Title
Essays on Retirement and Medicare Policy

Author
Leganza, Jonathan Michael

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on Retirement and Medicare Policy

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Jonathan M. Leganza

Committee in charge:

Professor Gordon Dahl, Chair
Professor Itzik Faldon, Co-Chair
Professor Julie Cullen
Professor Alexander Gelber
Professor Gaurav Khanna
Professor Krislert Samphantharak

2021

The dissertation of Jonathan M. Leganza is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To Alex.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I thank my advisors, Gordon Dahl and Itzik Fadlon, for their continuous guidance and support. Their mentorship has been instrumental in my growth as a researcher, as an academic, and as a person.

I am grateful to have been supported by the other members of my committee as well: Julie Cullen, Alex Gelber, Gaurav Khanna, and Krislert Samphantharak. I thank Julie especially for being generous with her time and for providing insightful feedback and comments on my work. I thank Alex for taking an interest in my research and for pushing me when I needed to be pushed. I also thank Jeff Clemens for additional advice and guidance, especially on the project that became Chapter 3.

I feel fortunate to have studied alongside UCSD classmates who quickly became friends, especially Bruno Lopez-Videla, Alex Masucci, and Sameem Siddiqui (I learned so much from our "applied coffees") as well as Nobuhiko Nakazawa and Stephanie Khoury. I thank Alex and Stephanie for being excellent co-authors. I thank Bruno for taking each step of the Ph.D. with me.

I have many to thank outside of UCSD. I am grateful that through my work set in Denmark, I gained an absolutely fantastic co-author and an even better friend in Esteban García-Miralles. I thank Claus Kreiner and Torben Heien Nielsen for their help with accessing the Danish data and for feedback on the projects that became Chapters 1 and 2. I thank the Center for Economic Behavior and Inequality (CEBI) at the University of Copenhagen for hosting me for a short visit as Esteban and I began working on our research together. I thank the National Bureau of Economic Research (NBER) for generous financial support through a Pre-Doctoral Fellowship in Retirement and Disability Policy Research while I worked on the project that became Chapter 1.

Finally, I thank my family for their unwavering support. I thank my parents for being a steady source of encouragement. I thank my wife, Alex, for always being right by side.

Chapter 1, in full, is currently being prepared for submission for publication of the material. García-Miralles, Esteban and Leganza, Jonathan M. "Public Pensions and Private Savings." The dissertation author was a primary investigator and an author of this material.

Chapter 1 was supported by an NBER Pre-Doctoral Fellowship in Retirement and Disability Policy Research. The research was performed pursuant to grant RDR18000003 from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States (U.S.) Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the U.S. Government or any agency thereof.

Chapter 2, in full, is currently being prepared for submission for publication of the material. García-Miralles and Leganza, Jonathan M. "Joint Retirement of Couples: Evidence from Discontinuities in Denmark." The dissertation author was a primary investigator and an author of this material.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Khoury, Stephanie, Leganza, Jonathan M., and Masucci, Alex. "Health Professional Shortage Areas and Physician Location Decisions." The dissertation author was a primary investigator and an author of this material.

VITA

2015	Bachelor of Arts, Economics, Indiana University
2015	Bachelor of Science, Mathematics, Indiana University
2021	Doctor of Philosophy, Economics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays on Retirement and Medicare Policy

by

Jonathan M. Leganza

Doctor of Philosophy in Economics

University of California San Diego, 2021

Professor Gordon Dahl, Chair
Professor Itzik Faldon, Co-Chair

This dissertation is comprised of three freestanding chapters, each of which studies a policy-motivated question related to the economics of aging. The first two chapters provide new evidence from Denmark on how social security policy impacts retirement, labor supply, and savings decisions of older workers, and the last chapter evaluates a U.S. Medicare policy that aims to improve access to healthcare for older Americans.

Chapter 1 studies how the provision of public pension benefits impacts private savings. We use administrative data and a regression discontinuity design to identify the causal effects of a Danish reform that increased social security eligibility ages. We find no evidence of anticipatory savings responses after the announcement of the reform, whereas we find large

increases in savings in retirement accounts when delayed benefit eligibility induces extended employment. The evidence suggests inertia is a leading mechanism: individuals continue to work and continue to save, and we show how employer default contribution rate policies mediate responses to the national reform.

Chapter 2 studies how social security influences joint retirement of couples. We first document joint retirement behavior around the early pension eligibility age in Denmark: we show that spouses are discontinuously more likely to retire when their partners first become eligible for social security benefits. We then explore underlying mechanisms and find age differences within couples to be crucial determinants of joint retirement behavior, which is primarily driven by older spouses continuing to work until their younger partners reach pension eligibility age.

Chapter 3 studies a U.S. Medicare policy that delivers bonus payments to physicians for practicing in areas with few doctors per capita. Using several sources of data from the Centers for Medicare and Medicaid Services and a matched difference-in-differences design, we find that counties designated as official shortage areas experience an increase in the number of early-career primary care physicians. However, we find no evidence that later-career physicians respond to shortage area designations, and we highlight how targeting bonus payments towards newer physicians may thus improve the effectiveness and cost-efficiency of the policy.

Chapter 1

Public Pensions and Private Savings

Abstract

How does the provision of public pension benefits impact private savings? We answer this question in the context of a Danish reform that increased social security eligibility ages contingent on birthdate. Using administrative data and a regression discontinuity design, we identify the causal effects of the policy. First, we show a lack of anticipatory savings responses after the reform was announced but before it was implemented. Second, we show large savings responses after implementation, when delayed benefit eligibility induces individuals to extend employment. Specifically, we find increased contributions to both employer-sponsored and personal retirement accounts, whereas we find no evidence of adjustments to other savings vehicles, such as bank or stock market accounts. Additional analyses point to inertia as a leading explanatory channel. The increased savings in personal retirement plans is entirely driven by those who made consistent contributions in the past. Moreover, the increased savings in employer-sponsored plans is largely explained by continuing to contribute at employer default rates, highlighting a role for firm policies in mediating responses to social security reform.

1.1 Introduction

A long-standing question in public finance asks how publicly-provided pension benefits impact private savings. Understanding the relationship between these two forms of retirement wealth is important for the optimal design of social security systems, which are some of the largest social insurance programs in the world. Classical work emphasizes that pension benefits should crowd out savings. Yet the effect of social security on savings is actually theoretically ambiguous after accounting for the effect of benefits on retirement decisions, since social security may induce earlier retirement and increase the time horizon over which assets are needed to finance consumption (Feldstein 1974). A principal task for empirical research is hence to investigate how public pension benefit schemes impact savings in practice.

Establishing convincing causal evidence on this question is difficult, due largely to two significant challenges. First, data availability is a major obstacle. A thorough analysis requires data that contain information on employment, earnings, and benefit receipt, as well as information on private savings, assets, and wealth. In most countries, these demands necessitate the use of survey data, which can suffer from small sample sizes and a lack of reliable and detailed information on assets. Second, identification requires a compelling source of exogenous variation in benefit payout structures.

In this paper, we overcome these challenges using administrative register data from Denmark and a regression discontinuity (RD) design. The context of our study is a major reform to the Danish retirement system announced in 2011 and implemented in 2014 that created a six-month discontinuous increase in pension eligibility ages for those born on or after January 1, 1954. Those born just after this cutoff date are similar in all aspects to those born just earlier, yet differ sharply in the ages at which they become eligible for pension benefits. We exploit the policy change to identify causal effects, estimating discontinuities in outcome variables by birthdate, and we exploit the breadth of our detailed data to study separately the effect of the reform on several types of savings vehicles.

Leveraging the timing of the policy, we distinguish between anticipatory responses (after the reform is announced but before it is implemented) and responses after implementation (when individuals navigate retirement years facing differences in benefit eligibility). In Denmark, there are three critical pension eligibility ages. The early retirement age (ERA) stipulates the age at which individuals first become eligible for early retirement benefits, two years later is an incentivized retirement age, and the Full Retirement Age (FRA) denotes the age at which individuals can transition to standard old-age benefits. These ages used to be 60, 62, and 65, respectively. The policy reform that we study initiated step-wise increases in each of these eligibility ages by birth cohort. We focus on the first phase of the reform, which creates the cleanest quasi-experiment. Those born on or just after January 1, 1954 learn in 2011, at age 57, that their critical pension eligibility ages are increasing to $60\frac{1}{2}$, $62\frac{1}{2}$, and $65\frac{1}{2}$ and constitute the treatment group. Those born just earlier experience no such change and constitute the control group. Our RD estimates over the years 2011 to 2013 capture the causal effects of *future* differences in pension eligibility. Our RD estimates over the years 2014 to 2018 capture the causal effects of *current* differences in pension eligibility, since it is during these years that our analysis sample navigates through the early retirement program. Note the data are not yet available to study behaviors around the FRA.[1]

We begin with an analysis of how retirement behavior changes in response to the reform. In the Danish setting, pension accrual incentives and high implicit taxes on work create strong incentives to retire either right at the ERA or right at the incentivized claiming age two years later. We show large corresponding spikes in retirement right at ages 60 and 62 for the control group. We then show how the reform causes the spikes in retirement to shift to the new eligibility ages in lockstep. The distribution of retirement ages for the treatment group contains large spikes in retirement right at $60\frac{1}{2}$ and $62\frac{1}{2}$, consistent with delayed retirement due to the reform-induced incentives.

[1]The birth cohorts we study are age 65 in 2019, and our data extend through 2018.

We then turn to our RD design to quantify the effects of the reform on savings. Our first set of RD results corresponds to the three-year anticipation period, as our analysis sample approaches age 60. We do not find any statistically significant or economically meaningful savings responses in anticipation of reaching pension eligibility ages. There is no evidence that individuals adjust savings through employer-sponsored retirement plans (analogous to 401(k)s), personal retirement plans (analogous to IRAs), bank accounts, stock market investments, or property wealth. These results are inconsistent with lifecycle models that call for forward-looking adjustments to savings after the announcement of the reform in response to future differences in pension benefit payouts.[2]

Our second set of RD results corresponds to the early retirement period, as our analysis sample ages from 60 to 64 and differences in benefit eligibility manifest themselves. During the first critical year of 2014, when the analysis sample is age 60 and the treatment group works longer in order to retire at the new ERA of $60\frac{1}{2}$, we document an increase in aggregate average earnings of \$6,117 (13%). We find concurrent and large increases in contributions to employer-sponsored retirement accounts, amounting to \$765 (15.5%) on average, that accompany this increase in earnings. We also find significant impacts on personal retirement accounts, as individuals are 3.9 percentage points (30%) more likely to contribute to these plans. During the second critical year of 2016, when treated individuals work longer to retire at the new incentivized age of $62\frac{1}{2}$, we find similar responses. In this year, earnings rise by 15%, contributions to employer-sponsored plans rise by 19%, and the likelihood of contributing to personal plans rises by 24%.

In contrast, during the non-critical years of 2015, 2017, and 2018, when the strong incentives for delayed retirement are not present, we find muted or null responses in earnings and savings in retirement accounts. Moreover, we consistently find no evidence of savings responses through any other financial vehicles, perhaps most notably bank accounts and

[2]This takeaway is broadly consistent with recent work that focuses on labor supply and earnings in the context of pension reform and finds a lack of forward-looking responses (Gelber et al. 2016 and Haller 2019).

stock market investments, in any year. That is, our results indicate savings respond *only* when the treatment group is induced to delay retirement to comply with the new pension eligibility ages and *only* in traditional retirement accounts, which are specifically earmarked for consumption in retirement.

What can explain our findings? To investigate mechanisms, we conduct a series of additional analyses, and the overall body of evidence points to inertial behavior. We first provide evidence against two alternative explanations for the lack of anticipatory responses. It is unlikely that a complete lack of awareness can explain the null responses after the reform is announced, as we show the policy was well-publicized and prompted large increases in relevant Google search activity. We also rule out an inability to respond as a leading explanatory channel; we find no evidence of anticipatory responses even for a subsample of individuals who have room to adjust contributions to voluntary retirement savings accounts and who may be more financially sophisticated.

Next, we unpack the positive savings responses in both personal and employer-sponsored retirement accounts during the critical years of extended employment, and we find evidence supporting inertia. Consistent with the reform leading to the continuation of previous savings behaviors, we show that the increases in contributions to personal retirement plans are entirely driven by those who had made frequent contributions to the accounts in the past. We then leverage our linked employee-employer data to show that the increases in employer-sponsored retirement plans are largely driven by continued contributions at employer default contribution rates during the policy-induced periods of extended employment. Employer contribution policies have been shown to be key drivers of savings in employer-sponsored retirement accounts (Madrian and Shea 2001, Choi et al. 2002, Beshears et al. 2009, Choi 2015), especially in Denmark (Chetty et al. 2014, Fadlon et al. 2016) where unions, employer associations, and firms have a major influence in setting contribution rates. We show how these types of policies can dictate responses to a national reform.

Taken together, our results show that in response to increases in pension eligibility ages, individuals extend employment and accumulate more savings. The lack of anticipatory responses, the lack of responses during non-critical years, the lack of adjustments to savings outside of retirement accounts, and the continuation of savings behaviors within retirement plans exhibited before the reform suggest inertia as the most likely mechanism.

Our paper relates most directly to the important literature that studies how private savings respond to the provision of public pension benefits.[3] Traditionally, papers aim to provide explicit estimates of the elasticity between public pension wealth and private savings. Earlier papers laid theoretical groundwork and provided empirical evidence mostly correlational in nature (e.g., Feldstein 1974, Feldstein and Pellechio 1979, Kotlikoff 1979, King and Dicks-Mireaux 1982, Diamond and Hausman 1984, Hubbard 1986, Pozo and Woodbury 1986, and Bernheim 1987). More recent papers have used difference-in-differences style estimators applied to survey datasets to study reforms and have produced a wide range of elasticity estimates from several different countries (e.g., Attanasio and Brugiavini 2003, Attanasio and Rohwedder 2003, Bottazzi et al. 2006, Aguila 2011, Feng et al. 2011, Lachowska and Myck 2018, and Slavov et al. 2019).[4] Finally, using an RD design, Lindeboom and Montizaan (2020) study how retirement expectations, retirement realizations, and savings decisions respond to a composite reform in the Netherlands which reduced pension wealth.

Our approach is to hone in on one prominent type of pension reform—namely changes in social security eligibility ages—and to unpack the causal effects of this policy on savings through the lens of a standard lifecycle framework.[5] In doing so, we make three main

[3]A second related literature studies pension eligibility ages and labor supply (e.g., Mastrobuoni 2009, Behaghel and Blau 2012, Staubli and Zweimüller 2013, Manoli and Weber 2016, Lalive et al. 2017, Geyer and Welteke 2019, Haller 2019, Nakazawa 2021, Deshpande et al. 2020, and Geyer et al. 2020). Our analysis also connects to the general literature on social security and retirement incentives, as reviewed by Krueger and Meyer (2002) and Blundell et al. (2016). For instance, Burtless and Moffitt (1985), Asch et al. (2005), Coile and Gruber (2007), Liebman et al. (2009), Brown (2013), and Manoli and Weber (2016) similarly analyze nonlinear budget constraints from pension systems.

[4]For cross-country empirical analyses on the topic, see Kapteyn and Panis (2005), Disney (2006), Hurd et al. (2012), and Alessie et al. (2013). In the context of Denmark, the most related findings come from Chetty et al. (2014), who show that a government mandatory savings program from 1998 to 2003 did not crowd out other savings among low-income individuals.

[5]Two working papers use approaches similar to ours. Etgeton et al. (2021) study anticipatory savings responses to a reform that increased the early retirement age of women using survey data from Germany,

contributions to the literature on social security and savings. First, we provide novel evidence on how savings respond to increases in social security eligibility ages using a compelling RD design and population-wide administrative data. Second, we leverage our data to analyze separate measures of third-party reported assets throughout essentially the entire financial portfolio, whereas the literature has been restricted to using survey measures of savings such as self-reported income minus self-reported consumption. We view this innovation as an important step forward, as different types of vehicles for savings are likely to differ in the extent to which they serve as natural substitutes for public pension wealth. We highlight in particular the distinction between retirement accounts and other savings, itself the subject of a related strand of literature.[6] Third, we exploit our setting to provide a more thorough exploration into mechanisms. We are able to uncover evidence suggesting inertia as an operative channel through our ability to study both anticipation and post-implementation time periods, through the panel structure of our data (which allows us to study how contributions to personal retirement plans differ by previous savings behaviors), and through the employer-employee linkages in our data (which allow us to incorporate firm default contribution rates into our analysis.)

Overall, our results have broad implications for social security policy and models of household behaviors. First, we find that the often-pulled policy lever of raising eligibility ages for public pensions leads to more savings set aside in retirement accounts for shorter retirement time horizons. Second, our results lend support to models that give rise to inertia in savings behaviors, such as those including fixed costs of adjustment, and they underscore a tight link between savings and employment. Third, our study emphasizes the importance of considering interactions with firm policies, such as employer retirement savings programs, when designing and predicting the effects of public policies.

---

and Nakazawa (2021) studies primarily how increasing pension eligibility ages impacts labor supply but also investigates physical and mental health, consumption, and savings using survey data from Japan.

[6]For earlier work on the relationship between tax-advantaged retirement accounts and total savings, see, e.g., Poterba et al. (1996), Engen et al. (1996), and Bernheim (2002). For more recent papers, see Gelber (2011), Chetty et al. (2014), and Andersen (2018).

The rest of this paper is organized as follows. Section 1.2 provides an overview of the institutional background. Section 1.3 grounds our empirical analysis with a conceptual framework and discusses the economic incentives. Section 1.4 describes the data. Section 1.5 lays out our identification strategy. Section 1.6 presents the main results, documenting the causal effects of the reform. Section 1.7 investigates underlying mechanisms and discusses potential explanations for our findings. We conclude in Section 1.8.

1.2 Institutional Background

The Danish retirement system is broadly typical of other OECD countries. Primary sources of retirement income include private retirement savings accounts and public pension benefits. In this section, we first discuss the central features of the retirement system before describing the policy reform. More background information can be found in Appendix 1.B.

1.2.1 Private Retirement Savings Accounts

As is typical of other modern economies, defined-contribution private retirement savings accounts dominate the retirement savings landscape in Denmark and constitute a key source of income in older age. Retirement savings plans can be either employer-sponsored accounts, analogous to 401(k)s in the U.S., or personal accounts, analogous to Individual Retirement Accounts (IRAs). The treatment of these savings accounts in the tax code is similar to the U.S setting: contributions are tax-deductible, returns are tax-advantaged, distributions from the accounts are taxed upon withdrawal, and penalties exist on early withdrawals.[7]

Broadly speaking, in Denmark participation in employer-sponsored retirement savings

[7]Our analysis focuses on these traditional retirement accounts. In 2013, Denmark introduced "Roth-style" retirement accounts to the economy. Contributions to these plans are not tax deductible, but distributions are not taxed. For completeness, we study these types of accounts in the appendix, though overall they are likely to make up a much smaller fraction of the financial portfolio for the birth cohorts we study, who were 59 years-old when the accounts were first introduced.

plans is often quasi-mandatory. Collective bargaining agreements between labor market unions and employer associations cover the majority of workers. These agreements frequently stipulate a minimum percentage of wages that are to be contributed to retirement savings accounts, and so contribution rates to employer-sponsored accounts tend to be similar for workers under the same agreement. For workers not covered by these agreements, firms often set their own default contribution rates. In contrast, contributing to personal retirement savings plans is completely voluntary.

1.2.2 Public Pension Benefits

Public old-age retirement benefits come from two main sources. The Old Age Pension (OAP) provides basic retirement income security, and the Voluntary Early Retirement Pension (VERP) provides early retirement benefits for those who choose to participate in the program. Participation in VERP requires making modest contributions to qualified Unemployment Insurance (UI) funds during working life, and the majority of workers—about 70% of the individuals in the birth cohorts we study—choose to participate. We focus our study on those participating in the VERP program, as it has historically played a major role in determining labor supply and retirement patterns of the Danish population, and as those not participating in VERP only just became eligible for the OAP in 2019 (for which the data does not yet exist). The two programs are closely connected; however, the provision of benefits from each program are governed by different rules and regulations.

Voluntary Early Retirement Pension

The VERP program grants participants access to up to five years of early retirement benefits, starting at the Early Retirement Age (ERA) of 60 and ending at the Full Retirement Age (FRA) of 65. The most important idea for our study is that the features of the VERP program produce very strong incentives to concurrently claim benefits and retire either right at the ERA or right at the incentivized age two years later. The following details explain why this is the case.

Workers claim into VERP, at which point they lock in their annual base benefits for the duration of the program. Benefits amount to roughly $27,000 (in 2010 U.S. dollars), which are then subject to strict means testing.[8] First, base benefits for the duration of the program are reduced against wealth held in private retirement accounts right before reaching age 60.[9] Second, benefit payouts are reduced against drawdown income from retirement accounts. Third, benefit payouts are additionally reduced against hours worked at a rate of 100%, which creates high implicit taxes on continued work after claiming. Even more, there are no actuarial adjustments for delaying claiming; deferring claiming simply forfeits benefits. For example, claiming at 61 results in only four years of benefits instead of five.

Two key rules drive the incentives to claim and retire either right at the ERA of 60, or the incentivized age of 62. First, the "transition rule" requires workers to be available to the labor force in order to be eligible to claim. An important implication of this rule is that retiring and dropping out of the workforce before reaching the ERA results in forgoing the entire five years of VERP eligibility. This rule creates strong incentives for workers to wait to retire until at least reaching the ERA (whereas the high implicit taxes and lack of adjustments for deferring claiming discourage working after the ERA). Second, the "two-year rule" creates financial incentives for some to claim VERP and retire at age 62. Most importantly, working and deferring claiming until age 62 results in the elimination of the means testing of VERP base benefits against private retirement account balances. Some additional but smaller financial incentives exist as well, though the means testing of benefit payouts against drawdown income and hours worked remain.[10] This relaxation of means

[8]Benefit amounts are determined through a formula linked to the UI system, but are capped at 91% of the maximum amount of UI benefits, which leads to base benefits that are in practice largely flat-rate.

[9]The government collects information on retirement account balances for VERP-eligible individuals around age $59\frac{1}{2}$, and the $27,000 base benefits are reduced using this information. The means testing rules depend on many factors, but roughly call for base benefits to be reduced by 60% of could-be annuitized income from retirement accounts.

[10]Satisfying the two-year rule results in a modest increase in base benefit amounts as well, to approximately

testing after age 62 can create strong financial incentives to wait to retire until age 62, especially for those with significant assets in private retirement accounts.

Old Age Pension

Upon reaching the FRA of 65, retirees transition from VERP to the OAP, which provides annual, flat-rate, old-age benefits until death. The key idea for our study is that OAP wealth largely does not depend on retirement age. Benefits are roughly $15,000 for married individuals and $20,000 for single individuals, but are reduced proportionally for those who have not lived in Denmark for at least 40 years. OAP benefits are means-tested against income, subject to an income test, though those wishing to continue to work can take advantage of approximately actuarially-fair adjustments for deferring claiming.

1.2.3 The 2011 Reform on Later Retirement

In response to population aging and budgetary concerns, the Danish government announced in May of 2011 a major reform to the retirement system. A key component of the reform stipulated the phasing in of stepwise 6-month increases in pension eligibility ages, contingent on birthdate. Figure 1.1 graphically illustrates how the reform indexed each of the three key eligibility ages to birthdate in a discontinuous fashion. We focus our entire analysis on the first birthdate discontinuity generated by the reform, which forms the cleanest quasi-experiment by creating a treatment and control group who differ only in their pension eligibility ages. The rules and regulations governing benefit amounts and means testing did not change for the sample we study.[11]

$29,600, as benefits become linked to 100% (rather than 91%) of maximum UI benefits. See Appendix 1.B for more details.

[11] The later phases of the reform continued to increase eligibility ages as illustrated in the figure, but also made more changes to the VERP program. The reform created more stringent VERP participation rules, slightly increased the standard base benefit amounts, and implemented even stricter means testing policies against assets held in private retirement accounts. Importantly, all of these changes were phased in to impact later birth cohorts, and none of them affect the individuals at the birthdate discontinuity that we study.

Specifically, in our analysis we exploit the fact that those born on January 1, 1954 learn in 2011 that their ERA has increased to age $60\frac{1}{2}$, that their incentivized retirement age has increased to age $62\frac{1}{2}$, and that their FRA has increased to age $65\frac{1}{2}$. In contrast, those born one day earlier, on December 31, 1953, experience no change in their pension eligibility ages, which remain constant at 60, 62, and 65. Our identification strategy exploits the discontinuous nature of the policy change; individuals born right around the birthdate cutoff should be similar in all aspects, yet face different retirement and savings incentives due to the reform.

1.3 Economic Framework

We use a simple lifecycle framework to model key features of the pension system as well as the changes in incentives brought on by the 2011 reform. Building directly on Laitner and Silverman (2007) and Hurd et al. (2012), we write down a standard dynamic model of consumption with an endogenous retirement decision and no uncertainty. We have two goals. First, we aim to ground our study in baseline theory to aid in the interpretation of our results. Second, we aim to provide benchmark predictions that can be mapped to our empirical analysis.

1.3.1 Model Setup and Solution

We borrow the initial setup from Hurd et al. (2012). Consider economic agents making decisions throughout continuous time $t \in [0, T]$. Agents choose consumption, c_t, and when to retire, $t = R$. Wages are constant while working so that $y_t = y$. Pension benefits received after retirement, $b_t(R)$, depend on the retirement age, and the present value of pension wealth is given by $B(R) = \int_R^T e^{-rt} b_t(R) dt$, where r is the interest rate. Utility during working life is given by $u(c_t)$, and utility in retirement is given by $u(c_t) + \Gamma$, where Γ is the utility gain from leisure. For simplicity, we assume the rate of time preference, ρ, equals the interest rate r.

Formally, agents solve the following optimization problem:

$$\begin{aligned}
\max_{R,\{c_s\}_{s=0}^{R}} \quad & \int_0^R e^{-\rho t} u(c_t)dt + \Psi(a_R + B(R), R) \\
\text{s.t.} \quad & \dot{a}_t = ra_t + y_t - c_t \\
& a_0 = 0,
\end{aligned} \tag{1.1}$$

where $\Psi(a_R + B(R), R)$ is the post-retirement indirect utility given by

$$\begin{aligned}
\Psi(a_R + B(R), R) = \max_{\{c_s\}_{s=R}^{T}} \quad & \int_R^T e^{-\rho t} \left(u(c_t) + \Gamma\right) dt \\
\text{s.t.} \quad & \dot{a}_t = ra_t - c_t \\
& a_T = 0.
\end{aligned} \tag{1.2}$$

For any given retirement age R, this formal problem has a familiar solution for consumption. After deriving first-order conditions, one can write:

$$\frac{u''(c_t)}{u'(c_t)} \dot{c}_t = \rho - r. \tag{1.3}$$

Since we assume the utility discount rate equals the interest rate, individuals should perfectly smooth consumption. Consumption in each period thus depends on lifetime resources, which depend on the timing of retirement:

$$c_t = c(Y(R), B(R)) = \frac{C^L}{T}, \tag{1.4}$$

where C^L is lifetime consumption and $Y(R) = y \int_0^R e^{-rs} ds$ is the present discounted value of lifetime earnings. The following first-order condition describes the optimal time of retirement:

$$(y + B'(R)) \cdot u'(c_R) = \Gamma. \tag{1.5}$$

The left-hand side is the marginal benefit of retiring later—the financial return to working longer converted to utility units using the marginal utility of consumption—and the right-hand side is the marginal cost of retiring later—foregone utility of leisure.

1.3.2 Retirement Incentives Before the Reform

This simple setup offers insight into retirement decisions in our setting. Assume that heterogeneous preferences for leisure are smoothly distributed. If individuals face a linear budget constraint, that is, if the financial return to work, $y + B'(R)$, is constant, then the distribution of optimal retirement ages would be governed by some smooth density function.

However, in our setting, pension wealth $B(R)$ is highly non-linear in retirement age R. Figure 1.2 illustrates this notion graphically by plotting public pension wealth against retirement age for a representative worker from the pre-reform birth cohort.[12] We can see that the key features of the system create two large spikes in pension wealth. The first spike occurs right at the ERA of 60. Retiring before this age results in a failure to satisfy the transition rule, and thus the inability to claim VERP benefits, which means public pension wealth is given by only the OAP.[13] Retiring right at 60 discontinuously increases pension wealth by the entire 5 years of VERP benefits. The second spike occurs right at age 62, the age at which means testing of VERP benefits against private retirement account balances is eliminated. Retiring one day before age 62 locks in three years of standard VERP benefits, whereas retiring one day later increases benefit payouts in each year due to reduced means testing.[14]

[12]For illustrative purposes, we abstract from discounting, and the benefit amounts depicted in the figure are for a worker who is married, who lives until age 85, and who has \$250,000 in private retirement savings accounts at age 60.

[13]The y-intercept in the stylized graph is \$300,000, which corresponds to 20 years (from age 65 to 85) of standard OAP benefits (\$15,000 per year).

[14]The negative slopes between 60 and 62 and between 62 and 65 result from the lack of actuarial adjustments when deferring claiming. Pension wealth for those who retire after age 65 is greater than just the OAP wealth due to quarterly bonus payments for working past age 62 (see Appendix 1.B). Note the size of each spike depends on assets held in retirement accounts; the greater the balances in retirement accounts, the smaller the first spike (due to more reductions in base VERP benefits) and the larger the second spike

The spikes in pension wealth at the critical ages translate to large discontinuities in lifetime consumption, C^L. Graph (a) of Figure 1.3 plots lifetime consumption against retirement age, for the same representative worker from the pre-reform cohort.[15] The discontinuities at 60 and 62 should induce bunching in the retirement distribution, as those who would have otherwise located either just to the left or just to the right of these ages find it optimal to retire right at the critical ages.[16]

We let the data speak to the strength of these bunching incentives in our setting. Graph (a) of Figure 1.4 plots the empirical distribution of retirement ages for those born before the January 1, 1954 birthdate cutoff.[17] There are few retirements before the ERA, and the spikes in retirement at the critical ages are large, indicating that the strong financial incentives to retire at either exactly the ERA or exactly two years after the ERA shape labor supply decisions of older workers.

1.3.3 Modeling the Reform: Benchmark Predictions

The 2011 reform increased pension eligibility ages. In the context of our framework, the major change is a shift in the location of the spikes in public pension wealth, $B(R)$, to $60\frac{1}{2}$ and $62\frac{1}{2}$, which changes the budget constraint as depicted by the maroon line in graph (b) of Figure 1.3. How should we expect individuals to respond to the reform? To ultimately provide benchmark predictions for savings, we first discuss changes in retirement incentives due to the reform. We then turn to the data to observe how the reform actually changed the retirement distribution. Finally, guided by these responses borne out in the data, we use

(due to greater gains from avoiding the means testing).

[15]For illustrative purposes, annual earnings are assumed to be \$55,000 and lifetime earnings are earnings after age 57, the age of our sample when the reform is announced.

[16]Note that incentive-induced bunching in retirement is not unique to the Danish system. Brown (2013) analyzes bunching in retirement at both kink and notch points created by incentives in the pension system for California teachers in the United States; similarly, Manoli and Weber (2016) study bunching at the early retirement age in Austria. For a general review of the bunching literature, see Kleven (2016).

[17]Details on the monthly data used to produce this graph can be found in Section 1.4; the underlying sample consists of workers born within six months of January 1, 1954. Retirement ages are defined using an absorbing state measure. We define monthly retirement age as the age of the individual in the last month during which earnings are positive, before permanently falling to zero.

our framework to assess how savings should respond.

Given the strong retirement incentives attached to VERP pension eligibility ages, we expect the dominant forces at play to essentially shift bunching masses at 60 and 62 to $60\frac{1}{2}$ and $62\frac{1}{2}$, respectively. We expect the influence of any other incentives to be minor. To examine whether this is the case, and to make headway on our predictions for savings, we directly evaluate the impact of the reform on retirement ages in the data.

Graph (b) of Figure 1.4 shows how the empirical distribution of retirement ages shifts after the reform. The maroon line depicts the behavior of those born after the January 1, 1954 birthdate cutoff, who are affected by the reform and face budget constraints corresponding to the maroon lines in graph (b) of Figure 1.3. The graph shows how the reform clearly induces a shift in bunching to the new pension eligibility ages and thus induced later retirement for many individuals.

Given these reform-induced labor supply responses, we can provide benchmark predictions for savings that are consistent with the lifecycle model. A key feature of the lifecycle framework is that future pension benefits and wages impact current consumption and savings, since individuals consider lifetime resources when determining optimal consumption paths. The reform induces later retirement, which represents an increase in lifetime income. The model calls for this extra income to be spread over the lifecycle in the form of increased consumption in every period. This change in the consumption profile yields two implications for savings (income less consumption), that can be directly mapped to our empirical analysis. First, during the anticipation period, after the announcement of the reform but before it is implemented, savings should *decrease* on average, as earnings during this period are unchanged but consumption has increased. Second, during the reform-induced periods of extended employment (e.g., between ages 60 and $60\frac{1}{2}$), savings should *increase* on average. Consumption is still elevated, but income is higher from continued employment, and the increase in consumption cannot be greater than the increase in income; some of the extra income should be saved to finance increased consumption throughout later stages of the lifecycle.

1.4 Data

To study empirically how raising pension eligibility ages impacts private savings, we use primarily annual administrative register data that cover the entire population of Denmark from 1985 to 2018. Attrition from the data is only due to migration out of Denmark or death. We use unique personal identifiers for individuals to link together population registers, which contain information on demographics (importantly including the exact date of birth), with labor-market registers, which contain detailed information on income and assets, in order to create a rich annual panel dataset. We use these data to conduct the bulk of our analyses.

We have also gained access to a complementary, monthly-level administrative dataset that contains information on all employees in Denmark from 2008 to 2017.[18] We use these data to more finely track exits from the labor force and to conduct the bunching analysis of retirement ages discussed above.

1.4.1 Key Variables

Our data constitute some of the highest quality data available on savings; they contain third-party reported variables on assets that essentially capture the entire financial portfolio, and thus form the ideal dataset for studying our research question. We avoid potential problems associated with using self-reported savings or imputed savings from self-reported income and consumption as outcome variables, and we exploit our data to study separately retirement savings accounts, bank accounts, stock market investments, and property values.

We observe flow variables that capture savings in traditional defined-contribution retirement accounts, which make up a dominant form of private saving in the economy and

[18]This dataset, known in Denmark as the *eIncome* register, contains information on wages and salaries that firms report to tax authorities at a monthly frequency. See Kreiner et al. (2016) and Kreiner et al. (2017) for more discussion on this relatively new dataset.

which might naturally be considered the closest substitutes to public pension wealth. We study as our main outcomes contributions to employer-sponsored accounts in levels and indicator variables for making positive contributions to personal accounts.[19] We also study annuitized distributions from these retirement accounts, but we are unable to distinguish between payments from employer-sponsored plans and personal plans. We winsorize contribution amounts at the 95th percentile, by year, in order to reduce the influence of outliers in our regressions, improve precision, and account for occasional observations of recorded contributions well-above annual contribution limits.[20]

For savings in bank accounts, stock market accounts, and property, we do not observe flow variables, but rather stock variables. Specifically, our measures of bank account balances and stock market account balances correspond to the value of assets held at the end of the calendar year, reported to tax authorities by financial institutions. Our measure of property corresponds to the year-end cash value of properties as assessed by the tax authorities directly. We use these measures to compute more noisy flow variables of savings in year t by subtracting year-end balances in year t with those from year $t-1$. We thus study changes in bank account balances, changes in stock market accounts, and changes in property values as our main outcomes. We winsorize these outcome variables (which unlike contributions to retirement accounts are not naturally bounded below by zero) at the 5th and 95th percentile in each year.[21]

Finally, we study as our main measure of labor supply pre-tax earnings, as defined by

[19] Our focus on extensive-margin responses to personal accounts is particularly informative in its own right, because contributions to personal plans are completely voluntary and thus less common than contributions to employer-sponsored plans. Mean contribution amounts in levels are often dominated by the large number of zeros. In Section 1.6, we discuss our approach to investigating contribution amounts to personal plans by using as outcomes indicators for making contributions of various sizes.

[20] Our analysis focuses on traditional retirement plans, though for completeness we analyze indicators for contributing to "Roth-style" retirement plans as well, in the appendix. As discussed in Section 1.2, Roth-style accounts were introduced to Denmark in 2013, when our analysis sample is 59 years-old, and thus likely form a substantially smaller part of the asset portfolio for the individuals we study.

[21] Still imprecision can present a challenge when studying these variables that capture changes in year-end assets within individuals, especially in relatively smaller samples. This general problem is discussed in more detail in Chetty et al. (2014); we follow their approach by additionally studying even more strictly winsorized versions of these outcome variables, at the 10th and 90th percentiles.

the amount of income on which individuals pay an 8% labor market tax. We also winsorize this variable by year at the 95th percentile for consistency. To define retirement ages, we use our monthly-level data. We use an absorbing state measure for retirement. We define monthly retirement age as the age of the individual in the last month during which earnings are positive, before permanently falling to zero. We study as our measure of benefit claiming annual VERP benefit amounts. We deflate all monetary values to 2010 levels and convert Danish kroner (DKK) to U.S. dollars. The exchange rate in 2010 was approximately 5.56 DKK to 1 USD.

1.4.2 Analysis Sample

Our analysis sample focuses on individuals participating in VERP who are born right around the first birthdate discontinuity generated by the 2011 reform. Specifically, starting with our data on the entire Danish population from 1985 to 2018, we carry out four main sample restrictions. First, we include only Danes born within six months of the cutoff date, January 1, 1954. Second, we keep only individuals who made regular participatory contributions to the VERP scheme before the reform was announced. Specifically, we keep those who made contributions in at least 70% of the pre-announcement years between 2001 and 2010.[22] Third, we balance the sample between the years 2006 and 2018. Fourth, we exclude the self-employed (defined during the pre-announcement period), who are subject to different rules and regulations concerning their early retirement options through the VERP scheme.

We are left with a sample of 40,042 individuals.[23] Table 2.1 presents summary statistics for calendar year 2010, the year before the reform is announced. Columns (1) and (2)

[22]We do not require contributions in 100% of the pre-announcement years in order to allow for short lapses in contributions, for which the program allows, as individuals in our analysis sample are required to contribute in 25 out of the last 30 years to be eligible for VERP.

[23]We conduct our analysis at the individual level because Denmark maintains individual-level tax and pension systems. See García-Miralles and Leganza (2021) for a study on joint retirement of spouses in Denmark.

display the mean and standard deviation of key variables for the entire analysis sample. Columns (3) and (4) provide the same information for the 12,020 individuals who will ultimately make up the main estimation sample in our RD design, namely those born within 56 days (8 weeks) of the January 1, 1954 birthdate cutoff. Our sample contains active older workers, most of whom are married. Average earnings in 2010 amount to approximately \$61,000. Most individuals (89%) make contributions to employer-sponsored retirement accounts, likely due to quasi-mandatory participation for many, and 41% of individuals contribute to personal retirement accounts. Average bank account balances amount to roughly \$26,000, whereas stock market account balances are smaller on average at just over \$7,000.

1.5 Identification Strategy

1.5.1 Regression Discontinuity Design

To identify the causal effects of increasing pension eligibility ages on savings and labor market outcomes, we employ a regression discontinuity (RD) design.[24] We derive identification from the policy-induced discontinuous change in eligibility ages contingent on birthdate. Due to the 2011 reform, individuals born on or after January 1, 1954 face pension eligibility ages of $60\frac{1}{2}$, $62\frac{1}{2}$, and $65\frac{1}{2}$, whereas those born just before face the previous eligibility ages of 60, 62, and 65. We use our RD design to estimate discontinuous changes in outcome variables at the birthdate cutoff.

Specifically, to implement our RD design, we estimate equations of the following form:

$$y_i = \alpha + \beta \cdot 1[x_i \geq c] + f(x_i - c) + 1[x_i \geq c] \cdot g(x_i - c) + Z_i\theta + \varepsilon_i, \tag{1.6}$$

where y_i is an outcome variable for individual i (such as contributions to retirement savings

[24] Imbens and Lemieux (2008), Lee and Lemieux (2010) and Cattaneo and Escanciano (2017) provide reviews of RD designs in economics.

accounts over some specified time period), x_i is birthdate, the running variable, c is the birthdate cutoff of January 1, 1954, Z_i is a vector of pre-determined control variables, f and g are functions, and ε_i is an error term. The coefficient of interest is β, which captures the average impact on the outcome of the six-month increase in pension eligibility ages for those born right around the birthdate cutoff.

In our baseline regression specification, we estimate separate linear polynomials in the running variable on either side of the cutoff, we use triangular weights, and we include as controls gender, pre-announcement marital status, and pre-announcement region of residence.[25] We choose our bandwidth to be eight weeks, or 56 days, on either side of the cutoff.

We probe the robustness of our results to these specification choices and discuss corresponding results in Section 1.6.3. In particular, we vary the bandwidth, drop the triangular weights, exclude controls, and estimate global linear polynomials in the running variable.

1.5.2 Threats to Identification and Assessment of Validity

The identifying assumption in our RD design is that other factors that could influence outcome variables do so smoothly in birthdate through the cutoff. In implementing our design, we estimate sharp jumps in outcomes right at the cutoff; causal interpretation of our results relies on the assumption that, in the absence of the policy-induced discontinuity in pension eligibility ages, outcome variables would have evolved smoothly through the cutoff.

The classical threat to identification in RD designs is manipulation of the running variable, which would typically generate a non-smooth density of the running variable. Manipulation in the usual sense is unlikely to be a potential problem in our setting, because

[25] We control for pre-announcement marital status using a dummy variable for being married or cohabiting in 2010. We control for pre-announcement region of residence using dummy variables for residing in 2010 in each of the five administrative regions of Denmark: Hovedstaden (the capital region containing Copenhagen), Sjælland, Syddanmark, Midtjylland (containing Aarhus), and Nordjylland.

our running variable is birthdate, which for our analysis group is determined long before the policy is announced. A separate threat to our design is the possibility of differential attrition by birthdate, as we ultimately balance our sample, selecting on being alive and in Denmark. If the reform impacts the propensity to drop out of the data (either due to death or leaving the country) in a way that is not as good as random as it relates to the outcome variables that we study, then balancing the sample as we do could bias our estimates.

We first note that while the literature on the mortality effects of social security income and pension eligibility ages across contexts is generally mixed (e.g., Snyder and Evans 2006, Kuhn et al. 2010, Hernaes et al. 2013, Fitzpatrick and Moore 2018), a recent paper finds no evidence that early retirement in Denmark impacts mortality (Nielsen 2019). Nonetheless, to more directly investigate the possibility of differential attrition in our study, we examine the density of our running variable in the spirit of McCrary (2008). Appendix Figure 1.A.1 plots a simple histogram of the running variable, birthdate, for the entire analysis sample. We also superimposed on top of the histogram smoothed values and confidence intervals from local polynomial regressions of the number of individuals on birthdate. A formal density test as proposed by Cattaneo et al. (2019) using our baseline choice of bandwidth results in a p-value of 0.97. Overall, we fail to find evidence indicating the presence of any problematic discontinuity in the density of the running variable at the birthdate cutoff.

As an additional check on the validity of our RD design, we investigate the smoothness of the (pre-determined) control variables through the birthdate cutoff. We estimate equation (1.6) without any covariates on the right-hand side, instead using each control variable as a left-hand side outcome variable. Appendix Table 1.A.1 presents these results. There are no statistically significant discontinuities in any of the control variables at the cutoff.

1.6 Main Results: Impact of Increasing Pension Eligibility Ages

In this section, we present our main results, which document the aggregate causal effects of increasing pension eligibility ages. We often lead with standard RD graphical analyses, which offer nonparametric representations of the causal effects of the reform. Specifically, we plot means of key outcome variables in one-week date-of-birth bins for individuals born around the birthdate cutoff, and we superimpose on these plots regression lines from estimating separate linear trends in the running variable for observations on either side of the cutoff. We then use regression-based estimates to quantify magnitudes and assess the statistical significance of our findings.

1.6.1 Anticipation Period

We begin our analysis by documenting impacts during the anticipation period. Recall that this period captures responses after the announcement, but before the implementation, of the reform. The individuals we study are 57 years old when the reform is announced, giving them time to make consumption and savings adjustments before they reach age 60, at which point differences in pension eligibility from the reform manifest themselves. The benchmark prediction laid out in Section 1.3 suggests a negative impact on savings over the anticipation period, as treated individuals should increase current consumption due to the net increase in lifetime income that will come from delayed retirement.

We find no evidence of any anticipatory savings responses though. Figure 1.5 illustrates this result graphically. Each graph corresponds to a different key outcome variable, where the variables of interest are averaged over the anticipation time period. For instance, graph (a) illustrates the RD estimate of the policy reform on average annual contributions to employer-sponsored retirement accounts between 2011 and 2013. Over this time period, average annual contributions to these types of accounts were around \$6,000 for the control group, and the graph shows no evidence of any discontinuous change in this outcome variable at the birthdate cutoff. Graph (b) shows no impact on contributions to personal plans, where here the extensive-margin outcome variable is the fraction of years contributing to personal plans. Likewise, graphs (c) through (e) show a lack of savings responses through changes in bank account balances, stocks market investments, and property wealth, respectively. Graph (f) shows that there are also no discontinuities in earnings over this time horizon. Overall, the graphs make a strong visual case for a lack of savings responses. The pattern of the binned means indicate that the savings of those born just to the left of the cutoff look no different than the savings of those born just to the right.

Table 1.2 presents results from corresponding regression analyses. We report in the table RD estimates of β from estimating equation (1.6) using our baseline specification. Not only are the point estimates statistically indistinguishable from zero, they are also economically insignificant. The point estimate on employer-sponsored retirement accounts, for example, is a positive \$20.32, which at face value represents a 0.33% increase off of the control group mean. The point estimate for contributions to personal retirement plans is small and positive, whereas the estimates for other savings vehicles are negative in sign, but small. To attempt to gain more precision, we follow Chetty et al. (2014) and further winsorize our non-retirement account savings outcomes at the 10th and 90th percentiles, and we report the results in Appendix Table 1.A.6. The first row presents the RD estimates for the anticipatory responses, which are very similar to our baseline results and more precise.

In general, a lack of anticipatory responses is not consistent with the notion that current savings respond to changes in future pension eligibility. We discuss potential explanations and underlying mechanisms for these results in Section 1.7, after first establishing the causal effects of the reform over the early retirement period, which then allows us to assess and discuss the overall body of evidence as a whole.

1.6.2 Early Retirement Period

Here we estimate the impact of the reform over the years 2014 to 2018. Discontinuities in these years reflect responses due to the implementation of the reform. Recall from Figure 1.4 that the reform induces extended employment to comply with the strong incentives now attached to the new pension eligibility ages. In our RD framework, we expect the shift in the spike in retirement at age 60 to age $60\frac{1}{2}$ to manifest itself as increases in earnings during 2014, the year during which our treatment and control group are both age 60, but when those in the treatment group retiring right at the ERA work six more months than their control group counterparts. Likewise, we expect the shift in the spike in retirement at age 62 to age $62\frac{1}{2}$ to be captured by the RD estimates in 2016. We call these two years "critical years," as they are the years during which individuals reach the two eligibility ages in the VERP scheme. Recall also that the benchmark lifecycle framework predicts increases in savings during these critical years, as individuals consume some of the extra income from continued work, but save some for future consumption.

Calendar year 2014 corresponds to the first critical year of the early retirement period, the first year during which differences in public pension eligibility present themselves. Figure 1.6 graphically depicts responses to the reform during this year. Graph (a) shows that the treatment group receives less VERP benefits during the year, almost exactly half of the average amount received by the control group, consistent with early retirees claiming right at $60\frac{1}{2}$, now that they are no longer eligible to claim at 60. Graph (b) shows a visually clear and large discontinuous increase in earnings amounting to just over \$6,000, which is a 13.7% increase off of a baseline mean of \$44,449. These results are entirely consistent with the delayed retirement documented in Figure 1.4.

Graph (c) of Figure 1.6 illustrates the effect of the reform on contributions to employer-sponsored retirement savings accounts. The RD estimate indicates an increase of \$765 to these retirement plans, which represents a meaningful 15.5% increase off of a mean of \$4,928.

Graph (d) illustrates how the treatment group is also 3.9 percentage points, or 27.9%, more likely to contribute to personal retirement accounts. Both of these point estimates are highly statistically significant, and the RD graphs provide visually compelling evidence that the reform causes individuals to save more in retirement accounts during the first critical year of policy-induced extended employment.

As mentioned in Section 1.4, we lead our analysis of contributions to personal plans with a binary indicator for contributing any positive amount. The large number of individuals contributing zero dollars makes it difficult to study contribution amounts in levels (see graph (a) of Appendix Figure 1.A.2). To overcome this challenge, we use as outcomes indicators for making contributions of various sizes to personal plans. Specifically, we use as outcome variables indicators for contributing between \$1 and \$X, where X starts at \$1,000 and increases until it captures contributions of all sizes. Graph (c) of Appendix Figure 1.A.2 plots the RD estimates and confidence intervals from estimating equation (1.6) on indicators for the various contribution amount bins. The point estimate furthest to the left mirrors the result in graph (d) of Figure 1.6: the policy causes a 3.9 percentage point decline in the likelihood of contributing \$0 to personal retirement plans. The subsequent point estimates show how in 2014 the reform caused increased contributions of meaningful amounts. The pattern of the point estimates, which are increasing as the contribution amount bins increase, suggests that the treatment group is more likely to make contributions of all sizes (except perhaps those over \$4,000).

We present regression-based results for all main outcomes in column (1) of Table 1.3. The reform not only results in greater contributions to both employer-sponsored and personal retirement accounts, it also leads to a decrease in annuitized distributions received from retirement accounts. Treatment individuals receive payments from retirement accounts that are about \$263 (16.6%) less on average.[26] Panel (c) of Table 1.3 reports RD estimates for

[26]Recall from Section 1.4 that we unfortunately cannot distinguish between distributions from employer-sponsored and personal accounts.

the other savings outcomes we study.[27] None of the estimates are statistically distinguishable from zero. The second row of Appendix Table 1.A.6 shows how additional winsorizing of these outcome variables produces small point estimates that are closer to zero and more precisely estimated. Overall, results from the first critical year show that in response to the increases in pension eligibility ages, individuals earn more from continuing to work, and this extended employment results in the accumulation of more savings in retirement accounts, whereas there is no evidence of adjustments to other types of savings.

Calendar year 2015 is not a critical year; in this year our analysis sample individuals are 61 years old. Those retiring right at the ERA have already done so, and those waiting to retire until the incentivized age must continue working until either age 62 or $62\frac{1}{2}$. The first column of Table 1.4 reports muted labor supply and savings responses during 2015; only one point estimate appears statistically distinguishable from zero.

In 2016, the second VERP critical year, our analysis sample individuals are 62 years old. Those who have continued to work in order to claim into VERP right when the means testing is relaxed retire during this year, either at age 62 for the control group or age $62\frac{1}{2}$ for the treatment group. Key results are graphically illustrated in Figure 1.7, and regression estimates for this year are reported in column (3) of Table 1.3. Similar to the first critical year, during 2016, treated individuals receive less VERP benefits and have 15.4% higher earnings. The extended employment again leads to more savings in retirement accounts: contributions to employer-sponsored plans increase by 18.8% and the likelihood of contributing to personal plans rises by 24.5%. Graph (c) of Appendix Figure 1.A.2 suggests that the increased contributions to personal plans are primarily contributions under $2,000. The point estimate on distributions from retirement accounts is negative and similar to the one in 2014, though more imprecisely estimated in this year. We again find no evidence of

[27] Results from analyzing indicators for contributing to Roth-style accounts, which were first introduced to the economy in 2013, are reported in Appendix Table 1.A.2; we find no evidence that the reform impacts contributing to these types of accounts (which likely make up a much smaller fraction of the retirement portfolio) in any year.

savings responses through bank accounts, stock market accounts, or property, as the main RD estimates (as well as those subject to more stringent winsorizations reported in Appendix Table 1.A.6) are statistically indistinguishable from zero.

Finally, in columns (3) and (4) of Table 1.4, we report RD estimates for calendar years 2017 and 2018, which are not critical years. During these years, individuals in our analysis sample are 63 and 64 years old. The majority of those retiring through the VERP scheme have already done so. Our RD estimates reported in the table show how responses in general have mostly dissipated during this time frame.[28]

Before moving on to further unpack our main results and investigate mechanisms, we first conduct a series of robustness checks, sensitivity analyses, and placebo exercises to further establish the validity of our main results. The upshot of these analyses is that our estimates are robust to standard RD specification checks, while several placebo tests provide reassuring evidence that our RD estimates indeed capture the causal effects of the policy reform.

1.6.3 Robustness and Specification Checks

We probe the robustness of our results along several dimensions by estimating our RD using various alternative specifications. We report results for the main outcomes in Appendix Table 1.A.3 (for the anticipation period), Appendix Table 1.A.4 (for critical year 2014), and Appendix Table 1.A.5 (for critical year 2016). The tables are constructed as follows. Each row indicates an alternative specification, and each column corresponds to a different outcome variable. Row A reproduces baseline estimates. In rows B through E, we vary the bandwidth, both increasing and decreasing the size of the bandwidth in one-week intervals. In row F, we use a global linear polynomial rather than separate linear polynomials on either side of the cutoff. In row G, we exclude controls, and in row H, we do not use

[28]The point estimates in 2017 and 2018 for changes in bank account balances are fairly large (around $600) but imprecisely estimated and statistically insignificant; additional winsorizing yields smaller point estimates (see Appendix Table 1.A.6).

triangular weights.

Overall, our results are stable. The point estimates for outcomes over the anticipation period are broadly similar to one another and never statistically distinguishable from zero. The point estimates during the critical years do not appear sensitive. The estimates for earnings as well as contributions to retirement accounts are almost always highly statistically significant and do not fluctuate meaningfully with specification choices, and the point estimates for other savings outcomes are never statistically distinguishable from zero.

1.6.4 Placebo Exercises

We additionally conduct three placebo exercises. First, we estimate our RD over a placebo time period. We test for discontinuous jumps in outcomes during the pre-announcement period from 2008 to 2010. There should be no discontinuities in outcomes due to the reform during this period, as the policy had not yet been announced. Indeed, Appendix Table 1.A.7 shows no statistically significant effects on any of the outcomes analyzed.

Second, we estimate our RD using placebo cutoffs around the true cutoff date. Appendix Figure 1.A.3 shows how our RD estimates for key outcome variables during each critical year shrink and become statistically insignificant as we use cutoffs further away from the true cutoff. We note that since we consistently use a bandwidth equal to 56 days on either side of the cutoff, the RD estimates corresponding to placebo cutoffs more than 56 days away from the true cutoff provide placebo estimates as proposed by Imbens and Lemieux (2008), since these estimates do not come from underlying data that contains a known discontinuity.

Finally, we replicate our entire analysis, but using placebo January 1 birthdate cutoffs for earlier birth cohorts who, to the best of our knowledge, are not impacted by policies that may result in discontinuities in outcomes as they age into the VERP program. Specifically, we implement our RD design first as if the cutoff was January 1, 1951, and then again as if the cutoff was January 1, 1952, testing for discontinuities in outcomes during the years these individuals reach their critical retirement ages of 60 and 62.[29] Appendix Table 1.A.8 reports the results; we find no evidence that being born just after these placebo January 1 cutoff dates impacts earnings or savings in retirement accounts at age 60 or 62.

1.7 Mechanisms

Taken together, the main results indicate deviations from benchmark theory and may point to inertial behavior as an underlying channel. We find that savings respond to the increase in eligibility ages only when the reform directly induces extended employment and only through retirement accounts. To explore mechanisms and directly assess the extent to which inertia might be driving the results, we first investigate the lack of anticipatory responses, and then we unpack the increases in contributions to retirement savings accounts during the two critical years.

1.7.1 Investigating the Lack of Anticipatory Savings Responses

Here we assess two natural alternative explanations for the lack of anticipatory responses other than inertia. First, it could be that a complete lack of awareness underlies the inaction: if individuals impacted by the reform are simply not aware of the changes to their eligibility ages until they reach age 60, then the lack of responses could be attributed to a deficiency of information. While we cannot rule out this explanation completely, we consider it an unlikely driving force behind the lack of anticipatory responses. In general, the major reform was well-publicized and a matter of political discourse. The later phases of the reform impact essentially all Danes younger than those that form our control group, and the reform is regarded as an initial push towards the gradual elimination of the VERP

[29] We do not use the January 1, 1953 birthdate as a placebo since a change in unemployment insurance policy for older individuals differentially impacted those born in 1953 compared to 1952 (OECD 2015).

program altogether.[30] Overall, we view our setting as one in which general awareness was likely high. For some reference, Appendix Figure 1.A.4 plots a Google search intensity index for "efterløn", which is the Danish word for the VERP program. The graph shows several large spikes in searches throughout the anticipation period.

A second candidate explanation could be the inability to respond. If "hand-to-mouth" or "wealthy hand-to-mouth" (Kaplan and Violante 2014, Kaplan et al. 2014) behavior is prevalent and individuals have little liquid financial assets, then it could be that they did not have room to adjust savings in response to the announcement of the reform. Two pieces of evidence suggest this is unlikely to be driving the null anticipatory responses in our context. First, average bank account balances for our analysis sample are relatively high (just over $26,000 in 2010) and constitute savings that are typically more liquid and easier to adjust. Second, we find no evidence of anticipatory responses when we estimate our RD using a subsample of individuals who are likely able to respond with more ease, namely those who had been using personal retirement plans before the announcement of the reform. These individuals have a natural way to respond—by adjusting their voluntary contributions to personal retirement plans—but also have higher bank account balances on average ($35,535) and may be more financially sophisticated. We report the corresponding results in Table 1.5. Column (1) shows no evidence of any anticipatory savings responses in any of the savings vehicles we study for this subsample.

[30]The prime minister of Denmark announced plans leading to the reform during his New Year's Day speech on the first day of 2011, while also suggesting an eventual elimination of the VERP program. Later phases of the reform make the entire scheme less financially attractive, and due to these changes, individuals wishing to opt out of the VERP program could in 2012 withdraw their contributions to the scheme. While likely a more attractive option for those younger than our analysis sample, we nonetheless investigate whether the reform impacted VERP participation at the birthdate cutoff we study. Appendix Table 1.A.9 reports results from estimating our RD on the likelihood of making participatory contributions to the VERP scheme and shows a lack of responses along this potential margin.

1.7.2 Investigating the Increased Savings in Retirement Accounts

We now turn to unpack the savings responses we find during the critical years, the large and meaningful increases in contributions to both employer-sponsored and personal retirement accounts.

Personal Retirement Savings Accounts

We start by investigating the increase in contributions to personal retirement plans. We study response heterogeneity by pre-announcement usage of these accounts. The goal is to assess whether the policy increases the likelihood of contributing for those using the accounts less regularly, or whether the average effect is mostly the result of continued contributions by those already using the accounts. To this end, we split the estimating sample into two groups: frequent users of personal plans (who contributed in either 2 or 3 years between 2008 and 2010) and infrequent users (who contributed in either 0 or 1 year between 2008 and 2010). We then estimate our RD on contributing to personal plans in each critical year separately for each group, and we report results in Table 1.6.

Consistent with inertia and the continuation of previous savings behaviors, we find that the savings response is driven entirely by frequent users. The point estimates for frequent users represent increases of around 30% for each critical year, and indicate that the policy results in continued contributions during periods of policy-induced extended employment from those who had been contributing before the announcement of the reform. The point estimates for infrequent users are small and statistically indistinguishable from zero; there is no evidence the reform spurs these individuals to take up contributing to personal plans.

Employer-Sponsored Retirement Savings Accounts

We next examine the increase in contributions to employer-sponsored retirement plans. The literature on retirement savings has shown firm policies such as firm default contribution rates to strongly influence wealth accumulation within retirement accounts (e.g., Madrian and Shea 2001, Beshears et al. 2009). This has been shown to be especially true in Denmark (Fadlon et al. 2016), where there is additional evidence that individuals save passively and that employer-sponsored plans can play a key role in driving overall wealth accumulation (Chetty et al. 2014). In Denmark, collective bargaining agreements between unions and employer associations often stipulate minimum contribution rates for workers, and among those not covered by these agreements, firms often set default contribution rates.

In the light of these institutional practices and the influential literature on firm savings policies, our findings of large increases in savings through employer-sponsored retirement plans in response to the reform inspires a natural question: to what extent do employers mediate savings responses to national reforms of social security systems? We exploit our linked employer-employee data to conduct two informative exercises that directly investigation this question. To this end, we use our population-wide data to construct firms, and we proxy for employer default contribution rates using the median contribution rates at firms. All of our analyses center on firm contribution rates defined in 2010, the year preceding the announcement of the reform, so as to avoid defining firm characteristics of an individual based on, e.g., the endogenous choice of workplace in periods after the announcement of the reform.[31]

Graphical Anlaysis. First, we conduct a graphical analysis that compares deviations from employer default contribution rates, for our treatment and control group, before and after the reform. Figure 1.8 depicts the results. Each graph plots the distribution of deviations from default contribute rates. For example, the large spikes around zero in graph

[31] Our approach to constructing firms and inferring firm-default contribution rates broadly follows related strategies in Chetty et al. (2014) and Fadlon et al. (2016). We construct firms using our data on all individuals in Denmark; we keep individuals over 18 years of age and assign them to firms. We then compute individual-specific contribution rates by dividing contributions to employer-sponsored retirement accounts by labor market earnings. We infer the default contribution rate of the firm as the median contribution rate among individuals at the firm. Our sample sizes decrease slightly for these analyses due to our inability to define workplaces in 2010 for every individual in our sample; roughly 6% of individuals did not have positive labor market earnings in 2010.

(a) show that individuals in both the treatment group and the control group tend to contribute at default rates; the fact that the two distributions lie on top of one other suggests that the propensity to deviate from the default rate did not differ by group in 2010, before the reform was announced. Graph (b) plots the same distributions during 2012; the graph shows no evidence that the behavior of the treatment and control group have diverged, despite the announcement of the reform. Graph (c) plots the distributions during 2014, the first critical year. The mass around zero has decreased more for the control group than the treatment group, with a corresponding rise in mass around negative ten percent, consistent with the control group beginning to retire and thus contributing less or not at all. (We note default contribution rates around 10% are common in Denmark.) In contrast, the mass of the treatment group remains higher around zero, suggesting they are more likely to still be contributing right around the default rate. The pattern continues in graph (d), the second critical year. This analysis points to an important role for employer defaults in shaping responses to the reform.

Regression Analysis. To better quantify the extent to which continuing to contribute at firm default rates can explain our findings, we conduct a regression-based analysis that compares actual contributions with predicted contributions according to default rates and earnings responses. Specifically, we define a new outcome variable, predicted contributions, as current earnings multiplied by the 2010 (pre-announcement period) firm default contribution rate, and we estimate our RD using this outcome. The RD estimate for predicted contributions captures the change in contributions to employer-sponsored plans that would arise from responding to the reform by continuing to work at the same firm, which increases earnings, and continuing to contribute out of those earnings at the default rate. We then compare the discontinuity in predicted contributions with the discontinuity in actual contributions. We report these results in Table 1.7. Column (1) reports the estimate for the impact of the policy on actual contributions in 2014, but for the subsample of individuals for whom we could define firm default contribution rates in 2010. The subsample is 93.7% of our main RD estimation sample, and the \$781 point estimate is very similar to our baseline estimate. Column (2) reports the estimate for the impact of the policy on predicted contributions in 2014, which is \$591. Taking these RD estimates at face value, the results indicate that in 2014, roughly $\frac{591}{781} = 76\%$ of the increase in contributions to employer-sponsored retirement accounts can be explained by continued contributions at firm default rates. Similarly, in 2016, the discontinuity in predicted contributions amounts to \$526, whereas the discontinuity in actual contributions is \$706, and thus firm default contribution rates can explain approximately 75% of the actual response during the second critical year. Overall, our results indicate that employers can play an important role in shaping how private savings ultimately respond to national social security reform.

1.8 Conclusion

In this paper, we provide novel evidence on the effects of increasing pension eligibility ages on private savings. We leverage rich, population-wide, linked employer-employee, administrative data on essentially the entire financial portfolio to study savings responses in a setting where strong labor supply incentives induce extended employment.

Our paper offers two main results. First, we find a lack of anticipatory responses, after the reform is announced but before it is implemented, inconsistent with the notion that future differences in pension eligibility impact current savings. Second, we find large and meaningful increases in contributions to retirement savings accounts—both personal plans and employer-sponsored plans—during periods of policy-induced extended employment. Then, through a series of additional analyses, we investigate mechanisms, and we view the overall body of evidence as pointing to inertia as a leading explanatory channel. In response to the reform, individuals continue working and continue saving in retirement accounts in a manner consistent with their behavior before the reform.

Our results carry important implications for policy. Pension eligibility ages are defining features of most social security systems, and similar reforms that increase eligibility ages have been enacted around the world in recent decades. A good deal of work investigates labor supply responses to these types of reforms, but understanding how raising eligibility ages will likely impact financial security throughout later stages of the lifecycle calls for an analysis of savings, a key resource used to finance consumption at older ages. We find that, in our setting, raising eligibility ages leads to longer working lives, increased earnings, and more private savings set aside in retirement accounts for shorter retirement time horizons.

1.9 Acknowledgements

Chapter 1, in full, is currently being prepared for submission for publication of the material. García-Miralles, Esteban and Leganza, Jonathan M. "Public Pensions and Private Savings." The dissertation author was a primary investigator and an author of this material.

Chapter 1 was supported by an NBER Pre-Doctoral Fellowship in Retirement and Disability Policy Research. The research was performed pursuant to grant RDR18000003 from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States (U.S.) Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the U.S. Government or any agency thereof.

1.10 Figures and Tables

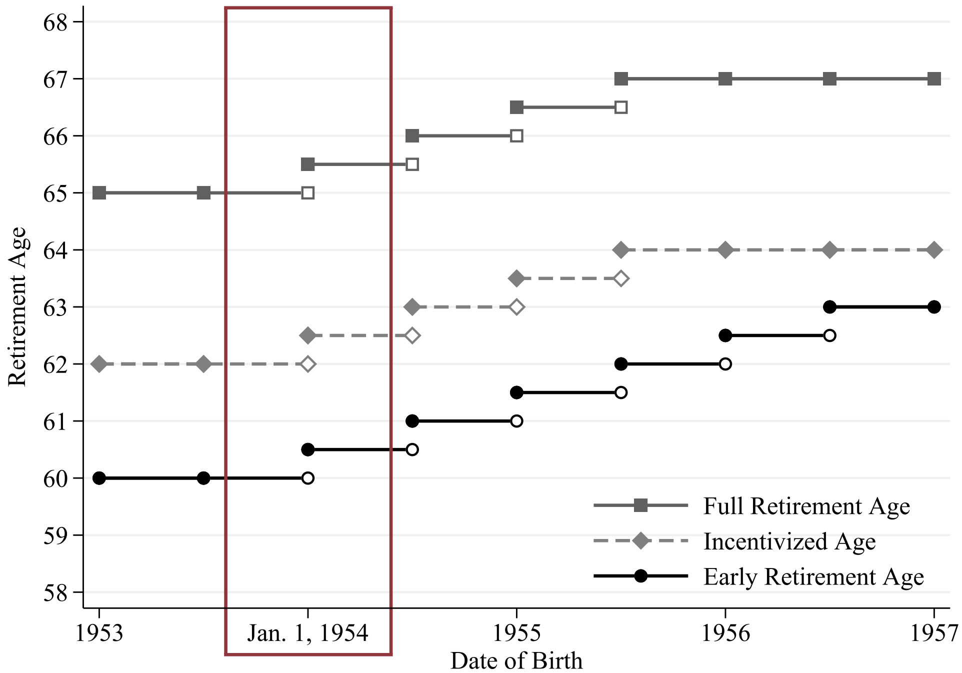


Figure 1.1: Pension Eligibility Ages by Birthdate

Notes: This figure graphically depicts the increases in pension eligibility ages due to the 2011 reform. Birth cohorts born before January 1, 1954 were unaffected by the reform. For these individuals, the key eligibility ages remained constant at 60, 62, and 65. Individuals born between January 1, 1954 and July 1, 1954 experience a six-month increase in each of the eligibility ages. Later phases of the reform introduced additional increases of eligibility ages as illustrated. The maroon rectangle highlights the birth cohorts relevant for our study.

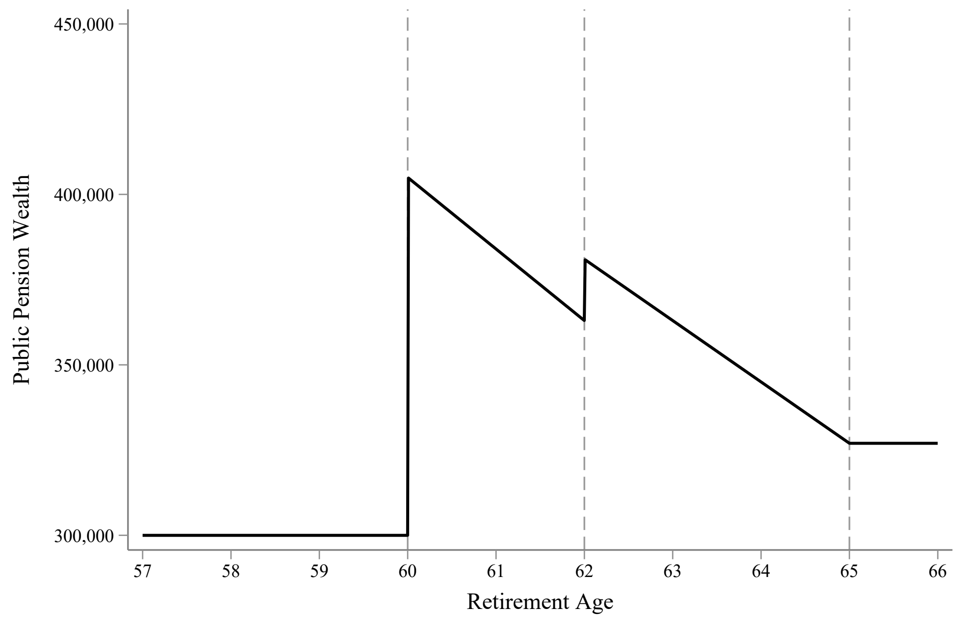


Figure 1.2: Pre-Reform Public Pension Wealth by Retirement Age

Notes: This figure plots public pension wealth against retirement age for a representative individual before the reform. For illustrative purposes, the benefit amounts depicted in the figure are for a worker who is married, who lives until age 85, and who has $250,000 in private retirement savings accounts at age 60. Note the y-intercept in the stylized graph is not zero, due to receiving OAP benefits after the early retirement program. The first spike in pension wealth at age 60 is due to the transition rule. Individuals retiring before 60 are not eligible to claim into the early retirement program and thus forfeit five years of early retirement benefits. The second spike in pension wealth at age 62 is due to the two-year rule. Retiring at age 62 eliminates the means-testing of early retirement benefits against private retirement savings accounts and produces higher benefits over the remaining three years of the early retirement program. The negative slopes between 60 and 62 and between 62 and 65 result from the lack of actuarial adjustments when deferring claiming. Pension wealth for those who retire after age 65 is greater than OAP wealth due to bonus payments for working past age 62 (see Appendix 1.B).

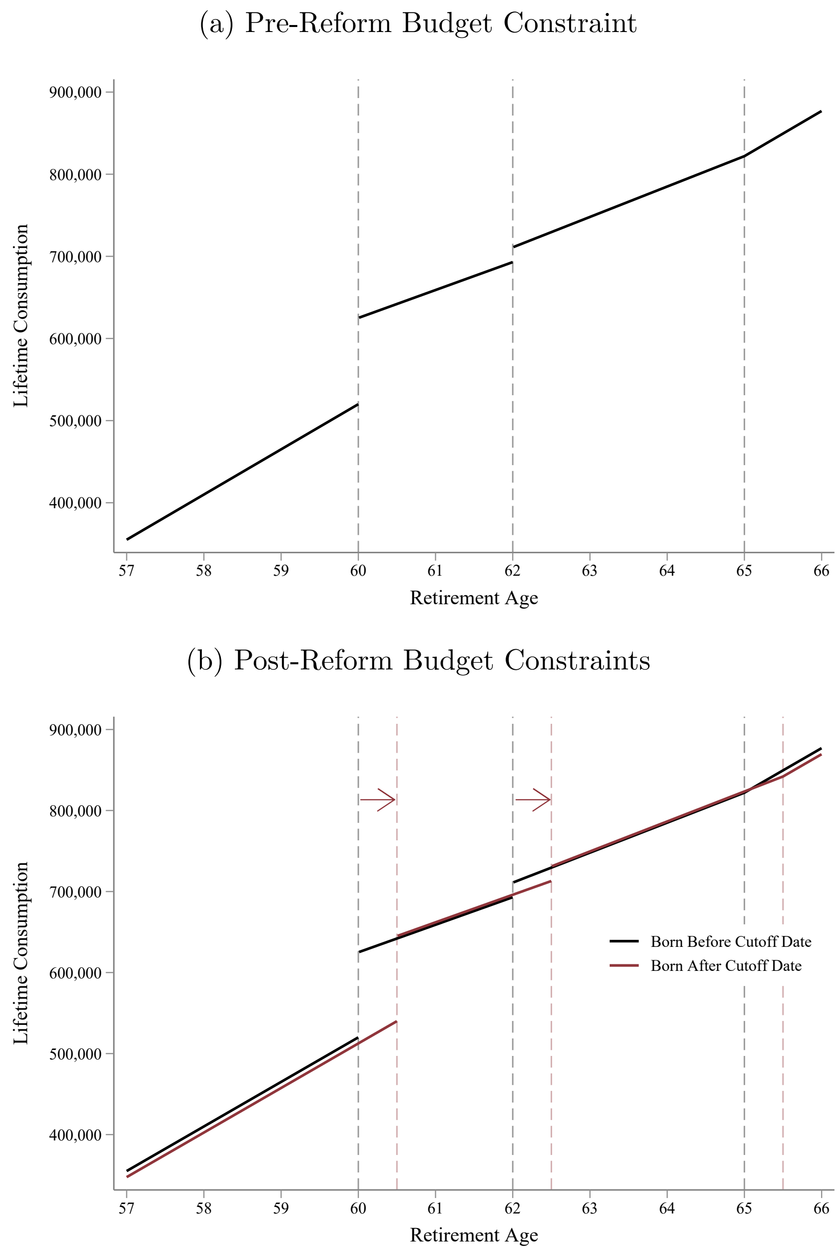
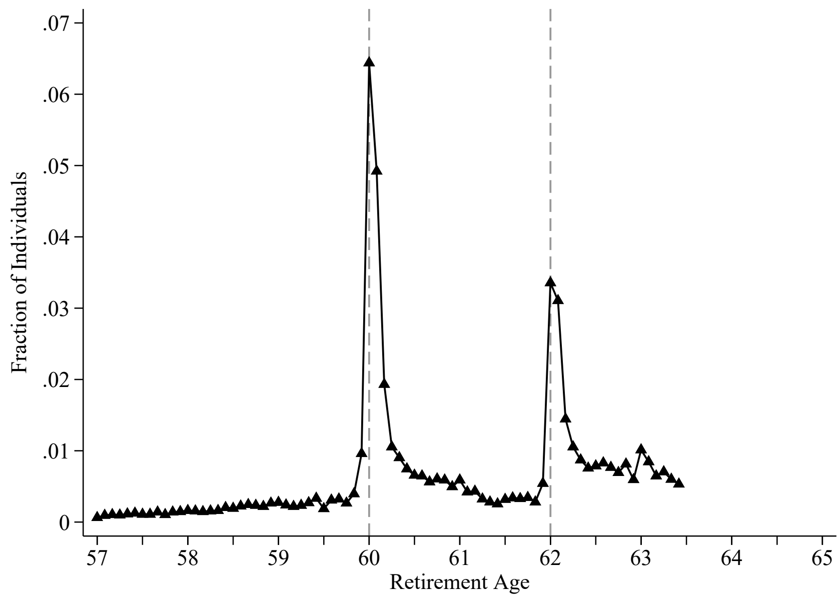


Figure 1.3: Lifetime Budget Constraints

Notes: This figure plots lifetime consumption against retirement age for the same representative worker as in Figure 1.2. Lifetime consumption is the sum of public pension wealth and lifetime earnings. For illustrative purposes, annual earnings are assumed to be $55,000 and lifetime earnings are earnings after age 57, the age of our sample when the reform is announced. Graph (a) depicts the lifetime budget constraint the worker faces before the reform. The spikes in pension wealth at age 60 and 62 translate to discontinuities in lifetime consumption. Graph (b) illustrates how the budget constraint changes due to the reform. If the worker was before the January 1, 1954 cutoff, the budget constraint is governed by the black line. If the worker was born on or after the cutoff, the budget constraint is governed by the maroon line. The key difference is the change in the location of the discontinuities in lifetime consumption.

(a) Retirement Distribution for the Control Group



(b) Retirement Distributions for Treatment and Control Groups

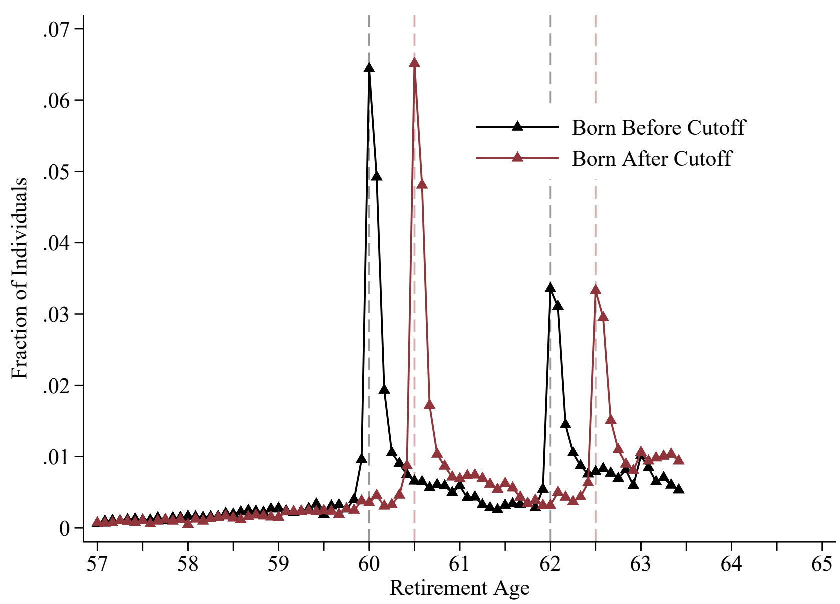


Figure 1.4: Empirical Distributions of Retirement Ages

Notes: This figure plots empirical distributions of retirement ages. Retirement is measured as an absorbing state. Monthly retirement age is defined as the age of the individual in the last month during which earnings are positive, before permanently falling to zero. Graph (a) shows how those born before the January 1, 1954 birthdate cutoff tend to either retire right around 60 or 62. Graph (b) shows how, in response to the reform, those born after the birthdate cutoff tend to retire right around $60\frac{1}{2}$ or $62\frac{1}{2}$.

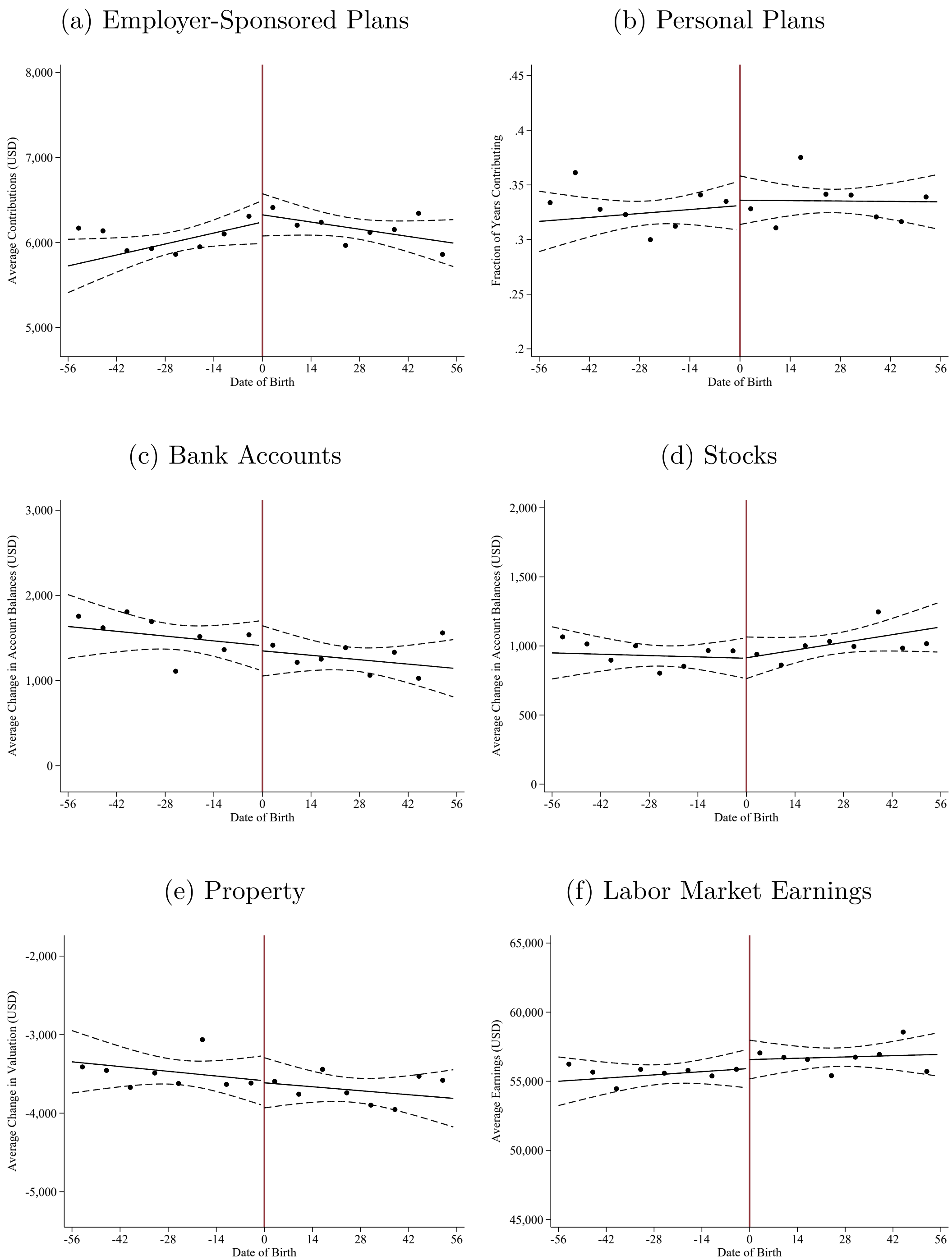


Figure 1.5: Responses Over the Anticipation Period

Notes: This figure illustrates the effect of the reform on key outcome variables over the anticipation time period. Each RD graph (a)–(f) corresponds to a separate outcome variable averaged over the three-year anticipation period, from 2011 to 2013. The graphs plot average outcomes in one-week date-of-birth bins. The maroon vertical lines designate the January 1, 1954 birthdate cutoff. The superimposed regression lines and 95-percent confidence intervals are based on the underlying unbinned data.

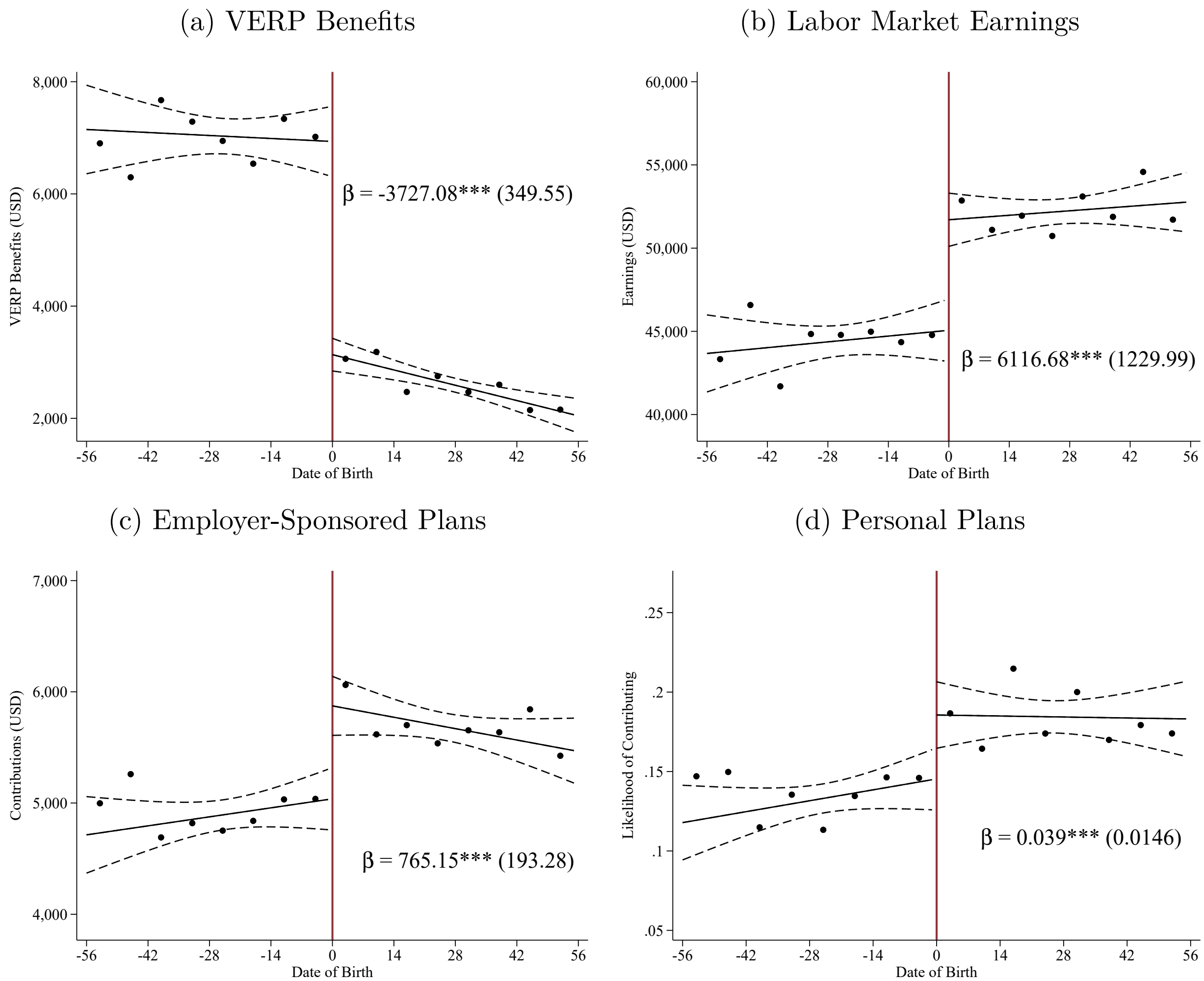


Figure 1.6: Responses During the First Critical Year 2014

Notes: This figure illustrates the effect of the reform on labor market outcomes and contributions to retirement accounts during the first critical year, when individuals born at the cutoff date are age 60. Each RD graph (a)–(d) plots average outcomes during 2014 in one-week date-of-birth bins. The maroon vertical lines indicate the January 1, 1954 birthdate cutoff. The superimposed regression lines and 95-percent confidence intervals are based on the underlying unbinned data. The RD estimates reported in the figures correspond to those in Table 1.3, and come from estimating equation (1.6).

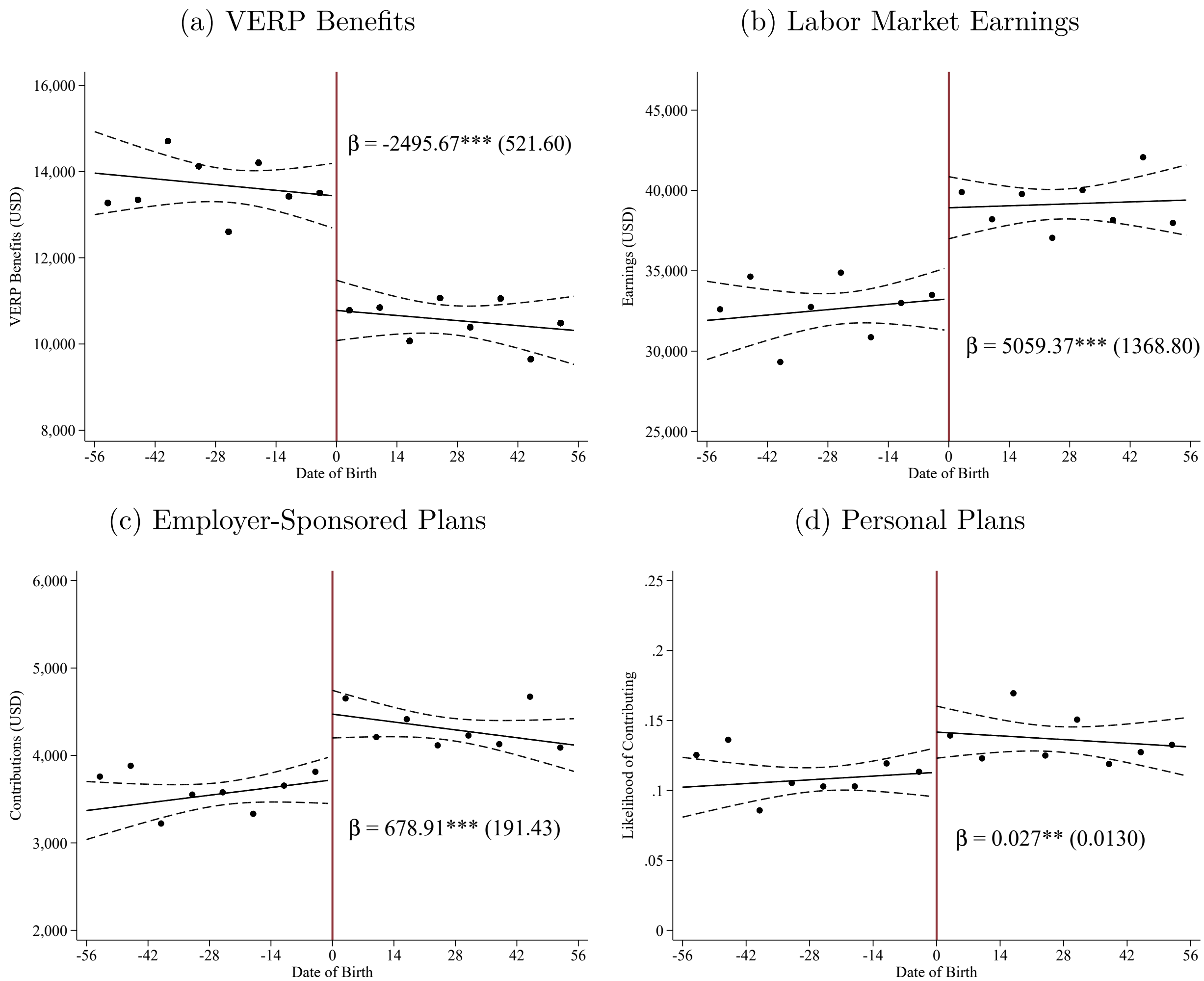
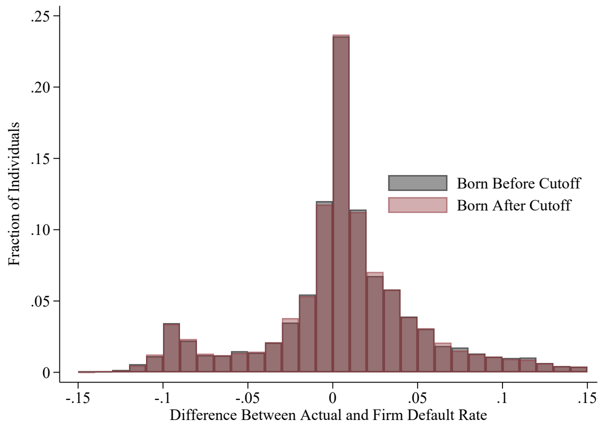


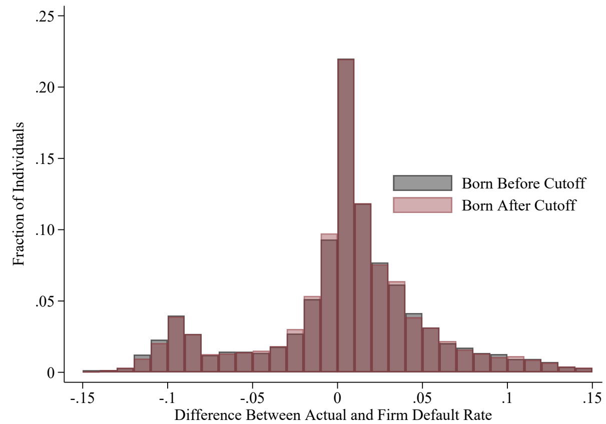
Figure 1.7: Responses During the Second Critical Year 2016

Notes: This figure illustrates the effect of the reform on labor market outcomes and contributions to retirement accounts during the second critical year, when individuals born at the cutoff date are age 62. Each RD graph (a)–(d) plots average outcomes during 2016 in one-week date-of-birth bins. The maroon vertical lines indicate the January 1, 1954 birthdate cutoff. The superimposed regression lines and 95-percent confidence intervals are based on the underlying unbinned data. The RD estimates reported in the figures correspond to those in Table 1.3, and come from estimating equation (1.6).

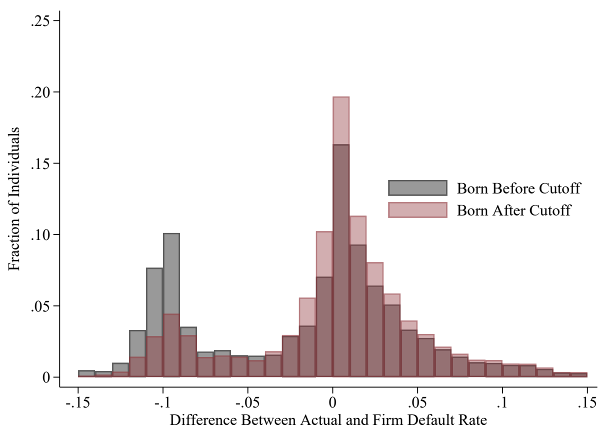
(a) Pre-Announcement Period: Year 2010



(b) Anticipation Period: Year 2012



(c) First VERP Critical Year: 2014



(d) Second VERP Critical Year: 2016

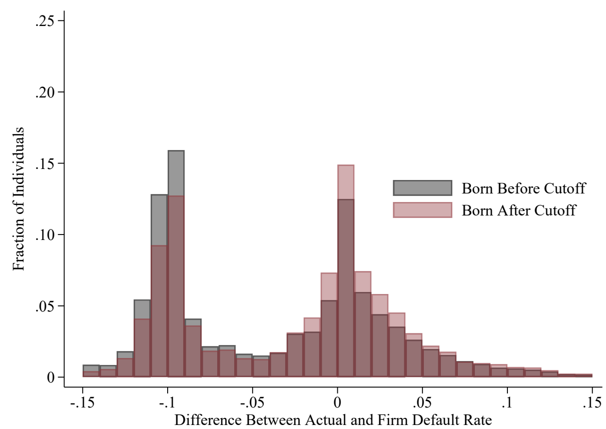


Figure 1.8: Differences Between Actual and Firm Default Contribution Rates

Notes: This figure illustrates how actual contribution rates to employer-sponsored retirement plans deviate from firm default contribution rates, over time, for both the treatment and control group. Firm default contribution rates are inferred as the median contribution rate among individuals working at the firm, as described in Section 1.7.2. Each graph (a)-(d) captures the distributions of deviations from firm default rates during a different year.

Table 1.1: Summary Statistics

	Analysis Sample		RD Sample	
	Mean (1)	SD (2)	Mean (3)	SD (4)
A: Demographics				
Age	56.99	0.29	56.99	0.09
Male	0.46	0.50	0.46	0.50
Married	0.72	0.45	0.72	0.45
Treated	0.52	0.50	0.52	0.50
B: Labor Market Earnings				
Any Earnings	0.94	0.23	0.94	0.24
Earnings	61,380	35,013	60,912	34,355
C: Retirement Savings (Flow Variables)				
Any Contribution to Employer Plans	0.89	0.32	0.89	0.32
Contributions to Employer Plans	6,508	4,951	6,430	4,888
Any Contribution to Personal Plans	0.41	0.49	0.41	0.49
Contributions to Personal Plans	1,192	2,130	1,171	2,111
D: Other Savings (Stock Variables)				
Bank Account Balances	26,505	46,790	26,238	45,558
Stock Market Account Balances	7,240	44,006	7,136	46,094
Property Wealth	152,541	189,923	151,354	182,384
Number of Individuals	40,042		12,020	

Notes: This table reports means and standard deviations of key variables, for the analysis sample and the main RD estimation sample, in 2010, the year before the reform. The analysis sample consists of a balanced panel of individuals born within six months of the January 1, 1954 birthdate cutoff who were making participatory contributions to the early retirement scheme and who were not self-employed. The main RD estimation sample consists of the subset of individuals from the analysis sample who were born within 56 days of the birthdate cutoff.

Table 1.2: Responses Over the Anticipation Period

	Years: 2011–2013	
	RD Estimate (1)	Mean (2)
A: Labor Supply		
Average Earnings	186.09	55,621
	(992.59)	
B: Retirement Accounts		
Average Contributions to Employer Plans	20.32	6,048
	(177.95)	
Fraction of Years Contributing to Personal Plans	0.005	0.33
	(0.016)	
C: Other Savings		
Average Change in Bank Accounts	-66.22	1,543
	(213.31)	
Average Change in Stock Market Accounts	-4.00	944
	(107.33)	
Average Change in Property Wealth	-31.048	-3,494
	(225.04)	
Obs.	12,020	

Notes: This table reports RD estimates for the impact of the reform on outcomes over the anticipation period. Outcome variables are averaged over 2011 to 2013. Panel A presents results for labor supply outcomes. Panel B presents results for contributions to retirement savings accounts. Panel C presents results for savings through bank accounts, stock market accounts, and property. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.3: Responses During Early Retirement Period Critical Years

	Critical Year: 2014		Critical Year: 2016	
	RD Estimate (1)	Mean (2)	RD Estimate (3)	Mean (4)
A: Labor Supply				
VERP Benefits	-3727.08***	6,995	-2495.67***	13,634
	(349.55)		(521.60)	
Earnings	6116.68***	44,449	5059.37***	32,737
	(1229.99)		(1368.80)	
B: Retirement Accounts				
Contributions to Employer Plans	765.15***	4,928	678.91***	3,603
	(193.28)		(191.43)	
Any Contribution to Personal Plans	0.039***	0.14	0.027**	0.11
	(0.0146)		(0.0130)	
Distributions from Retirement Plans	-262.92***	1,584	-236.23	2,467
	(88.22)		(163.73)	
C: Other Savings				
Change in Bank Accounts	-120.84	1,876	370.12	801
	(469.46)		(468.54)	
Change in Stock Market Accounts	-295.57	1,843	31.56	312
	(211.15)		(86.43)	
Change in Property Wealth	-6.54	-522	0.40	-649
	(22.03)		(27.09)	
Obs.	12,020		12,020	

Notes: This table reports RD estimates for the impact of the reform on outcomes during the early retirement period critical years. Column (1) displays results during 2014, when individuals born at the cutoff date are age 60. Column (3) displays results during 2016, when individuals born at the cutoff date are age 62. Panel A presents results for labor supply outcomes. Panel B presents results for contributions to (and distributions from) retirement savings accounts. Panel C presents results for savings through bank accounts, stock market accounts, and property. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

Table 1.4: Responses During Early Retirement Period Non-Critical Years

	Year: 2015		Year: 2017		Year: 2018	
	RD Estimate (1)	Mean (2)	RD Estimate (3)	Mean (4)	RD Estimate (5)	Mean (6)
A. Labor Supply						
VERP Benefits	-548.92	8,262	-1006.78**	16,872	-856.75	17,236
	(481.58)		(583.33)		(583.33)	
Earnings	1925.14	41,251	2780.50**	27,032	805.75	24,133
	(1387.34)		(1356.15)		(1329.58)	
B: Retirement Accounts						
Contributions to Employer Plans	327.76*	4,575	258.31	3,023	36.68	2,476
	(198.93)		(182.52)		(170.67)	
Any Contribution to Personal Plans	0.015	0.12	0.006	0.10	0.004	0.10
	(0.014)		(0.012)		(0.012)	
Distributions from Retirement Plans	-141.34	1,956	-123.96	2,834	-51.86	3,282
	(132.87)		(195.70)		(213.84)	
C: Other Savings						
Change in Bank Accounts	-414.15	1,192	622.01	-17	610.25	4,229
	(476.21)		(467.66)		(557.27)	
Change in Stock Market Accounts	92.70	1,738	-51.86	1,193	-61.06	-1,754
	(236.40)		(163.55)		(184.00)	
Change in Property Wealth	15.30	-960	-18.78	-1,313	-56.47	-1,040
	(42.32)		(56.41)		(45.07)	
Obs.	12,020		12,020		12,020	

Notes: This table reports RD estimates for the impact of the reform on outcomes during the early retirement period non-critical years. Column (1) displays results during 2015, when individuals born at the cutoff date are age 61. Column (3) displays results during 2017, when individuals born at the cutoff date are age 63. Column (5) displays results during 2018, when individuals born at the cutoff date are age 64. Panel A presents results for labor supply outcomes. Panel B presents results for contributions to (and distributions from) retirement savings accounts. Panel C presents results for savings through bank accounts, stock market accounts, and property. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.1$

Table 1.5: Anticipatory Responses for Users of Personal Retirement Plans

	RD Estimate (1)	Mean (2)
A: Labor Supply		
Earnings	-84.24	56,739
	(1486.16)	
B: Retirement Accounts		
Contributions to Employer Plans	319.51	5,962
	(265.25)	
Any Contribution to Personal Plans	0.001	0.71
	(0.019)	
C: Other Savings		
Change in Bank Accounts	68.07	1,554
	(347.15)	
Change in Stock Market Accounts	70.29	1,157
	(174.67)	
Change in Property Wealth	115.99	-3,712
	(344.96)	
Obs.	5,015	

Notes: This table reports RD estimates for the impact of the reform on outcomes over the anticipation time period for the subsample of individuals who had been using personal retirement plans before the announcement of the reform. The subsample is defined as those who made contributions to personal plans in either two or three of the years between 2008 and 2010. Outcome variables are averaged over 2011 to 2013. Panel A presents results for labor supply outcomes. Panel B presents results for contributions to retirement savings accounts. Panel C presents results for savings through bank accounts, stock market accounts, and property. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.6: Contributions to Personal Retirement Plans by Previous Use

	RD Estimate (1)	Mean (2)
A. Frequent Users		
Any Contribution to Personal Plans in 2014	0.095***	0.28
	(0.029)	
Any Contribution to Personal Plans in 2016	0.062**	0.21
	(0.026)	
Obs.	5,015	
B. Infrequent Users		
Any Contribution to Personal Plans in 2014	-0.001	0.04
	(0.011)	
Any Contribution to Personal Plans in 2016	0.003	0.04
	(0.010)	
Obs.	7,005	

Notes: This table reports RD estimates for the impact of the reform on contributions to personal retirement plans during critical years 2014 and 2016, by previous use of the accounts. Panel A reports results for the subsample of individuals who made contributions to personal plans in either two or three of the years between 2008 and 2010. Panel B reports results for the subsample of individuals who made contributions in either 0 or 1 year between 2008 and 2010. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

Table 1.7: Actual vs. Predicted Contributions to Employer Retirement Plans

	RD Estimates	
	Actual Contributions (1)	Predicted Contributions (2)
Contributions in 2014	781.32***	590.74***
	(198.93)	(172.85)
Contributions in 2016	705.64***	525.63***
	(199.05)	(185.82)
Obs.	11,259	11,259

Notes: This table reports RD estimates for the impact of the reform on actual contributions to employer-sponsored retirement plans as well as predicted contributions to employer-sponsored retirement plans, during both critical years 2014 and 2016. Predicted contributions are defined as current earnings multiplied by the 2010 inferred firm default contribution rate. Firm default contribution rates are inferred as the median contribution rate among individuals working at the firm, as described in Section 1.7.2. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}\ p < 0.01$, $^{**}\ p < 0.05$, $^{*}\ p < 0.1$

1.A Appendix: Additional Figures and Tables

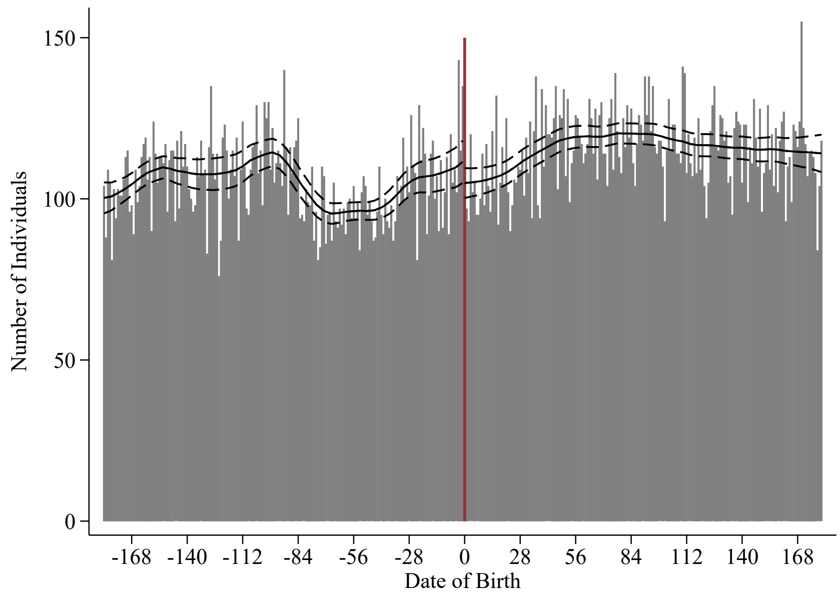


Figure 1.A.1: Histogram of the Running Variable

Notes: This figure depicts the density of the running variable, birthdate. The graph plots a histogram of the running variable for the entire analysis sample. Superimposed on top of the histogram are smoothed values and confidence intervals from local polynomial regressions of the number of individuals on birthdate. A formal density test as proposed by Cattaneo et al. (2019) using our baseline RD bandwidth of 56 days results in a p-value of 0.97.

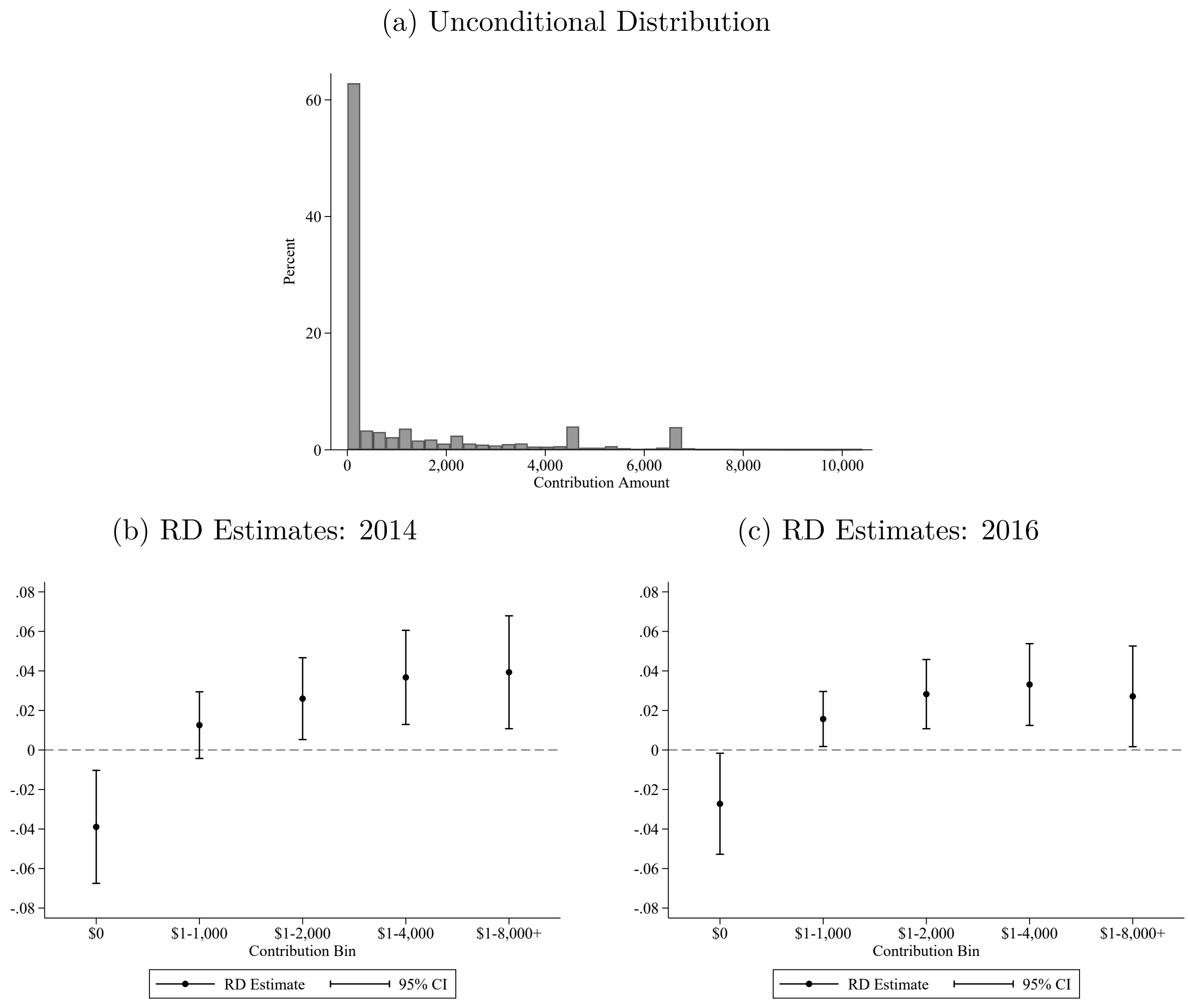


Figure 1.A.2: Analyzing Contribution Amounts to Personal Retirement Plans

Notes: This figure illustrates the method of analyzing contribution amounts to personal retirement plans. Graph (a) plots the unconditional distribution of contribution amounts in 2010. The large number of small and zero contributions show why analyzing average contributions in levels is difficult. We use five indicator variables that capture contributions (i) that amount to $0, (ii) that are between $1 and the $1,000, (iii) that are between $1 and $2,000, (iv) that are between $1 and $4,000, and (v) that are greater than $1. Graph (b) plots the RD estimates from estimating equation (1.6) using as outcomes these indicator variables in 2014. Graph (d) plots the results for 2016.

Figure 1.A.3: Placebo Exercise: Pseudo Birthdate Cutoffs

Notes: This figure illustrates how the RD estimates for labor market earnings and contributions to retirement plans, during each of the two critical years, change when placebo cutoffs are used rather than the true cutoff. Each graph (a)–(f) plots RD estimates and 95-percent confidence intervals from using the baseline RD estimating specification at various pseudo cutoffs.

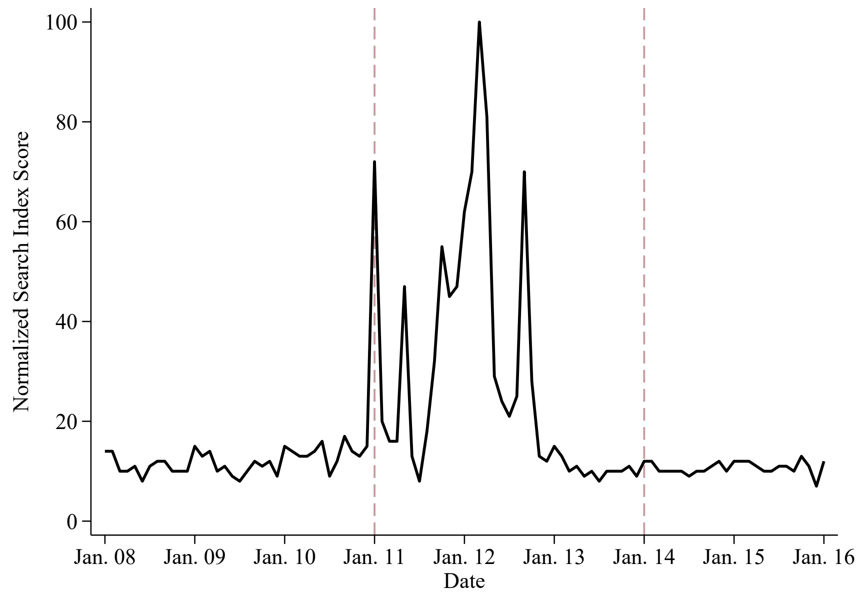


Figure 1.A.4: Google Searches for Efterløn

Notes: This figure plots a Google Trends search intensity index for "efterløn," which is the Danish word for the VERP program, between January 1, 2008 and January 1, 2016.

Table 1.A.1: RD Estimates for Control Variables as Outcomes

	RD Estimate (1)	Mean (2)
Male	0.026	0.47
	(0.020)	
Married	0.018	0.69
	(0.018)	
Hovedstaden	-0.003	0.12
	(0.013)	
Sjælland	-0.010	0.25
	(0.017)	
Syddanmark	-0.005	0.24
	(0.017)	
Midtjylland	0.022	0.24
	(0.017)	
Nordjylland	-0.005	0.15
	(0.014)	
Obs.	12,020	

Notes: This table reports RD estimates for the impact of the reform on (pre-determined) control variables. Control variables include an indicator for being male, an indicator for being married in 2010, and indicators for residing in each of the five regions of Denmark in 2010. The five regions are Hovedstaden (the capital region containing Copenhagen), Sjælland, Syddanmark, Midtjylland (containing Aarhus), and Nordjylland. The RD estimates come from estimating equation (1.6), except without any control variables on the right-hand side, but rather control variables on the left-hand side as outcomes. The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff and employ triangular weights. Robust standard errors are in parentheses. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.1$

Table 1.A.2: RD Estimates for Contributions to Roth-Style Plans

	Personal Plans		Employer Plans	
	RD Estimate (1)	Mean (2)	RD Estimate (3)	Mean (4)
Contribute in 2013	0.001	0.08	-0.003	0.02
	(0.011)		(0.004)	
Contribute in 2014	-0.010	0.12	0.003	0.01
	(0.013)		(0.004)	
Contribute in 2015	-0.007	0.14	0.001	0.01
	(0.014)		(0.004)	
Contribute in 2016	-0.015	0.15	0.000	0.01
	(0.014)		(0.004)	
Contribute in 2017	-0.004	0.16	0.002	0.01
	(0.015)		(0.004)	
Contribute in 2018	-0.022	0.18	-0.000	0.06
	(0.015)		(0.010)	
Obs.	12,020		12,020	

Notes: This table reports RD estimates for the impact of the reform on the likelihood of making any contribution to "Roth-style" retirement accounts. Outcome variables for both contributions to employer-sponsored and personal accounts are indicator variables for making any contribution to the plans. Roth-style plans were first introduced to the Danish economy in 2013. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.A.3: Robustness to Alternative Specifications: Anticipatory Responses

	Employer Plans (1)	Personal Plans (2)	Bank Accounts (3)	Stocks (4)	Property (5)	Earnings (6)
A. Baseline	20.32	0.005	-66.22	-4.00	-31.05	186.09
	(177.95)	(0.016)	(213.31)	(107.33)	(225.04)	(992.59)
B. 70 Day Bandwidth	98.64	0.011	-60.89	32.61	-120.58	569.84
	(159.24)	(0.014)	(190.07)	(96.09)	(201.64)	(891.63)
C. 63 Day Bandwidth	71.97	0.009	-69.92	16.33	-75.47	392.29
	(167.83)	(0.015)	(200.72)	(101.23)	(212.36)	(938.25)
D. 49 Day Bandwidth	-32.61	-0.003	-94.29	-37.65	10.80	64.51
	(190.26)	(0.017)	(228.75)	(114.77)	(240.53)	(1058.08)
E. 42 Day Bandwidth	-55.72	-0.013	-142.11	-48.97	50.48	114.30
	(205.40)	(0.019)	(247.95)	(123.95)	(259.55)	(1138.09)
F. Global Polynomial	32.87	0.005	-66.58	-8.23	-31.31	190.27
	(177.95)	(0.016)	(213.34)	(107.39)	(225.10)	(992.24)
G. No Controls	84.95	0.005	-60.92	3.34	-26.18	645.80
	(180.98)	(0.016)	(213.34)	(107.62)	(230.82)	(1016.33)
H. No Triangular Weights	158.40	0.017	-89.89	55.05	-138.87	712.59
	(163.18)	(0.015)	(195.47)	(98.94)	(207.14)	(917.50)

Notes: This table reports results from assessing the sensitivity of the RD estimates over the anticipation time period to various specification checks. Each column corresponds to a different main outcome variable. Each row indicates the specification choice and how it differs from the baseline specification. Row A reproduces baseline estimates for ease of comparison. Row B increases the bandwidth by two weeks. Row C increases the bandwidth by one week. Row D decreases the bandwidth by one week. Row E decreases the bandwidth by two weeks. Row F uses a global linear polynomial rather than two separate linear polynomials on either side of the cutoff. Row G drops control variables from the regressions. Row H does not use triangular weights. Robust standard errors are in parentheses. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

Table 1.A.4: Robustness to Alternative Specifications: Critical Year 2014

	Employer Plans (1)	Personal Plans (2)	Bank Accounts (3)	Stocks (4)	Property (5)	Earnings (6)
A. Baseline	765.15***	0.039***	-120.84	-295.57	-6.54	6116.68***
	(193.28)	(0.0146)	(469.46)	(211.15)	(22.03)	(1229.99)
B. 70 Day Bandwidth	797.83***	0.046***	-128.95	-206.62	-16.54	6275.60***
	(172.67)	(0.0131)	(420.78)	(188.02)	(19.69)	(1101.65)
C. 63 Day Bandwidth	793.50***	0.043***	-135.19	-247.66	-10.98	6203.09***
	(182.14)	(0.0138)	(443.17)	(198.54)	(20.77)	(1160.65)
D. 49 Day Bandwidth	733.37***	0.034**	-46.97	-366.64	-4.36	6079.54***
	(206.73)	(0.0156)	(501.41)	(226.63)	(23.57)	(1313.79)
E. 42 Day Bandwidth	725.82***	0.029*	102.11	-398.01	-2.07	6183.88***
	(223.17)	(0.0168)	(540.60)	(245.76)	(25.49)	(1415.75)
F. Global Polynomial	775.62***	0.039***	-63.97	-300.77	-7.47	6114.95***
	(193.07)	(0.015)	(469.43)	(210.48)	(22.06)	(1224.64)
G. No Controls	835.91***	0.040***	-118.17	-274.96	-15.12	6641.61***
	(196.79)	(0.015)	(469.41)	(211.75)	(22.49)	(1257.63)
H. No Triangular Weights	859.49***	0.051***	-108.84	-160.06	-11.55	6387.30***
	(176.47)	(0.0134)	(431.89)	(191.71)	(20.16)	(1130.06)

Notes: This table reports results from assessing the sensitivity of the RD estimates during the first critical year of 2014 to various specification checks. Each column corresponds to a different main outcome variable. Each row indicates the specification choice and how it differs from the baseline specification. Row A reproduces baseline estimates for ease of comparison. Row B increases the bandwidth by two weeks. Row C increases the bandwidth by one week. Row D decreases the bandwidth by one week. Row E decreases the bandwidth by two weeks. Row F uses a global linear polynomial rather than two separate linear polynomials on either side of the cutoff. Row G drops control variables from the regressions. Row H does not use triangular weights. Robust standard errors are in parentheses. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

Table 1.A.5: Robustness to Alternative Specifications: Critical Year 2016

	Employer Plans (1)	Personal Plans (2)	Bank Accounts (3)	Stocks (4)	Property (5)	Earnings (6)
A. Baseline	678.91***	0.027**	370.12	31.56	0.40	5059.37***
	(191.44)	(0.0130)	(468.54)	(86.43)	(27.09)	(1368.80)
B. 70 Day Bandwidth	721.07***	0.031***	388.02	53.57	-5.59	5289.61***
	(171.07)	(0.0117)	(418.65)	(77.16)	(24.33)	(1226.75)
C. 63 Day Bandwidth	716.87***	0.029**	388.26	39.28	-1.50	5251.99***
	(180.44)	(0.0123)	(441.54)	(81.41)	(25.60)	(1292.08)
D. 49 Day Bandwidth	649.15***	0.023*	370.55	35.03	-0.60	4959.92***
	(204.72)	(0.0139)	(501.45)	(92.61)	(28.92)	(1461.54)
E. 42 Day Bandwidth	647.06***	0.019	359.95	43.67	-3.03	5063.54***
	(220.98)	(0.0150)	(542.03)	(100.34)	(31.20)	(1574.97)
F. Global Polynomial	688.72***	0.028**	369.31	35.70	0.68	5062.84***
	(191.49)	(0.0131)	(467.28)	(86.52)	(27.12)	(1368.00)
G. No Controls	751.88***	0.029**	390.65	34.99	-10.00	5672.21***
	(196.15)	(0.0131)	(468.85)	(86.53)	(27.64)	(1410.79)
H. No Triangular Weights	766.20***	0.037***	412.22	30.601	2.82	5535.10***
	(175.00)	(0.0121)	(427.67)	(78.68)	(25.09)	(1260.00)

Notes: This table reports results from assessing the sensitivity of the RD estimates during the second critical year of 2016 to various specification checks. Each column corresponds to a different main outcome variable. Each row indicates the specification choice and how it differs from the baseline specification. Row A reproduces baseline estimates for ease of comparison. Row B increases the bandwidth by two weeks. Row C increases the bandwidth by one week. Row D decreases the bandwidth by one week. Row E decreases the bandwidth by two weeks. Row F uses a global linear polynomial rather than two separate linear polynomials on either side of the cutoff. Row G drops control variables from the regressions. Row H does not use triangular weights. Robust standard errors are in parentheses. $^{***} p < 0.01$, $^{**} p < 0.05$, $^{*} p < 0.1$

Table 1.A.6: Additional Winsorizing of Flow Savings Variables Computed From Stock Variables

	Bank Accounts (1)	Stocks (2)	Property (3)
Anticipation	-59.41	-16.51	34.53
	(151.26)	(32.90)	(174.59)
2014	-37.27	-48.33	4.96
	(331.04)	(57.80)	(17.95)
2015	-293.30	32.47	20.91
	(328.95)	(53.75)	(32.73)
2016	423.54	5.52	14.94
	(328.00)	(20.14)	(22.15)
2017	473.24	-3.43	5.12
	(327.48)	(35.83)	(44.33)
2018	301.88	-59.63	-10.86
	(408.59)	(89.76)	(34.54)
Obs.		12,020	

Notes: This table reports additional RD estimates for the impact of the reform on savings in bank accounts, stock market accounts, and property, where outcome variables are more-stringently winsorized at the 10th and 90th percentiles. The columns denote the different type of savings vehicle, and the rows indicate the time period. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.A.7: Placebo Exercise: Pre-Announcement Period

	Years: 2008–2010	
	RD Estimate (1)	Mean (2)
A: Labor Supply		
Earnings	692.77	59,778
	(890.49)	
B: Retirement Accounts		
Contributions to Employer-Sponsored Plans	-4.76	6,607
	(195.79)	
Any Contribution to Personal Plans	-0.003	0.25
	(0.018)	
C: Other Savings		
Change in Bank Accounts	-110.57	1,427
	(209.89)	
Change in Stock Market Accounts	-29.54	-186
	(45.04)	
Change in Property Wealth	-122.54	-12,614
	(615.83)	
Obs.	12,020	

Notes: This table reports RD estimates on outcomes over the pre-announcement placebo time period. Outcome variables are averaged over 2008 to 2010. Panel A presents results for labor supply outcomes. Panel B presents results for contributions to retirement savings accounts. Panel C presents results for savings through bank accounts, stock market accounts, and property. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.A.8: Placebo Exercise: Previous Birth Cohorts

	First Critical Year	Second Critical Year
	RD Estimate (1)	RD Estimate (2)
A: 1950/1951 Birth Cohorts		
Earnings	-729.20	-1194.96
	(1283.84)	(1331.95)
Contributions to Employer Plans	-215.25	-131.75
	(204.14)	(179.62)
Any Contribution to Personal Plans	0.013	-0.004
	(0.0192)	(0.0137)
Obs.	11,788	11,788
B: 1951/1952 Birth Cohorts		
Earnings	706.59	1243.32
	(1293.11)	(1344.74)
Contributions to Employer Plans	166.52	101.42
	(197.36)	(184.75)
Any Contribution to Personal Plans	0.016	0.004
	(0.019)	(0.014)
Obs.	11,810	11,810

Notes: This table reports RD estimates during "critical years" for placebo birth cohorts. Panel A presents results for earnings and contributions to retirement savings accounts using January 1, 1951 as a placebo birthdate cutoff. Column (1) presents results for the year that individuals born on this placebo birthdate cutoff are age 60. Column (2) presents results for the year that individuals born on this placebo birthdate cutoff are age 62. Panel B presents results when using January 1, 1952 as a placebo birthdate cutoff. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, (pre-determined) marital status, and (pre-determined) indicators for region of residence. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.A.9: RD Estimates for VERP Participation

	RD Estimate (1)	Mean (2)
Participate in 2011	-0.003	0.94
	(0.0090)	
Participate in 2012	0.005	0.93
	(0.0099)	
Participate in 2013	-0.009	0.92
	(0.0106)	
Obs.	12,020	

Notes: This table reports RD estimates for the impact of the reform on participatory VERP contributions. The outcome variables are indicators for making qualified contributions to UI funds in each of the three years leading up to the implementation of the reform. The RD estimates come from estimating equation (1.6). The regressions use separate linear polynomials in the running variable on either side of the birthdate cutoff, employ triangular weights, and include as controls gender, marital status as of 2010, and indicators for region of residence as of 2010. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

1.B Appendix: Additional Institutional Details

This section provides additional institutional details. The particular rules and regulations discussed pertain to our analysis time period and the birth cohorts relevant for our study.

Additional Information on Retirement Savings Accounts

Traditional defined contribution retirement savings plans in Denmark can be either employer-sponsored plans or personal plans. Within each type of plan, there are also three main types of accounts, which differ in the way that they are paid out. Life annuity accounts pay out as annuities for the rest of the account holder's life. Fixed-term annuity accounts pay out as income streams for a designated time period, typically either ten or twenty-five years. Capital accounts pay out as lump sum distributions.

Similar to the U.S. setting, the accounts are tax-advantaged. Contributions to the accounts are tax-deductible. Capital gains in the accounts are taxed upon accrual at approximately 15%, which is typically favorable compared to taxation of capital gains on savings outside of retirement accounts. Payments from life annuity and fixed-term annuity accounts are taxed as regular income, whereas distributions from capital accounts are taxed at approximately 40%.

In 2013, Denmark introduced "Roth-style" retirement plans. Contributions to these accounts are not tax-deductible, but lump sum distributions from the accounts are tax-free. These accounts aimed to replace the traditional capital accounts, as starting in 2013 contributions to capital accounts are no longer tax-deductible.

Additional Information on the Voluntary Early Retirement Pension

Participating in VERP requires making fixed contributions to qualified unemployment insurance (UI) funds during working life. These contributions amount to roughly $1,000 per year. To be eligible to claim, individuals must have contributed in 25 out of the previous 30 years.

VERP benefits are linked to the UI benefit schedule, but are typically viewed as flat-rate in practice, since they are capped at 91% of the maximum UI benefits. Typically benefit amounts are calculated using the highest twelve months of earnings over the previous two years. Monthly benefits correspond to 90% of these earnings divided by 12. Base benefits are then the minimum of either this amount or 91% of the maximum UI benefits. The maximum VERP benefits amount to roughly $27,000 per year, in 2010 USD.

Benefits are then subject to means testing, first against assets held in private retirement accounts, which determines base payments for the duration of the program. The government collects information on account balances from banking and financial institutions, usually when workers contributing to VERP are around age $59\frac{1}{2}$. This information is used to compute base benefits depending on claiming age. Benefits are reduced against assets in retirement accounts at approximately 60% of "could-be annuitized" payments.

In addition to this means testing, benefit payouts are further means tested against income after claiming. Benefits are means tested against drawdown from private retirement accounts, at a rate of around 50%. Benefits are also means tested against hours worked at a rate of 100%. VERP benefits are linked to an hourly rate per month, and each hour of work while on the program reduces VERP benefits by one hour.

Two key rules serve as defining features of the VERP program. The "transition rule" stipulates conditions under which individuals can transition to the VERP program. The regulation states that, to be eligible to claim VERP benefits, one must be "available to the labor force." Individuals can transition to VERP either from employment or from formal unemployment, which involves meeting UI requirements such as searching for jobs. An important implication of this rule is that an individual who retires and exits the labor force before reaching VERP eligibility age will not satisfy the transition rule and will not be eligible for benefits.

The "two-year rule" provides incentives for individuals to retire and transition to the VERP program two years after the earliest eligibility age. To satisfy the rule, individuals must work through the first two years of the VERP program. It is not enough to simply delay claiming of benefits. Satisfying the rule leads to three financial bonuses. First, base benefits for the duration of the VERP program are no longer means-tested against wealth held in private retirement accounts. Second, benefit amounts are weakly increased, as benefits become tied to 100% of the maximum UI benefits, rather than 91%. Third, every additional quarter worked after satisfying the two-year rule results in a tax-free lump sum payment equal to approximately $2,250.

Additional Information on the Old Age Pension

The OAP provides near-universal old-age benefits for Danes. Benefits are proportionally reduced for individuals that have lived in Denmark fewer than forty years. Benefit amounts are comprised of three main components. First, a base benefit of approximately $10,000 per year is provided to all individuals. This amount is subject to an earnings test where benefits are reduced at a rate of 30% against earnings above roughly $40,000. Second, a pension allowance is provided. The allowance is approximately $10,000 per year for single individuals and $5,000 for married individuals. This amount is subject to an income test where benefits are reduced at a rate of roughly 30% against earnings above $9,500. Third, there is a pension supplement available for the poorest pensioners. This amounts to about $1,000 per year but is delivered to only those with low levels of assets. In general, due to a 2004 reform, OAP benefits can be deferred with adjustments that are approximately actuarially fair.

Bibliography

Aguila, E. (2011). Personal retirement accounts and saving. *American Economic Journal: Economic Policy 3*(4), 1–24.

Alessie, R., V. Angelini, and P. van Santen (2013). Pension wealth and household savings in Europe: Evidence from sharelife. *European Economic Review 63*, 308–328.

Andersen, H. Y. (2018). Do tax incentives for saving in pension accounts cause debt accumulation? Evidence from Danish register data. *European Economic Review 106*, 35–53.

Asch, B., S. J. Haider, and J. Zissimopoulos (2005). Financial incentives and retirement: Evidence from federal civil service workers. *Journal of Public Economics 89*(2-3), 427–440.

Attanasio, O. P. and A. Brugiavini (2003). Social security and households' saving. *Quarterly Journal of Economics 118*(3), 1075–1119.

Attanasio, O. P. and S. Rohwedder (2003). Pension wealth and household saving: Evidence from pension reforms in the United Kingdom. *American Economic Review 93*(5), 1499–1521.

Behaghel, L. and D. M. Blau (2012). Framing social security reform: Behavioral responses to changes in the full retirement age. *American Economic Journal: Economic Policy 4*(4), 41–67.

Bernheim, B. D. (1987). The economic effects of social security: Toward a reconciliation of theory and measurement. *Journal of Public Economics 33*(3), 273–304.

Bernheim, B. D. (2002). Taxation and saving. In *Handbook of Public Economics*, Volume 3, pp. 1173–1249. Elsevier.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2009). The importance of default options for retirement saving outcomes: Evidence from the United States. In *Social Security Policy in a Changing Environment*, pp. 167–195. University of Chicago Press.

Blundell, R., E. French, and G. Tetlow (2016). Retirement incentives and labor supply. In *Handbook of the Economics of Population Aging*, Volume 1, pp. 457–566. Elsevier.

Bottazzi, R., T. Jappelli, and M. Padula (2006). Retirement expectations, pension reforms, and their impact on private wealth accumulation. *Journal of Public Economics 90*(12), 2187–2212.

Brown, K. M. (2013). The link between pensions and retirement timing: Lessons from California teachers. *Journal of Public Economics 98*, 1–14.

Burtless, G. and R. A. Moffitt (1985). The joint choice of retirement age and postretirement hours of work. *Journal of Labor Economics 3*(2), 209–236.

Cattaneo, M. and J. C. Escanciano (Eds.) (2017). *Regression Discontinuity Designs: Theory and Applications*, Volume 38. Emerald Publishing Ltd.

Cattaneo, M. D., M. Jansson, and X. Ma (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 1–7.

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

Choi, J. J. (2015). Contributions to defined contribution pension plans. *Annual Review of Financial Economics 7*, 161–178.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2002). Defined contribution pensions: Plan rules, participant choices, and the path of least resistance. *Tax policy and the Economy 16*, 67–113.

Coile, C. and J. Gruber (2007). Future social security entitlements and the retirement decision. *Review of Economics and Statistics 89*(2), 234–246.

Deshpande, M., I. Fadlon, and C. Gray (2020). How sticky is retirement behavior in the U.S.? Responses to changes in the full retirement age. *NBER Working Paper No. w27190*.

Diamond, P. A. and J. A. Hausman (1984). Individual retirement and savings behavior. *Journal of Public Economics 23*(1-2), 81–114.

Disney, R. (2006). Household saving rates and the design of public pension programmes: Cross-country evidence. *National Institute Economic Review 198*(1), 61–74.

Engen, E. M., W. G. Gale, and J. K. Scholz (1996). The illusory effects of saving incentives on saving. *Journal of Economic Perspectives 10*(4), 113–138.

Etgeton, S., B. Fischer, H. Ye, et al. (2021). The effect of increasing retirement age on households' savings and consumption expenditures. *Working Paper*.

Fadlon, I., J. Laird, and T. H. Nielsen (2016). Do employer pension contributions reflect employee preferences? Evidence from a retirement savings reform in Denmark. *American Economic Journal: Applied Economics 8*(3), 196–216.

Feldstein, M. (1974). Social security, induced retirement, and aggregate capital accumulation. *Journal of Political Economy 82*(5), 905–926.

Feldstein, M. and A. Pellechio (1979). Social security and household accumulation: New microeconometric evidence. *Review of Economics and Statistics 61*(3).

Feng, J., L. He, and H. Sato (2011). Public pension and household saving: Evidence from urban China. *Journal of Comparative Economics 39*(4), 470–485.

Fitzpatrick, M. D. and T. J. Moore (2018). The mortality effects of retirement: Evidence from social security eligibility at age 62. *Journal of Public Economics 157*, 121–137.

García-Miralles, E. and J. M. Leganza (2021). Joint retirement of couples: Evidence from discontinuities in Denmark. *CEBI Working Paper No. 06/21*.

Gelber, A. M. (2011). How do 401(k)s affect saving? Evidence from changes in 401(k) eligibility. *American Economic Journal: Economic Policy 3*(4), 103–22.

Gelber, A. M., A. Isen, and J. Song (2016). The effect of pension income on elderly earnings: Evidence from social security and full population data.

Geyer, J., P. Haan, A. Hammerschmid, and M. Peters (2020). Labor market and distributional effects of an increase in the retirement age. *Labour Economics*, 101817.

Geyer, J. and C. Welteke (2019). Closing routes to retirement for women: How do they respond? *Journal of Human Resources*.

Haller, A. (2019). Welfare effects of pension reforms.

Hernaes, E., S. Markussen, J. Piggott, and O. L. Vestad (2013). Does retirement age impact mortality? *Journal of Health Economics 32*(3), 586–598.

Hubbard, R. G. (1986). Pension wealth and individual saving: Some new evidence. *Journal of Money, Credit and Banking 18*(2), 167–178.

Hurd, M., P.-C. Michaud, and S. Rohwedder (2012). The displacement effect of public pensions on the accumulation of financial assets. *Fiscal Studies 33*(1), 107–128.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Kaplan, G. and G. L. Violante (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica 82*(4), 1199–1239.

Kaplan, G., G. L. Violante, and J. Weidner (2014). The wealthy hand-to-mouth. *Brookings Papers on Economic Activity* (1), 77–153.

Kapteyn, A. and C. Panis (2005). Institutions and saving for retirement: comparing the United States, Italy, and the Netherlands. In *Analyses in the Economics of Aging*, pp. 281–316. University of Chicago Press.

King, M. and L. L. Dicks-Mireaux (1982). Asset holdings and the life-cycle. *The Economic Journal*, 247–267.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*, 435–464.

Kotlikoff, L. J. (1979). Testing the theory of social security and life cycle accumulation. *American Economic Review 69*(3), 396–410.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016). Tax reforms and intertemporal shifting of wage income: Evidence from Danish monthly payroll records. *American Economic Journal: Economic Policy 8*(3), 233–57.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2017). Pension saving responses to anticipated tax changes: Evidence from monthly pension contribution records. *Economics Letters 150*, 104–107.

Krueger, A. B. and B. D. Meyer (2002). Labor supply effects of social insurance. In *Handbook of Public Economics*, Volume 4, pp. 2327–2392. Elsevier.

Kuhn, A., J.-P. Wuellrich, and J. Zweimüller (2010). Fatal attraction? Access to early retirement and mortality.

Lachowska, M. and M. Myck (2018). The effect of public pension wealth on saving and expenditure. *American Economic Journal: Economic Policy 10*(3), 284–308.

Laitner, J. and D. Silverman (2007). Life-cycle models: Lifetime earnings and the timing of retirement. *Michigan Retirement Research Center Working Paper No. 165*.

Lalive, R., A. Magesan, and S. Staubli (2017). Raising the full retirement age: Defaults vs incentives.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Liebman, J. B., E. F. Luttmer, and D. G. Seif (2009). Labor supply responses to marginal social security benefits: Evidence from discontinuities. *Journal of Public Economics 93*(11-12), 1208–1223.

Lindeboom, M. and R. Montizaan (2020). Disentangling retirement and savings responses. *Journal of Public Economics 192*.

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

Manoli, D. S. and A. Weber (2016). The effects of the early retirement age on retirement decisions. *NBER Working Paper No. w22561*.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics 93*(11-12), 1224–1233.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Nakazawa, N. (2021). The effects of increasing the eligibility age for public pension on individual labor supply: Evidence from Japan.

Nielsen, N. F. (2019). Sick of retirement? *Journal of Health Economics 65*, 133–152.

OECD (2015). Ageing and employment policies: Denmark 2015: Working better with age. *OECD Publishing, Paris.*

Poterba, J. M., S. F. Venti, and D. A. Wise (1996). How retirement saving programs increase saving. *Journal of Economic Perspectives 10*(4), 91–112.

Pozo, S. and S. A. Woodbury (1986). Pensions, social security, and asset accumulation. *Eastern Economic Journal 12*(3), 273–281.

Slavov, S., D. Gorry, A. Gorry, and F. N. Caliendo (2019). Social security and saving: An update. *Public Finance Review 47*(2), 312–348.

Snyder, S. E. and W. N. Evans (2006). The effect of income on mortality: Evidence from the social security notch. *Review of Economics and Statistics 88*(3), 482–495.

Staubli, S. and J. Zweimüller (2013). Does raising the early retirement age increase employment of older workers? *Journal of Public Economics 108*, 17–32.

Chapter 2

Joint Retirement of Couples: Evidence from Discontinuities in Denmark

Abstract

We study joint retirement behavior and document underlying mechanisms. Exploiting administrative data and the discontinuous increase in retirement when individuals reach pension eligibility age, we estimate sizable spillover effects to their spouses. We show that age differences within couples are crucial determinants of joint retirement, which is primarily driven by older spouses working longer. Controlling for these age differences reveals that female spouses respond more, even controlling for relative earnings. Relative earnings play a role consistent with collective models of household behavior. A complementary analysis shows that a reform increasing eligibility ages induces similar spillovers, suggesting no significant adjustment costs.

2.1 Introduction

In recent decades, aging populations have led to widespread pension reform. These reforms, and pension systems in general, are often designed at the individual level, however, the presence of significant spillovers within couples will have implications for projections of labor supply, budgetary estimations, and welfare analyses. Therefore, understanding the retirement behavior of couples is crucial for the design and evaluation of social security policy. In line with this reasoning, recent work on household finances is shifting attention towards interactions within couples, particularly in models of labor supply and retirement decisions (Gustman and Steinmeier, 2000, 2004; An et al., 2004; Bingley and Lanot, 2007; Van der Klaauw and Wolpin, 2008; Casanova, 2010; Michaud and Vermeulen, 2011; Honoré and de Paula, 2018; Honoré et al., 2020). These structural models illustrate two opposing forces determining joint retirement: household budget constraints (i.e. income effects) and household preferences (i.e. leisure complementarities), often finding a dominant role for leisure complementarities within the household.

However, there is limited work providing convincing causal evidence of joint retirement to guide policy and model design, particularly in regards to the mechanisms that underlie these behaviors. Providing causal estimates of joint retirement is challenged by the existence of unobserved covariates, such as preferences for leisure or types of jobs, and confounded factors, such as age, health, income shocks or shared assets. The empirical task is further hampered by the lack of suitable data and the complex design of public pension systems that sometimes affect spouses jointly, making the identification exercise infeasible or complicating the interpretation of the estimates. For example, the U.S. context faces some of these challenges, since pension benefits are linked between spouses, as is taxation. This might explain the lack of reduced-form evidence on joint retirement decisions from this country.[1]

[1] Hurd (1990) and Blau (1998) provide early evidence on the associations between spouses' retirement age in the U.S.

In this paper, we estimate the causal effects of pension eligibility ages on the retirement behavior of couples and provide evidence on the mechanisms that explain these behaviors. In our main analysis we exploit over two decades of administrative data from Denmark and the discontinuous increase in retirement that occurs when individuals reach their pension eligibility age to identify the effects on their spouses, controlling flexibly for the effect of spousal age. We study the period 1991–2013, where the early pension eligibility age remained constant at age 60, and was therefore known by couples well in advance. We show that one year after reaching their own early pension eligibility age, individuals are 20 percentage points more likely to be retired. We then find a sizable spillover effect on spouses, as we document a sharp 1.5 percentage point increase in the likelihood of spouses to be retired when their partners reach pension eligibility age. This amounts to a scaled spillover effect of 7.5%.

Next, we explore mechanisms that underlie joint retirement behavior and find four relevant dimensions. First, age differences between spouses are a crucial determinant of joint retirement. Joint retirement is primarily driven by older spouses who work past their own pension eligibility age, while waiting for their younger spouse to become eligible as well. Therefore, joint retirement behavior has a positive effect on aggregate labor supply. Second, we document a strong gender difference; female spouses are more likely to adjust their retirement to make it coincide with the pension eligibility age of their male partners. Importantly, this result is only revealed when we control for the age composition of the couple, since older partners are disproportionally males, which confounds the results from a simple comparison of male and female spouses. This gender difference prevails even after controlling for relative earnings within the couple, suggesting that gender norms may be playing a role. Third, a closer analysis of heterogeneous responses by relative earnings shows joint retirement patterns consistent with a collective model of household decisions, where couples in which the primary earner values joint leisure more are more likely to retire jointly. We also find patterns consistent with couples considering the opportunity cost of retirement, as we observe that younger spouses who are secondary earners are more likely to retire jointly by retiring earlier, while older spouses who are primary earners are more likely to retire jointly by retiring later. Fourth, we study joint retirement in the context of a reform that increased pension eligibility ages to investigate how couples adjust to a policy change. In a complementary analysis using a local difference-in-differences design, we find a 9% spillover effect to spouses, which is similar to our estimate from the previous, stable period. This suggests that spouses do not face any significant adjustment costs in response to the reform.

Our paper is primarily related to a small number of recent studies that explore the effect of pension eligibility ages on joint retirement. Of these, two stand out as closest to our paper. Lalive and Parrotta (2017) exploit 10 years of survey data from a Swiss census and the sharp change in retirement induced by gender-specific pension eligibility ages, finding evidence of significant spillover effects on female spouses and inconclusive results for males. Willén et al. (2020) exploit administrative data and a Norwegian reform that lowered pension eligibility ages for workers in specific firms to study spillovers across spouses and across programs; they restrict their analysis of spillovers to younger spouses and find an effect on female spouses only. Three other papers study reforms to pension eligibility ages. Selin (2017) and Bloemen et al. (2019) study reforms that affected public sector workers in Sweden and the Netherlands respectively, and Atalay et al. (2019) studies an increase in female pension eligibility ages using Australian survey data. Finally, Banks et al. (2010) and Hospido and Zamarro (2014) exploit cross-country differences in statutory retirement ages and find spillover effects to British men and to European women respectively.[2,3]

[2]Other studies on joint retirement have considered reforms that indirectly affect retirement through changes in the pension design. Baker (2002) investigates a Canadian spouse allowance that is means-tested jointly with the partner's wage giving them shared financial incentives and finds evidence of joint retirement. Coile (2004) explores the financial incentives to retire of each spouse and its interrelation, using the Health and Retirement Study. Stancanelli (2017) studies a reform that increases the contribution period needed to claim full pension benefits in France, finding very small effects for joint retirement. Kruse (2020) studies the removal of the earnings test on early pension benefits of private sector workers in Norway and finds significant spillovers to spouses working in the public sector.

[3]We also relate to the large literature that studies the impact of pension eligibility ages on own retirement:

The main contribution of our paper is to provide novel evidence on the mechanisms that explain joint retirement, which have implications for policy and model design. We show that age differences between spouses are crucial determinants of joint retirement behavior. We document gender differences that are not confounded by these age differences, whereas the previous literature is limited to simple gender splits and reports mixed results. In addition, our long panel data allows us to study the effect of relative earnings based on predetermined earnings shares. Lastly, we are able to complement the analysis with an evaluation of a pension reform that illustrates the lack of adjustment costs and has direct implications for policy.

The second contribution of our paper is to provide clear quasi-experimental evidence from administrative data for a representative population and a representative pension system. Our analysis includes male and female spouses as well as spouses that are relatively younger or older. Furthermore, as in most modern pension systems, the pension eligibility age of males and females is the same, and taxation and pension benefits are independent between spouses.[4] Finally, we study a major reform that is being adopted in many other countries and that affects a majority of the population, as opposed to a particular subgroup.

The paper is structured as follows. Section 2.2 describes the institutional background. Section 2.3 presents the data and the samples of analysis. Section 2.4 lays out our empirical strategy for estimating the effect of reaching a stable pension eligibility age and reports the results. Section 2.5 analyzes the reform that increased pension eligibility ages. Section 2.6 concludes.

E.g. Mastrobuoni (2009), Behaghel and Blau (2012), Staubli and Zweimüller (2013), Cribb et al. (2016), Manoli and Weber (2016), Geyer and Welteke (2019), Haller (2019), Nakazawa (2019), and Deshpande et al. (2020).

[4]In the past, many pension systems had different pension eligibility ages for males and females, but currently most developed countries have the same pension eligibility age for both genders or are in a process of convergence (OECD, 2015).

2.2 Institutional Background

The Danish retirement system is broadly typical of other developed countries (OECD, 2019). The two primary sources of retirement income are benefit payments from public pensions and savings in private retirement accounts, with the latter coming from personal or employer contributions during working life.

Pension benefits come from two main sources. The Old Age Pension (OAP) provides universal retirement income security at old ages, and the Voluntary Early Retirement Pension (VERP) provides early retirement benefits for those who choose to participate in the program. The majority of workers participate, about 80% of the birth cohorts we study. As VERP plays a major role in determining labor supply and retirement patterns of the Danish population, we focus our analysis on the VERP eligibility age.

2.2.1 Voluntary Early Retirement Pension

The VERP program, introduced in 1979, provides access to early retirement benefits, traditionally from age 60. Participating in VERP requires making modest contributions to qualified unemployment insurance funds during working life. Benefits are flat-rate and result in a fixed amount paid to all workers equal to roughly $27,000 annually (in 2010 USD).

The decision to claim VERP benefits is tightly linked to retirement, although they are technically separate decisions. The reason for this tight link is that the design of VERP produces strong incentives to retire at the same time as claiming. First, individuals must be "available to the labor market" in order to transition to VERP, that is they must be employed or actively searching for jobs or on a special transition pension (*delpension*). Hence, if individuals choose to leave the labor market before reaching VERP eligibility age, they will potentially forgo 5 years of benefits. Second, there are no actuarial adjustments for deferring claiming, so delaying claiming by one year amounts to a foregone year of benefits. Third, benefits are also subject to substantial means testing against labor market earnings at essentially 100%, which creates strong disincentives to keep working after VERP benefits are claimed, and against private retirement accounts.

The VERP program has remained fairly stable over time. Importantly, during the period 1991–2013, which we use in our first analysis, the VERP eligibility age remained constant at age 60. Two changes occurred during this period that are worth mentioning. First, the number of years that an individual has to contribute to an unemployment fund to qualify for VERP increased over time.[5] Second, a pension reform in 1999 introduced incentives for individuals to delay claiming of VERP benefits by two years, to age 62. By postponing claiming to age 62 the flat-rate benefits are slightly increased (from approximately $27,000 to $29,600) and they are no longer means-tested against private pension accounts. The effect of the reform was a mild decrease in the number of people claiming at age 60, and a new discontinuous increase at age 62. Across our different analyses we show that this reform does not meaningfully affect our results.[6]

In 2011 the Danish government announced a pension reform increasing pension eligibility ages in 6 month steps contingent on birthdate. Both the VERP and OAP ages increased, as well as the incentivized VERP age, while all other characteristics of the program remained unchanged. In Section 2.5 we describe this reform in detail, and we exploit the first discontinuity created by the reform to study the effect on joint retirement. We focus on the first cohort affected, those born after the cutoff date of January 1, 1954, whose VERP eligibility age was raised from 60 to $60\frac{1}{2}$, and who are first impacted in 2014 when they turn 60.

Two features of the VERP program make it ideal to study joint retirement behavior. First, the pension benefits are independent between spouses. The decision to claim or retire does not have any direct effect on the pension benefits of the spouse. Therefore, we can rule

[5]From 1985, individuals had to contribute for 15 years out of the last 20 years. In 1990 the number of years increased to 20 out of the last 25, and in 1995 it increased to 25 out of the last 30.

[6]While not a reform of VERP, between 1992 and 1996 a transitional benefits program allowed long-term unemployed above age 55 (and above age 50 from 1994) to retire with similar conditions as the VERP program.

out direct effects on the pension benefits of spouses as a mechanism for joint retirement in our analyses.[7] Second, the pension eligibility age is the same for men and women over the entire period considered, which has two advantages. First, our setting is representative of modern systems in most OECD countries that have eliminated the gender gap in statutory pension eligibility ages over the last decades (OECD, 2015, 2017). Second, we can study heterogeneous effects by gender, age composition and income shares within the couple that are not affected by differential pension eligibility ages.[8]

2.2.2 Old Age Pension

The OAP provides universal old-age benefits. The eligibility age was traditionally 67, and it was lowered to 65 by the 1999 reform. Therefore, less than 5% of the spouses in our samples of analysis are old enough to be eligible for OAP. Benefits are roughly $15,000 for married or cohabiting individuals and $20,000 for single individuals. Individuals are eligible for full OAP benefits if they have resided in Denmark for at least 40 years, and benefits are reduced proportionally if individuals have resided for a shorter period. Claiming benefits is an active choice, and the decision to claim is separate from the decision to cease working. From 2004, individuals can defer claiming OAP benefits and receive (approximately) actuarially-fair increases in benefits. Also, the means testing of OAP is less strict than that of VERP.

[7]This is in contrast to Baker (2002) who studies exactly these direct links between spouses' pension benefits, and also to the second empirical design of Atalay et al. (2019) which is based on the characteristics of Vietnam veterans' pension system.

[8]Note that this is in contrast to the two closest related papers to ours. Atalay et al. (2019) exploit a reform that raises women's pension eligibility ages to converge to that of men's. Lalive and Parrotta (2017) study a stable period where retirement ages were different beween men and women.

2.3 Data and Sample of Analysis

2.3.1 Data

We use administrative data covering the entire population of Denmark over the period 1986–2014. Using personal identifiers for each individual, we combine different registers with information on labor market outcomes, pension benefits, socio-demographics and family linkages. Variables are third-party reported on an annual basis and contain a large degree of disaggregation. Individuals cannot select themselves out of the registers, and they only exit the registers if they migrate out of the country or die.

In addition, we also use monthly-frequency register data on earnings for all employees in Denmark and on pension benefits for the entire population, both of them available from 2008. We combine this data with the annual-frequency registers using the same individual identifiers. This allows us to define retirement ages with more precision, which is crucial for the analysis of the 2014 reform that increased the pension eligibility age by 6 months.[9]

2.3.2 Key Variables.

One advantage of our data is that we can measure different margins of labor supply and retirement behavior. We consider three main outcomes, which are defined either at the end of each calendar year (when using the annual data in the first, age-based setting) or as half-year measures (when using the monthly data in the second, reform-based setting, since the reform increased the VERP eligibility age by 6 months).

Retirement: We define retirement as ceasing to earn labor market income. For the age-based design we use the annual data to define retirement as the year in which individuals earned income for the last time.[10] Therefore, we define retirement as an absorbing state where

[9]This new dataset, often referred to as *eIncome*, is described in more detail in Kreiner et al. (2016).

[10]We allow for some small positive income, equivalent to 1 month of average earnings, to accommodate the fact that individuals can receive some labor income after they have retired, such as holidays payments or delayed wages.

the retirement variable takes the value one thereafter. In the robustness section we show that the results are robust to using a flow definition of retirement where we allow individuals to retire multiple times. These definitions are standard in the retirement literature (Coile and Gruber, 2007; Deshpande et al., 2020). For the reform-based design, we use the monthly data to define a dummy that takes the value one if an individual works past the first half of the year (that is, past July 1) in a given year. This accommodates the fact that individuals unaffected by the reform become eligible for benefits at the beginning of the reform year (2014) when they turn 60, whereas individuals affected by the reform become eligible at least 6 months later, when they reach age $60\frac{1}{2}$.

Claiming: We define claiming as receiving pension income, either VERP or OAP. For the age-based design we define an indicator equal to one if an individual receives any pension income in a given year. For the reform-based design we define an indicator that takes the value one if an individual received pension income before July 1 in a given year.

Earnings: In both research designs we use taxable annual labor market earnings from the annual registers. We winsorize this variable at the 1^{st} and 99^{th} percentile to reduce the influence of outliers. We adjust this variable for inflation using 2010 as a baseline and convert Danish kroner to U.S. Dollars using the exchange rate 1 USD = 5.56 DKK.

2.3.3 Samples of Analysis

We define two samples of analysis, one for each research design. For both of our research designs we start with the full population of Danish couples who reside in Denmark between 1991 and 2014. We define couples as those who are either married, or in a registered partnership, or cohabiting. To avoid endogenous changes in marital status around the time of pension eligibility we identify couples when they are both below age 60 and observe them for as long as they remain together. We restrict the analysis to couples who are up to 8 years apart from each other, which excludes around 5% of the sample on each side of the distribution. We illustrate the distribution of age differences within couples in panel (a) of Appendix Figure 2.A.4, and we show that our results are robust to dropping this restriction in Section 2.4.6.

We focus the analysis on dual-earner couples. First, we restrict the sample to couples where the reference individual (that is, the focal partner who reaches their own pension eligibility age) has earned labor income at least once between ages 55 and 59. All cohorts in our sample of analysis are observed back to age 55 since we have data from 1986. We also exclude reference individuals who are self-employed or on disability benefits at least once between ages 50 and 59, as they are subject to different rules and regulations of the VERP scheme. Second, we restrict the sample to couples where the spouse has earned labor income at least once between ages 50 and 59. We use this longer period for spouses to ensure that our sample does not exclude younger spouses who retire in their early 50s, as they can potentially retire jointly with their older partners.[11]

Age-based sample. For our age-based design, we consider the period 1991–2013, where the early pension eligibility age remained stable at age 60. This provides us with more than two decades of observations from individuals who faced the same pension eligibility age. We focus the analysis on couples where the reference individual is 57 to 60 years old, which leads to a sample size of 367,585 couples and 2,206,044 couple-year observations.

Reform-based sample. For our reform-based design, we consider the period 2008–2014, starting in 2008 because the monthly-frequency data is only recorded from that year. To focus on individuals who are more likely to be impacted by the reform, we restrict this sample to reference individuals who have made qualifying contributions to the VERP program at least once between ages 50 and 59. Note that we cannot impose this restriction on the full age-based sample because we do not observe contributions far back in time, but in the

[11] Note that there are four cohorts of spouses that we cannot observe before age 60 to impose the restriction, and therefore we keep all those spouses, who represent 0.4% of the sample. Similarly, there are nine cohorts of spouses that we cannot observe during the entire period between ages 50 to 59. In this case, we impose the restriction based on the years that we observe. This affects 12% of the spouses, of which 80% are observed for 5 or more years.

robustness section we show that our results from both designs are robust to this decision.[12] In our baseline specification, we focus on individuals born within a 3-month window on either side of the January 1, 1954 cutoff, and we balance the sample, leading to a sample size of 10,321 couples and 73,395 couple-year observations.

Table 2.1 presents summary statistics for the two samples and for the corresponding unrestricted population. The first four columns correspond to the age-based period of analysis (1991–2013) and the last four columns correspond to the reform-based period of analysis (2008–2014). First, we can compare the analysis samples to their corresponding population samples. We note that both reference individuals and spouses in the analysis samples have higher earnings, higher education, and are less likely to be retired before age 60. This is mainly a consequence of restricting the analysis to dual-earner couples and to those who did not receive disability benefits in the past. Also note that the age difference between spouses is similar between the analysis sample and the population, but the standard deviation is smaller due to the restriction that drops spouses who are more than 8 years apart. Second, we can compare the two analysis samples. Overall the two samples are similar, but the reform-based sample has a smaller share of males (47% against 52%), higher earnings ($64,156 against $60,289) and is slightly more likely to be retired before age 60 (16% against 14%), but these differences are not statistically significant. These differences are in line with the effect of restricting the reform-based sample to VERP contributors, as females are more likely to contribute to the program. The age difference between partners in both analysis samples is similar and so are the standard deviations.

[12] Specifically, we show that our age-based results are robust to imposing the restriction for the subsample of observations over 2008 to 2013, for whom we can observe past contributions. We also show that the reform-based results are robust to not imposing the restriction.

2.4 The Effect of Reaching Pension Eligibility Age

2.4.1 Age-Based Discontinuity Design

To identify the causal effects of individuals reaching pension eligibility age on their own retirement and on their spouses, we exploit the discontinuity that occurs around the early pension eligibility age. Specifically, we study the retirement patterns of reference individuals and their spouses around the eligibility age of the reference individuals, that is around age 60. Importantly, when analyzing spouses' retirement patterns we control flexibly for the effect of own age on their own retirement behavior.

We lead our analysis with a graphical illustration of the retirement patterns of the reference individuals and their spouses, which then guides our estimation strategy and allows us to evaluate the assumptions of the estimation model.

Note that each member of a couple can potentially appear both as the reference individual and as the spouse in the analysis, as long as they are observed at ages 57–60 during the period considered. This reflects the dual nature of the couples' decision, and our design allows us to study their retirement behavior from both sides, observing them as reference individuals when they reach their pension eligibility age and as spouses, with respect to their partners' eligibility age. In the heterogeneity analysis we will, nevertheless, split the sample by age composition and gender and each member of the couple will appear only as either the reference individual or the spouse.

2.4.2 The Effect of Reaching Pension Eligibility on Own Retirement

We begin by analyzing the retirement behavior of reference individuals around their own pension eligibility age. Specifically, in Figure 2.1 we pool individuals for the period 1991–2013 and plot raw means of each outcome variable for the reference individual against their own age. As expected, given the strong incentives to retire exactly at the pension eligibility age, we observe a clear discontinuous jump in all outcomes at age 60. An important feature of the data is that the outcome variables are measured at the end of each calendar year, and so is age, which we round up to months. Hence, individuals who turn 60 early in the year can claim their pension earlier that year than those who turn 60 later in the year. This induces a gradual phasing-in of the exposure to early retirement eligibility as monthly age increases from 60 to 61, a pattern captured by Figure 2.1.

We are interested in the "full-exposure" effect of being eligible for one entire calendar year. Individuals who are fully exposed are those who turn 60 at the beginning of January, becoming eligible for early retirement at that moment. These individuals are exposed to early pension eligibility for 12 months by the time their information is recorded in the administrative data in December. In contrast, individuals who turn 60 later in the year are eligible for a shorter period of time that year, so they are only partially exposed. Our estimation strategy exploits information from both partially and fully exposed individuals to estimate the full-exposure effect with greater precision.

We quantify the full-exposure effect by estimating the following piecewise linear regression, which is closely guided by the graphical analysis:

$$y_{it} = \alpha + \beta_1\, age_{it} + \beta_2\, \mathbb{1}\{age_{it} \geq 60\} + \beta_3\, \mathbb{1}\{age_{it} \geq 60\} \cdot age_{it} + \sum_{c=1991}^{2013} \kappa_c \cdot D_c + \epsilon_{it} \quad (2.1)$$

where y_{it} is the outcome of interest for reference individual i at time t, age_{it} is monthly age of the reference individual at the end of the calendar year, and $\mathbb{1}\{age_{it} \geq 60\}$ is an indicator variable that takes the value one if the monthly age of the reference individual is 60 or above and zero otherwise. The model therefore estimates a discontinuous jump at monthly age 60 and a differential trend thereafter, as suggested by the graphical analysis. D_c are calendar year dummies. We estimate this regression for individuals between monthly ages 57 and just below 61.[13]

The full-exposure effect is then given by $\beta_2 + \frac{11}{12} \cdot \beta_3$. This estimator captures the treatment effect of being eligible for early pension during one full calendar year. It is composed of a sharp change in levels at the eligibility-age cutoff, captured by β_2, and a change in trends, captured by the slope parameter β_3 that captures the effect of one year of eligibility from age 60. We plot the parametric fit of this model in Figure 2.1. The full-exposure effect corresponds to the vertical distance between the solid line and the dashed line just below age 61.[14]

The first row of Table 2.2 reports the full-exposure estimates for the different outcomes of the reference individual. The first column reports the full-exposure effect on retirement. The estimate is 0.2034, which means that reaching pension eligibility age increases the share of retired individuals by around 20 percentage points. Note that the share of retired individuals before they reach pension eligibility is also positive, around 16% before age 60, as illustrated in panel (a) of Figure 2.1. This shows that individuals can also retire before they reach pension eligibility.[15] The second outcome of interest, pension claiming, is reported in the second column. The point estimate is 0.35, so around 35% of individuals claim VERP benefits by the end of their first year of eligibility. The effect for claiming is larger than for retirement for two reasons. First, it is not possible to claim VERP benefits before age 60, as illustrated in panel (b) of Figure 2.1, and second, individuals who claim can still have posi-

[13]Because the outcome variables are measured in December, individuals who turn 60 in December often do not have time to receive pension income until the next year. This is clearly seen in Figure 2.1, panel (b), where the dot for December is much lower. To prevent this from biasing our estimates we exclude these individuals by adding a dummy variable that takes the value one if their monthly age is exactly 60. In Table 2.5 of the robustness section we show that the results are largely unaffected if these individuals are kept.

[14]A similar methodology is used by Fadlon et al. (2019) to study the effect of Social Security's survivors benefits on labor supply in the U.S. Also, Nielsen (2019) studies the effect of retirement on health exploiting the same age-discontinuity in Denmark.

[15]We have argued in Section 2.2 that there exists strong incentives to claim right at the early pension eligibility age, but individuals might cease to earn labor income earlier than 60 for a number of reasons: they might become unemployed or claim a partial pension until they turn 60, they might voluntarily stop working even if that implies the inability to claim VERP later on, and lastly, not all individuals in our sample qualify for VERP, as explained in Section 2.3, around 80% of the individuals in the age-based sample of analysis made contributions to qualify for VERP.

tive earnings in the same year. Finally, the third column reports the full-exposure effect on annual labor market earnings, which can potentially reflect responses both on the extensive margin and on the intensive margin. We estimate a decrease of \$8,642 in annual earnings after one year of exposure to pension eligibility.

Overall our results show that reaching pension eligibility leads to a strong first stage. Individuals are discontinuously more likely to retire after age 60. We now turn to estimate the causal effects of pension eligibility on spousal retirement behavior.

2.4.3 The Effect of Reaching Pension Eligibility on Spouses

For the spillover effect on spouses, we follow a similar empirical strategy as for reference individuals. The main difference is that we need to control for the effect of spouse's own age on their retirement behavior so that we can isolate the causal effect of their partner's pension eligibility.

We lead the analysis with a nonparametric illustration of spouse retirement patterns around their partners' age, cleaned from the effect of the spouses' own age. Specifically, we plot the residuals from the following regression:

$$y_{it}^s = \alpha + \sum_{a=49}^{69} \delta_a \cdot D_a^s + \sum_{a=49}^{69} \gamma_a \cdot D_a^s \cdot D_g + \sum_{c=1991}^{2013} \kappa_c \cdot D_c + \varepsilon_{it} \tag{2.2}$$

where y_{it}^s is the outcome variable of interest for spouse s of individual i at time t, D_a^s are dummy variables for spouses' monthly age, and D_g is a gender dummy. The residuals $\hat{\varepsilon}_{st}$ therefore capture the spouses' retirement behavior that is not explained by their own age and gender.[16]

The dots in Figure 2.2 plot spousal residuals $\hat{\varepsilon}_{it}$ binned over the monthly age of reference individuals. This illustrates the spouses' retirement patterns that are driven by

[16]An alternative approach to this methodology would be to estimate equation (2.2) adding age dummies for the reference individual and plot those coefficients. We show that the result is similar in Appendix Figure 2.A.1.

their partner's age. We observe that spousal residuals change discontinuously right when their partner becomes eligible for early pension at age 60, resembling the same pattern we observed for the reference individuals themselves.

Guided by this graphical analysis, we estimate a parametric model that quantifies the causal effect of one partner reaching pension eligibility age on the retirement behavior of their spouse. The estimating equation is similar to equation (2.1) for the reference individual, but with spouses' outcomes as the dependent variables and additional controls for spouses' age and gender that do not impose any functional form. The estimating equation is:

$$\begin{aligned} y^s_{it} = \alpha + \beta_1\, age_{it} + \beta_2\, \mathbb{1}\{age_{it} \geq 60\} + \beta_3\, \mathbb{1}\{age_{it} \geq 60\} \cdot age_{it} + \\ \sum_{a=49}^{69} \delta_a \cdot D^s_a + \sum_{a=49}^{69} \gamma_a \cdot D^s_a \cdot D_g + \sum_{c=1991}^{2013} \kappa_c \cdot D_c + \epsilon_{it} \end{aligned} \tag{2.3}$$

where y^s_{it} is the outcome of interest for spouse s of individual i, age_{it} is age of the reference individual in months, and $\mathbb{1}\{age \geq 60\}$ is an indicator variable that takes the value one if the reference individual is 60 or older (in terms of monthly age) and zero otherwise. D^s_a are dummy variables for spouses' monthly age, and D_g is a gender dummy. We estimate this regression for the same sample of reference individuals, between ages 57 to 61, as before.

The full-exposure effect is again given by $\beta_2 + \frac{11}{12} \cdot \beta_3$. For illustrative purposes, Figure 2.2 superimposes the parametric fit of the model estimated in equation (2.3) over the residuals from equation (2.2). The full-exposure effect corresponds to the vertical distance between the solid and dashed lines just below age 61. The second row of Table 2.2 reports the full-exposure effect on spouses from their partner reaching pension eligibility age. The effects on all three spousal outcomes are statistically significant at the 1% level. These point estimates can be viewed as the reduced-form effects on spouses.

To judge the size of joint retirement behaviors, we report "scaled effects" in the last row of Table 2.2, defined as the full-exposure effect on the spouse divided by the full-exposure effect on the reference individual. These scaled effects are our preferred measure for reporting and interpreting joint retirement spillovers, as they are comparable across different outcomes, samples of analysis, and empirical strategies, including our reform-based design. We compute standard errors for these scaled estimates by bootstrapping (Andrews and Buchinsky, 2000; MacKinnon, 2006).[17]

The scaled effect on the retirement outcome is 7.5%. That is, for every 100 individuals who retire right when they reach their early pension eligibility age, about 8 of their spouses are induced to retire as well. This is after controlling for the effect of the spouses' age on their own retirement behavior.

Claiming leads to scaled effect of 3.4%. This effect is smaller than the one for retirement for two reasons. First, the denominator is larger, that is, the full-exposure effect on the reference individual is larger for claiming than for retirement as discussed earlier. Second, the numerator is slightly smaller, the full-exposure effect on the spouses is smaller because of spouses who retire but do not claim. Knowing the joint retirement effect on claiming is important for policy and fiscal estimations, but for the reasons mentioned above it does not fully capture joint retirement behavior.[18] In the next subsection we explore the interaction between claiming and the age composition of couples and its implications for heterogeneous joint retirement responses.

For earnings, the scaled effect is 9.8%. Note that this outcome potentially captures both extensive margin responses and intensive margin adjustments that can be in the form of hours worked, choice of job, or effort. However, we cannot conclude that there are significant intensive margin responses based on the larger size of the scaled effect for earnings compared to retirement. Note that the size of the scaled effect for earnings depends on the relative

[17] Note that these scaled effects are conceptually similar to the estimates from an instrumental variables approach. We use scaled effects because they allow for a more flexible estimation of the second stage (the spouses' full-exposure effect) by estimating the jump at 60 and the differential trend separately. An instrumental variables approach, instead, imposes the same functional form as the first stage (the instrumented outcome of the reference individuals).

[18] For an analysis of retirement and claiming in the U.S. see Deshpande et al. (2020). Note that while deferring claiming in the U.S. leads to actuarial adjustments of future pension benefits, in the VERP program there is no such actuarially fair updating, and therefore the decision to claim and retire are more closely related.

earnings within couples, and the scaled effect will increase if the spouses who adapt their behavior to retire jointly are mainly the primary earners, even if adjustments occur only through the extensive margin. This in turn depends on the response heterogeneity, which we analyze in the following section.

2.4.4 Explaining Joint Retirement: Heterogeneity and Mechanisms

The aggregate results from the previous section, reported in Table 2.2, mask important differences across different types of couples. In this section, we explore differences across three characteristics: age differences within the couple, gender, and primary earner status. We are in an exceptional position to do so, due to our large sample size and the symmetric design of the Danish pension system, where men and women face the same pension rules and pension benefits are independent between spouses.

Age differences within the couple

We study the effect that relative age within partners has on joint retirement and find that it plays a crucial role. We begin our analysis by splitting our sample based on whether spouses are older or younger than their partners who are reaching age 60. For each of these subsamples we replicate the analysis and report the results in columns (1) and (2) of Table 2.3. Focusing on the retirement outcome, we observe that the scaled effect is 10% for older spouses and only 2.9% for younger spouses, while still highly significant. These results suggest an important role for the ability to claim own pension benefits in the decision to retire jointly. Older spouses who retire right when their younger partners reach pension eligibility age must continue working past their own pension eligibility, and can then claim benefits themselves. In contrast, younger spouses who retire right when their older partners reach pension eligibility age cannot claim their own benefits, since they themselves have not yet become eligible. This is a potentially financially costly decision, especially if it entails giving up the right to claim VERP later on due to the "transition to VERP" rules explained in Section 2.2. Overall, our results show that in this context couples favor the joint retirement path where the older spouse works past their eligibility age rather than the younger spouse retiring before reaching pension eligibility age.

Next, we explore the effect of age differences in more detail. Specifically, we define subsamples based on smaller intervals of their age differences and estimate joint retirement spillover for these subsamples. The results are reported in Figure 2.3, where we plot the scaled effects as spouses' age increases relative to their partners. We observe that the largest scaled effects are concentrated among spouses who are older, but not too far apart from their partners' age. Specifically, focusing on the retirement outcome, reported in panel (a), we find the largest effect (above 10%) for spouses up to 2 years older, followed by spouses who are between 2 and 4 years older. The effect decreases for spouses who are more than 4 years older than their partners. For younger spouses, we do not find evidence of differential spillovers in joint retirement as the difference between partners' age increases. The point estimates remain small and stable around the same size as for the pooled subsample of younger spouses (2.9%), although less precisely estimated due to the smaller sample size.

Overall, these results point to age differences between partners and the ability to claim as crucial determinants of joint retirement. Policies that aim to account for the joint retirement of couples must account for the economic incentives faced by each age group, and particularly for the ability of each partner to claim. In our setting, younger spouses cannot claim benefits of their own if they retire when their partner reaches early pension eligibility. However, in other settings, such as those centered on later pension eligibility thresholds where younger spouses can also claim their own benefits when they retire at the same time as their older partners, the joint retirement spillover of younger spouses might be larger. In addition, economic analysis of intra-household behavior should account for the effect of the age-composition of couples.

The effect of gender

Next, we explore heterogeneity by gender, a dimension where previous studies have found particularly mixed results. Some of the difficulties faced by the literature include pension systems where eligibility ages differ by gender or where a reform affected one gender only, lack of statistical power that hampered the estimation of small effects, and failure to account for confounded effects between age differences and gender. Our analysis overcomes these challenges, as there are no gender differences in the Danish pension system, benefits and taxation are independent between spouses, and we have statistical power to estimate gender differences controlling for other confounding factors such as age differences.

We begin by replicating our analysis over a simple split by gender. Column (3) of Table 2.3 presents results for the subsample of male spouses and column (4) for female spouses. The scaled effects for both male and female spouses is 7.5%, which could erroneously lead us to conclude that both genders are equally likely to adapt their behavior to retire jointly with their partner.

However, this simple split by gender masks important differences in the composition of relative age between spouses among the two groups. As in most countries, Danish men tend to be the older member of the couple.[19] Specifically, in our analysis sample males are around two years older than females, as we illustrate in panel (b) of Appendix Figure 2.A.4. We have shown that older spouses are much more likely to retire jointly, therefore the estimate found for men confounds the fact that the subsample of male spouses is composed by a larger share of older spouses. Therefore, to explore gender differences in joint retirement, we must control for the confounded age differences. We address this by reweighting the subsample of female spouses to match the distribution of age differences from the subsample of male spouses, and then re-estimate the spillover effect. The result is shown in column (5), where we observe that the scaled effect for females rises from 7.5% to 13%. We can then compare this scaled

[19] Hospido and Zamarro (2014) and Coile (2004) consistently find similar age differences, of around two years, for different European countries and for the U.S. respectively.

effect to the scaled effect for male spouses, reassured that the difference is not driven by the age-difference composition of both subsamples. Interestingly, we find that females clearly respond more, contrary to the conclusion that we could have reached from the simple split by gender.[20] The reweighting strategy assumes that couples where females are the older spouse are comparable to couples where females are the younger spouse. We explore this in Appendix Table 2.A.2, and show that these two types of couples are remarkably similar along observable characteristics such as labor market earnings, educational attainment, retirement probability, or whether they live in the Copenhagen region, all measured before age 57. Specifically, female spouses are very similar to each other regardless of whether they are the younger or the older member in the couple, and so are males.

A potential explanation for these gender differences in behavior is that relative earnings within couples confound joint retirement and gender. We study the role of relative earnings in detail in the next section, but regarding its impact on the gender gap, we show that the gender gap found is robust to further reweighting the sample of female spouses to have the same distribution of earnings shares as male spouses. The results are reported in column (6) of Table 2.3, where the scaled effect estimate for retirement remains high at 13.6%.[21] Our results therefore unveil a gender gap that cannot be explained by age or relative earnings within couples, suggesting a role for gender norms. This result adds to recent findings of gender differences that cannot be explained by traditional economic incentives (Daly and Groes, 2017; Kleven et al., 2019; Gørtz et al., 2020; Lassen, 2020).

Our results also document a new source of gender differences in earnings and labor supply which, unlike previous studies that focus on childbearing and childcare, originates in

[20]Furthermore, we find this gender gap both for couples were the female partner is the younger member as well as for couples where the female partner is the older member. We show this in Appendix Table 2.A.1 where, as an alternative to the reweighting strategy, we split the sample in four, by gender and by relative age between partners.

[21]The gender gap also remains when we further reweight the subsample of female spouses to ensure that the share who made contributions to qualify for VERP in the past is the same as in the subsample of male spouses. Note that we only observe VERP contributions for the most recent period of time and hence we perform this test for the period 2008–2013 only. The scaled spillover for retirement is 7% for males and 11.4% for females after reweighing.

the dynamics of family formation combined with the joint retirement behavior of couples, manifesting itself at the end of working life. Because males tend to be older than their female partners, couples who retire together most often achieve this either by males retiring later or by females retiring earlier, therefore increasing males lifetime earnings relative to females. Note that the "grandchild penalty" found by Gørtz et al. (2020) could explain part of the gap we identify, as grandmothers retire earlier to take care of their grand children, but it does not explain it all, as we also find that older female spouses are more likely to retire later, waiting to retire together with their younger partners.

The effect of relative earnings within couples

We now study the role of relative earnings within couples for joint retirement. To define the relative earnings of each member of the couple we compute predetermined earnings shares based on the average labor market earnings of each partner between ages 55 and 57, and report the distribution of these shares in panel (d) of Appendix Figure 2.A.4. We define an indicator for who is the primary earner in the couple based on these shares, excluding couples with very similar earnings shares (those between 47.5% and 52.5%, which represent 14% of the sample), although the results are robust to keeping them.

The interaction between relative earnings and gender. A growing literature studies the decision-making process of households through the lens of a collective model (Chiappori, 1992; Browning and Chiappori, 1998; Donni and Chiappori, 2011), where members with more negotiation power have more weight in the decision-making process of the household (Browning et al., 1994). If males and females differ in their preference for joint leisure, we would expect that the member with more power, the primary earner, will have a bigger influence in the joint retirement decision. We explore this in Table 2.4, where we replicate our analysis to estimate spillover effects over four different subsamples, distinguishing by whether the spouse is the primary or secondary earner and by gender. To avoid composition effects confounding our results, we reweight each primary-earner subsample so that it matches the distribution of the secondary-earner subsample of the same gender in terms of age differences. We report results only for the retirement outcome.

We find that couples where males are the primary earner are more likely to retire jointly, consistent with the finding of Browning et al. (2020) that males value joint leisure more than women, and in further support of the collective model as an explanation of couples labor supply.[22] Specifically, we find that male spouses who are secondary earners, reported in column (1), are much less likely to adjust their behavior to retire jointly than male spouses who are primary earners, as reported in column (2). The scaled effect is 4.3% against 9.1%. Correspondingly, female spouses who are secondary earners are much more likely to adjust their behavior to retire together than female spouses who are primary earners, as we see from comparing column (3) to column (4), with scaled effects of 8.2% and 2.3% respectively. These results also suggest that, among couples where males are the primary earner, both men and women are equally likely to be the ones adjusting their behavior to retire jointly, either delaying or anticipating their retirement, as the scaled effects from columns (2) and (3) are very similar.

The interaction between relative earnings and age differences. We now explore whether the interaction between relative earnings and age differences within couples affect their preferred route to joint retirement. Specifically, one might expect that older members who are primary earners are more likely to extend their employment while younger members who are secondary earners are more likely to retire earlier, consistent with the opportunity cost of retirement as foregone labor market earnings. We study this by replicating our analysis to estimate spillover effects over four different subsamples, distinguishing by whether the spouse is the primary or secondary earner and whether the spouse is the younger or older member of the couple. To avoid composition effects confounding our results, we reweight the primary-earner subsamples so that they match the distribution of the secondary-earner

[22]Note that the finding that males value joint leisure more than women can also be interpreted as males disliking some forms of independent leisure more than women, such as staying at home while their partners go to work.

subsamples in terms of age differences and gender.

The results are reported in Table 2.4. Overall, the primary-earner status does not seem to be a major determinant of joint retirement, as the differences between primary and secondary earner spouses is small and not statistically significant. However, interpreting the estimates at face value, we observe patterns consistent with the opportunity cost of retirement. We see that among older spouses, shown in columns (1) and (2), primary earners are 1.1 percentage points more likely to retire jointly. That is, they are more likely to work past their retirement age waiting for their younger spouses to reach their own pension eligibility age. On the contrary, among younger spouses, shown in columns (3) and (4), secondary earners are 2.7 percentage points more likely to retire jointly, that is they are more likely to stop working before they reach their own pension eligibility age to retire when their older partner becomes eligible. These results are consistent with the opportunity cost of retirement seen as foregone earnings. The returns to continued employment are higher for primary earners, who therefore are more likely to work longer, while the foregone earnings from secondary earners are smaller, making it less costly to stop working earlier.

2.4.5 The Evolution of Joint Retirement Over Time

In our analyses we have pooled two decades of observations to obtain precise estimates of the causal effects of reaching pension eligibility age on joint retirement of spouses. In this section we provide evidence on the evolution of these estimates over time. To do so, we replicate the previous analysis over 5-year running windows. We report the evolution of the scaled effects for the three outcomes of interest in Figure 2.4, where each dot at year t corresponds to the scaled effect estimated for the period $t-4$ to t. For instance, the last dot from 2013 reports the scaled effects estimated for the period 2009–2013.

Overall, we observe that joint retirement has been stable over time, which allows us to interpret the scaled effect estimates for the full period as reflecting a stable spillover behavior, as opposed to the average of an estimate that has been changing over time. As such, the size of the full-period estimates is also representative of the effect in most recent years, which are of more interest for policy and also the relevant period for comparison with the reform-based estimates derived from the 2011 reform that we present in Section 2.5.

2.4.6 Threats to Identification and Robustness

Our identification strategy relies on the assumption that, once we control flexibly for the spouses' age on their own retirement, the discontinuous behavior that occurs when their partner reaches pension eligibility age is caused by that event, and nothing else. In this section we provide a number of tests to assess the validity of our design.

Placebo test. To be reassured that we successfully control for the effect of the spouses' age, we carry out a placebo test. We repeat the analysis for the same sample of reference individuals, but we randomly assign them a fake spouse of similar age. Specifically, we assign a spouse of the same age to half of the reference individuals, and we assign spouses who are between 1 and 3 years younger or older to the other half of the reference individuals.[23] In this sample, spouses are likely to retire at the same time because their ages are highly correlated and most of them reach pension eligibility age right around the same time. However, we should not observe any joint retirement behavior beyond the one due to this age correlation between spouses, given that fake spouses cannot influence each other. If our empirical strategy successfully controls for the effect of age correlations, then we should not find any evidence of joint retirement in this placebo sample. Reassuringly we do not find any, as reported in Appendix Table 2.A.3 and Appendix Figure 2.A.2.

Alternative specifications. In Table 2.5 we show that the results are robust to a series of changes in the model specification and in the sample definition. Row A reports the baseline estimates for comparison. In row B we extend the sample of analysis to include reference

[23]Note that we do not use only spouses of the same age to avoid collinearity between the age of both partners.

individuals of ages 55 and 56. In row C we exclude reference individuals aged 59 by adding a dummy variable to the model that takes one if the reference individual is 59 or older. This excludes monthly ages between 59 and 60 from the estimation of the counterfactual behavior. In row D we keep couples with partners that are more than 8 years apart from each other. Row E drops the dummy that identifies reference individuals who turn 60 in December, so that they are included in the estimation of the jump at 60 and the differential trend afterwards. Row F allows for a nonlinear counterfactual before age 60 by adding a second order polynomial of the reference individuals' age to the model. This nonlinear specification reduces our point estimates (e.g. the scaled effect on retirement becomes 4.1%), but note that we are fitting a second order polynomial over a short period of three years (ages 57 to 60). To account for this, in row G we increase the age range to include reference individuals of ages 55 and 56 (as in B) and fit a second order polynomial (as in F), obtaining spillover effects much closer to our baseline estimates. Row H controls for predetermined region and education of the reference individual and their spouse. Row I adds a dummy for individuals born after 1939, who are therefore affected by the 1999 reform that introduced incentives to claim VERP at age 62 and lowered the OAP to age 65. Row J estimates the effect over the period 2008–2013, which is almost the same period considered in the reform-based design that we present in Section 2.5. In row K we present estimates over the same period as in J, and restrict the sample to reference individuals who have made contributions to qualify for VERP at least once between ages 50 and 59. Note that we can only impose this restriction for these later calendar years as we do not observe contributions far back in the past. Finally, in row L we report the scaled effect for retirement defined as a flow variable, which allows individuals to retire multiple times (see Appendix Figure 2.A.3 for the full-exposure effects). Reassuringly, our results are robust to all these changes.

Attrition. Individuals cannot self-select out of the registers. The only two reasons for an individual to exit the registers are either migrating out of Denmark or dying. If reaching pension eligibility caused any of these two things to happen, we would miss that individual from the sample, but in no case would they be wrongly considered as retired. Note also that Nielsen (2019) finds no evidence of increased mortality at retirement studying the same age discontinuity in Denmark.

2.5 Impact of Increasing Retirement Ages

We have shown that spouses are more likely to retire right when their partners reach pension eligibility age. What happens to the joint retirement of couples when the pension eligibility age of one partner changes? In this section we study a major reform that discontinuously increased the early pension eligibility age of selected cohorts. This analysis complements the previous analyses by testing whether the joint retirement spillover that occurs in a stable setting carries over to a reform setting, or whether couples face adjustment costs that limit their capacity to retire together.

2.5.1 The 2011 Pension Reform

In May of 2011 the Danish government announced a pension reform that discontinuously increased retirement ages in six-month increments contingent on birthdate. The first increase introduced by the reform provides us with the clearest natural experiment: the early pension eligibility age (that is, the VERP eligibility age) was increased from 60 to $60\frac{1}{2}$ for those born from January 1, 1954, while it remained at 60 for those born right before. The reform also introduced six-month increments in the incentivized early retirement age that was traditionally at age 62 and in the OAP age that was traditionally at age 65, but we maintain our original focus on the prominent early pension eligibility age.[24] Other characteristics of the VERP program remained the same, including the pension benefits and its independence between spouses. The duration of VERP remained 5 years in length because

[24]Cohorts born later than July 1, 1954 experienced additional increases in their pension eligibility ages that we illustrate in Appendix Figure 2.B.1.

the OAP age increased as well.

The design of the VERP program, which we introduced in Section 2.2, creates strong incentives to retire right at the VERP eligibility age. Hence, the reform induced strong shifts in claiming and retirement ages of the affected individuals that we can use as a first stage to study spillover effects to their spouses. For more details on this reform and an analysis of savings responses of individuals directly affected by it see García-Miralles and Leganza (2020).

2.5.2 Reform-Based Discontinuity Design

To identify the casual effect of increasing individuals' pension eligibility age, we use a local difference-in-differences framework. The treatment group is composed of individuals born on January 1, 1954 or soon after, whose pension eligibility ages increase by 6 months due to the reform. The control group is composed of individuals born right before January 1, 1954, whose pension eligibility ages remain the same. In our main analysis we consider a bandwidth of three months around January 1, 1954 but we show that our results are robust to different bandwidth choices.

We asses the parallel-trends assumption and the dynamics around the announcement and implementation of the reform by estimating a dynamic difference-in-differences model over the period 2008–2014 of the form:

$$y_{it}^{(s)} = \alpha + \delta \cdot treat_i + \sum_{c \neq 2010} \kappa_c \cdot D_c + \sum_{c \neq 2010} \beta_c \cdot D_c \cdot treat_i + X_{it}' \cdot \psi + \epsilon_{it} \qquad (2.4)$$

where $y_{it}^{(s)}$ is the outcome variable of interest, either for the reference individual (y_{it}) or for their spouses (y_{it}^s). D_c are calendar year dummies, and $treat_i$ is an indicator for individuals in the treatment group. The matrix X_{it}' is a set of controls that includes spousal age rounded to quarters interacted with gender when the model is estimated for spousal outcomes.

The results from these dynamic difference-in-differences (Figures 2.5 and 2.6) are discussed in detail in the next section. Note that to assess the parallel-trends assumption we must consider the pre-announcement period (2008–2010), for which we find no evidence of differential trends. During the period between announcement and implementation (2011–2013) treated individuals and their spouses could adjust behaviors in anticipation of reaching increased pension eligibility ages. However, we find no evidence of anticipatory responses for the reference individuals, nor for the spouses despite a slight change in the coefficient for 2013, the year just before implementation. Nevertheless, to be on the safe side we quantify the effects of the reform with respect to the pre-announcement period only, and show that our results are robust to including the anticipation period in the pre-period. Specifically, we estimate the following model to quantify the causal effects of the reform:

$$y_{it}^{(s)} = \beta_0 + \beta_1 \cdot treat_i + \beta_2 \cdot ant_{it} + \beta_3 \cdot post_{it} + \beta_4 \cdot treat_i \cdot ant_{it} + \beta_5 \cdot treat_i \cdot post_{it} + X'_{it} \cdot \psi + \epsilon_{it} \quad (2.5)$$

where $y_{it}^{(s)}$ is the outcome variable of interest, either for the reference individual (y_{it}) or for their spouses (y_{it}^{s}), $treat_i$ is an indicator for individuals in the treatment group, ant_{it} is an indicator for years in the anticipation period (2011-2013), $post_{it}$ is an indicator for implementation year 2014, and X'_{it} is a set of controls that includes spousal age rounded to quarters interacted with spousal gender. When this equation is estimated for the reference individual, the coefficient β_5 identifies the causal effect of the reform on the reference individual (the first stage). When the equation is estimated for the spousal outcomes, the coefficient β_5 identifies the causal effect on the spouses (the reduced-form).

To obtain scaled effects for the spillover of the reform to spouses (Local Average Treatment Effects), we estimate a 2SLS model where the retirement outcomes of the reference individual are instrumented by their treatment status interacted with the calendar year where the reform directly affects them ($treat_i \cdot post_{it}$). The first stage of the 2SLS model corresponds to equation (2.5) when it is estimated for the reference individual's outcomes.

The second-stage equation is the following:

$$y^s_{it} = \beta_0 + \beta_1 \cdot \hat{y}_{it} + \beta_2 \cdot treat_i + \beta_3 \cdot ant_{it} + \beta_4 \cdot post_{it} + \beta_5 \cdot treat_i \cdot ant_{it} + X'_{it}\psi + u_{it} \quad (2.6)$$

where $\hat{y}_{it}$ is the predicted outcome for the reference individual estimated in the first-stage and the coefficient β_1 identifies the scaled spillover effect. We show the validity of the instrument as a strong predictor of the reference individuals' outcomes in the following section. The exclusion restriction is discussed in the robustness Section 2.5.5 along with other specification tests.

2.5.3 The Effect of Increasing the Pension Eligibility Age on Own Retirement

The reform induced a strong response from individuals directly affected by the increase in pension eligibility ages. Figure 2.5 shows the results of the dynamic difference-in-differences model on the retirement outcomes of individuals directly affected by the reform. We confirm that the behavior of the treated and control groups along the three outcomes considered is similar during the period before announcement (2008–2010) as well as before implementation of the reform (2011–2013). The trends of both groups move in parallel and we can rule out any significant anticipatory response.

During the implementation year of 2014, individuals in the treatment group respond to the reform by delaying retirement, consistent with the strong incentives built into the VERP program. Panel (a) of Figure 2.5 shows that individuals in the treatment group are around 19 percentage points less likely to retire during the first half of the year. Note that the reform increased the pension eligibility age by 6 months and hence we define retirement as stopping to work during the first half of the year, as explained in Section 2.3. Individuals affected by the reform are also 26 percentage points less likely to claim benefits, and have higher annual labor market earnings, around $8,140, during the implementation year. In the first row of Table 2.6 we report estimates from the pooled difference-in-differences model, which quantify the large and significant effect of the reform on individuals directly affected, providing a strong first stage to analyze spillover effects to spouses.

2.5.4 The Effect of Increasing the Pension Eligibility Age on Spouses

We now study the effect of the reform on spousal retirement behavior. Figure 2.6 reports the dynamic effects. In the period preceding the announcement of the reform (2008–2010), spouses from both treatment and control individuals behave similarly, providing evidence in support of the parallel trends assumption. After announcement and before implementation (2011–2013), no coefficient is significantly different from zero, suggesting that spouses do not respond differentially in anticipation of their partners' increased pension eligibility age, in line with the lack of anticipation of the reference individuals who are themselves affected directly by the reform.[25] In the implementation year, 2014, we observe that spouses of individuals who are affected by the reform are induced to delay their retirement, consistent with extending employment in order to retire jointly with their partner. We find evidence of spouses adjusting their behavior along the other two margins as well; spouses claim later and increase their annual earnings.

The second row of Table 2.6 reports the difference-in-differences estimates that quantify these spousal effects. The estimates are statistically significant for retirement and claiming, but not for earnings ($690) due to the larger variance of this outcome. We report scaled effects from the 2SLS model in the third row of Table 2.6. The scaled effect on retirement is

[25] Although we do not find evidence of anticipatory responses from spouses, we do observe that in 2013, the year just before implementation, the coefficients tend to move slightly, perhaps suggesting a mild, and not significant, anticipatory response by spouses. This is the reason why in our main model specification to quantify the effect of the reform (equations 2.5 and 2.6) we include an indicator variable for the years between announcement and anticipation of the reform.

9%, indicating that for every 100 individuals who postpone their retirement due to the reform, around 9 spouses will delay their own retirement to make it coincide with that of their partner. The spillover in claiming is 4.2% and the spillover in earnings is 8.5%, although the later is not statistically significant.

Overall, our findings show that the reform induced similar spillover effects as the ones we estimated in a stable context where pension eligibility ages did not change and were know by the couples well in advance. These results are consistent with a lack of significant frictions that prevent couples from adjusting their behavior to retire jointly. This may be of particular interest to policy makers trying to predict short-run responses of social security reforms based on estimates from stable settings. Conversely, it helps with interpreting other reform-based estimates in the literature, as it shows that couples' joint retirement behavior can adjust relatively quickly to changes that affect the retirement age of one partner.

We also explore heterogeneity in responses to the reform. Despite the relatively large sample size of our reform-based design (a panel of 10,321 individuals), we are unable to explore heterogeneous responses in as much depth as in the age-based design, where we estimated effects on reweighted samples and from more granular sample splits. However, the results from a simple age and gender split go in the same direction as the effect we found in the previous section using the age-based design. We report the results in Appendix Table 2.B.1. Older spouses respond the most, with a 12% spillover in retirement against 3% for the younger spouses. The result from a simple split between male and female spouses returns estimates of similar size (9% and 9.1% respectively) as was the case in the age-based design. This suggest again that female spouses respond more once we account for the fact that females are most often the younger member of the couple (around 1.8 years younger in this analysis sample).

2.5.5 Threats to Identification and Robustness

Identifying assumption. The validity of our empirical approach relies on the assumption that in the absence of the reform, spousal outcomes of the treated and control individuals would move in parallel across time. We already showed that trends are parallel in the years preceding the implementation of the reform. However, in interpreting our outcomes as causal, we also assume that spouse behaviors differ in 2014 only because their partners are diferentially affected by the reform. A violation of this assumption occurs if the spouses themselves are directly, and *differentially*, impacted by the same reform.

By construction, treated individuals are 3 months older on average than control individuals, and so are their spouses. Therefore, because the reform affects individuals based on their birth date, older spouses are more likely to be directly impacted by the reform themselves. In this section we show that the differential impact of the reform on the spouses is small and that our results are robust to a series of tests that address this concern.

First, note that only spouses born during the first 6 months of 1954 are affected by the reform that increases their eligibility age from 60 to $60\frac{1}{2}$ and impacts them in 2014. In Appendix Figure 2.B.2 we plot the distribution of spouses' birth dates and show that spouses of treated individuals are only 1.3 percentage points more likely to be born during those 6 months than spouses of control individuals (6.5% against 5.2%). To ensure that our results are not driven by this difference, we do the following two tests. First, we replicate our analysis reweighting the sample of treated individuals so that they have the same distribution of spousal date of birth as the control group. Second, we replicate the analysis excluding individuals whose spouses are born in the first half of 1954, both from the treatment and control groups. The results are reported in rows B and C of Table 2.B.2 and are very similar to the baseline results.

We also note that spouses born after July 1, 1954 are affected by the reform by experiencing larger increases in their pension eligibility ages (as illustrated by Appendix

Figure 2.B.1), but these increases only affect them directly after 2014, and we do not include those years in our analysis. Spouses in the control group are 2.2% more likely to be born after July 1, 1954 (44.3% against 42.1%). Importantly, this differential impact of the reform on the spouses would only affect our results if the reference individuals or their spouses responded in anticipation to future changes in their pension eligibility age. We address this concern in two ways. First, we note that across our analyses, we do not find evidence of anticipatory responses (see Figures 2.5 and 2.6). Second, we replicate our analysis for the subsample of individuals whose spouses are more than 3 months older. This subsample ensures that all spouses are born before January 1, 1954 and therefore are totally unaffected by the reform. The results, reported in Appendix Table 2.B.3 show even larger spillover effects. This is to be expected, as we have shown earlier that older spouses are the ones that respond the most. Overall, these tests make us confident that the small share of spouses who are diferentially impacted by the reform do not have a substantive impact on our results.

Robustness. We perform a series of robustness tests including changes to the model specification and to the sample definition. Table 2.B.2 shows the results. Row D shows that the results are unaffected by estimating the model without the anticipation variable. Rows E and F report the results from decreasing and increasing the bandwidth around the cutoff date of January 1, 1954 by two weeks. Row G shows the results when we do not balance the sample of analysis. Row H adds controls for region and education of the reference individuals and their spouses, defined when they are 57 years old. Finally, row I extends the sample to include reference individuals who did not contribute to the VERP program between ages 50-59. Overall our results are robust to all these changes. We note, however, that although the size of the estimates for claiming remains stable, they turn insignificant in some cases, and the same happens to the estimates for earnings, which remain insignificant in most cases.

2.6 Conclusion

Spouses adjust their behavior to retire together, which implies a significant role for leisure complementarities within couples. We estimate joint retirement spillovers induced by pension eligibility ages in two complementary settings. In the first setting the pension eligibility age is stable and known by couples well in advance, whereas in the second setting the pension eligibility age increases due to a reform that discontinuously affects selected cohorts. We find similar joint retirement spillovers in both settings, suggesting that joint retirement behavior prevails in a reform context and is not hampered by adjustment costs. Specifically, we find that for every 100 individuals who retire upon reaching pension eligibility age, around 8 of their spouses are induced to retire as well.

Our data allow us to advance the understanding of mechanisms and behaviors that underlie joint retirement decisions. We explore different margins of adjustment such as claiming and annual earnings, and we document strong heterogeneous responses. Joint retirement is largely driven by older spouses who work past their own pension eligibility age, waiting for their younger spouses to become eligible for their own pension benefits. We uncover a significant and consistent gender gap, where female spouses are more likely to adjust their retirement age to make it coincide with that of their male partner. This gender gap emerges after controlling for the age composition of couples, since men tend to be older than females and this confounds the effect from a simple gender split. The gender gap is not explained by differences in relative earnings within couples. Relative earnings within couples do not seem to be major determinants of joint retirement, but we find patterns consistent with the opportunity cost of retirement.

Our results, which are derived in the context of a representative pension system, have implications for the design and evaluation of public policies. We find that policies that delay retirement ages of individuals can have spillover effects to spouses, and the size of these effects depends crucially on the age of spouses relative to their partners and on their capacity to claim benefits of their own. Our findings suggest that increasing the retirement age of younger partners (who are traditionally females) will generate the largest spillover effects in the form of delayed retirement of their older spouses. This is particularly relevant for countries whose statutory retirement ages are still lower for females. Our findings may also inform models of intra-household decision making more generally, which are increasingly the subject of theoretical and structural work on labor supply and retirement.

2.7 Acknowledgements

Chapter 2, in full, is currently being prepared for submission for publication of the material. García-Miralles and Leganza, Jonathan M. "Joint Retirement of Couples: Evidence from Discontinuities in Denmark." The dissertation author was a primary investigator and an author of this material.

2.8 Figures and Tables

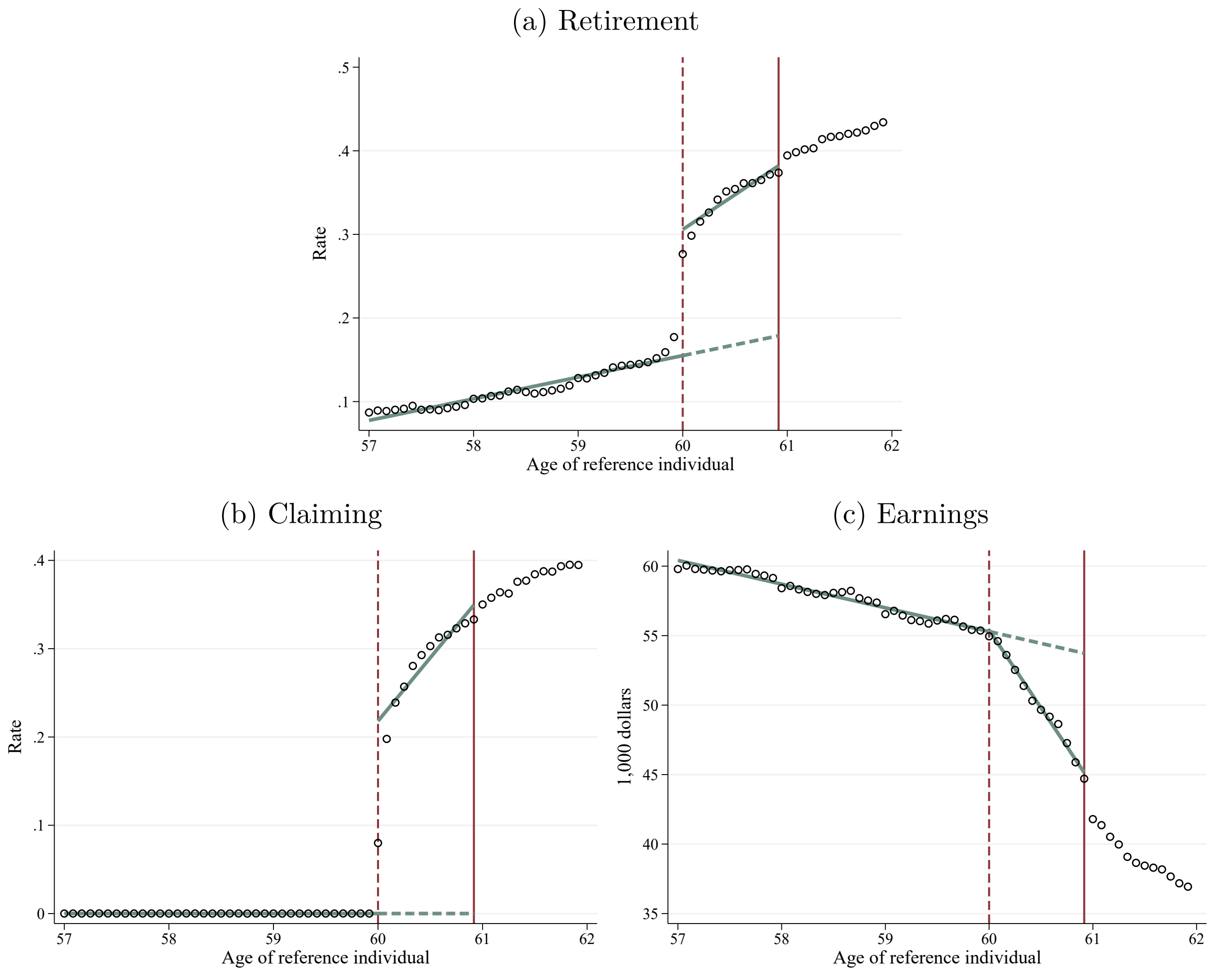


Figure 2.1: The Effect of Reaching Pension Eligibility Age on Own Retirement

Notes: These figures plot different outcomes for individuals around their own pension eligibility age of 60, pooling individuals over the period 1991–2013. The hollow circles are raw means of the outcome variable measured at the end of each calendar year, grouped in monthly age bins. The solid lines plot the parametric fit estimated with the piecewise linear regression model (2.1). The dashed line represents the counterfactual behavior in the absence of pension eligibility, based on a linear extrapolation from the observed outcome before age 60. The full-exposure effect of being eligible for early retirement pension during an entire year is represented by the vertical distance between the solid and dashed lines just below age 61.

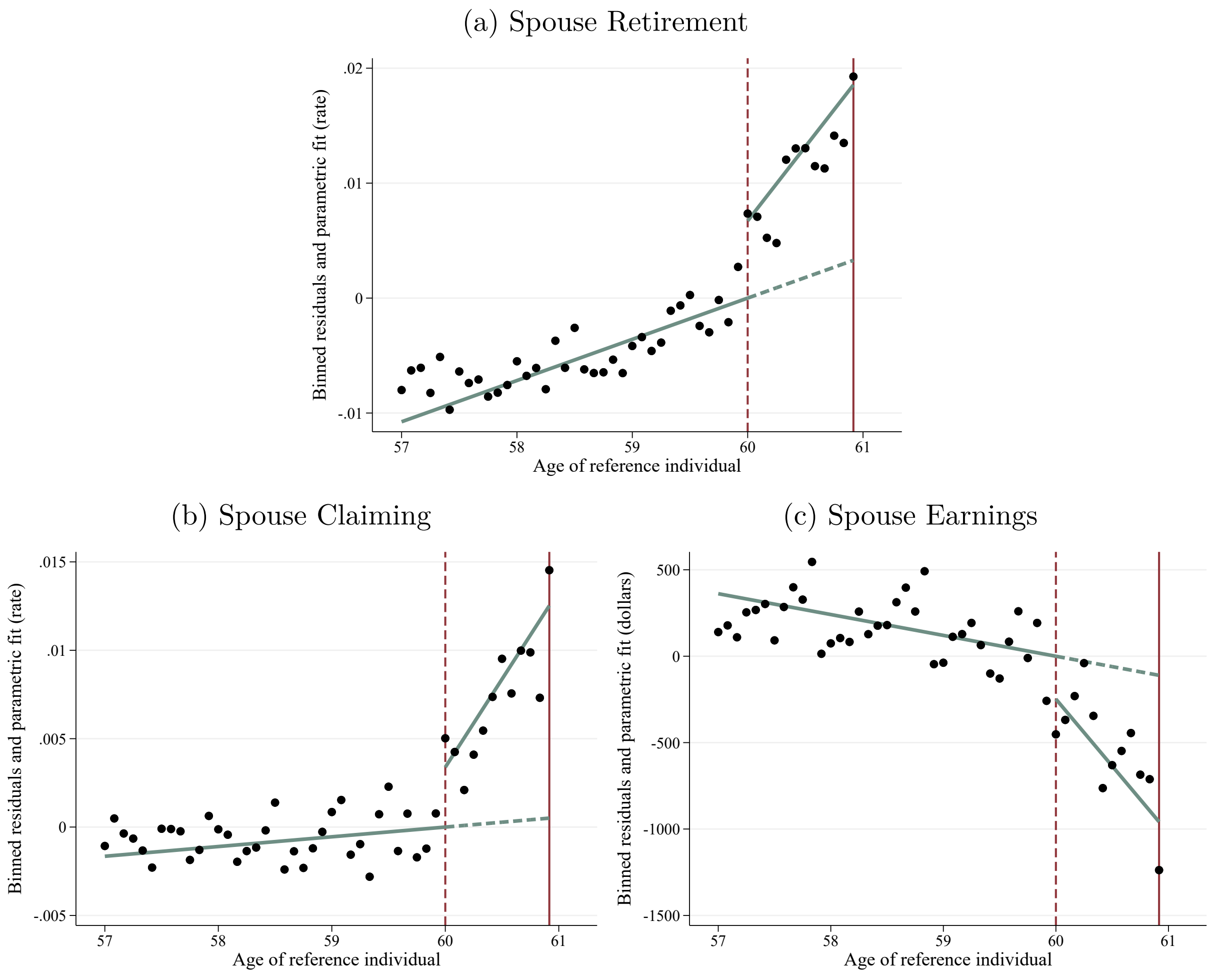


Figure 2.2: The Effect of Reaching Pension Eligibility Age on Spouses

Notes: These figures plot different outcomes for spouses around the pension eligibility age of their partner. The dots are the residuals estimated in equation (2.2) where the spousal outcome is regressed on their own age and gender. The residuals are grouped in monthly bins of the reference individual's age. The solid lines plot the parametric fit estimated with the piecewise linear regression model (2.3). The dashed line represents the counterfactual behavior in the absence of pension eligibility, based on a linear extrapolation from the observed outcome before age 60. The full-exposure effect on the spouses of their partners being eligible for early retirement pension during an entire year is represented by the vertical distance between the solid and dashed lines just below age 61.

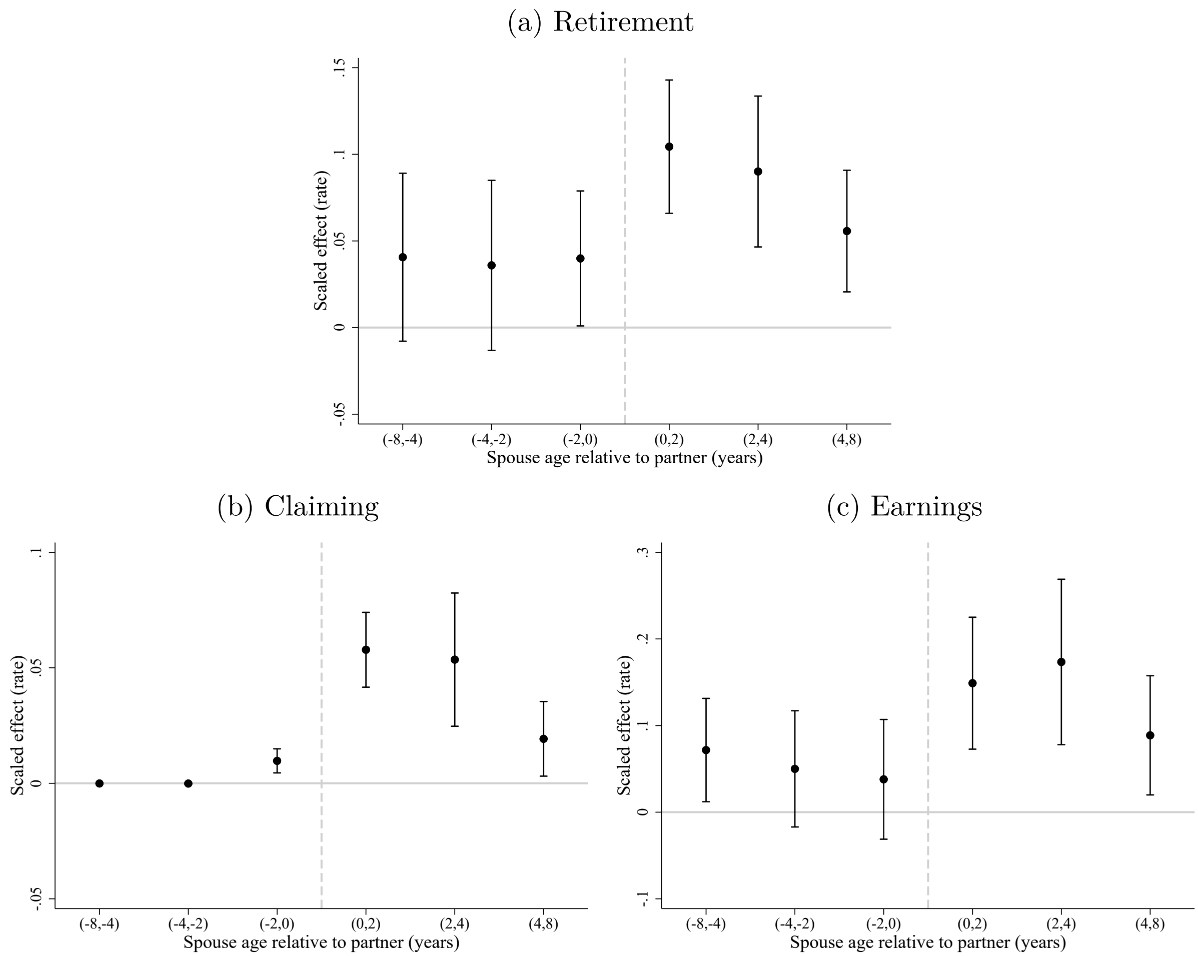


Figure 2.3: Joint Retirement Behavior by Age Differences Within Couples

Notes: These figures plot the scaled estimates of joint retirement for different subsamples of couples based on the age difference between spouses. These scaled effects are estimated using the same methodology as for the full sample: first estimating models (2.1) and (2.3) to obtain full-exposure effects and then dividing the full-exposure effect on spouses by the full-exposure effect on reference individuals. We report 95% confidence intervals calculated from bootstrapped standard errors.

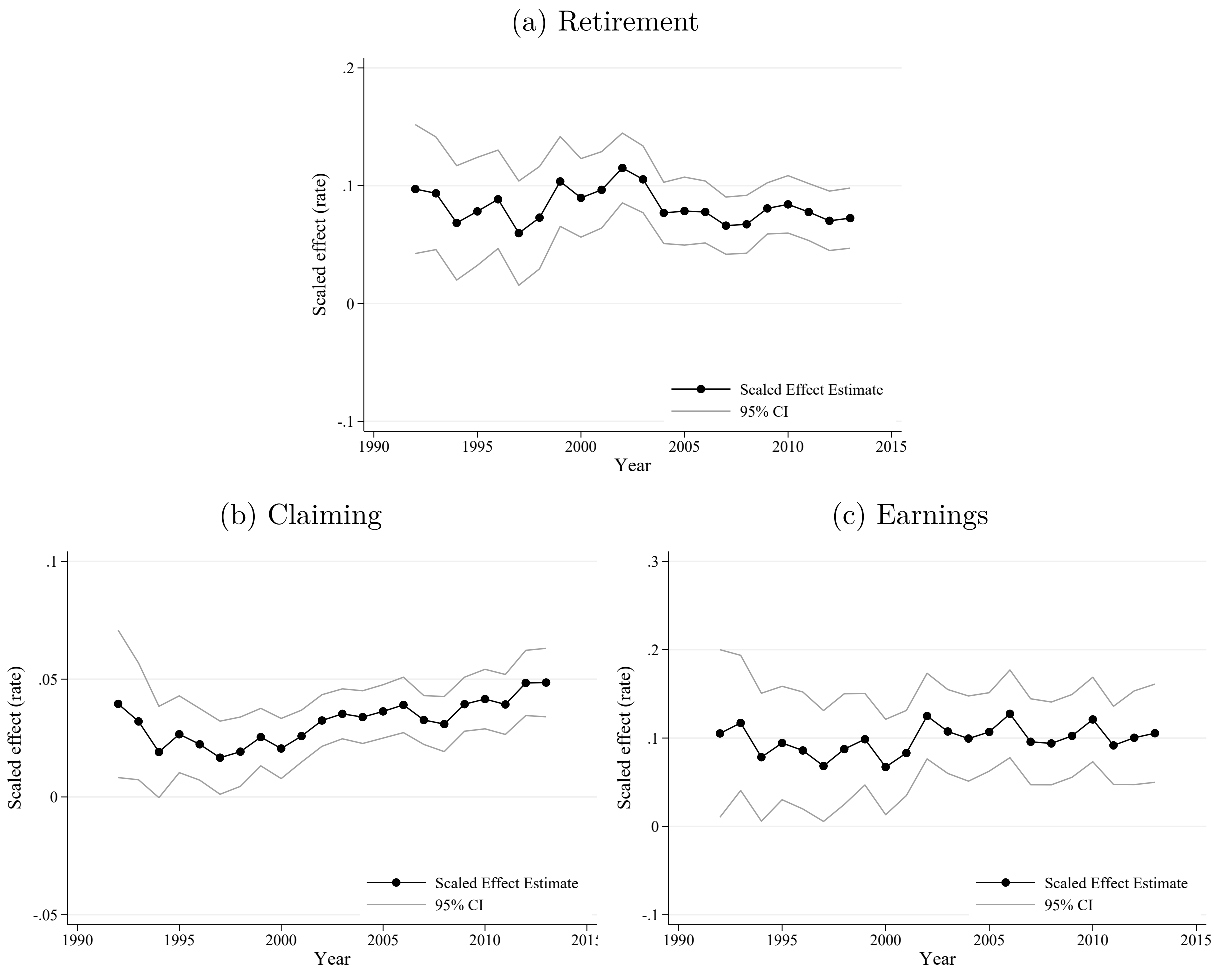


Figure 2.4: The Evolution of Joint Retirement Over Time

Notes: These figures plot the evolution over time of the scaled estimates of joint retirement for different outcomes. Scaled effects are estimated over a 5-year running window using the same methodology as for the full time period: first estimating models (2.1) and (2.3) to obtain full-exposure effects and then dividing the full-exposure effect on spouses by the full-exposure effect on reference individuals. The scaled effects and the full-exposure effects for the whole period 1991–2013 are reported in Table 2.2. We report 95% confidence intervals calculated from bootstrapped standard errors.

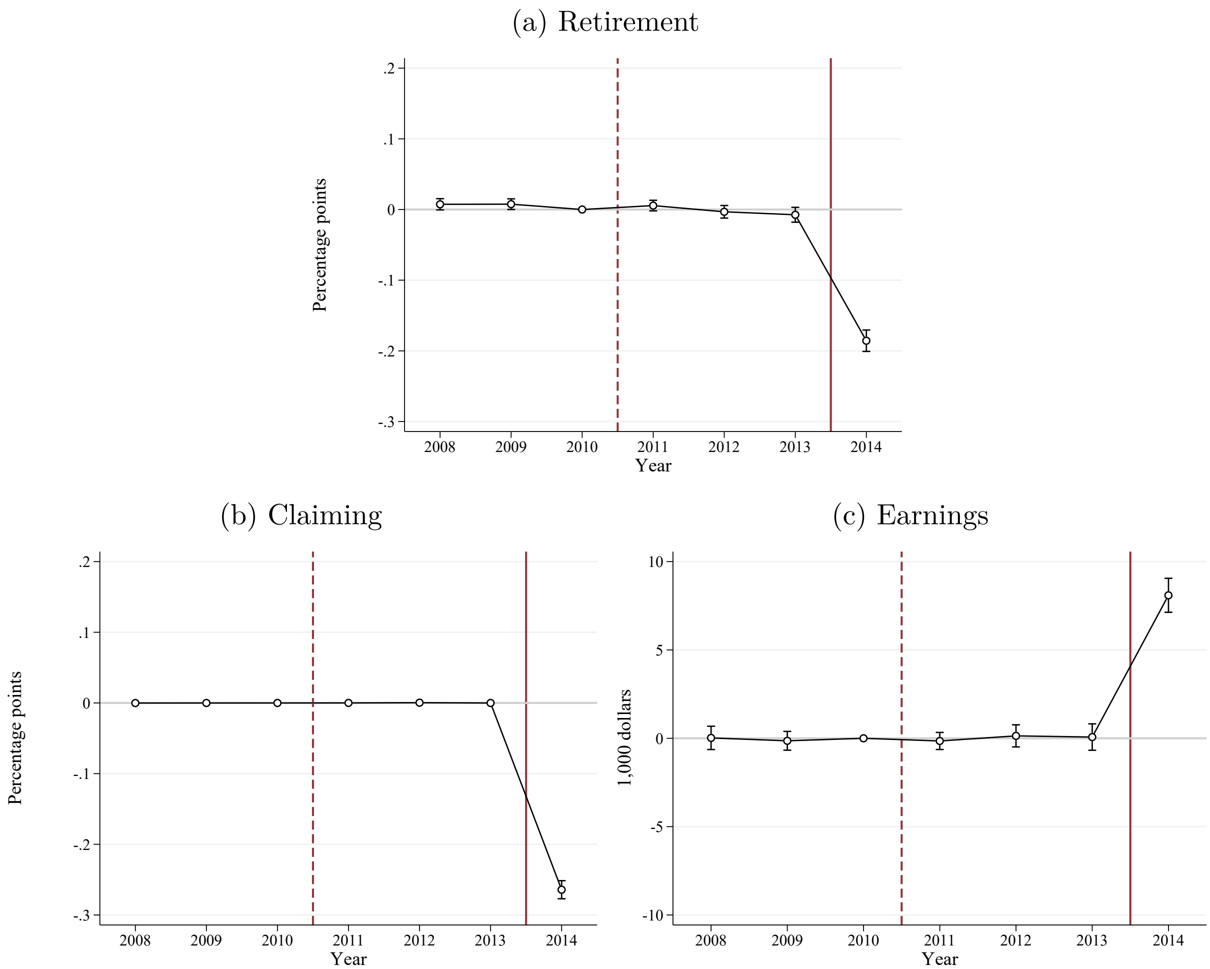


Figure 2.5: The Effect of Increasing Pension Eligibility Age on Own Retirement

Notes: These figures plot the β_c coefficients from the dynamic difference-in-differences model (2.4), estimated on different outcomes for reference individuals. Each coefficient shows the difference between the treated group (whose pension eligibility age increases by 6 months, to age $60\frac{1}{2}$) and the control group (whose pension eligibility age remains at age 60). Individuals turn 60 around the beginning of 2014, therefore the coefficient for 2014 identifies the causal effect of the reform during the implementation year. We report confidence intervals at the 95% level, calculated from robust standard errors clustered at the couple level.

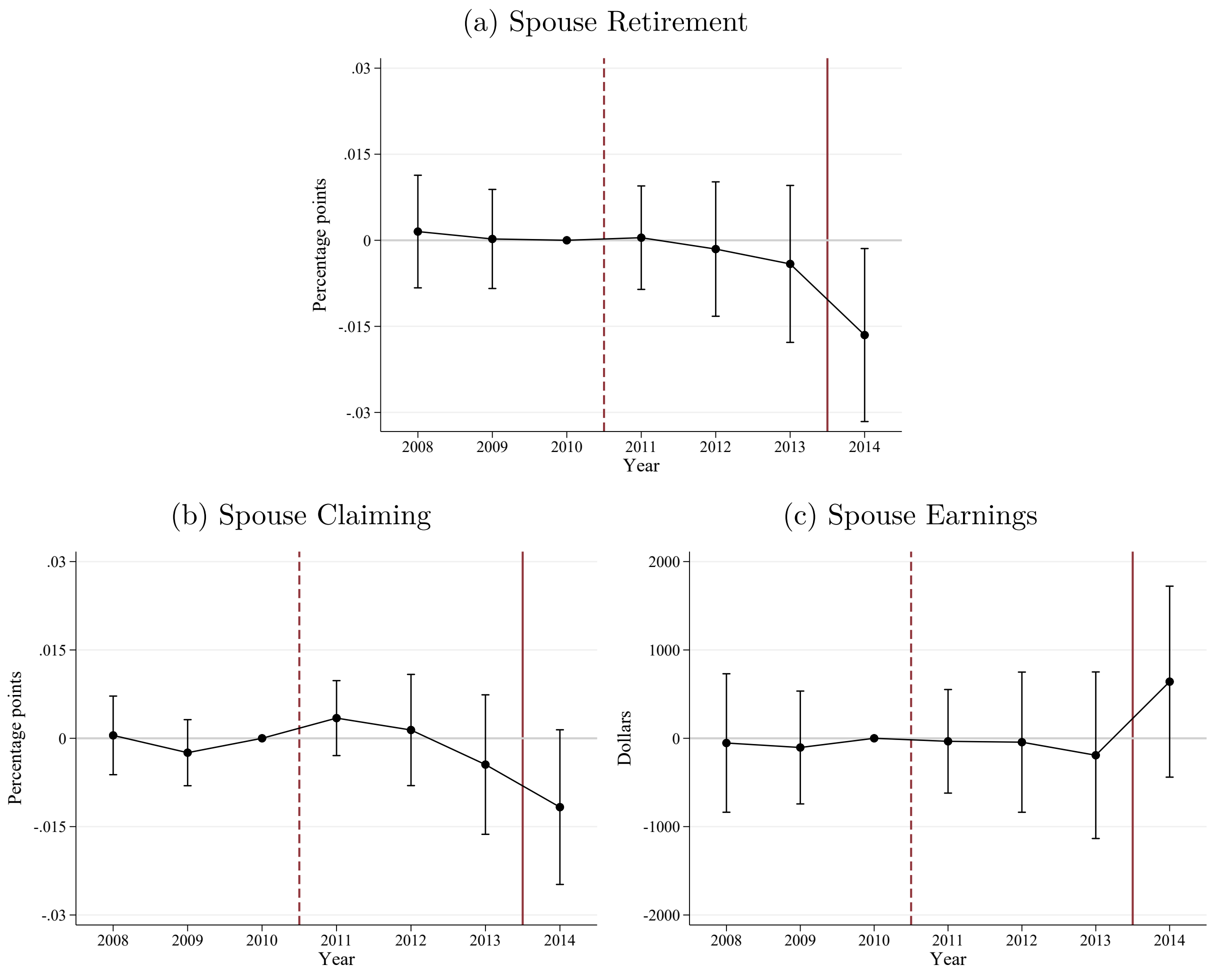


Figure 2.6: The Effect of Increasing Pension Eligibility Age on Spouses

Notes: These figures plot the β_c coefficients from the dynamic difference-in-differences model from equation (2.4), estimated on different outcomes for spouses of reference individuals. Each coefficient shows the difference between the treatment group (spouses whose partners' pension eligibility age increases by 6 months, to age $60\frac{1}{2}$) and the control group (spouses whose partners' pension eligibility age remains at 60). The coefficient for 2014 identifies the causal effect of the reform on the spouses on the implementation year. We report confidence intervals at the 95% level, calculated from robust standard errors clustered at the couple level.

Table 2.1: Summary Statistics

	Age-Based Design Period (1991–2013)				Reform-Based Design Period (2008–2014)			
	Population		Analysis Sample		Population		Analysis Sample	
	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)	Mean (7)	SD (8)
A: Reference Individuals								
Age	58.45	1.12	58.44	1.12	57.45	2.04	57.47	2.06
Male	0.51	0.50	0.52	0.50	0.50	0.50	0.47	0.50
Dane	0.98	0.15	1.00	0.00	0.97	0.18	1.00	0.00
Copenhagen region	0.26	0.44	0.27	0.44	0.25	0.43	0.22	0.41
Educ. Primary	0.37	0.48	0.29	0.45	0.30	0.46	0.25	0.43
Educ. Secondary	0.41	0.49	0.45	0.50	0.41	0.49	0.45	0.50
Educ. Tertiary	0.03	0.18	0.04	0.19	0.04	0.20	0.04	0.20
Educ. Bachelor	0.14	0.34	0.17	0.37	0.18	0.39	0.20	0.40
Educ. Master	0.05	0.22	0.05	0.22	0.07	0.26	0.05	0.23
Earnings age 55-57	45,268	41,165	60,289	35,186	55,582	41,780	64,156	32,218
Retired by age 57	0.20	0.40	0.09	0.29	0.25	0.43	0.12	0.32
Retired by age 58	0.22	0.41	0.11	0.31	0.26	0.44	0.13	0.34
Retired by age 59	0.24	0.43	0.14	0.35	0.29	0.45	0.16	0.37
Retired by age 60	0.39	0.49	0.34	0.47	0.43	0.49	0.35	0.48
B: Spouses								
Age difference (years)	0.34	5.23	0.25	3.46	0.19	5.26	-0.10	3.50
Age	58.11	5.36	58.19	3.64	57.26	5.62	57.57	4.04
Male	0.49	0.50	0.48	0.50	0.50	0.50	0.53	0.50
Dane	0.99	0.08	1.00	0.06	0.98	0.12	0.99	0.08
Copenhagen region	0.26	0.44	0.27	0.44	0.25	0.43	0.22	0.41
Educ. Primary	0.37	0.48	0.29	0.46	0.28	0.45	0.23	0.42
Educ. Secondary	0.41	0.49	0.44	0.50	0.42	0.49	0.45	0.50
Educ. Tertiary	0.03	0.18	0.04	0.19	0.04	0.21	0.05	0.23
Educ. Bachelor	0.14	0.35	0.17	0.38	0.18	0.39	0.20	0.40
Educ. Master	0.05	0.22	0.05	0.22	0.07	0.26	0.06	0.24
Earnings age 55-57	45,877	39,995	58,419	34,725	56,091	43,924	66,224	34,921
Retired by age 57	0.20	0.40	0.12	0.33	0.26	0.44	0.15	0.36
Retired by age 58	0.21	0.41	0.13	0.34	0.26	0.44	0.14	0.35
Retired by age 59	0.22	0.42	0.15	0.35	0.26	0.44	0.15	0.35
Retired by age 60	0.34	0.48	0.30	0.46	0.35	0.48	0.27	0.44
Number of Observations	4,366,996		2,206,044		166,554		73,395	

Notes: This table reports means and standard deviations of relevant variables for different samples of interest. The first four columns correspond to the age-based period of analysis (1991–2013) where the pension eligibility age remained stable, and it includes individuals of age 57 to 60. The last four columns correspond to the reform-based period of analysis (2008–2014) where the pension eligibility age was increased starting in 2014, and it includes individuals born between July 1, 1953 and June 30, 1954. Columns denoted "Population" correspond to the full population without applying any sample restriction. Columns denoted "Analysis sample" correspond to our baseline samples of analysis, after applying the restrictions described in Section 2.3.3.

Table 2.2: The Effect of Reaching Pension Eligibility Age

	Retirement	Claiming	Earnings
Reference Individual	0.2034***	0.3496***	-8,642***
	(0.001)	(0.001)	(69.431)
Spouse	0.0153***	0.0120***	-848***
	(0.001)	(0.001)	(61.165)
Scaled Effect	0.0750***	0.0344***	0.0981***
	(0.0071)	(0.0031)	(0.012)
N. of clusters	367,585	367,585	367,585
Observations	2,206,044	2,206,044	2,206,044

Notes: This table reports the effect of reference individuals reaching pension eligibility age on their own retirement and on their spouses' retirement. Each column reports the results for a different outcome. The first row reports the full-exposure effect to pension eligibility on own retirement estimated in equation (2.1). The second row reports the full-exposure effect on the spouses from their partners becoming eligible for pension, estimated in equation (2.3). The third row reports the scaled effect resulting from diving the spouse full-exposure effect by the reference individual full-exposure effect. Robust standard errors in parentheses, clustered at the couple level. Bootstrapped standard errors for scaled effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2.3: Heterogeneity in the Effect of Reaching Pension Eligibility Age on Retirement by Age Difference and Gender

Reference Individual	Young	Old	Female	Male	Male (w)	Male (w)
Spouse	Old	Young	Male	Female	Female (w)	Female (w)
	(1)	(2)	(3)	(4)	(5)	(6)
A. Retirement						
Reference Individual	0.2562***	0.1588***	0.2668***	0.1479***	0.1616***	0.1795***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
Spouse	0.026***	0.004***	0.020***	0.011***	0.021***	0.024***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Scaled Effect	0.0994***	0.0287***	0.0745***	0.0751***	0.130***	0.136***
	(0.0088)	(0.010)	(0.0070)	(0.013)	(0.018)	(0.017)
B. Claiming						
Reference Individual	0.4307***	0.28***	0.4567***	0.2544***	0.2632***	0.295***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Spouse	0.021***	0.000	0.017***	0.008***	0.018***	0.020***
	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)	(0.003)
Scaled Effect	0.0495***	0.00350***	0.0374***	0.0301***	0.0674***	0.0691***
	(0.0053)	(0.0010)	(0.0044)	(0.0038)	(0.0092)	(0.010)
C. Earnings						
Reference Individual	-9,558***	-7,970***	-9,081***	-8,408***	-9,035***	-9,160***
	(93.657)	(97.971)	(81.417)	(104.024)	(140.987)	(162.946)
Spouse	-1,856***	-510***	-1,168***	-602***	-589***	-769***
	(117.525)	(79.457)	(160.724)	(68.661)	(97.229)	(151.466)
Scaled Effect	0.184***	0.0608***	0.122***	0.0680***	0.0661***	0.0849***
	(0.020)	(0.017)	(0.021)	(0.014)	(0.018)	(0.025)
N. of clusters	297,686	334,966	302,589	330,172	330,172	330,172
Observations	1,038,096	1,167,948	1,054,359	1,151,685	1,151,685	1,151,685

Notes: This table reports the effect of the reference individuals reaching pension eligibility age on their own retirement and on their spouses' retirement, distinguishing heterogeneous responses by gender and age differences within the couple. Each column shows results for a different subsample. The subsample in column (5) is reweighted to have the same distribution of age differences as the subsample from column (3) and the subsample in column (6) is further reweighted to have the same distribution of earnings shares as (3). Each panel reports results for a different outcome variable. Within each panel, the first row reports the full-exposure effect of pension eligibility on own retirement. The second row reports the full-exposure effect on spouses of their partners being eligible for retirement pension. The third row reports the scaled effect resulting from diving the spouse full-exposure effect by the reference individual full-exposure effect. Robust standard errors in parentheses, clustered at the couple level. Bootstrapped standard errors for scaled effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.4: Heterogeneity in the Effect of Reaching Pension Eligibility Age on Retirement by Relative Earnings

A. By Gender

Reference Individual	Female Primary	Female Sec. (w)	Male Primary	Male Secondary (w)
Spouse	Male Secondary	Male Primary (w)	Female Secondary	Female Primary (w)
	(1)	(2)	(3)	(4)
Reference Individual	0.2475***	0.2745***	0.1434***	0.1426***
	(0.004)	(0.002)	(0.002)	(0.004)
Spouse	0.0111***	0.025***	0.0117***	0.003
	(0.003)	(0.002)	(0.001)	(0.003)
Scaled Effect	0.0434**	0.0909***	0.0816***	0.0225
	(0.018)	(0.011)	(0.0074)	(0.02)
N. of clusters	58,311	201,541	229,321	53,949
Observations	191,681	713,870	800,843	185,860

B. By Age Differences

Reference Individual	Young Primary	Young Sec. (w)	Old Primary	Old Second. (w)
Spouse	Old Secondary	Old Prim. (w)	Young Secondary	Young Prim. (w)
	(1)	(2)	(3)	(4)
Reference Individual	0.2094***	0.2651***	0.1412***	0.1541***
	(0.003)	(0.003)	(0.002)	(0.004)
Spouse	0.0197***	0.028***	0.0073***	0.004
	(0.003)	(0.002)	(0.002)	(0.002)
Scaled effect	0.0928***	0.104***	0.0537***	0.0265
	(0.022)	(0.014)	(0.019)	(0.023)
N. of clusters	94,735	161,573	193,106	93,917
Observations	321,816	571,978	671,295	327,752

Notes: The table reports the effect of the reference individuals reaching pension eligibility age on their own retirement and on their spouses' retirement, distinguishing heterogeneous responses by primary earner status within the couple. Panel A further distinguish by gender and Panel B by age differences. Each column contains results for a subsample of the population. In Panel A, the subsamples in columns (2) and (4) are reweighed to have the same distribution of age differences as columns (1) and (3), respectively. In Panel B the subsamples in columns (2) and (4) are reweighed to have the same distribution of gender and age differences as columns (1) and (3), respectively. Within each panel, the first row reports the full-exposure effect of pension eligibility on own retirement. The second row reports the full exposure-effect on spouses of their partners being eligible for retirement pension. The third row reports the scaled effect resulting from diving the spouse full-exposure effect by the reference individual full-exposure effect. Robust standard errors in parentheses, clustered at the couple level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2.5: Robustness to Alternative Specifications for the Effect of Reaching Pension Eligibility Age

	Retirement	Claiming	Earnings
A. Baseline	0.0750***	0.0344***	0.0981***
	(0.0071)	(0.0031)	(0.012)
B. Including Younger Ages	0.0752***	0.0360***	0.0934***
	(0.0052)	(0.0026)	(0.010)
C. Excluding Age 59	0.0904***	0.0387***	0.120***
	(0.011)	(0.0047)	(0.018)
D. Unrestricted Age Difference	0.0690***	0.0311***	0.0905***
	(0.0070)	(0.0031)	(0.012)
E. No Donut December	0.0730***	0.0321***	0.0926***
	(0.0068)	(0.0031)	(0.012)
F. Nonlinear Counterfactual	0.0407***	0.0356***	0.0545*
	(0.016)	(0.0053)	(0.030)
G. Nonlinear & Incl. Younger	0.0691***	0.0327***	0.0943***
	(0.0083)	(0.0035)	(0.016)
H. Adding Controls	0.0747***	0.0339***	0.0924***
	(0.0069)	(0.0033)	(0.011)
I. Dummy 1999 Reform	0.0746***	0.0344***	0.0978***
	(0.0065)	(0.0028)	(0.012)
J. Period 2008–2013	0.0760***	0.0463***	0.107***
	(0.012)	(0.0073)	(0.024)
K. 2008–2013 & VERP Eligible	0.0705***	0.0438***	0.107***
	(0.011)	(0.0067)	(0.025)
L. Retirement Flow Variable	0.0573***	–	–
	(0.0055)		

Notes: This table reports the scaled effect estimates from replicating our main analysis over different sample definitions and over different specifications of the estimation models (equations 2.1 and 2.3). Row A reproduces results from our baseline specification, which correspond to those reported in Table 2.2. Row B replicates the analysis over a sample extended to include reference individuals of ages 55 and 56. Row C excludes reference individuals aged 59. Row D keeps couples with partners that are more than 8 years apart from each other. Row E keeps reference individuals who turn 60 in December. Row F allows for a nonlinear counterfactual by adding a second order polynomial. Row G implements the two changes applied in B and F. Row H controls for predetermined region and education of reference individuals and spouses. Row I adds a dummy for individuals born after 1939, who are affected by the 1999 reform. Row J estimates the effect over the period 2008–2013. Row K estimates the effect over the same period as J and restricts the sample to reference individuals who have contributed to VERP at least once between ages 50 and 59. Row L reports the estimate for retirement defined as a flow variable, allowing individuals to retire multiple times. Bootstrapped standard errors in parentheses. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

Table 2.6: The Effect of Increasing Pension Eligibility Age

	Retirement	Claiming	Earnings
Reference Indiv.	-0.191***	-0.264***	8,140***
	(0.0074)	(0.0065)	(479)
Spouse	-0.0172**	-0.0110*	690
	(0.0073)	(0.0064)	(532)
Scaled Effect	0.0902**	0.0418*	0.0847
	(0.038)	(0.024)	(0.065)
F-test instr.	662.3	1643.6	288.8
N. of clusters	10,321	10,321	10,321
Observations	73,395	73,395	73,395

Notes: This table reports the effect of the 2011 reform, which increased the pension eligibility age. Each column reports results for a different outcome. The first row reports the effect on the individuals affected by the reform (the first stage) and the second row reports the spillover effect to their spouses (the reduced-form effect), which are estimated using equation (2.5). The third row reports the scaled effect (the LATE) resulting from the 2SLS model estimated in equation (2.6). Robust standard errors in parentheses, clustered at the couple level. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

2.A Appendix: Age Discontinuity Design

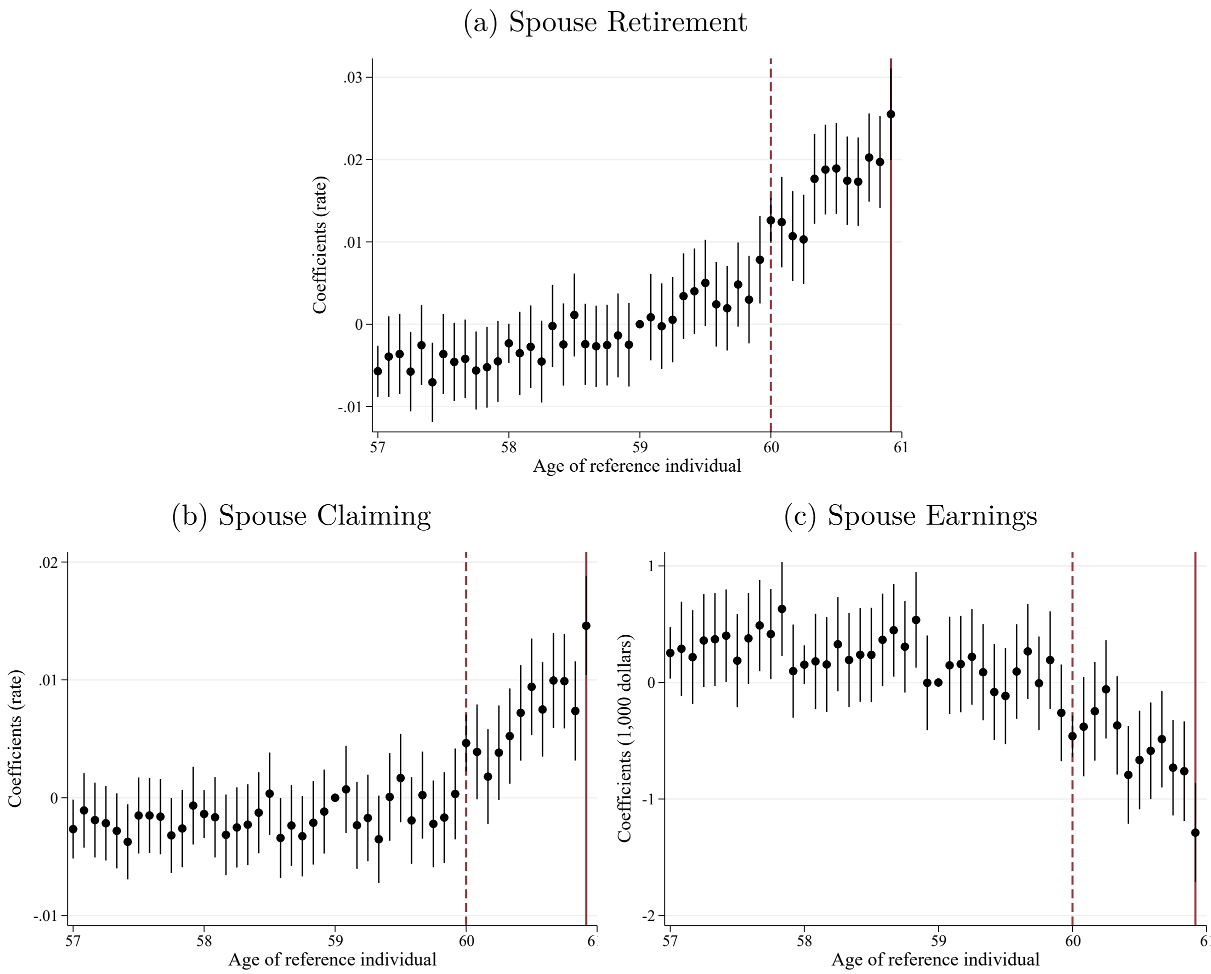


Figure 2.A.1: Alternative Graphical Evidence of the Effect of Pension Eligibility Age on Spouses

Notes: These figures show an alternative approach to obtain nonparametric evidence on spouses behavior around the pension eligibility age of their partner. They plot the δ^r_a coefficients from estimating the regression $y^s_{it} = \alpha + \sum_{a=57}^{62} \delta^r_a \cdot D^r_a + \sum_{a=49}^{69} \delta^s_a \cdot D^s_a + \sum_{a=49}^{69} \gamma_a \cdot D^s_a \cdot D_g + \sum_{c=1991}^{2013} \kappa_c \cdot D_c + \epsilon_{st}$, where y^s_{it} are the different outcomes plotted in each figure, D^r_a are age dummies for the reference individual, D^s_a are age dummies for the spouse, D_g is a gender dummy for the spouse, and D_c are calendar year dummies.

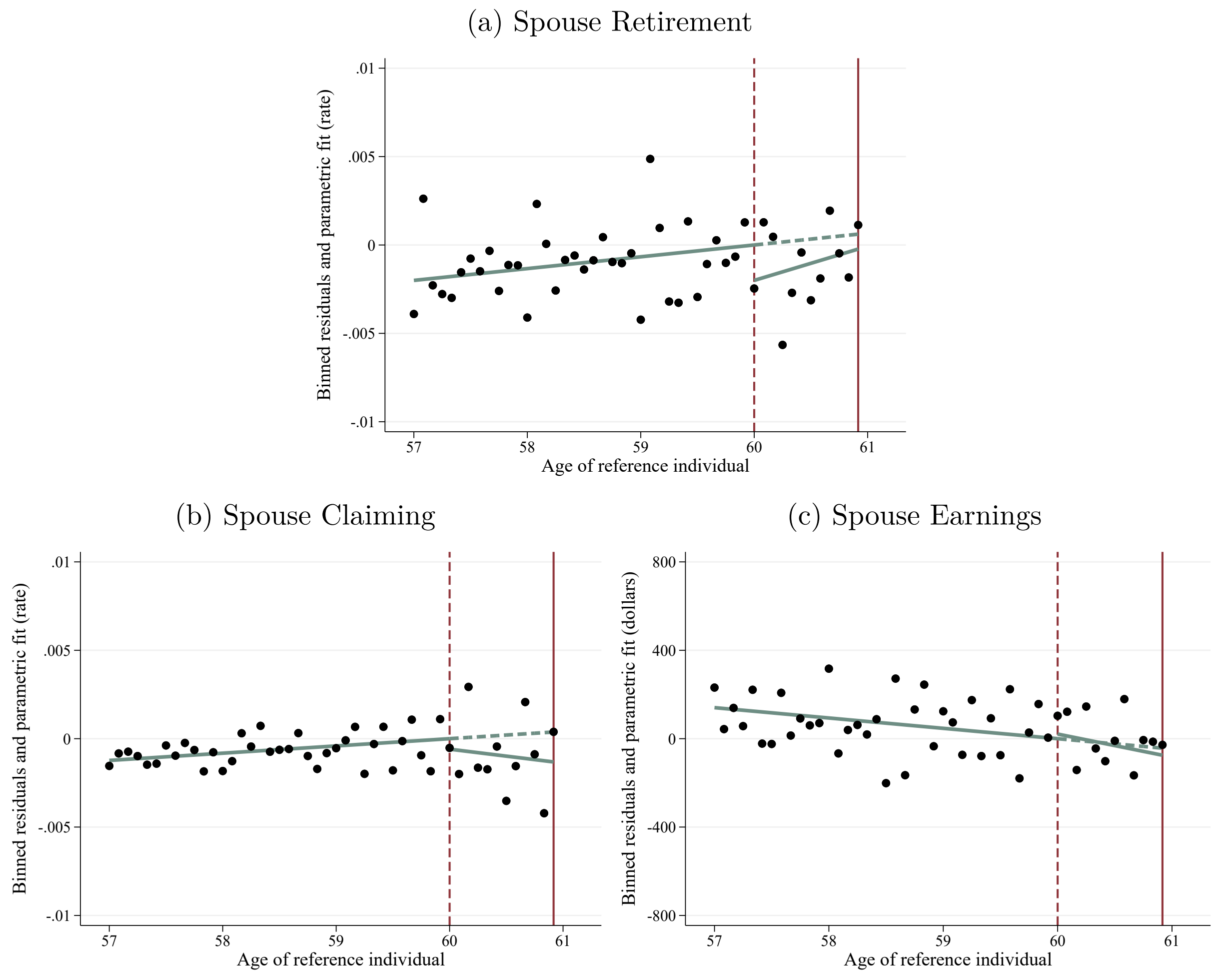


Figure 2.A.2: Placebo Test Assigning Fake Spouses of Similar Age for the Effect of Reaching Pension Eligibility Age

Notes: These figures plot results from replicating the analysis over a placebo sample where the reference individuals are the same as in the main analysis, but they are matched to fake spouses of similar age. The figures show no evidence of joint retirement, as is expected if the research design is valid: fake spouses cannot affect each other's retirement behavior, and the effect coming from the correlation between their ages is controlled for by the empirical design. For more details on the construction of this figure, see the notes of Figure 2.2. See Appendix Table 2.A.3 for the placebo point estimates.

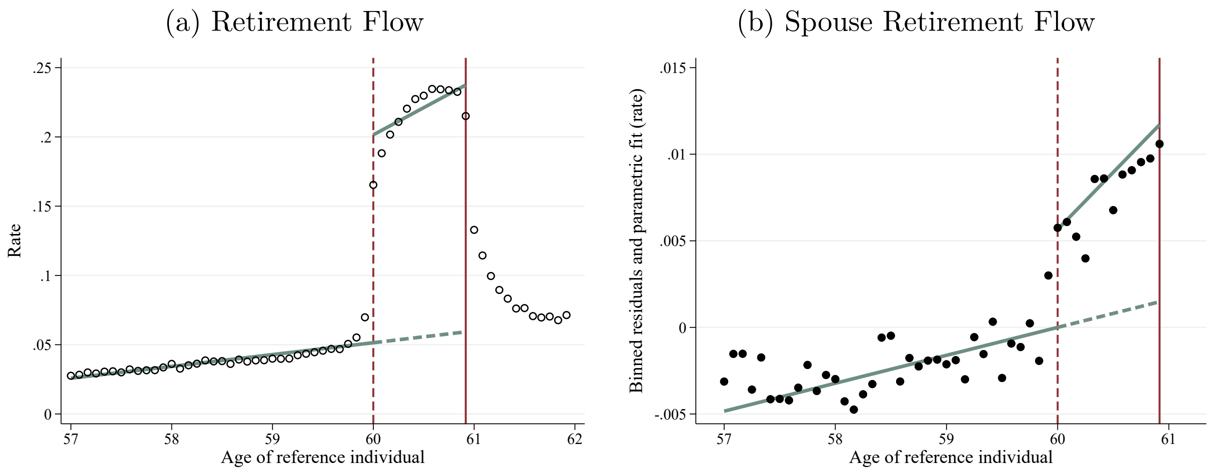


Figure 2.A.3: The Effect of Reaching Pension Eligibility Age on Retirement Defined as Flow

Notes: These figures plot an alternative definition of the retirement outcome, defined as a flow variable that takes the value one in the year in which an individual retires and zero otherwise. For more details on the construction of these figures see notes of Figures 2.1 and 2.2. The scaled effect estimate resulting from this outcome is reported in Table 2.5.

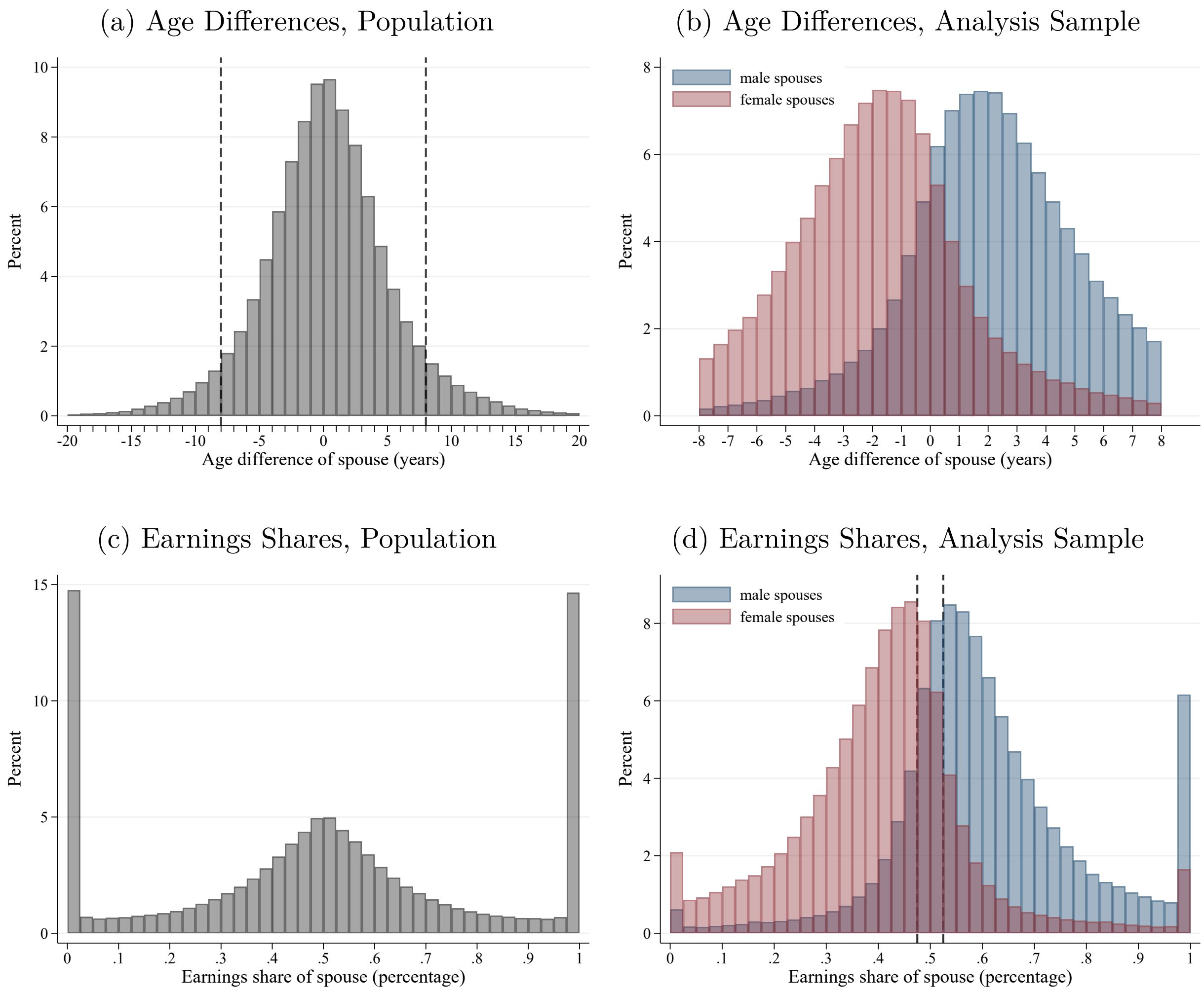


Figure 2.A.4: Distribution of Spouses' Age Differences and Earnings Shares

Notes: Panel (a) plots the distribution of age differences within spouses for the population of Danish couples between 1991 and 2013, before applying the sample restrictions described in Section 2.3.3. The vertical dashed lines mark the tails that are excluded from the sample of analysis, corresponding to couples with more than 8 years difference in age. Panel (b) plots the distribution of age differences for the age-based sample of analysis resulting from imposing the restrictions described in Section 2.3.3. Panel (c) plots the distribution of earnings shares within the couple, based on average annual labor market earnings of each partner between ages 55 and 57, for the full Danish population between 1991 and 2013. Panel (d) plots earnings shares for the age-based sample of analysis. The vertical dashed lines mark the interval of couples with very similar earnings shares (between 0.475 and 0.525) who are excluded in the heterogeneity analysis that defines an indicator variable to identify which member of the couple is the primary earner.

Table 2.A.1: Heterogeneity in the Effect of Reaching Pension Eligibility Age on Retirement. Alternative to Reweighting: Split by Age Differences and Gender

Reference Indiv.	Young Female	Young Male	Old Female	Old Male
Spouse	Old Male	Old Female	Young Male	Young Female
	(1)	(2)	(3)	(4)
A. Retirement				
Reference Indiv.	0.2801***	0.1765***	0.2257***	0.1409***
	(0.002)	(0.004)	(0.004)	(0.002)
Spouse	0.024***	0.030***	0.002	0.005***
	(0.002)	(0.003)	(0.002)	(0.001)
Scaled Effect	0.0872***	0.167***	0.00954	0.0359***
	(0.0095)	(0.030)	(0.015)	(0.013)
B. Claiming				
Reference Indiv.	0.4758***	0.2793***	0.3975***	0.2482***
	(0.002)	(0.004)	(0.004)	(0.002)
Spouse	0.020***	0.025***	0.000	0.000
	(0.002)	(0.004)	(0.001)	(0.000)
Scaled Effect	0.0428***	0.0878***	0.00413*	0.00331***
	(0.0061)	(0.018)	(0.0025)	(0.0013)
C. Earnings				
Reference Indiv.	-9,579***	-9,740***	-7,571***	-8,076***
	(93.052)	(248.045)	(166.764)	(114.302)
Spouse	-1,881***	-1,200***	-284	-583***
	(131.71)	(191.19)	(226.94)	(75.48)
Scaled Effect	0.197***	0.127***	0.0450	0.0725***
	(0.023)	(0.032)	(0.052)	(0.016)
N. of clusters	228,199	69,596	74,390	260,576
Observations	797,667	240,429	256,692	911,256

Notes: This table reports the effect of the reference individuals reaching pension eligibility age on their own retirement and on their spouses' retirement, distinguishing heterogeneous responses by gender and age composition of the couple. Each column contains results for a different subsample. Each panel reports results for a different outcome variable. Within each panel, the first row reports the full exposure effect of pension eligibility on own retirement as estimated in equation (2.1). The second row reports the full exposure effect on the spouses of their partners being eligible for retirement pension estimated in equation (2.3). The third row reports the scaled effect resulting from diving the spouse effect by the own effect. Robust standard errors in parentheses, clustered at the couple level. Bootstrapped standard errors for scaled effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2.A.2: Descriptive Statistics by Gender and Age Differences

	Female				Male			
	Younger		Older		Younger		Older	
	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)	Mean (7)	SD (8)
Earnings age 55-57	48,213	23,886	50,393	24,445	74,823	45,510	72,220	39,940
College education	0.22	0.41	0.28	0.45	0.25	0.43	0.21	0.41
Retired by age 57	0.11	0.32	0.12	0.32	0.08	0.27	0.07	0.25
Copenhagen region	0.26	0.44	0.30	0.46	0.30	0.46	0.26	0.44
Numer of Observations	213,862		69,661		65,431		240,733	

Notes: This table reports means and standard deviations of relevant variables for all reference individuals in the sample of analysis used for the age-based empirical design. Column (1) corresponds to females who are younger than their partner, whereas column (2) corresponds to females that are older than their partners. Columns (3) and (4) do the same for males. Labor market earnings are computed as the average between ages 55 and 57. Retirement, education, and whether they live in the capital region, are measured at age 57.

Table 2.A.3: Placebo Test with Fake Spouses for the Effect of Reaching Pension Eligibility Age

	Retirement	Claiming	Earnings
Reference Individual	0.2034***	0.3496***	-8,642***
	(0.001)	(0.001)	(69.431)
Spouse	-0.001	-0.002	-32
	(0.001)	(0.001)	(78.79)
Scaled Effect	-0.00415	-0.00484	0.00370
	(0.0079)	(0.0035)	(0.017)
N. of clusters	367,585	367,585	367,585
Observations	2,206,044	2,206,044	2,206,044

Notes: This table reports the results of replicating the analysis over a placebo sample where the reference individuals are the same as in the main analysis, but they are matched to fake spouses of similar age. The placebo test finds no evidence of joint retirement, as should be expected if the empirical strategy is valid. Fake spouses cannot affect each other's retirement behavior, and the effect coming from the correlation between their ages is controlled for by the empirical design. See the notes of Table 2.2 for a detailed explanation of the content of the table. Robust standard errors in parentheses, clustered at the couple level. Bootstrapped standard errors for scaled effects. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.1$

2.B Appendix: Reform Discontinuity Design

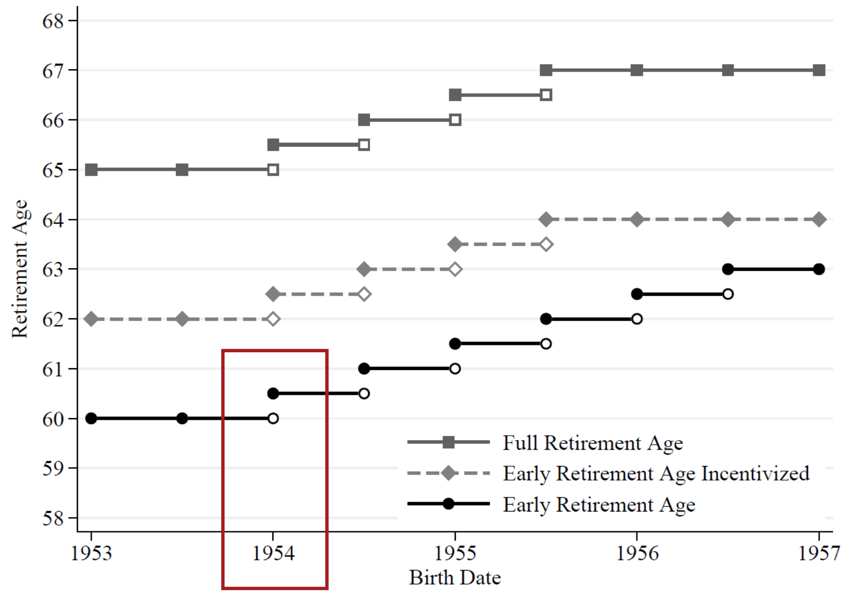


Figure 2.B.1: Graphical Depiction of the 2011 Reform

Notes: This figure depicts the 2011 reform that increased retirement ages in 6-month steps contingent on birth date. Cohorts born before January 1, 1954 were unaffected by the reform. Cohorts born between January 1, 1954 and July 1, 1954 experienced an increase of 6 months in their pension eligibility ages. Their early pension eligibility age increased from 60 to $60\frac{1}{2}$, their incentivized early pension eligibility age increased from 62 to $62\frac{1}{2}$ and their full retirement pension increased from 65 to $65\frac{1}{2}$. The red square marks the discontinuity that we exploit in our reform-based research design, where we study the effect of increasing pension eligibility ages. Later cohorts experienced larger increases.

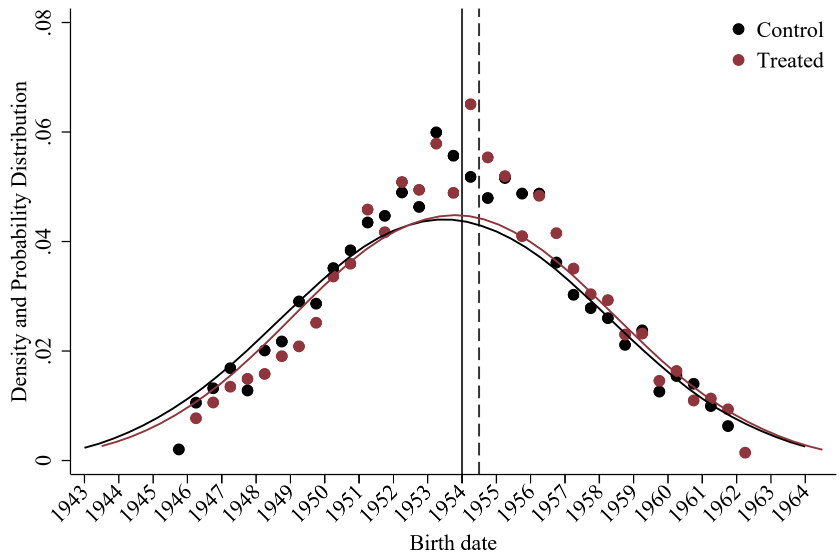


Figure 2.B.2: Birth Date of Spouses by Treatment Group for the Reform Sample

Notes: This graph plots the kernel density function and the probability distribution of the birth date of spouses in the treatment and control groups. Spouses in the treatment group are slightly younger than those in the control group, as a consequence of defining the treatment and control groups based on whether the reference individual was born, respectively, after or before January 1, 1954. Spouses that are born between January 1 and June 30, 1954 (indicated by the solid and dashed vertical lines) are directly impacted by the reform in 2014. We can see from the probability distribution, which is depicted by the dots, that spouses in the treatment group are 1.3 percentage points more likely to be born within those dates than the spouses from the control group (6.5% against 5.2%). Spouses born after June 30, 1954 (dashed vertical line) are impacted by the reform only after 2014. Spouses in the treatment group are 2.2 percentage points more likely to be born after June 30, 1954 (44.3% against 42.1%).

Table 2.B.1: Heterogeneity in the Effect of Increasing Pension Eligibility Age

Reference Individual	Young	Old	Female	Male
Spouse	Old	Young	Male	Female
	(1)	(2)	(3)	(4)
A. Retirement				
Reference Individual	-0.259***	-0.118***	-0.258***	-0.118***
	(0.011)	(0.0098)	(0.011)	(0.0097)
Spouse	-0.0315**	-0.00354	-0.0232**	-0.0107
	(0.013)	(0.0069)	(0.011)	(0.0090)
Scaled Effect	0.122**	0.0301	0.0898**	0.0907
	(0.049)	(0.058)	(0.044)	(0.076)
B. Claiming				
Reference Individual	-0.327***	-0.197***	-0.338***	-0.184***
	(0.0097)	(0.0086)	(0.0097)	(0.0084)
Spouse	-0.0219*	-0.000220	-0.0185*	-0.00297
	(0.012)	(0.00066)	(0.011)	(0.0066)
Scaled Effect	0.0669*	0.00112	0.0546*	0.0162
	(0.038)	(0.0034)	(0.031)	(0.036)
C. Earnings				
Reference Individual	10,885***	5,381***	10,678***	5,381***
	(667.6)	(695.5)	(612.7)	(743.1)
Spouse	1,366	-29.97	928.1	381
	(905.9)	(546.9)	(872.6)	(567.6)
Scaled Effect	0.126	-0.0056	0.0869	0.0707
	(0.083)	(0.10)	(0.081)	(0.11)
N. of clusters	5,385	5,161	5,541	5,008
Observations	37,541	35,854	38,542	34,853

Notes: This table reports the effect of the 2011 reform, which increased the pension eligibility age, distinguishing heterogeneous responses by age composition and gender of the couple. Each column contains results for a different subsample. Each panel reports results for a different outcome variable. Within each panel, the first row reports the effect on the individuals affected by the reform and the second row reports the spillover effect on their spouses, which are both estimated in equation (2.5). The third row reports the scaled effect (the LATE) resulting from the 2SLS model estimated in equation (2.6). F-tests for the strength of the instruments are all well above 10. Robust standard errors in parentheses, clustered at the couple level. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.1$

Table 2.B.2: Robustness to Alternative Specifications for the Effect of Increasing Pension Eligibility Age

	Retirement	Claiming	Earnings
A. Baseline	0.0902**	0.0418*	0.0847
	(0.038)	(0.024)	(0.065)
B. Reweight Spouses Birth	0.0966**	0.0324	0.0638
	(0.040)	(0.026)	(0.068)
C. Donut Affected Spouses	0.0954**	0.0380	0.0895
	(0.039)	(0.025)	(0.067)
D. Without Anticipation	0.0858***	0.0432**	0.0871
	(0.032)	(0.021)	(0.053)
E. Smaller Bandwidth	0.101**	0.0494*	0.0932
	(0.042)	(0.027)	(0.073)
F. Larger Bandwidth	0.0590*	0.0320	0.0847
	(0.035)	(0.022)	(0.065)
G. Not Balancing	0.0932**	0.0462*	0.0664
	(0.039)	(0.024)	(0.069)
H. Adding Controls	0.0901**	0.0415*	0.0822
	(0.038)	(0.024)	(0.065)
I. No VERP restriction	0.104***	0.0317	0.107**
	(0.034)	(0.021)	(0.055)

Notes: This table reports the scaled effect estimates (2SLS estimates) from replicating our main analysis using different sample definitions and different specifications of the estimation model (equation 2.6). Row A reproduces results from our baseline specification, which correspond to those reported in Table 2.6. Row B reweighs the observations so that the treated and control group have the same distribution of spouses' birth date. Row C excludes spouses born in the first half of 1954. Row D does not estimate the anticipation period separately. Row E reduces the bandwidth by 2 weeks. Row F extends the bandwidth by 2 weeks. Row G does not balance the sample. Row H controls for region and education of reference individuals and their spouses. Row I extends the sample to include individuals who did not contribute to the VERP program between ages 50-59. Robust standard errors in parentheses, clustered at the couple level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2.B.3: The Effect of Increasing Pension Eligibility Age. Replication Over Sample of Spouses At Least 3 Months Older

	Retirement	Claiming	Earnings
Reference Individual	-0.258***	-0.326***	10,718***
	(0.011)	(0.0099)	(680.8)
Spouse	-0.0280**	-0.0179	1,480
	(0.013)	(0.013)	(938.1)
Scaled Effect	0.109**	0.0550	0.138
	(0.051)	(0.039)	(0.087)
F-test instr.	523.4	1078.6	247.8
N of clusters	5,096	5,096	5,096
Observations	35,511	35,511	35,511

Notes: This table replicates the analysis for a subsample where spouses are at least 3 months older than their partners. This ensures that all spouses are born before January 1, 1954, and therefore are totally unaffected by the 2011 reform. This rules out the possibility that the spillover effect to spouses is driven by spouses in the treated and control groups being diferentially impacted by the reform. See Table 2.6 for notes on the construction of this table.

Bibliography

An, M. Y., B. J. Christensen, and N. D. Gupta (2004, oct). Multivariate mixed proportional hazard modelling of the joint retirement of married couples. *Journal of Applied Econometrics 19*(6), 687–704.

Andrews, D. W. and M. Buchinsky (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica 68*(1), 23–51.

Atalay, K., G. F. Barrett, and P. Siminski (2019, jul). Pension incentives and the joint retirement of couples: Evidence from two natural experiments. *Journal of Population Economics 32*(3), 735–767.

Baker, M. (2002). The retirement behavior of married couples: Evidence from the spouse's allowance. *The Journal of Human Resources 37*(1), 1–34.

Banks, J., R. Blundell, and M. C. Rivas (2010). The dynamics of retirement behavior in couples: Reduced-form evidence from England and the US . *University College London, mimeo.*

Behaghel, L. and D. M. Blau (2012). Framing social security reform: Behavioral responses to changes in the full retirement age. *American Economic Journal: Economic Policy 4*(4), 41–67.

Bingley, P. and G. Lanot (2007). Public pension programmes and the retirement of married couples in Denmark. *Journal of Public Economics 91*(10), 1878–1901.

Blau, D. M. (1998). Labor force dynamics of older married couples. *Journal of Labor Economics 16*(3), 595–629.

Bloemen, H., S. Hochguertel, and J. Zweerink (2019). The effect of incentive-induced retirement on spousal retirement rates: Evidence from a natural experiment. *Economic Inquiry 57*(2), 910–930.

Browning, M., F. Bourguignon, P.-A. Chiappori, and V. Lechene (1994). Income and outcomes: a structural model of intrahousehold allocation. *Journal of Political Economy 102*(6), 1067–1096.

Browning, M. and P.-A. Chiappori (1998). Efficient intra-household allocations: a general characterization and empirical tests. *Econometrica*, 1241–1278.

Browning, M., O. Donni, and M. Gørtz (2020, 10). Do you have time to take a walk together? Private and joint time within the household. *The Economic Journal*. ueaa118.

Casanova, M. (2010). Happy together: A structural model of couples' joint retirement choices. *Working Paper, Department of Economics, University of California* (5), 1–53.

Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of Political Economy 100*(3), 437–467.

Coile, C. (2004). Retirement incentives and couples' retirement decisions. *Topics in Economic Analysis and Policy 4*(1), 441–470.

Coile, C. and J. Gruber (2007). Future social security entitlements and the retirement decision. *The Review of Economics and Statistics 89*(2), 234–246.

Cribb, J., C. Emmerson, and G. Tetlow (2016). Signals matter? Large retirement responses to limited financial incentives. *Labour Economics 42*, 203–212.

Daly, M. and F. Groes (2017). Who takes the child to the doctor? Mom, pretty much all of the time. *Applied Economics Letters 24*(17), 1267–1276.

Deshpande, M., I. Fadlon, and C. Gray (2020). How sticky is retirement behavior in the U.S.? Responses to changes in the full retirement age. *NBER Working Paper No. w27190*.

Donni, O. and P.-A. Chiappori (2011). Nonunitary models of household behavior: a survey of the literature. *Household Economic Behaviors*, 1–40.

Fadlon, I., S. P. Ramnath, and P. K. Tong (2019). Market inefficiency and household labor supply: Evidence from social security's survivors benefits. *NBER Working Paper No. w25586*.

García-Miralles, E. and J. M. Leganza (2020). Public pensions and private savings. *Working Paper 06/21. The Center for Economic Behavior and Inequality (CEBI). University of Copenhagen.*.

Geyer, J. and C. Welteke (2019). Closing Routes to Retirement for Women: How Do They Respond? *Journal of Human Resources*.

Gørtz, M., S. Sander, and A. Sevilla (2020). Does the child penalty strike twice?

Gustman, A. L. and T. L. Steinmeier (2000). Retirement in dual-career families: A structural model. *Journal of Labor Economics 18*(3), 503–545.

Gustman, A. L. and T. L. Steinmeier (2004). Social security, pensions and retirement behaviour within the family. *Journal of Applied Econometrics 19*(6), 723–737.

Haller, A. (2019). Welfare effects of pension reforms.

Honoré, B. E. and Á. de Paula (2018). A new model for interdependent durations. *Quantitative Economics 9*(3), 1299–1333.

Honoré, B. E., T. H. Jørgensen, and Á. Paula (2020). The Informativeness of Estimation Moments. *Journal of Applied Econometrics*.

Hospido, L. and G. Zamarro (2014). Retirement patterns of couples in Europe. *IZA Journal of European Labor Studies 3*(1), 12.

Hurd, M. D. (1990). The joint retirement decision of husbands and wives. In *Issues in the Economics of Aging*, pp. 231–258. University of Chicago Press, 1990.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016). Tax reforms and intertemporal shifting of wage income: Evidence from Danish monthly payroll records. *American Economic Journal: Economic Policy 8*(3), 233–57.

Kruse, H. (2020). Joint retirement in couples: Evidence of complementarity in leisure. *The Scandinavian Journal of Economics*.

Lalive, R. and P. Parrotta (2017, jun). How does pension eligibility affect labor supply in couples? *Labour Economics 46*, 177–188.

Lassen, A. S. (2020). Gender norms and specialization in household production: Evidence from a danish parental leave reform.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record 82*, S2–S18.

Manoli, D. S. and A. Weber (2016). The effects of the early retirement age on retirement decisions. *NBER Working Paper No. w22561*.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics 93*(11-12), 1224–1233.

Michaud, P.-C. and F. Vermeulen (2011). A collective labor supply model with complementarities in leisure: Identification and estimation by means of panel data. *Labour Economics 18*(2), 159–167.

Nakazawa, N. (2019). The effects of increasing the eligibility age for public pension on individual labor supply: Evidence from Japan. *Available at SSRN 3438100*.

Nielsen, N. F. (2019, may). Sick of retirement? *Journal of Health Economics 65*, 133–152.

OECD (2015). *Pensions at a Glance: OECD and G20 Indicators*. OECD.

OECD (2017). *Pensions at a Glance: OECD and G20 Indicators*. OECD.

OECD (2019). *Pensions at a Glance: OECD and G20 Indicators*. OECD.

Selin, H. (2017, jun). What happens to the husband's retirement decision when the wife's retirement incentives change? *International Tax and Public Finance 24*(3), 432–458.

Stancanelli, E. (2017, dec). Couples' retirement under individual pension design: A regression discontinuity study for France. *Labour Economics 49*, 14–26.

Staubli, S. and J. Zweimüller (2013). Does raising the early retirement age increase employment of older workers? *Journal of Public Economics 108*, 17–32.

Van der Klaauw, W. and K. I. Wolpin (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics 145*(1-2), 21–42.

Willén, A., K. Vaage, and J. Johnsen (2020). Interactions in public policies: spousal responses and program spillovers of welfare reforms. *NHH Dept. of Economics Discussion Paper* (20).

Chapter 3

Health Professional Shortage Areas and Physician Location Decisions

Abstract

To address geographic disparities in healthcare provision, the U.S. government designates primary care Health Professional Shortage Areas (HPSAs), and the Centers for Medicare and Medicaid Services (CMS) provide 10% bonus payments to physicians billing in these areas. We use administrative data from CMS and a matched difference-in-differences design to study the effects of shortage area designations on physician location decisions. We find that counties designated as HPSAs experience a 23% increase in the number of early-career primary care physicians. The increase is driven entirely by physicians who attended ranked medical schools. However, we find no evidence that physicians in later career stages relocate to shortage areas. Overall, our findings suggest that targeting incentive payments towards newer physicians may improve the effectiveness and cost-efficiency of policies aimed at addressing physician shortages.

3.1 Introduction

There exists wide regional variation in healthcare spending and utilization, as well as health outcomes across the United States (Skinner 2011). While the literature seeks to understand and debates the relative importance of supply side factors versus demand side factors in causing this phenomenon, a closely-related fact has captured the interest of researchers and policymakers alike: some areas have significantly fewer doctors per capita than other areas. Individuals living in these so-called “shortage areas” may face higher costs of obtaining medical treatment and may be less likely to seek preventive care.

To address potential problems associated with the presence of physician shortages, the U.S. government identifies areas in need and attempts to increase resources available to residents of these areas. A particularly prominent policy aims to improve access to primary care through financially incentivizing physicians to practice in areas deemed to have too few doctors. Specifically, the Health Resources and Services Administration works with state agencies to manage official designations of Health Professional Shortage Areas (HPSAs), and through the Centers for Medicare and Medicaid Services (CMS), physicians receive a 10 percent bonus payment on the Medicare services they bill in designated HPSAs.

In this paper, we ask whether Health Professional Shortage Area designations influence the location decisions of primary care physicians (PCPs). To answer this question, we study the effect of a county being designated as a HPSA on the stock of Medicare-billing primary care doctors practicing in that county. We first link together several sources of administrative data from CMS using unique physician identifiers to create a county-level panel dataset that contains information on physician counts (by doctor characteristics such as graduation date and medical school attended), as well as HPSA designation status. We then supplement these data, which capture the near-universe of physicians who bill Medicare Part B, with county-level information from the Area Health Resource File.[1] Using this panel

[1] Note that the vast majority of primary care physicians bill to Medicare; more than 90% of non-pediatric primary care physicians accept Medicare patients (Kaiser Foundation 2015).

dataset, which spans the years 2012 to 2017, we employ a matched difference-in-differences design to identify the causal effect of HPSA designations on the stock of Medicare-billing PCPs.

We use a matching strategy in order to overcome a significant challenge associated with studying the impact of shortage area designations. To identify causal effects, one needs a valid counterfactual for the evolution of PCP counts in HPSA counties. Yet designations are not random; they are in part directly due to declines in the number of physicians practicing in a county. Thus comparing a control group of all non-HPSA counties with a treatment group of HPSA counties is unlikely to be a credible approach. Our matching strategy, which uses variables defined over a baseline time period that capture information directly relevant for official shortage area designations, addresses this concern by selecting counties similar to HPSAs to serve as controls.

Specifically, to each county designated as a HPSA during our analysis time period, we match similar counties that are not designated as HPSAs. We then use a difference-in-differences framework to compare the stock of PCPs in HPSAs before and after the official designation with that of the matched control counties. Importantly, we exploit our data to analyze physician responses separately by career stage. Early-career physicians likely making initial location decisions after completing their residencies may face substantially lower costs of moving compared to later-career physicians, and the degree to which they respond may have particularly important consequences for evaluating the efficacy of the program, if physicians who locate in shortage areas tend to continue practicing there for the duration of their career. We also use information on medical school rankings to proxy for physician quality, and we assess whether physician responses differ along this dimension.

Our main result is that designated counties experience an increase in the number of early-career PCPs. The pattern of our dynamic difference-in-differences estimates suggests a relatively quick rise in the count of early-career physicians during the first two years of designation, which then stabilizes at a higher level. Our preferred estimate indicates that designated counties experience an average increase of approximately 0.114 physicians per 10,000 residents, which roughly amounts to 0.67 physicians per county and represents a 23% increase off of a modest baseline mean. We then show that the increase is entirely driven by an influx of early-career PCPs who attended ranked medical schools, perhaps reflecting the ability of the program to attract high-quality physicians to areas in need.

In contrast, we find no evidence of an increase in counts of later-career physicians, who are likely more settled and may face higher costs of relocating already-established practices. Our results are consistent with the notion that bonus payments for billing in HPSAs may be more attractive to newer physicians—who are likely already considering (re)location decisions as it relates to the timing of recently completed residencies or initial career trajectories.

Our findings have direct implications for policy. The 10 percent bonus payment attached to HPSAs is provided to all PCPs billing services to Medicare, but the majority of these are later-career doctors, who we find to be generally unresponsive. A more effective and cost-efficient way to increase physician counts in underserved areas may be to target a higher percentage bonus payment at the subset of physicians we find to be responsive. For instance, using a simple and stylized policy exercise, we show that a 20 percent bonus payment offered to PCPs who relocate to a HPSA in the first 10 years of their career may induce even more movement of early-career physicians than the current program while substantially reducing overall payments to inframarginal doctors who would practice in a HPSA under either regime.

This paper relates broadly to the large literature that studies physician responses to financial incentives, often analyzing how payment rates and prices impact provision of care (e.g., Ellis and McGuire 1986, McGuire and Pauly 1991, McGuire 2000, and Chandra et al. 2011) and physician labor supply more generally (e.g., Nicholson and Propper 2011).[2] We contribute to this literature by providing new evidence on how financial incentives impact a

[2]For additional work in the U.S. setting, see Hadley and Reschovsky (2006), Clemens and Gottlieb (2014), Alexander (2015), Johnson and Rehavi (2016), Clemens et al. (2018), and Gottlieb et al. (2020). For evidence from other countries, see Sørensen and Grytten (2003), Kantarevic et al. (2008), Devlin and Sarma (2008), Sarma et al. (2010), and Brekke et al. (2017).

key component of physician labor supply: practice location.

We thus relate most closely to other papers that investigate physician location decisions, especially in the context of physician shortages.[3] Despite the importance and policy-relevance of the topic, there is limited causal evidence informing the issues. In a review of research on shortage area programs, Bärnighausen and Bloom (2009) discuss several observational studies and conclude that, mostly due to selection effects, none allow for credible causal inference. More recently, a series of working papers develop models of physician location decisions, simulate the effects of various incentive policies designed to combat shortages, and find generally that physicians are not very responsive to financial and salary incentives (Zhou 2017, Falcettoni 2018, and Kulka and McWeeny 2019).[4] Of these papers, Kulka and McWeeny (2019) is the most similar to ours, as they complement their structural analysis with a reduced-form evaluation of state-level student loan forgiveness programs and find small positive effects. We contribute to this strand of the literature by offering causal evidence on the effectiveness of a large, nation-wide program designed to address shortage areas through direct monetary payments. Furthermore, in exploiting our data to study how responses vary by career stage, we are able to uncover evidence that early-career PCPs are more responsive to shortage area designations.

Finally, our findings connect to an important discussion in the literature on how payment policies influence the overall capacity of the healthcare system, particularly as it relates to the allocation of human capital to and within the health sector. Existing work shows that Medicare policy can increase investments in medical technology (Finkelstein 2007, Acemoglu and Finkelstein 2008, and Clemens and Gottlieb 2014) as well as physician on-the-job investments (Clemens et al. 2018), and other papers highlight an important role for financial incentives in shaping the decision to become a doctor (Chen et al. 2020 and Gottlieb

[3]More generally, papers have documented factors such as the location and type of medical training as influencing practice locations (e.g., Burfield et al. 1986 and Chen et al. 2010). Another related paper set in a different context is Huh (2018), who finds that Medicaid expansions can attract dentists to poorer areas.

[4]These papers advance earlier work that modeled physician location decisions in the U.S. (Hurley 1991 and Holmes 2005) and Canada (Bolduc et al. 1996).

et al. 2020).[5] In finding that the HPSA program brings physicians to designated counties, we present evidence of a government payment policy expanding access to healthcare in specific geographies and influencing the distribution of health-sector human capital across space.[6]

The rest of this paper is organized as follows. Section 3.2 provides an exposition of the policy environment. Section 3.3 describes the data sources and highlights how we construct our dataset. Section 3.4 lays out our matched difference-in-differences framework. Section 3.5 presents our results. Section 3.6 discusses policy implications. We conclude in Section 3.7.

3.2 Policy Environment

Overview of Health Professional Shortage Areas. The Health Resources and Services Administration (HRSA), which is an agency of the United States Department of Health and Human Services, strives to "improve health outcomes and address health disparities through access to quality services, a skilled health workforce, and innovative, high-value programs."[7] In order to bring federal resources to people in need, HRSA creates shortage designations. Health Professional Shortage Areas (HPSAs) are one type of shortage designation, and it is this particular type on which CMS bases their Medicare bonus payment program.[8] HPSA designations can be made for three disciplines (primary care, mental health, and dental health) at three different levels (geographic area, population group, and facilities). Because primary care physicians (PCPs) play such a central role in the provision of healthcare in

[5] Another set of related papers show that specialty choice may also be influenced by financial incentives (e.g., Sloan 1970, Bazzoli 1985, Hurley 1991, Nicholson and Souleles 2001, Nicholson 2002, Bhattacharya 2005, Gagné and Léger 2005, and Sivey et al. 2012).

[6] Our analysis thus also connects to the influential research concerned with assessing causes and implications of regional differences in healthcare utilization, expenditures, and physician practice styles (e.g. Fisher et al. 2003a, Fisher et al. 2003b, Sutherland et al. 2009, Gottlieb et al. 2010, Song et al. 2010, Zuckerman et al. 2010, Skinner 2011, Finkelstein et al. 2016, Molitor 2018, and Cutler et al. 2019).

[7] See their mission statement on the following website: `https://www.hrsa.gov/about/index.html`.

[8] Other types of shortage area designations maintained by HRSA include: Medically Underserved Areas (MUAs), Medically Underserved Populations (MUPs), and Governor's Designated Secretary Certified Shortage Areas for Rural Health Clinics.

the United States, and because the CMS Medicare incentive payment program that we study in this paper does not apply to population group or facility shortage designations, we restrict our attention to HPSAs designated for the primary care discipline at the geographic level. Unless otherwise specified, hereafter we use the more general terms, "HPSAs" and "designations," to refer to this specific type of shortage designation.

HPSA Designation Process. While HRSA manages and grants HPSA designations, the responsibility to identify potential shortage areas falls on state Primary Care Offices (PCOs), who generally submit applications on behalf of geographic areas in their state to HRSA. State PCOs do not all operate in the same manner. For instance, depending on the PCO, areas identified as potential HPSAs can be census tracts, minor civil divisions (e.g., townships), or entire counties. Nonetheless, once HRSA receives an application, they work with the applying PCO to gather objective data used to both determine HPSA eligibility status and to calculate a score intended to quantify the severity of the shortage.[9] The score is primarily determined by an area's population-to-provider ratio, but it also depends on the fraction of the population below the federal poverty line, an infant health index, and travel time to the nearest source of care outside of the proposed HPSA. While the actual score may be informative for programs beyond the scope of our paper, the Medicare bonus payments provided by CMS depend only on overall designation status, and they do not depend on the score-based severity of the shortage.

Medicare Bonus Payments from CMS. The Centers for Medicare and Medicaid Services provide 10 percent bonus payments on Medicare services furnished by physicians in primary care geographic HPSAs designated by December 31 of the previous year. Bonuses are paid quarterly and are generated automatically when physicians provide services in a CMS-maintained list of HPSA ZIP codes, which consists of ZIP codes that fall entirely within a designated HPSA (e.g., all ZIP codes completely contained in a county that is

[9]As a general benchmark, HRSA typically considers an area to have a shortage of providers if they have a population to provider ratio of 3,500:1 or more.

a designated HPSA). Physicians providing services in designated areas not on the CMS-maintained ZIP code list can still receive the HPSA bonus payment by appending a modifier to their claims; these physicians are responsible for determining the HPSA status of their area based on tools provided by HRSA. Due to the data availability discussed in Section 3.3 (and because CMS relies primarily on their own list of HPSA ZIP codes), we use as our source of variation designations that result in automatically-billed HPSA ZIP codes. The 10% bonus payment program produces the major incentive for locating in HPSAs and applies to all physicians in HPSAs; though for some groups of doctors, other related programs may interact with designations to create additional incentives.[10]

3.3 Data

To analyze the impact of HPSA designations on the location decisions of Medicare-billing PCPs, we draw on five main data sources to assemble a detailed, county-level, panel dataset. In this section, we provide an overview of the data sources, highlight our approach to creating the county panel, and discuss key variables for our analysis.

3.3.1 Data Sources and Creating the County Panel

To construct a county panel suitable for our analysis, we start by linking together three physician-level datasets developed by CMS. The first, *Medicare Provider Utilization and Payment Data: Physician and Other Supplier* (MPUP), contains detailed information on Medicare services provided by healthcare professionals at the physician-code-location

[10]A variety of smaller federal incentive programs aim to bring physician and non-physician healthcare providers to shortage areas. For example, loan forgiveness and scholarship programs through the National Health Service Corps (NHSC) and the NURSE Corps, Rural Health Clinic Programs through CMS, and the J-1 visa waiver program for foreign medical graduates may use HPSA criteria to determine eligibility in their contexts. Some primary care physicians may also participate in these programs and thus may face additional incentives above and beyond the bonus program. In addition, most states have some form of a loan forgiveness program for practicing in rural areas (Kulka and McWeeny 2019) which could potentially interact with HPSA designations. For more information on HPSA designations in general and additional related programs, see `https://bhw.hrsa.gov/shortage-designation/hpsas`.

level from 2012–2017.[11] It is based on CMS administrative claims data for Medicare Part B fee-for-service beneficiaries, and it represents the near-universe of Medicare billing physicians. Only Medicare-billing doctors who do not bill any HCPCS code at least 10 times in a given year are omitted from the data for that year. Of note, more than 90% of non-pediatric primary care physicians accept Medicare patients (Boccuti et al., 2015). We extract from this dataset the unique physician identification numbers, National Provider Identifiers (NPIs), of Medicare-billing doctors and information regarding their specialty. From annual disseminations of a second physician-level dataset, the *National Plan and Provider Enumeration System* (NPPES), we extract information on the primary practice location for the Medicare-billing physicians.[12] Linking these two datasets yields panel data for Medicare-billing physicians spanning the years 2012 to 2017, with information on physician specialty and practice location.

The third physician-level dataset we employ is the *Physician Compare* dataset, which CMS began publishing in 2014 for the use of patients who wish to gather information about doctors who accept Medicare. From these data we extract physician graduation dates and medical school attendance, which allows us to analyze doctor responses by career stage and quality of medical school (as proxied for by medical school rankings). The ability to incorporate this information in our analysis is important for policy. For example, the effectiveness of the program in alleviating concerns regarding the provision of medical care in the longer run may depend on the types of physicians ultimately induced to locate in shortage areas.

[11] Specifically, one observation in the dataset is defined by (1) a National Provider Identifier, the unique physician identification number, (2) a Healthcare Common Procedure Coding System (HCPCS) code, which are specific codes detailing the procedure undertaken by the physician, and (3) place of service.

[12] The MPUP does contain information on practice location; however, the variables contained in this dataset are not suitable for our analysis. Specifically, location variables in the MPUP data are updated to be the location of the physician in the subsequent calendar year. For example, the 2014 MPUP data contain billing information for physicians who billed Medicare in 2014, but the location variable captures locations at the end of the 2015 calendar year. It is for this reason that we use the NPPES data to accurately define physician location for the calendar years for which we have billing information. We define location as a physician's primary practice location in December of the year of observation.

The main drawback of the Physician Compare dataset lies in the fact that it is a snapshot in time of currently-billing physicians. While we make use of all available archived data from 2014 onward, we do not have a snapshot of the Medicare-billing physicians before the initial publication of the data in 2014. For the most part, this drawback is rather harmless, as the information pulled from Physician Compare (i.e. graduation year and medical school) is time-invariant, and most doctors in our panel of Medicare-billing physicians appear in all waves of the data. However, after we link the Physician Compare data to our panel data, graduation year and medical school are mechanically missing for physicians that practice and bill to Medicare *only* in 2012 or 2013 (because those doctors are never observed in a year for which Physician Compare exists).[13] While it is perhaps more likely that the physicians who are observed only in 2012 and/or 2013 are late-career physicians who have retired by 2014, our leading analysis does not count these physicians as belonging to any career stage (and it also does not count them as having attended ranked or unranked medical schools). We show that the rate of missing data does not differ significantly between the treatment group and the control group before or after designation in Appendix Figure 3.A.3.

After linking together the three physician-level data sources, we aggregate the data up to the county level. That is, we create a county-level dataset with counts of primary care Medicare-billing physicians spanning the years 2012 to 2017.[14] Finally, into our newly-constructed panel we merge data derived from two more sources. First, for information regarding HPSA status, we use the official, CMS-maintained list of ZIP codes that define automatically billed HPSAs. We aggregate this data up to the county level by simply counting the number of HPSA ZIP codes in a county. Second, for more information on county characteristics, we pull variables from the *Area Health Resources File* (AHRF), which contains a wide range of county-level, health-related variables derived from the American

[13]There are 16,873 (7.23%) primary care physicians who only appear in the data in 2012 and 2013, overall, and 2,563 (6.63%) in our analysis counties.

[14]We define a doctor as a primary care physician if her specialty is any of the following: "family practice," "general practice," "internal medicine," "geriatric medicine," or "pediatric medicine."

Medical Association Masterfile and county-level demographic and economic variables derived from the American Community Survey. Linking together all of the data sources, we create a county panel containing information on population demographics, economic conditions, HPSA designations, and the stock of Medicare-billing primary care physicians.

3.3.2 Key Variables

The outcome variables of interest for our analysis are per-capita counts of Medicare-billing primary care physicians. We analyze the evolution of the total count of these doctors in counties across time, but we also break down the stock of physicians into counts by career stage and by quality of medical school. In any given year, we define early-career PCPs, who may have higher elasticities governing their labor supply (and practice location) decisions, as those who graduated from medical school 5 to 10 years prior. Our definition of early-career physicians intends to capture those likely making initial location decisions for their practice after completing their residencies. Our choice of 5 years after graduating is also driven by the data: the vast majority of physicians are not assigned an NPI until about 5 years after finishing medical school.[15] We then define later-career PCPs as those who graduated more than 10 years ago.

We also analyze physician counts by quality of medical school. HRSA designates shortage areas with the goal of bringing resources to areas in need. From a policy perspective, the types of physicians the program brings in may have important consequences. We therefore break down counts of physicians along this dimension. Specifically, we study counts of PCPs who attended ranked medical schools separately from counts of PCPs who attended unranked medical schools. To define the relevant variables, we use the 2018 rankings of med-

[15]In any given year, the data contain a very small number of physicians who report having graduated less than 5 years earlier. The counts of physicians by medical school cohort do not approach the typical cohort size until 5 years after graduation. This is because physicians typically spend their years immediately after graduation completing their residency and likely do not yet have an NPI. To maintain a consistent interpretation of our definition of early-career physicians, we exclude from our count of early-career PCPs the handful of physicians in the data who are not likely to have completed their residency by defining early-career PCPs as those graduating 5 to 10 years earlier.

ical schools for primary care from the U.S. News & World Report, and we consider a medical school to be ranked if it is any one of the 95 schools receiving an official ranking.[16]

We use several additional variables in our matched difference-in-differences design. In particular, we define our treatment variables based on whether or not a county contains at least one automatically-billed designated HPSA ZIP code.[17] We also use county-level variables from the AHRF indicating the total number of active physicians per capita and the percent of the population below the federal poverty line to carry out our matching procedure, and we employ three more variables from the AHRF specifying the population, unemployment rate, and median household income of counties as controls. In Section 3.4, we describe specifically how these variables enter our design.

3.4 Empirical Strategy

Our goal is to estimate the causal effect of HPSA designations on physician location decisions. An ideal experiment would randomly assign HPSA designations to some counties and track the counts of physicians in these counties compared to a control group of non-designated counties. A potentially-naive difference-in-differences framework that aims to approximate this ideal would involve the comparison of designated counties (i.e., the treatment group), in which 10% bonus payments are made to Medicare-billing PCPs, to counties that are not designated (i.e., the control group), in which there are no 10% bonus payments for Medicare-billing PCPs. Such a comparison is not without problems, as counties designated as HPSAs are likely very different in observable and unobservable ways than counties that are not designated.

[16]About 36% of PCPs in the sample report a medical school of "Other," which we classify as unranked. Some PCPs reporting "Other" may have attended medical school outside of the U.S.

[17]While some counties are only "partially" HPSA-designated, meaning only some of its zip codes are on the CMS list of automatically billed HPSAs, the majority of HPSA-designated counties in our sample are fully designated. There are 79 (36.4%) partially designated counties in our analysis data. Of those, 20% are at least 50% designated. We assess the robustness of our results to the exclusion of partially designated counties in Section 3.5.2.

Indeed, Figure 3.1 illustrates exactly this concern. The solid line depicts the average count of PCPs in HPSAs, where time on the x-axis is relative to designation year. The stock of physicians in HPSA counties tends to fall leading up to the designation year, which is not unexpected. In contrast, the dotted line depicts the average count of PCPs for the potential control group that consists of all other counties. Relative time for this comparison group is defined by matching to each HPSA all other counties, and then assigning a placebo designation year to the comparison counties equal to the actual designation year for the HPSA county to which they are matched. The stock of physicians in all other counties is not falling in the years before placebo designation, which would raise concerns about the validity of a straightforward difference-in-differences estimator.

For these reasons, we use a matched difference-in-differences approach to select a control group of non-designated counties that are more similar to HPSAs. In Section 3.4.1, we detail our procedure for selecting the control group and discuss our analysis sample. In Section 3.4.2, we describe the specifics of how we implement our matched difference-in-differences design.

3.4.1 Matched County Design

Matching Procedure. To select our control group, we borrow a matching procedure from Deryugina et al. (2018) to identify counties that are similar to our treatment group comprised of HPSAs.[18] We match to each treated county three control counties, and we assign the matched controls a placebo designation year equal to the actual designation year of their corresponding treated county.

To select the three control counties for each treated county, we use as our set of matching variables $\mathbf{X}_{ct}$ three variables defined at a baseline: number of active physicians per capita, annual percentage change in active physicians per capita, and percent of the

[18] Deryugina et al. (2018) study the long-run effects of Hurricane Katrina; we broadly base our matching procedure off of the one they employ, which selects cities similar to New Orleans.

population below the federal poverty line. We use these variables (pulled from the AHRF) from 2010 and 2011, which corresponds to two or three years before any of the earliest designations that we study. HRSA uses both the stock of physicians and the poverty rate to determine the score of proposed HPSAs, and designations are largely due to declines in physician counts; therefore, we view these variables as a reasonable and natural benchmark set on which to match.

For each treated county, we use our matching variables to compute a measure of "closeness" to each potential control county, where the pool of potential controls consists of the counties that are never designated as HPSAs in our sample period. To compute the closeness between a treatment county c^* and a control county c, we sum the squared difference between counties of each variable $x_{ct} \in \mathbf{X}_{ct}$ (normalized by that variable's standard deviation in the pool of counties σ_{x_t}) across both years in the baseline period 2010–2011.[19] That is,

$$\text{Closeness}(c^*, c) = \sum_{t=2010}^{2011} \sum_{x_{ct} \in \mathbf{X}_{ct}} \left(\frac{x_{ct} - x_{c^*,t}}{\sigma_{x_t}} \right)^2 . \tag{3.1}$$

In addition to the variables included in the closeness measure, matching on region is important given that the existing literature has indicated that geography has an influence on physician residential choices (Burfield et al., 1986; Chen et al., 2010). For this reason, we stipulate that a treatment county can only be matched to control counties that are in its geographic region.[20] The three counties from the pool of potential controls with the smallest value of this match measure for a given treatment county are included in the control sample with a placebo designation year equal to the actual designation year of the treatment county to whom they are matched.

We probe the robustness of our results to changing different aspects of the matching

[19] Note that while the other match variables are defined for both 2010 and 2011, the percentage change in number of physicians is only calculated for the annual change from 2010 to 2011 since these are our designated baseline years. Thus, the closeness measure includes two values for the stock of active physicians, two values for the poverty rate, and one value for the percentage change in active physicians.

[20] We define four distinct regions roughly corresponding to South, Northeast, Midwest, and West.

procedure in Section 3.5.2. Specifically, we vary the combination of baseline variables used to construct the match, and we vary the number of control counties matched to each treatment county.

Analysis Sample. The treatment group consists of the 217 counties that we see become designated between 2013 and 2017. The matching method described above generates a control group from the sample of counties that are never designated as HPSAs between 2012 and 2017. Three counties are matched to each treatment county to serve as controls, and counties are allowed to be matched to more than one treatment county; the resulting analysis sample thus includes 651 control counties, 470 of which are unique.[21]

Table 3.1 presents summary statistics for descriptive variables, for the treatment and control groups separately. The statistics come from the year preceding (actual or placebo) designation. The table shows that HPSAs generally look similar to control counties in terms of descriptive observables, although they are less populous and have slightly fewer physicians. Figure 3.1 makes it clear that the matched sample improves upon the non-matched sample in terms of assessing the validity of a difference-in-differences estimator through examination of parallel pre-trends. The dashed line plots the average counts of PCPs in our control group constructed using the matching procedure. The group experiences a decline in the stock of PCPs before placebo designation year similar to that in HPSAs, which allows us to more confidently use the evolution of PCP counts in the control group as a counterfactual for the evolution of PCP counts in the treatment group.

3.4.2 Implementation

We use the matching procedure described above to construct a suitable control group for counties within-whom an automatically-billed, primary care geographic HPSA is designated. To then analyze the effect of designations, we use a standard difference-in-differences

[21] Our panel is unbalanced due to the fact that the number of lead and lag years we see for a county depends on the year it was treated. By design, we exclude those counties that are always designated and study only those designated counties for which we see the year before and year of designation.

framework. Specifically, to document the dynamic impacts, we estimate the following equation:

$$y_{ct} = \alpha + \beta treat_c + \sum_{\tau \neq -1} \gamma_\tau I_\tau + \sum_{\tau \neq -1} \delta_\tau treat_c \times I_\tau + Z_{ct}\theta + \varepsilon_{ct}, \tag{3.2}$$

where y_{ct} is an outcome for county c in year t (e.g., the total number of Medicare-billing PCPs per 10,000 county residents), $treat_c$ is an indicator that equals one for counties receiving a designation over our sample period, the I_τ's are indicators for years relative to (actual or placebo) designation, Z_{ct} is a vector of controls, and the δ_τ's are the parameters of interest, which capture the average difference in y between the treatment and control groups relative to the omitted year.[22]

The identifying assumption asserts that, in the absence of HPSA designations, the stock of Medicare-billing PCPs in treated counties would have evolved in parallel with that in control counties. Analyzing the estimated δ_τ's from equation (3.2) provides an assessment on the validity of the design; specifically, we test whether the δ_τ's for $\tau < 0$ are different from zero, which would indicate the presence of pre-trends and might raise concerns regarding our difference-in-differences approach. Encouragingly, we consistently find no evidence of pre-trends that might invalidate the design.

Estimating the fully dynamic specification permits an evaluation of the key parallel trends assumption, but it also shows how the stock of doctors evolves over time; that is, results from estimating equation (3.2) shed light on how immediate or delayed, as well as how persistent or temporary, any physician responses to designations might be. After assessing the dynamic impact of HPSA designations, to better quantify the magnitudes of the mean

[22]Based on our data, $\tau \in \{-5, -4, \ldots, 4\}$ because the earliest year we can observe a change from not designated to designated is 2013 and our data goes through 2017; however, we pool together observations three or more years away from designation due to low observation counts.

treatment effect, we estimate the usual difference-in-differences estimating equation:

$$y_{ct} = \alpha + \beta treat_c + \gamma post_{ct} + \delta(treat_c \times post_{ct}) + Z_{ct}\theta + \varepsilon_{ct}, \tag{3.3}$$

where $post_{ct}$ is an indicator that equals one if for county c year t is a post-designation (or post-placebo-designation) year and δ is the parameter of interest.

Finally, while estimating equation (3.3) pools all pre-period years together and all post-period years together in order to quantify the overall effect, we employ one related additional specification. Guided by the graphical analysis of the dynamic impact, we split the post-designation period into two: a short-run period and a medium-run period. Specifically, we estimate

$$\begin{aligned} y_{ct} = \alpha + \beta treat_c + \gamma^{SR} post_{ct}^{SR} + \gamma^{MR} post_{ct}^{MR} \\ + \delta^{SR}(treat_c \times post_{ct}^{SR}) + \delta^{MR}(treat_c \times post_{ct}^{MR}) + Z_{ct}\theta + \varepsilon_{ct}, \end{aligned} \tag{3.4}$$

where $post_{ct}^{SR}$ is a (post-period short-run) indicator that equals one if for county c year t is in the year of the designation, and $post_{ct}^{MR}$ is a (post-period medium-run) indicator that equals one if for county c year t is after the immediate year of designation. Estimating equation (3.4) allows us to split up the post period and quantify short-run and medium-run effects, captured by δ^{SR} and δ^{MR} respectively. We often highlight the medium run estimates, which capture the impact on counts of doctors practicing in a county after allowing for the stock to evolve over a brief transition period.

3.5 Results

In this section, we first discuss our main results. We then discuss various robustness and specification checks. In general, we lead our analysis with graphical representations of dynamic effects before quantifying average magnitudes. In our leading regression specifications, all outcome variables are normalized per 10,000 population at baseline and winsorized at the 95th percentile, and we include county-level controls for household income, population, and the unemployment rate.[23]

3.5.1 Main Results

Figure 3.2 presents the results of estimating equation (3.2) for early-career and later-career PCPs.[24] The estimates for each parameter δ_τ are plotted along with 95% confidence intervals. These point estimates allow us to assess the validity of the identifying assumption and examine dynamic impacts.

The left-hand-side graph presents estimates of the impact of HPSA designation on counts of early-career doctors. The point estimates for δ_τ where $\tau < 0$ are not statistically different from zero and do not appear to be trending in any direction before the year of designation, which lends support to the parallel trends assumption. After designation, we see a relatively quick rise in the stock of these physicians practicing in HPSAS relative to non-HPSAs. The point estimate in year 0 is slightly elevated, whereas each of the point estimates on the indicators for the later post periods are positive and very similar to one another. The pattern of the dynamic estimates is consistent with a brief transition period over which the stock of doctors increases in response to the reform before stabilizing at the new level; this pattern also motivates a particular focus on the medium run estimates, which will quantify the effect of the policy on the stock of doctors after this brief transition period. Results from estimating equations (3.3) and (3.4) to quantify magnitudes are reported in Table 3.2. Column (1) summarizes the responses of early-career doctors. Panel A shows a statistically significant average medium-run increase of 0.114 early-career doctors per 10,000

[23]We measure baseline population in 2011. We include as controls indicators for $5,000 average household-income bins, current population, current population squared, and the unemployment rate.

[24]The corresponding graphs of raw means for these outcomes can be found in Appendix Figure 3.A.1. As defined in Section 3.3.2, early-career PCPs are those who graduated 5 to 10 years ago.

(s.e. 0.0570). This estimate corresponds to an increase of about 23% when compared to the baseline mean of 0.49 in the period before designation, and given that the average population of a treated county in our sample is around 59,000, it translates to approximately 0.67 more doctors per county on average. Panel B reports the average treatment effect for the entire post period, which includes the transition year as seen in the dynamics, thus resulting in a slightly smaller point estimate.

In contrast, the right-hand-side graph of Figure 3.2 shows no evidence of responses from later-career physicians. None of the dynamic point estimates are statistically distinguishable from zero, and the graph shows no discernible pattern or trend. Column (2) of Table 3.2 presents estimates for later-career PCPs; the magnitudes of the point estimates are comparatively smaller than those for early-career physicians, and the baseline mean is larger. At face value, the standard pooled difference-in-differences estimate for this outcome would represent a 0.13% increase in later-career doctor counts.

These results are consistent with PCPs in later career stages facing higher barriers to relocating. The cost of leaving behind a business that has already been established may be high, especially when considered with any implicit costs of moving to a potentially less desirable area. PCPs at the beginning of their career, however, might have fewer professional ties binding them to a given area, particularly when making initial location decisions after completing residencies.

Given the responsiveness of early-career doctors to HPSA designation, one may wonder which types of physicians are most likely to be induced to practice in a HPSA—in particular, whether they tend to be of higher or lower quality. Successfully attracting doctors to HPSAs that are young and high quality may increase both the quantity and quality of care in medically underserved areas. To proxy for physician quality, we use medical school rankings, and we analyze separate counts of early-career PCPs by whether the doctors attended a medical school that is included in the 2018 U.S. News Primary Care medical school rankings.

The dynamic effects on the stock of early-career doctors, split up by ranked and unranked medical schools, are presented in Figure 3.3, with corresponding graphs of means in Appendix Figure 3.A.2. First, we note the impacts in pre-designation years (on both counts of ranked and unranked doctors) are statistically indistinguishable from zero and do not exhibit any concerning trend. Next, we can see from comparing the left-hand-side graph and the right-hand-side graph that the entire post-designation increase in early-career doctors is driven by those who attended ranked medical schools. The dynamics for ranked physicians point to the same brief transition period followed by a period of stability, whereas the dynamics for unranked physicians reveal a lack of responses over the entire period. Corresponding point estimates are presented in Table 3.3; the estimates for early-career ranked doctors resemble those for the total number of early-career doctors, and are more precisely estimated. The medium run estimate indicates that treated counties gain 0.100 early-career, ranked PCPs per 10,000 population on average following HPSA designation (column (1) of Panel A), which corresponds to about 0.59 doctors in the average treated county, a 40% increase off of a small baseline mean. Mean treatment effects for early-career unranked physicians are much smaller and indistinguishable from zero (column (2)). Unfortunately, we lack the data to further investigate underlying mechanisms that could explain this dichotomy. Among other potential explanations, it could be that information about HPSAs is more widely disseminated at ranked schools, that students from these schools graduate with more debt, or that these doctors are more motivated to alleviate geographic shortages in care.

Lastly, to provide a gauge for the overall impact of designations, we present estimates on the per capita stock of all Medicare-billing PCPs. Figure 3.4 shows no evidence that designations have an impact on total PCP counts. This is not surprising, as the majority of PCPs are later-career PCPs, whom we have found to be unresponsive to HPSA status. We quantify corresponding magnitudes in Table 3.4. Columns (2) and (3) report separate estimates for the total stocks of ranked and unranked PCPs, both of which are statistically indistinguishable from zero.

3.5.2 Robustness and Specification Checks

We assess the robustness of our results along several dimensions. For simplicity, we focus on treatment effects from estimating equation (3.3) and medium run effects from estimating equation (3.4), for each of our main outcome variables: early-career PCPs; early-career PCPs from ranked schools; early-career PCPs from unranked schools; and later-career PCPs.

First, we probe the sensitivity of our results to various regression specifications. Table 3.5 displays results for the medium run effects, and Table 3.6 displays results for the mean treatment effects over all post-designation years. Each table is constructed as follows. Row A reproduces the baseline estimates. Rows B through D vary the approach to censoring the data for outliers. Rows E and F assess the sensitivity to inclusion of control variables. Overall, across both tables, we see that our results are not too sensitive to the choice of winsorization; point estimates are similar if we winsorize more stringently, winsorize less stringently, or do not winsorize at all, though we tend to experience precision gains when winsorizing more of the data. Further, results appear robust to both omitting all of the control variables as well as adding additional controls (year and state fixed effects).

Second, we assess the robustness of our results to removing partially designated counties from our treatment group. Appendix Table 3.B.1 reports point estimates for the medium run effects as well as overall pooled estimates. The first column reproduces our baseline estimates from studying all partially designated counties, and the remaining three columns report estimates from studying only counties that are at least 10%, 50%, and 100% designated. The point estimates remain generally consistent across columns. Results for later-career PCPs seem to vary more than others, though the effects are relatively small and are never statistically distinguishable from zero. We note that the number of observations drops by about 36% from column (1) to column (4).

Third, we vary our matching strategy. Appendix Table 3.B.2 reports results from altering the number of control counties that we match to each treatment county. Point estimates are broadly stable, though those for later-career PCPs appear more sensitive. Appendix Table 3.B.3 reports results from changing the variables used in our matching procedure. Column (1) reproduces estimates from our leading procedure. Column (2) does not match on the baseline trends in physician counts, and column (3) does not match on the baseline number of physicians. Column (4) matches only on geography and poverty rate. Column (5) matches on the baseline level of physicians along with a baseline trend in the poverty rate, rather than using the trend in physician counts. Overall our results appear mostly stable, especially the results on early-career ranked PCPs, and alternative matching procedures may address potential concerns about matching on both baseline levels and trends of physicians while also selecting a control group of counties that are themselves not designated over our time period.

3.6 Policy Discussion

Responsiveness to HPSA designation varies significantly by career stage: there is evidence for an increase in the stock of early-career PCPs, but no evidence of any effect for PCPs in later career stages. The 10% HPSA bonus payments are made to all physicians regardless of career stage, and the majority of PCPs in HPSA-designated counties in our sample are later-career PCPs. Thus, millions of dollars in bonus payments are spent on doctors who the empirical evidence suggests are unlikely to change their practice location in response to the program. The cost effectiveness of the HPSA bonus payment program may be improved by targeting the incentive payment exclusively to those who do respond, namely early-career PCPs.[25] In this case, even a bonus payment higher than 10% could result in a

[25]Note that these targeted groups can feasibly be identified by policymakers, as career stages are defined by readily observable physician characteristics: graduation date and age.

lower cost per additional PCP in shortage areas and an overall lower cost of the program.

To illustrate this, we walk through a simple policy analysis that compares the estimated cost effectiveness of the 10% bonus payment program to that of a hypothetical alternative program that offers larger bonus payments to only early-career PCPs. This exercise requires some caveats, as we make a handful of simplifying assumptions. Importantly, we assume that the entirety of the effect of HPSA designation on the stock of early-career PCPs stems from the bonus payments. However, other programs connected to HPSA designations as well as potential interactions between private insurance payments and HPSA status may contribute to the total incentives associated with designations.[26] We also focus just on the costs and effects of the program for PCPs, even though all physicians practicing in HPSAs receive the bonus payments. We make back-of-the-envelope calculations that take our point estimates at face value and assume that effects scale linearly with the size of the bonus payments. Our aim is to conduct a simple yet informative exercise that draws from our main findings to highlight policy implications.

Focusing on our analysis sample of 217 designated counties, in the year before treatment, the average designated county has 0.49 early-career PCPs and 3.15 later-career PCPs per 10,000. Taking the point estimates in Panel B of Table 3.2 at face value, the stock of early-career PCPs becomes 0.59 per 10,000 in the average post-treatment year while the stock of later-career PCPs remains unchanged. The claims data imply post-treatment bonus payments to PCPs totaling \$226,900 per year per county, resulting in an annual cost of \$2,268,600 per additional PCP per 10,000 in the average HPSA-designated county.[27]

[26] The 10% bonus payment is a salient and major incentive that impacts all doctors in HPSAs, and our estimates come from studying designations defined using CMS data on automatically-billed HPSAs. To the extent that official HPSA designations interact with other various government programs related to shortage areas though, there could be additional incentives for locating in a HPSA. For instance, most states maintain loan forgiveness programs for practicing in rural areas, some of which may use criteria related to official HPSA designations. (See Kulka and McWeeny (2019) for a more detailed discussion of state loan forgiveness programs.) Additionally, to the extent that private insurance companies follow the lead of Medicare (Clemens and Gottlieb 2017, Clemens et al. 2017) and offer bonus payments for providing services in shortage areas, the direct financial incentives for locating in a HPSA could be even greater.

[27] The figure of \$2,268,600 per year for 1 additional PCP per 10,000 comes from dividing the average annual bonus payment at the county level (\$226,900) by the average increase in early-career PCPs attributed to

Suppose instead that a 20% bonus payment is offered to all early-career PCPs who practice in a HPSA-designated county. The bonus payment would remain available to these PCPs as long as the county remains designated, while no bonus would be paid to PCPs who graduated from medical school more than 10 years before the time of designation. Assuming that the response scales linearly with respect to the size of the bonus payment, the stock of early-career PCPs would increase to 0.69 per 10,000 following treatment and the stock of later-career PCPs would remain constant at 3.15 per 10,000. So the new regime would be predicted to yield 0.20 additional PCPs per 10,000, but (according to the claims data) at a reduced total annual cost of $57,100 per county, or $285,600 per additional PCP per 10,000.[28] This amounts to nearly an eight-fold decrease in costs per PCP.

As explained above, we make several simplifying assumptions in arriving at these results. Most notably, if HPSA incentives other than the 10% bonus payments are contributing to the increase in early-career PCPs, we may be overestimating the reduction in costs per additional PCP that would result from altering the bonus payment program as described. Nonetheless, it seems likely that there is significant scope for reducing costs and improving the effectiveness of the bonus payment program by adjusting it to target the subset of physicians we find to be responsive to relocation incentives.

3.7 Conclusion

This paper studies how physician location decisions respond to 10 percent Medicare bonus payments for practicing in "shortage areas." We find that while the majority of primary

HPSA designation (about 0.1 PCPs per 10,000). Note that the MPUP dataset omits line items for services provided by an NPI to 10 or fewer beneficiaries in a given year, so all cost figures slightly understate the true totals.

[28]While this analysis assumes no effect of HPSA designation for later-career PCPs, note that the proposed regime of targeted 20% payments would result in increased cost-effectiveness even under less generous assumptions. For instance, we could assume a positive effect of 10% bonus payments on later-career PCPs of 0.26 PCPs per 10,000, which is the top of the 95% confidence interval on the point estimate for this career group. In this case the cost per an additional PCP per 10,000 under the standard 10% bonus payment program would be $630,200, still greater than the $285,600 under our proposed targeted 20% bonus payment program.

care physicians do not appear to respond to the policy, an important subset of doctors do respond. Designated counties, on average, experience an increase in the stock of early-career physicians that amounts to roughly 23% and corresponds to about 0.67 more doctors per county. Results indicate that this increase occurs rather quickly, is stable over time, and is driven by increases in counts of PCPs who attended ranked medical schools.

Our findings can inform policymakers tasked with alleviating physician shortages. Accounting for response heterogeneity by career stage of doctors might improve the cost-effectiveness of bonus payment programs. For instance, to avoid paying bonuses to infra-marginal physicians already located in shortage areas, an alternative program offered solely to physicians in the first 10 years of their career that pays an even greater bonus amount for Medicare procedures provided in HPSAs might attract more doctors and reduce costs.

3.8 Acknowledgements

Chapter 3, in full, is currently being prepared for submission for publication of the material. Khoury, Stephanie, Leganza, Jonathan M., and Masucci, Alex. "Health Professional Shortage Areas and Physician Location Decisions." The dissertation author was a primary investigator and an author of this material.

3.9 Figures and Tables

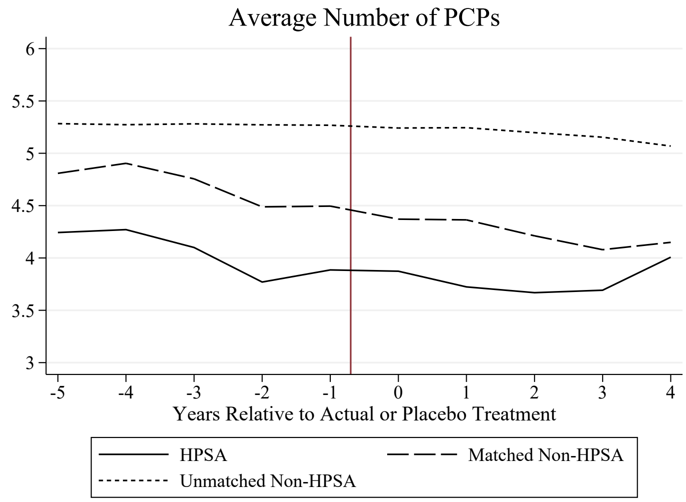


Figure 3.1: Average Number of PCPs for HPSA and Non-HPSA Counties

Notes: This graph plots the average number of PCPs per 10,000 population for treatment HPSA counties and potential non-HPSA control counties around actual or placebo designation year. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The matched control sample consists of the non-HPSA counties that are matched to HPSA counties using the method described in Section 3.4. The unmatched control sample consists of all counties that are never designated as a HPSA during 2012–2017, assigned as controls to and given placebo designation years from all counties in the treatment sample. Three control counties are matched to each treatment county, resulting in 217 treatment counties, 651 matched control counties (470 of which are unique), and 1,606 unmatched control counties. The x-axis shows the years relative to actual or placebo HPSA designation.

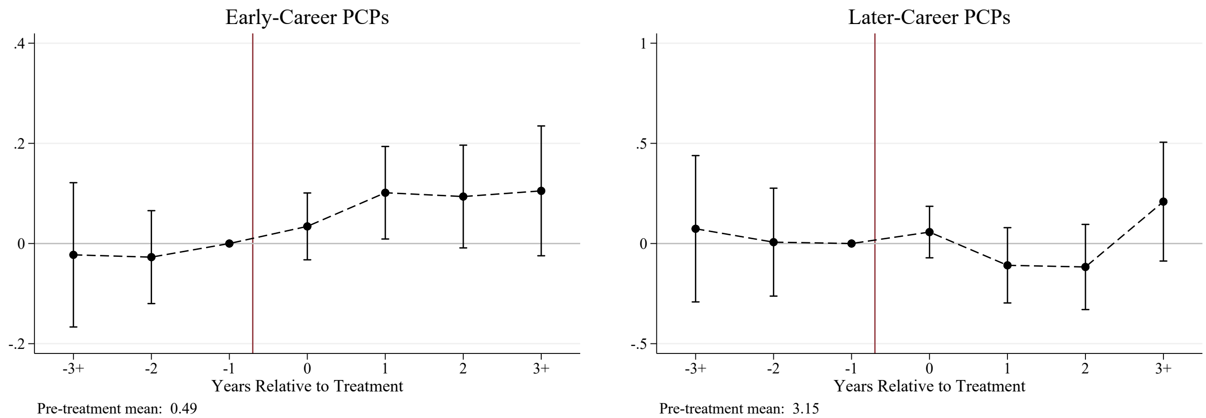


Figure 3.2: Impact of HPSA Designation on PCP Counts by Career Stage

Notes: These graphs plot the point estimates of the δ_τ's and their 95% confidence intervals from estimating equation (3.2), where the outcome y_{ct} is the stock of PCPs in the indicated career stage per 10,000 population in a county. Early-career PCPs are those graduating 5-10 years earlier and later-career PCPs are those graduating more than 10 years earlier. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. The x-axis shows the years relative to HPSA designation. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression.

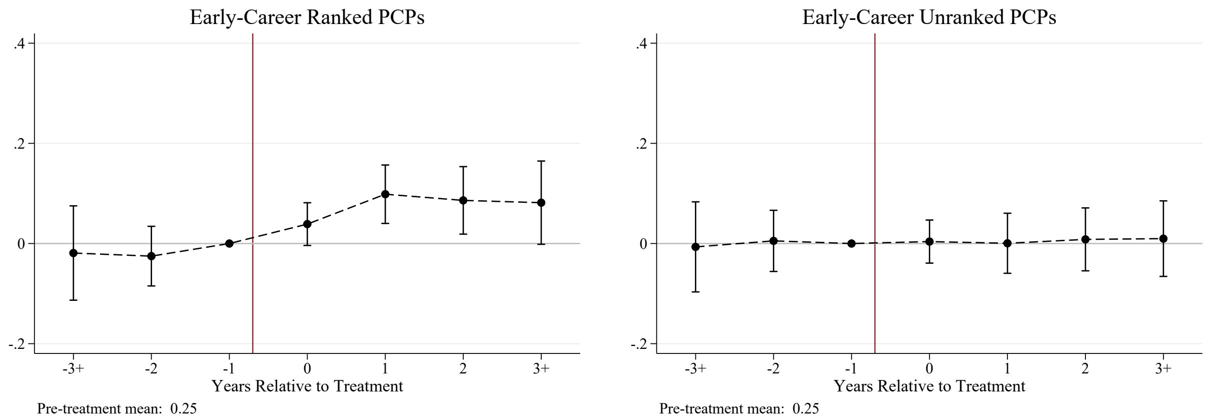


Figure 3.3: Impact of Designation on Early-Career PCP Counts by Medical School Rank

Notes: These graphs plot the point estimates of the δ_τ's and their 95% confidence intervals from estimating equation (3.2), where the outcome y_{ct} is the stock of early-career PCPs who attended ranked or unranked medical schools per 10,000 population in a county. Early-career PCPs are those graduating 5-10 years earlier. The 95 schools included in the 2018 U.S. News Primary Care medical school rankings are defined as ranked, and all other medical schools are defined as unranked. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. The x-axis shows the years relative to HPSA designation. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression.

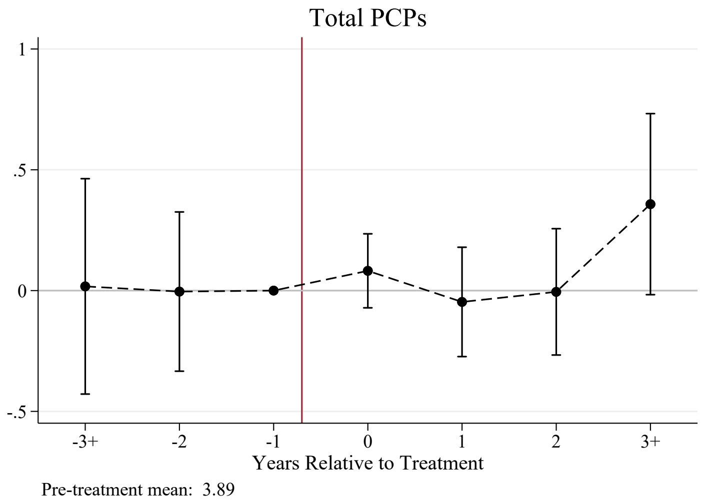


Figure 3.4: Impact of HPSA Designation on Total PCP Counts

Notes: This graph plots the point estimates of the δ_τ's and their 95% confidence intervals from estimating equation (3.2), where the outcome y_{ct} is the stock of PCPs per 10,000 population in a county. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. The x-axis shows the years relative to HPSA designation. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression.

Table 3.1: Summary Statistics for Descriptive Variables

	Treatment $\tau = -1$			Control $\tau = -1$		
	mean	min	max	mean	min	max
Physicians Per 10k	9.95	0.00	87.63	10.40	0.00	89.65
Percent Persons in Poverty	17.3	4.2	42.0	17.4	7.2	44.8
Population	58,969	690	1,265,111	67,568	589	1,919,402
Unemployment Rate	7.3	1.8	20.0	6.9	2.1	16.9
Median Household Income	44,479	22,834	86,703	44,161	23,837	110,843
Observations		217			651	

Notes: This table presents summary statistics for the analysis sample. Statistics are presented separately for the treatment group and the control group. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Data for each variable in the table is obtained for each county from the Area Health Resources File in the year before treatment for treatment counties and the year before the assigned treatment year for control counties. Physicians Per 10k (and its percentage change) and Percent Persons in Poverty are the variables used in the matching procedure to determine the closeness of eligible control counties to treatment counties.

Table 3.2: Impact of HPSA Designation on PCP Counts by Career Stage

	(1) Early-Career PCPs	(2) Later-Career PCPs
Panel A. Split Post-Period		
$treat_c \times post_{ct}^{SR}$	0.0476	0.0349
	(0.0431)	(0.0947)
$treat_c \times post_{ct}^{MR}$	0.114**	-0.00913
	(0.0570)	(0.146)
Panel B. Pooled Post-Period		
$treat_c \times post_{ct}$	0.0968*	0.00400
	(0.0509)	(0.128)
Dep. Mean	0.49	3.15
Clusters	687	687
Observations	5208	5208

Notes: This table presents the point estimates of δ^{SR} and δ^{MR} from estimating equation (3.4) in Panel A, and the point estimate of δ from estimating equation (3.3) in Panel B, where the outcome y_{ct} is the stock of PCPs in the indicated career stage per 10,000 population in a county. Early-career PCPs are those graduating 5-10 years earlier and later-career PCPs are those graduating more than 10 years earlier. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Impact of Designation on Early-Career PCPs by Medical School Rank

	(1) Early-Career Ranked PCPs	(2) Early-Career Unranked PCPs
Panel A. Split Post-Period		
$treat_c \times post_{ct}^{SR}$	0.0507*	0.00446
	(0.0278)	(0.0264)
$treat_c \times post_{ct}^{MR}$	0.100***	0.00694
	(0.0361)	(0.0335)
Panel B. Pooled Post-Period		
$treat_c \times post_{ct}$	0.0873***	0.00625
	(0.0323)	(0.0299)
Dep. Mean	0.25	0.25
Clusters	687	687
Observations	5208	5208

Notes: This table presents the point estimates of δ^{SR} and δ^{MR} from estimating equation (3.4) in Panel A, and the point estimate of δ from estimating equation (3.3) in Panel B, where the outcome y_{ct} is the stock of early-career PCPs who attended ranked or unranked medical schools per 10,000 population in a county. Early-career PCPs are those graduating 5-10 years earlier. The 95 schools included in the 2018 U.S. News Primary Care medical school rankings are defined as ranked, and all other medical schools are defined as unranked. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Impact of HPSA Designation on PCPs by Medical School Rank

	(1) Total PCPs	(2) Ranked PCPs	(3) Unranked PCPs
Panel A. Split Post-Period			
$treat_c \times post_{ct}^{SR}$	0.0786	0.0381	0.0322
	(0.115)	(0.0917)	(0.0803)
$treat_c \times post_{ct}^{MR}$	0.121	0.163	-0.0106
	(0.180)	(0.136)	(0.118)
Panel B. Pooled Post-Period			
$treat_c \times post_{ct}$	0.111	0.131	0.000776
	(0.157)	(0.120)	(0.105)
Dep. Mean	3.89	1.89	1.88
Clusters	687	687	687
Observations	5208	5208	5208

Notes: This table presents the point estimates of δ^{SR} and δ^{MR} from estimating equation (3.4) in Panel A, and the point estimate of δ from estimating equation (3.3) in Panel B. The outcome y_{ct} is the stock of PCPs per 10,000 population in a county in column 1, and this outcome is split up into PCPs who attended ranked or unranked medical schools in columns 2 and 3. The 95 schools included in the 2018 U.S. News Primary Care medical school rankings are defined as ranked, and all other medical schools are defined as unranked. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Robustness of Medium-Run Estimates to Alternative Regression Specifications

	(1) Early-Career PCPs	(2) Early-Career Ranked PCPs	(3) Early-Career Unranked PCPs	Later-Career PCPs
A. Baseline	0.114**	0.100***	0.00694	-0.00913
	(0.0431)	(0.0361)	(0.0335)	(0.146)
B. Winsor. 99	0.115*	0.116**	0.00594	0.0485
	(0.0691)	(0.0529)	(0.0380)	(0.161)
C. Winsor. 90	0.107*	0.0772***	0.00470	-0.0501
	(0.0490)	(0.0293)	(0.0288)	(0.136)
D. No Censoring	0.113	0.116**	-0.00269	0.0477
	(0.0712)	(0.0570)	(0.0418)	(0.168)
E. No Controls	0.111*	0.0988***	0.00251	-0.0134
	(0.0578)	(0.0360)	(0.0344)	(0.150)
F. More Controls	0.109*	0.0979***	0.00547	-0.0422
	(0.0553)	(0.0341)	(0.0327)	(0.144)
Clusters	687	687	687	687
Observations	5208	5208	5208	5208

Notes: This table presents point estimates of δ^{MR} from estimating equation (3.4) for the main outcomes as we vary the regression specification. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Row A reproduces our baseline estimates. Row B winsorizes outcome variables at the 99th percentile. Row C winsorizes outcome variables at the 90th percentile. Row D does not winsorize outcome variables. Row E drops controls from the regression. Row F adds year and state fixed effects to the regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Robustness of Pooled Estimates to Alternative Regression Specifications

	(1) Early-Career PCPs	(2) Early-Career Ranked PCPs	(3) Early-Career Unranked PCPs	Later-Career PCPs
A. Baseline	0.0968*	0.0873***	0.00625	0.0349
	(0.0509)	(0.0323)	(0.0299)	(0.0947)
B. Winsor. 99	0.0969	0.0987**	0.00434	0.0604
	(0.0616)	(0.0478)	(0.0341)	(0.141)
C. Winsor. 90	0.0902**	0.0659**	0.00483	-0.0368
	(0.0436)	(0.0261)	(0.0256)	(0.119)
D. No Censoring	0.0989	0.103**	-0.00409	0.0605
	(0.0641)	(0.0521)	(0.0374)	(0.147)
E. No Controls	0.0946*	0.0865***	0.00270	0.0000441
	(0.0515)	(0.0322)	(0.0307)	(0.131)
F. More Controls	0.0924*	0.0851***	0.00522	-0.0233
	(0.0495)	(0.0306)	(0.0291)	(0.125)
Clusters	687	687	687	687
Observations	5208	5208	5208	5208

Notes: This table presents point estimates of δ from estimating equation (3.3) for the main outcomes as we vary the regression specification. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Row A reproduces our baseline estimates. Row B winsorizes outcome variables at the 99th percentile. Row C winsorizes outcome variables at the 90th percentile. Row D does not winsorize outcome variables. Row E drops controls from the regression. Row F adds year and state fixed effects to the regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.A Appendix: Additional Figures and Tables

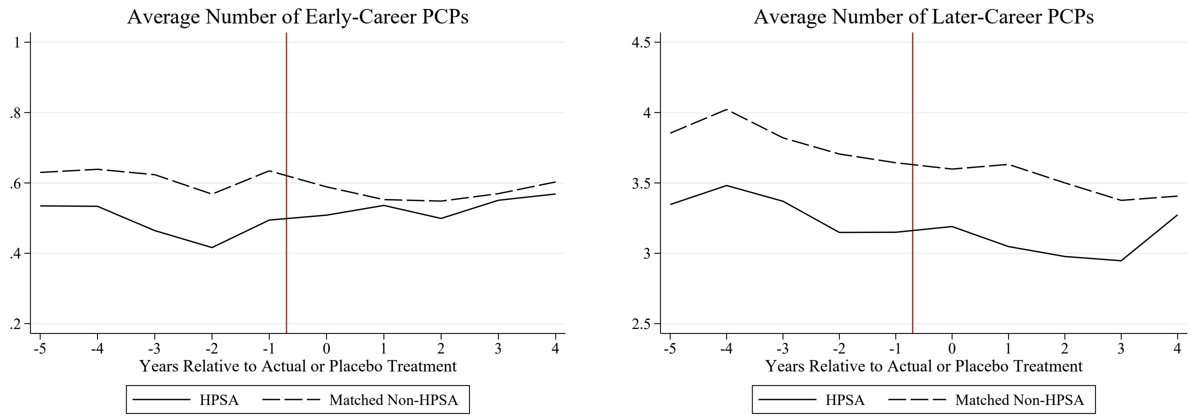


Figure 3.A.1: Average PCP Counts by Career Stage

Notes: These graphs plot the average number of PCPs in the indicated career stage per 10,000 population in a county in the sample of treatment HPSA counties and the non-HPSA control counties around actual or placebo treatment. Early-career PCPs are those graduating 5-10 years earlier and later-career PCPs are those graduating more than 10 years earlier. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. The x-axis shows the years relative to HPSA designation.

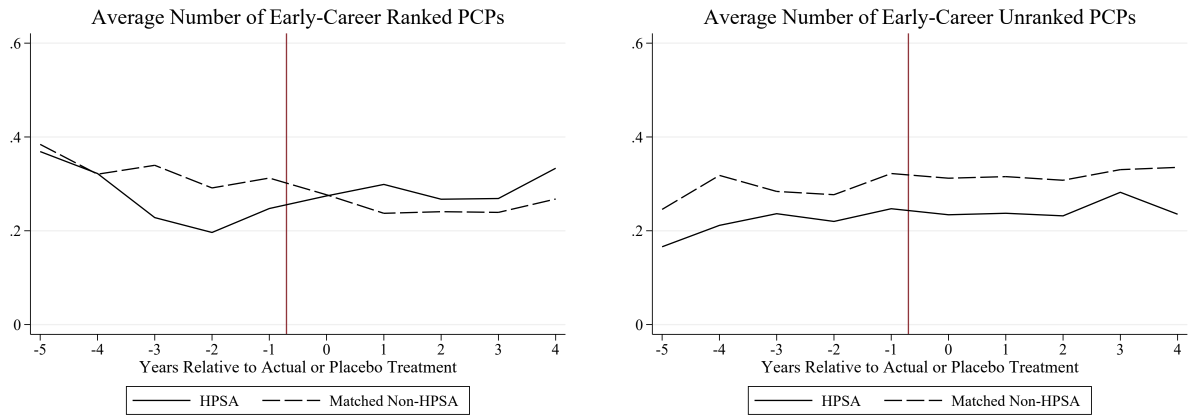


Figure 3.A.2: Average Early-Career PCP Counts by Medical School Rank

Notes: These graphs plot the average number of early-career PCPs who attended ranked or unranked medical schools per 10,000 population in a county in the sample of treatment HPSA counties and the non-HPSA control counties around actual or placebo treatment. Early-career PCPs are those graduating 5-10 years earlier. The 95 schools included in the 2018 U.S. News Primary Care medical school rankings are defined as ranked, and all other medical schools are defined as unranked. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. The x-axis shows the years relative to HPSA designation.

Table 3.A.1: Dynamic Impact of Designations on PCP Counts by Career Stage

	(1) Early-Career PCPs	(2) Later-Career PCPs
$treat_c \times -5$	0.00993	0.0390
	(0.131)	(0.299)
$treat_c \times -4$	0.0368	0.0938
	(0.104)	(0.246)
$treat_c \times -3$	-0.0623	0.0764
	(0.0683)	(0.186)
$treat_c \times -2$	-0.0272	0.00735
	(0.0473)	(0.137)
$treat_c \times -1$	0	0
	0	0
$treat_c \times 0$	0.0342	0.0572
	(0.0340)	(0.0655)
$treat_c \times 1$	0.101**	-0.109
	(0.0471)	(0.0958)
$treat_c \times 2$	0.0938*	-0.117
	(0.0523)	(0.108)
$treat_c \times 3$	0.126*	0.145
	(0.0657)	(0.156)
$treat_c \times 4$	0.0747	0.398**
	(0.0770)	(0.190)
Clusters	687	687
Observations	5208	5208

Notes: This table presents the δ_τ point estimates from estimating equation (3.2), where the outcome y_{ct} is the stock of PCPs in the indicated career stage per 10,000 population in a county. Early-career PCPs are those graduating 5–10 years earlier and later-career PCPs are those graduating more than 10 years earlier. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.2: Dynamic Impact of Designations on Early-Career PCPs by Medical School Rank

	(1) Early-Career Ranked PCPs	(2) Early-Career Unranked PCPs
$treat_c \times -5$	0.0536	-0.0282
	(0.0830)	(0.0713)
$treat_c \times -4$	0.0435	-0.00875
	(0.0751)	(0.0626)
$treat_c \times -3$	-0.0733*	0.00129
	(0.0426)	(0.0433)
$treat_c \times -2$	-0.0252	0.00529
	(0.0304)	(0.0311)
$treat_c \times -1$	0	0
	0	0
$treat_c \times 0$	0.0387*	0.00388
	(0.0217)	(0.0219)
$treat_c \times 1$	0.0984***	0.000520
	(0.0297)	(0.0306)
$treat_c \times 2$	0.0861**	0.00830
	(0.0343)	(0.0320)
$treat_c \times 3$	0.0789*	0.0286
	(0.0409)	(0.0403)
$treat_c \times 4$	0.0855	-0.0179
	(0.0520)	(0.0434)
Clusters	687	687
Observations	5208	5208

Notes: This table presents the δ_τ point estimates from estimating equation (3.2), where the outcome y_{ct} is the stock of early-career PCPs who attended ranked or unranked medical schools per 10,000 population in a county. Early-career PCPs are those graduating 5–10 years earlier. The 95 schools included in the 2018 U.S. News Primary Care medical school rankings are defined as ranked, and all other medical schools are defined as unranked. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012-2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.3: Dynamic Impact of Designations on PCPs by Medical School Rank

	(1) Total PCPs	(2) Ranked PCPs	(3) Unranked PCPs
$treat_c \times -5$	0.0804	0.0154	0.145
	(0.360)	(0.280)	(0.275)
$treat_c \times -4$	0.115	0.0422	0.102
	(0.300)	(0.252)	(0.213)
$treat_c \times -3$	-0.0497	-0.0400	0.0287
	(0.224)	(0.175)	(0.147)
$treat_c \times -2$	-0.00367	0.00363	0.0234
	(0.168)	(0.124)	(0.105)
$treat_c \times -1$	0	0	0
	0	0	0
$treat_c \times 0$	0.0818	0.0361	0.0578
	(0.0780)	(0.0560)	(0.0520)
$treat_c \times 1$	-0.0469	0.0889	-0.0911
	(0.115)	(0.0792)	(0.0777)
$treat_c \times 2$	-0.00539	0.0773	-0.00838
	(0.133)	(0.0943)	(0.0906)
$treat_c \times 3$	0.262	0.246*	0.0833
	(0.186)	(0.127)	(0.121)
$treat_c \times 4$	0.498**	0.331*	0.171
	(0.236)	(0.170)	(0.141)
Clusters	687	687	687
Observations	5208	5208	5208

Notes: This table presents the δ_τ point estimates from estimating equation (3.2), where the outcome y_{ct} is the stock of PCPs per 10,000 population in a county in column 1, and this outcome is split up into PCPs who attended ranked or unranked medical schools in columns 2 and 3. The 95 schools included in the 2018 U.S. News Primary Care medical school rankings are defined as ranked, and all other medical schools are defined as unranked. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013-2017. The control sample consists of all counties that are never designated as a HPSA during 2012-2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

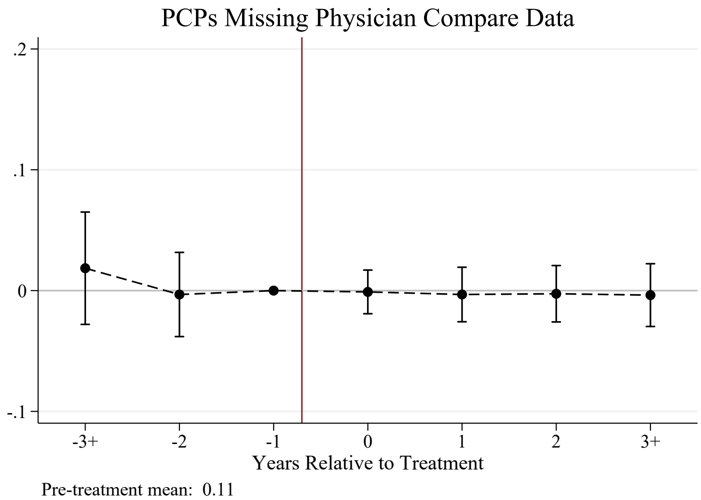


Figure 3.A.3: PCP Missing Data Relative to Designation

Notes: This graph plots the point estimates of the δ_τ's and their 95% confidence intervals from estimating equation (3.2), where the outcome y_{ct} is the stock of PCPs per 10,000 population in a county that are missing data on graduation year or medical school from the Physician Compare dataset. Almost all PCPs missing data on one of these variables are also missing data on the other variable. The treatment sample consists of all counties that become designated as a primary care HPSA in some year 2013–2017. The control sample consists of all counties that are never designated as a HPSA during 2012–2017 and are matched to the treatment counties using the matching procedure described in Section 3.4.1. Three control counties are matched to each treatment county, resulting in a sample size of 217 treatment counties and 651 control counties. 470 of the 651 control counties are unique, as control counties can be matched to multiple treatment counties. The x-axis shows the years relative to HPSA designation. Controls for unemployment rate, median household income, and population at the county-year level are included in the regression.

3.B Appendix: Additional Robustness Checks

Table 3.B.1: Robustness to Partially Designated County Inclusion

	(1) HPSA > 0%	(2) HPSA > 10%	(3) HPSA > 50%	(4) HPSA = 100%
Panel A. Medium Run Estimates				
Early-Career PCPs	0.114**	0.0946	0.112	0.101
	(0.0570)	(0.0600)	(0.0688)	(0.0721)
Early-Career Ranked PCPs	0.100***	0.102***	0.110**	0.105**
	(0.0361)	(0.0386)	(0.0439)	(0.0454)
Early-Career Unranked PCPs	0.00694	-0.0114	-0.0108	-0.00990
	(0.0335)	(0.0349)	(0.0399)	(0.0425)
Later-Career PCPs	-0.00913	-0.0561	-0.108	-0.184
	(0.146)	(0.151)	(0.161)	(0.168)
Panel B. Pooled Estimates				
Early-Career PCPs	0.0968*	0.0789	0.0893	0.0779
	(0.0509)	(0.0537)	(0.0624)	(0.0655)
Early-Career Ranked PCPs	0.0873***	0.0902***	0.0947**	0.0893**
	(0.0323)	(0.0347)	(0.0400)	(0.0415)
Early-Career Unranked PCPs	0.00625	-0.0118	-0.0114	-0.0110
	(0.0299)	(0.0312)	(0.0356)	(0.0379)
Later-Career PCPs	0.00400	-0.0444	-0.0937	-0.161
	(0.128)	(0.132)	(0.142)	(0.148)
Obs.	5,208	4,728	3,696	3,312

Notes: This table presents the point estimates of δ^{MR} from estimating equation (3.4) in Panel A and the point estimates of δ from estimating equation (3.3) in Panel B, for the main outcome variables as we vary the definition of HPSA designation. The columns designate the level at which a county must be designated to be included in the treatment group as a HPSA. Column (1) reproduces our preferred definition of designation, which includes all partially designated counties as treated counties. Columns (2), (3), and (4) include as treatment counties those with at least 10 percent, 50 percent, and 100 percent of zip codes designated, respectively. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.2: Robustness to Number of Matched Control Counties

	(1) $n_{control} = 1$	(2) $n_{control} = 2$	(3) $n_{control} = 3$	(4) $n_{control} = 4$	(5) $n_{control} = 5$
Panel A. Medium Run Estimates					
Early-Career PCPs	0.0932	0.107*	0.114**	0.0995*	0.0986*
	(0.0688)	(0.0613)	(0.0570)	(0.0560)	(0.0533)
Early-Career Ranked PCPs	0.101**	0.106***	0.100***	0.0946***	0.0925***
	(0.0415)	(0.0392)	(0.0361)	(0.0350)	(0.0336)
Early-Career Unranked PCPs	-0.00993	-0.00563	0.00694	0.00131	0.00276
	(0.0409)	(0.0359)	(0.0335)	(0.0333)	(0.0323)
Later-Career PCPs	0.118	0.0464	-0.00913	-0.0220	-0.0586
	(0.187)	(0.160)	(0.146)	(0.139)	(0.135)
Panel B. Pooled Estimates					
Early-Career PCPs	0.0768	0.0870	0.0968*	0.0839*	0.0832*
	(0.0614)	(0.0547)	(0.0509)	(0.0500)	(0.0477)
Early-Career Ranked PCPs	0.0858**	0.0901**	0.0873***	0.0833***	0.0809***
	(0.0371)	(0.0352)	(0.0323)	(0.0313)	(0.0301)
Early-Career Unranked PCPs	-0.00751	-0.00590	0.00625	0.000261	0.00142
	(0.0369)	(0.0320)	(0.0299)	(0.0296)	(0.0287)
Later-Career PCPs	0.0932	0.0512	0.00400	-0.00494	-0.0418
	(0.165)	(0.141)	(0.128)	(0.122)	(0.118)

Notes: This table presents the point estimates of δ^{MR} from estimating equation (3.4) in Panel A and the point estimates of δ from estimating equation (3.3) in Panel B, for the main outcome variables as we vary the number of controls matched to each treatment county. Column (3) reproduces our preferred matching procedure, in which we match 3 controls to each treatment county. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.3: Robustness to Match Variables

	(1)	(2)	(3)	(4)	(5)
Panel A. Medium Run Estimates					
Early-Career PCPs	0.114**	0.0961*	0.0996*	0.0556	0.0690
	(0.0570)	(0.0537)	(0.0597)	(0.0600)	(0.0503)
Early-Career Ranked PCPs	0.100***	0.0943***	0.0862**	0.0809**	0.0857***
	(0.0361)	(0.0344)	(0.0375)	(0.0373)	(0.0319)
Early-Career Unranked PCPs	0.00694	-0.00467	0.0116	-0.0249	-0.0208
	(0.0335)	(0.0321)	(0.0355)	(0.0352)	(0.0326)
Later-Career PCPs	-0.00913	-0.0539	-0.0761	-0.122	-0.103
	(0.146)	(0.142)	(0.146)	(0.146)	(0.137)
Panel B. Pooled Estimates					
Early-Career PCPs	0.0968*	0.0797*	0.0793	0.0441	0.0629
	(0.0509)	(0.0478)	(0.0534)	(0.0537)	(0.0449)
Early-Career Ranked PCPs	0.0873***	0.0802***	0.0719**	0.0724**	0.0735**
	(0.0323)	(0.0307)	(0.0336)	(0.0335)	(0.0287)
Early-Career Unranked PCPs	0.00625	-0.00308	0.00831	-0.0272	-0.0146
	(0.0299)	(0.0284)	(0.0314)	(0.0310)	(0.0285)
Later-Career PCPs	0.00400	-0.0417	-0.0569	-0.0953	-0.0842
	(0.128)	(0.124)	(0.128)	(0.127)	(0.120)
Match Variables:					
# Physicians	✓	✓	✗	✗	✓
%Δ Physicians	✓	✗	✓	✗	✗
Poverty Rate	✓	✓	✓	✓	✗
%Δ Poverty Rate	✗	✗	✗	✗	✓
Geographic Region	✓	✓	✓	✓	✓

Notes: This table presents the point estimates of δ^{MR} from estimating equation (3.4) in Panel A and the point estimates of δ from estimating equation (3.3) in Panel B, for the main outcome variables as we vary the variables used in the matching procedure. Column (1) reproduces our preferred matching procedure, in which we match on the baseline variables corresponding to the level of total physicians, trends in total physicians, and the poverty rate. Column (2) does not match on the baseline trends in physician counts. Column (3) does not match on the baseline number of physicians. Column (4) excludes both baseline trends and numbers of total physicians from the match. Column (5) matches on the baseline number of total physicians along with a baseline trend in the poverty rate. Controls for unemployment rate, median household income, and population at the county-year level are included in each regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Acemoglu, D. and A. Finkelstein (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy 116*(5), 837–880.

Alexander, D. (2015). Does physician pay affect procedure choice and patient health? evidence from medicaid c-section use. *FRB of Chicago Working Paper No. WP-2017-7*.

Bärnighausen, T. and D. E. Bloom (2009). Financial incentives for return of service in underserved areas: a systematic review. *BMC health services research 9*(1), 86.

Bazzoli, G. J. (1985). Does educational indebtedness affect physician specialty choice? *Journal of Health Economics 4*(1), 1–19.

Bhattacharya, J. (2005). Specialty selection and lifetime returns to specialization within medicine. *Journal of Human Resources 40*(1), 115–143.

Boccuti, C., C. Fields, G. Casillas, and L. Hamel (2015). Primary care physicians accepting medicare: A snapshot. *Kaiser Family Foundation Issue Briefs*.

Bolduc, D., B. Fortin, and M.-A. Fournier (1996). The effect of incentive policies on the practice location of doctors: a multinomial probit analysis. *Journal of Labor Economics 14*(4), 703–732.

Brekke, K. R., T. H. Holmås, K. Monstad, and O. R. Straume (2017). Do treatment decisions depend on physicians' financial incentives? *Journal of Public Economics 155*, 74–92.

Burfield, W., D. Hough, and W. Marder (1986). Location of medical education and choice of location of practice. *Academic Medicine 61*(7), 545–54.

Chandra, A., D. Cutler, and Z. Song (2011). Who ordered that? the economics of treatment choices in medical care. In *Handbook of Health Economics*, Volume 2, pp. 397–432. Elsevier.

Chen, F., M. Fordyce, S. Andes, and L. G. Hart (2010). Which medical schools produce rural physicians? a 15-year update. *Academic Medicine 85*(4), 594–598.

Chen, Y., P. Persson, and M. Polyakova (2020). The roots of health inequality and the value of intra-family expertise. *NBER Working Paper No. 25618*.

Clemens, J. and J. D. Gottlieb (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review 104*(4), 1320–49.

Clemens, J. and J. D. Gottlieb (2017). In the shadow of a giant: Medicare's influence on private physician payments. *Journal of Political Economy 125*(1), 1–39.

Clemens, J., J. D. Gottlieb, and J. Hicks (2018). Do medicare payments influence physicians' on-the-job investments? *Working Paper*.

Clemens, J., J. D. Gottlieb, and T. L. Molnár (2017). Do health insurers innovate? evidence from the anatomy of physician payments. *Journal of Health Economics 55*, 153–167.

Cutler, D., J. S. Skinner, A. D. Stern, and D. Wennberg (2019). Physician beliefs and patient preferences: a new look at regional variation in health care spending. *American Economic Journal: Economic Policy 11*(1), 192–221.

Deryugina, T., L. Kawano, and S. Levitt (2018). The economic impact of hurricane katrina on its victims: evidence from individual tax returns. *American Economic Journal: Applied Economics 10*(2), 202–33.

Devlin, R. A. and S. Sarma (2008). Do physician remuneration schemes matter? the case of canadian family physicians. *Journal of Health Economics 27*(5), 1168–1181.

Ellis, R. P. and T. G. McGuire (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics 5*(2), 129–151.

Falcettoni, E. (2018). The determinants of physicians' location choice: Understanding the rural shortage. *Working Paper*.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *The Quarterly Journal of Economics 122*(1), 1–37.

Finkelstein, A., M. Gentzkow, and H. Williams (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics 131*(4), 1681–1726.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003a). The implications of regional variations in medicare spending. part 1: the content, quality, and accessibility of care. *Annals of Internal Medicine 138*(4), 273–287.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003b). The implications of regional variations in medicare spending. part 2: health outcomes and satisfaction with care. *Annals of internal medicine 138*(4), 288–298.

Gagné, R. and P. T. Léger (2005). Determinants of physicians' decisions to specialize. *Health Economics 14*(7), 721–735.

Gottlieb, D. J., W. Zhou, Y. Song, K. G. Andrews, J. S. Skinner, and J. M. Sutherland (2010). Prices don't drive regional medicare spending variations. *Health Affairs 29*(3), 537–543.

Gottlieb, J. D., M. Polyakova, K. Rinz, H. Shiplett, V. Udalova, et al. (2020). Who values human capitalists' human capital? healthcare spending and physician earnings. *Working Paper*.

Hadley, J. and J. D. Reschovsky (2006). Medicare fees and physicians' medicare service volume: Beneficiaries treated and services per beneficiary. *International Journal of Health Care Finance and Economics 6*(2), 131–150.

Holmes, G. M. (2005). Increasing physician supply in medically underserved areas. *Labour Economics 12*(5), 697–725.

Huh, J. (2018). Medicaid and provider supply. *Working Paper*.

Hurley, J. E. (1991). Physicians' choices of specialty, location, and mode. *Journal of Human Resources 26*(1).

Johnson, E. M. and M. M. Rehavi (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy 8*(1), 115–41.

Kantarevic, J., B. Kralj, and D. Weinkauf (2008). Income effects and physician labour supply: evidence from the threshold system in ontario. *Canadian Journal of Economics 41*(4), 1262–1284.

Kulka, A. and D. B. McWeeny (2019). Rural physician shortages and policy intervention. *Working Paper*.

McGuire, T. G. (2000). Physician agency. In *Handbook of Health Economics*, Volume 1, pp. 461–536. Elsevier.

McGuire, T. G. and M. V. Pauly (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics 10*(4), 385–410.

Molitor, D. (2018). The evolution of physician practice styles: evidence from cardiologist migration. *American Economic Journal: Economic Policy 10*(1), 326–56.

Nicholson, S. (2002). Physician specialty choice under uncertainty. *Journal of Labor Economics 20*(4), 816–847.

Nicholson, S. and C. Propper (2011). Medical workforce. In *Handbook of Health Economics*, Volume 2, pp. 873–925. Elsevier.

Nicholson, S. and N. S. Souleles (2001). Physician income expectations and specialty choice. *NBER Working Paper No. 8536*.

Sarma, S., R. A. Devlin, B. Belhadji, and A. Thind (2010). Does the way physicians are paid influence the way they practice? the case of canadian family physicians' work activity. *Health Policy 98*(2-3), 203–217.

Sivey, P., A. Scott, J. Witt, C. Joyce, and J. Humphreys (2012). Junior doctors' preferences for specialty choice. *Journal of Health Economics 31*(6), 813–823.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In *Handbook of Health Economics*, Volume 2, pp. 45–93. Elsevier.

Sloan, F. A. (1970). Lifetime earnings and physicians' choice of specialty. *Industrial and Labor Relations Review 24*(1), 47–56.

Song, Y., J. Skinner, J. Bynum, J. Sutherland, J. E. Wennberg, and E. S. Fisher (2010). Regional variations in diagnostic practices. *New England Journal of Medicine 363*(1), 45–53.

Sørensen, R. J. and J. Grytten (2003). Service production and contract choice in primary physician services. *Health Policy 66*(1), 73–93.

Sutherland, J. M., E. S. Fisher, and J. S. Skinner (2009). Getting past denial - the high cost of health care in the united states. *New England Journal of Medicine 361*(13), 1227–1230.

Zhou, T. J. (2017). The doctor is in/out: Determinants of physician labor supply dynamics. *Working Paper*.

Zuckerman, S., T. Waidmann, R. Berenson, and J. Hadley (2010). Clarifying sources of geographic differences in medicare spending. *New England Journal of Medicine 363*(1), 54–62.