NBER WORKING PAPER SERIES

INCORPORATING EMPLOYEE HETEROGENEITY INTO DEFAULT RULES FOR RETIREMENT PLAN SELECTION

Gopi Shah Goda
Colleen Flaherty Manchester

Working Paper 16099
http://www.nber.org/papers/w16099

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2010

We would like to thank James Robertson for excellent research assistance, and Kristine Brown, Maria Fitzpatrick, Caroline Hoxby, Damon Jones, Ron Laschever, Brigitte Madrian, Olivia Mitchell, Kevin Mumford, Anita Alves Pena, Greg Rosston, John Shoven, and Sita Slavov for helpful comments as well as seminar participants at Stanford University, the University of Minnesota and the NBER Summer Institute. This research was supported by the Center for Retirement Research at Boston College pursuant to a grant from the U.S. Social Security Administration through the Steven H. Sandell 2009-2010 Grant Program for Junior Scholars in Retirement Research. The opinions and conclusions are solely those of the authors and should not be construed as representing the opinions or policy of the Social Security Administration, any agency of the Federal Government, the Center for Retirement Research at Boston College, or the National Bureau of Economic Research.

Incorporating Employee Heterogeneity into Default Rules for Retirement Plan Selection
Gopi Shah Goda and Colleen Flaherty Manchester
NBER Working Paper No. 16099
July 2010
JEL No. H8,J26

ABSTRACT

We study the effect of incorporating heterogeneity into default rules by examining the choice between retirement plans at a firm which transitioned from a defined benefit (DB) to a defined contribution (DC) plan. The default plan for existing employees varied discontinuously depending on their age. Using regression discontinuity techniques, we find that the default increased the probability of enrollment in the default plan by 60 percentage points. We develop a framework to solve for the optimal default rule analytically and numerically and find that considerable welfare gains are possible if defaults vary by observable characteristics.

Gopi Shah Goda
Stanford University
SIEPR
366 Galvez St.
Stanford, CA 94305
and NBER
gopi@stanford.edu

Colleen Flaherty Manchester
Human Resources and Industrial Relations
University of Minnesota
Room 3-300R
321 - 19th Avenue South
Minneapolis, MN 55455
cmanch@umn.edu

1 Introduction

The impact of default rules, which specify an individual's outcome when no choice is made, has been well-documented over the last decade. From dramatically increasing participation in 401(k) plans (e.g., Madrian and Shea 2001) to influencing organ donor status (e.g., Johnson and Goldstein 2003), default rules have a large effect on outcomes despite holding the set of choices constant. They have attracted widespread attention because their influential effects challenge the standard economic assumption of rational decision-making (DellaVigna 2009). Furthermore, the strong effects of defaults open up the possibility for welfare-enhancing policy interventions.

In this paper, we examine the role of default provisions in a new setting: a one-time, irrevocable choice between two alternative pension plans, a defined benefit (DB) plan and a defined contribution (DC) plan. While DB plans provide retirement benefits that are a set formula based on earnings and service, DC plans provide benefits based on tax-advantaged contributions and subsequent investment performance. Because the plans differ in their accrual patterns and risk characteristics, decisions between the two plans can lead to substantially different retirement income profiles. Furthermore, the fact that the choice cannot be changed after a particular date represents an important distinction from much of the previous work on defaults in the context of planning for retirement. In particular, decisions regarding asset allocation, the rate of contributions, or even participation in 401(k) plans are choices that affect one's financial circumstances over a horizon as short as a single pay period and can be corrected if mistakes are thought to have been made. By contrast, an irreversible choice between retirement plans can lead to significantly different amounts of wealth.

We examine a particular firm's transition from a DB plan to a DC plan. While all new employees were offered only the DC plan, the firm offered existing workers a one-time opportunity to make an irreversible choice between the DB plan and the DC plan for all future benefit accruals while employed at the firm. Employees who did not actively choose a retirement plan were defaulted into one plan or the other depending on their age at the time of the transition: individuals age 45 or older were defaulted to remain in the DB plan, while the default for employees under age 45 was to switch to the DC plan.

The unique nature of the default, namely that it varied discontinuously by age, allows us to estimate the causal effect using a regression discontinuity framework. This methodology differs from past studies of the effect of defaults, which have exploited changes in a default policy regime over time or across localities to identify the effect of the default. Figure 1 depicts the percent enrollment in the DC plan as a function of age. By visual inspection, the large discontinuity at age 45 supports the hypothesis that the default had a substantial effect on plan enrollment. Formally, our regression estimates likewise indicate a strong effect: individuals just under the age of 45 are approximately 60 percentage points more likely to choose the DC plan relative to those just over age 45 at the time of plan transition. Given the permanent nature of the decision and the large amount of wealth at stake, the effect is even more dramatic than previous findings of the effects of defaults.

The variation in the default plan across employees allows us to consider the potential gains from having a heterogeneous default rule. While there are a few examples of heterogeneous defaults in the realm of retirement savings (e.g., default distribution options for DB plans that vary by marital status and target-date funds which vary default asset allocation by age), there are often potential legal consequences of treating employees differently. As a result, data limitations have thus far prevented the explicit examination of the effect of incorporating heterogeneity into default provisions (Carroll, Choi, Laibson, Madrian and Mertrick 2009). We develop a framework for evaluating the potential welfare implications of default provisions for employees and applying it to the firm in our setting. We propose that an optimal default policy maximizes the "aggregate default wealth" or the risk-adjusted value that each employee receives if he or she defaults, aggregated over all employees at the firm. Maximizing the aggregate default wealth is likely to improve employee welfare in many models of household behavior given the strong effects of defaults in this and in other contexts.

We examine default rules that are characterized by a cutoff age where employees younger than the age cutoff are defaulted into the DC plan and employees older than this cutoff age are defaulted into the DB plan. We solve for the optimal age cutoff analytically and provide comparative statics of how the optimal age-based default varies with pension plan, firm, and employee characteristics. We show that the sign of the relationship between the optimal age cutoff and a parameter, such as the level of risk aversion, depends on how the parameter affects the relative value of the DC plan over the DB plan. We also define measures to evaluate the gain from implementing the optimal age-based default against alternative universal default policies that default all employees into the same plan or a policy that incorporates additional characteristics in the determination of the default.

We illustrate properties of the optimal age cutoff numerically for the firm in our analysis that underwent the transition. We find that, under baseline assumptions, the firm's chosen age cutoff of 45 is within the range of optimal cutoff ages under reasonable assumptions regarding risk aversion, and show how these results change with the value of different parameters, confirming our analytical results. Finally, we find that an optimal age-based default policy is far superior to a universal policy and steers over 99 percent of employees to their optimal plan over a broad range of plausible levels of risk aversion. These results suggest that conditioning the default on observable characteristics in addition to age, such as gender and income, would have a negligible effect on worker welfare. However, the relative value of each retirement plan is strongly influenced by the level of risk aversion assumed. Therefore, if the employer misestimates the level of risk aversion among employees or if there is substantial heterogeneity in risk aversion across the workforce, the benefits of conditioning the default on age are reduced. However, the benefits are still greater than those under a universal DC default and comparable to those under a universal DB default.

The remainder of the paper is organized as follows. Section 2 provides an overview of the changing pension landscape and describes how our analysis builds upon the literature on retirement plan choice and default provisions. Section 3 presents the regression discontinuity analysis to identify the causal effect of the default on plan enrollment. Section 4 analytically solves for the optimal age-based default rule and its properties for the general case, while Section 5 empirically illustrates the results from Section 4 using simulated retirement wealth and variability for the employees at the firm in our setting. Section 6 concludes the paper.

2 Background and Related Literature

Over the last 30 years, there has been a pronounced shift from DB plans to DC plans as a result of legislative changes that increased the administrative burden of offering DB plans and changing demand from an increasingly mobile workforce. These changes led employers to replace terminated DB plans with DC plans (Papke 1999; GAO 2008) and made DC plans the most common type of newly implemented employer-provided retirement plan (Kruse 1995). In a recent study, the GAO found that approximately 3.3 million workers who actively participate in a DB plan were affected by a DB plan "freeze," or the discontinuation of future benefit accruals (GAO 2008). Among these, approximately 83 percent implemented alternative retirement plans, usually a DC plan, implying that a substantial number of employees have experienced a plan migration. A 2003 report by Towers-Perrin indicates that 14 percent of programs that implemented a plan change in recent years allowed current employees to choose between the old and new plan (Towers-Perrin 2003).

This paper builds off both the literature on determinants of retirement plan selection and the literature on the role of default provisions on individual's retirement savings behavior. The two types of retirement plans differ in the risk faced by participants: participants in DC plans bear the risk of poor investment experience, while DB participants are exposed to the labor market risk of unexpected separations as well as the risk of their employer defaulting on pension obligations. The consensus in the literature on plan selection is that the expected relative value of retirement wealth under either plan depends on an individual's risk preferences, demographic characteristics, and expected job mobility (Bodie, Marcus and Merton 1988; Papke 1999; Clark and Pitts 1999; Clark, Ghent and McDermed 2006; Brown and Weisbenner 2007). Studies in this area typically evaluate whether the individual determinants of plan selection are consistent with economic theory in terms of which worker characteristics are associated with higher expected retirement wealth for a DC plan relative to a DB plan, or vice versa. These papers use data from institutions, such as public university systems, that offer individuals a choice between enrolling in either a DC or a DB plan (Clark and Pitts 1999; Clark, Ghent and McDermed 2006; Brown and Weisbenner 2007; Manchester 2010) or those that have changed plan offerings, for example, replacing a DB plan with a DC plan (Papke 2004; Yang 2005).

The second branch of literature – the role of default rules in retirement savings behavior – developed from the fact that DC plans shift the responsibility of enrollment, contribution, and investment decisions to the individual. There is now a large established body of literature that shows that default provisions can have a large impact on savings decisions (for summary, see Beshears, Choi, Laibson and Madrian (2006)). This literature has investigated the role of defaults in decisions regarding participation (e.g., Madrian and Shea 2001), contribution rates (e.g., Choi, Laibson, Madrian and Metrick 2004), asset allocation (e.g., Choi, Laibson and Madrian 2005) and, recently, distributions from DC plans (Mitchell, Mottola, Utkus and Yamaguchi 2009). Typically these studies use changes in 401(k) plan characteristics and default rules at a particular employer as a type of "natural experiment" to investigate the causal effect of defaults on employee retirement saving, particularly among new hires. This literature has found that despite the fact that a default does not change the menu of options, switching the default option has substantial effects on retirement savings decisions at each juncture. It is important to note, however, that these previous studies estimate changes in an outcome variable that has a limited time horizon because participants can change their enrollment decisions at future dates. On the other hand, plan choice at the time of a transition from DB to DC plans is irreversible. As pointed out by Beshears et al. (2006), this permanent feature of the decision may mitigate the effect of defaults in the context of plan choice relative to previous results in the literature.

While the role of default rules in savings plans has been well-studied, the literature on the role of default provisions on plan choice is much more limited. In order to determine the causal effect of a default in this context, there must be a choice of plans and variation in the default plan either across employees or over time. Past research on firms transitioning from DB plans to DC plans has been in settings where the authors were unable to exploit any variation in the default rule to estimate a causal effect. For example, Papke (2004) and Yang (2005) both examined a one-time plan transition with a universal DB default, which precluded the identification of the default's causal effect on plan choice.[1]

Recent work on defaults has recognized the presence of worker heterogeneity in the design of welfare-enhancing policy interventions. Carroll et al. (2009) find that a third type of enrollment, "active decision," in which individuals must actively choose between enrolling or not is preferable to standard or automatic enrollment even when there is a large degree of heterogeneity in preferred savings rates if individuals have a strong propensity to procrastinate. When individuals are forced to make an active choice, the effect on enrollment rates in 401(k) plans is comparable to a default policy of automatic enrollment. A recent study by Handel (2009) suggests that removing switching costs by requiring active choice may allow people to make more appropriate health plan choices, but may also lead to worse outcomes in health care markets by increasing adverse selection into health plans.

We contribute to the existing literature in several ways. First, we quantify the effect of default provisions in a new context: an irreversible decision regarding plan choice. Second, we are able to exploit the variation in the default rule across employees as a new source of identification, and the particular type of variation allows for estimation using a regression discontinuity framework. Third, we provide an analytical framework for solving for the optimal age-based default rule and for assessing the gains from incorporating heterogeneity relative to alternative default policies. Finally, we illustrate these results empirically using data from the firm in our setting.

3 Causal Effect of Default on Plan Enrollment

This section outlines the regression discontinuity methodology used to estimate the causal effect of the default, provides additional details on the employees involved in the transition, and reports the results from estimating the effect of this particular age-based default on plan enrollment.

[1] Both Lachance, Mitchell and Smetters (2003) and Milevsky and Promislow (2004) evaluate the option instituted by the State of Florida in which public employees were given the opportunity to convert their accrued DB benefits into an individually-managed DC plan. The plan choice was revocable in that employees had a one-time option to "buy back" their DB benefit prior to retirement and both papers focus on analytically evaluating the value of this option, rather than the determinants of plan choice.

3.1 Regression Discontinuity Methodology

We implement a standard regression discontinuity framework (see, for example, Imbens and Lemieux (2008) and Lee and Lemieux (2009)) to examine the role of the default rule on plan choice. In particular, we are interested in estimating how being assigned the DC plan as the default relative to the DB plan affects the probability of switching to the DC plan. In this methodology, the treatment (being assigned the DC plan as the default) is determined by the value of a forcing variable (the employee's age as of September 1, 2002) being above or below a fixed cutoff value (age 45). Therefore, employees under age 45 received the treatment, while those age 45 or older did not. Because the treatment is a deterministic function of the forcing variable, age, this framework is known as a "sharp" regression discontinuity design. This methodology allows for age to be correlated with plan choice; however, the key assumption is that the relationship between age and the outcome is a smooth, continuous function. This assumption allows any discontinuity at the cutoff value to be interpreted as a causal effect (Imbens and Lemieux 2008). Therefore, the discontinuity in plan enrollment at the cutoff age is the causal effect of the DC default on the probability of enrolling in the DC plan.

Formally, the treatment is the default assignment, given by d_i, and is a deterministic function of the participant's age as of September 1, 2002, A_i,

$$d_i = \mathbf{1}\{A_i < c\}, \tag{1}$$

where the variable c denotes the cutoff value and is equal to 45 in this context. A value of $d_i = 1$ implies that the participant is assigned the DC plan as the default and $d_i = 0$ corresponds to being assigned the DB plan as the default.

The outcome of interest, enrollment into the DC plan, is given by the variable Y_i. We are interested in the causal impact of d_i on Y_i. This treatment effect, given by τ, is estimated as follows:

$$\tau = \lim_{a \uparrow c} E[Y_i|X_i, A_i = c] - \lim_{a \downarrow c} E[Y_i|X_i, A_i = c] = E[Y_i^{d_i=1} - Y_i^{d_i=0}|X_i, A_i = c], \tag{2}$$

where X_i represents other observable characteristics that are correlated with the outcome variable Y_i. The estimate of τ is thus obtained by estimating the following regression equation:

$$Y_i = \alpha + \beta(A_i - c) + \tau d_i + \gamma(A_i - c)d_i + X_i\pi + \varepsilon_i \tag{3}$$

for individuals whose age is in the interval $[c-h, c+h]$ for a bandwidth value of h. We examine the sensitivity of our results to different choices of bandwidths and to including higher order powers of $(A_i - c)$.

3.2 Data

To estimate the regression discontinuity estimate of the effect of the default as outlined above, we utilize administrative data from a large non-profit firm that offered 925 existing union employees an opportunity to remain in a DB plan or switch to a DC plan.[2,3] While we do not know the exact motives of the default rule, the firm's decision to discontinue the DB plan to new hires was likely due to increased administrative costs and follows the widespread movement toward DC plans.

These existing employees had approximately six months to make the election. As of September 1, 2002, employees were enrolled in their chosen plan or, if they had failed to make a choice, were defaulted into a plan depending on whether they were older or younger than age 45 on that date. We refer to those who formally chose their plan as "active" participants, and those who were defaulted into a plan as "passive" participants. Table 1 shows the distribution of the type of choice (active vs. passive) and the enrollment decision for the two age groups. Of those employees eligible for the transition, just under half made an active choice. However, the vast majority (70 percent) of employees who made an active choice mimicked the default rule, which is consistent with employees taking the default provisions as advice (Beshears et al. 2006).

[2]While the firm employs over 5,000 workers, only existing employees covered by a collective bargaining agreement were eligible for this transition; the existing non-unionized employees were offered a similar one-time opportunity to transition out of the DB plan into the DC plan in 1997. Data are not available for this earlier transition. However, the default rule for these workers was also an age-based default with a cutoff value of age 45.

[3]The DB and DC plans were of roughly similar generosity; the specific features of each plan are discussed in greater detail in Section 5.

After the plan transition, 48 percent of employees were enrolled in the DC plan (Table 2). The dataset contains information regarding the participant's age, gender, ethnicity/race, hourly wage, tenure, and hours per week. In addition, the employees are divided between two nearby campuses, a primary location where approximately two-thirds of the employees work and a secondary smaller location a few miles away. Summary statistics for these additional variables are provided in Table 2. We restrict the sample to workers under age 65, the normal retirement age at this firm under the DB plan; the average age is approximately 46 years.

One potential concern is that the sample used for the subsequent empirical evaluation is comprised of only unionized workers who may differ from non-unionized workers. Given that unions have historically favored DB plans over DC plans, the overall participation rates in the DB plan may be higher than in a non-unionized sample. Unionized workers may also have different propensities to enroll in the default plan. However, the estimation of the causal effect of the default, as outlined in Section 3.1, is not affected by factors that influence the sample's overall behavior because it is identified from the outcomes of workers on either side of age 45.

3.3 Regression Discontinuity Results

In order to validate using regression discontinuity methods to estimate a causal effect, we first verify that there is no discontinuity in the density of age at the cutoff value of age 45, which would suggest manipulation of the receipt of treatment.[4] As shown in Figure 2, there is a relatively constant density of employees surrounding the cutoff value of age 45. Second, we confirm that other observable covariates are a smooth function of the forcing variable and do not experience a discontinuity at the cutoff value as shown in the six panels of Figure 3.

After having verified these conditions for causal inference, we estimate the coefficient τ in Equation (3) by fitting linear probability models to the data assuming a rectangular kernel. Table 3 summarizes regression discontinuity estimates of the effect of the DC default on enrollment into the DC plan assuming a bandwidth of five years. Column (1), the local linear estimate, reports the

[4]Given that the forcing variable is based on birth date, manipulation is not a concern.

estimate of τ with no other right-hand side regressors. Column (2) allows the relationship between age and DC enrollment to be linear, with a different slope below and above the discontinuity. Column (3) adds in control variables summarized in Table 2. Columns (4) and (5) provide estimates assuming the relationship between age and DC enrollment follows a cubic function, and differ only in the addition of control variables in Column (5).

These estimates indicate that the default had a strong effect on the enrollment into the DC plan, confirming the initial evidence in Figure 1. An employee just under age 45 was approximately 60 percentage points more likely to enroll in the DC plan than an employee just over age 45. The estimated effect is larger when higher order powers of $(A_i - c)$ are included in the regression. We prefer the estimates in Column (3) because the linear specification fits the data well, as shown in Figure 1, which plots the best fit line and cubic polynomial on either side of the discontinuity. Quantitatively similar results are found when estimating the regression using the nonlinear probit model which accounts for the dichotomous nature of the dependent variable and the effects are robust to the choice of bandwidth as shown in Appendix A.[5]

Adding demographic control variables does not substantially improve the fit of the model after the default provision is taken into account. Coefficients on gender and race binary variables (not shown) are statistically insignificant, as are coefficient on hours, hourly wage, and whether the employee works in the primary location. Only the coefficient on length of past tenure in the firm is statistically significant, though the estimate is quantitatively small. We also examine whether the causal effect of the default on plan enrollment differs by gender to investigate whether females are more susceptible to the default and we find no difference.[6]

In summary, we find a substantial effect of the default on plan choice: the default increases the probability of enrolling in one plan over the other by 60 percentage points. This effect is, to our knowledge, the first estimate of causal effect of the default in the context of the choice between DB and DC plans. While the magnitude of the estimated effect is comparable to prior work on

[5]Cross-validation methods are often used to find the optimal bandwidth in regression discontinuity estimation. However, because the function on either side of the discontinuity is approximately linear, the results are inherently less susceptible to the bandwidth choice (see Figure 1).

[6]These results are not reported but are available upon request.

defaults in the retirement savings literature, the effect estimated here is somewhat surprising given the potentially large differences in wealth accrual across the two plans and the irreversible nature of the decision (Beshears et al. 2006). In addition, the heterogeneity in the default across age provides a unique source of identification for estimating the causal effect of the default.

4 Solving for the Optimal Age-Based Default Rule

The estimates of the default shown in the previous section along with those found in prior literature indicate that defaults can be powerful tools to steer economic behavior. An important question that has been broached in the literature on defaults and in policy circles is how defaults may be constructed to improve welfare. In this section we explore: (1) how to set an optimal default based on observable characteristics; (2) properties of the optimal default as a function of known parameters; and (3) how to determine the value of incorporating heterogeneity into default provisions relative to alternative default policies.[7]

4.1 Characterizing retirement wealth in the DB and DC plan

We begin by characterizing the *future* benefits that accrue in both the DB and the DC plan.[8] The DB formula is defined by $b_j(w_j)$ for all years j between 0 and $r-a$, where r is the age of exit from the firm and a is the worker's current age, as a function of annual wages in year j, w_j. The function b_j is completely general in that it can allow for various types of accrual patterns.[9] The

[7]Note that the results in this section need not be limited to the particular firm in this paper; in theory, any firm's employee and pension plan characteristics could be utilized to obtain the optimal age-based default rule for that firm.

[8]At the firm in our setting, while the employee had also accrued years of service prior to the plan migration, the employee received that benefit stream regardless of the plan they were enrolled in after the transition. Therefore, the value of past benefits accrued are not included in wealth calculations for either the DB or DC plan and are not relevant for the comparison of future retirement wealth across the two types of plans.

[9]For example, a firm that offers 2 percent of average wages for each year of service as the annual retirement benefit would have $b_j(w_j) = 0.02w_j$ for all j. Similarly, a firm that offers its employees 2 percent $\times$ years of service $\times$ the average salary over the last five years at the firm as the annual retirement benefit would be characterized by $b_j(w_j) = 0$ if $j < r-a-5$ and $b_j(w_j) = 0.02(r-a)\frac{w_j}{5}$ if $j \geq r-a-5$.

wealth evaluated at retirement age ρ in the DB plan as a function of age a is then:

$$w^{DB}(a) = \int_0^{r-a} b_j(w_j) A_\rho dj, \tag{4}$$

where A_ρ is the actuarial present value of a stream of \$1 annual payments commencing at age ρ and paid until death.

The wealth in the DC plan at retirement age ρ is equal to the contributions made, accumulated with returns from investment experience. Contributions are typically a percentage of annual wages, though they can vary from year to year. We denote employer contributions into the employee's account in year j by c_j and the sequence of returns in all subsequent years by $\delta(k)$ for $k \in [j, \rho - a]$.[10] The wealth evaluated at retirement age ρ in the DC plan as a function of age a is then:

$$w^{DC}(a) = \int_0^{r-a} c_j w_j e^{\int_j^{\rho-a} \delta(k) dk} dj. \tag{5}$$

We next compute the expected discounted utility of retirement wealth for each plan by explicitly modeling two sources of uncertainty: separation risk, which affects the age of exit r, and investment risk, which affects the sequence of returns $\delta(\cdot)$.[11] We assume r and $\delta(\cdot)$ are drawn from a joint distribution $h^p(r, \delta | a < r \leq \bar{r})$ for plan $p \in \{DB, DC\}$. While separation risk affects the value of the DB and the DC plan, because DB wealth does not depend on $\delta(\cdot)$, investment risk only affects DC retirement wealth.[12]

[10] Note we assume c_j includes only employer contributions, not employee contributions. This distinction is made so both plans' values reflect the benefits provided by the employer only.

[11] We do not explicitly model other sources of risk, such as the risk that the employer goes bankrupt and can no longer honor DB pension obligations, or the risk of dying prior to retirement. However, it can easily be shown that the results presented here are general to any types of risk.

[12] In our numerical results that follow, we assume independence between separation risk and investment risk and that the distribution does not vary by plan choice. We assume r is drawn from a distribution $f(r|a < r \leq \bar{r}) = \frac{f(r)}{1-F(a)}$ where F is the cumulative distribution function of f and that the sequence $\delta(\cdot)$ is determined by a draw from a distribution $g(\delta)$. However, these assumptions do not affect our analytical results, which are all in terms of the certainty equivalent.

Assuming a discount rate d, the expected utilities for each plan are given by:

$$EU(w^{DB}(a)) = \int_a^{\bar{r}} \frac{U(w^{DB}(a))}{(1+d)^{\rho-a}} h^{DB}(r,\delta|a<r\leq\bar{r})dr \tag{6}$$

$$EU(w^{DC}(a)) = \int_{-\infty}^{\infty}\int_a^{\bar{r}} \frac{U(w^{DC}(a))}{(1+d)^{\rho-a}} h^{DC}(r,\delta|a<r\leq\bar{r})drd\delta. \tag{7}$$

We then define the certainty equivalent for plan $p \in \{DB, DC\}$ as:

$$CE^p(a) = U^{-1}(EU(w^p(a))). \tag{8}$$

The certainty equivalent $CE^p(a)$ for plan p is the amount that makes the individual indifferent between receiving the amount $CE^p(a)$ for certain and the gamble characterized by the uncertain income stream from plan p. Therefore, plan $\tilde{p}$ is preferable to plan $\hat{p}$ if and only if $CE^{\tilde{p}}(a) > CE^{\hat{p}}(a)$.

4.2 Default rules that maximize aggregate default wealth

We posit that the optimal default policy for workers is one that maximizes the "aggregate default wealth." The aggregate default wealth represents the certainty equivalent that each employee receives if he or she defaults, aggregated over all employees at the firm. Maximizing the aggregate default wealth is likely to improve aggregate welfare because the results in the previous section (and prior research on the effects of defaults) indicate that employees are likely to enroll in their default plan, either due to inertia, transaction costs, complexity, procrastination, or the perception of endorsement (Beshears et al. 2006). Further justification for this objective function comes from the fact that the cost of a worker enrolling in the plan that is not the default plan is low relative to the cost of optimizing over the choice of plans. Therefore, because workers who determine which plan is best can easily enroll in their optimal plan even if it is not their default, choosing a policy that maximizes the aggregate default wealth is likely to improve overall outcomes. In addition, this method does not require making assumptions of realized plan choices and setting the default option represents a lever that firms can use while preserving employee choice. While in this section, we focus on policies that are likely to maximize employee welfare, we recognize that firms may have alternative objective functions that take into account the relative costs of the two plans. We revisit this subject briefly in Appendix B.

We consider age-based default policies that take the following form for a given cutoff age $\tilde{a}$: (1) individuals younger than $\tilde{a}$ are defaulted into the DC plan, and (2) individuals older than $\tilde{a}$ are defaulted into the DB plan. We solve for the optimal age-based default policy that maximizes the aggregate default wealth under these constraints.

Define the minimum age among employees at the firm as $\underline{a}$ and the maximum age as $\bar{a}$. Formally, the optimal age-based default policy that satisfies conditions (1) and (2) above is defined by a^*, where a^* maximizes:

$$\begin{aligned} L(a^*) &= \int_{\underline{a}}^{a^*} CE^{DC}(a)da + \int_{a^*}^{\bar{a}} CE^{DB}(a)da \\ &= \int_{\underline{a}}^{\bar{a}} CE^{DB}(a)da + \int_{\underline{a}}^{a^*} \left(CE^{DC}(a) - CE^{DB}(a) \right) da. \end{aligned} \tag{9}$$

The first term of Equation (9) does not depend on a^*. Therefore we can solve for a^* as:

$$a^* = \arg\max_{\tilde{a}} \int_{\underline{a}}^{\tilde{a}} \left(CE^{DC}(a) - CE^{DB}(a) \right) da. \tag{10}$$

Solving the problem in Equation (10) yields the first-order condition for an interior solution:

$$H(a^*;\gamma) \equiv \left(CE^{DC}(a^*) - CE^{DB}(a^*) \right) = 0, \tag{11}$$

where γ is the vector of parameters $[\gamma_l]$ that define the solution a^*.

By the implicit function theorem, we can derive the comparative statics of a^* as follows:

$$\frac{\partial a^*}{\partial \gamma_l} = -\frac{H_{\gamma_l}(a^*;\gamma)}{H_{a^*}(a^*;\gamma)} = -\left.\frac{\frac{\partial CE(a)^{DC}}{\partial \gamma_l} - \frac{\partial CE(a)^{DB}}{\partial \gamma_l}}{\frac{\partial CE(a)^{DC}}{\partial a} - \frac{\partial CE(a)^{DB}}{\partial a}}\right|_{a=a^*}. \tag{12}$$

Because a^* maximizes $L(a^*)$, the denominator in Equation (12) is negative. Therefore,

$$\operatorname{sign}\left(\frac{\partial a^*}{\partial \gamma_l}\right) = \operatorname{sign}\left(\frac{\partial CE(a)^{DC}}{\partial \gamma_l} - \frac{\partial CE(a)^{DB}}{\partial \gamma_l}\right)\Bigg|_{a=a^*}. \tag{13}$$

Intuitively, Equation (13) shows that the optimal age-based default rule defined by a^* is increasing in the value of parameters that improve the relative value of the DC plan over the DB plan. For example, an increase in the generosity of the employer contribution to the DC plan would lead to an increase in the optimal cutoff age, thereby defaulting more employees into the DC plan. Similarly, a more generous DB benefit plan formula would lead to a decrease in the optimal cutoff age. More details on comparative statics can be found in Appendix C.

The ability to set the default based on an observable characteristic rather having than the same default for all employees at the firm is at worst welfare-neutral since the choice of setting the default to be equal for all employees (i.e., a^* equal to $\overline{a}$ or $\underline{a}$) is a choice available under the optimization procedure outlined above. But how much do employees gain from a heterogeneous default that differs by age? Moreover, how much would be gained by conditioning the default on additional observable characteristics?

We evaluate the performance of the optimal age-based default policy using two measures, N_π and $loss_\pi$, where π denotes one of three potential policies: $\{$universal DB default policy, universal DC default policy, optimal age-based default policy$\}$. Universal policies are important to consider not only for their ease of implementation, but because a universal DB default policy has been a common policy used by firms that have underwent a plan transition (e.g., the transitions evaluated by Papke (2004) and Yang (2005)) and a universal DC default may be attractive to firms who wish to minimize future DB liabilities. Define N_π to be the number of employees defaulted into a suboptimal plan under policy π, where

$$N_\pi \equiv \int_{\underline{a}}^{\overline{a}} \mathbf{1}_{[\overline{CE} > CE_\pi]} da \tag{14}$$

and $loss_\pi$ to be the average loss in certainty equivalent for these employees, relative to the certainty equivalent in their optimal plan:

$$loss_{\pi} \equiv \frac{\int_{\underline{a}}^{\overline{a}} \frac{\overline{CE}-CE_{\pi}}{\overline{CE}} da}{N_{\pi}} \tag{15}$$

where $\overline{CE} = \max(CE^{DB}, CE^{DC})$ and CE_{π} represents the certainty equivalent of the plan specified as the default under policy π. By comparing the value of N_{π} under the optimal age-based default policy to that under either universal policy, we can determine the default policy's ability to allocate employees into their optimal plan. In addition, the value of N_{π} under the optimal age-based default policy allows us to evaluate the potential gain from conditioning the default on other observable characteristics that affect the relative value of the two plans for employees.[13] Specifically, because a default based on all characteristics that influence the value of each plan would result in no employees defaulted into a suboptimal plan, if N_{π} is small, not many employees would be affected by conditioning the default on additional characteristics. The magnitude of the difference in the certainty equivalents for those defaulted into a suboptimal plan is encompassed in the measure $loss_{\pi}$.

It is important to note that our results implicitly assume that utility is additively separable in other sources of retirement wealth (for example, from past or future employer plans or personal savings accounts) and ignore the potential value of other benefits from DB or DC plans, such as the ability to borrow against DC plan balances. We also ignore any potential effects on other components of total compensation, for instance, the possibility that more workers enrolling in the plan that gives them higher benefits increases the cost to the employer, thereby causing the employer to reduce wages or other fringe benefits. Despite these limitations, our framework provides valuable insights into how heterogeneity may be introduced into default provisions.

5 Numerical Simulations of the Optimal Age-Based Default Rule

In this section we empirically evaluate the optimal age-based default using data from the large, non-profit firm that was the focus of the analysis in Section 3. This allows us to evaluate how

[13] In our numerical results that follow, these include gender and income.

the optimal cutoff age compares to the firm's chosen cutoff age of 45 and how the optimal cutoff age changes with values of relevant parameters.[14] We also quantify the value of conditioning the default on age relative to alternative default policies, such as defaulting all employees to the same plan or setting defaults that vary with additional characteristics.

5.1 Simulation methods and assumptions

We discretize the optimization procedure in the previous section and consider age-based default policies where a^* is an integer. We model uncertainty in r, the exit age from the firm, and uncertainty in the sequence of investment returns. Separation probabilities are assumed to follow a constant hazard rate by age, and summarized by probabilities $p_r(a)$ where $\sum_{r=a+1}^{\rho} p_r(a)$ for each a. To model investment risk ($\delta(k)$ from our analytical framework), we simulate S draws of investment return sequences using a Monte Carlo method outlined and used by Shoven and Sialm (1998, 2003) that draws series of asset returns for three different asset classes: stocks, bonds, and money market. The distribution of returns is assumed to be lognormal and draws are assumed to be serially-independent.[15] Our numerical results assume that separation risk and investment risk are independent.[16]

The assumptions used to simulate DB and DC retirement wealth are based on the known characteristics of the firm in our setting and are reported in Table 4. The firm's DB formula provides workers a stream of payments equal to a constant percentage of the employee's average wage for each year of service at the firm, so $b_j(w_j) = bw_j$ for all j, where the DB multiplier b is equal to 2 percent. The employer's contribution to the DC plan is comprised of 5 percent of the employ-

[14] We choose to illustrate the solution to the problem which maximizes employee welfare rather than the alternative objective functions that take into account firm costs (described in Appendix B) for two main reasons. First, we believe the firm in our setting had the objective of maximizing the welfare of existing employees at the time of the plan transition due to the anecdotal evidence and the non-profit nature of the firm. Second, solving for the optimal age cutoff using Equation (B-1) or (B-3) requires a method for calculating firm costs. Firm costs would include not only the present value of DB and DC benefits, but would also depend on the uncertainty in funding DB benefits, administrative costs to either plan, and the firm's cost of raising capital.

[15] For more details regarding the simulation methods, please see Appendix D.

[16] If we assume that these risks are correlated, perhaps because poor economic conditions reduce investment returns and increase involuntary separations, these two sources of risk would tend to counteract each other, as high separation risk reduces the relative value of the DB plan and low investment returns reduce the relative value of the DC plan.

ees' annual salary plus additional matching contributions, up to a total contribution of 10 percent. We model our results assuming the median match percentage of 3.5 percent of salary for a total employer contribution rate c equal to 8.5 percent.[17] In evaluating both the DB and DC retirement wealth, we assume real wages grow at a constant 2 percent per year, and a 2.5 percent rate of inflation. We assume a constant real discount rate d of 1 percent, and a constant separation hazard of 5 percent, taken from the data. The normal retirement age at the firm is 65, so we assume $\rho = 65$. Annuity values are taken from Social Security Administration mortality tables for males and females born in 1960 and assume a 2.9 percent interest rate, implying an annuity value of 14.48 for women and 13.15 for men.

The mean, standard deviations, and covariances for different asset classes used in the simulation are calibrated based on historical real returns reported by Ibbotson Associates (Ibbotson 2008) and summarized in the bottom panel of Table 4. We assume asset allocation follows the pattern of Fidelity Investments target-date retirement funds, the default fund allocation for DC participants at this firm.[18] Target-date funds have become increasingly popular default options among employer-sponsored savings accounts; Vanguard (2009) reports that 9 out of 10 plans with automatic enrollment have target-date funds as the default asset allocation. When illustrating the comparative statics of the optimal age-based default policy, we vary each of these parameters from their baseline values independently.

We use the standard constant relative risk aversion (CRRA) utility function to evaluate the expected discounted utility for each plan and to then translate it into a certainty equivalent for each individual as in Equation (8). The functional form of the utility function is given by:

$$U(w) = \frac{w^{1-\alpha}}{1-\alpha}, \tag{16}$$

where α is the measure of relative risk aversion. Estimates of α in the literature vary from 1

[17]In order to obtain a match percentage of 3.5 percent of salary, employees must contribute approximately 2.5% percent of their salary each year.

[18]We use the allocation for ages between 20 and 65 in 5-year intervals assumed by current funds to fit a fractional multinomial logit model with fourth order age terms in order to estimate an implied asset allocation across the three classes for ages between the 5-year intervals.

to 10, or higher depending on the context (e.g., Mehra and Prescott 1985; Kocherlakota 1996; Chetty 2006). Recent work by Goldstein, Johnson and Sharpe (2008) estimate $\alpha = 6.1$ from an experimental setting in which participants indicated their preferences for a distribution of future retirement income given a cost constraint. In our empirical analysis we vary α from 0 to 10 to evaluate how a^* changes with the level of risk aversion among employees at the firm.

5.2 Main simulation results

The panels in Figure 4 show the certainty equivalent by age for each pension plan across different levels of risk aversion. The absolute levels of the certainty equivalent are decreasing in the relative risk aversion, ranging from an average of \$112,106 (DC) and \$98,390 (DB) for $\alpha = 0$ to \$7,261 (DC) and \$18,015 (DB) for $\alpha = 10$. For lower levels of risk aversion, the certainty equivalent is higher in the DC plan for younger workers as shown in panels (a) and (b) in Figure 4. For the highest level of risk aversion shown (panel (d)), the certainty equivalent from the DB plan exceeds the DC plan for all ages.

As outlined in Section 4, we define the optimal cutoff age a^* for the simple, age-based default as the one which maximizes the aggregate default wealth (Equation (9)). We use our baseline parameter assumptions to solve for a^* for different levels of risk aversion. The objective function in Equation (9) is plotted against age for different values of α in the panels of Figure 6. The age at which the objective function is maximized is a^*, and is reported in Table 6 for different values of α. The optimal cutoff value ranges from 44 for $\alpha = 2$ to 20 for $\alpha = 10$, and is non-monotonic in nature due to the opposing sources of risk: the presence of investment risk implies that more risk-averse individuals would prefer the DC plan, but separation risk affects the value of the DB plan more than the DC plan and thus pushes the optimal cutoff age in the opposite direction. For levels of risk aversion less than 4, the optimal age cutoff is between 42 and 47, largely in line with the firm's chosen age cutoff of 45. Therefore, our simulation results indicate that the firm's cutoff age is close to optimal for levels of risk aversion less than 4 under the assumption that their objective was to maximize employee utility. The lowest value for α such that a universal DB default is optimal is approximately equal to 8. Therefore, we find that for a sizable range of risk aversion, an age-based default rule produces a higher aggregate default wealth than a universal DB default under baseline assumptions.

5.3 Sensitivity of Cutoff Age to Alternative Assumptions

In Table 6 we illustrate how the optimal age cutoff changes with different assumptions regarding pension plan characteristics, asset allocation, separation hazards, investment risk, wage growth, the discount rate, and the rate of inflation. Our results confirm the general predictions given in Section 4: changes in the parameter values that increase the value of the DC plan relative to the DB plan increase the optimal cutoff age.

We find that increasing the generosity of the DC contribution rate or reducing the generosity of the DB plan increases the optimal cutoff age across all values of α. The optimal cutoff age depends on the asset allocation chosen in the DC plan such that when the portfolio is invested solely in stocks, the riskiest asset class, the optimal cutoff age is higher for lower levels of risk aversion and is lower for higher levels of risk aversion. Analogously, investing only in bonds reduces the optimal cutoff age; however, a^* does not vary as dramatically across different levels of risk aversion. Under the assumption that DC participants are invested solely in low-yielding risk-free assets, the DB plan performs better even under high levels of risk aversion.

Eliminating separation risk entirely illustrates that isolating investment risk produces a monotonically decreasing optimal cutoff age. Similarly, eliminating investment risk while maintaining the base assumption for separation risk yields a monotonically increasing optimal cutoff age. As separation risk increases, the certainty equivalent in the DB plan is reduced more than the certainty equivalent in the DC plan, and, therefore, the optimal cutoff age increases, defaulting additional employees into the DC plan. If the standard deviations of the three asset classes are doubled, the optimal age cutoff increases for low levels of risk aversion and decreases for high levels of risk aversion relative to the baseline.

The optimal cutoff age is weakly decreasing in real wage growth, indicating that higher wage growth increases the relative value of DB plan benefits over DC plan benefits, though the effect is small. The real discount rate does not affect the relative values of the DB and DC certainty equivalents; therefore, the optimal cutoff age does not change with respect to the real discount rate. The inflation rate assumption affects the optimal cutoff age because inflation differentially affects the value of the DB relative to the DC plan. DB plan participants' benefits are based on a formula that includes their nominal wages, from what may be many years prior to retirement. By contrast, early contributions in DC accounts are invested in markets that implicitly adjust for inflation. High inflation therefore reduces the value of DB benefits relative to DC benefits and increases the optimal cutoff age.[19]

In our analysis, we do not explicitly value the differences in death benefits and vesting requirements between the two plans. In our firm, the DB plan provides a survivor benefit to a named beneficiary equal to 50 percent of the accrued retirement benefit, while the DC plan provides the survivor the worker's entire DC account. This makes the DB plan less attractive for all workers, with a greater differential for younger workers who have a higher likelihood of dying at some point before retirement relative to older workers who are closer to retirement. Similarly, because DB benefits vest after 5 years of service while employer contributions to DC accounts vest after only 1 year, incorporating vesting requirements into our analysis would have an analogous effect, raising the relative value of the DC plan for all workers and particularly for younger workers who are more likely to have less tenure at the firm. The combination of these effects would slightly increase the optimal cutoff age, defaulting more workers into the DC plan.

5.4 Evaluation of Optimal Age-Based Default Policy

We quantify the gain from an age-based default policy using the measures N_π and $loss_\pi$ as outlined in Equations (14) and (15) and report the results in Table 7. We find that under the optimal

[19] Note that the low sensitivity to wage growth and the higher sensitivity to the inflation rate is likely due to the fact that our firm's DB benefit formula uses wages throughout a worker's career rather than common alternatives, such as a formula that bases benefits on final average salary. This type of formula would be more likely to protect DB participants from inflation and would reward higher wage growth more than a DC plan. On the other hand, the relative effect of separation risk on the DB plan is likely to be greater with a DB formula that uses a final average salary.

age-based default, at most 9 employees (one percent) are allocated to a suboptimal plan with an average loss in certainty equivalent of four percent or less. In comparison, a universal DC default misclassifies at least 50 percent of employees into a suboptimal plan with an average loss in certainty equivalent of approximately 34 percent. A universal DB default performs notably better than the universal DC, particularly for higher values of risk aversion, but is inferior to the optimal age-based policy until the two policies coincide (i.e. when risk aversion levels are approximately equal to 8 or higher). Furthermore, because the optimal age-based policy allocates 99 percent of employees to the plan that provides their higher certainty equivalent, conditioning the default policy on other observable characteristics that affect the relative value of the two plans, namely gender and income, would have a negligible effect on plan outcomes.[20] Because employees are predisposed to choosing the default plan, as shown in Section 3, it is likely that a heterogeneous default would lead to a substantial increase in welfare. It is important to note that these loss estimates are likely an upper bound as employees can always elect out of the default plan if it is suboptimal.

One limitation of the measure N_π is that it masks the composition of employees who get defaulted into the plan with the lower certainty equivalent. In particular, we find that all of the workers who lose from the optimal default policy are women, and the majority are workers with below-median income. This finding is likely due to the low weight that women and lower income workers receive in the aggregate default wealth due to their smaller numbers (in the case of women) or lower certainty equivalents (for lower-income workers). Modifying the objective function to place more weight on these potentially more vulnerable groups could address this issue.

Thus far, our analysis has presumed that risk aversion is both known and homogenous across employees. However, these assumptions are tenuous given that risk aversion is not a readily observable characteristic. Moreover, the DB and DC certainty equivalents are sensitive to the employees' level of risk aversion. Therefore, we briefly consider how the benefit of an age-based default policy relative to a universal policy is affected by: (1) misestimating the level of risk aversion; and (2) heterogeneity in risk aversion across employees.

[20] Legal restrictions that prevent treating employees differently by gender and income may also prevent employers from conditioning the default on these factors.

In the top panel of Table 8, we simulate N_π and $loss_\pi$ under the assumption that the employer misestimates the level of risk aversion by one when the true level of relative risk aversion is $\alpha = 5$. These results can be compared to those found in Table 7 when the firm correctly assumes that $\alpha = 5$. If the firm underestimates risk aversion by one and assumes $\alpha = 4$, then the firm would set the optimal cutoff age to be 42 rather than 36, resulting in 144 employees being classified into their suboptimal plan. If the firm overestimates the parameter by one, and believes $\alpha = 6$, then the procedure outlined previously would choose the optimal cutoff age as 25, resulting in the misclassfication of 120 employees. When we compare this misclassification to that under the universal default policy, we see that even when risk aversion is misestimated by one, an age-based default policy performs better than the policy of defaulting all employees into the DC plan. In fact, because the optimal cut-off age is less than 65 for all non-negative values of risk aversion (see Table 6), the age-based policy is always superior to a universal DC default regardless of the assumed value of risk aversion. When we compare the misclassification to that under the universal DB default, we see that, in this case, misestimating α by one results in an outcome that is comparable to the degree of misclassification when all employees are defaulted into the DB plan (144 and 120 versus 143), but the average loss in certainty equivalent is less relative to the universal DB default (9.9% and 13.2% versus 15.4%). Overall, misestimating the level of risk aversion defaults more employees into suboptimal plans relative to when risk aversion is known.

Next, we consider heterogeneity in risk aversion across employees. To illustrate the sensitivity of the results to different types of heterogeneity, we assume an average value of α equal to five, and we model the heterogeneity in two ways. First, we model risk aversion as randomly distributed across employees using a discrete probability distribution where α takes on one of three values (3, 4, or 5) with equal probability. Second, we model risk aversion as a deterministic linear function of age, $\alpha_i = f(a)$, based on Barsky, Juster, Kimball and Shapiro (1997) and Kimball, Sahm and Shapiro (2009), which find evidence of a positive correlation between age and risk aversion.

When risk aversion is randomly distributed, this form of heterogeneity increases the optimal cutoff age from 36 when $\alpha = 5$ for all employees to age 39, and results in the misclassification of

118 employees into their suboptimal plan. When risk aversion is tied to age, the optimal cutoff age is also 39 and just 19 employees are misclassified into their suboptimal plan. We conclude that heterogeneity in risk aversion across employees defaults more employees into their suboptimal plan; however, assuming that the variation in risk aversion is due to age reduces the number of employees in their suboptimal plan and the average losses incurred by these individuals. Because age is a strong predictor of the relative value of the two plans, assuming risk aversion is a monotonic function of age results in a substantially smaller number of misclassified employees than assuming risk aversion is more random (i.e. where there may be younger employees with high levels of risk aversion who would prefer the DB plan and older employees with low levels of risk aversion who would prefer the DC plan). We do not consider heterogeneity in risk aversion over other observable characteristics, namely gender, because the literature has failed to find consistent evidence that risk aversion is strongly related to gender (see Arano, Parker and Terry (2010) for review).

6 Conclusion

The idea that individual choices can depend on institutional arrangements, such as the default provisions that dictate what happens when an individual fails to make an active decision, is now widespread in economics. This idea has given rise to a notion of "libertarian paternalism," an approach that encourages private and public bodies to steer individuals to choices that are likely to make them better off while still preserving freedom of choice (Thaler and Sunstein 2003). The current study provides strong evidence that a firm can significantly influence the choice of a pension plan by its decision of which plan is the default: our estimates indicate that employees who were subject to a DC plan default were 60 percentage points more likely to enroll in a DC plan relative to a DB plan. The default was an overwhelming determinant of plan enrollment, alone explaining a substantial amount of the variability in plan choice. This evaluation differs from past studies of defaults in that we examine a one-time, irreversible decision with potentially substantial retirement wealth at stake and we are able to use a regression discontinuity framework to estimate the causal effect of the default rule.

We also develop a framework for finding the optimal age-based default rule for pension plan choice. We solve for the age cutoff that maximizes the aggregate default wealth, or the certainty equivalent that each employee receives upon failing to make an active decision, aggregated across all employees. For the firm in our setting, we numerically simulate the optimal cutoff age and find that over a broad range of plausible levels of risk aversion, a heterogeneous default “nudges” employees into plans with a higher risk-adjusted value relative to a universal policy of defaulting all employees into the DB plan, a common default policy used by firms transitioning from a DB plan to a DC plan, or the DC plan. Our results also suggest that age is an important determinant of the relative value of a DC plan over a DB plan, and that incorporating observable characteristics in addition to age into the determination of the default would not significantly alter employee outcomes. We also demonstrate how the optimal age-based default rule varies with different levels of risk aversion, asset allocations strategies, and plan characteristics.

It is important to note that the relative values of the two plans are sensitive to the level and distribution of risk aversion among employees at the firm. If employers systematically underestimate or overestimate the level of risk aversion among their employees when constructing an optimal age-based default policy, or if there is substantial heterogeneity in risk aversion across employees, more employees would be defaulted into a suboptimal plan. We find that the optimal age-based default policy is superior to a universal DC default regardless of the assumed level or distribution of risk aversion, but that the benefit of an age-based default policy relative to a universal DB default policy decreases if risk aversion is misestimated or if it is randomly distributed across employees. While incorporating risk aversion into a heterogeneous default policy would produce outcomes that are welfare-enhancing, such an approach may be difficult to implement given that risk aversion is difficult to observe. However, there have been advances in survey methods to assess risk aversion that could be utilized by firms to learn about their employees’ levels of risk aversion. Furthermore, we show that the optimal age cutoff is not sensitive to incorporating heterogeneity in risk aversion deterministically across age.

Overall, our results suggest that substantial welfare gains are possible by varying defaults by observable characteristics. In particular, when a set of observable characteristics strongly predicts the value of one choice over others, conditioning a default on these characteristics can "nudge" decision-makers into choices that are likely to raise their expected utility.

References

Arano, Kathleen, Carl Parker, and Rory L. Terry, "Gender-Based Risk Aversion and Retirement Asset Allocation," *Economic Inquiry*, 2010.

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, 1997, *112* (2), 537–579.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian, "The Importance of Default Options for Retirement Savings Outcomes: Evidence from the United States," *NBER Working Paper No. 12009*, 2006.

Bodie, Zvi, Alan J. Marcus, and Robert C. Merton, "Defined Benefit versus Defined Contribution Pension Plans. What Are the Real Trade-offs?," in Zvi Bodie, John Shoven, and David Wise, eds., *Pensions in the U.S. Economy*, Chicago, IL: University of Chicago Press, 1988.

Brown, Jeffrey R. and Scott J. Weisbenner, "Who Chooses Defined Contribution Plans?," *NBER Working Paper No. 12842*, 2007.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Mertrick, "Optimal Defaults and Active Decisions," *Quarterly Journal of Economics*, 2009, *124* (4), 1639–74.

Chetty, Raj, "A New Method of Estimating Risk Aversion," *American Economic Review*, 2006, *96* (5), 1821–1834.

Choi, James J., David Laibson, and Brigitte C. Madrian, "Are Empowerment and Education Enough? Underdiversification in 401(k) Plans," *Brookings Papers on Economic Activity*, 2005, (2), 151–198.

____ , ____ , ____ , and Andrew Metrick, "For Better or For Worse: Default Effect and 401(k) Savings Behavior," in David A. Wise, ed., *Perspectives in the Economics of Aging*, Chicago, IL: University of Chicago Press, 2004.

Clark, Robert L. and M. Melinda Pitts, "Faculty Choice of Pension Plan: Defined Benefit versus Defined Contribution," *Industrial Relations*, 1999, *38* (1), 18–45.

____ , Linda S. Ghent, and Ann A. McDermed, "Pension Plan Choice among University Faculty," *Southern Economic Journal*, 2006, *72* (3), 560–577.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, *47* (2), 315–372.

GAO, "Defined Benefit Pensions: Plan Freeze Affect Millions of Participants and May Post Retirement Income Challenges," Report Number: GAO-08-817, General Accounting Office, Washington DC 2008.

Goldstein, Daniel, Eric J. Johnson, and William F. Sharpe, "Choosing Outcomes versus Choosing Products: Consumer-Focused Retirement Investment Advice," *Journal of Consumer Research*, October 2008, *35*, 440–456.

Handel, Benjamin, "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," mimeo., Northwestern University 2009.

Ibbotson, *Stocks, Bonds, Bills, and Inflation 2008 Yearbook*, Chicago, IL: Ibbotson Associates, 2008.

Imbens, Guido and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

Johnson, Eric J. and Daniel Goldstein, "Do Defaults Save Lives?," *Science*, 2003, *302*, 1338–1339.

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro, "Risk Preferences in the PSID: Individual Imputations and Family Covariation," *American Economic Review*, May 2009, *99* (2), 363–368.

Kocherlakota, Narayana R., "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 1996, *34* (1), 42–71.

Kruse, Douglas L., "Pension Substitution in the 1980s: Why the Shift toward Defined Contribution?," *Industrial Relations*, 1995, *34* (2), 218–241.

Lachance, Marie-Eve, Olivia S. Mitchell, and Kent Smetters, "Guaranteeing Defined Contribution Pensions: The Option to Buy Back a Defined Benefit Promise," *Journal of Risk and Insurance*, 2003, *70* (1), 1–16.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *NBER Working Paper No. 14723*, 2009.

Madrian, Brigitte C. and Dennis F. Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 2001, *116* (4), 1149–1525.

Manchester, Colleen F., "The Effect of Pension Plan Type on Retirement Age: Distinguishing Plan Incentives from Preferences," *Southern Economic Journal*, 2010, *77* (1), 104–125.

Mehra, Rajnish and Edward C. Prescott, "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 1985, *15*, 145–161.

Milevsky, Moshe A. and S. David Promislow, "Florida's Pension Election: From DB to DC and Back," *Journal of Risk and Insurance*, 2004, *71* (3), 381–404.

Mitchell, Olivia S., Gary R. Mottola, Stephen P. Utkus, and Takeshi Yamaguchi, "Default, Framing, and Spillover Effects: The Case of Lifecycle Funds in 401(K) Plans," NBER Working Paper 15108, National Bureau of Economic Research 2009.

Papke, Leslie, "Are 401(k) Plans Replacing Other Employer-Provided Pensions? Evidence from Panel Data," *Journal of Human Resources*, 1999, *34* (2), 346–369.

____ , "Pension Plan Choice in the Public Sector: The Case of Michigan State Employees," *National Tax Journal*, 2004, *57* (2), 329–339.

Shoven, John B. and Clemens Sialm, "Long Run Asset Allocation for Retirement Savings," *Journal of Private Portfolio Management*, 1998, pp. 13–26.

____ **and** ____ , "Asset Location in Tax-Deferred and Conventional Savings Accounts," *Journal of Public Economics*, 2003, *88*, 23–38.

Thaler, Richard H. and Cass R. Sunstein, "Libertarian Paternalism," *American Economic Review*, May 2003, *93* (2), 175–179.

Towers-Perrin, "Back to the Future: Redefining Retirement in the 21st Century," 2003.

Vanguard, "How American Saves 2009," 2009.

Yang, Tongxuan, "Understanding the Defined Benefit versus Defined Contribution Choice," Pension Research Council Working Paper Series 2005-04, The Wharton School, University of Pennsylvania 2005.

Table 1: Distribution of Age Group by Choice (*Active vs. Passive Choice*)

Age Group (*default*)	Passive Choice	Active Choice	
		Consistent with Default	Inconsistent with Default
Under 45 (*DC Plan*)	194	162	41
45 or Over (*DB Plan*)	308	134	86
Total	502	296	127

Notes: Sample restricted to employees less than 65 years of age. N=925.

Table 2: Summary Statistics: Employee Characteristics

Variable	Mean	Std. Dev.	Min	Max
Enrolled in DC Plan	0.478	0.500	0	1
Made Passive Choice	0.543	0.498	0	1
Age	46.07	9.72	21.88	64.96
Female	0.188	0.391	0	1
White	0.425	0.495	0	1
Black	0.114	0.317	0	1
Hispanic	0.303	0.460	0	1
Other Race/Ethnicity	0.159	0.366	0	1
Weekly Hours	39.35	3.27	20.00	55.00
Hourly Wage	23.98	6.63	10.24	36.81
Tenure (years)	12.08	9.12	0.75	43.41
Primary Work Location	0.699	0.459	0	1

Notes: Sample restricted to employees less than 65 years of age. N=925.

Table 3: Regression Discontinuity Estimates of Default Effect on Plan Choice (*Linear Probability Model*)

	(1)	(2)	(3)	(4)	(5)
Under 45	0.605***	0.576***	0.578***	0.777***	0.801***
	(0.042)	(0.080)	(0.081)	(0.141)	(0.143)
(Age - 45)		-0.017	-0.011	0.369**	0.437**
		(0.020)	(0.020)	(0.181)	(0.179)
$(\text{Age - 45})^2$				-0.169**	-0.198**
				(0.085)	(0.084)
$(\text{Age - 45})^3$				0.020*	0.024**
				(0.011)	(0.011)
(Age - 45)$\times$Under 45		0.025	0.023	-0.4	-0.462*
		(0.029)	(0.029)	(0.266)	(0.261)
$(\text{Age - 45})^2\times$Under 45				0.129	0.167
				(0.130)	(0.130)
$(\text{Age - 45})^3\times$Under 45				-0.028	-0.030*
				(0.018)	(0.018)
Hours			-0.004		-0.005
			(0.005)		(0.006)
Hourly Wage			0.005		0.006
			(0.005)		(0.005)
Tenure			-0.006**		-0.007**
			(0.003)		(0.003)
Primary Work Location			0.068		0.075
			(0.065)		(0.065)
Constant	0.227***	0.272***	0.356	0.077	0.113
	(0.030)	(0.061)	(0.268)	(0.101)	(0.302)
R^2	0.359	0.357	0.365	0.358	0.368
N	353	353	353	353	353

Notes: Dependent variable is enrollment in the DC plan. Regression discontinuity estimate is coefficient on “Under 45,” which estimates the change in DC plan enrollment at the age cutoff. Robust standard errors in parentheses. Bandwidth of 5 years. Columns (3) and (5) also include gender and race binary variables. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4: Baseline Assumptions for Optimal Age-Based Default

	Assumption	
Plan Characteristics:		
DB Multiplier (b)	2.0%	
DC Contribution Rate (c)	8.5%	
Other Parameters:		
Real Wage Growth Rate (g)	2.0%	
Real Discount Rate (d)	1.0%	
Separation Hazard ($\Rightarrow p_r^a$)	5.0%	
Inflation (i)	2.5%	
Real Asset Returns:	μ	σ
Stocks	6.4%	18.8%
Bonds	2.7%	9.2%
Money Market	0.7%	3.9%
Asset Covariances:		
Stocks-Bonds	0.4065%	
Bonds-Money Market	0.2033%	
Money Market-Stocks	0.0763%	

Table 5: Age Cutoff under Optimal Age-Based Default Rule

	(1) $\alpha = 0$	(2) $\alpha = 2$	(3) $\alpha = 5$	(4) $\alpha = 10$
Optimal age cutoff	44	47	36	20

Notes: The optimal age cutoff is the integer value that maximizes Equation (9) for different levels of risk aversion under the baseline assumptions (as shown in Table 4).

Table 6: Comparative Statics for Optimal Age-Based Default Rule

Assumptions	$\alpha = 0$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
Baseline	44	47	36	20
5% DC Contribution Rate	36	40	20	20
10% DC Contribution Rate	47	49	42	20
1% DB Multiplier	56	57	56	30
3% DB Multiplier	38	42	20	20
100% Stocks	48	48	20	20
100% Bonds	32	39	33	20
100% Cash	20	26	23	20
0% Separation Hazard	40	36	23	20
10% Separation Hazard	46	48	36	20
No Investment Risk	40	47	50	50
Double Investment Risk	50	46	20	20
0% Real Wage Growth	45	47	36	20
4% Real Wage Growth	43	47	36	20
0% Real Discount Rate	44	47	36	20
2% Real Discount Rate	44	47	36	20
1.5% Inflation	42	45	20	20
3.5% Inflation	46	49	42	20

Notes: Each row gives the optimal cutoff age, a^*, for different levels of risk aversion for deviations from the baseline assumptions along one parameter dimension. Baseline assumptions are shown in Table 4.

Table 7: Measures to Evaluate Optimal Age-Based Default

α	Policy	Age Cutoff	N_π	$loss_\pi$
$\alpha = 0$	Universal DB Default	20	344	38.0%
	Universal DC Default	65	581	34.7%
	Optimal Age-Based Default	44	9	4.0%
$\alpha = 2$	Universal DB Default	20	462	41.8%
	Universal DC Default	65	463	34.0%
	Optimal Age-Based Default	47	6	3.6%
$\alpha = 5$	Universal DB Default	20	143	15.4%
	Universal DC Default	65	782	33.8%
	Optimal Age-Based Default	36	9	1.0%
$\alpha = 10$	Universal DB Default	20	–	–
	Universal DC Default	65	925	59.5%
	Optimal Age-Based Default	20	–	–

Notes: N_π denotes the number of employees defaulted into a suboptimal plan across different levels of risk aversion for three alternative policies. $loss_\pi$ denotes the average loss in certainty equivalent among the N_π individuals defaulted into a suboptimal plan, as a percentage of the higher plan's certainty equivalent. Universal DB default assumes that all employees are defaulted into the DB plan, universal DB default assumes that all employees are defaulted into the DC plan, and optimal age-based default denotes solution to maximizing Equation (9).

Table 8: Measures to Evaluate Optimal Age-Based Default

Panel A: Misestimation of Risk Aversion

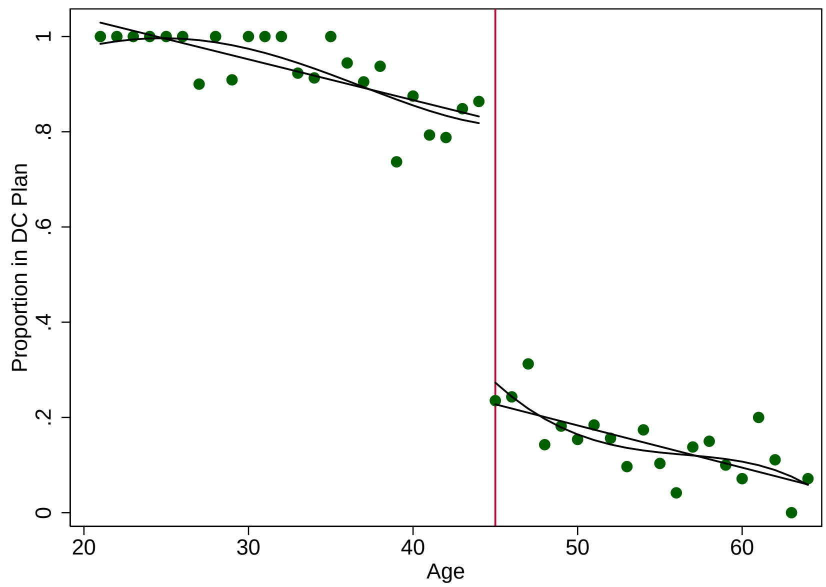
True α	Assumed α	Policy	Age Cutoff	N_π	$loss_\pi$
$\alpha = 5$	$\alpha = 4$	Optimal Age-Based Default	42	144	9.9%
$\alpha = 5$	$\alpha = 6$	Optimal Age-Based Default	25	120	13.2%

Panel B: Heterogeneity in Risk Aversion

α	Policy	Age Cutoff	N_π	$loss_\pi$
$\alpha_i \sim$ iid	Universal DB Default	20	141	23.8%
p(4)=p(5)=p(6)=1/3	Universal DC Default	65	784	34.8%
	Optimal Age-Based Default	39	118	14.2%
$\alpha_i = f(a)$	Universal DB Default	20	204	23.3%
$\alpha_i \in [4,6]$ with $\overline{\alpha}$=5	Universal DC Default	65	721	36.2%
	Optimal Age-Based Default	39	19	5.4%

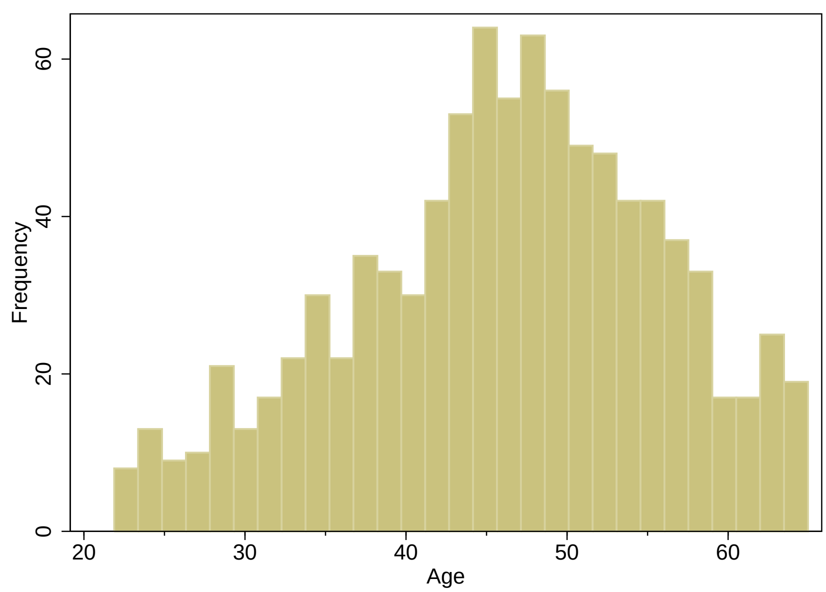
Notes: Panel A shows N_π and $loss_\pi$ when risk aversion parameter is under- or over-estimated by 1. Panel B shows N_π and $loss_\pi$ when risk aversion varies randomly across employees or varies deterministically as a function of age.

Figure 1: DC Enrollment Rate by Age



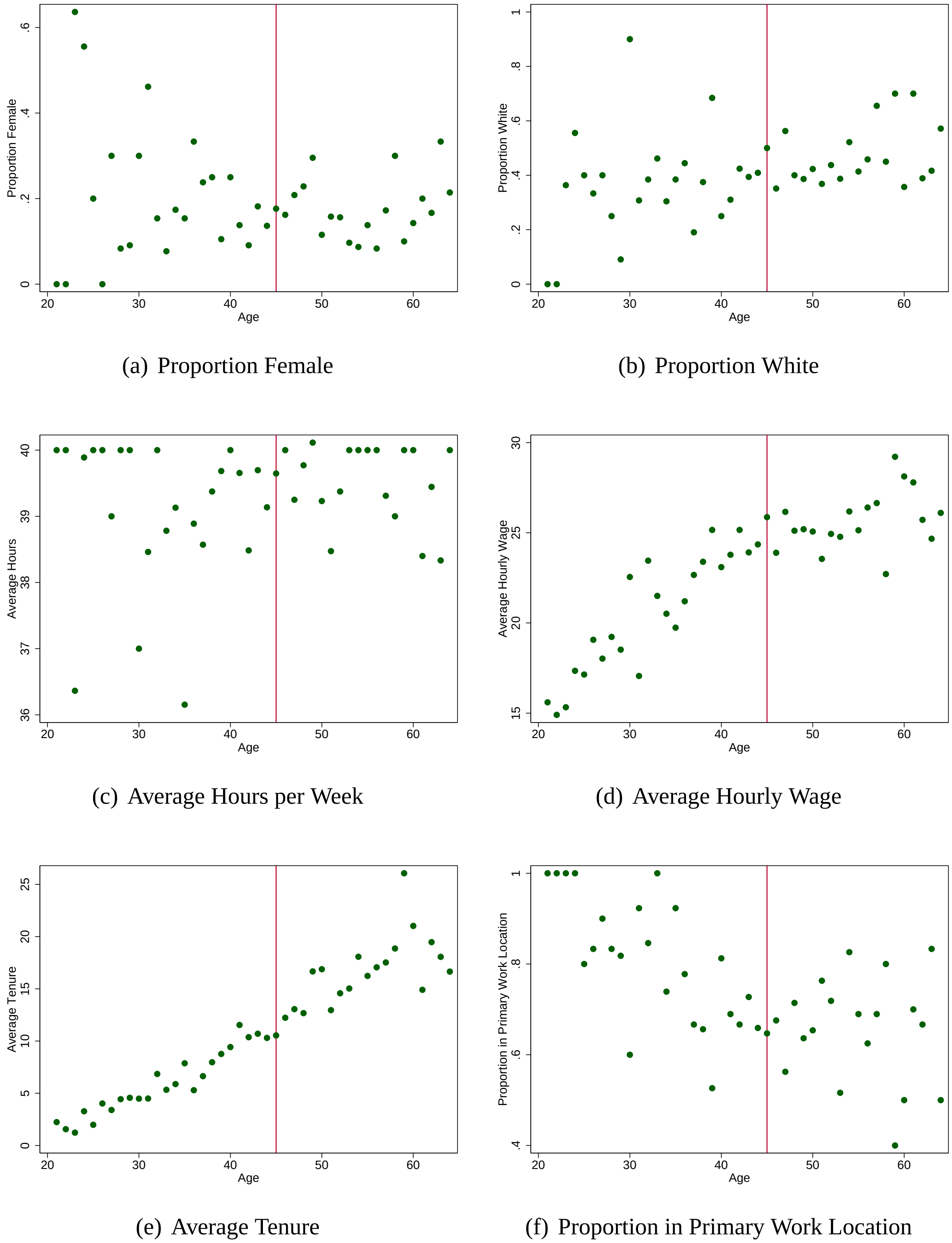
Notes: DC enrollment rates computed for each single year of age. Best fit line and cubic are shown for each side of the age 45 cutoff.

Figure 2: Distribution of Employee Age



Notes: Histogram of employee age as of September 1, 2002, using one-year bins.

Figure 3: Average Value of Covariates by Single Year of Age



(a) Proportion Female

(b) Proportion White

(c) Average Hours per Week

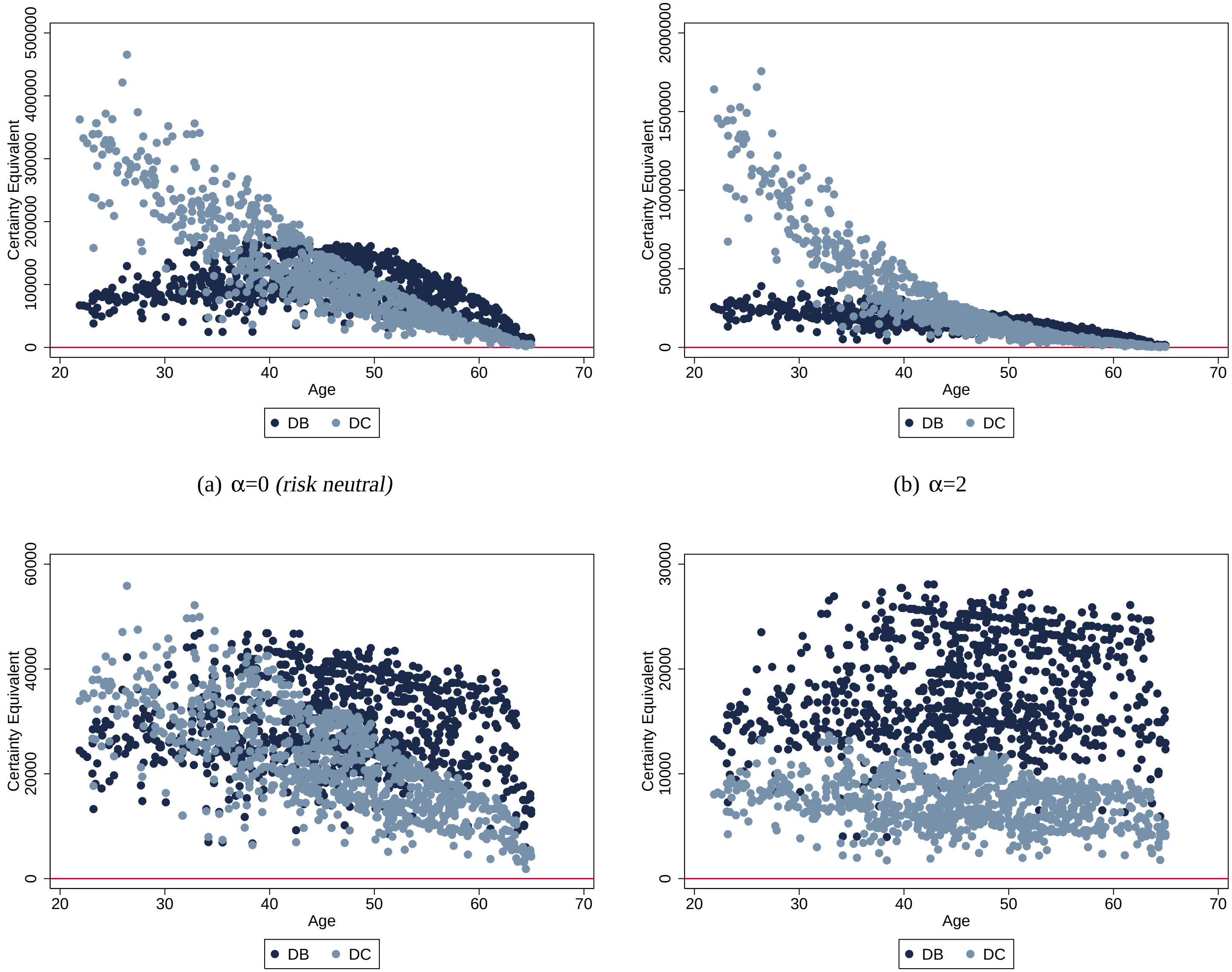
(d) Average Hourly Wage

(e) Average Tenure

(f) Proportion in Primary Work Location

Notes: Panels used to verify no discontinuity at age 45 for other observable characteristics; vertical line marks age 45.

Figure 4: Certainty Equivalent by Age for Different Levels of Risk Aversion



(a) α=0 *(risk neutral)*

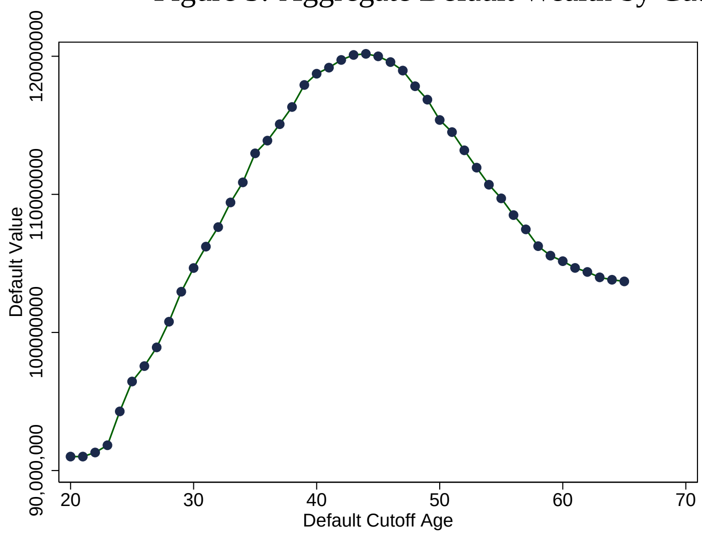
(b) α=2

(c) α=5

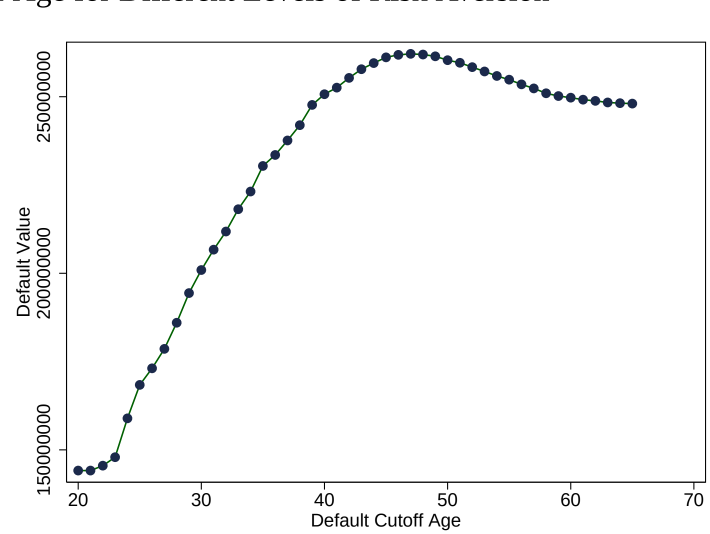
(d) α=10

Notes: Each panel shows the certainty equivalent for each plan by age for different levels of risk aversion (α).

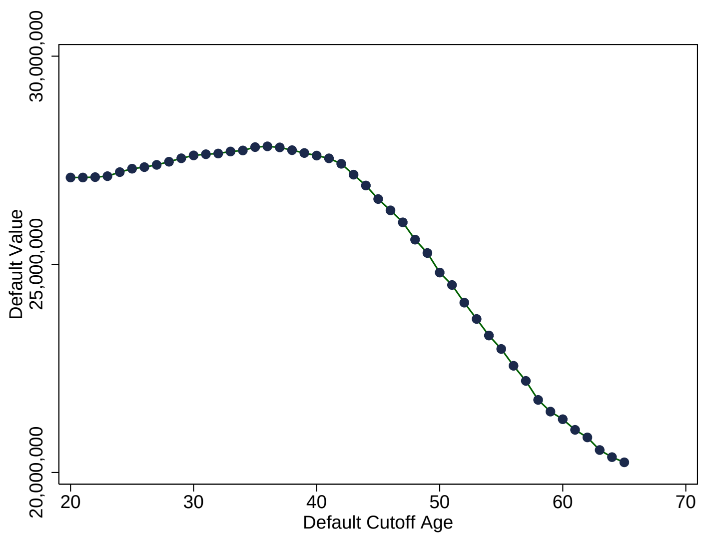
Figure 5: Aggregate Default Wealth by Cutoff Age for Different Levels of Risk Aversion



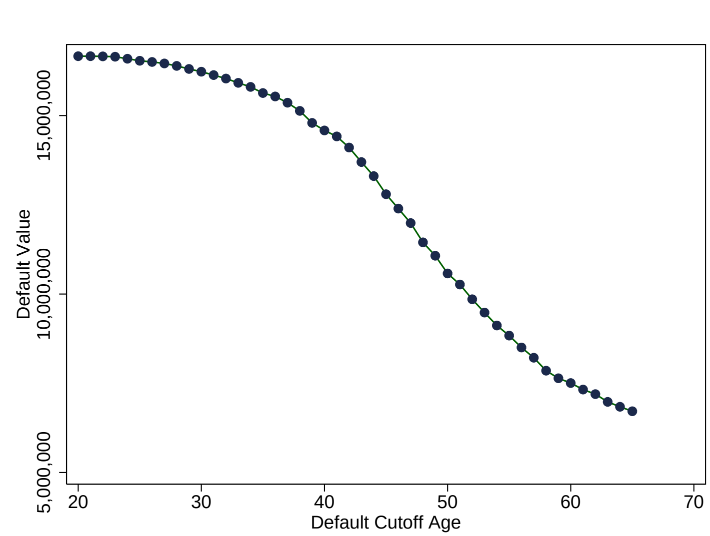
(a) α=0 *(risk neutral)*



(b) α=2



(c) α=5



(d) α=10

Notes: Each panel plots the aggregate default wealth by cutoff age for different levels of risk aversion (α). The optimal cutoff age maximizes the aggregate default wealth.

Appendix A: Robustness to Choice of Bandwidth

The regression discontinuity results shown in Table 3 were estimated with a bandwidth of 5 years. The table below summarizes linear and probit estimates of the regression discontinuity for different values of the bandwidth across different specifications.

	(1)	(2)	(3)	(4)	(5)
LPM	0.584***	0.690***	0.703***	0.739***	0.737***
($h = 2.5$)	(0.058)	(0.108)	(0.112)	(0.196)	(0.200)
LPM	0.636***	0.544***	0.532***	0.697***	0.709***
($h = 7.5$)	(0.035)	(0.066)	(0.066)	(0.122)	(0.125)
LPM	0.665***	0.540***	0.538***	0.685***	0.694***
($h = 10$)	(0.030)	(0.059)	(0.059)	(0.108)	(0.110)
Probit	0.584***	0.687***	0.718***	0.728***	0.741***
($h = 2.5$)	(0.058)	(0.099)	(0.096)	(0.172)	(0.171)
Probit	0.605***	0.581***	0.601***	0.758***	0.784***
($h = 5$)	(0.042)	(0.082)	(0.083)	(0.116)	(0.110)
Probit	0.636***	0.538***	0.542***	0.679***	0.708***
($h = 7.5$)	(0.035)	(0.071)	(0.074)	(0.114)	(0.114)
Probit	0.665***	0.527***	0.547***	0.683***	0.722***
($h = 10$)	(0.030)	(0.065)	(0.069)	(0.101)	(0.098)
Treatment of Age	Constant	Linear	Linear	Cubic	Cubic
Controls	No	No	Yes	No	Yes

Notes: Dependent variable is enrollment in the DC plan. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. LPM represents linear probability models. Probit models report marginal effects.

Appendix B: Alternative Objective Functions

Section 4 assumed that the firm's objective was to maximize employee welfare. As an alternative, suppose firms operate in a two-stage manner where they choose how much to spend on deferred compensation in the first stage and then choose a policy to maximize employee welfare subject to this constraint in the second stage. Our maximization problem in Equation (10) can be easily modified to accommodate this optimization strategy as follows:

$$a^{**} = \arg\max_{\tilde{a}} \int_{\underline{a}}^{\tilde{a}} \left(CE^{DC}(a) - CE^{DB}(a) \right) da \tag{B-1}$$

subject to the firm budget constraint:

$$\int_{\underline{a}}^{a^{**}} FC^{DC}(a)da + \int_{a^{**}}^{\overline{a}} FC^{DB}(a)da \leq B, \tag{B-2}$$

where $FC^p(a)$ denotes net firm costs for plan p for a worker age a and B is the firm's budget for deferred compensation. Net firm costs could include the present value of future retirement benefits offset by any benefit fueled by the retirement plan, e.g. differences in turnover costs across the two plans. This modified problem is equivalent to the original problem if the budget constraint is not binding, i.e., $B \geq \int_{\underline{a}}^{a^*} FC^{DC}(a)da + \int_{a^*}^{\overline{a}} FC^{DB}(a)da$, where a^* denotes the optimal cutoff found in Section 4.2. If the budget constraint is binding, the optimal cutoff would be chosen from the feasible set of cutoff policies.

A second alternative is to consider a social planner's problem which maximizes total surplus. A social planner would choose to find the optimal cutoff policy that maximizes overall benefits less costs:

$$a^{***} = \arg\max_{\tilde{a}} \int_{\underline{a}}^{\tilde{a}} \left[\left(CE^{DC}(a) - FC^{DC}(a) \right) - \left(CE^{DB}(a) - FC^{DB}(a) \right) \right] da. \tag{B-3}$$

The first order condition of the social planner's problem equates marginal benefits to marginal costs:

$$CE^{DC}(a^{***}) - FC^{DC}(a^{***}) = CE^{DB}(a^{***}) - FC^{DB}(a^{***}). \tag{B-4}$$

In summary, an optimal age-based default rule may be obtained if the firm's objective differed from one which only considered employee welfare.

Appendix C: Comparitive Statics for the Optimal Age-Based Default Policy

As shown in Equation (13), the optimization problem and the implicit function theorem together provide a formula to sign the direction of changes in the optimal age cutoff for changes in known parameters. Below, we examine several special cases to illustrate the intuition in this result.

Case 1: Suppose $c_j = c$ for all j. The optimal cutoff age a^* is increasing in c:

$$\text{sign}\left(\frac{\partial a^*}{\partial c}\right) = \text{sign}\left(\underbrace{\frac{\partial CE^{DC}}{\partial c}}_{>0} - \underbrace{\frac{\partial CE^{DB}}{\partial c}}_{=0}\right) > 0. \tag{C-1}$$

Case 2: Similarly, suppose $b_j(w_j) = b$ for all j. The optimal cutoff age a^* is decreasing in b:

$$\text{sign}\left(\frac{\partial a^*}{\partial b}\right) = \text{sign}\left(\underbrace{\frac{\partial CE^{DC}}{\partial b}}_{=0} - \underbrace{\frac{\partial CE^{DB}}{\partial b}}_{>0}\right) < 0. \tag{C-2}$$

Case 3: Suppose the utility function U is such that $U^{-1}(\beta w) = h(\beta) \cdot U^{-1}(w)$ for some function h. Then the maximization problem in Equation (10) does not depend on the discount rate d. Therefore:

$$\frac{\partial a^*}{\partial d} = 0. \tag{C-3}$$

Appendix D: Monte Carlo Simulations

Monte Carlo simulations were performed to obtain 1,000 simulations of 45 years of asset returns. This methodology follows Shoven (1999). Suppose $r_i = ln(1 + R_i)$, where R_i denotes the simple real return of one of three types of assets (stocks ($i = s$), bonds ($i = b$), and money market accounts ($i = m$)). We assume r_i is distributed normally, i.e. $1 + R_i$ is distributed lognormally. The lognormal distribution is skewed to the right and ensures that simple returns cannot fall below -100%. Let m_i,

s_i, and s_{ij} denote the moments of R_i, and μ_i, σ_i, and σ_{ij} denote the moments of r_i.

The Monte Carlo simulation is done in three steps. First, μ_i, σ_i, and σ_{ij} are obtained for all classes of assets by the following set of equations:

$$\mu_i = log\left(\frac{1+m_i}{\sqrt{1+\left(\frac{s_i}{1+m_i}\right)^2}}\right) \tag{D-1}$$

$$\sigma_i^2 = log\left(1+\frac{s_i^2}{(1+m_i)^2}\right) \tag{D-2}$$

$$\sigma_{ij} = log\left(1+\frac{s_{ij}}{(1+m_i)(1+m_j)}\right) \tag{D-3}$$

Next, three independent standard-normal random variables z are generated for each simulation using Matlab's random number generator. These three random numbers are combined such that the returns have the desired variances and covariances.

$$r_s = \mu_s + \sigma_s z_1 \tag{D-4}$$

$$r_b = \mu_b + \frac{\sigma_{sb}}{\sigma_s} z_1 + z_2 \sqrt{\sigma_b^2 - \left(\frac{\sigma_{sb}}{\sigma_s}\right)^2} \tag{D-5}$$

$$r_m = \mu_m + a z_1 + b z_2 + c z_3 \tag{D-6}$$

The constants a, b, and c are given by:

$$a = \frac{\sigma_{sm}}{\sigma_s} \tag{D-7}$$

$$b = \frac{\sigma_{bm} - \frac{\sigma_{sm}\sigma_{sb}}{\sigma_s^2}}{\sqrt{\sigma_b^2 - \left(\frac{\sigma_{sb}}{\sigma_s}\right)^2}} \tag{D-8}$$

$$c = \sqrt{\sigma_m^2 - a^2 - b^2}. \tag{D-9}$$

Finally, the simple returns R are determined by using the transformation $R_i = \exp(r_i) - 1$.