9	10	11	12	13	

Notes: Table shows coefficients on ln(Account Value) from separate regressions of whether the account was claimed within 1, 2,..., or 13 years (denoted by the Years Unclaimed row), including property code and year reported unclaimed fixed effects. Since the columns subsequently increase the number of years of possible claiming from 1 to 13, the sample size decreases due to the window of relevant data. Stars denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *Data Source:* Massachusetts Unclaimed and Claimed Property Data, 1998 to 2018.