Crowdout in the Decumulation Phase: Evidence from the First Year of Required Minimum Distributions

Lucas Goodman*

May 14, 2019

Abstract

This study estimates the extent to which a policy-induced increase in distributions at retirement crowds out dissaving from taxable assets. This parameter informs the policy effectiveness of the underlying policy and sheds light on how retirees manage the decumulation of their assets. In particular, this study analyzes distributions induced by Required Minimum Distributions (RMDs), primarily from Individual Retirement Accounts (IRAs). The study shows that retirement distributions at age 70 vary discontinuously at a statutory threshold in exact day of birth. Using this discontinuity as a first stage, the study finds that taxable saving becomes more positive (or less negative) among those whose retirement distributions are induced to be higher. In the baseline specification, ratio at which induced retirement contributions crowd out taxable dissaving is estimated to be 0.42.

Keywords: Required minimum distributions, Individual Retirement Arrangements, retirement, decumulation, crowdout, regression discontinuity.

JEL Codes: D14, D15, H24

[1] I am grateful to Adam Isen, Shanthi Ramnath, Lesley Turner, and Melissa Kearney for helpful comments. All errors are my own.

* Office of Tax Analysis, U.S. Department of the Treasury. The views expressed in this paper are those of the author and do not necessarily reflect the views of the U.S. Department of the Treasury or the Office of Tax Analysis.

Introduction

In 2012, over one million 69-year olds owned Individual Retirement Accounts (IRAs), with a total value of approximately $209 billion. As these individuals begin or continue their retirement, they will make decisions regarding how to draw down these IRA assets, taking into account their other assets and annuitized income flows (such as Social Security and defined benefit pensions). This "decumulation" phase of retirement saving, as it is known, has received comparatively less attention from economists, policymakers, and financial advisors than the pre-retirement,"accumulation" phase.

This paper analyzes the effect of an added constraint to this optimization problem. In particular, individuals who turn 70.5 prior to the end of a calendar year are required to take annual distributions of a certain percent of their prior year balance, known as Required Minimum Distributions (RMDs).[1] These rules were put in place in order to raise revenue by limiting the tax benefit associated with IRAs. Recent research (Mortenson, Schramm, and Whitten (2018), Brown, Poterba, and Richardson (2017)) has shown that these rules are binding for a large share of IRA-holders: a considerable share of IRA-holders make a distribution exactly equal to their RMD. In addition, IRA distributions fell substantially in a year when the RMD rules were temporarily suspended. However, little is known regarding whether IRA distributions induced by RMDs causes a change in net worth (e.g., by altering consumption), or whether the distributed amounts are simply reallocated to other accounts. This has first-order implications for the policy effectiveness of RMDs and also sheds light on how retirees make decisions regarding the decumulation of their savings.

In this paper, I use a regression discontinuity (RD) design to estimate the effect of RMDs on "taxable saving" (i.e., in accounts that produce taxable dividend or interest income, or in home equity which reduces potentially-deductible mortgage interest expense), exploiting a discontinuity in exact date of birth. I find a point estimate of 0.42, suggesting that RMDs are only partially reallocated to other accounts. This modest degree of crowdout is consistent with some combination of basic neoclassical models of retirement decumulation and behavioral frictions. Additionally, this evidence suggests that RMDs are somewhat, though not completely, successful at achieving their policy aim of increasing present-value tax revenue without harming retirement security by too much. In the main specification, I can reject both crowdout of zero (which would require strong behavioral biases) and crowdout of one (which would be the approximate prediction of the most bare neoclassical model) at the 99 percent confidence level.

The RD design in this paper exploits the fact that RMD status depends discontinuously on age at the end of the year. To illustrate, consider two individuals. The first turns 70.5 on December 30 in year τ (i.e., with a birthday of June 30, $\tau - 70$); she is required to take an RMD for year τ. The second turns

[1] RMD rules apply generally to IRAs and some defined contribution retirement plans. In general, RMDs for DC plans, but not IRAs, can be deferred for individuals who are still in the workforce.

70.5 on January 1 in year $\tau+1$ (i.e., with a birthday of July 1, $\tau-70$); she is not required to take an RMD for year τ.[2] The RD design generalizes this example to estimate the discontinuity in the conditional mean of some outcome as the running variable (date of birth, relative to June 30, $\tau-70$) crosses the threshold (zero). Importantly, both sets of individuals will be required to take an RMD beginning in $\tau+1$ and in later years, so long as their IRA balance remains positive; thus, the "treatment" in this RD design is the requirement to begin taking RMDs one year sooner.[3] Using this RD design, I confirm the previous literature that RMDs appear to be binding. Being just older than 70.5 at year τ increases IRA distributions by $2,980 (or about 0.57 of the potential RMD amount) and total retirement distributions by $3,450 (or 0.62 of the potential RMD amount).

I then use this result as the first stage to estimate a crowdout parameter: the rate at which induced retirement distributions are offset by an increase in taxable saving. In particular, I construct taxable saving by capitalizing changes in income flows from time t to $t+1$, following a method analogous to Saez and Zucman (2016). In the baseline specification, I find that taxable saving increases by a total of $1,450 in the two years bracketing the treatment year, which is significantly different than zero at the 1 percent level. Using a quantile treatment effects estimator (using the method of Frandsen, Frolich, and Melly (2012)), I find that this positive effect is present throughout the entire distribution of savings. Additionally, I reduce noise by using the inverse hyperbolic sine transformation (Pence (2006)); under this transformation, I calculate the median treatment effect to be $1,300.

Estimating the magnitude of the crowdout parameter is more challenging. This is partially due to the fact that theory requires the regression to be run using savings in levels, which requires Winsorization to provide informative estimates. After aggregating over the two years most plausibly affected by treatment, I estimate a crowdout ratio of approximately 0.42, with a 95 percent confidence interval spanning 0.12 to 0.73 in the baseline specification. To corroborate these estimates using a more precise estimator, I additionally use a threshold-based approach suggested by Chetty, et al (2014). This approach estimates the effect of the discontinuity on the share of individuals with taxable saving greater than some amount y^* and compares that estimate to the effect that one would predict if crowdout were unity. Under some additional assumptions, the ratio of the estimated effect to the predicted effect represents the crowdout rate. Furthermore, this method estimates crowdout parameter locally – i.e., for those with potential saving (i.e., saving when treatment equals zero) near y^*. As y^* varies, we can trace out the heterogeneity in this parameter. Under this approach, I estimate that the crowdout rate is approximately 0.2 to

[2]Under a grace rule, individuals can defer their first RMD as late as April of the following year. This will generally complicate the relationship between exact date of birth and retirement distributions in a given year. Fortunately, a clear majority of individuals appear not to take advantage of the grace rule.

[3]A consequence of this identification strategy is that the estimates will only be relevant for those near age 70.

0.4 for those with relatively small amounts of potential saving relative to the potential RMD, while it tends to be larger for those with larger magnitudes of potential saving, especially for those with large amounts of negative potential saving.

These results can provide evidence on the answers to two broader questions. The first question is a simple policy analysis question: are RMDs successful at raising revenue in a present value sense? Contributions to IRAs are (generally) made on a pre-tax basis, and distributions from IRAs are (generally) subject to income tax as they are withdrawn. Under the assumption that tax rates are flat and constant over time, the present value tax benefit of an IRA is that the investment income is not taxed as it is earned, effectively increasing the net-of-tax rate of return (Crain and Austin (1997)). For RMDs to raise present-value revenue, then it must be the case that RMDs cause an increase in saving (or decrease in dissaving) in taxable accounts, creating taxable interest, dividend, or capital gains income, or reducing mortgage interest deductions.[4] The modest offset implied by the RD point estimates suggests that RMDs are somewhat successful at achieving this aim, though the bottom (0.12) or the top (0.73) of the confidence interval would have different implications. Second, these empirical results can shed light on the broader question of how retirees draw down their assets. To fix ideas, I present a simple neoclassical model, in which retirees choose distributions from IRAs and taxable (dis)saving, subject to RMD constraints. Quite intuitively, the optimal distribution strategy under this model is to exhaust taxable accounts before taking distributions from IRAs, since IRAs offer a higher rate of return. Furthermore, a binding RMD would reduce taxable dissaving on essentially a one-for-one basis. The empirical analysis is only partially consistent with this prediction. This could mean that some individuals are altering consumption as a result of the RMD, perhaps because the RMD is perceived as an increase in wealth. Alternatively, the RMDs could increase saving in assets with less-than-average income yield (such as zero-interest bank accounts), which would cause me to underestimate crowdout.

This paper relates to three strains of the literature. First, there are two recent papers which analyze RMDs. Mortenson, Schramm, and Whitten (2018) analyze tax data to compare the observed distribution of IRA distributions, as a share of the RMD, in 2009 (when RMD rules were suspended) and all other years. They find large spikes in the share of individuals taking a normalized distribution (i.e., distribution divided by previous year's value) within 0.05 percent of their RMD, suggesting that they are bound by the RMD. Additionally, approximately 26 percent of individuals continue to bunch in 2009, perhaps because of optimization frictions or because RMDs are perceived as guidance for an appropriate rate of drawdown. In sum, they estimate that 52 percent of individuals are bound by RMDs. Brown, Poterba, and Richardson (2017) study account-holders at TIAA-CREF, a large IRA custodian. They also find that many individuals suspended their RMDs in 2009; interestingly, however,

[4]RMDs could also raise present value revenue without creating taxable investment income by causing the income to be taxed at a higher marginal rate than it would have been taxed if withdrawn later.

suspension probabilities were not very different between those whose 2008 distributions equaled their RMD and those with larger distributions. They also find that suspension was less likely for individuals that appear more reliant on IRA distributions for living expenses: those who are more elderly or who have fewer outside financial resources. Lastly, consistent with the evidence in Mortenson, Schramm, and Whitten (2018), they find survey evidence that some who chose not to suspend their distribution in 2009 did so because they perceived RMDs as implicit guidance on the appropriate rate of drawdown.[5] I build on these papers by showing the effect of RMDs on a broader picture of individuals' balance sheets. In particular, I use the fact that RMDs are binding for many individuals – a result that I confirm – as a first stage to show that induced retirement distributions are causing modest changes to net worth, while also being partially reallocated to other accounts.

Second, there are a number of papers that analyze the drawdown of retirement assets. Poterba, Venti, and Wise (2011) use the Health and Retirement Survey (HRS) to document the low amount of non-annuitized wealth held by retirees. They find that retirees tend not to consume out of non-annuitized wealth, such as IRAs but also imputed home equity; rather, such assets are held as precautionary savings. Similarly, Poterba, Venti, and Wise (2013) study the Survey of Income and Program Participation (SIPP) and find that IRA distributions as a share of IRA wealth are small for retirees through their 70s, suggesting that IRA values peak around age 80.[6] De Nardi, French, and Jones (2015) argue that the apparently slow drawdown of retirement assets can be explained by some combination of bequest motives and anticipation of large medical expenses at the end of life, though they cannot distinguish between these two factors. This literature generally leaves open the question of how individuals make these distribution decisions. Given that asset decumulation is a very complicated decision to make, with uncertainty regarding investment returns, longevity, and other marginal utility shocks, people might rely strongly on defaults, or perceived guidance such as RMDs. My findings help to inform whether this is true in the case of RMDs. A high amount of crowdout of taxable saving by RMDs would imply that RMDs are not substantially affecting total asset drawdown (at least at age 70), which would suggest that individuals are not treating RMDs as guidance for optimal asset decumulation. The intermediate amount of crowdout found in the present paper is thus somewhat inconclusive in this regard.

Third, there is a long literature estimating crowdout of taxable saving by retirement saving in the acumulation phase. Many early, observation studies reviewed by Bernheim (2002) attempted to compare 401(k)-eligible and -ineligible individuals to estimate the effect of eligibility on total saving, using various econometric methods to make eligible and ineligible individuals more similar.[7]

[5] Alonso-Garcia et al (2018) also find experimental evidence that RMD-like provisions influence individuals' desired drawdown rates.

[6] By contrast, Mortenson, Schramm, and Whitten (2018) find in the tax data that IRA values tend to peak near the beginning of the RMDs (age 70).

[7] Such studies include, but are not limited to, Poterba, Venti, and Wise (1994, 1995, 1996),

More recently, Gelber (2011) compares the increase in savings for workers who become eligible in their second year at a firm relative to the increase for workers who were eligible beginning in their first year; the point estimates suggest no crowdout, but the estimates lack the statistical precision to rule out substantially nonzero crowdout. Chetty, et al (2014) analyzes administrative Danish data and develops a key distinction between "passive savers" whose behavior is heavily influenced by default rules and "active savers" whose behavior is much closer to the neoclassical model. They find that crowdout of taxable savings by retirement savings depends heavily on whether active savers or passive savers are responding. They estimate that a policy change that affected only active savers (a change to the subsidy rate) yielded approximately full crowdout; crowdout for policies that affect passive savers, as measured by changes in taxable saving, tend to be much smaller. Beshears, et al (2017) find mixed evidence for crowdout in the case of automatic enrollment, which increases retirement saving by passive savers. They find that auto-enrollment does not increase unsecuritized debt, such as credit card debt, but that it does increase automobile debt and mortgage debt. I contribute to this literature by studying crowdout in the decumulation phase, for which little is known, rather than the accumulation phase. One might expect different amounts of crowdout in this setting. On the one hand, individuals are already in or near retirement, perhaps making retirement saving/decumulation decisions more salient, increasing the share of active "savers." Furthermore, the choice architecture could lead to crowdout even among passive savers, since IRA custodians have an incentive to encourage clients to deposit their RMD into a taxable account at the same firm.

The remainder of this paper is structured as follows. Section 1 discusses the institutional background of IRAs and RMDs. Section 2 presents a conceptual framework for understanding how RMDs might affect decumulation. Section 3 describes the data set that I use. Section 4 describes the regression discontinuity method in general, including several extensions. Section 5 presents the first stage and reduced form estimates. Section 6 presents the estimates of crowdout. Section 7 examines several potential threats to identification. Section 8 considers sensitivity analyses to the main results. Finally, Section 9 concludes.

1 Policy Background

1.1 The role of IRAs in the U.S. retirement system

The retirement system in the United States is made up of several pillars. Some of these pillars take the form of life-contingent income streams, such as Social Security income and defined benefit pensions. Other pillars, notably defined contribution pension plans, IRAs, and other forms of taxable savings, are generally in the form of a lump sum of assets, whose drawdown must be managed by the retiree (or to an agent delegated that responsibility by the retiree).

Engen, Gale, and Scholz (1994, 1996), and Benjamin (2003).

IRAs represent a modest portion of total retirement income in the population. Goodman, et al (2019) estimate that (traditional) IRAs distributed \$190 billion (in 2009 dollars) to individuals aged 55 and up in 2014. This compares to \$64 billion from defined contribution retirement plans, \$436 billion from defined benefit retirement plans, and \$652 billion in (Old-Age) Social Security payments.[8] Additionally, Goodman, et al (2019) show that the vast majority of IRA assets represent assets rolled over from DC plans. They estimate that regular individual contributions to IRAs were only \$16 billion in 2014, while rollovers from DC plans were \$365 billion. Thus, one should think about IRAs as largely being an extension of the defined contribution retirement system.

The main tax advantage of an IRA relative to a taxable account holding the same assets is that the investment income is not taxed, leading to a higher net-of-tax rate of return. This can be seen as follows. Generally, contributions to IRAs are made on a pre-tax basis (whether the original contribution was to the IRA or to a DC plan that was subsequently rolled over), but distributions are subject to income tax when withdrawn. Suppose that one were willing to forego one dollar in current consumption and one has access to an investment that will earn a gross return R between now and the date of retirement. One could contribute $\frac{1}{1-T_0'}$ dollars to an IRA, where T_0' is the marginal tax rate in the year of contribution; this foregoes one dollar of current consumption. Investment income earned in the IRA is not directly taxed, so the IRA would have an account balance of $\frac{R}{1-T_0'}$ at retirement. Since the distribution is subject to income tax, the after-tax amount of the distribution would be $R \times \frac{1-T_1'}{1-T_0'}$, where T_1' is the marginal tax rate at the time of distribution. If tax rates are flat and constant over time, then that dollar of foregone current consumption would translate to R dollars of consumption at retirement.

Suppose instead that the foregone dollar in consumption were invested in a taxable account. The initial contribution would be one dollar, since the contribution is made after-tax. The taxable account has access to same investments, creating a pre-tax return of R; however, the investment income itself (in the form of interest, dividends, and capital gains) is taxable, leading to an effective rate of return $R^T < R$. The account would be worth R^T at retirement; since there is no tax on distribution, this account could finance R^T of consumption at retirement. Comparing the IRA to the taxable account, we see that the IRA advantage is $R - R^T$, which reflects exactly the fact that the investment returns in IRAs are not directly taxed.

1.2 Required minimum distributions (RMDs)

RMDs have been part of the tax code since 1962, when Congress required participants of newly-established Keogh plans to begin taking distributions beginning in the year in which the age of 70.5 is attained. Congress has generally applied

[8]However, among the analysis sample discussed below (which is conditioned on owning an IRA at age 69), IRAs unsurprisingly represent a larger share of retirement income.

these rules to new forms of tax-advantaged savings accounts as they have created them. For the purpose of this paper, the most important types of accounts to which RMDs apply are (traditional) IRAs, workplace defined contribution plans, and two special types of IRAs for small businesses (known as SEP IRAs and SIMPLE IRAs).[9] Roth IRAs are not subject to RMDs.

Individuals must generally begin making RMDs *for* the calendar year in which they turn 70.5. The required distribution is equal to a certain fraction of the aggregated IRA balance (including SEP and SIMPLE IRAs) at the end of time $t-1$. For most individuals, this fraction is a simple function of age (in years) at time t; for 70-year olds (i.e., those who turn 70.5 but not 71 in a given year), this fraction is $\frac{1}{27.4} \approx 3.65$ percent. This fraction increases with age, in a manner concordant with the reduction is remaining life expectancy.[10] An individual is entitled to a "grace period" in his or her first year of being subject to an RMD: the RMD for that first year may be taken through April 1 of the following year. This grace period does not apply in any other year. [11]

In general, IRA fiduciaries are required to notify account holders of the presence and amount of an RMD. The tax code specifies that failure to make an RMD will trigger a 50 percent excise tax due on the undistributed amount. However, as shown by Mortenson, Schramm, and Whitten (2018), this excise tax is rarely applied, since the IRS has the authority to waive the penalty if it was caused by "reasonable error" and the taxpayer took reasonable steps to correct that error. Nevertheless, Mortenson, Schramm, and Whitten (2018) show that RMD rules appear to be substantially binding, as a large number of individuals bunch exactly at the RMD amount and reduced their distributions in a year when RMD rules were suspended.

2 Conceptual Framework

In this section, I put forward a simple model to illustrate some predicted effects of the RMD treatment. In particular, I assume that individuals arrive at time t with "pension" assets P_t (including IRAs) and taxable assets A_t. Pension assets earn a gross yearly return R_P while taxable assets earn a gross yearly return R_A, with $R_P > R_A$. For all individuals, consumption at time t is given by the following budget constraint:

$$c_t = \underbrace{R_P P_t - P_{t+1}}_{=d_t\text{,Pension distributions}} + \underbrace{R_A A_t - A_{t+1}}_{=-s_t\text{,Taxable dissaving}} \tag{1}$$

[9]In general, an individual's RMD from a DC plan may be deferred until the date when the individual separates from the employer that sponsors that DC plan. Additionally, inherited IRAs are subject to RMDs prior to age 70.5 as well.

[10]This formula is more complicated for individuals with a spouse that is more than ten years younger that is the sole beneficiary of the IRA.

[11]As a result, an individual may end up making two RMDs in the year in which they turn 71.5: the one required to be made for year t and the one required to be made for $t-1$.

Importantly, all individuals face the following RMD constraint:

$$R_P P_t - P_{t+1} \geq \theta_t P_t \tag{2}$$

Lastly, I assume that all individuals face nonnegativity constraints on P_{t+1} and A_{t+1}.[12]

An individual behaving according to neoclassical assumptions would solve the following Bellman equation at time t, subject to the nonnegativity constraints and the constraints in (7) and (8).

$$V^t(A_t, P_t) = \max_{A_{t+1}, P_{t+1}} u(c_t) + \beta V^{t+1}(A_{t+1}, P_{t+1}) \tag{3}$$

As I show in the Appendix, the solution to this model has a few salient features. First, it is optimal to exhaust taxable assets before taking pension distributions in excess of the RMD; this is quite intuitive, since the two types of assets are otherwise identical, but pension assets earn a higher return. Second, so long as taxable assets at $t + 1$ are positive (which will be the usual case when RMDs bind), the usual consumption Euler equation holds with R_A being the relevant interest rate. This means that changing θ_t from θ to θ' can only affect consumption through a wealth effect, so long as RMDs are binding at both θ and θ'.[13] From the budget constraint, consumption is equal to pension distributions plus taxable dissaving. If the consumption response to a change in θ_t is small, then it must be the case that the taxable dissaving response to a change in θ_t roughly offsets the "first-stage" response on pension distributions. Therefore, this neoclassical model predicts a crowd-out ratio of approximately one.

However, under certain extensions of this model, we might not expect crowd-out to be approximately one. For instance, there might be a third asset ("money", M_t) which earns no return (and thus does not appear as saving in the data) but that individuals hold due to some preference for liquidity. Thus, the budget constraint $c_t = d_t - s_t$ becomes $c_t = d_t - s_t - s_t^M$, where s_t^M is saving in the form of money. In this case, even if c_t is little affected by a change in θ_t, the effect on d_t could be offset by some combination of an effect on s_t and s_t^M, potentially reducing crowdout to be less than one.[14] Finally, there are a variety of behavioral models that suggest that consumption *might* change in response to an induced increase in d_t. In particular, a mental accounting model in which individuals perceive the RMD distribution as a windfall gain would tend to cause a substantial consumption response in the present period, which would tend to reduce estimated crowdout.

[12] Note that this concept of taxable assets is slightly different than in the empirical section. In the empirical section, it is possible to have negative taxable net assets due to mortgage debt. In reality, such individuals presumably have positive net assets ("equity") accounting for the value of the property secured by that mortgage (which I do not observe in the data).

[13] This is an important sub-case. For instance, $\theta = 0$ represents the case where "RMDs" are not in effect, but where individuals are not allowed to contribute to IRAs or similar accounts. This will generally be the case for retirees prior to the age of 70.5, since IRA contributions are only allowed up to taxable compensation. To the extent that individuals prior to age 70.5 are not taking IRA distributions, then they are (weakly) bound by this constraint.

[14] Empirically, this would be similar to the case when taxable assets of the analysis sample earn a yield lower than what is implied by the capitalization factors.

3 Data

I draw data from the database which holds the near-universe of most major tax forms from 1999 to the present. For a given experiment year τ, I construct the sample as follows. I require individuals to turn 70 in year τ and to hold an IRA (with positive value) as of $\tau - 1$. These individuals will generally be required to make an RMD for year τ (the treatment group) and others will not (the control group). I then collect data for these individuals and their spouses from times $t = \tau - 3$ to $t = \tau + 6$. For the main analysis, I collect (1) time-invariant characteristics, namely exact date of birth, (2) 1040-level information on retirement distributions, interest income, and dividend income, (3) information on IRA holdings from Form 5498, (4) information on mortgage interest paid from Form 1098, and (5) enough information from various forms to make a rough marginal tax rate calculation. I then repeat this process for $\tau =$ 2002 through $\tau = 2014$, excluding $\tau = 2009$ when RMD rules were suspended.[15]

I calculate saving by capitalizing changes in three different income flows. The three flows are (taxable) interest, dividends, and mortgage interest.[16] For each income type j, I calculate saving s_{ijt} as equal to $(z_{ij,t+1} - z_{ijt})c_{jt}$, where z_{ijt} is income of type j during calendar year t and c_{jt} is a capitalization factor.[17] I calculate capitalization factors following closely the method of Saez & Zucman (2016). These capitalization factors are equal to a ratio. The numerator of that ratio is the aggregate assets of type j in the economy at time t, as reported in the Financial Accounts of the United States, that is owned in accounts that could produce taxable income for individuals. The denominator aggregate amount of income of type j in the economy at time t.[18] For taxable interest and dividends, I calculate the aggregate amount of income using cleaned, representative cross-sections created by the Statistics of Income division of the IRS. I calculate the aggregate amount of mortgage interest appearing on Form 1098 using the same underlying database (which represents the near-universe of Forms 1098) from which I draw the main sample. I plot the time series of each capitalization factor in Appendix Figure A1.

[15] The most recent year of quality data is 2015 (due to how the capitalization factors below are calculated). Thus, for $\tau \geq 2011$, I am not able to retrieve data through $\tau + 6$. However, the main specifications of this paper require data through $\tau + 1$ only; the data represent a fully balanced panel over this period.

[16] Mortgage interest is calculated from Forms 1098 received by an individual and his or her spouse (if filing jointly). An individual need not itemize deductions in order to receive Form 1098. Taxable interest and dividends are taken from Form 1040.

[17] In the case of mortgage interest, s_{ijt} is defined as the opposite of the expression given, since mortgage debt represents a liability, not an asset.

[18] The denominator of this ratio is proportional to the average amount of assets held throughout the 12 months of calendar year t, while the numerator represents the aggregate amount held at the end of time t. Therefore, prior to calculating the ratio, I take the mean of the numerator at times t and $t-1$. This ratio, c_{jt}^M, represents the capitalization factor that applies at the midpoint of time t. However, we are interested in the capitalization factor that applies at the end of time t, not the midpoint (since we are comparing income flows in t to those in $t + 1$). Therefore, I define c_{jt} as equal to the mean of c_{jt}^M and $c_{j,t+1}^M$. Since capitalization factors do not change substantially from one year to the next, this adjustment has little effect on the estimates.

I define IRA distributions as distributions on Form 1099-R identified as being from an IRA (including SEP and SIMPLE IRAs, but not Roth IRAs).[19] I define total retirement distributions as IRA distributions (as measured above) of the individual and his or her spouse, plus taxable pensions from Form 1040. Since we are interested in comparing taxable saving (which is made with post-tax dollars) and retirement distributions (which I assume to be made with pre-tax dollars), it is necessary to convert them into the same units. I do so by computing the net-of-tax amount of retirement distributions. Using a simplified tax calculator, I compute the incremental tax liability owed on the outcome of interest (e.g., IRA distributions, or retirement distributions more generally) and subtract this from the gross distribution. This simplified tax calculator accounts for the standard rate structure, the partial taxation of Social Security benefits, the interaction of the standard rate structure with capital gains, and a simplified version of the Alternative Minimum Tax.

Table 1 reports summary statistics at time $\tau - 1$, the year in which each individual attains 69. Half of this sample is male and 71 percent is married filing jointly. Non-retirement income is unsurprisingly right-skewed, with a mean ($65,592) that exceeds the 75th percentile ($57,400).[20] The mean and median IRA values (at the end of $\tau - 1$) are $168,596 and $60,300 respectively; this values would correspond in most cases to a pre-tax "potential RMD" of approximately $6,150 and $2,200, respectively. Most individuals are not taking an IRA distribution at $\tau - 1$, however most tax units are taking some form of IRA or pension distribution. Net taxable assets and taxable saving are, of course, quite noisy; the standard deviations are orders of magnitude larger than means and the three quantiles reported. This previews the necessity of noise-mitigation strategies discussed in Section 4 below. In the final rows of the table, I report each income flow that is part of the saving calculation. Over 82 percent of individuals have positive interest income, with a median of about $500. A somewhat smaller share, 59 percent, have positive dividend income. Only 41 percent are paying mortgage interest. In sum, approximately 92 percent of individuals have some measured income flow with which I can calculate saving.

4 Method: Regression discontinuity in general

I use a regression discontinuity (RD) design to estimate the effect of being subject to RMDs at age 70 (year τ) on (1) outcomes related to retirement distributions and (2) taxable saving. These broadly correspond to the first stage and reduced form, respectively, of a fuzzy regression discontinuity to estimate the rate at which retirement distributions crowd out taxable saving in this context. I will also estimate versions of this fuzzy RD. In this section, I present

[19] A Roth conversion would generally trigger a 1099-R which appears similar to a regular distribution. Therefore, I subtract observed Roth conversions from observed IRA distributions.

[20] Throughout this paper, all dollar amounts are reported in 2016 dollars, as inflated/deflated by the PCE deflator. All quantiles are rounded to the nearest $100 to protect taxpayer privacy.

Table 1: Summary statistics

	Mean (1)	25th percentile (2)	Median (3)	75th percentile (4)
Male	0.491			
	(0.500)			
Married	0.707			
	(0.455)			
Any wage income	0.420			
	(0.494)			
Non-retirement income	65,592	5,100	22,100	57,400
	(718,993)			
IRA value	168,596	19,500	60,300	163,100
	(430,618)			
IRA distributions (individual)	5,675	0	0	2,800
	(32,029)			
IRA and pension distributions (tax unit)	24,957	2,500	16,000	34,300
	(49,538)			
Net taxable assets	416,051	0	35,200	258,800
	(9,007,299)			
Taxable saving	12,524	-22,400	0	23,800
	(7,715,971)			
Interest income	5,250	(—)	500	3,100
	(84,740)			
Share with positive amount:	0.825			
Dividend income	5,781	0	100	2,300
	(163,655)			
Share with positive amount:	0.586			
Mortgage interest	3,928	0	0	3,800
	(39,530)			
Share with positive amount:	0.406			
Share with any investment income	0.916			
	(0.277)			
Observations	10,882,700			

Notes: This table reports summary statistics for the analysis sample, where quantities are measured as of the age-69 year ($\tau - 1$). Observations with birthdays between June 29 and July 2 are dropped. Quantiles and observation counts are rounded to the nearest 100 to protect taxpayer privacy. An entry of "(—)" indicates that the value is positive but less than 50. All dollar amounts are adjusted for inflation and expressed in 2016 dollars. See Section 3 for a discussion of variable construction.

the baseline RD strategy for a general outcome y_i when treatment status D_i changes discretely and sharply by one unit as some running variable x_i increases beyond some cutoff c. The RD estimate for the effect of treatment D_i is given as follows.

$$\hat{\beta} = lim_{x \to c^+} \hat{E}(y|x) - lim_{x \to c^-} \hat{E}(y|x) \tag{4}$$

In the present context, x represents age at the end of year τ, relative to 70.5. I scale x in days. To be symmetric, I define $x = -0.5$ for being one day younger than 70.5 (i.e., having a birthday on July 1) and $x = 0.5$ for being one day older than 70.5; thus, x takes on the values $\{\ldots, -1.5, -0.5, 0.5, 1.5, \ldots\}$. When $x > c = 0$, individuals are "treated", meaning that they are required to take an RMD for year τ.

To recover the point estimate of the treatment effect, we need to estimate the difference in (4). The most common way of doing so is via local linear regression. In particular, the estimator for $lim_{x \to c^+} \hat{E}(y|x)$ is the constant from a regression of y on x, using observations only to the right of the cutoff, using a kernel weight $K\left(\frac{x-c}{h}\right)$ for some bandwidth h. The kernel weight gives a higher weight to observations near $x = c$; for most choices of kernel (such as the common triangular kernel, which I use), it puts a zero weight on observations where $|x - c| > h$. The estimator for the left-hand-side limit, $lim_{x \to c^-} \hat{E}(y|x)$ is defined similarly. In fact, this RD estimate can equivalently be recovered as $\hat{\beta}$ from the following regression, weighted by $K_h\left(\frac{x_i - 0}{h}\right)$, where D_i is an indicator for $x_i \geq 0$:

$$y_i = \alpha + \gamma_1 x_i + \gamma_2 x_i D_i + \beta D_i + \epsilon_i \tag{5}$$

The coefficient for the fuzzy RD can be recovered from a similar regression, replacing the D_i covariate with the endogenous outcome (retirement distributions), and instrumenting for that endogenous outcome with D_i. In practice, I estimate these regressions using data collapsed to the birthday (x) level, and weight the regressions appropriately. I use heteroskedasticity-robust standard errors (White (1980)), which is analogous to using standard errors clustered by x on a regression on the underlying microdata.[21]

In order for Equation (4) to recover the causal effect of D_i, it must be the case that the conditional expectations of potential outcomes are continuous around the threshold. In other words, $E(y|x, D = 0)$ and $E(y|x, D = 1)$ must each be continuous in x at $x = 0$. One particular such threat is manipulation of the running variable – either retiming of births or manipulation of recorded dates of births, typically over very short time horizons.[22] To caution against this threat, I exclude observations with birthdays between June 29 and July 2 in all main specifications. I discuss in detail this and other threats to identification in Section 7.

4.1 Choice of bandwidth

In general, there is a bias/variance tradeoff in the choice of the bandwidth h. A larger h will increase the effective sample size, reducing variance. But to the extent that the true data generating process is not linear, a larger h will also introduce a greater amount of bias from misspecification of (5). There are a

[21]As a sensitivity check reported in Section 8, I also use the standard errors (and bias-correction) proposed by Calonico, Cattaneo, and Titiunik (2014). This leaves the main results qualitatively unchanged.

[22]While such manipulation would obviously not be taken in response to RMDs, there may be other eligibility thresholds that depend on having a birthday before or after July 1.

variety of ways to choose h, all designed to manage this tradeoff. In the baseline analysis, I use the "optimal" bandwidth of Calonico, Cattaneo, and Titiunik (2014); I will refer to this bandwidth hereafter as the CCT bandwidth.[23] In each of the main tables, I will additionally report results for bandwidths of 200 percent and 50 percent of the CCT bandwidth.

4.2 Quantile treatment estimates

In addition to the effect of some treatment on mean levels some outcome, policymakers are often interested in the effect of that treatment on different quantiles of the distribution of that outcome. In the context of a randomized experiment, the q^{th} "quantile treatment effect" (QTE) is equal to the difference between the q^{th} quantile of the treated population and the q^{th} quantile of the control population. Such QTEs are admittedly somewhat difficult to interpret without the addition of implausible assumptions. The QTE represents neither the q^{th} quantile of the treatment effect nor the treatment effect on the q^{th}-quantile individual (ranked by the dependent variable). Nevertheless, QTEs have some intuitive implications. In particular, if some QTE is positive, then it must be the case that the treatment effect is positive for some positive-measure interval of the (untreated) distribution of the dependent variable (Heckman, et al (1997)).

To estimate the QTEs in the RD setting, I follow the method of Frandsen, Frolich, and Melly (2012). The procedure is as follows.

1. For some value y of the dependent variable Y_i, define $U_i(y)$ as an indicator for $Y_i \leq y$.

2. Estimate separately $\hat{F}_1(y) \equiv lim_{x \to 0^+} U_i(y)$ and $\hat{F}_0(y) = lim_{x \to 0^-} U_i(y)$ using local linear regression.[24] This recovers the estimated empirical distribution of Y_i at each side of the discontinuity.

3. Repeat steps (1) and (2) for all y.[25]

4. Invert $\hat{F}_1(y)$ and $\hat{F}_0(y)$ to get the quantile functions $\hat{Y}_1(q)$ and $\hat{Y}_0(q)$.[26] The q^{th} QTE, denoted by $\Delta(q)$, is equal to $\hat{Y}_1(q) - \hat{Y}_0(q)$.

To compute confidence intervals, I repeat steps (1)-(4) for B bootstrap samples (where individuals with the same value of the running variable are resampled together), each time computing $\Delta_b(q)$. For each q, the bottom of the 95 percent confidence interval is given by the 2.5th percentile of the distribution of $\Delta_b(q)$, and the top is given by the 97.5th percentile.

[23] The CCT bandwidth is a refinement to the well-known Imbens and Kalyanaraman (2012) optimal bandwidth.

[24] I approximately follow the suggestion of Frandsen, Frolich, and Melly (2012) in setting the bandwidth for this regression as a function of the optimal bandwidth for estimating the limits of the analogous conditional means.

[25] In practice, I repeat steps (1) and (2) for 200 quantiles of the Y_i distribution.

[26] I linearly interpolate between the evaluated grid points of y in order to obtain the quantile function for an arbitrary $0 < q < 1$.

4.3 Winsorization and inverse hyperbolic sine transformation

Taxable saving is very noisy – both because of noise in the underlying variable and because of measurement error caused by the capitalization method. In particular, the standard deviation of measured taxable saving (in year τ) is more than ten times that of retirement distributions. As a result, estimating the effect on savings in raw levels is uninformative. There are several ways to reduce noise in order to recover informative estimates. First, one can Winsorize the data. In general, to Winsorize a variable, observations greater than the $(1-q)^{th}$ quantile are coded as equal to the $(1-q)^{th}$ quantile, and observations less than the q^{th} quantile are coded as equal to the q^{th} quantile.[27] This diminishes the role of extreme values, though the choice of q is somewhat ad hoc.

As an additional strategy, I also apply a monotonic transformation to that dependent variable prior to estimation in some specifications. If saving were uniquely positive, the natural log would be an obvious choice. However, saving is often non-positive. Therefore, I use the inverse hyperbolic sine (IHS) (Pence (2006)). Let s denote saving. In general, let $g(s;\theta) = \frac{1}{\theta} sinh^{-1}(\theta s) = \frac{1}{\theta}\ln(\theta s + \sqrt{\theta^2 s^2 + 1})$. Note that the derivative of this function is $g'(s;\theta) = \frac{1}{\theta^2 s^2 + 1}$. Thus, when s is large (relative to θ and relative to $\frac{1}{\theta}$), $g(s;\theta)$ has approximately the same shape as $\ln(s)$.[28] When s is smaller in magnitude, the function is more linear than $\ln(s)$. Of course, this functional form transformation will tend to yield different results depending on θ. I follow a maximum likelihood method proposed by Burbidge, Magee, and Robb (1988) to choose the value of θ. In particular, the method assumes that $g_i \equiv (s_i;\theta) = z_i'\beta + u_i$, with the strong assumption that u_i is i.i.d. normal with variance σ^2.[29] Since g_i is an invertible function of s_i, there is a well-defined likelihood of observing the set of s_i conditional on β, σ^2, θ. The nuisance parameters β and σ^2 can be concentrated out, and the resulting concentrated likelihood can be maximized to find the optimal θ. Given this optimal θ, I compute the transformation and estimate the regression discontinuity as usual. I compute standard errors via bootstrap (clustered by exact values of the running variable).[30]

To interpret the magnitude of the coefficient estimates, I report the median implied treatment effect (in terms of saving or scaled saving) across all individuals. Using a Rubin potential outcomes framework (Rubin (1974)), the implied treatment effect for an individual i is equal to the estimated value of s_i when $D_i = 1$ (denoted $s_i(1)$) minus the estimated value of s_i when $D_i = 0$ (denoted

[27] I do this Winsorization separately for each value of the running variable, separately by event time.

[28] Additionally, $g(0;\theta) = 0$ and $g(-s;\theta) = -g(s;\theta)$, meaning that large negative values are also shrunk toward zero.

[29] In this case, z_i includes a constant, the treatment variable D_i, the running variable x_i, and its interaction with treatment $x_i D_i$.

[30] For this specification, I hold the bandwidth fixed at 75 percent of the optimal value for the 10-percent Winsorized specification, which crudely accounts for the fact that this transformation will lead to a dependent variable with smaller variance. I leave to future work the more difficult problem of jointly determining the optimal θ and optimal bandwidth.

$s_i(0)$. We observe the actual value of $s_i(1)$ for treated individuals and the actual value of $s_i(0)$ for untreated individuals, and thus we can use that in place of the estimated value. For untreated individuals, the estimated value of the unobserved $s_i(1)$ is equal to $g^{-1}\big(g(s_i(0);\theta^*)+\beta\big)$. Similarly, for treated individuals, the estimated value of the unobserved $s_i(0)$ is equal to $g^{-1}\big(g(s_i(1);\theta)-\beta\big)$.

These two noise-reduction measures have different advantages and disadvantages. On the one hand, the IHS method has the advantage of being more principled, in that there is no requirement to choose a Winsorization threshold q. On the other hand, the magnitudes of the estimated effects are somewhat difficult to interpret. Furthermore, the theoretical crowdout parameter refers to the increase in taxable saving, in dollars, caused by the (exogenous) increase in retirement distributions, in dollars. For this reason, the two methods (IHS and Winsorization) complement each other. The former is most useful for showing whether a reduced form effect exists – that is, whether taxable saving changes discontinuously at $x = 0$ – while the latter is most directly used for estimating the magnitude of crowdout.[31]

5 First stage and reduced form

In this section, I evaluate the effect of the RMD requirement at age 70 on retirement distributions (the "first stage") and taxable saving (the "reduced form"). I do this broadly in two ways. Let s_i denote the outcome of interest, such as net-of-tax retirement distributions, or taxable saving. The simplest estimation is to define the dependent variable as s_i, in levels. This recovers the average effect the RMD treatment on retirement distributions, in dollars. However, basic introspection suggests that the treatment effect might be approximately linear in the amount of the (actual or potential) RMD, since this reflects the amount of actual treatment.[32] Thus, I also estimate a version of (5) with the dependent variable equal to s_i divided by RMD_i, where RMD_i is the (net-of-tax) RMD that individual i would face if she were in the treatment group. I weight this regression by RMD_i, except that to avoid excessive influence by a very small number of individuals, I censor this weight at \$100,000, which represents approximately the 99.9th percentile of the distribution of RMD_i. The results of this regression can be interpreted as (approximately) the effect of the RMD treatment on s_i, as a share of the potential IRA RMD.

[31] I will also use other threshold-based methods to estimate crowdout, as discussed below.

[32] Ideally, the potential RMD would include the RMD that would apply to any defined contribution retirement plan; unfortunately, this is not observed. However, the aggregate flows reported in Goodman, et al (2019), combined with the fact that the sample is conditioned on holding an IRA at age 69, suggest that the IRA component makes up the vast majority of the total RMD.

5.1 First stage: Effect of RMD treatment on retirement distributions

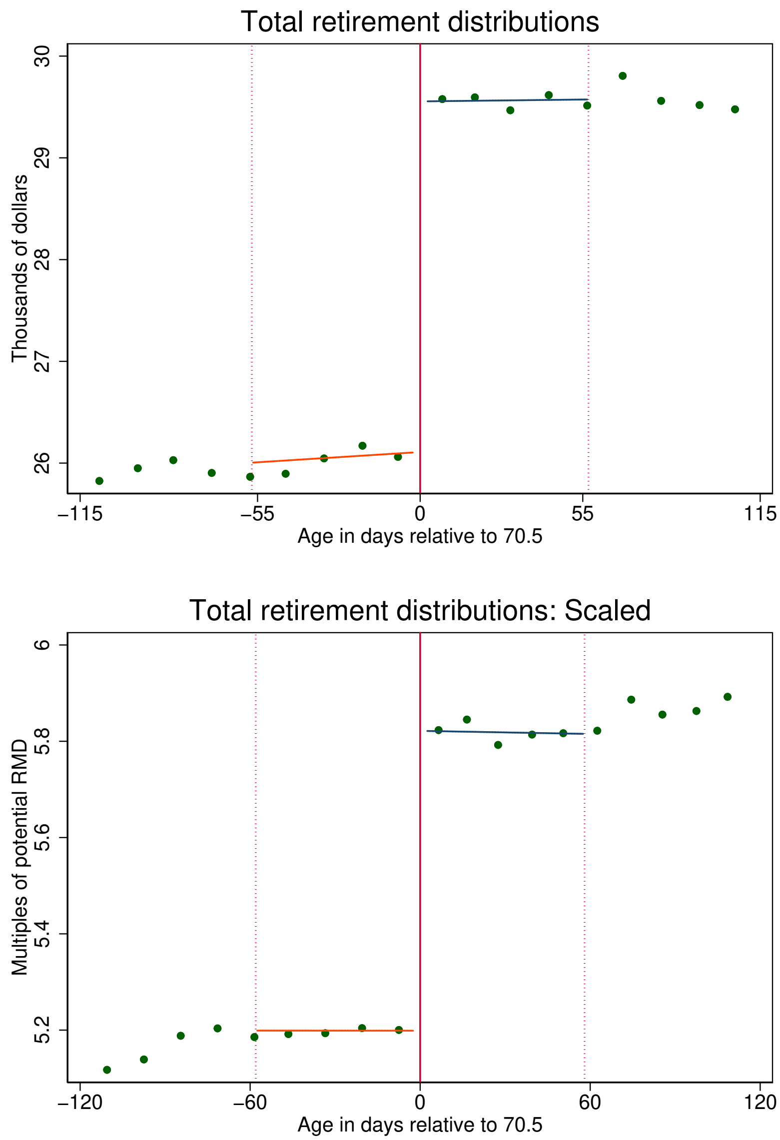
The "first stage" represents the extent to which the treatment caused an increase in (net-of-tax) retirement distributions. Retirement distributions include distributions from one's own IRA, as indicated on Form 1099-R. This component is most directly affected by the RMD treatment. Additionally, retirement distributions include distributions from the spouse's IRA, as well as taxable pensions as reported on Form 1040. The RMD treatment does not directly affect distributions from the spouse's IRA; however, couples could potentially react to the RMD treatment by shifting distributions from the spouse's IRA to the individual's. Furthermore, taxable pensions could be affected directly and indirectly by the RMD treatment – indirectly through a similar substitution channel as spousal IRAs, and directly through the application of RMDs to defined contribution pensions. For the sake of estimating crowdout, we want to incorporate such responses in the first stage.

In the top panel Figure 1, I plot retirement distributions (d_i) at age 70 as a function of the running variable. The bottom panel shows a similar plot, but with d_i scaled by the potential RMD. The points in the Figure are simple means taken within bins, where the bin size is selected to minimize an integrated mean squared error criterion using the method of Calonico, Cattaneo, and Titiunik (2015).[33] Additionally, I show the fits from (5), estimated using the optimal CCT bandwidth, which in this case is 56.92 days, as bracketed by the dotted vertical lines.[34] This Figure indicates that there is a strong "first stage" for our crowdout estimate: there is a clear increase in retirement distributions at $x = 0$. This is an important result in and of itself, confirming the work by Mortenson, Schramm, and Whitten (2018) and Brown, Poterba, and Richardson (2017) using other sources of variation that RMDs do appear to be binding for some portion of the IRA-holding population.

[33]This procedure is implemented in the Stata command `rdplot` (Calonico, et al (2017)).

[34]Note that, at $x = x' \neq 0$, these fits do not represent the local linear regression estimate for $E(y|x = x')$. Such an estimate would require recentering the kernel at $x = x'$. However, predictions away from $x = 0$ are not directly relevant for estimating (5). In order to illustrate the estimation of the discontinuity, I opt to plot the fits from (5) instead of the pointwise predictions.

Figure 1: First stage: R.D. estimates of RMD status on IRA distributions at age 70



Notes: This figure plots the average value of retirement distributions as a function of the running variable, x (age in days relative to 70.5). Those to the right of 0 are required to take an RMD for age 70; those to the left are not. Retirement distributions are (net-of-tax) total IRA distributions of the reference individual and his or her spouse, as measured on Form 1099-R, plus taxable pensions as reported on Form 1040. In the top panel, the dependent variable is retirement distributions in levels. In the bottom panel, the dependent variable is total retirement distributions divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). The specification in the bottom panel weighted by the potential RMD, censored at $100,000. Bin size is selected using the `rdplot` command in STATA. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. All dollar amounts are adjusted for inflation and expressed in 2016 dollars.

Table 2: First stage: Effect of RMD status at age 70 on retirement distributions at age 70

	In levels (000s)			Scaled by potential RMD		
		Bandwidths			**Bandwidths**	
	200 percent	Baseline	50 percent	200 percent	Baseline	50 percent
	(1)	(2)	(3)	(4)	(5)	(6)
Total retirement distributions	3.439***	**3.446*** **	3.532***	0.599***	**0.623*** **	0.601***
	(0.094)	**(0.141)**	(0.232)	(0.024)	**(0.035)**	(0.052)
	[113.83 days]	**[56.92 days]**	[28.46 days]	[116.04 days]	**[58.02 days]**	[29.01 days]
Own IRA distributions	2.987***	**2.982*** **	3.010***	0.574***	**0.572*** **	0.569***
	(0.054)	**(0.066)**	(0.082)	(0.008)	**(0.012)**	(0.017)
	[94.87 days]	**[47.43 days]**	[23.72 days]	[115.01 days]	**[57.51 days]**	[28.75 days]
Spouse's IRA distributions	-0.068*	**-0.060**	-0.080	-0.023***	**-0.020**	-0.039**
	(0.036)	**(0.050)**	(0.081)	(0.009)	**(0.013)**	(0.020)
	[141.14 days]	**[70.57 days]**	[35.28 days]	[109.32 days]	**[54.66 days]**	[27.33 days]
Taxable pensions of tax unit	0.479***	**0.506*** **	0.585***	0.050***	**0.070** **	0.069
	(0.053)	**(0.088)**	(0.161)	(0.017)	**(0.028)**	(0.047)
	[153.35 days]	**[76.67 days]**	[38.34 days]	[114.56 days]	**[57.28 days]**	[28.64 days]

Notes: This table reports regression discontinuity estimates of being required to take an RMD for the age-70 year (year τ) on retirement distributions in that year. Each cell refers to a separate regression. In the first row, the dependent variable is "total retirement distributions," which includes IRA distributions of the reference individual and his or her spouse, as measured on Form 1099-R, as well as taxable pensions as reported on Form 1040. In the remaining rows, the dependent variable is the indicated component of total retirement distributions. In columns 1-3, the dependent variable is retirement distributions in thousands of dollars. In columns 4-6, the dependent variable is retirement distributions divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). Specifications in columns 4-6 are weighted by the potential RMD, censored at $100,000. In columns 2 and 5, the bandwidth is selected using the method of Calonico, Cattaneo, and Titiunik (2014). In columns 1 and 4, the bandwidth is 200 percent of the baseline bandwidth; in columns 3 and 6, it is 50 percent of the baseline bandwidths. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors allow arbitrary correlation of errors within individuals with the same value of the running variable. All dollar amounts are adjusted for inflation and expressed in 2016 dollars. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

I report coefficient estimates for the specification plotted in Figure 1 in the top row of Table 2. The left three columns correspond to the top panel of Figure 1 and columns 4 through 6 correspond to the bottom panel. I vary the bandwidths within each set. Columns 2 and 5 (in bold) represent the baseline bandwidth (which was used in Figure 1), while columns 1 and 4 (3 and 6) 3 use 200 percent (50 percent) of this baseline bandwidth. The baseline result is that retirement distributions at age 70 increase by $3,446 when subject to RMDs at age 70, or about 62 percent of the RMD. Both results are quite robust to bandwidth.

In the remainder of Table 2, I break out the total effect on retirement distributions into their three subcomponents.[35] The effect on distributions from one's own IRA represents the vast majority of the total effect on retirement distributions. This is unsurprising, given that IRAs represent the bulk of assets subject to RMD in the general population of 70-year-olds (Goodman, et al (2019)) and because the present sample is conditioned on holding an IRA at age 69. The effect on spouse's IRA distributions is negative, consistent with a substitution response, but the magnitude is small and the estimates are mostly statistically insignificant. This small effect on spousal IRAs can be explained by considering the relatively narrow set of tax units that would be likely to respond in this manner. For such a response to occur, the tax unit must be (1) bound by RMDs once the reference individual becomes subject to them and (2) *not* bound by RMDs if the reference individual were not subject to them – i.e., if the potential spousal distribution when not treated is greater than the spouse's RMD (or zero, if the spouse is not subject to an RMD him or herself) but less than the spouse's RMD plus the reference individual's RMD. This is a relatively knife-edge circumstance, explaining the small (if any) effect on spousal IRAs. The effect on taxable pensions is positive, though much smaller than the effect on IRAs. Recall that this latter estimate combines two effects: first, a positive treatment effect due to most DC pensions also being subject to RMDs with the same discontinuity, and second, a negative treatment effect due to adjustment. In sum, there is strong evidence that the RMD treatment increased retirement distributions in general, and IRA distributions in particular, at age 70.

In Figure 2, I repeat the same analysis, separately by event time k, equal to $t - \tau$.[36] In addition to the effect at event time 0, which simply replicates the results above, there are several features of these results to note. First, the estimated effects prior to treatment (at event times -3, -2, and -1) are negative, but small in magnitude and generally insignificant. Second, the effect is approximately zero at event time 1 and is negative (and significant) at event time 2. Fourth, the estimated effect is modestly negative in event times 3 through the end of the sample window (though somewhat less negative than at event time 2). I explore each of these features in more detail, below.

The first result is that the effect of the RMD treatment is zero (using raw

[35] These specifications are illustrated graphically in Appendix Figure A2.

[36] This figure uses the baseline sample. This sample is unbalanced: the later experiment years do not have six years of data after the experiment year. I show a similar plot using a balanced sample (i.e., $2002 \leq \tau \leq 2010$) in Appendix Figure A3. Results are quite similar.

distributions) or slightly negative (using scaled distributions) in the three years prior to treatment. The fact that these effects are very small relative to the effect at τ is reassuring: despite potential concerns about manipulation of the running variable and other threats to identification considered in Section 7, IRA distributions prior to age 70 appear to be mostly unaffected. The slight negative effect seen using scaled distributions could reflect a small amount of residual composition bias, or potentially forward-looking behavior. Given that these effects are small, they are unlikely to substantially affect the crowdout estimates.

The second result – the effects seen at event times 1 and 2 – are related to two additional features of the institutional setting and the types of responses. The first feature is the "grace rule": individuals in their first year of being subject to an RMD may choose to defer that RMD until April 1 of the following year (in which case the distribution will be observed in the data in that following year). However, in all other years, the RMD must be taken prior to December 31. Even in the absence of any behavioral adjustment, we would expect the application of the grace rule to cause an estimated effect at event times 1 and 2. At event time 1, some share of treated individuals will take both their τ and $\tau+1$ RMDs, while some share of control individuals will choose to defer their $\tau+1$ RMD into the following year. This would generally lead to a positive treatment effect at event time 1. Similarly, some share of control individuals would take both their $\tau+1$ and $\tau+2$ RMDs at event time 2, leading to a negative treatment effect at event time 2. This latter negative effect at event time 2 is apparent in Figure 2. However, there is little evidence of a positive effect at event time 1; while the coefficient estimates are positive, they are small and (for scaled distributions) insignificant. This is likely caused by an additional margin of response: the extensive margin. In Appendix Figure A4, I show that being treated appears to cause a reduction in the probability of holding an IRA (at the end of year τ) by about one percentage point.[37] Since by construction the entire sample holds an IRA at the end of $\tau-1$, such an effect is driven entirely by account closures. This account closure response has the opposite effect at time $k=1$, when a larger portion of the *control* group closes their account, offsetting the mechanical application of the grace rules.

Lastly, the third result – the slightly negative effects at event times 3 through 6 – can be partially explained by a mechanical effect of the treatment group having lower RMDs because they took an extra distribution at time 0 (since RMDs are a fraction of IRA balance). Let p denote the unknown fraction of "compliers" – those who increase their distributions at time τ because they were subject to an RMD at that time. These are the individuals that are bound by RMDs. Further, assume that this status is approximately time-invariant over this number of years, meaning that p is constant over time. From Table 2, the increase in retirement distributions was \$3,446, or 0.62 multiples of the potential RMD, at $k=0$. This implies that the increase in retirement distributions

[37] Mortenson, Schramm, and Whitten (2018) also find such extensive margin responses to RMDs.

among compliers was $\frac{\$3446}{p}$ dollars. The RMD percentage is approximately 4.5 percent over ths age range, so these individuals would have reduced their distributions by approximately $0.045 \times \frac{\$3446}{p}$. Thus, this effect alone would case a reduction in distributions at event times 3 through 6 of approximately $0.045 \times \$3,446 \times \frac{p}{p} \approx \$160$. The analogous calculation for the scaled specification is $\approx$ 0.03 of the potential RMD. In general, -\$155 is near the bottom of the confidence interval for the each specification in levels at event times 3 to 6, meaning that one cannot rule out a simple mechanical application of the RMD rules explaining the bulk of these results.[38] However, -0.03 is outside the confidence interval for the scaled specifications, which are estimated somewhat more precisely.[39] This suggests that some form of behavioral response explains a portion of this effect, potentially consistent with the neoclassical model presented in Section 2. Under that model, the optimal choice is to exhaust taxable assets first (since they earn a lower rate of return), followed by IRA assets (or retirement assets more generally). An increase in IRA distributions at event time 0 could cause a marginal increase in the date at which taxable assets are exhausted, reducing IRA distributions during that marginal period. This would also be consistent with a less tractable but more realistic extension of that model which replaces the nonnegativity constraint on taxable assets with a convex flow utility of holding positive taxable assets in a given period (i.e., due to a liquidity preference). An increase in retirement distributions at time 0 increases the taxable asset stock, which allows agents to take smaller retirement distributions in the future.

5.2 Reduced form: Effect of RMD treatment on taxable saving

The previous subsection found clear evidence that being subject to RMD rules increases IRA distributions in particular and retirement distributions in general. In this subsection, we test whether there was also an effect on taxable saving. As discussed previously, one would expect an effect on taxable saving unless (1) the RMD treatment affects the path of consumption or (2) the RMD treatment tends to be absorbed in a non-interest-bearing account such as a standard checking account.

For the purpose of estimating crowdout, taxable saving should be calculated including the entire period over which we might expect a response. However, even under the assumption that the response to an exogenous increase in retirement distributions is contemporaneous, the timing of measured taxable saving creates a complication. Recall that retirement distributions at time τ reflect distributions taken during the calendar year. By contrast, taxable saving at time τ reflects the difference between average assets during calendar year $\tau + 1$ and average assets during calendar year τ. If the induced retirement distributions

[38]Such results are presented in Appendix Table A1 using this baseline sample. The analogous results for the balanced sample are reported in Appendix Table A2.

[39]Results for the scaled sample are presented in Appendix Table A3 using the baseline sample and in Appendix Table A4 for the balanced sample.

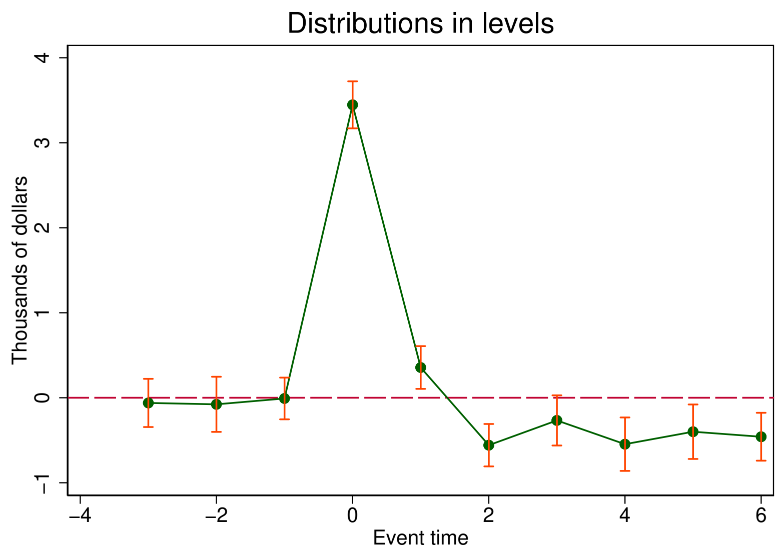
are taken at any time prior to the end of the year (with a contemporaneous savings response), then treatment will tend to increase average assets at τ to some degree, reducing measured saving at τ. The solution to this problem is to aggregate taxable saving over an additional year, effectively comparing $\tau+1$ assets to $\tau-1$ assets. This strategy requires the first stage effect on retirement distributions to be 0 at $\tau-1$ and $\tau+1$, which is approximately true – see Figure 2. Therefore, I will proceed under such aggregation.[40]

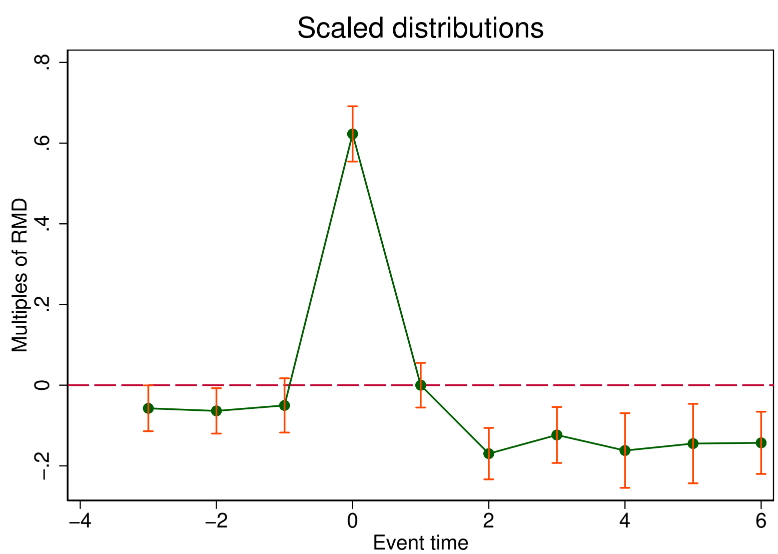
The top row Table 3 reports the coefficient results for taxable saving aggregated between times τ and $\tau-1$, in raw levels. Due to the large amount of noise in the data, this result is uninformative. The remainder of the three rows of Table 3 estimate (5) with the dependent variable Winsorized, with $q=0.01$, $q=0.05$, and $q=0.1$, respectively. As q increases, the coefficients and standard errors tend to shrink toward zero. With 10 percent Winsorization, the baseline estimate (column 2) in levels is \$1,452. This estimate is similar when the bandwidth is doubled, though it is less robust to the smaller bandwidth in column 3 (which has a 40 percent larger standard error). I show these four estimates graphically in the upper-left, upper-right, and lower-left panels of Figure 3. The discontinuity is visually apparent, especially for the 10-percent Winsorized data (lower-left panel), though the remaining noise is also evident. The analogous specifications using the scaled data yield positive results, but the standard errors are sufficiently large to make the scaled estimates statistically insignificant. These estimates are shown graphically in Appendix Figure A5.

Next, I estimate quantile treatment effects for saving and scaled saving (aggregated between τ and $\tau-1$). In the context of this reduced form, the primary purpose of the QTEs is to increase precision, as quantiles (away from 0 and 1) tend to be estimated more precisely than means. These results are shown in Figure 4. The top panel uses saving in levels while the bottom panel uses saving scaled by the potential RMD. For both outcomes, there is an interval (which includes the 50th percentile) where the QTE is very close to zero. This corresponds to the interval where the left-hand-side limit and the right-hand-side limit are each estimated to be very close to zero – this corresponds mostly to the small but non-negligible portion of this population that does not have any taxable interest or dividend income or mortgage payments. Outside of this interval, the QTE is positive and significantly different from zero for saving in levels from the 25th to the 90th percentiles (Portions below the 10th and above the 90th percentiles are noisily estimated and omitted from the figure for the sake of readability). The QTE is smaller than the \$1,452 mean effect (from

[40] However, the assumption of zero first stage effect at $\tau+1$ is unlikely to be true at a finer level of time. For instance, the share of the treatment group that is responding to the grace rule will tend to take their distributions early in the year (prior to April 1), while the share of the control group taking their extra distributions might do so more uniformly throughout the year. These would have different effects on average assets at $\tau+1$. This suggests the need to aggregate both the first stage and the reduced form over additional years to encompass all years when there is a potential effect. Unfortunately, since the first stage effect is nonzero over all event times greater than or equal to 2, this strategy is infeasible. To the extent that average assets at $\tau+1$ are larger for the treatment group because their first stage response at $\tau+1$ is earlier in the year, then the estimates presented in Section 6 will be slightly overstated.

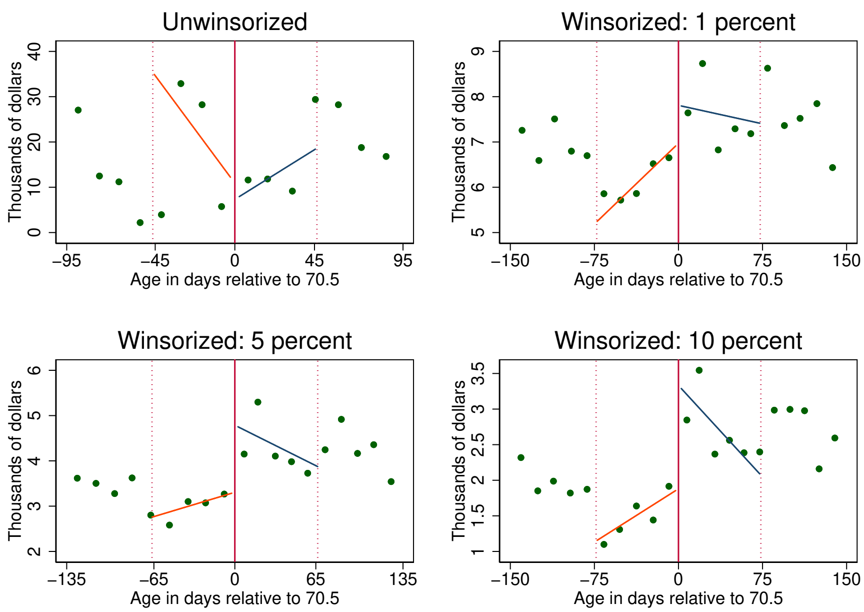
Figure 2: Effect RMD treatment on retirement distributions over time





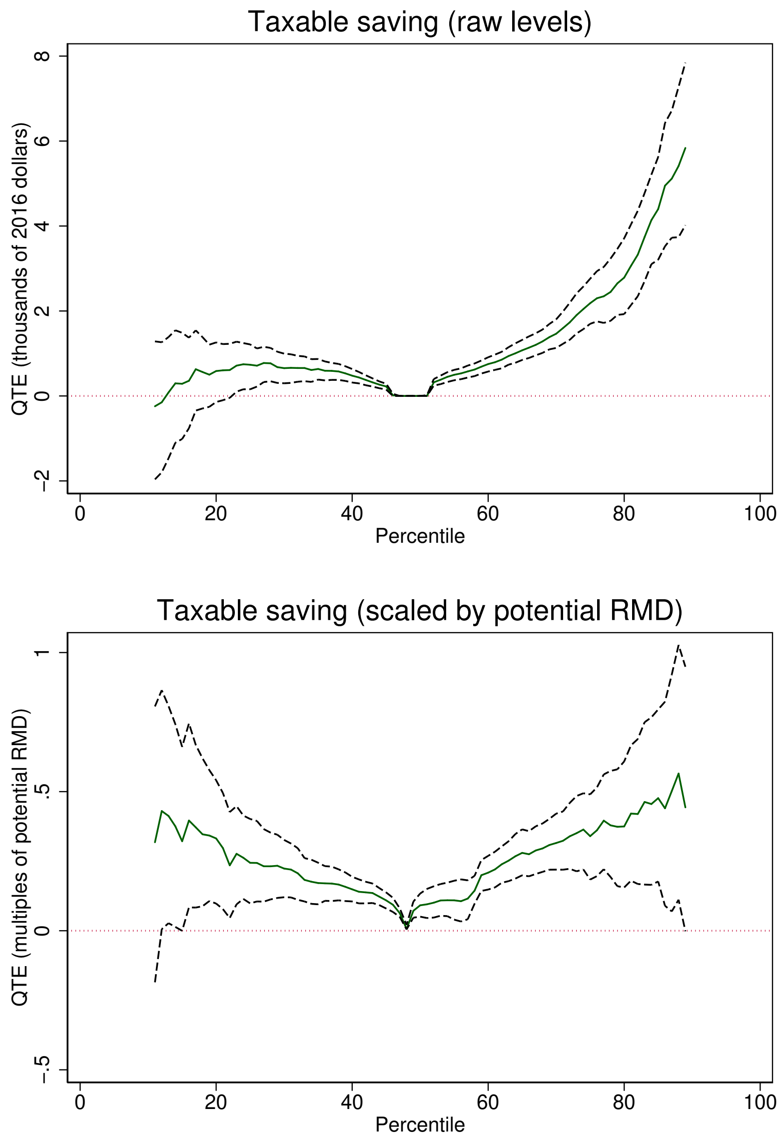
Notes: This figure plots the estimated coefficients and standard errors of the effect of RMD treatment at age 70 on retirement distributions, separately by event time. Retirement distributions are (net-of-tax) total IRA distributions of the reference individual and his or her spouse, as measured on Form 1099-R, plus taxable pensions as reported on Form 1040. In the top panel, the dependent variable is retirement distributions in levels. In the bottom panel, the dependent variable is total retirement distributions divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). The specification in the bottom panel weighted by the potential RMD, censored at $100,000. This figure uses the optimal CCT bandwidth, estimated separately by event time. See Appendix Tables A1 and A3 for the values of the coefficients and standard errors at each point. All dollar amounts are adjusted for inflation and expressed in 2016 dollars.

Figure 3: Reduced form: R.D. estimates of RMD status on taxable saving near RMD treatment



Notes: Each panel of this figure plots the average value of taxable savings as a function of the running variable, x (age in days relative to 70.5). Those to the right of 0 are required to take an RMD for age 70; those to the left are not. Saving is calculated using changes in mortgage interest, taxable interest income, and taxable dividend income aggregated between time $\tau - 1$ and $\tau + 1$, where τ is the year the individual turns 70; see text for details. In the upper-right panel, the dependent variable is taxable saving in levels (in thousands of dollars). In the remaining panels, taxable saving is Winsorized at the level indicated. Bin size is selected using the `rdplot` command in STATA. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. All dollar amounts are adjusted for inflation and expressed in 2016 dollars.

Figure 4: Reduced form: Quantile treatment effects of RMD treatment on taxable saving



Notes: This figure presents quantile treatment estimates for the effect of the RMD treatment on taxable saving. Saving is calculated using changes in mortgage interest, taxable interest income, and taxable dividend income aggregated between time $\tau - 1$ and $\tau + 1$, where τ is the year the individual turns 70; see text for details. In the upper panel, the dependent variable is taxable saving in thousands of dollars. In the bottom panel, the dependent variable is scaled by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). The specification in the bottom panel weighted by the potential RMD, censored at $100,000. The QTE is estimated using the method of Frandsen, Frolich, and Melly (2012). The dotted lines show the bounds of a 95 percent confidence interval, calculated via block bootstrap at the level of the running variable. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. All dollar amounts are adjusted for inflation and expressed in 2016 dollars. Percentiles above 90 and below 10 are not shown.

Table 3: Reduced form: Effect of RMD status at age 70 on taxable saving near RMD treatment

	Saving in levels (000s)			Scaled saving		
	Bandwidths			**Bandwidths**		
	200 percent	Baseline	50 percent	200 percent	Baseline	50 percent
	(1)	(2)	(3)	(4)	(5)	(6)
Unwinsorized	-13.185	**-3.578**	22.994	-2.491	**0.531**	5.072*
	(10.173)	**(11.553)**	(14.863)	(1.810)	**(2.102)**	(2.896)
	[92.76 days]	**[46.38 days]**	[23.19 days]	[86.09 days]	**[43.05 days]**	[21.52 days]
Winsorized: 1 percent	1.466**	**0.837**	-0.422	0.259	**-0.014**	-0.182
	(0.740)	**(1.070)**	(1.541)	(0.426)	**(0.609)**	(0.826)
	[145.84 days]	**[72.92 days]**	[36.46 days]	[103.32 days]	**[51.66 days]**	[25.83 days]
Winsorized: 5 percent	1.510***	**1.481****	0.492	0.226	**0.092**	-0.187
	(0.439)	**(0.652)**	(0.926)	(0.167)	**(0.240)**	(0.320)
	[133.05 days]	**[66.53 days]**	[33.26 days]	[137.59 days]	**[68.79 days]**	[34.40 days]
Winsorized: 10 percent	1.429***	**1.452*****	0.993	0.214*	**0.124**	-0.222
	(0.311)	**(0.432)**	(0.610)	(0.129)	**(0.186)**	(0.258)
	[146.98 days]	**[73.49 days]**	[36.74 days]	[112.62 days]	**[56.31 days]**	[28.15 days]

Notes: This table reports regression discontinuity estimates of being required to take an RMD for the age-70 year (year τ) on taxable saving aggregated between $\tau - 1$ and τ, which is related to the change in assets from $\tau - 1$ to $\tau + 1$. Each cell refers to a separate regression. In columns 1-3, the dependent variable is taxable saving in thousands of dollars. In columns 4-6, the dependent variable is taxable saving divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). Specifications in columns 4-6 are weighted by the potential RMD, censored at $100,000. In rows 2, 3, and 4, the dependent variable is Winsorized from above and below at the 1st, 5th, and 10th percentiles, respectively. The Winsorization is performed separately for each value of the running variable. In columns 2 and 5, the bandwidth is selected using the method of Calonico, Cattaneo, and Titiunik (2014). In columns 1 and 4, the bandwidth is 200 percent of the baseline bandwidth; in columns 3 and 6, it is 50 percent of the baseline bandwidths. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors allow arbitrary correlation of errors within individuals with the same value of the running variable. All dollar amounts are adjusted for inflation and expressed in 2016 dollars. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

Table 3) below the 60th percentiles, and mostly larger than $1,452 above the 60th percentile. The QTEs for scaled saving are broadly similar, though more uniformly positive and significant. The increase in precision is quite evident in this specification. In particular, while the estimated mean effects for scaled saving were insignificantly different from zero, the QTE for scaled saving is positive and significant above the 10th and below the 90th percentiles.

As an additional strategy to reduce noise, I apply the inverse hyperbolic sine transformation as discussed in Section 4.3. I present these results in Figure 5 and Table 4. In Figure 5, the discontinuity in transformed saving is clear for savings in levels (top panel) and is much less apparent for scaled saving (bottom panel). The top row of Table 4 reports the coefficient estimate, with analytical standard errors. Columns 1-3 report results for saving in levels. The coefficient is positive and highly significant at all considered bandwidths. Because the magnitude of these estimates is difficult to interpret, I also report the median implied effect, in thousands of dollars. This median implied effect for the baseline bandwidth is approximately $1,170, similar to the mean effect found in the 10-percent Winsorized sample in Table 3. Columns 4-6 repeat the same procedure for scaled saving. As was the case in Table 3, this specification is estimated somewhat less precisely. While the magnitude of the median implied effect (which is scaled in multiples of the potential RMD) is quite close to the result from Table 3, the confidence interval is wide.

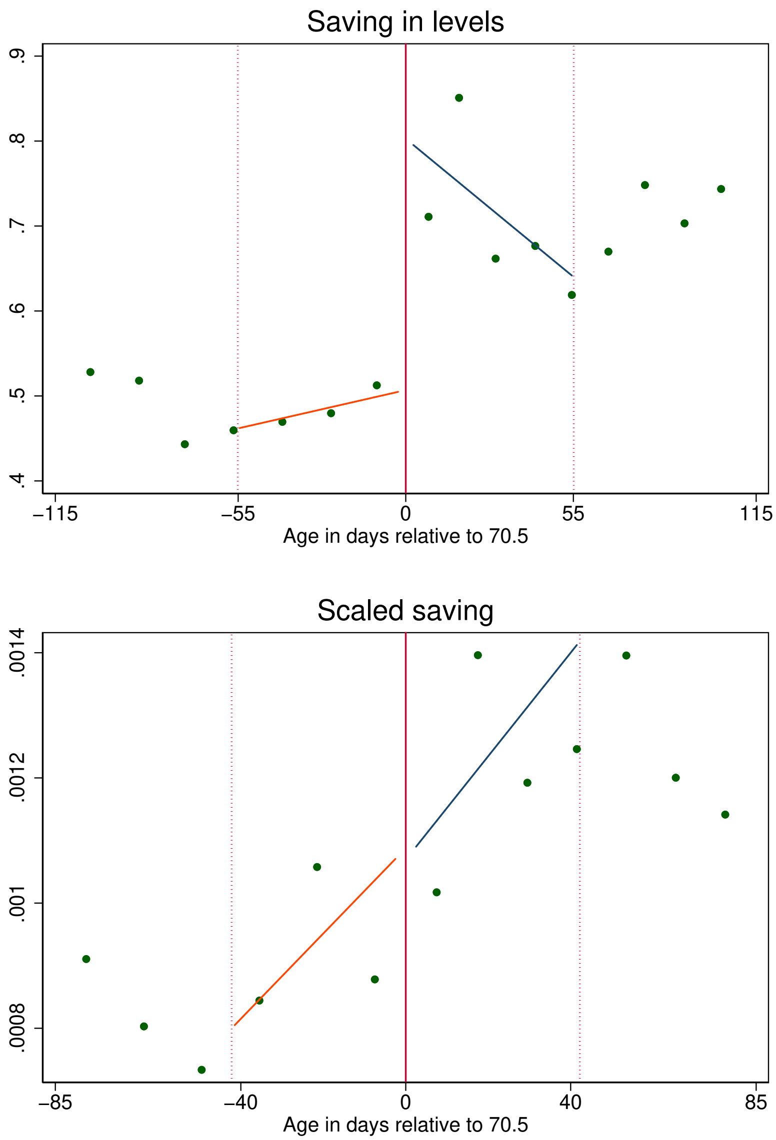
Finally, I use the IHS specification to explore the dynamic effect of the RMD treatment on saving. To simplify computation, I hold θ fixed at its optimal value as found above, and I ignore uncertainty in θ when computing confidence intervals. These results are plotted in Figure 6. The figure shows that the saving effect is significant only at event time 0, corresponding to the change in assets between τ and $\tau + 1$. This would be consistent with a contemporaneous saving response to induced distributions if such induced distributions were primarily taken near the end of year τ. Additionally, the lack of an effect prior to treatment (especially at event times -3 and -2) is reassuring, suggesting that whatever threats to identification exist (as discussed in Section 7) appear to not be affecting saving.

6 Estimating crowd-out of taxable dissaving by RMDs

6.1 Regression-based approach

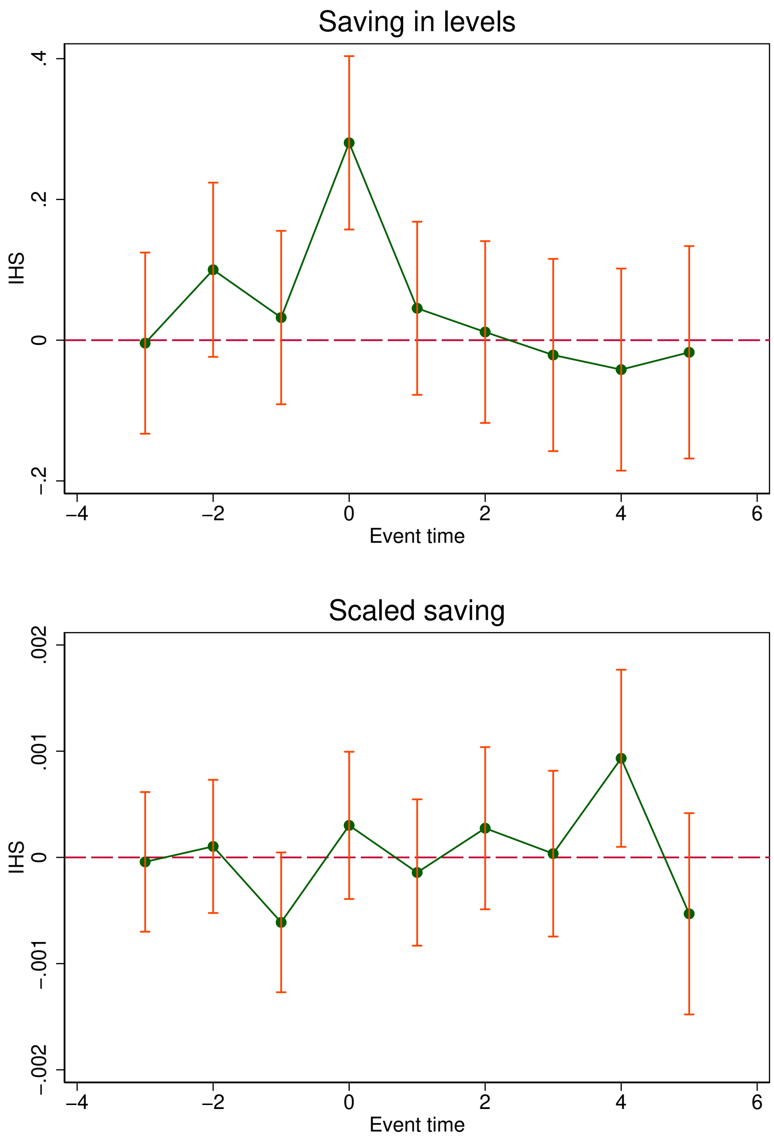
The previous section reported evidence that individuals who are just old enough to be subject to RMDs at age 70 have higher (1) retirement distributions at age 70 and (2) taxable saving at age 70 than those who are slightly younger. In this section, we convert these two estimates into an estimate of crowd-out. In a typical setting of retirement saving, crowd-out would represent the extent to which an exogenous increase in retirement savings causes a reduction in saving in other (usually taxable) accounts. In the setting of the paper, the situation is

Figure 5: Reduced form: R.D. estimates of RMD status on taxable saving near RMD treatment: inverse hyperbolic sine transformation



Notes: This figure plots the average value of the inverse hyperbolic sine of taxable savings as a function of the running variable, x (age in days relative to 70.5). Those to the right of 0 are required to take an RMD for age 70; those to the left are not. Prior to estimation, taxable saving (y) is transformed into g according to $g(y;\theta) = \frac{1}{\theta} ln\Big(\theta \frac{y}{1000} + \big(\theta^2 (\frac{y}{1000})^2 + 1\big)\Big)$. The parameter θ is chosen by a maximum likelihood estimation procedure discussed in Section 4.3. Saving is calculated using changes in mortgage interest, taxable interest income, and taxable dividend income aggregated between time $\tau - 1$ and $\tau + 1$, where τ is the year the individual turns 70; see text for details. In the upper panel, the dependent variable is taxable saving in thousands of dollars. In the bottom panel, the dependent variable is scaled by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). This specification is weighted by the potential RMD, censored at \$100,000. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped.

Figure 6: R.D. estimates of RMD status on taxable saving over time: inverse hyperbolic sine transformation



Notes: This figure plots the estimated coefficients and standard errors of the effect of RMD treatment at age 70 on taxable saving, separately by event time. Prior to estimation, taxable saving (y) is transformed into g according to $g(y;\theta) = \frac{1}{\theta} ln\Big(\theta \frac{y}{1000} + \big(\theta^2 (\frac{y}{1000})^2 + 1\big)\Big)$. The parameter θ is chosen by a maximum likelihood estimation procedure discussed in Section 4.3. In the top panel, the y is taxable saving in levels. In the bottom panel, y is taxable saving divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). The specification in the bottom panel weighted by the potential RMD, censored at $100,000. This figure uses the optimal CCT bandwidth from Table 3. See Appendix Tables A1 and A3 for the values of the coefficients and standard errors at each point. All dollar amounts are adjusted for inflation and expressed in 2016 dollars.

Table 4: Reduced form: Effect of RMD status at age 70 on taxable saving at age 70, inverse hyperbolic sine transformation

	Saving in levels			Scaled saving		
	Bandwidths			**Bandwidths**		
	200 percent (1)	Baseline (2)	50 percent (3)	200 percent (4)	Baseline (5)	50 percent (6)
Coefficient	0.269***	**0.296*** **	0.169	0.00010	**-0.00002**	-0.00020
	(0.050)	**(0.073)**	(0.113)	(0.00027)	**(0.00040)**	(0.00064)
Median effect (000s)	1.074**	**1.181** **	0.674	0.103	**-0.018**	-0.203
	[0.670, 1.563]	**[0.638, 1.810]**	[-0.151, 1.383]	[-0.359, 0.657]	**[-0.608, 0.792]**	[-0.959, 1.133]
Optimal θ		0.097			2.530	
		[0.096, 0.098]			[2.088, 237.662]	

Notes: This table reports regression discontinuity estimates of being required to take an RMD for the age-70 year (year τ) on taxable saving aggregated between $\tau - 1$ and τ, which is related to the change in assets from $\tau - 1$ to $\tau + 1$. Prior to estimation, the dependent variable y is transformed into g according to $g(y;\theta) = \frac{1}{\theta} ln\Big(\theta \frac{y}{1000} + \big(\theta^2 (\frac{y}{1000})^2 + 1\big)\Big)$. The parameter θ is chosen by a maximum likelihood estimation procedure discussed in Section 4.3. The top row reports the coefficient and analytical standard error, which has no ready interpretation. The second row reports the median effect (in terms of y) across all individuals implied by the coefficient. The final row reports the optimal θ. Confidence intervals for the final two rows are computed via block bootstrap over the running variable. In columns 1-3, y is taxable saving in dollars. In columns 4-6, the y is taxable saving divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). Specifications are *not* weighted. Columns 2 and 5 use the optimal bandwidth found in the analogous columns of Table 3. In columns 1 and 4, the bandwidth is 200 percent of the baseline bandwidth; in columns 3 and 6, it is 50 percent of the baseline bandwidths. The optimal θ is determined using the baseline bandwidth, and is held fixed for the other bandwidths. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. In row 1: * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. In other rows: ** $p < 0.05$

the opposite: the quasi-experimental variation causes an increase in retirement *distributions*. Thus, we are interested in the extent to which the increase in retirement distributions caused by RMDs leads to taxable saving that is less negative – in other words, the crowd-out of taxable dissaving by retirement distributions.

This parameter can be estimated via a Fuzzy RD, similar to (5). However, on the right hand side, D_i is replaced by retirement distributions (y_i^R), and the regression is estimated using instrumental variables, where y_i^R is instrumented by D_i. This mechanically recovers the "reduced form" effect of D_i on saving, divided by the "first stage" effect of D_i on retirement distributions.[41]

The basic fuzzy RD estimates are presented in Table 5. In all rows, y_i^R represents retirement saving, in levels (columns 1-3) or scaled by the potential RMD (columns 4-6). In a manner analogous to Table 3, I vary the Winsorization level of taxable saving, from $q = 0.01$ to $q = 0.05$ and 0.1. Using the baseline bandwidth (column 2), the specifications in levels find a crowd-out parameter of approximately 0.4, with statistical significance for $q = 0.05$ and $q = 0.1$. While these estimates are estimated with sufficient precision to rule out zero crowdout (in rows 2 and 3), the confidence intervals remain fairly wide. In particular, the most precise baseline estimate (row 3, column 2) yields a confidence interval of $[0.12, 0.73]$. The coefficients are little effected when the bandwidth is increased, while the precision improves. With a bandwidth that is 200 percent of the optimal CCT bandwidth, the confidence interval shrinks to $[0.19, 0.62]$. The point estimates are not robust to 50 percent of the baseline bandwidth; however, the visual evidence in Figure 3 (lower right panel) suggests that a longer bandwidth is indeed more appropriate. The specifications used scaled saving (in columns 4-6) also yield qualitatively similar coefficients, but with large standard errors. In sum, the point estimates of the specification in levels suggest modest crowd-out of taxable dissaving by induced retirement distributions; one can rule out crowdout of both 0 and 1 at the 1 percent level in the baseline specification.

6.2 Threshold-based approach

While the estimates reported in Table 5 are fairly consistent and economically plausible, the Winsorization that is necessary to obtain informative estimates is somewhat unsatisfying. Additionally, the estimates are somewhat imprecise. For this reason, I consider an additional approach using thresholds, which refines a method used by Chetty, et al (2014). This approach requires some stronger assumptions, but does not require an ad-hoc decision to be made regarding Winsorization. It will also generally be estimated with more precision. This approach is motivated by the fact that we can fairly precisely estimate the effect of RMD treatment, D_i, on the probability that saving (or scaled saving) exceeds some threshold y^*. We can then compare this observed quantity to what we would expect to see if the crowd-out parameter were one. Under the

[41] These estimates are not literally equal to the ratio of the two coefficients presented in Tables 2 and 3, since the bandwidths used in the estimation of those two coefficients were not the same.

Table 5: Fuzzy R.D. estimates of crowd-out of taxable dissaving by retirement distributions

	Saving in levels			Scaled saving		
		Bandwidths			**Bandwidths**	
	200 percent	Baseline	50 percent	200 percent	Baseline	50 percent
Winsorized: 1 percent	0.334	**0.183**	-0.452	0.530	**-0.018**	-0.283
	(0.255)	**(0.359)**	(0.532)	(0.688)	**(0.963)**	(1.356)
	[109.10 days]	**[54.55 days]**	[27.27 days]	[108.83 days]	**[54.41 days]**	[27.21 days]
Winsorized: 5 percent	0.429***	**0.418****	0.042	0.251	**0.030**	-0.405
	(0.139)	**(0.206)**	(0.283)	(0.329)	**(0.444)**	(0.638)
	[117.23 days]	**[58.61 days]**	[29.31 days]	[100.45 days]	**[50.22 days]**	[25.11 days]
Winsorized: 10 percent	0.405***	**0.422*****	0.067	0.351	**0.171**	-0.386
	(0.109)	**(0.155)**	(0.214)	(0.223)	**(0.308)**	(0.438)
	[106.10 days]	**[53.05 days]**	[26.53 days]	[106.99 days]	**[53.50 days]**	[26.75 days]

Notes: This table reports estimated crowdout of taxable dissaving by retirement distributions induced by RMDs. These estimates are constructed using a fuzzy RD, where the dependent variable is taxable saving aggregated between $\tau - 1$ and τ and the endogenous regressor is retirement distributions at time τ. Each cell refers to a separate regression. In columns 1-3, the dependent variable is taxable saving in thousands of dollars. In columns 4-6, the dependent variable is taxable saving divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). Specifications in columns 4-6 are weighted by the potential RMD, censored at $100,000. The dependent variable is Winsorized from above and below at the 1st, 5th, and 10th percentiles in rows 1, 2, and 3, respectively. The Winsorization is performed separately for each value of the running variable. In columns 2 and 5, the bandwidth is selected using the method of Calonico, Cattaneo, and Titiunik (2014). In columns 1 and 4, the bandwidth is 200 percent of the baseline bandwidth; in columns 3 and 6, it is 50 percent of the baseline bandwidths. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors allow arbitrary correlation of errors within individuals with the same value of the running variable. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

(strong) assumption that the saving response is binary – either individuals do not change their taxable saving, or they increase their taxable saving exactly enough to offset the increased retirement distributions – then the ratio of the observed quantity to the theoretical quantity represents the share of individuals who respond.

More formally, let y_i denote the outcome of interest (i.e., saving or scaled saving). Let $G_i(y^*)$ denote an indicator for $y_i \geq y^*$ for some arbitrary threshold y^*. Using the usual RD strategy, we can recover empirically the effect of RMD treatment D_i on $G_i(y^*)$.

We can recover the theoretically predicted amount as follows. Consider a potential outcomes framework (Rubin (1974)). Let $y_i(D_i)$ denote potential outcomes for saving (or scaled saving). If the crowdout parameter were one, $y_i(D_i)$ could be written as $y_i(0) + T_i D_i$. In this expression, T_i is the first-stage treatment effect: the extent to which retirement distributions increase when D_i flips from 0 to 1. We observe $y_i = y_i(1)D_i + y_i(0)(1 - D_i)$. The effect of D_i on the fraction of individuals with $y_i \geq y^*$ is equal to the share of individuals who have $y_i^0 \in (y^* - T_i, y^*)$. These are the individuals who have $y_i < y^*$ when $D_i = 0$ but $y_i \geq y^*$ when $D_i = 1$. In the setting considered by Chetty (2014), it was natural to assume that T_i and $y_i(0)$ were independent. Under that independence assumption (and making an additional approximation that the density of y_i^0 is locally constant), the predicted discontinuity under full crowdout would be equal to $E(T_i) f_{y^0}(y^*)$. However, this independence assumption is less plausible in the present setting. Consider the simplest case when y_i and T_i are measured in levels. T_i will tend to be larger when the RMD is larger, which will tend to be correlated with the amount of taxable wealth held by individual i, which is unlikely to be independent of taxable saving. Even when y_i and T_i are scaled relative to the RMD, T_i will be larger when RMDs are more binding, which theory suggests to be the case when taxable assets are larger. Appendix Table A5 confirms this hypothesis. This table reports the first stage coefficients (for scaled retirement distributions) separately by six bins of taxable assets, as measured at $\tau - 1$.[42] While the first stage effects are positive and significant in all bins, the magnitude of the first stage effect is monotonically increasing in baseline assets.

I relax the assumption of independence. However, it remains necessary for tractability to impose more structure to the nature of the dependency. I assume that the first stage effect T_i is constant within the asset bin a_i, as motivated by theory. Under these assumptions, the predicted increase can be calculated using iterated expectations as follows, where the summation is over the 6 asset bins:

$$p \times \sum_a \Pr\left[y_i(0) \in \left(y^* - T(a), y^*\right) | a_i = a\right] \Pr(a_i = a) \tag{6}$$

$T(a)$ is estimated via separate first-stage regressions by asset bin (and are equal to the values reported in Appendix Table A5). Given $T(a)$, I estimate

[42] For the purpose of this table, taxable assets includes only bonds and stocks, not reduced by mortgage indebtedness.

$\Pr\left[y_i(0) \in \left(y^* - T(a), y^*\right)\right]$ by computing the empirical analogue at each value of the running variable, and taking the predicted value just to the left side of $x = 0$.[43] I estimate $\Pr(a_i = a)$ in a similar fashion.[44] Crowdout is computed as the p that is needed to equalize the predicted discontinuity with the observed discontinuity.

I present the estimates of crowdout derived from this method in Table 6. I present estimates for nine different values of y^*, which correspond to the 10th, 20th, ..., 90th percentiles of the distribution of scaled saving (for untreated individuals). Under the maintained assumptions above – in particular, that response is binary – these estimates can be interpreted as the estimated amount of crowdout for individuals with $y_i(0)$ near y^*.[45] The crowdout estimates are significant and positive for the 2nd through 7th decile breaks, ranging from 0.16 at the 5th decile break (the median, or approximately zero saving) to 0.90 at the 2nd decile break (substantial dissaving). The effects at the 1st, 8th, and 9th decile breaks are sufficiently imprecise as to be essentially uninformative. These results broadly corroborate the crowdout estimates in Table 5 that crowdout is positive in the neighborhood of 0.4.

7 Threats to Identification

In order for Equation (4) to recover the causal effect of D_i, it must be the case that $E(y|x, D = 0)$ and $E(y|x, D = 1)$ are each be continuous in x at $x = 0$. There are two reasons why this assumption could be violated. First, being born just before June 30 (rather than just after) could affect outcomes y through some channel other than RMDs. In particular, such individuals might be discontinuously eligible to start school one year earlier than their peers, which would cause "relative age" (i.e., age relative to classmates) to be discontinuously different. There is a substantial recent literature showing that relative age can have substantial contemporaneous and long-lasting effects.[46] However, at least in present day, July 1 is not typically the date cutoff for entry into Kintergarten;

[43] That is, I regress the empirical analogue on a constant and x, using triangular kernel weights and using observations only to the left of 0. I opt not to make the approximation that the density is locally constant, as that is demonstrably false near $y_i(0) = 0$.

[44] There remains a decision to make regarding the bandwidths to use. In general, the CCT bandwidths are individually MSE-optimal for each RD specification, meaning that they optimally trade off bias and variance for that specification. It is not obvious that those CCT bandwidths will be optimal for the entire object in (6). In particular, I conjecture that the combination of multiple RD coefficients will tend to increase noise, but not necessarily bias. This suggests that one should use a larger bandwidth than would be individually optimal for each regression. For this reason, for each RD specification in Table 6, I use a bandwidth that is 50 percent larger than the optimal bandwidth for estimating the mean effect on scaled saving.

[45] Standard errors are computed via bootstrap.

[46] For example, see Dhuey & Lipscomb (2008) for the effect of relative age on high school leadership activities, Dhuey, et al (2017) for the effect of relative age on cognitive development, Musch & Grondin (2001) for a review of the effect of relative age on sports performance, and Layton, et al (2018) for the effect of relative age on ADHD diagnoses.

Table 6: Estimates of crowd-out of taxable dissaving by retirement distributions: threshold approach

	Scaled saving threshold (1)	Predicted discontinuity (2)	Observed discontinuity (3)	Estimated crowdout (4)
Decile break 1:	-27.554	0.0017	0.0013	0.751
			(0.0011)	[-0.520, 2.092]
Decile break 2:	-9.152	0.0069	0.0062***	0.895**
			(0.0019)	[0.332, 1.466]
Decile break 3:	-3.397	0.0159	0.0078***	0.491**
			(0.0022)	[0.200, 0.761]
Decile break 4:	-0.913	0.0288	0.0086***	0.299**
			(0.0023)	[0.124, 0.479]
Decile break 5:	0.035	0.0661	0.0106***	0.160**
			(0.0022)	[0.088, 0.228]
Decile break 6:	1.223	0.0323	0.0089***	0.274**
			(0.0025)	[0.102, 0.438]
Decile break 7:	3.737	0.0166	0.0059***	0.357**
			(0.0020)	[0.092, 0.612]
Decile break 8:	9.626	0.0070	0.0013	0.180
			(0.0020)	[-0.383, 0.756]
Decile break 9:	28.745	0.0016	-0.0016	-1.013
			(0.0011)	[-2.277, 0.463]

Notes: This table reports estimated crowdout using the threshold approach described in Section 6.2. The dependent variable is taxable saving divided by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). All specifications are weighted by the potential RMD, censored at $100,000. The first column indicates the threshold y^* under consideration. The second column indicates the discontinuity in the share of individuals above y^* would observe at $x = 0$ if crowdout were one. Column 3 reports the actual observed discontinuity using the RD approach. Column 4 reports the ratio of these two columns, which represents the crowdout ratio under assumptions discussed in the text. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. Standard errors in Column 3 are analytical. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. Confidence intervals in column 4 are estimated via block bootstrap by the running variable. ** $p < 0.05$

September 1 (or thereabouts) is a much more common date.[47] To the best of my knowledge, date cutoffs (if they existed) for entry into Kintergarten have not been tabulated for the time period when my sample would have been school-age.

Second, the composition of individuals (in terms of unobservables that may affect y) may be different on each side of the cutoff. This could arise due to manipulation of the running variable, either by mothers or medical professionals changing the actual timing of birth, or by administrators adjusting the recorded date of birth after the fact. Suppose without loss of generality that such manipulation is toward late June birthdays. Those with birthdays in late June would then be a mix of those with an "unmanipulated" late June birthday and those with a "manipulated" late June birthday, while those with an early July birthday would consist only of the "unmanipulated." Even though this manipulation would obviously not have been in response to the RMD treatment,[48] having a manipulated birthday might be correlated with other unobservables that affect y indirectly.

Another potential cause of composition bias comes from the fact that I condition the sample on holding an IRA prior to treatment (since potential treatment is a function of IRA balance at the end of the age-69 year). It is conceivable that forward-looking agents adjust their behavior prior to age 70 in response to the RMD rules that come into effect at age 70.5. For instance, a member of the treatment group (subject to an RMD at age 70) might find saving in an IRA to be less attractive than a member of the control group, given the slightly nearer-in-time RMD requirements. This could cause the treatment group to include slightly fewer marginal IRA savers; such marginal savers could have different unobservables than the rest of the population, potentially leading to bias. This concern may become more severe as the conditioning year moves closer to the treatment year. On the other hand, moving the conditioning year further away from treatment makes the sample less well-targeted. While the baseline specification conditions on holding an IRA at age 69, I show in Section 8 that results are very similar when conditioning instead on holding an IRA at age 68.

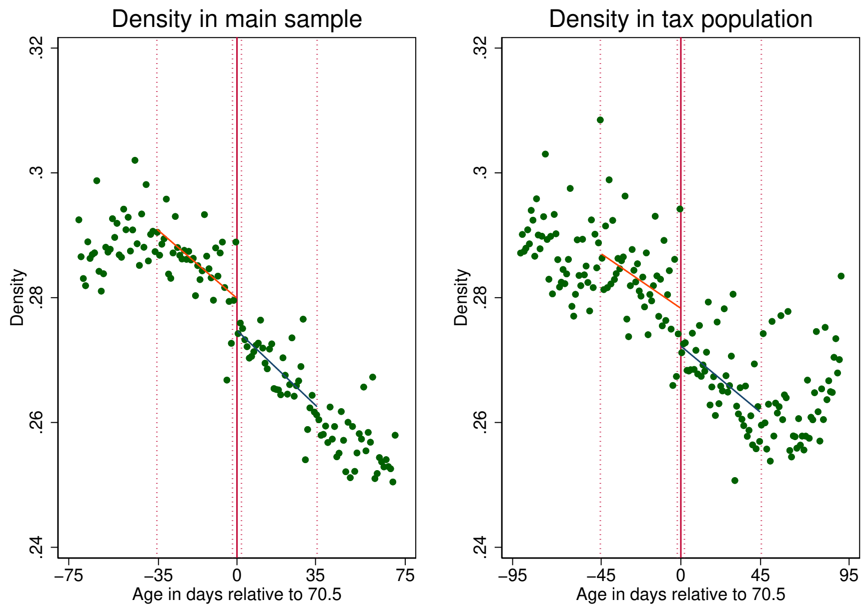
More generally, I explore these identification assumptions in two ways. First, as is standard, I analyze the density of individuals in my sample as a function of the running variable (McCrary (2008)). This will in general uncover any manipulation of the running variable, which would tend to show itself as excess mass on one side of the discontinuity. A composition bias caused by the sample conditioning would cause a similar effect. Second, I will test for discontinuities in other pre-treatment observable factors at the cutoff; if the cutoff has an effect on y through some channel other than RMDs, then the cutoff would presumably have an effect on these other pre-determined factors as well.

To begin, the left panel of Figure 7 plots the density of individuals in the sample (multiplied by 100), aggregated across experiment years τ. The figure shows that there does appear to be a small discontinuity in the density at

[47]Kintergarten entry dates have been tabulated by the National Center for Education Statistics, and are available at `https://nces.ed.gov/programs/statereform/tab5_3.asp`.

[48]One simple reason why this is the case is that RMD requirements were not put in place until after all of the individuals in my sample were born.

Figure 7: Density of 69-year olds in the sample and in the population as a function of the running variable



Notes: The left panel of this figure plots the density of individuals (mutiplied by 100) in the main sample, with bin size equal to a single day. The right panel plots the analogous density for the broader tax population of 69-year olds (aggregated over the same years as the main sample). The tax population includes all individuals who appear on a tax return or who receive an information return in a given year; see text for further discussion. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Additionally, the dotted lines closest to zero indicate $|x| = 1.5$; in the baseline specifications in this paper, observations with $|x| \leq 1.5$ are dropped.

the cutoff: there is slightly more mass on the early July side (left of zero) relative to the late June side. This is potentially concerning, since it could reflect manipulation of the running variable or composition bias due to sample conditioning. The first three columns of the first row of Table 7 report the magnitude of this jump. The second column corresponds to the specification illustrated in the left panel of Figure 7 – i.e., the result from estimating (5) by OLS and computing heteroskedasticity-robust standard errors; columns 1 and 3 use a different bandwidth. The baseline specification in column 2 finds that the discontinuity is 0.0052 percentage points, which is significant at the 5 percent level. Given that the baseline density is about 0.28 percent, this effect represents an increase of about 1.9 percent.

To address this threat to identification, I first investigate whether the same discontinuity appears in the general population of 69-year olds. If so, that would suggest that the discontinuity is not the result of behavior of IRA-holders (or

Table 7: McCrary Tests

	Full sample			Donut hole		
	Bandwidths			**Bandwidths**		
	50 percent	Baseline	200 percent	50 percent	Baseline	200 percent
	(1)	(2)	(3)	(4)	(5)	(6)
Main sample density	-0.0042	-0.0052**	-0.0071***	0.0023	-0.0031	-0.0055***
	(0.0039)	(0.0026)	(0.0017)	(0.0038)	(0.0031)	(0.0020)
	[17.86 days]	[35.71 days]	[71.42 days]	[15.66 days]	[31.33 days]	[62.66 days]
Tax population density	-0.0071	-0.0060*	-0.0088***	-0.0021	-0.0042	-0.0076***
	(0.0051)	(0.0032)	(0.0020)	(0.0051)	(0.0031)	(0.0020)
	[22.73 days]	[45.46 days]	[90.93 days]	[21.21 days]	[42.43 days]	[84.86 days]
Fraction in main sample	0.0038	0.0012	-0.0005	0.0025	-0.0000	-0.0013
	(0.0035)	(0.0024)	(0.0016)	(0.0043)	(0.0026)	(0.0016)
	[22.12 days]	[44.24 days]	[88.47 days]	[22.38 days]	[44.76 days]	[89.51 days]

Notes: This table reports regression discontinuity estimates of being required to take an RMD for the age-70 year (year τ) on the density of individuals on each side of the threshold. In columns 4-6, I follow the main specifications and drop individuals with birthdays between June 29 and July 2, inclusive. In columns 1-3, the entire sample is used. Each cell refers to a separate regression. In the first row, the dependent variable is the density of individuals in the main sample. In the second row, the dependent variable is the density of individuals in the broader tax population of 69-year-olds over the same time period. The tax population includes all those individuals who, in a given year, appear on a tax return (as primary, secondary, or dependent) or who receive an information return such as Form W2. In the third row, the dependent variable is the ratio between the number of observations in the main sample and the number of observations in the tax population. In columns 2 and 5, the bandwidth is selected using the method of Calonico, Cattaneo, and Titiunik (2014). In columns 1 and 4, the bandwidth is 200 percent of the baseline bandwidth; in columns 3 and 6, it is 50 percent of the baseline bandwidths. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors are robust to heteroskedasticity.* $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$.

potential IRA-holders) prior to age 70 in anticipation of RMD rules. The challenge here is to define the relevant population. I follow Cilke (2014) and others by using the "tax population": those who interact with the tax system in some form during the course of the year. This condition is satisfied in a given year if an individual (1) files a tax return, either as primary or secondary, (2) appears as a dependent on a tax form, or (3) receives any information return, such as a Form W2 (for wage earnings), a Form 1099-R (for distributions from pensions and similar accounts), or a Form 1099-SSA (for Social Security income). Almost all individuals either file a tax return or appear on an information return, so this measure approximates the full population of individuals. To be consistent with the construction of the main sample, I restrict myself to 69-year olds between 2000 and 2013, excluding 2008 (i.e., those individuals that would be 70 between 2001 and 2014, excluding 2009).

The resulting density among the 69-year-old tax population is plotted in the right panel of Figure 7. This figure shows that the discontinuity in density is also apparent in this much broader sample, with in fact a slightly larger magnitude. In particular, as reported in the second row of Table 7, this population experiences a discontinuity in density of -0.0060 percentage points using the baseline bandwidth (compared to -0.0052 percentage points in the main sample), which is significant at the 10 percent level. Thus, the discontinuous change in density in the main sample appears not to be caused by differential behavior in anticipation of RMD rules, which is reassuring. Confirming this observation, I redefine the dependent variable in the third row as the ratio of the main sample population (at a given x) to the tax population. This discontinuity is small and insignificant.[49]

Nevertheless, the discontinuity in the sample population is still concerning, even if the same discontinuity is present in the fuller population as well. Such discontinuities could be consistent with classical manipulation of the running variable. The most plausible such manipulation could involve manipulating date of birth by one or two days. In columns 4-6 of Table 7, I repeat the same plots, but drop birth dates corresponding to June 29 through July 2. The magnitudes of the discontinuities in each population shrink toward zero and become statistically insignificant. For this reason, the main empirical specification excludes these observations, effectively adding a donut hole to the regression discontinuity design. Yet, even after removing the donut hole, there remains a slight excess mass on the left (untreated) side of the discontinuity. The most likely explanation for this small discontinuity is manipulation (e.g., timing of conception) on a slightly longer time scale. This is not fatal to the empirical design, since the manipulation (if it exists) is not directly related to the outcome of interest; however, this manipulation suggests that we should investigate carefully whether the treatment and control groups look similar according to characteristics that might more directly affect outcomes.

To do so, I estimate whether there are any discontinuities in pre-determined characteristics at the cutoff. Three relevant (observed) characteristics are pre-

[49] Appendix Figure A7 illustrates this specification graphically.

Table 8: Discontinuities in baseline covariates

	IRA value at 69	**Taxable assets**	**Non-retirement income**
	(1)	(2)	(3)
Coefficient	0.131**	0.051	-0.009
	(0.062)	(0.205)	(0.032)
Median effect (000s)	0.524	0.162	-0.031
	[-0.011, 0.915]	[-1.246, 1.939]	[-0.328, 0.230]
Optimal θ	0.062	0.031	0.124
	[0.055, 0.072]	[0.031, 0.032]	[0.115, 0.137]

Notes: This table reports regression discontinuity estimates of being required to take an RMD for the age-70 year (year τ) on pre-determined outcomes. The three outcomes are (1) IRA value, (2) taxable assets (not reduced by mortgage indebtedness), and (3) non-retirement income. All variables are person-specfiic averages over the three years prior to treatment. Non-retirement income is adjusted gross income less taxable pensions and taxable IRA distributions. Prior to estimation, the dependent variable y is transformed into g according to $g(y;\theta) = \frac{1}{\theta} ln\Big(\theta \frac{y}{1000} + \big(\theta^2 (\frac{y}{1000})^2 + 1\big)\Big)$. The parameter θ is chosen by a maximum likelihood estimation procedure discussed in Section 4. The top row reports the coefficient and analytical standard error, which has no ready interpretation. The second row reports the median effect (in terms of y) across all individuals implied by the coefficient. The final row reports the optimal θ. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. Confidence intervals for the final two rows are computed via block bootstrap over the running variable. In columns 1-3, y is taxable saving in dollars. I use a bandwidth of 45 days. See Appendix Tables A6 through A8 for alternative bandwidths. In row 1: * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. In other rows: ** $p < 0.05$

treatment values of (1) taxable assets, (2) income, and (3) IRA value. For each of these outcomes, I aggregate within each individual over the three years prior to treatment. Then, I define the dependent variable y_i to be the inverse hyperbolic sine of this 3-year average, as discussed in Section 4.3. I then estimate Equation (5) using that dependent variable. The results are presented in Table 8, and graphically in Figure 8. The estimated discontinuities are essentially zero for non-retirement income and taxable assets. The estimated effect is slightly positive for IRA value prior to treatment (indicating better outcomes just to the right of the discontinuity). However, the magnitude of this effect is modest: the median implied effect is only \$520 (relative to the mean amount of approximately \$165,000 as found in Table 1). Furthermore, the bootstrapped confidence interval (which accounts for uncertainty in the inverse hyperbolic sine parameter θ) includes zero. Thus, while individuals just to the left and just to the right of the age cutoff do not appear to be identical, the differences between them seem to be small in magnitude.[50]

[50] Additionally, as I have shown above, I estimate that there is no difference in taxable saving or retirement distributions in years $\tau - 3$ through $\tau - 1$. That is, the small amount of manipulation – if it exists – does not appear to be causing differences in the outcomes of interest prior to treatment.

Figure 8: Values of pre-determined covariates as a function of the running variable: Inverse hyperbolic sine transformation

Notes: Each panel of this figure plots the average value of the inverse hyperbolic sine of a predetermined otucome as a function of the running variable, x (age in days relative to 70.5). Those to the right of 0 are required to take an RMD for age 70; those to the left are not. Prior to estimation, the outcome (y) is transformed into g according to $g(y;\theta) = \frac{1}{\theta} ln\Big(\theta\frac{y}{1000} + \big(\theta^2(\frac{y}{1000})^2+1\big)\Big)$. The parameter θ is chosen by a maximum likelihood estimation procedure discussed in Section 4.3. In the upper-left panel, the outcome is average IRA value at ages 67-69. In the upper-right panel, the outcome is adjusted gross income less taxable pensions and taxable IRA distributions (averaged over ages 67-69). The outcome in the lower left panel is taxable assets, not reduced by mortgage debt (averaged over 67-69), which is calculated by capitalizing observed taxable dividends and taxable interest. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. All dollar amounts are adjusted for inflation and expressed in thousands of 2016 dollars.

8 Sensitivity Analyses

In this section, I consider the effect of three specification choices on the main estimates. First, one might be concerned that the sample construction – specifically, the restriction to those with positive IRA values at the end of $\tau-1$ – might cause the sample composition to change discontinuously at $x = 0$. This would occur, for instance, if forward-looking individuals who are treated at age 70 are more likely to close their accounts at age 69. The evidence in Appendix Figure A7 belays this concern, as the number of individuals in the sample as a share of the broader tax population is continuous in x through the threshold. Nevertheless, this concern would be further mitigated if the sample were conditioned on having a positive IRA value at age 68, assuming that the forward-looking behavior is most evident in the year prior to treatment. Second, as discussed in Section 7, there appears to be a slight excess mass just to the left of the threshold, both in the sample as well as in the broader tax filing population. This is potentially consistent with manipulation, especially for those with birthdays very close to June 30. To address this concern, the main specifications exclude those with birthdays on June 29, June 30, July 1, and July 2, creating a donut hole around $x = 0$. In this section, I consider the effect of eliminating that donut hole. Third, Calonico, Catteneo, and Titiunik (2014) have proposed an alternative estimator for treatment effects in RD settings, both for the point estimate and the confidence intervals. This alternative estimator is designed to incorporate additional curvature in the conditional mean function, as well as more rigorously account for the uncertainty in the functional form. In this section, I repeat the main estimates using that alternative estimator.

The results are presented in Table 9. For this purpose, I define the "main estimates" to be (1) the contemporaneous (year-τ) effect of RMD treatment on retirement distributions, (2) the effect of RMD treatment on taxable saving (aggregated between years $\tau-1$ and τ), using the 10-percent Winsorized sample, and (3) the crowdout estimate with using the 10-percent Winsorized sample. The first row reproduces the baseline estimates. The second row conditions on having a positive IRA value at the end of $\tau-2$ rather than the end of $\tau-1$. The third row eliminates the donut hole for birthdays June 29 through July 2. The fourth row uses the Calonico, Catteneo, and Titiunik (2014) estimator. The results in all six columns are quite consistent across the rows. The results for retirement distributions are within \$60 in levels and 4 percent of the potential RMD. The results for taxable saving in levels are within \$100; when scaled by the potential RMD, there is slightly more variability, but the differences are small relative to the standard error. Similarly, the crowdout estimates in levels are very close to each other. The Calonico, Catteneo, and Titiunik (2014) standard error is about 12 percent larger than the baseline standard error, which causes the estimate to be significant at only the 5 percent level rather than the 1 percent level. Crowdout in the scaled specification again exhibits slightly more variability, which is unsurprising given the lack of precision in this specification.

9 Conclusion

I estimate the extent to which an exogenous increase in retirement distributions caused by RMDs crowds out taxable dissaving. Using an RD strategy that exploits variation in exact date of birth, I find clear evidence that taxable saving increases in the years bracketing age 70 among those who are just old enough to be required to take an RMD for their age-70 year. In the baseline specification, the crowdout estimate is approximately 0.4. I corroborate this estimate using an alternative threshold-based approach, which yields similar (if slightly smaller)

Table 9: Sensitivity of main results to alternative specification choices

	Retirement Distributions		Taxable saving		Crowdout	
	In levels	Scaled	In levels	Scaled	In levels	Scaled
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline	3.446***	0.618***	1.452***	0.124	0.422***	0.171
	(0.141)	(0.035)	(0.432)	(0.186)	(0.155)	(0.308)
	[56.92 days]	[57.99 days]	[73.49 days]	[56.31 days]	[53.05 days]	[53.50 days]
Condition on $V_{68} > 0$	3.390***	0.651***	1.535***	0.294	0.465***	0.461
	(0.133)	(0.035)	(0.445)	(0.179)	(0.159)	(0.299)
	[57.46 days]	[61.35 days]	[74.20 days]	[78.88 days]	[54.91 days]	[68.07 days]
No donut hole at $x <= 1.5$	3.477***	0.625***	1.551***	0.222	0.461***	0.303
	(0.154)	(0.029)	(0.420)	(0.155)	(0.151)	(0.297)
	[61.52 days]	[59.16 days]	[73.67 days]	[79.18 days]	[55.69 days]	[56.89 days]
CCT(2014) method	3.456***	0.625***	1.468***	0.077	0.417**	0.091
	(0.183)	(0.045)	(0.486)	(0.218)	(0.186)	(0.362)
	[56.92 days]	[57.99 days]	[73.49 days]	[56.31 days]	[53.05 days]	[53.50 days]

Notes: This table reports sensitivity analyses to the main estimates. Columns 1 and 2 report the first stage: the effect of being subject to an RMD requirement at age 70 on retirement distributions in that year. Columns 3 and 4 report the reduced form: the effect of being subject to an RMD requirement at age 70 on taxable saving over the two years bracketing age 70. Columns 5 and 6 report the associated crowdout estimated via a fuzzy RD with that reduced form and first stage. The crowdout estimate is not mechanically equal to the reduced form divided by the first stage because of different bandwidths in each specification. Columns 1, 3, and 5 define retirement distributions and taxable saving in levels. Columns 2, 4, and 6 scale by the IRA RMD that would apply at age 70 if the individual were subject to RMDs (the "potential RMD"). Specifications in these columns are weighted by the potential RMD, censored at $100,000. The first row repeats the baseline estimates found in Tables 2, 3, and 5. The second row changes the sample to be conditioned on having a positive IRA value at age 68, rather than age 69. The third row adds back in observations with birthdays between June 29 and July 2 which are dropped in the baseline sample. The fourth row uses the estimator (and standard errors) proposed by Calonico, Cattaneo, and Titiunik (2014). In all rows, the optimal bandwidth is chosen using the method of Calonico, Cattaneo, and Titiunik (2014). * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

crowdout effects.

This crowdout parameter is important for two reasons. First, it informs whether RMDs are successful at achieving their policy goal of raising present value tax revenue; a higher degree of crowdout increases revenue gains by creating more taxable investment income (or reducing deductible mortgage interest). Second, the parameter sheds light on how retirees make decisions regarding the decumulation of their assets. Under a neoclassical model, RMDs do not affect consumption to first order. A crowdout parameter less than 1 suggests one of the following possibilities: (1) RMDs do not affect consumption, but some RMDs are deposited into (and kept in) non-interest-bearing accounts, leading me to underestimate true crowdout or (2) distributions induced by RMDs are perceived as an increase in wealth and therefore affect consumption non-trivially. Further research is needed to disentangle these two explanations.

References

[1] Benjamin, Daniel J. (2003). "Does 401(k) eligibility increase saving? Evidence from propensity score subclassification." *Journal of Public Economics* 87(5-6), 1259-1290.

[2] Bernheim, Douglas B. (2002). "Taxation and Saving." In *Handbook of Public Economics, Volume 3*, Auerbach, Alan J. and Feldstein, Martin (eds): Elsevier Science B.V., Amsterdam, 1173-1249.

[3] Beshears, John, Choi, James J., Laibson, David, Madrian, Brigitte C., Skimmyhorn, William L. (2017). "Borrowing to Save? The Impact of Automatic Enrollment on Debt." Unpublished Manuscript.

[4] Brown, Jeffrey R., Poterba, James, and Richardson, David P. (2017). "Do Required Minimum Distributions Matter? The Effect on the 2009 Holiday on Retirement Plan Distributions." *Journal of Public Economics* 151(C), 96-109.

[5] Burbidge, John B., Magee, Lonnie, and Robb, A. Leslie (1988). "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association* 83(401), 123-127.

[6] Calonico, Sebastian, Cattaneo, Matias D., Farrell, Max H., and Titiunik, Rocio (2017). "rdrobust: Software for regression-discontinuity designs." *The State Journal* 17(2), 372-404.

[7] Calonico, Sebastian, Cattaneo, Matias D., and Titiunik, Rocio (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6), 2295-2326.

[8] Chetty, Raj, Friedman, John N., Leth-Petersen, Soren, Nielsen, Torben Heien, and Olsen, Tore (2014). Active vs. Passive Decisions and Crowd-Out

in Retirement Savings Accounts: Evidence from Denmark. *The Quarterly Journal of Economics* 129(3) 1141-1219.

[9] Cilke, James (2014). "The Case of the Missing Strangers: What we Know and Don't Know About Non-Filers." Unpublished manuscript.

[10] Crain, Terry L. and Austin, Jeffrey R. (1997). "An Analysis of the Tradeoff Between Tax Deferred Earnings in IRAs and Preferential Capital Gains." *Financial Services Review* 6(4), 227-242.

[11] De Nardi, Mariacristina, French, Eric, and Jones, John B. (2016). "Savings After Retirement: A Survey." *Annual Review of Economics* 8, 177-204.

[12] Dhuey, Elizabeth, Figlio, David, Karbownik, Krzysztof, and Roth, Jeffrey (2017). "School Starting Age and Cognitive Development." NBER Working Paper 23660.

[13] Dhuey, Elizabeth and Lipscomb, Stephen (2008). "What makes a leader? Relative age and high school leadership." *Economics of Education Review* 27(2), 173-183.

[14] Engen, Eric M., Gale, William G., and Scholz, John K. (1994). "Do Saving Incentives Work?" *Brookings Papers on Economic Activity* 1994(1), 85-180.

[15] Engen, Eric M., Gale, William G., and Scholz, John K. (1996). "The illusory effects of saving incentives on saving?" *Journal of Economic Perspectives* 10(4), 113-138.

[16] Frandsen, Brigham R., Frolich, Markus, and Melly, Blaise (2012). "Quantile treatment effects in the regression discontinuity design." *Journal of Econometrics* 168(2), 382-395.

[17] Imbens, Guido and Kalyanaraman, Karthik (2011). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79(3), 933-959.

[18] Gelber, Alexander M. (2011). "How do 401(k)s Affect Saving? Evidence from Changes in 401(k) Eligiblity." *American Economic Journal: Economic Policy* 3(4), 103-122.

[19] Goodman, Lucas, Mackie, Kathleen, Mortenson, Jacob A., and Schramm, Heidi R. (2019). "The Evolution of Leakage and Retirement Asset Flows in the U.S." Unpublished manuscript.

[20] Layton, Timothy J., Barnett, Michael L., Hicks, Tanner R., and Jena, Anupam B. (2018). "Attention Deficit-Hyperactivity Disorder and Month of School Enrollment." *The New England Journal of Medicine* 379(22), 2122-2130.

[21] McCrary, Justin (2008). "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142(2), 698-714.

[22] Mortenson, Jacob A., Schramm, Heidi R., and Whitten, Andrew (2018). "The Effects of Required Minimum Distribution Rules on Withdrawals from Traditional IRAs." Unpublished manuscript.

[23] Musch, Jochen and Grondin, Simon (2001). "Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport." *Developmental Review* 21(2), 147-167.

[24] Pence, Karen M. (2006). "The Role of Wealth Transformations: An Application to Estimating the Effect of Tasx Incentives on Saving." *The B.E. Journal of Economic Analysis & Policy* 5(1), 1-26.

[25] Poterba, J., Venti, Steven, and Wise, David (1994). "401(k) plans and tax-deferred saving." In *Studies in the Economics of Aging*, Wise, D.A. (ed.): University of Chicago Press, Chicago, 105-138.

[26] Poterba, J., Venti, Steven, and Wise, David (1995). "Do 401(k) contributions crowd out other personal saving?" *Journal of Public Economics* 58(1), 1-32.

[27] Poterba, J., Venti, Steven, and Wise, David (1996). "How retirement saving programs increase saving." *Journal of Economic Perspectives* 10(4), 91-112.

[28] Poterba, J., Venti, Steven, and Wise, David (2011). "The Composition and Drawdown of Wealth in Retirement." *Journal of Economic Perspectives* 25(4), 95-118.

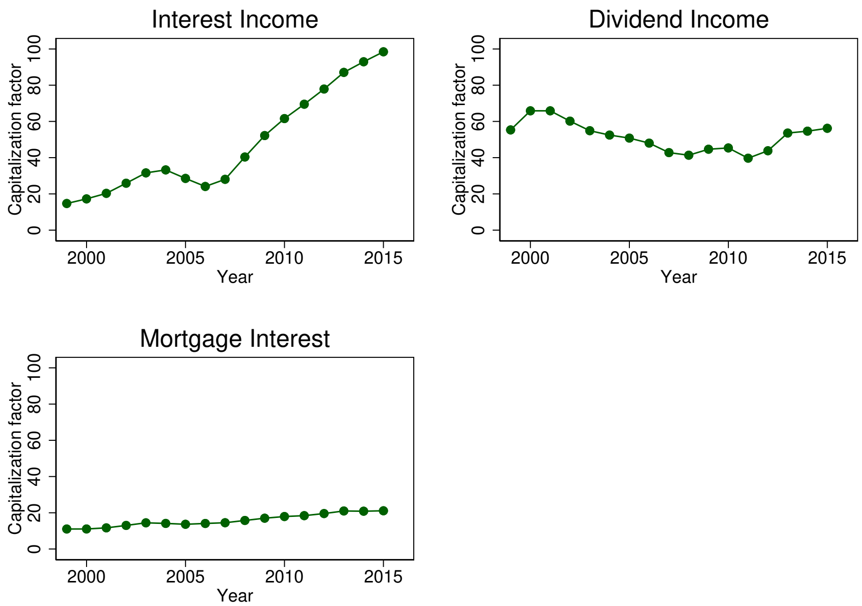
[29] Poterba, J., Venti, Steven, and Wise, David (2013). "The Drawdown of Personal Retirement Assets: Husbanding or Squandering." Unpublished Manuscript.

[30] Rubin, Donald B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66(5), 688-701.

[31] Saez, Emmanuel and Zucman, Gabriel (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." *The Quarterly Journal of Economics* 131(2), 519-578.

A Appendix

Figure A1: Time series of capitalization factors



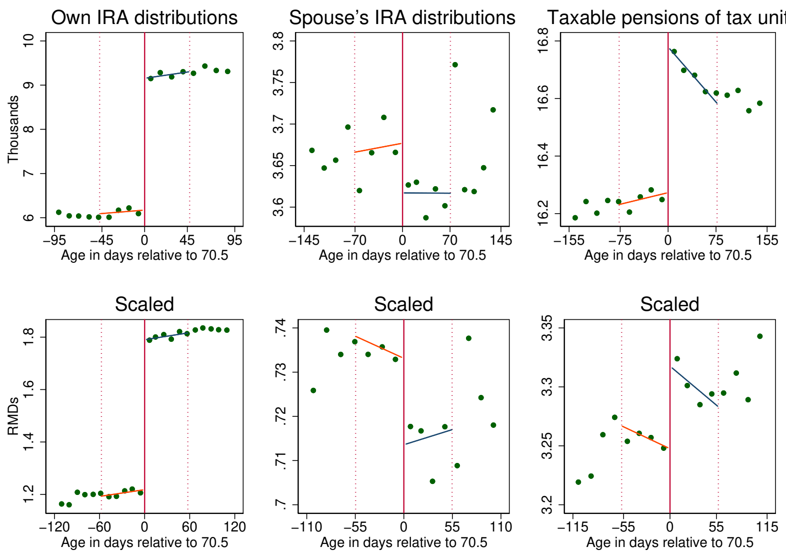
Notes: This figure plots the evolution of the capitalization factors over time. Broadly, the capitalization factors are computed as the ratio of the aggregate stock of the relevant asset in the economy (held in accounts that could generate the relevant type of taxable income) to the observed income reported on tax returns (in the case of interest and dividend income) or reported in information returns (in the case of mortgage interest). See text for further details of the calculation of these capitalization factors.

Notes: This figure plots outcomes at age 70 as a function of the running variable. In the leftmost two panels, the outcome is IRA distributions of the reference individual. In the middle two panels, the outcomes is IRA distributions of the spouse; this outcome is zero for unmarried individuals. In the rightmost two panels, the outcome is taxable pensions as reported on Form 1040. These three outcomes sum to total retirement distributions, as discussed in the main text. In the top three panels, the outcome is in thousands of dollars. In the bottom three panels, the outcome is scaled by the potential RMD (and the specification is weighted by such). See also the notes to Figure 1.

Notes: This figure plots the share of individuals who hold a positive-value IRA at the end of the age-70 year as a function of the running variable, x (age in days relative to 70.5). Those to the right of 0 are required to take an RMD for age 70; those to the left are not. Bin size is selected using the `rdplot` command in STATA. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped.

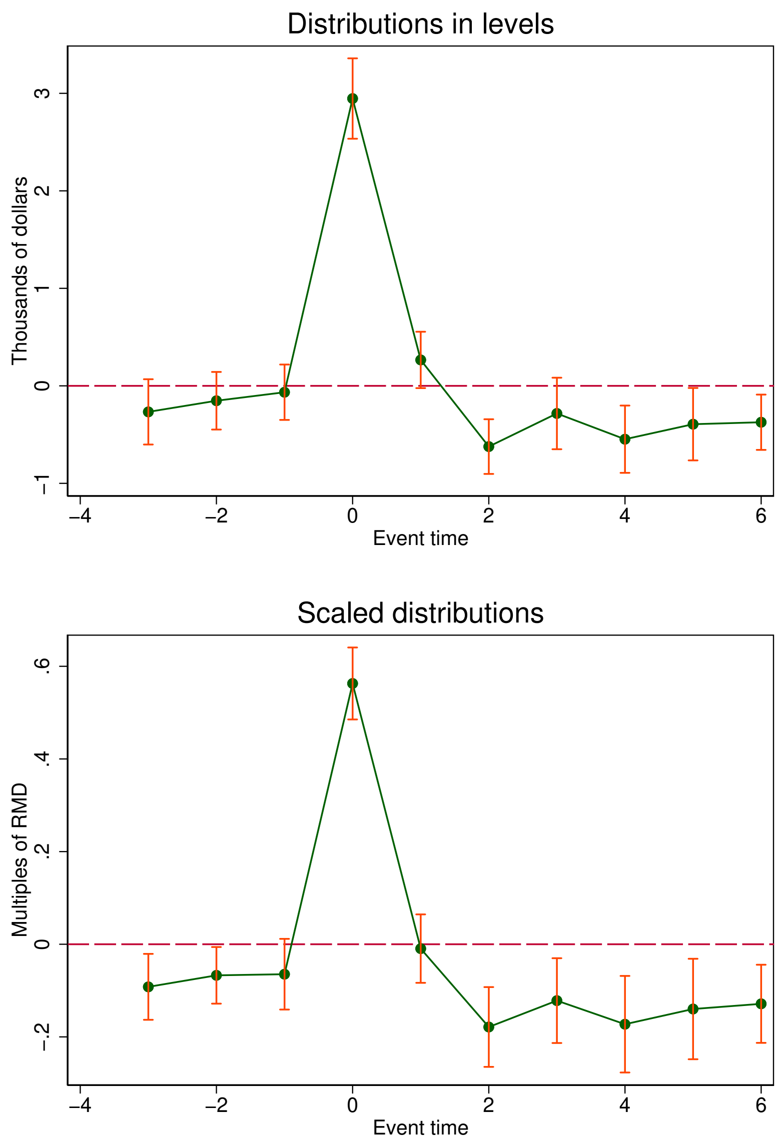
Notes: This figure plots the ratio of the number of individuals in the main sample to the number of individuals in the tax population that are the same age (69) in the same year, as a function of the running variable. The tax population includes all individuals who appear on a tax return or who receive an information return in a given year; see text for further discussion.

Figure A2: Effect of RMD treatment on retirement distributions, broken down between own IRA, spouse's IRA, and taxable pensions



The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Additionally, the dotted lines closest to zero indicate $|x| = 1.5$; in the baseline specifications in this paper, observations with $|x| \leq 1.5$ are dropped.

Figure A3: Effect RMD treatment on retirement distributions over time: balanced panel



Notes: This figure is analogous to Figure 2, except that the sample is restricted to those with $\tau \leq 2010$ in order to balance the panel. See also the notes to Figure 2.

Figure A4: Effect of RMD treatment on holding a positive-value IRA at the end of event time 0

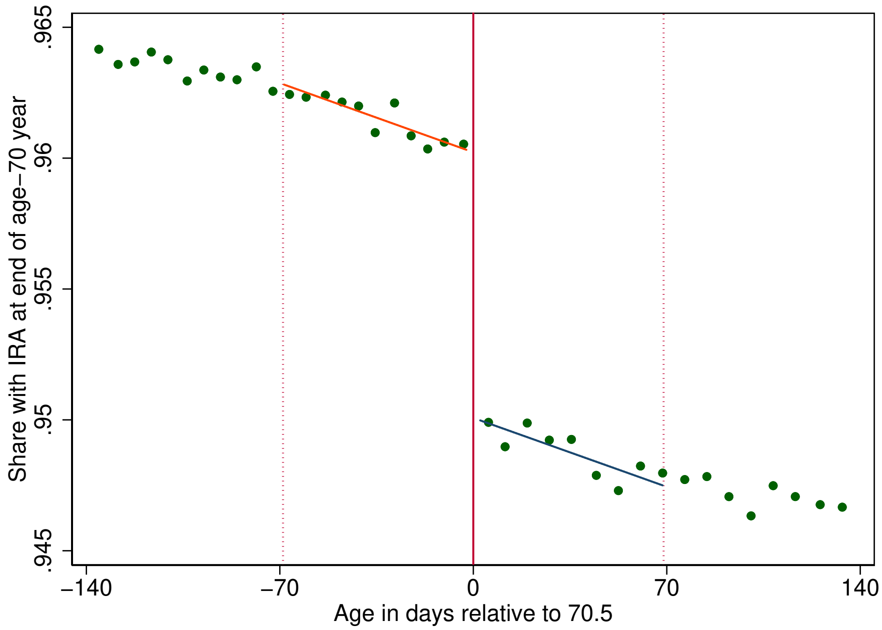
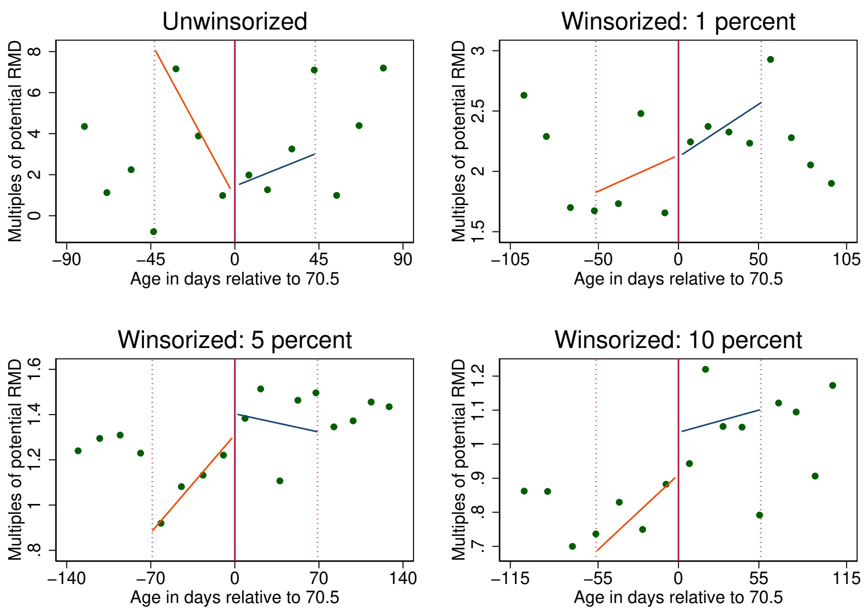
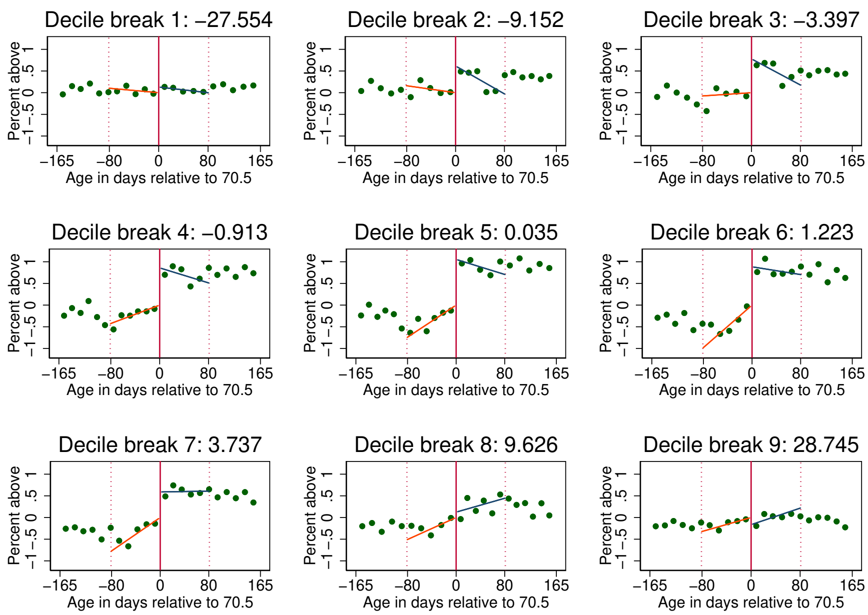


Figure A5: R.D. estimates of RMD status on taxable saving near RMD treatment, scaled by potential RMD



Notes: This figure is analogous to Figure 3, except that the dependent variable is scaled by the IRA RMD that would apply at time τ if the individual were subject to RMDs (the "potential RMD"). The specification is weighted by that potential RMD, censored at $100,000. See also the notes to Figure 3.

Figure A6: R.D. estimates of RMD status on scaled saving exceeding some threshold y^*



Notes: This figure plots the share of individuals with taxable saving in excess of various thresholds y^* as a function of the running variable, x (age in days relative to 70.5). Bin size is selected using the `rdplot` command in STATA. The fits from Equation (5) are plotted within the optimal CCT bandwidth, whose bounds are indicated by dotted lines. Observations with birthdays between June 29 and July 2 ($|x| \leq 1.5$) are dropped. See also the notes to Figure 3.

Figure A7: Sample observations as a share of tax population

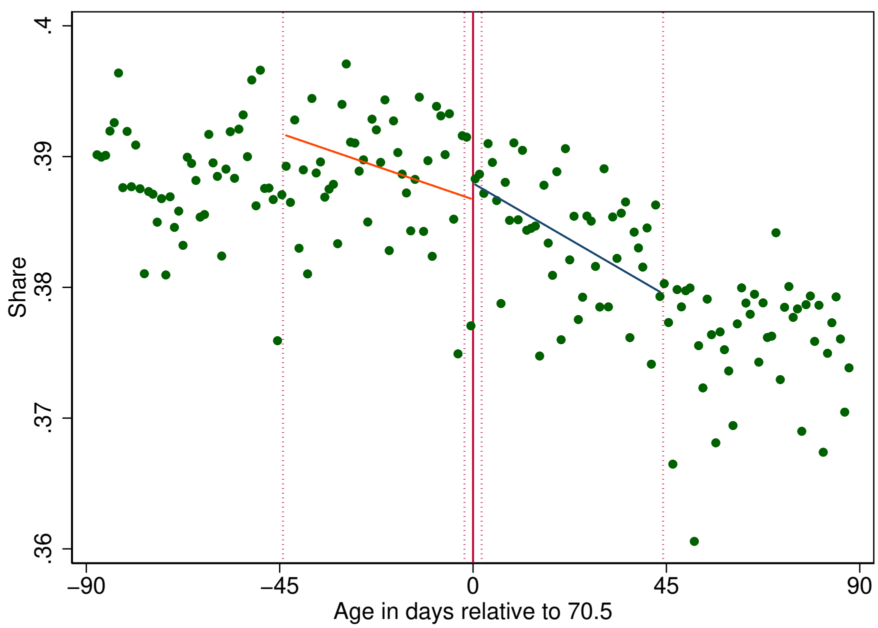


Table A1: Effect of RMD status at age 70 on retirement distributions over time: raw levels (thousands)

		Bandwidths	
	200 percent	Baseline	50 percent
Event time	(1)	(2)	(3)
-3	-0.018	**-0.061**	-0.070
	(0.104)	**(0.145)**	(0.183)
	[97.20 days]	**[48.60 days]**	[24.30 days]
-2	-0.032	**-0.077**	-0.028
	(0.102)	**(0.166)**	(0.274)
	[87.52 days]	**[43.76 days]**	[21.88 days]
-1	-0.023	**-0.008**	-0.002
	(0.078)	**(0.125)**	(0.218)
	[143.40 days]	**[71.70 days]**	[35.85 days]
0	3.439***	**3.446*****	3.532***
	(0.094)	**(0.141)**	(0.232)
	[113.83 days]	**[56.92 days]**	[28.46 days]
1	0.376***	**0.356*****	0.405**
	(0.085)	**(0.128)**	(0.177)
	[112.06 days]	**[56.03 days]**	[28.02 days]
2	-0.509***	**-0.558*****	-0.581***
	(0.087)	**(0.127)**	(0.201)
	[100.64 days]	**[50.32 days]**	[25.16 days]
3	-0.174*	**-0.267***	-0.305
	(0.097)	**(0.151)**	(0.239)
	[97.63 days]	**[48.81 days]**	[24.41 days]
4	-0.391***	**-0.546*****	-0.662**
	(0.102)	**(0.160)**	(0.269)
	[96.40 days]	**[48.20 days]**	[24.10 days]
5	-0.241**	**-0.400****	-0.465*
	(0.101)	**(0.164)**	(0.276)
	[97.99 days]	**[49.00 days]**	[24.50 days]
6	-0.311***	**-0.459*****	-0.582**
	(0.102)	**(0.144)**	(0.237)
	[93.07 days]	**[46.54 days]**	[23.27 days]

Notes: The second column of this table reports the coefficient estimates and standard errors underlying the top panel of Figure 2. Columns 1 and 3 report analogous coefficient estimates for bandwidths that are 200 percent and 50 percent, respectively, of the baseline bandwidth. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors are robust to heteroskedasticity.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See also the notes to Figure 2.

Table A2: Effect of RMD status at age 70 on retirement distributions over time: raw levels (thousands), balanced panel

	Bandwidths		
	200 percent	Baseline	50 percent
Event time	(1)	(2)	(3)
-3	-0.140	**-0.267**	-0.262
	(0.123)	**(0.171)**	(0.221)
	[78.35 days]	**[39.17 days]**	[19.59 days]
-2	-0.049	**-0.153**	-0.058
	(0.102)	**(0.151)**	(0.224)
	[90.83 days]	**[45.42 days]**	[22.71 days]
-1	-0.077	**-0.065**	-0.179
	(0.096)	**(0.145)**	(0.244)
	[115.60 days]	**[57.80 days]**	[28.90 days]
0	3.060***	**2.947*** **	2.573***
	(0.152)	**(0.210)**	(0.277)
	[77.38 days]	**[38.69 days]**	[19.35 days]
1	0.321***	**0.266***	0.173
	(0.105)	**(0.147)**	(0.186)
	[124.65 days]	**[62.32 days]**	[31.16 days]
2	-0.538***	**-0.623*** **	-0.815***
	(0.119)	**(0.143)**	(0.167)
	[96.31 days]	**[48.16 days]**	[24.08 days]
3	-0.177	**-0.283**	-0.357
	(0.126)	**(0.187)**	(0.286)
	[93.24 days]	**[46.62 days]**	[23.31 days]
4	-0.411***	**-0.547*** **	-0.696**
	(0.113)	**(0.176)**	(0.290)
	[106.22 days]	**[53.11 days]**	[26.55 days]
5	-0.259**	**-0.393** **	-0.493
	(0.119)	**(0.189)**	(0.308)
	[99.01 days]	**[49.51 days]**	[24.75 days]
6	-0.252**	**-0.373*** **	-0.430*
	(0.100)	**(0.145)**	(0.231)
	[107.25 days]	**[53.62 days]**	[26.81 days]

Notes: This table is analogous to Appendix Table A1, except that the sample is restricted to $\tau \leq 2010$, ensuring that the panel is balanced. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors are robust to heteroskedasticity.* $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. See also the notes to Appendix Table A1 and Figure 2.

Table A3: Effect of RMD status at age 70 on IRA distributions over time: scaled

		Bandwidths	
	200 percent	Baseline	50 percent
Event time	(1)	(2)	(3)
-3	-0.064***	**-0.057****	-0.075**
	(0.020)	**(0.029)**	(0.034)
	[104.88 days]	**[52.44 days]**	[26.22 days]
-2	-0.059***	**-0.064****	-0.085**
	(0.019)	**(0.029)**	(0.042)
	[93.24 days]	**[46.62 days]**	[23.31 days]
-1	-0.068***	**-0.050**	-0.069
	(0.021)	**(0.034)**	(0.056)
	[108.93 days]	**[54.46 days]**	[27.23 days]
0	0.599***	**0.623*****	0.601***
	(0.024)	**(0.035)**	(0.052)
	[116.04 days]	**[58.02 days]**	[29.01 days]
1	-0.015	**-0.000**	-0.034
	(0.020)	**(0.028)**	(0.037)
	[114.35 days]	**[57.17 days]**	[28.59 days]
2	-0.177***	**-0.170*****	-0.204***
	(0.023)	**(0.033)**	(0.044)
	[112.83 days]	**[56.41 days]**	[28.21 days]
3	-0.127***	**-0.123*****	-0.167***
	(0.023)	**(0.035)**	(0.049)
	[133.49 days]	**[66.74 days]**	[33.37 days]
4	-0.147***	**-0.162*****	-0.216***
	(0.030)	**(0.047)**	(0.069)
	[109.81 days]	**[54.91 days]**	[27.45 days]
5	-0.128***	**-0.145*****	-0.192**
	(0.031)	**(0.050)**	(0.078)
	[104.72 days]	**[52.36 days]**	[26.18 days]
6	-0.141***	**-0.143*****	-0.198***
	(0.026)	**(0.039)**	(0.061)
	[151.11 days]	**[75.56 days]**	[37.78 days]

Notes: The second column of this table reports the coefficient estimates and standard errors underlying the bottom panel of Figure 2. Columns 1 and 3 report analogous coefficient estimates for bandwidths that are 200 percent and 50 percent, respectively, of the baseline bandwidth. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors are robust to heteroskedasticity.* $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. See also the notes to Figure 2.

Table A4: Effect of RMD status at age 70 on retirement distributions over time: scaled, balanced panell

	Bandwidths		
	200 percent	Baseline	50 percent
Event time	(1)	(2)	(3)
-3	-0.077***	**-0.092****	-0.100**
	(0.025)	**(0.036)**	(0.051)
	[105.57 days]	**[52.79 days]**	[26.39 days]
-2	-0.078***	**-0.067****	-0.095**
	(0.022)	**(0.031)**	(0.048)
	[117.38 days]	**[58.69 days]**	[29.35 days]
-1	-0.091***	**-0.065***	-0.099
	(0.026)	**(0.039)**	(0.062)
	[120.11 days]	**[60.06 days]**	[30.03 days]
0	0.554***	**0.563*****	0.512***
	(0.029)	**(0.040)**	(0.049)
	[123.08 days]	**[61.54 days]**	[30.77 days]
1	-0.030	**-0.009**	-0.048
	(0.027)	**(0.038)**	(0.054)
	[115.10 days]	**[57.55 days]**	[28.77 days]
2	-0.187***	**-0.179*****	-0.220***
	(0.032)	**(0.044)**	(0.056)
	[110.86 days]	**[55.43 days]**	[27.72 days]
3	-0.130***	**-0.122*****	-0.149**
	(0.030)	**(0.047)**	(0.070)
	[118.55 days]	**[59.28 days]**	[29.64 days]
4	-0.168***	**-0.173*****	-0.223***
	(0.033)	**(0.053)**	(0.081)
	[121.72 days]	**[60.86 days]**	[30.43 days]
5	-0.135***	**-0.140****	-0.169*
	(0.034)	**(0.055)**	(0.089)
	[106.27 days]	**[53.14 days]**	[26.57 days]
6	-0.142***	**-0.129*****	-0.169**
	(0.028)	**(0.043)**	(0.068)
	[145.18 days]	**[72.59 days]**	[36.29 days]

Notes: This table is analogous to Appendix Table A2, except that the sample is restricted to $\tau \leq 2010$, ensuring that the panel is balanced. Bandwidths are reported in brackets below each estimate. The regressions are estimated using local linear regression with a triangular kernel. Standard errors are robust to heteroskedasticity.* $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. See also the notes to Appendix Table A2 and Figure 2.

Table A5: First stage of threshold approach: effect of RMD treatment on retirement distributions, by asset quintiles

	First stage coefficient (1)
Zero assets	0.300***
	(0.100)
Quintile 1	0.277***
	(0.065)
Quintile 2	0.365***
	(0.059)
Quintile 3	0.534***
	(0.049)
Quintile 4	0.653***
	(0.052)
Quintile 5	0.894***
	(0.055)

Notes: This table reports the estimated first stage coefficient, separately by six asset bins, where asset bins are defined based on taxable assets (not deprecated by mortgage debt) as of time $\tau - 1$. The first row includes those with zero assets. The remaining rows represent quintiles of those with positive assets. The first stage coefficient is the effect of the RMD treatment on retirement distributions, scaled by the potential IRA (and weighted by the potential IRA censored at $100,000). These coefficients enter into the calculation of crowdout using the threshold method, as reported in Table 6. This table uses a fixed bandwidth of xx days. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$. See also the notes to Table 6.

Table A6: Discontinuities in baseline covariates: assets

	Bandwidths		
	90 days	45 days	22.5 days
	(1)	(2)	(3)
Coefficient	0.090	**0.051**	0.308
	(0.138)	**(0.205)**	(0.323)
Median effect (000s)	0.287	**0.162**	0.982
	[-0.580, 1.404]	**[-1.246, 1.939]**	[-1.399, 4.171]
Optimal θ		0.031	
		[0.031, 0.032]	

Notes: This table reports results analogous to Table 8 for taxable assets, with additional bandwidths. The optimal θ is calculated using the baseline bandwidth and is held fixed for other bandwidths. See also the notes to Table 8.

Table A7: Discontinuities in baseline covariates: IRA FMV

	Bandwidths		
	90 days	45 days	22.5 days
	(1)	(2)	(3)
Coefficient	0.092**	**0.131****	0.233**
	(0.041)	**(0.062)**	(0.097)
Median effect (000s)	0.368**	**0.524**	0.927**
	[0.045, 0.671]	**[-0.011, 0.915]**	[0.188, 1.428]
Optimal θ		0.062	
		[0.055, 0.072]	

Notes: This table reports results analogous to Table 8 for IRA value, with additional bandwidths. The optimal θ is calculated using the baseline bandwidth and is held fixed for other bandwidths. See also the notes to Table 8.

Table A8: Discontinuities in baseline covariates: Taxable assets

	Bandwidths		
	90 days	45 days	22.5 days
	(1)	(2)	(3)
Coefficient	0.027	**-0.009**	0.035
	(0.022)	**(0.032)**	(0.051)
Median effect (000s)	0.094	**-0.031**	0.120
	[-0.066, 0.263]	**[-0.328, 0.230]**	[-0.432, 0.561]
Optimal θ		0.124	
		[0.115, 0.137]	

Notes: This table reports results analogous to Table 8 for taxable assets, with additional bandwidths. The optimal θ is calculated using the baseline bandwidth and is held fixed for other bandwidths. See also the notes to Table 8.

B Conceptual Framework: More Details

In this Appendix, I prove the two assertions made in the text: (1) it is optimal to exhaust taxable assets before taking pension distributions in excess of the RMD and (2) when assets are positive, the usual consumption Euler equation holds with R_A as the relevant interest rate. To review, the model is as follows.

The budget constraint is:

$$c_t = \underbrace{R_P P_t - P_{t+1}}_{=d_t,\text{Pension distributions}} + \underbrace{R_A A_t - A_{t+1}}_{=-s_t,\text{Taxable dissaving}} \tag{7}$$

The RMD constraint is as follows.

$$R_P P_t - P_{t+1} \geq \theta_t P_t \tag{8}$$

Lastly, there are nonnegativity constraints on P_{t+1} and A_{t+1}.

The Lagrangian for the Bellman equation is:

$$\begin{aligned} L = u(R_P P_t - P_{t+1} + R_A A_t - A_{t+1}) + \\ \lambda_t\big((R_P - \theta_t)P_t - P_{t+1}\big) + \\ \mu^A_{t+1} A_{t+1} + \mu^P_{t+1} P_{t+1} + \\ \beta V^{t+1}(A_{t+1}, P_{t+1}) \end{aligned} \tag{9}$$

The first order conditions are as follows, where $V^{t'}_A$ and $V^{t'}_P$ refer to the partial derivative of $V^{t'}(A_{t'}, P_{t'})$ with respect to $A_{t'}$ and $P_{t'}$ respectively.

$$\begin{aligned} A_{t+1}&: u'(c_t) = \mu^A_{t+1} + \beta V^{t+1}_A \\ P_{t+1}&: u'(c_t) + \lambda_t = \mu^P_{t+1} + \beta V^{t+1}_P \end{aligned} \tag{10}$$

The envelope conditions (updated to $t+1$) are:

$$\begin{aligned} V^{t+1}_A &= R_A u'(c_{t+1}) \\ V^{t+1}_P &= R_p u'(c_{t+1}) + \lambda_{t+1}(R_P - \theta_{t+1}) \end{aligned} \tag{11}$$

Combining, we have the following two equations for $u'(c_t)$:

$$\begin{aligned} u'(c_t) &= \mu^A_{t+1} + \beta R_A u'(c_{t+1}) \\ &= \mu^P_{t+1} + \beta R_P u'(c_{t+1}) - \lambda_t + \lambda_{t+1}(R_P - \theta_{t+1}) \end{aligned} \tag{12}$$

The usual complementarity slackness conditions also hold.

Assertion (2) is immediate. When assets are positive, μ^A_{t+1} equals zero, and the usual consumption Euler equation holds with R^A as the relevant interest rate. ■ Intuitively this must be the case because the marginal source of (dis-)saving is taxable assets, which earns an interest rate R_A.

Assertion (1) can be restated equivalently as follows: "If RMDs do not bind, then assets (A_{t+1}) must equal zero." When RMDs do not bind, $\lambda_t = \lambda_{t+1} = 0$. Then, rearranging the two equations in 12 yields:

$$\beta(R_P - R_A)u'(c_{t+1}) = \mu^A_{t+1} - \mu^P_{t+1} \tag{13}$$

Since the left hand side is positive (because $R_P > R_A$) and $\mu^P_{t+1} \geq 0$, it must be the case that $\mu^A_{t+1} > 0$, which implies $A_{t+1} = 0$. ■