UNIVERSITY OF COPENHAGEN
DEPARTMENT OF ECONOMICS

PhD thesis

Esteban García-Miralles

Essays on Retirement, Savings, Taxation, and Skill Formation

Advisors: Mette Gørtz and Miriam Gensowski

January 31, 2021

Contents

Acknowledgments

I write with joy as I recall all the people that have helped me to write this dissertation. First, I would like to thank my two advisors. To Mette Gørtz for her continuous and generous support, and for her effective supervision. And to Miriam Gensowski, with whom I have written the fourth chapter of this thesis, for her energetic guidance and for teaching me the rewards of perseverance.

I am also very grateful to my co-author Jonathan Leganza, with whom I have written the first and second chapters of this dissertation. Not only have I grown as a researcher through our collaboration, but I also have had the most stimulating and fun times. Jonathan and I worked together during my visit to the University of California, San Diego during the fall of 2019. I would like to thank my hosts Gordon Dahl and Itzik Fadlon for their extremely generous and useful feedback, and for their support during and after my visit.

I also thank two of my PhD colleagues and friends with whom I have shared so many good times, and a few worries, over the last years, Patrick Thöni and Benedikt Busch. Many more people from the Economics department at the University of Copenhagen have also supported me with generous and useful feedback and discussions. I thank Pol Campos, Kristoffer B. Hvidberg, Claus Kreiner, Søren Leth-Petersen, Torben H. Nielsen, Benjamin Ly Serena, Sonja Settele, and Jakob Egholt Søgaard. As well as Emiliano Santoro for his keen guidance through the Job Market.

Finally, I would also like to thank a number of people that have influenced my earlier research career, such as my undergraduate professors Jorge Onrubia and Lourdes Moreno, and researchers at Bank of Spain from my time as a research assistant, Olympia Bover, Laura Crespo, José María Casado, José María Labeaga, and Roberto Ramos.

Esteban García-Miralles
Copenhagen, January 2021.

Introduction

This PhD dissertation consists of four self-contained chapters. The first two chapters study how the design of public pension systems affect individuals' behavior. The first chapter studies how pension eligibility ages impact the retirement behavior of couples. The second chapter explores how public pension provision impacts private savings. The third chapter of this dissertation studies how optimal inheritance taxation depends on the social welfare function assumed and on the underlying individual heterogeneity. Finally, the fourth chapter explores human capital formation by estimating the causal effect of parental shocks on their children's non-cognitive skills.

Chapter I. Joint Retirement of Couples: Evidence from Discontinuities in Denmark
with Jonathan Leganza

This chapter studies how social security influences joint retirement of couples. We exploit three decades of administrative data from Denmark to explore joint retirement in two complementary settings. In the first setting, we exploit the discontinuous increase in retirement observed when individuals become eligible for public pension benefits to identify the causal effects on their spouses. We find that spouses are more likely to retire right when their partners reach pension eligibility age, with a spillover effect across spouses of 7.5%.

We further unpack this result by studying additional margins of adjustment such as benefit claiming and earnings, and by documenting meaningful response heterogeneity. We find age differences within couples to be a crucial determinant of joint retirement, which is primarily driven by older spouses who continue to work until their younger partners reach pension eligibility. This means that joint retirement behavior increases aggregate labor supply. Controlling for these age differences uncovers a gender gap where female spouses are more likely to adjust their behavior to retire jointly. This gender gap remains after controlling for earnings shares within couples, suggesting that it may be due to gender norms.

In the second setting, we study to what extent couples adapt their behavior to retire jointly after a reform increases pension eligibility ages. We find spillover effects across spouses comparable to those from the first setting, in which eligibility ages were stable and known by couples well in advance. This suggests that spouses do not face adjustment costs limiting their capacity to retire together after the reform.

Chapter II. Public Pensions and Private Savings

with Jonathan Leganza

This chapter studies how the provision of public pension benefits impact private savings. We answer this question in the context of a reform in Denmark that altered old-age benefit payouts through a discontinuous increase in pension eligibility ages contingent on birthdate. Using detailed administrative data and a regression discontinuity design, we identify the causal effects of the policy, leveraging our setting to study essentially the entire financial portfolio.

We document responses over two distinct time horizons. First, we show a lack of responses after the reform was announced but before it was implemented, inconsistent with the notion that future differences in pension eligibility impact savings. 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.

Chapter III. The Crucial Role of Social Welfare Criteria and Individual Heterogeneity for Optimal Inheritance Taxation

This chapter extends the calibrations of Piketty and Saez (2013. "A Theory of Optimal Inheritance Taxation." Econometrica 81 (5): 1851–86) to unveil the importance of the assumed social welfare criteria and its interplay with individual heterogeneity on optimal inheritance taxation. I calibrate the full social optimal tax rate and find that it is highly sensitive to the assumed social welfare criteria. The optimal tax rate ranges from negative (under a utilitarian criterion) to positive and large (even assuming joy of giving motives). A decreasing marginal utility of consumption does not affect the results qualitatively, given the underlying distribution of wealth and income.

I also calibrate the optimal tax rate by percentile of the distribution of bequest received, as in Piketty and Saez, but accounting for heterogeneity in wealth and labor income. This leads to significant variation in the optimal tax rate among zero-bequest receivers, contrary to their finding of a constant tax rate.

Chapter IV. Are Children's Socio-Emotional Skills Shaped by Parental Health Shocks?
with Miriam Gensowski

This chapter explores the formation of human capital with a focus on socio-emotional skills. Child skills are shaped by parental investments, when parents experience a health shock, their investments and therefore their children's skills may be affected. We estimate causal effects of severe parental health shocks on child socio-emotional skills. We leverage a large-scale survey distributed to all public schools in Denmark, which we link to hospital records for the entire Danish population. We find that socio-emotional skills of 11-16 year-olds are robust to parental health shocks, with the exception of statistically significant but very small reductions in conscientiousness. We study short-run effects of these shocks with a child-fixed effects model, and dynamics around the shocks with event studies. Finally, a sibling comparison suggests some long-run build-up of effects of early shocks.

Introduktion

Denne ph.d.-afhandling består af fire selvstændige kapitler. De to første kapitler undersøger, hvordan udformningen af offentlige pensionssystemer påvirker individers adfærd. Det første kapitel undersøger, hvordan pensionsalderen påvirker tilbagetrækningsadfærden for par. Det andet kapitel udforsker, hvordan udbuddet af offentlige pensioner påvirker private opsparinger. Afhandlingens tredje kapitel undersøger, hvordan optimal arvebeskatning afhænger af den samfundsvelfærdsfunktion, der antages og den underliggende individuelle heterogenitet. Til slut udforsker det fjerde kapitel udvikling af human kapital ved at estimere den kausale effekt af stød til forældrene på deres børns ikke-kognitive evner.

Kapitel I. Fælles Tilbagetrækning for Par: Evidens fra Diskontinuiteter i Danmark
med Jonathan Leganza

Dette kapitel undersøger, hvordan social sikring påvirker fælles tilbagetrækning for par. Vi udnytter tre årtiers administrative data fra Danmark til at udforske fælles tilbagetrækning gennem to komplimentære rammer. Inden for den første ramme udnytter vi den diskontinuerlige stigning i pensionsalderen, der observeres for individer for at identificere den kausale effekt på partnere. Vi finder at partnere er mere tilbøjelige til at trække sig, tilbage når deres partnere har nået tilbagetrækningsalderen med en spillover-effekt på tværs af partnere på 7,5%.

Vi klarlægger videre resultatet ved at undersøge yderligere justerbare marginer såsom ansøgning om ydelser og indtjening og ved at dokumentere betydelig reaktionsheterogenitet. Vi finder, at aldersdifferencen inden for parrene er en afgørende determinant for fælles tilbagetrækning, hvilket primært er drevet af ældre partnere, der fortsætter med at arbejde indtil deres yngre partnere når pensionsalderen. Dette betyder at fælles tilbagetrækningsadfærd øger det samlede arbejdsudbud. Ved at kontrollere for disse aldersforskelle afdækkes en forskel mellem mænd og kvinder, hvor kvindelige partnere er mere tilbøjelige til at tilpasse deres adfærd for at trække sig fælles tilbage. Denne forskel mellem kønnene vedbliver efter vi kontrollerer for indtjeningsandele mellem partnerne, hvilket antyder at det kan skyldes kønsnormer.

Inden for den anden ramme undersøger vi, i hvilken grad par tilpasser deres tilbagetrækningsadfærd efter en reform, der forhøjer pensionsalderen. Vi finder spillovereffekter på tværs af par, der er sammenlignelige med dem fundet inden for den første ramme, hvor pensionsalderen var stabil og kendt af parrene på forhånd. Dette antyder at partnere ikke har justeringsomkostninger, der begrænser deres mulighed for at trække sig fælles tilbage efter reformen.

Kapitel II. Offentlige Pensioner og Private Opsparinger

med Jonathan Leganza

Dette kapitel undersøger, hvordan udbuddet af offentlige pensioner påvirker private opsparinger. Vi besvarer dette spørgsmål i konteksten af en reform i Danmark, der ændrede pensionsudbetalingerne gennem en diskontinuert stigning i pensionsalderen. Ved at benytte detaljerede administrative data og et "regression discontinuity design" identificerer vi den kausale effekt af reformen og udnytter dette til reelt at undersøge hele den finansielle portefølje.

Vi dokumenterer reaktioner over to forskellige tidshorisonter. Først viser vi mangel på reaktion efter reformen blev annonceret, men før den blev implementeret, hvilket er inkonsistent med opfattelsen om, at fremtidige forskelle i pensionsalderen skulle påvirke opsparing. Dernæst viser vi, at der er store opsparingsreaktioner efter implementeringen, hvor udskudt pensionsalder motiverer individer til at forlænge deres beskæftigelse. Konkret finder vi stigende bidrag til både arbejdsgiverbetalte og personlige pensionsopsparinger, mens vi ikke finder tegn på justeringer i andre opsparingstyper som bank- og aktieopsparinger.

Yderligere analyse peger på træghed som en af de vigtigste forklarende kanaler. Forøgelsen af opsparing i individuelle pensioner drives udelukkende af dem, der har foretaget konsistente bidrag tidligere. Desuden kan stigningen i arbejdsgiverbetalt pensioner overvejende forklares af en fortsættelse af arbejdsgivers standardindbetalinger, hvilket viser en rolle for firmapolitikker i at mediere reaktioner af reformer af sociale pensioner.

Kapitel III. Den Afgørende Rolle Velfærdskriterier og Individuel Heterogenitet Spiller for Optimal Formue Beskatning

Dette kapitel udvider Piketty and Saez' (2013. "A Theory of Optimal Inheritance Taxation." Econometrica 81 (5): 1851–86) kalibreringer for at afdække vigtigheden af det antagne velfærdskriterie og dets samspil med individuel heterogenitet for optimal formuebeskatning. Jeg kalibrerer den fuldt socialt optimale skatterate og finder, at den i høj grad er sensitiv til det antagne velfærdskriterie. Den optimale skatterate går fra negativ (under et utilitaristisk kriterie) til positiv og stor (selv hvis 'glæde ved at give' antages). Givet den underliggende fordeling af formue og indkomst påvirker aftagende marginalnytte af forbrug ikke kvalitativt resultaterne.

Jeg kalibrerer også den optimale skatterate for percentiler af fordelingen af arv modtaget som i Piketty og Saez, bortset fra at jeg tager højde for heterogenitet i formue- og arbejdsindkomst. Dette leder til signifikant variation i den optimale skatterate blandt dem, der ikke arver, hvilket står i modsætning til deres resultat om en konstant skatterate.

Kapitel IV. Er Børns Socio-Emotionelle Færdigheder Formet af Stød til Forældres Sundhedstilstand?

med Miriam Gensowski

Dette kapitel udforsker udviklingen af human kapital med fokus på socio-emotionelle færdigheder. Børns færdigheder er formet af forældres investeringer, så når forældre oplever stød til deres sundhedstilstand, kan deres investeringer og derfor deres børns færdigheder blive påvirket. Dette kapitel estimerer kausale effekter fra alvorlige stød til forældres helbred på børns socio-emotionelle færdigheder. Vi udnytter et omfattende spørgeskema, der er distribueret til alle offentlige skoler i Danmark, hvilket vi kobler til hospitalsregistrer for hele den danske befolkning. Vi finder at socio-emotionelle færdigheder for 11 til 16-årige er robuste over for stød til forældres sundhedstilstand med undtagelse af statistisk signifikante men meget små reduktioner i samvittighedsfuldhed. Vi undersøger kortsigtseffekterne af disse stød med en "barn-fixed-effects"-model og dynamikken omkring stødene med "event studies". Sammenligning af søskende antyder til slut en langsigtet opbygningseffekter af tidlige stød.

Chapter I

Joint Retirement of Couples: Evidence from Discontinuities in Denmark

Joint Retirement of Couples:
Evidence from Discontinuities in Denmark*

Esteban García-Miralles† Jonathan M. Leganza‡

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.

Keywords: Joint Retirement, Pension Eligibility Age, Couples Labor Supply
JEL Classification: J14, J26, D10, H55

*We thank our advisors, Gordon Dahl, Itzik Fadlon, Miriam Gensowski and Mette Gørtz for support and guidance. We also thank Martin Browning, Mette Ejrnæs, Alex Gelber, Thomas Høgholm Jørgensen, Claus Kreiner, Søren Leth-Petersen, Torben Heien Nielsen, Benjamin Ly Serena, Adam Sheridan, Jakob Egholt Søgaard, and participants at the CESifo Area Conference on Public Economics, at the CEBI and Graduate Seminars at the University of Copenhagen, at the DGPE conference, and at the 45th Spanish Economic Association Meeting for helpful comments. García-Miralles gratefully acknowledges funding from the Novo Nordisk Foundation (grant no. NNF17OC0026542) and from the the Danish National Research Foundation through its grant (DNRF-134) to the Center for Economic Behavior and Inequality (CEBI).

†University of Copenhagen and CEBI. (email: egm@econ.ku.dk)
‡University of California, San Diego. Department of Economics. (email: jleganza@ucsd.edu)

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]

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

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

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]

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 describes the institutional background.

[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: 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).

Section 3 presents the data and the samples of analysis. Section 4 lays out our empirical strategy for estimating the effect of reaching a stable pension eligibility age and reports the results. Section 5 analyzes the reform that increased pension eligibility ages. Section 6 concludes.

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.

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 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 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

[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.

[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.

are not affected by differential pension eligibility ages.[8]

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.

3 Data and Sample of Analysis

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]

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

[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.

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

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 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.

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

[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.

distribution. We illustrate the distribution of age differences within couples in panel (a) of Appendix Figure A.4, and we show that our results are robust to dropping this restriction in Section 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 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.

[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.

[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.

Table 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.

4 The Effect of Reaching Pension Eligibility Age

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.

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 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 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:

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

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 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 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 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 1, and second, individuals who claim can still have positive 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.

[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 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 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 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 3, around 80% of the individuals in the age-based sample of analysis made contributions to qualify for VERP.

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.

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:

$$y_{it}^s = \alpha + \sum_{a=49}^{69} \delta_a \cdot D_a^s + \sum_{a=49}^{69} \gamma_a \cdot D_a^s \cdot D_g + \sum_{c=1991}^{2013} \kappa_c \cdot D_c + \varepsilon_{it} \tag{2}$$

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

$$\begin{aligned} y_{it}^s = \alpha + \beta_1\, age_{it} + \beta_2\, \mathbb{1}\{age_{it} \geq 60\} + \beta_3\, \mathbb{1}\{age_{it} \geq 60\} \cdot age_{it} + \\ \sum_{a=49}^{69} \delta_a \cdot D_a^s + \sum_{a=49}^{69} \gamma_a \cdot D_a^s \cdot D_g + \sum_{c=1991}^{2013} \kappa_c \cdot D_c + \epsilon_{it} \end{aligned} \tag{3}$$

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

where y_{it}^s 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_a^s 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 superimposes the parametric fit of the model estimated in equation (3) over the residuals from equation (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 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, 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

[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).

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 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.

4.4 Explaining Joint Retirement: Heterogeneity and Mechanisms

The aggregate results from the previous section, reported in Table 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.

4.4.1 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 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

[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.

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. 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 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.

4.4.2 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 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 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 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 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.

[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.

[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 A.1 where, as an alternative to the reweighting strategy, we split the sample in four, by gender and by relative age between partners.

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 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 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.

4.4.3 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 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

[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 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 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

[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.

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 subsamples in terms of age differences and gender.

The results are reported in Table 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.

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 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 5.

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 A.3 and Appendix Figure A.2.

Alternative specifications. In Table 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 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

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

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 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 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.

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.

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 the OAP age increased as well.

The design of the VERP program, which we introduced in Section 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).

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

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

the period 2008–2014 of the form:

$$y_{it}^{(s)} = \alpha + \delta \cdot treat_i + \sum_{c \neq 2010} \kappa_c \cdot D_c + \sum_{c \neq 2010} \beta_c \cdot D_c \cdot treat_i + X'_{it} \cdot \psi + \epsilon_{it} \tag{4}$$

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 5 and 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:

$$y_{it}^{(s)} = \beta_0 + \beta_1 \cdot treat_i + \beta_2 \cdot ant_{it} + \beta_3 \cdot post_{it} + \beta_4 \cdot treat_i \cdot ant_{it} + \beta_5 \cdot treat_i \cdot post_{it} + X'_{it} \cdot \psi + \epsilon_{it} \tag{5}$$

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 (5) when it is estimated for the reference individual's outcomes. The second-stage equation is the following:

$$y_{it}^{s} = \beta_0 + \beta_1 \cdot \hat{y}_{it} + \beta_2 \cdot treat_i + \beta_3 \cdot ant_{it} + \beta_4 \cdot post_{it} + \beta_5 \cdot treat_i \cdot ant_{it} + X_{it}'\psi + u_{it} \quad (6)$$

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 5.5 along with other specification tests.

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 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 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 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 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.

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 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 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 6. The scaled effect on retirement is 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,

[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 5 and 6) we include an indicator variable for the years between announcement and anticipation of the reform.

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 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).

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 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 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 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 5 and 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 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 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.

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.

References

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

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

Atalay, K., Barrett, G. F., and Siminski, P. (2019). 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., Blundell, R., and Rivas, M. C. (2010). The dynamics of retirement behavior in couples: Reduced-form evidence from England and the US . *University College London, mimeo.*

Behaghel, L. and Blau, D. M. (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 Lanot, G. (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., Hochguertel, S., and Zweerink, J. (2019). The effect of incentive-induced retirement on spousal retirement rates: Evidence from a natural experiment. *Economic Inquiry*, 57(2):910–930.

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

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

Browning, M., Donni, O., and Gørtz, M. (2020). 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 Gruber, J. (2007). Future social security entitlements and the retirement decision. *The Review of Economics and Statistics*, 89(2):234–246.

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

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

Deshpande, M., Fadlon, I., and Gray, C. (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 Chiappori, P.-A. (2011). Nonunitary models of household behavior: a survey of the literature. *Household Economic Behaviors*, pages 1–40.

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

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

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

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

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

Gustman, A. L. and Steinmeier, T. L. (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 de Paula, Á. (2018). A new model for interdependent durations. *Quantitative Economics*, 9(3):1299–1333.

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

Hospido, L. and Zamarro, G. (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*, pages 231–258. University of Chicago Press, 1990.

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

Kreiner, C. T., Leth-Petersen, S., and Skov, P. E. (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 Parrotta, P. (2017). 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 Weber, A. (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 Vermeulen, F. (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). 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). What happens to the husband's retirement decision when the wife's retirement incentives change? *International Tax and Public Finance*, 24(3):432–458.

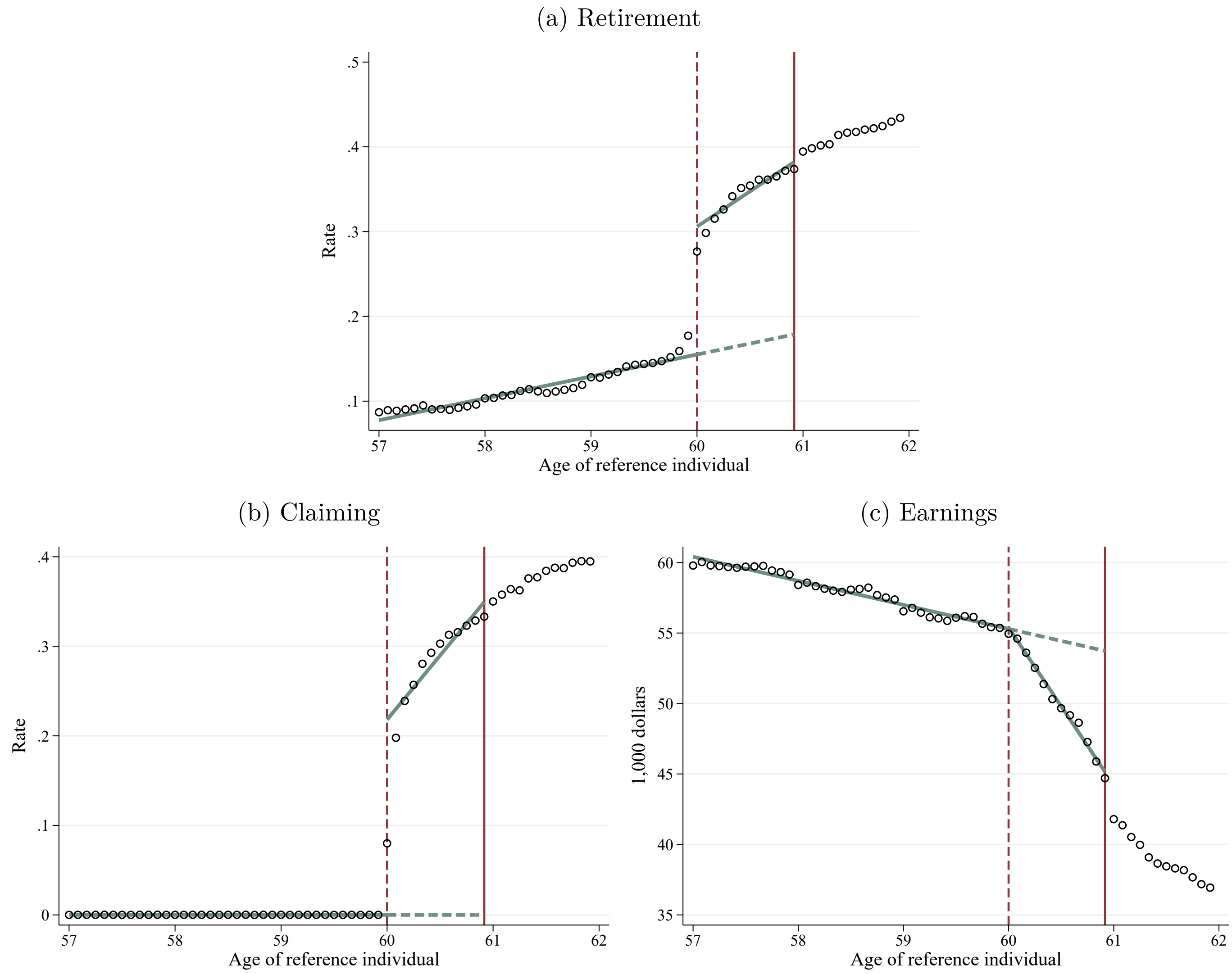
Stancanelli, E. (2017). Couples' retirement under individual pension design: A regression discontinuity study for France. *Labour Economics*, 49:14–26.

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

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

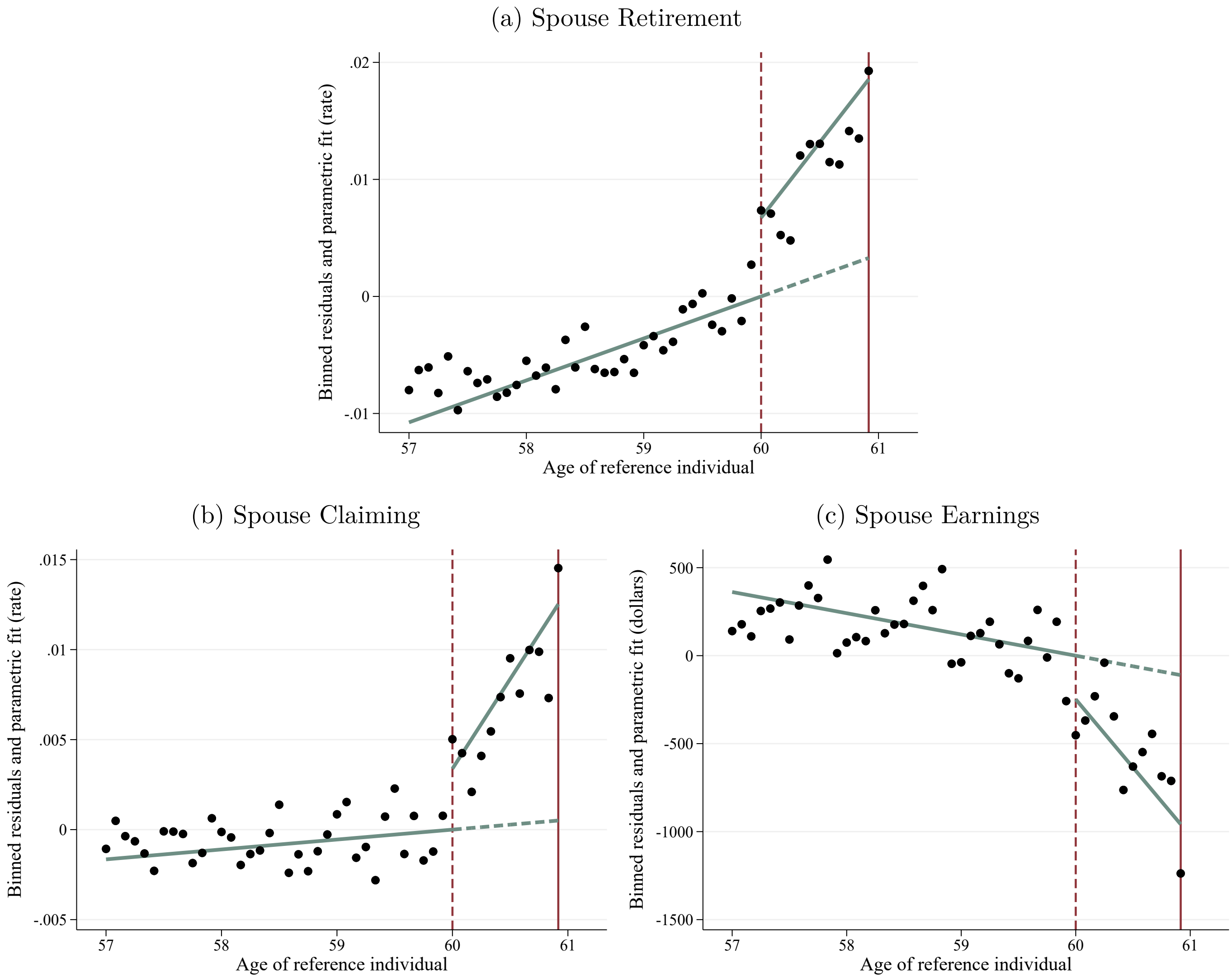
Willén, A., Vaage, K., and Johnsen, J. (2020). Interactions in public policies: spousal responses and program spillovers of welfare reforms. *NHH Dept. of Economics Discussion Paper*, (20).

Figure 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 (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.

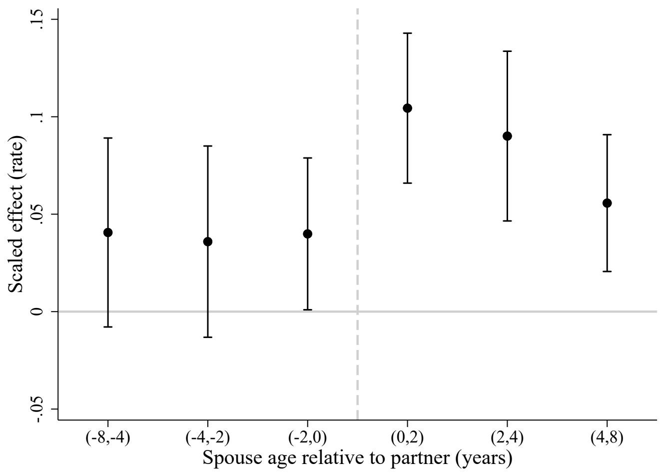
Figure 2: The Effect of Reaching Pension Eligibility Age on Spouses



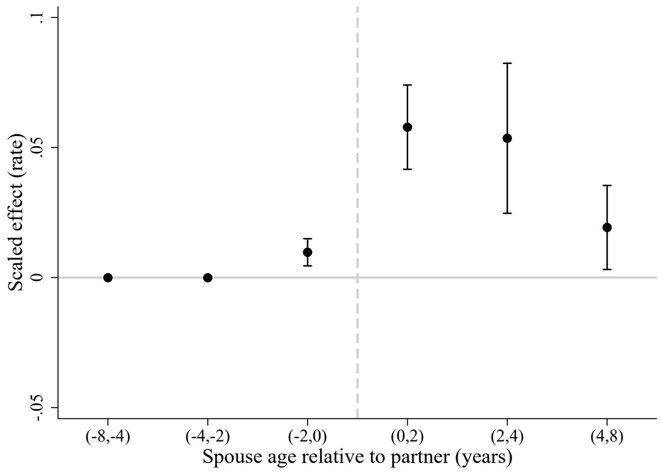
Notes: These figures plot different outcomes for spouses around the pension eligibility age of their partner. The dots are the residuals estimated in equation (2) where the spousal outcome is regressed on their own age and gender. The residuals are grouped in monthly bins of the reference individual's age. The solid lines plot the parametric fit estimated with the piecewise linear regression model (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 3: Joint Retirement Behavior by Age Differences Within Couples

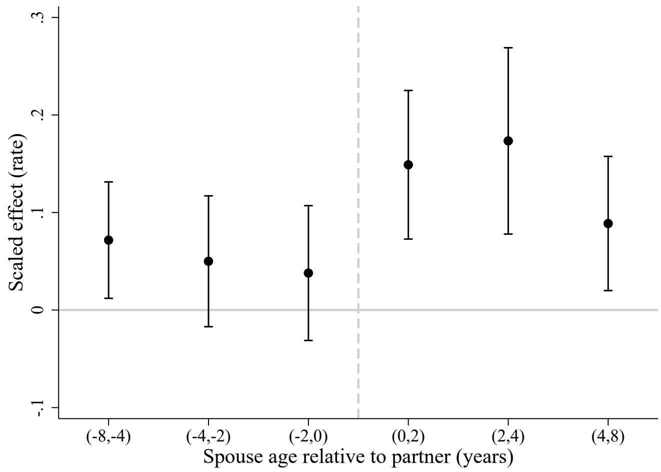
(a) Retirement



(b) Claiming

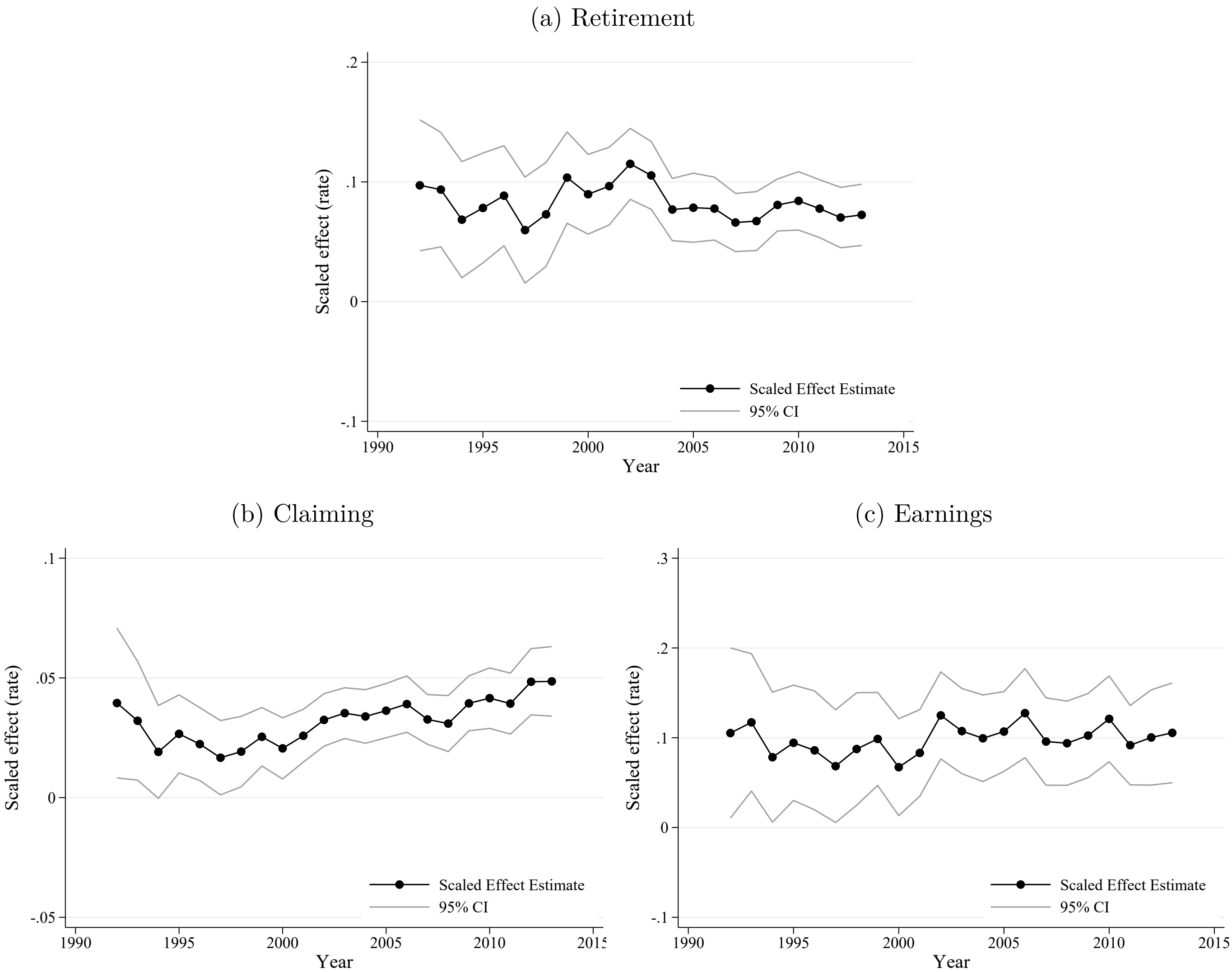


(c) Earnings



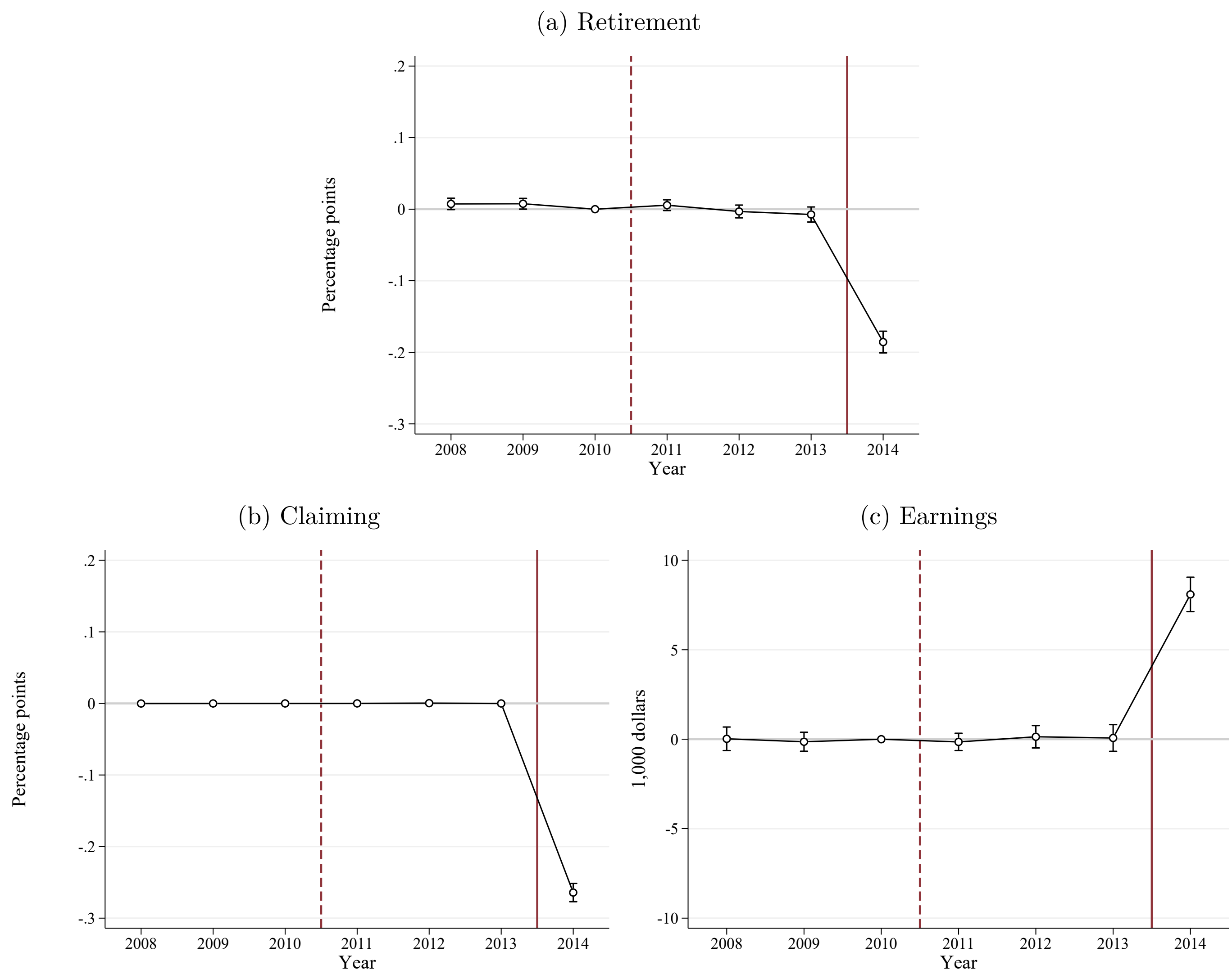
Notes: These figures plot the scaled estimates of joint retirement for different subsamples of couples based on the age difference between spouses. These scaled effects are estimated using the same methodology as for the full sample: first estimating models (1) and (3) to obtain full-exposure effects 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 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 (1) and (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. We report 95% confidence intervals calculated from bootstrapped standard errors.

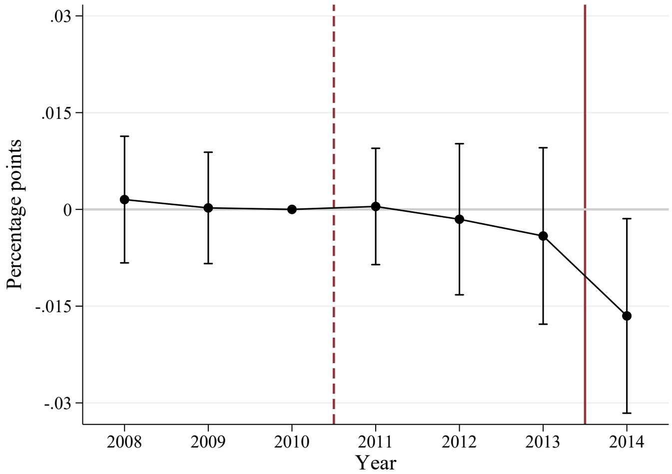
Figure 5: The Effect of Increasing Pension Eligibility Age on Own Retirement



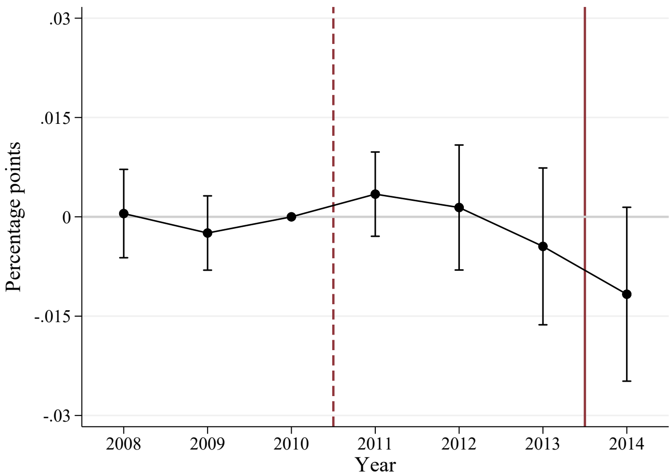
Notes: These figures plot the β_c coefficients from the dynamic difference-in-differences model (4), estimated on different outcomes for reference individuals. Each coefficient shows the difference between the treated group (whose pension eligibility age increases by 6 months, to age $60\frac{1}{2}$) 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 6: The Effect of Increasing Pension Eligibility Age on Spouses

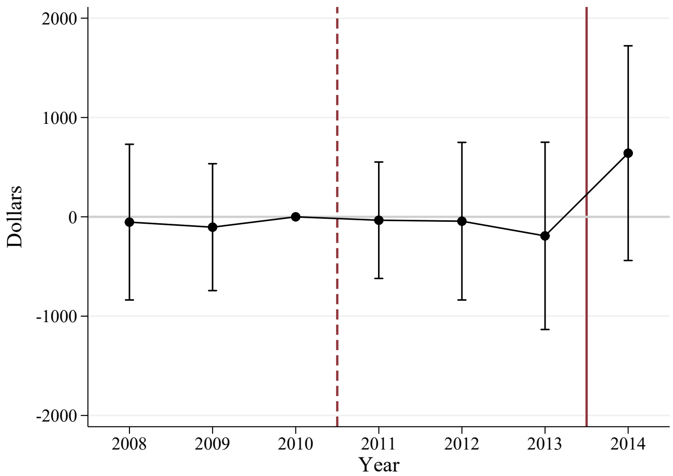
(a) Spouse Retirement



(b) Spouse Claiming



(c) Spouse Earnings



Notes: These figures plot the β_c coefficients from the dynamic difference-in-differences model from equation (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 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 3.3.

Table 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 (1). The second row reports the full-exposure effect on the spouses from their partners becoming eligible for pension, estimated in equation (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 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 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 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 1 and 3). Row A reproduces results from our baseline specification, which correspond to those reported in Table 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 6: The Effect of Increasing Pension Eligibility Age

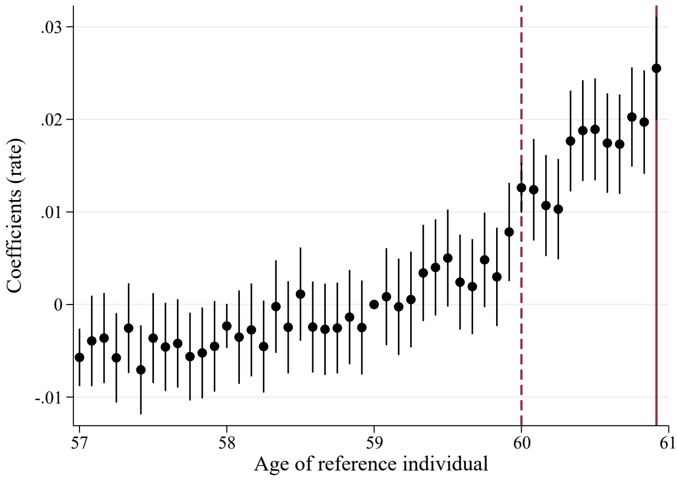
	Retirement	Claiming	Earnings
Reference Indiv.	-0.191***	-0.264***	8,140***
	(0.0074)	(0.0065)	(479)
Spouse	-0.0172**	-0.0110*	690
	(0.0073)	(0.0064)	(532)
Scaled Effect	0.0902**	0.0418*	0.0847
	(0.038)	(0.024)	(0.065)
F-test instr.	662.3	1643.6	288.8
N. of clusters	10,321	10,321	10,321
Observations	73,395	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 (5). The third row reports the scaled effect (the LATE) resulting from the 2SLS model estimated in equation (6). Robust standard errors in parentheses, clustered at the couple level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

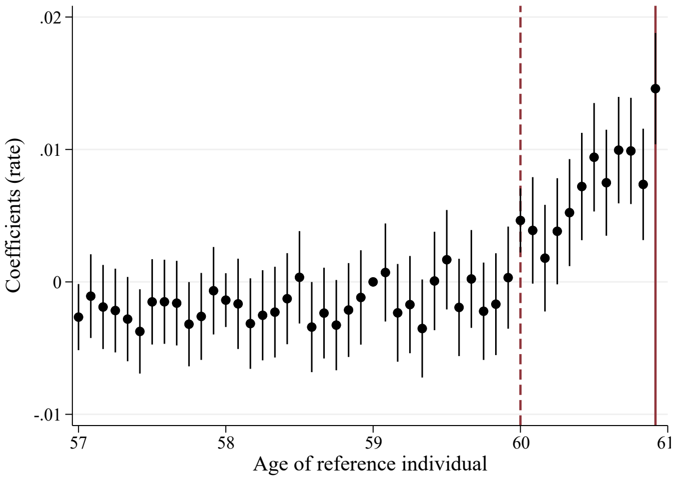
Appendix A Age Discontinuity Design

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

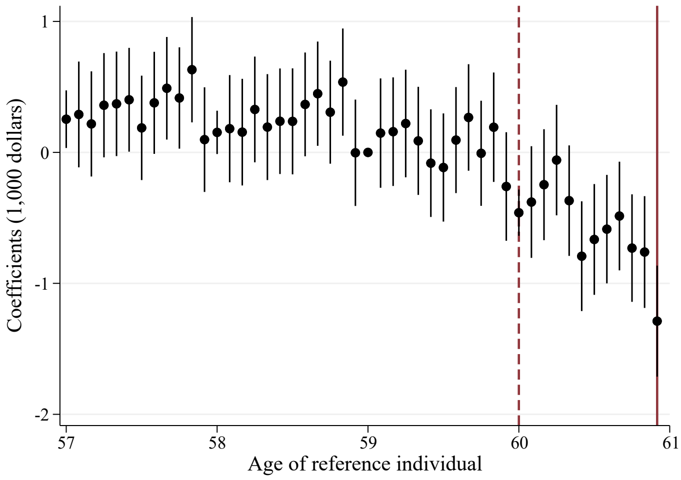
(a) Spouse Retirement



(b) Spouse Claiming



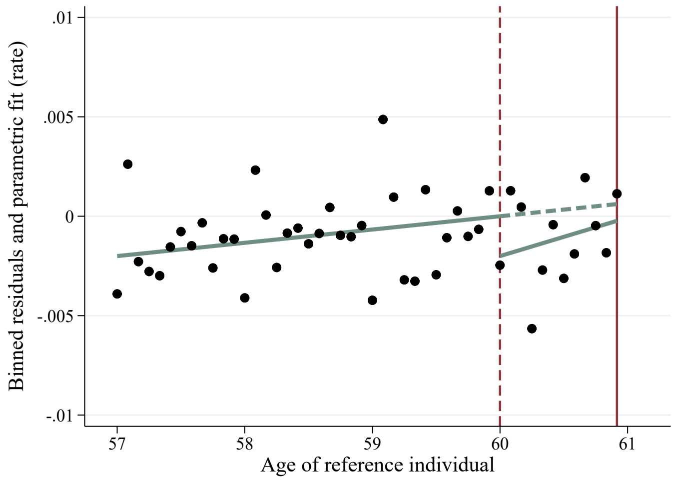
(c) Spouse Earnings



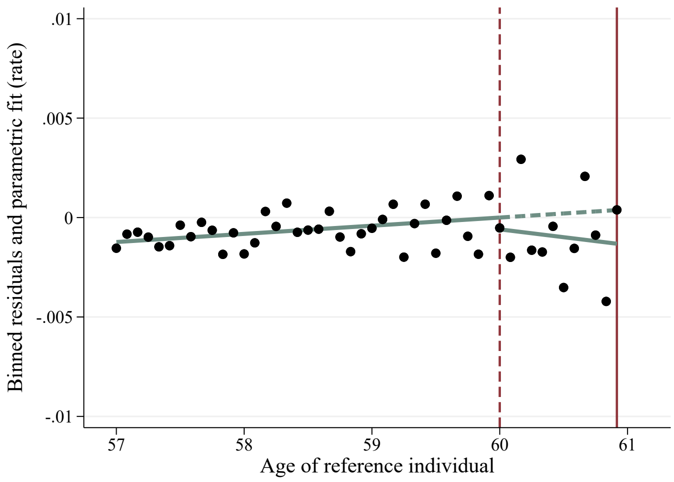
Notes: These figures show an alternative approach to obtain nonparametric evidence on spouses behavior around the pension eligibility age of their partner. They plot the δ_a^r coefficients from estimating the regression $y_{it}^s = \alpha + \sum_{a=57}^{62} \delta_a^r \cdot D_a^r + \sum_{a=49}^{69} \delta_a^s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

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

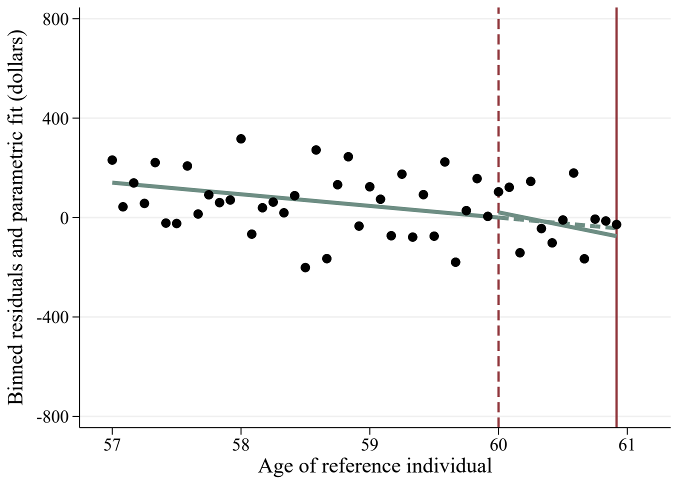
(a) Spouse Retirement



(b) Spouse Claiming



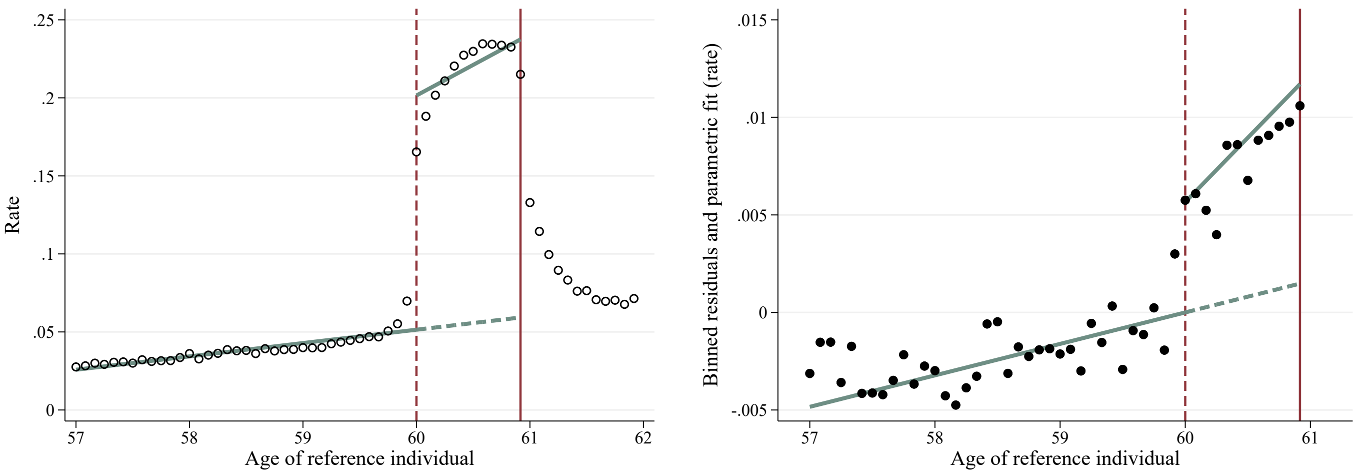
(c) Spouse Earnings



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. See Appendix Table A.3 for the placebo point estimates.

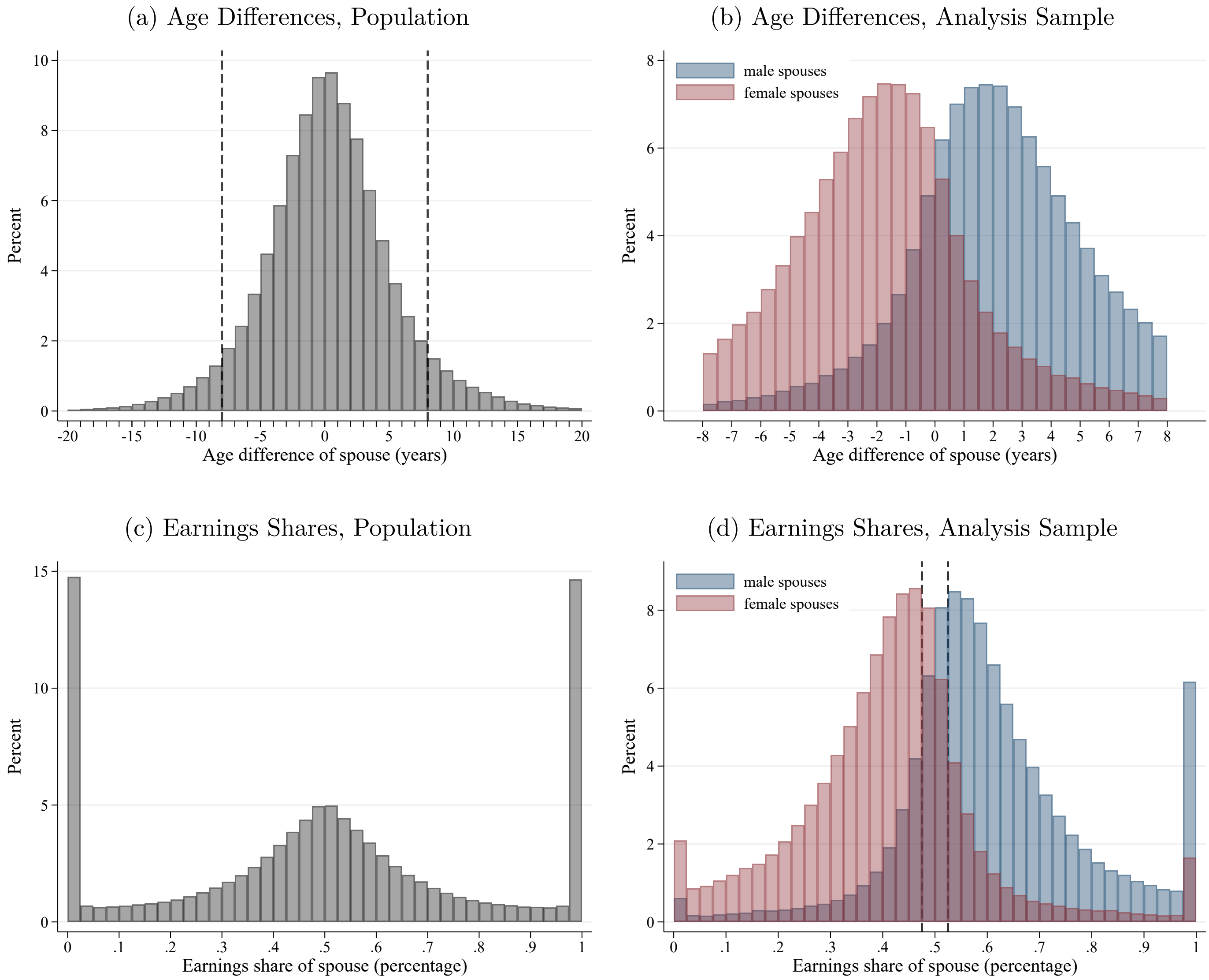
Figure A.3: The Effect of Reaching Pension Eligibility Age on Retirement Defined as Flow

(a) Retirement Flow (b) Spouse Retirement 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 1 and 2. The scaled effect estimate resulting from this outcome is reported in Table 5.

Figure 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 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 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 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 (1). The second row reports the full exposure effect on the spouses of their partners being eligible for retirement pension estimated in equation (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 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.

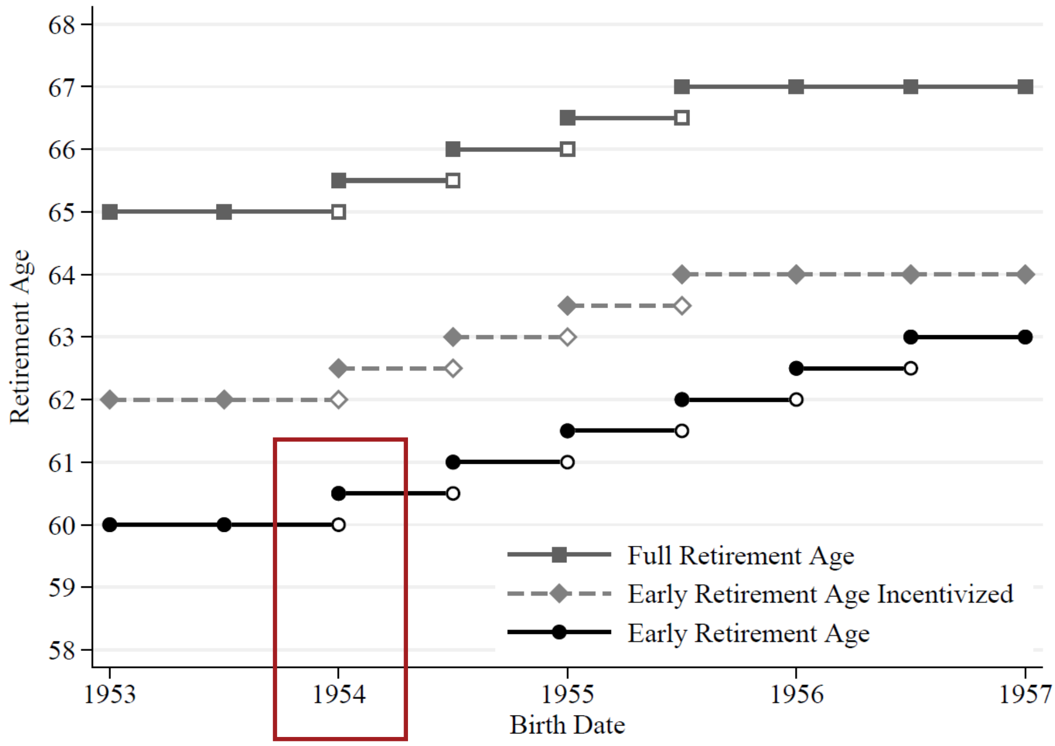
Table A.3: Placebo Test with Fake Spouses
for the Effect of Reaching Pension Eligibility Age

	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 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. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

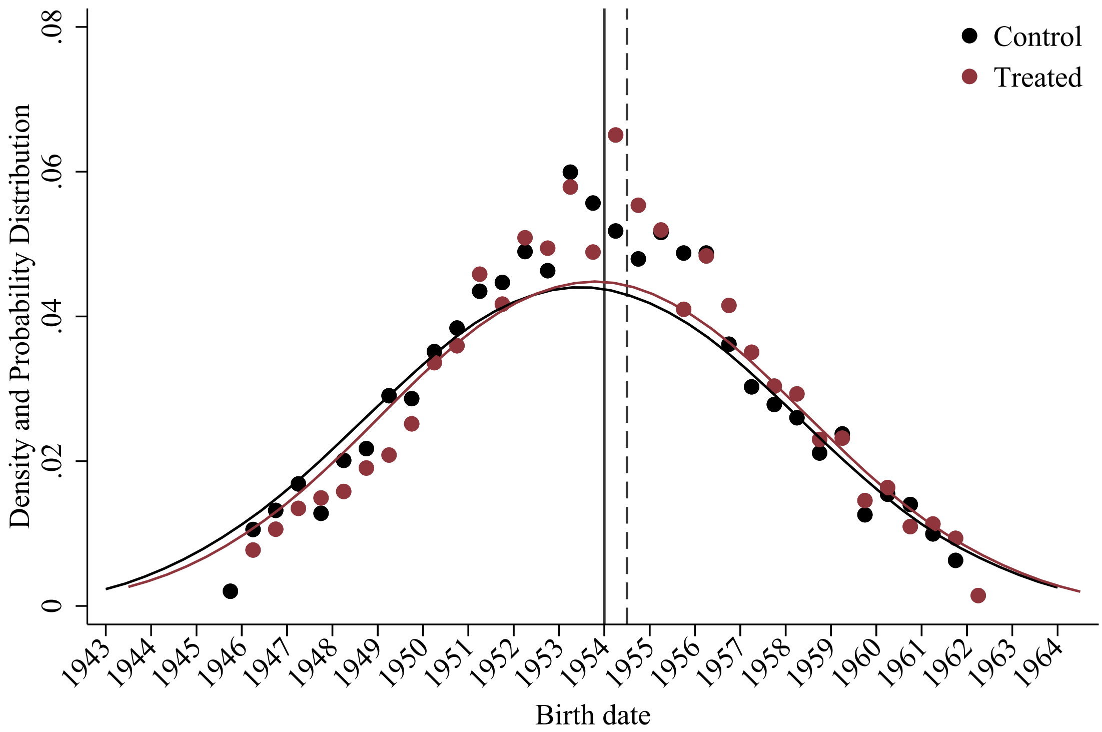
Appendix B Reform Discontinuity Design

Figure 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 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 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 (5). The third row reports the scaled effect (the LATE) resulting from the 2SLS model estimated in equation (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 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 6). Row A reproduces results from our baseline specification, which correspond to those reported in Table 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 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 6 for notes on the construction of this table.

Chapter II

Public Pensions and Private Savings

Public Pensions and Private Savings*

Esteban García-Miralles† Jonathan M. Leganza‡

Abstract

How does the provision of public pension benefits impact private savings? We answer this question in the context of a reform in Denmark that altered old-age benefit payouts through a discontinuous increase in pension eligibility ages contingent on birthdate. Using detailed administrative data and a regression discontinuity design, we identify the causal effects of the policy, leveraging our setting to study essentially the entire financial portfolio. We document responses over two distinct time horizons. First, we show a lack of responses after the reform was announced but before it was implemented, inconsistent with the notion that future differences in pension eligibility impact savings. 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.

Keywords: social security, private savings, pension reform
JEL codes: H55, D14, J26

*We thank our advisors, Gordon Dahl, Itzik Fadlon, Miriam Gensowski, and Mette Gørtz, for support and guidance, and we are grateful to Julie Cullen and Alex Gelber for valuable feedback. We also thank Jeff Clemens, Roger Gordon, Gaurav Khanna, Claus Thustrup Kreiner, Søren Leth-Petersen, Bruno Lopez-Videla, Torben Heien Nielsen, Mette Rasmussen, Benjamin Ly Serena, and Jakob Egholt Søgaard for helpful comments. Leganza gratefully acknowledges financial support from the NBER Pre-Doctoral Fellowship in Retirement and Disability Policy Research (under Director Nicole Maestas). García-Miralles gratefully acknowledges funding from the Novo Nordisk Foundation (grant no. NNF17OC0026542) and from the the Danish National Research Foundation through its grant (DNRF-134) to the Center for Economic Behavior and Inequality (CEBI) at the University of Copenhagen. The research reported herein 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 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 United States Government or any agency thereof.

†University of Copenhagen and CEBI. (email: egm@econ.ku.dk)

‡University of California, San Diego. Department of Economics. (email: jleganza@ucsd.edu)

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.

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

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

[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).

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 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. 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.

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.

The rest of this paper is organized as follows. Section 2 provides an overview of the institutional background. Section 3 grounds our empirical analysis with a conceptual framework and discusses the economic incentives. Section 4 describes the data. Section 5 lays out our identification strategy. Section 6 presents the main results, documenting the causal effects of the reform. Section 7 investigates underlying mechanisms and discusses potential explanations for our findings. We conclude in Section 8.

[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).

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 B.

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 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.

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%

[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.

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.

2.2.1 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

[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.

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 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.

2.2.2 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.

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 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]

[10] Satisfying the two-year rule results in a modest increase in base benefit amounts as well, to approximately \$29,600, as benefits become linked to 100% (rather than 91%) of maximum UI benefits. See Appendix 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.

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.

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:

$$\begin{aligned} \max_{R, \{c_s\}_{s=0}^{R}} \quad & \int_0^R e^{-\rho t} u(c_t) dt + \Psi(a_R + B(R), R) \\ \text{s.t.} \quad & \dot{a}_t = r a_t + y_t - c_t \\ & a_0 = 0, \end{aligned} \tag{1}$$

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

$$\begin{aligned} \Psi(a_R + B(R), R) = \max_{\{c_s\}_{s=R}^T} \quad & \int_R^T e^{-\rho t} \left(u(c_t) + \Gamma \right) dt \\ \text{s.t.} \quad & \dot{a}_t = r a_t - c_t \\ & a_T = 0. \end{aligned} \tag{2}$$

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

$$\frac{u''(c_t)}{u'(c_t)} \dot{c}_t = \rho - r. \tag{3}$$

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:

$$c_t = c(Y(R), B(R)) = \frac{C^L}{T}, \tag{4}$$

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:

$$(y + B'(R)) \cdot u'(c_R) = \Gamma. \tag{5}$$

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.

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 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]

The spikes in pension wealth at the critical ages translate to large discontinuities in lifetime consumption, C^L. Graph (a) of Figure 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 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

[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 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 (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 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.

retire at either exactly the ERA or exactly two years after the ERA shape labor supply decisions of older workers.

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 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 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 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 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.

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.

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 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 in-

[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.

dicator 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 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

[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 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 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.

DKK to 1 USD.

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 1 presents summary statistics for calendar year 2010, the year before the reform is announced. Columns (1) and (2) 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.

5 Identification Strategy

[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.

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:

$$y_i = \alpha + \beta \cdot 1[x_i \geq c] + f(x_i - c) + 1[x_i \geq c] \cdot g(x_i - c) + Z_i\theta + \varepsilon_i, \tag{6}$$

where y_i is an outcome variable for individual i (such as contributions to retirement savings 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 6.3. In particular, we vary the bandwidth, drop the triangular weights, exclude controls, and estimate global linear polynomials in the running variable.

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

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

[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.

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 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 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 (6) without any covariates on the right-hand side, instead using each control variable as a left-hand side outcome variable. Appendix Table A.1 presents these results. There are no statistically significant discontinuities in any of the control variables at the cutoff.

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.

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 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 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 2 presents results from corresponding regression analyses. We report in the table RD estimates of β from estimating equation (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 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 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.

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 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 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 4.

Graph (c) of Figure 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 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 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 A.2 plots the RD estimates and confidence intervals from estimating equation (6) on indicators for the various contribution amount bins. The point estimate furthest to the left mirrors the result in graph (d) of Figure 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 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 3 reports RD estimates for the other savings outcomes we study.[27] None of the estimates are statistically distinguishable

[26]Recall from Section 4 that we unfortunately cannot distinguish between distributions from employer-sponsored and personal accounts.

[27]Results from analyzing indicators for contributing to Roth-style accounts, which were first introduced

from zero. The second row of Appendix Table 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 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 7, and regression estimates for this year are reported in column (3) of Table 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 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 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 A.6) are statistically indistinguishable from zero.

Finally, in columns (3) and (4) of Table 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]

to the economy in 2013, are reported in Appendix Table 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.

[28]The point estimates in 2017 and 2018 for changes in bank account balances are fairly large (around

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.

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 A.3 (for the anticipation period), Appendix Table A.4 (for critical year 2014), and Appendix Table 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 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.

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 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 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

$600) but imprecisely estimated and statistically insignificant; additional winsorizing yields smaller point estimates (see Appendix Table A.6).

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 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.

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.

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

[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).

reform is regarded as an initial push towards the gradual elimination of the VERP program altogether.[30] Overall, we view our setting as one in which general awareness was likely high. For some reference, Appendix Figure 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 5. Column (1) shows no evidence of any anticipatory savings responses in any of the savings vehicles we study for this subsample.

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.

[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 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.

7.2.1 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 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.

7.2.2 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 8 depicts the results. Each graph plots the distribution of deviations from default contribute rates. For example, the large spikes around zero in graph (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 contri-

[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.

bution 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 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.

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.

References

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.

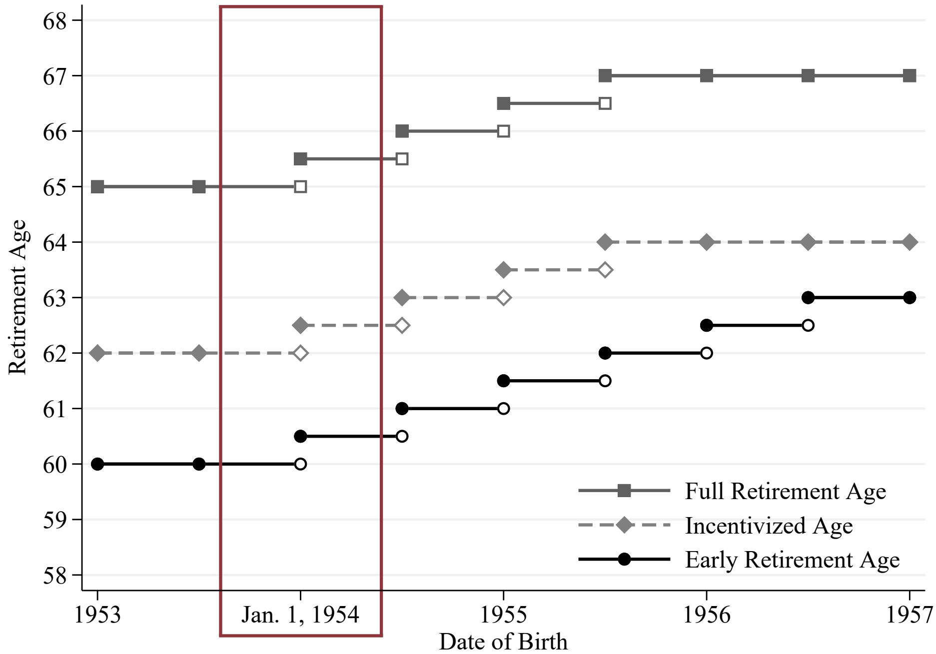
Pozo, S. and S. A. Woodbury (1986). Pensions, social security, and asset accumulation. *Eastern Economic Journal 12*(3), 273–281.

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.

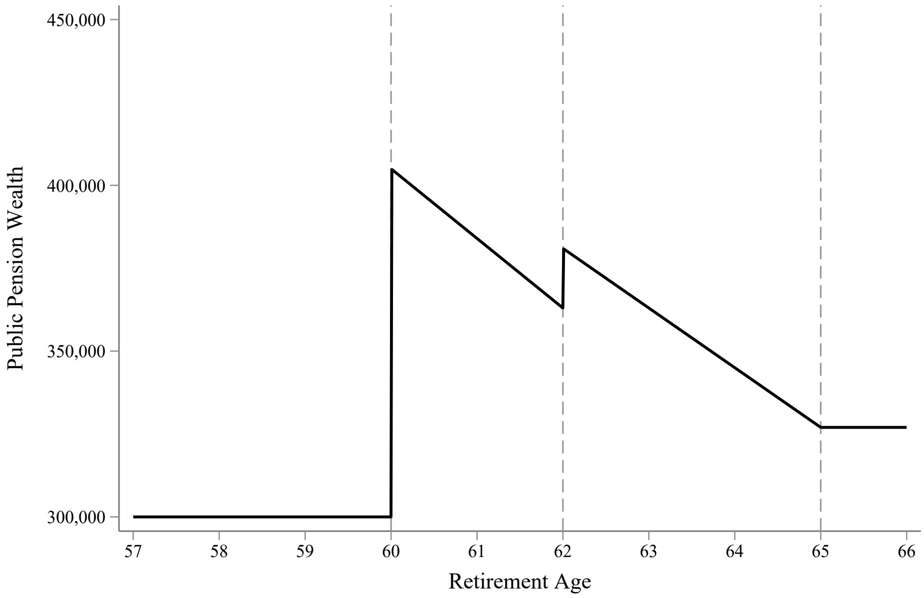
Staubli, S. and J. Zweimüller (2013). Does raising the early retirement age increase employment of older workers? *Journal of Public Economics 108*, 17–32.

Figure 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.

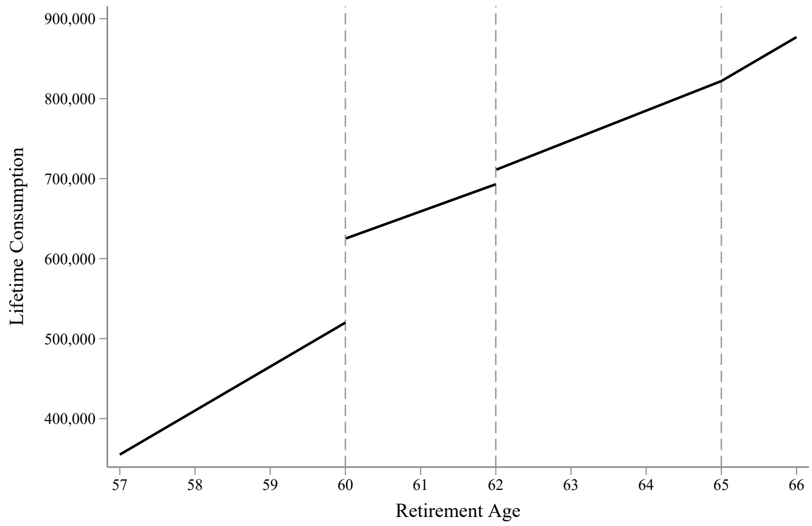
Figure 2: Pre-Reform Public Pension Wealth by Retirement Age



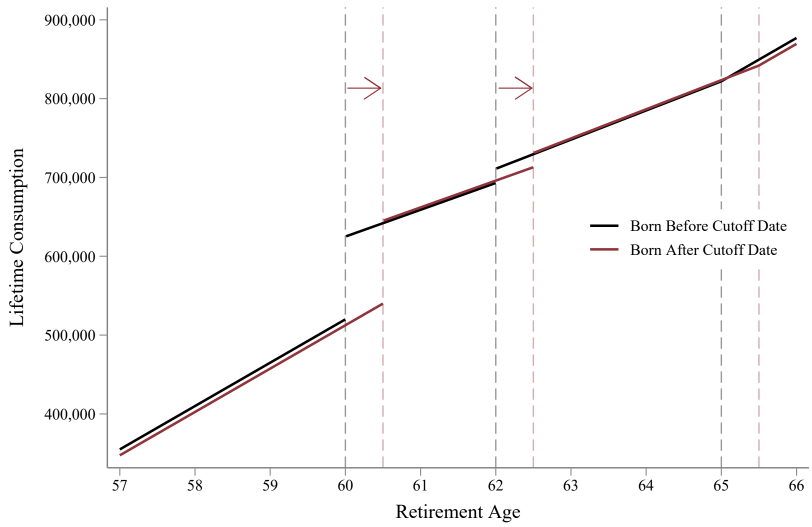
Notes: This figure plots public pension wealth against retirement age for a representative individual before the reform. For illustrative purposes, the benefit amounts depicted in the figure are for a worker who is married, who lives until age 85, and who has \$250,000 in private retirement savings accounts 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 B).

Figure 3: Lifetime Budget Constraints

(a) Pre-Reform Budget Constraint



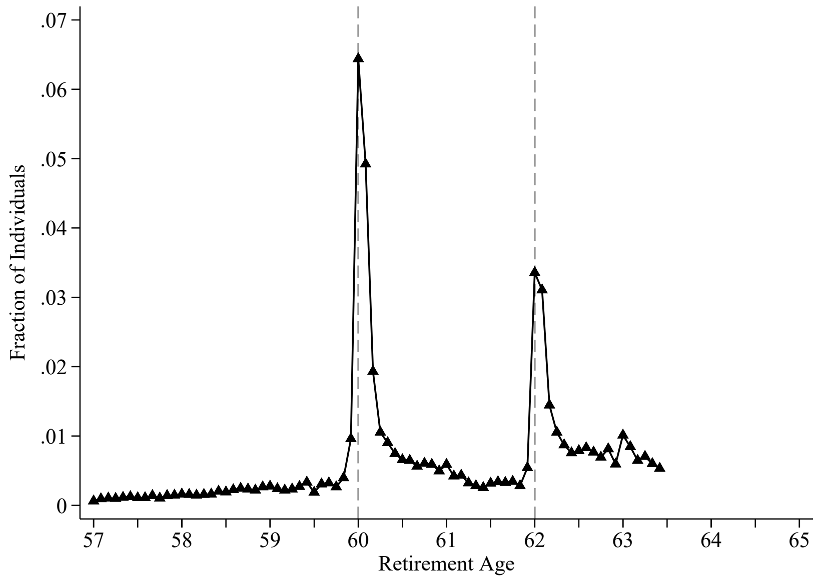
(b) Post-Reform Budget Constraints



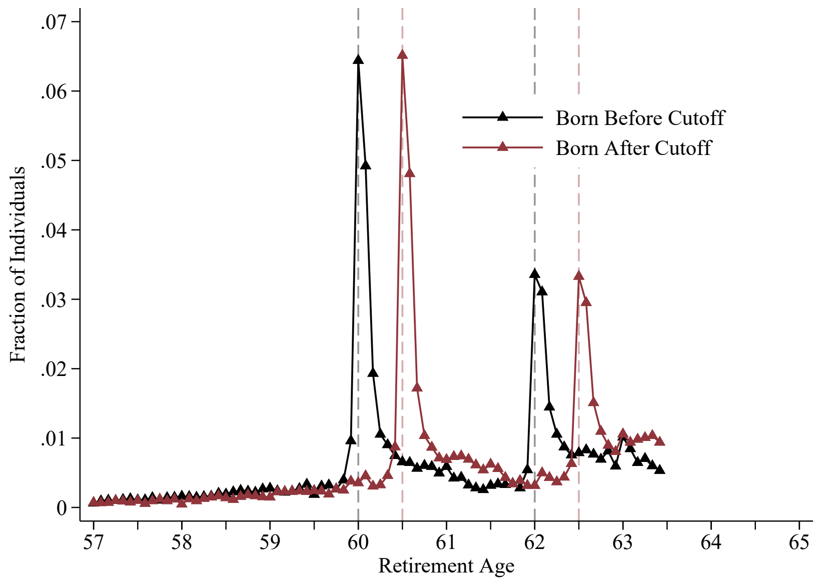
Notes: This figure plots lifetime consumption against retirement age for the same representative worker as in Figure 2. Lifetime consumption is the sum of public pension wealth and lifetime earnings. For illustrative purposes, annual earnings are assumed to be $55,000 and lifetime earnings are earnings 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.

Figure 4: Empirical Distributions of Retirement Ages

(a) Retirement Distribution for the Control Group



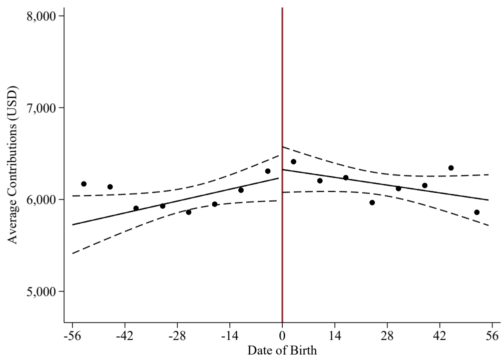
(b) Retirement Distributions for Treatment and Control Groups



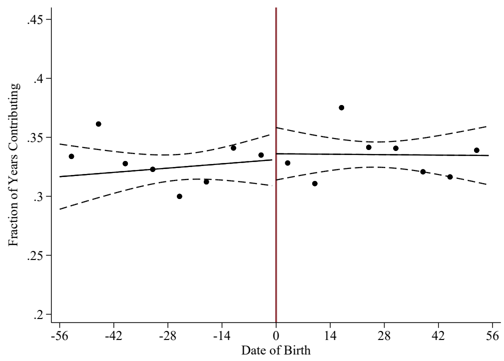
Notes: This figure plots empirical distributions of retirement ages. Retirement is measured as an absorbing state. Monthly retirement age is defined as the age of the individual in the last month during which earnings are positive, before permanently falling to zero. Graph (a) shows how those born before 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 5: Responses Over the Anticipation Period

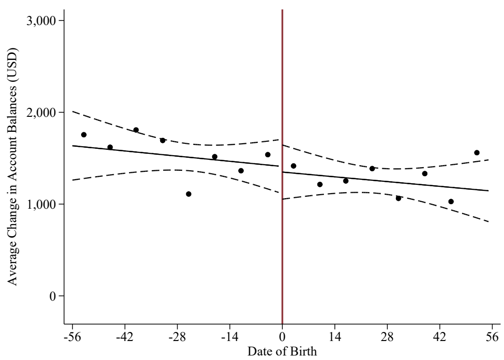
(a) Employer-Sponsored Plans



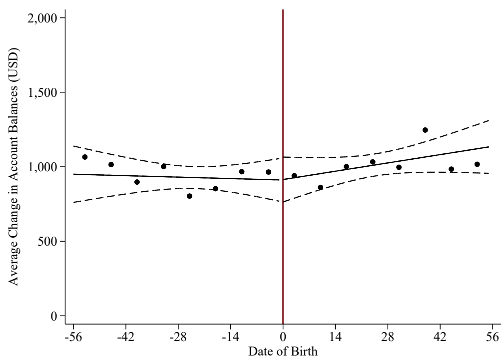
(b) Personal Plans



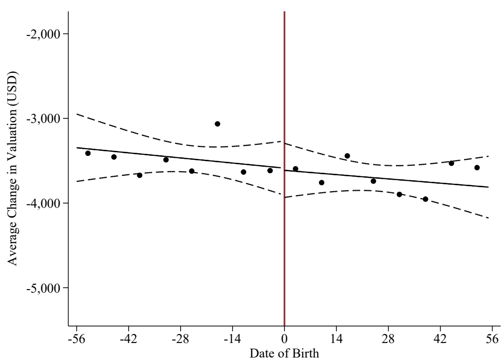
(c) Bank Accounts



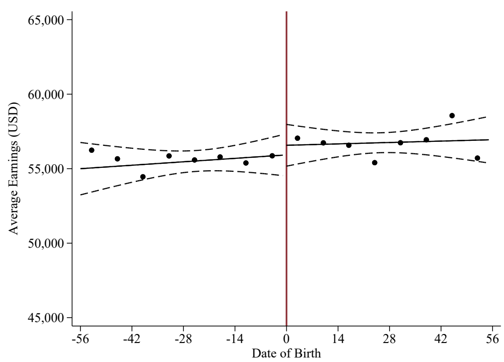
(d) Stocks



(e) Property

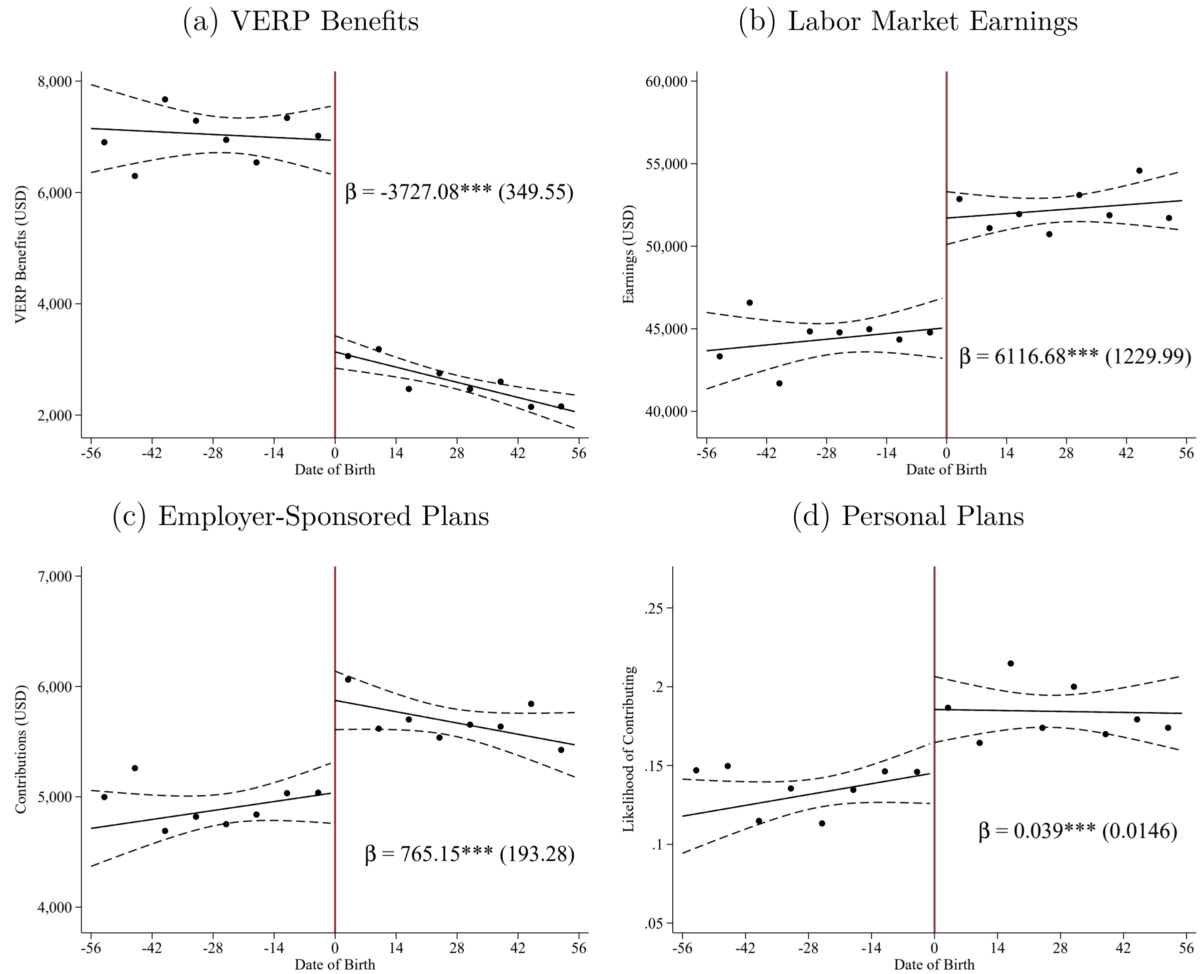


(f) Labor Market Earnings



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.

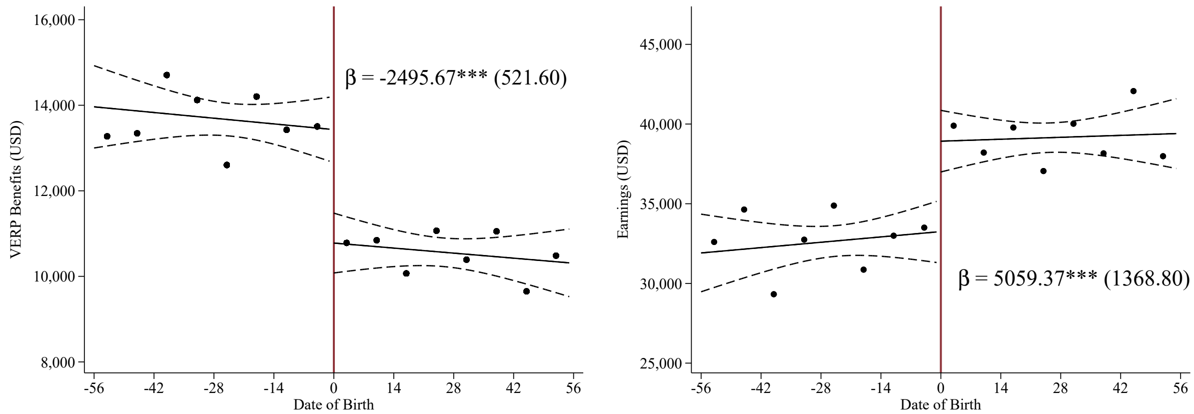
Figure 6: Responses During the First Critical Year 2014



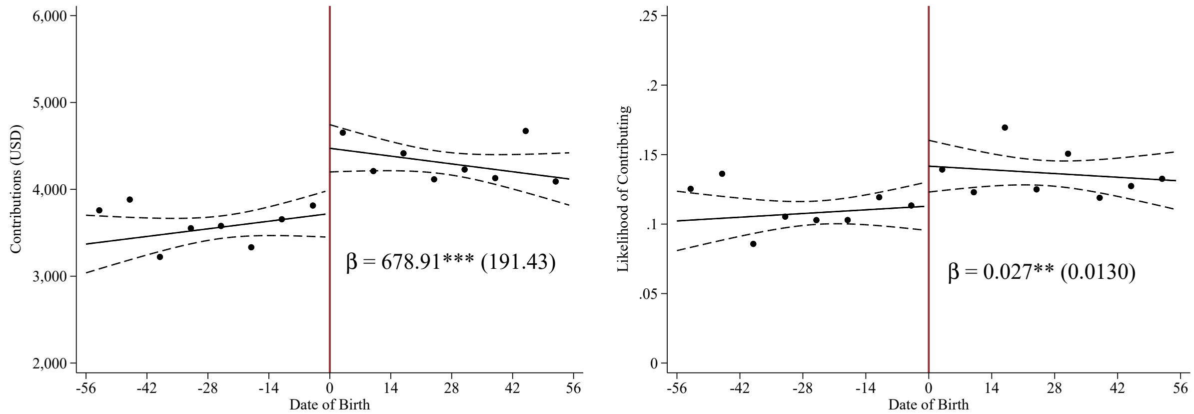
Notes: This figure illustrates the effect of the reform on labor market outcomes and contributions to retirement accounts during the first critical year, when individuals born at the cutoff date are age 60. Each RD graph (a)–(d) plots average outcomes during 2014 in one-week date-of-birth bins. The maroon 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 3, and come from estimating equation (6).

Figure 7: Responses During the Second Critical Year 2016

(a) VERP Benefits (b) Labor Market Earnings



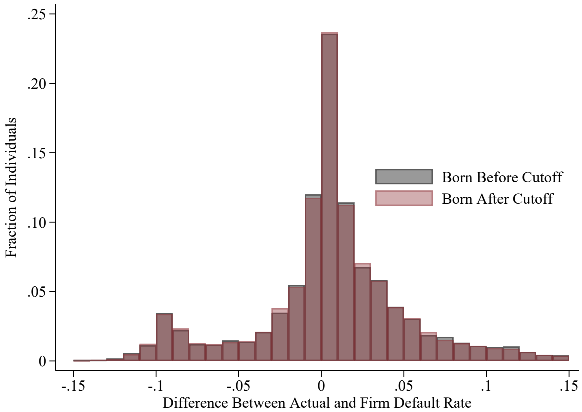
(c) Employer-Sponsored Plans (d) Personal Plans



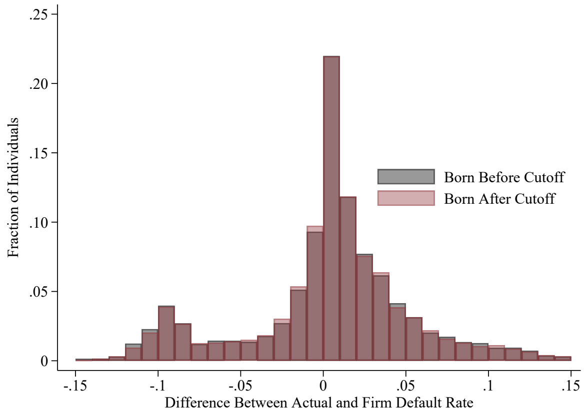
Notes: This figure illustrates the effect of the reform on labor market outcomes and contributions to retirement accounts during the second critical year, when individuals born at the cutoff date are age 62. Each RD graph (a)–(d) plots average outcomes during 2016 in one-week date-of-birth bins. The 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 3, and come from estimating equation (6).

Figure 8: Differences Between Actual and Firm Default Contribution Rates

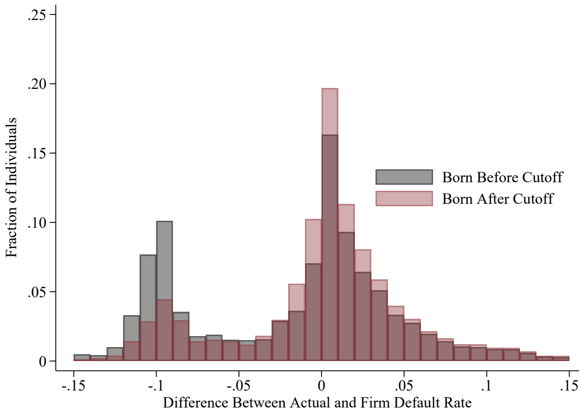
(a) Pre-Announcement Period: Year 2010



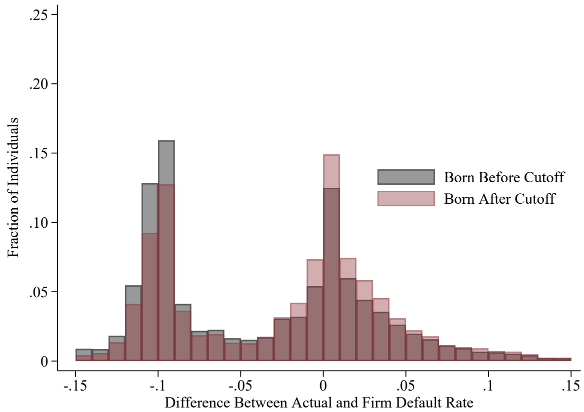
(b) Anticipation Period: Year 2012



(c) First VERP Critical Year: 2014



(d) Second VERP Critical Year: 2016



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 7.2.2. Each graph (a)-(d) captures the distributions of deviations from firm default rates during a different year.

Table 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 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 (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 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 (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 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 (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 5: Anticipatory Responses for Users of Personal Retirement Plans

	RD Estimate (1)	Mean (2)
A: Labor Supply		
Earnings	-84.24 (1486.16)	56,739
B: Retirement Accounts		
Contributions to Employer Plans	319.51 (265.25)	5,962
Any Contribution to Personal Plans	0.001 (0.019)	0.71
C: Other Savings		
Change in Bank Accounts	68.07 (347.15)	1,554
Change in Stock Market Accounts	70.29 (174.67)	1,157
Change in Property Wealth	115.99 (344.96)	-3,712
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 (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 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 (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. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.1$

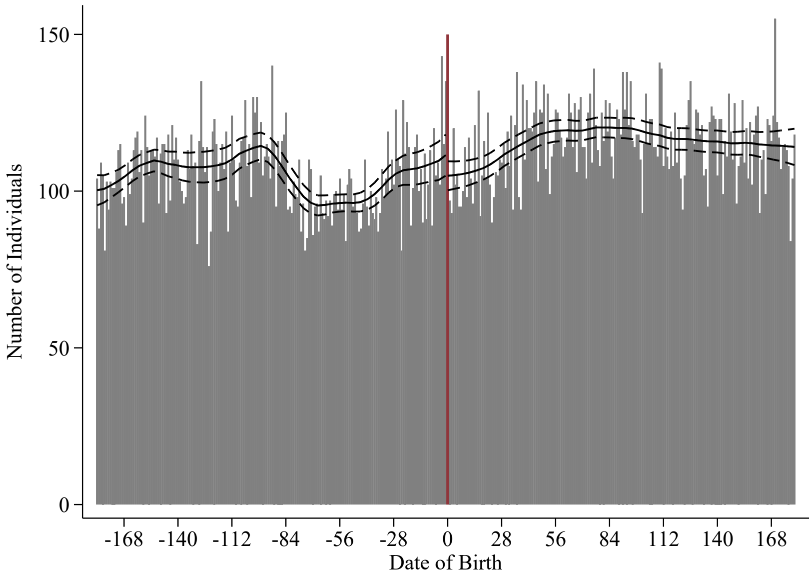
Table 7: Actual vs. Predicted Contributions to Employer Retirement Plans

	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 7.2.2. The RD estimates come from estimating equation (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. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.1$

Appendix A Additional Figures and Tables

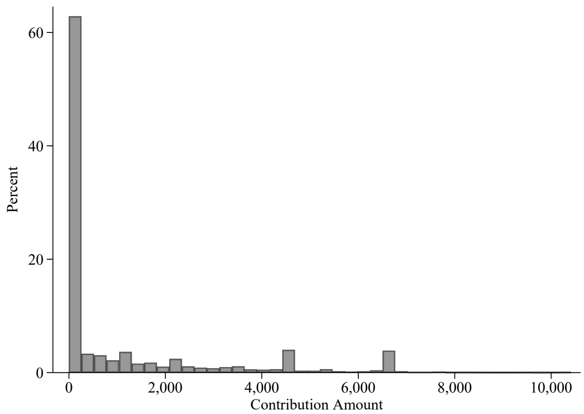
Figure A.1: Histogram of the Running Variable



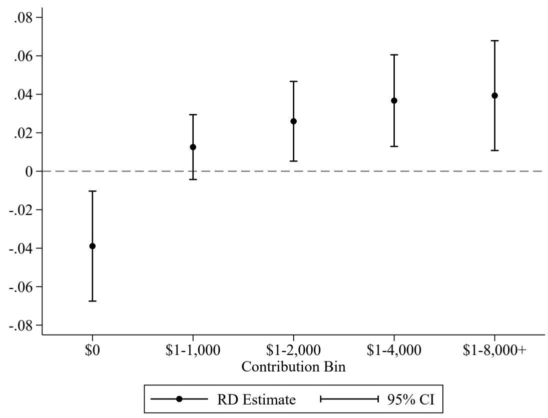
Notes: This figure depicts the density of the running variable, birthdate. The graph plots a histogram of the running variable for the entire analysis sample. Superimposed on top of the histogram are smoothed values and confidence intervals from local polynomial regressions of the number of individuals 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 A.2: Analyzing Contribution Amounts to Personal Retirement Plans

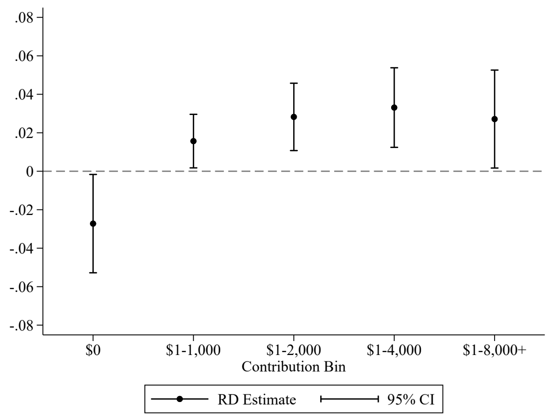
(a) Unconditional Distribution



(b) RD Estimates: 2014



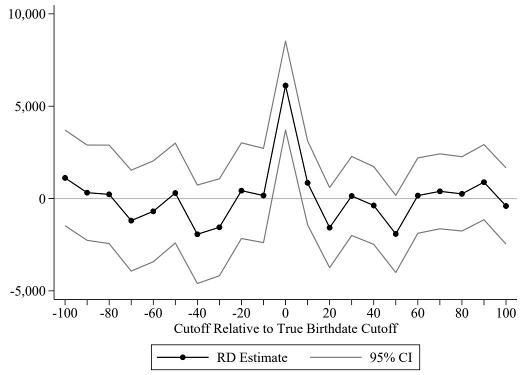
(c) RD Estimates: 2016



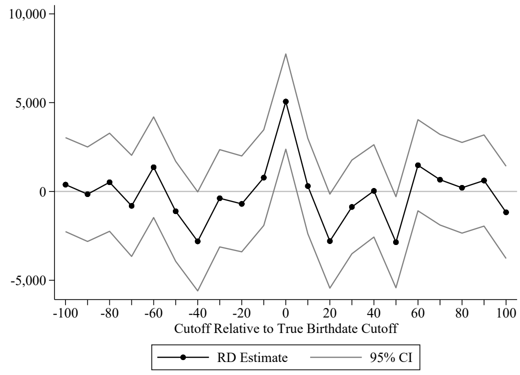
Notes: This figure illustrates the method of analyzing contribution amounts to personal retirement plans. Graph (a) plots the unconditional distribution of contribution amounts in 2010. The large number of small and zero contributions show why analyzing average contributions in levels is difficult. We 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 (6) using as outcomes these indicator variables in 2014. Graph (d) plots the results for 2016.

Figure A.3: Placebo Exercise: Pseudo Birthdate Cutoffs

(a) Labor Market Earnings: Year 2014

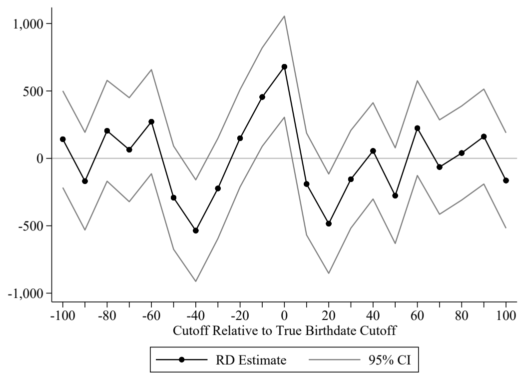


(b) Labor Market Earnings: Year 2016

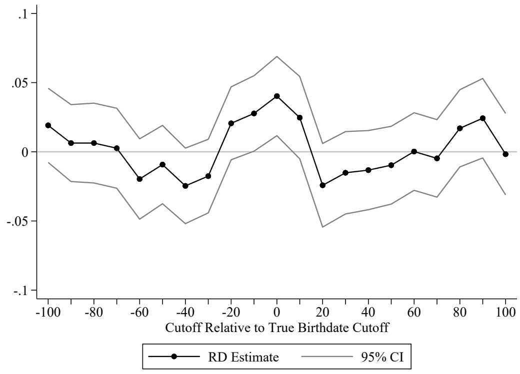


(c) Employer Plans: Year 2014

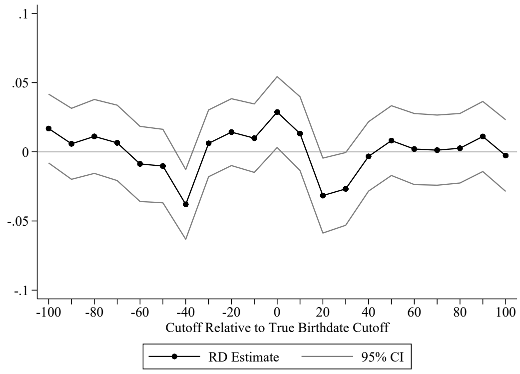
(d) Employer Plans: Year 2016



(e) Personal Plans: Year 2014

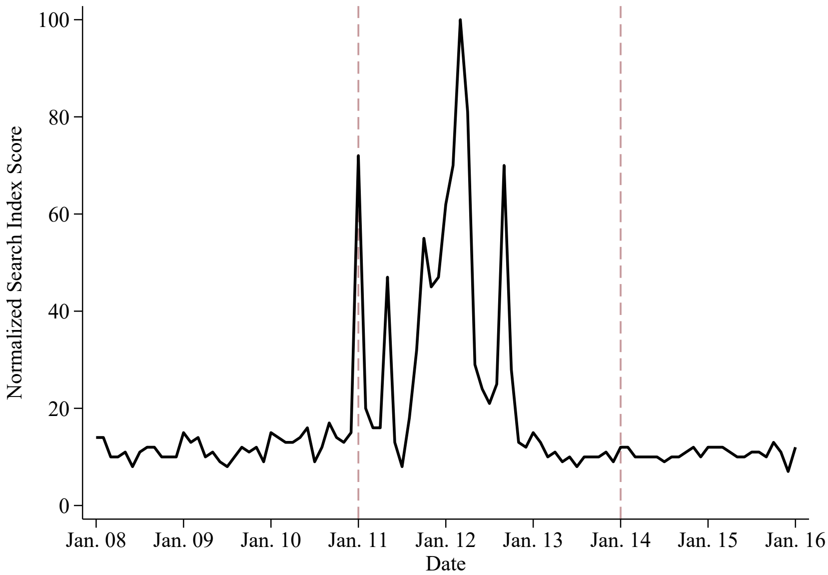


(f) Personal Plans: Year 2016



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 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 A.1: RD Estimates for Control Variables as Outcomes

	RD Estimate (1)	Mean (2)
Male	0.026 (0.020)	0.47
Married	0.018 (0.018)	0.69
Hovedstaden	-0.003 (0.013)	0.12
Sjælland	-0.010 (0.017)	0.25
Syddanmark	-0.005 (0.017)	0.24
Midtjylland	0.022 (0.017)	0.24
Nordjylland	-0.005 (0.014)	0.15
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 (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 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 (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 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 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 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 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 (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 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 (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 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 (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 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 (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. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Appendix B 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.

B.1 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.

B.2 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.

B.3 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.

Chapter III

The Crucial Role of Social Welfare Criteria and Individual Heterogeneity for Optimal Inheritance Taxation

The Crucial Role of Social Welfare Criteria and Individual Heterogeneity for Optimal Inheritance Taxation *

Esteban García-Miralles †

Abstract

This paper extends the calibrations of Piketty and Saez (2013) to unveil the importance of the assumed social welfare criteria and its interplay with individual heterogeneity on optimal inheritance taxation. I calibrate the full social optimal tax rate and find that it is highly sensitive to the assumed social welfare criteria. The optimal tax rate ranges from negative (under a utilitarian criterion) to positive and large (even assuming joy of giving motives). A decreasing marginal utility of consumption does not affect the results qualitatively given the underlying distribution of wealth and income. I also calibrate the optimal tax rate by percentile of the distribution of bequest received, as in Piketty and Saez, but accounting for heterogeneity in wealth and labor income. This leads to significant variation in the optimal tax rate among zero-bequest receivers, contrary to their finding of a constant tax rate.

JEL codes: H21, H23, D31, D63

Keywords: Optimal inheritance taxation, social welfare function, wealth inequality

*I would like to thank Claus T. Kreiner for his guidance and insightful discussions. This work also benefited from the advice kindly provided by Bertrand Garbinti for the replication analysis. I also thank Pol Campos, José María Casado, Laura Crespo, Miriam Gensowski, Mette Gørtz, Essi Kujansuu, José María Labeaga, Lourdes Moreno, Jorge Onrubia, Roberto Ramos, Ernesto Villanueva and seminar participants at Banco de España, at the 43rd Symposium of the Spanish Economic Association (SAEe, Madrid), at the XX Applied Economics Meeting (Valencia) and at the XXIV Encuentro de Economía Pública (Toledo) for useful comments and suggestions. I gratefully acknowledge financial support from Fundación Ramón Areces. The activities of CEBI are financed by a grant from the Danish National Research Foundation. Grant number #DNRF134

†García-Miralles: Department of Economics, University of Copenhagen, and Center for Economic Behavior and Inequality (CEBI), Øster Farimagsgade 5, 1353 Copenhagen K, Denmark. E-mail: egm@econ.ku.dk

1 Introduction

Taxation of wealth is currently at the center of many academic and political debates. For the case of inheritance taxation, policy makers are discussing reforms across many European countries, and in the U.S. the estate tax has been modified almost every year since 2001, currently operating with a 40% top marginal rate. This paper presents a positive analysis of two crucial features that underlie the design of optimal inheritance taxation, namely the assumed preference for redistribution (the social welfare function —SWF—) and the large variation across individuals regarding their preferred optimal tax rate (the underlying individual heterogeneity).

Most studies on inheritance taxation assume a utilitarian SWF. While this is a standard approach in the literature of optimal taxation, it has important consequences, as noted by Fleurbaey and Maniquet (2018) for the case of labor income taxation. I show that inheritance taxation is particularly affected by this assumption due to the interaction between the positive externalities that can arise from joy of giving bequest motives and the high concentration of bequests at the top of the distribution.

The model derived by Piketty and Saez (2013) —henceforth PS13— allows for different SWFs, which can be used to calibrate the optimal tax rate under different social welfare criteria. However, they opt for calibrating the optimal tax rate from the perspective of each percentile of the distribution of bequest received rather than the full social optimum under standard social welfare criteria. While their approach is informative of the role of heterogeneity in bequests received on inheritance taxation, it does not result in a single tax rate applicable to the entire population, and it does not fully capture heterogeneity in wealth and labor income.

This paper presents two contributions. First, I show that different assumptions on the SWF lead to very different full social optimal tax rates due to the high concentration of bequests at the top of the distribution and the existence of positive externalities. To do so, I revisit the model of PS13 and calibrate their optimal tax formula for the U.S.

under three different standard social welfare criteria.[1] I obtain that under a utilitarian criterion the optimal tax rate is always negative, even with fully accidental bequests. Under the responsibility and compensation criterion, the optimal tax rate is positive and very sensitive to other parameters of the model, particularly to the bequest elasticity. Under a Rawlsian criterion the optimal tax rate is positive and large, solely limited by the bequest elasticity. Interestingly, the concavity of the individual utility function does not have a qualitatively important impact on the optimal tax rate, due to offsetting effects from the underlying distribution of wealth and labor income and the trade-off between bequest and labor taxes.

Second, I extend the calibration by percentile of the distribution of bequest received to include heterogeneity in wealth and labor income. I find that the optimal tax rate for those who do not receive any bequests (70% of the population) varies significantly, from an 83.3% tax rate for the worst-off individuals to negative tax rates for those who, despite not having received any bequest, have accumulated wealth through high labor income. This result differs from the one obtained by PS13, in which the tax rate remains fairly constant around 50% for all zero-bequest receivers.

Altogether, these results show that the optimal inheritance tax rate depends heavily on the assumed SWF and the underlying distribution of bequests, income, and wealth. These two findings are crucial for the design of optimal inheritance taxes. Policy makers must account for the effect of different SWFs and the utilitarian framework is not a neutral benchmark. The percentile calibrations show a large variation on the optimal tax rate from the individual point of view. This helps to explain the public debate around inheritance taxation given the large heterogeneity in individual preferences and highlights its dependence on the social planner's welfare function.

The paper proceeds as follows: Section 2 introduces the literature on inheritance taxation with a focus on the assumed bequest motives and social welfare criteria. Section 3 summarizes the model of Piketty and Saez (2013). Section 4 presents the results from calibrating the full social optimal tax rate under standard social welfare criteria. Section

[1] PS13 (p.15 of supplementary material) write: "It would be interesting to use our estimates to compute the full social optimum implied by various SWFs ..."

5 presents the results from calibrating the optimal tax by percentiles accounting for heterogeneity in wealth and labor income. Section 6 concludes.

2 Review of the literature

The study of optimal inheritance taxation needs to account for two relevant features of inheritance taxation. This section presents an overview of how they have been addressed in the literature. The first feature is the bequest motive, that is, the motivation for the donor to leave a bequest. With *altruistic* motives, donors care about the lifetime utility of their heirs and therefore internalize the effects of bequests on the donees. Under *joy of giving* motives, the donors' utility function depends on the after-tax bequest left, but not on the utility of the donees, which can lead to a positive externality because donors do not internalize the effect of their actions on the donees.[2] Finally, *accidental* motives lead to unplanned bequests and in this case the tax rate has no effect on the donors' utility.[3]

A second crucial feature for the study of optimal inheritance taxation is the assumption imposed on how individual utilities are weighted in the SWF. Frequently a utilitarian criterion is assumed. This turns out to be particularly relevant due to the high concentration of bequests at the top of the distribution and the presence of externalities of giving that increase proportionally with the amount bequeathed. Hence, even small variations in the social weights of individuals at the top of the distribution can cause significant changes in the optimal tax.

These two features are unremarked in the most prominent results of the literature. For example, the model of Atkinson and Stiglitz (1976) has been extrapolated to the study of inheritance taxation reinterpreting consumption of different commodities as consumption at different points in time, and taxation of future consumption as a tax on bequests, which should therefore be zero. This model implicitly assumes joy of giving bequest motives because it is the bequests left, and not the utility of the heirs, that enters the utility of the first generation. The social planner of this model maximizes a utilitarian SWF.

[2] This 'externality of giving' differs from a standard atmospheric externality because it is interpersonal, requiring differentiated Pigouvian taxes.

[3] Kopczuk and Lupton (2007) estimate that over 30% of bequests are accidental.

Chamley (1986) and Judd (1985) study capital taxation using an infinite-life model, measuring social welfare from the first generation. They assume altruistic bequest motives and since it is a representative agent model, the implicit SWF is utilitarian. They conclude that the optimal tax rate is zero, however Straub and Werning (2020) have overturned this result, obtaining a positive tax rate.

Farhi and Werning (2010) extend the model of Atkinson and Stiglitz (1976) to explicitly model inheritance taxation considering two generations. The first generation of donors have joy of giving motives and starts with no wealth inequality but heterogeneous productivity, so that the inheritance received by the second generation and labor inequality are perfectly correlated. The second generation only consumes what they inherited and do not work. If the social planner (with a utilitarian SWF) only considers the utility of the first generation, the optimal tax rate is zero. However, when the utility of the second generation is included in the social welfare the optimal inheritance tax rate becomes negative to correct for the positive externality caused by joy of giving motives.

Cremer and Pestieau (2011) use an overlaping generations model based on Diamond (1965) and extend it to model inheritances, showing how the optimal inheritance tax rate depends on the bequest motives. If bequests are fully accidental, a tax rate of 100% is optimal. If bequest motives are altruistic, the utility function of the representative individual fully captures the utility of next generations internalizing the positive externality of giving. In this case, the optimal tax rate in the long run is zero. With joy of giving motives, the positive externality appears and the optimal tax rate is negative. Note, however, that in all these cases the SWF is utilitarian.

Brunner and Pech (2012a) and Brunner and Pech (2012b) improve upon previous models by including initial wealth inequality. They find that the optimal tax can increase social welfare if initial wealth and earning abilities are correlated. They consider altruistic and joy of giving motives, but control for double-counting of utilities between generations. They assume that the SWF puts more weight on low ability individuals to ensure a preference for redistribution but the implications of this assumption are not further explored.

Recent contributions to the literature of inheritance taxation emphasize the labor supply response of inheritors. Kopczuk (2013) finds that an increase on bequests will reduce total labor supply and revenue from labor income taxes, generating a negative 'fiscal externality' that can be counteracted with a tax on bequests. In this line, Kindermann et al. (2018) further develop and calibrate a life-cycle model that accounts for the labor supply of heirs and find sizable 'fiscal externalities'. Both models assume joy of giving motives and utilitarian SWF.[4]

Closely related to this paper is the model of Farhi and Werning (2013), which introduces heterogeneity in altruistic motives and also considers different SWFs. They find that "optimal estate taxes depend crucially on redistributive objectives. Different welfare criteria lead to results ranging from taxes to subsidies" (p.490). Their results therefore constitute a theoretical basis for the empirical calibrations that I present here, based on the model of PS13.

3 The model of Piketty and Saez

The model of PS13 contributes to the literature allowing for alternative SWFs and for a combination of bequest motives. The authors present a dynamic stochastic model with a discrete set of generations that do not overlap, with heterogeneous bequest tastes and labor productivities. There is labor augmenting economic growth at rate $G > 1$ per generation. The government has a given budgetary need E that is financed with linear taxes on labor income at rate τ_{Lt} and on capitalized bequest at rate τ_{Bt}. This revenue is then equally distributed across individuals as a lump-sum grant per individual, E_t.

Each individual, ti, lives in generation t and belongs to dynasty i. Each receives a pre-tax bequest b_{ti} that earns an exogenous gross rate of return R and at death leaves a pre-tax bequest b_{t+1i} to the next generation. There is an unequal initial distribution of bequests b_0 given exogenously. Each individual works l_{ti} hours at a pre-tax wage rate w_{ti} drawn from an arbitrary but stationary distribution, earning $y_{Lti} = w_{ti} l_{ti}$.

Individuals have a utility function $V^{ti}(c_{ti}, b, \underline{b}, l_{ti})$, increasing in consumption c_{ti}, in

[4] Elinder et al. (2012) provide empirical evidence on such labor supply responses to inheritances.

pre-tax bequest left b (capturing accidental motives), and in after-tax capitalized bequest left $\underline{b} = R \cdot b_{t+1i}(1 - \tau_{Bt+1})$ (capturing joy of giving motives[5]) and decreasing in labor l_{ti}. Note that the donor's utility function includes the after-tax capitalized bequest left but not the utility of the bequest receivers, resulting in a positive externality. Individuals use their net-of-taxes lifetime resources on consumption c_{ti} and bequest left b_{t+1i}. Hence, the individual maximization problem is

$$\max_{l_{ti}, c_{ti}, b_{t+1i} \geq 0} V^{ti}(c_{ti}, b, \underline{b}, l_{ti}) \qquad \text{s.t.} \tag{1}$$

$$c_{ti} + b_{t+1i} = Rb_{ti}(1 - \tau_{Bt}) + w_{ti}l_{ti}(1 - \tau_{Lt}) + E_t$$

The utility functions V^{ti} and the wage rates w_{ti} are assumed to follow an ergodic stochastic process such that with constant tax rates τ_B and τ_L, and government revenue E, the economy converges to a unique ergodic steady-state equilibrium independent of the initial distribution of bequests b_{0i}. In equilibrium individuals maximize utility as in (1) and this results in a steady-state ergodic equilibrium distribution of bequests and earning (b_{ti}, y_{Lti}).

The steady-state SWF is defined as the sum of individual utilities weighted by Pareto weights $\omega_{ti} \geq 0$. Hence, a normative social welfare criterion must be assumed. The government must solve

$$SWF = \max_{\tau_L, \tau_B} \int_i \omega_{ti} V^{ti}(c_{ti}, b, \underline{b}, l_{ti}) \qquad \text{s.t.} \tag{2}$$

$$E = Rb_t\tau_B + w_t l_t \tau_L$$

The derivation of the optimal tax rate on bequests τ_B takes the linear marginal tax on labor income τ_L as given. In the steady-state equilibrium the government's financial needs E will be constant ($dE = 0$) and with no government debt, the two taxes, τ_B and τ_L, will be linked to each other in order to satisfy the government's budget constrain. The

[5] PS13 denote these bequests as altruistic (as opposed to accidental), however it corresponds to joy of giving motives as defined above.

optimal linear tax on bequests that maximizes steady-state social welfare is

$$\tau_B = \frac{1 - \left[1 - \frac{e_L \tau_L}{1-\tau_L}\right] \cdot \left[\frac{\bar{b}^{\text{received}}}{\bar{y}_L}(1+\hat{e}_B) + \frac{\nu}{R/G}\frac{\bar{b}^{\text{left}}}{\bar{y}_L}\right]}{1 + e_B - \left[1 - \frac{e_L \tau_L}{1-\tau_L}\right]\frac{\bar{b}^{\text{received}}}{\bar{y}_L}(1+\hat{e}_B)} \tag{3}$$

where ν is the share of joy of giving bequests and e_B and e_L are the long-run elasticities that capture behavioral responses of bequest flows b_t and of the aggregated labor supply in terms of earning y_{Lt} with respect to the corresponding net-of-tax rates $(1 - \tau_B)$ and $(1 - \tau_L)$. Because the two taxes, τ_B and τ_L, are linked to satisfy the government budget constraint, the elasticities capture the effect of a joint and budget-neutral change in both taxes. The elasticities are defined as

$$e_B = \frac{1-\tau_B}{b_t}\frac{db_t}{d(1-\tau_B)}\bigg|_E \qquad \text{and} \qquad e_L = \frac{1-\tau_L}{y_{Lt}}\frac{dy_{Lt}}{d(1-\tau_L)}\bigg|_E \tag{4}$$

The distributional parameters $\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$ and $\bar{y}_L$ capture two elements. First, the degree of inequality of bequests received, bequests left, and labor income observed in the data. And second, the normative weighting of the individuals in the SWF.

$$\bar{b}^{\text{received}} = \frac{\int_i g_{ti} b_{ti}}{b_t} \quad , \quad \bar{b}^{\text{left}} = \frac{\int_i g_{ti} b_{t+1i}}{b_{t+1}} \quad \text{and} \quad \bar{y}_L = \frac{\int_i g_{ti} y_{Lti}}{y_{Lt}} \tag{5}$$

The three parameters are defined as the ratios of the population average weighted by the social welfare weights g_{ti} (defined below) to the unweighted population averages. The ratios will be smaller than 1 if the social welfare weights g_{ti} put more weight on individuals that are worse-off and will be equal to 1 when these weights are equally distributed.

The social welfare weights g_{ti} (Saez and Stantcheva, 2016) are defined as each individual's marginal utility of consumption, V_c^{ti}, weighted by the Pareto weight ω_{ti} and divided by the weighted average of the marginal utility of consumption for the entire population to normalize them. They measure the social value of increasing consumption of individual

ti by one unit relative to distributing that unit equally across all individuals.

$$g_{ti} = \frac{\omega_{ti} V_c^{ti}}{\int_j \omega_{tj} V_c^{tj}} \tag{6}$$

Calibration

The strategy followed by PS13 for the calibration of the optimal tax rate is to calibrate it for each percentile of the distribution of bequest received. In other words, they sequentially calibrate the optimal tax from the perspective of each 1% interval of the distribution of bequest received, as if the social planner only cared for those individuals. In terms of the social welfare weights, g_{ti}, their approach is equivalent to recursively setting the weights of all individuals to zero except for those belonging to percentile p.[6,7]

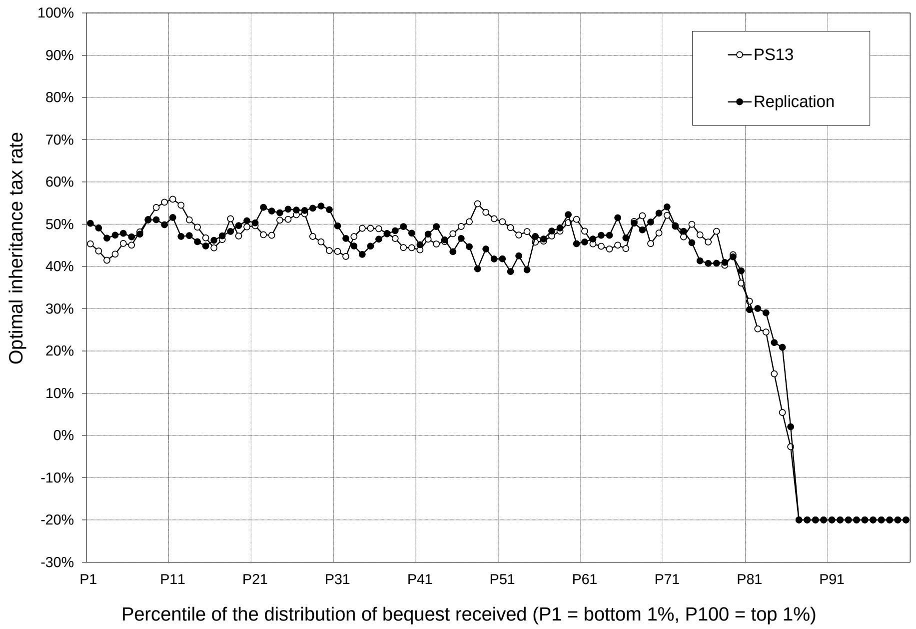
Using U.S. micro-data from the Survey of Consumer Finances (SCF) 2010 and focusing on individuals aged 70+, PS13 obtain the optimal tax rate by percentile of bequest received, which is shown in figure 1a along with my own replication.[8] The figure reports the optimal linear tax rate τ_B from the point of view of each percentile of bequest receivers based on (3) and given the benchmark parameters $e_b = 0.2$, $e_L = 0.2$, $\tau_L = 30\%$, $\nu = 1$, $R/G = 1.8$ and a capitalization rate $r = 3$. We observe that the optimal tax rate remains constant around 50% until percentile 70, corresponding to individuals who have not received any bequest. It then drops rapidly as the inheritance received, and to a lesser extent wealth and income, increase. For percentiles above 85 the optimal tax turns negative (a subsidy), growing to minus infinity. Note that the figure is constructed with a lower bound of −20%.

In figure 1b I show the three distributional parameters $\bar{b}_p^{\text{received}}$, $\bar{b}_p^{\text{left}}$, and $\bar{y}_{Lp}$ that underlie my replication of the optimal tax rate. We observe that they remain fairly constant until percentile 70, causing the constant 50% optimal tax rate for the first 70 percentiles. In Section 5, I account for heterogeneity in wealth and labor income, obtaining a different result.

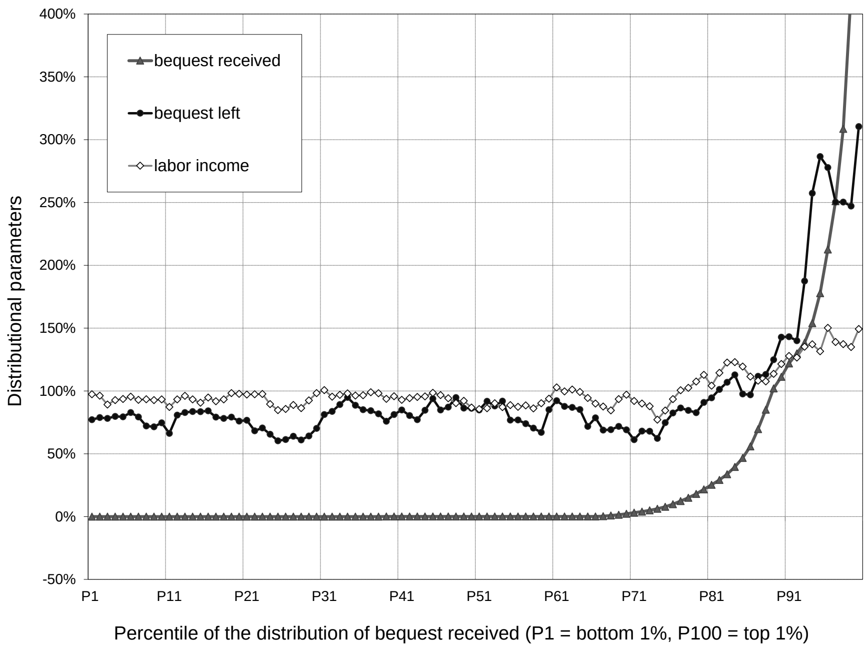
[6] In their own words: "To be agnostic and explore heterogeneity in optimal τ_B across the distribution, we consider percentile p-weights which concentrate uniformly the weights g_{ti} on percentile p of the distribution of *bequest received.*" (PS13, p.1873).

[7] PS13 also calibrate the optimal tax rate for larger groups of de distribution of bequest received (0-50, 50-70, 70-90 and 90-95).

[8] Note that the replication for the first 70 percentiles cannot be exact because individuals are randomly assigned to each percentile, as discussed in Section 5.



(a) Optimal linear inheritance tax rate by percentile of bequest received. Calibration of PS13 for the U.S. and replication.



(b) Distributional parameters by percentile of bequest received for the U.S. using data from SCF 2010. Own calculation.

Figure 1: Replication of the optimal tax rate by percentile of bequest received and distributional parameters.

4 Calibration of the Full Social Optimum

This section shows the results from calibrating the full social optimal tax rate under three standard social welfare criteria. First, the *utilitarian* criterion, which corresponds to a social planer with no preference for redistribution that weights individuals equally in the SWF, with ω_{ti} equally distributed. Second, the *responsibility and compensation criterion*, which sets ω_{ti} to 1 for individuals who did not receive any bequests, and to zero for those who did, arguing that this source of inequality is unmerited. And third, the *Rawlsian* criterion, which has the strongest preference for redistribution, considering only the worst-off individual in the SWF, and setting ω_{ti} to zero for all individuals except for the individual with the lowest utility.[9]

The individual utility V^{ti} enters the social welfare weights g_{ti} through the individual marginal utility of consumption V_c^{ti}. I consider a utility function that is additively separable in consumption c_{ti}, i.e. $V^{ti}(c_{ti}, b, \underline{b}, l_{ti}) = u(c_{ti}) + h^{ti}(b, \underline{b}, l_{ti})$. First, I consider $u(c_{ti})$ being linear and hence a marginal utility $V_c^{ti} = \alpha$.[10] Second, I consider $u(c_{ti})$ being isoelastic, with $V_c^{ti} = c_{ti}^{-\rho}$ which is strictly concave for $\rho > 0$. I evaluate this function for a range of values of ρ between 0 and 1.4 based on the estimates of Chetty (2006).

The social welfare weights g_{ti} resulting from the different combinations of the three social welfare criteria and the different utility functions are shown in the appendix (figure A1).[11] These welfare weights are then used to compute the distributional parameters of bequest received, bequest left, and labor income, defined in (5), which determine the full social optimal tax rate defined in (3).

Utilitarian

Table 1 presents the resulting full social optimal tax rates. The first panel shows the results under the utilitarian criterion and different levels of concavity of the individual utility function. Under the utilitarian criterion, the pareto weights ω_{ti} are equally distributed for

[9] PS13 calibrate the optimal tax rate under a "meritocratic Rawlsian" criterion, which is equivalent to the responsibility and compensation criterion but setting the welfare weights to zero for about half the population.

[10] In this case, the marginal utility of bequest left must be non-constant to obtain an interior solution.

[11] Note that under a Rawlsian criterion, the welfare weights are the same for all specifications of individual utilities, since only one individual has positive weight.

all individuals and the welfare weights g_{ti} are only unequally distributed when the utility function is strictly concave. I obtain that under the utilitarian criterion the optimal tax rate is negative irrespective of the concavity of the functional form.

This negative-tax result is caused by the positive externality that originates in the joy of giving motive. Note that $V^{ti}(c_{ti}, b, \underline{b}, l_{ti})$ increases with the after-tax bequest left $\underline{b}$, that is, the utility of the donors increases due to the act of bequeathing alone, regardless of its positive effect on the utility of the donees. In a steady-state equilibrium with a social planner that cares about the utility of all generations, this produces a positive externality and the optimal tax rate internalizes it by means of a negative tax.

Importantly, the negative-tax result hinges also on the assumption of a utilitarian SWF. The reason is that the positive externality grows proportionally with bequest received and the latter is highly concentrated at the top of the distribution, leading to very large positive externalities for individuals who receive the largest bequests. Because all individuals are weighted equally by the utilitarian criterion, the positive externality present at the top of the distribution dominates the full social optimum. Therefore, when the full social optimum derived by PS13 is calibrated under a utilitarian criterion it reaches the same result as previous models who derived the optimal tax rate under joy of giving motives and a utilitarian criterion (Farhi and Werning, 2010).[12]

The result that the tax rate becomes more negative as the utility function becomes more concave might be counter-intuitive at first sight. If we increase the welfare weights of the poor, shouldn't the bequest subsidy decrease? However, the government's margin of decision is to trade off tax rates on labor income and on inheritances, conditional on raising the revenue E_t. Therefore, individuals with high marginal utility of consumption (who are the ones with less income)[13] will be weighted more by the government and these

[12]Note that strictly speaking, under a utilitarian criterion with linear individual utility the first best solution would be to use a lump-sum tax rather than a distortive labor income tax to finance the bequest subsidy . In this case, however, the lump-sum grant would not be optimal. We must hence assume that in a second-best world even a government with equal welfare weights is forced to use labor income taxation to collect a given amount E.

[13]In the absence of consumption data, I use labor income as a proxy for consumption. An alternative would be to construct a measure of overall budget combining lifetime income and wealth. Both measures, however, are likely to incur in some measurement error that will be concentrated at the top and bottom percentiles of the distribution.

individuals prefer a high rate on labor rather than on bequests. Crucially, this statement hinges on the underlying distribution of income and wealth observed in the data. If we look at the distributional parameter reported in the central columns of Table 1 we observe that as the concavity of the utility function increases, the distributional parameter of labor income decreases more than that of bequest received and bequest left. In other words, the labor incomes that the government weights in its welfare function represent a decreasing share. The distributional parameters of bequest received and bequest left also decrease, but less so, leading to a decrease in the tax, that is, an increase in the subsidy.

Table 1: Full social optimal tax rate under different welfare criteria

Welfare criterion	Utility	$\bar{b}_p^{\text{received}}$	$\bar{b}_p^{\text{left}}$	$\bar{y}_{Lp}$	Optimal tax rate
Utilitarian	$\rho = 0$ (linear)	1.00	1.00	1.00	-582.3%
	$\rho = 0.3$	0.99	0.88	0.91	-5040.1%
	$\rho = 0.7$	0.92	0.77	0.80	-10000.0%
	$\rho = 1.4$	0.78	0.71	0.61	-10000.0%
Resp. & compens.	$\rho = 0$ (linear)	0.00	0.83	0.99	48.1%
	$\rho = 0.3$	0.00	0.72	0.91	50.0%
	$\rho = 0.7$	0.00	0.63	0.79	49.9%
	$\rho = 1.4$	0.00	0.64	0.60	38.5%
Rawlsian	linear = isoel.	0.00	0.00	0.23	83.3%

Note: Own calculations using SCF 2010. Lower bound -10000%.
Benchmark parameters: $e_b = 0.2$, $e_L = 0.2$, $\tau_L = 30\%$, $\nu = 1$, $R/G = 1.8$.

Responsibility and compensation

Under the responsibility and compensation criterion, individuals who received a positive bequest (around 30%) are weighted out of the SWF, and those who did not, have positive weights either equally distributed when the utility function is linear or diminishing in labor income when the utility function is isoelastic. Under this criterion the optimal tax rate becomes positive and, for the linear utility case, equal to 48.1%.

The positive-tax result highlights the importance of the SWF for the optimal tax rate. By excluding individuals from the top percentiles the externality of giving disappears and the optimal tax rate becomes positive. This is driven by the distributional parameter of bequest received which, by definition, drops to zero.

Interestingly, the concavity of the utility function impacts the optimal tax rate non-monotonically, and this is driven by the distribution of wealth, which does not increase monotonically with labor income (and hence with the marginal utility of consumption). With bequest receivers weighted out of the social welfare function, the ratio between the distributional parameters of bequest left and labor income ($\bar{b}_p^{\text{left}}/\bar{y}_{Lp}$) is what determines the government's choice of taxes on bequests and labor income. In this case, moving from a linear utility function to a slightly concave function ($\rho = 0.3$) reduces that ratio, therefore the share of wealth that the government cares about (the one weighted in its social welfare function) decreases less than the share of labor income that the government weights in. However, for further degrees of concavity the effect is the opposite and the ratio increases, leading to a decrease in the optimal tax rate.

As an illustration of these forces, note that the distributional parameter of bequest left decreases as ρ goes from 0 to 0.3 and to 0.7, but then increases when $\rho = 1.4$. In this later case, the weight given to the individuals at the bottom of the income distribution is an order of magnitude of 10 times the weight when $\rho = 0.7$. These individuals have a comparatively high net wealth (see percentile 1 of figure 3a) which makes them prefer a low or even negative tax on wealth despite not having received any bequest. This increases the ratio $\bar{b}_p^{\text{left}}/\bar{y}_{Lp}$ and pushes the full social optimal tax rate down.

Rawlsian

The Rawlsian criterion assigns the full Pareto weight ω_{ti} to the worst-off individual and sets it to zero elsewhere. Since only one individual has positive weight, the specification of this individuals's utility function is redundant, and therefore the welfare weights g_{ti} are identical for both the linear and the isoelastic specifications. Hence, the full social optimal tax rate under any specification of the individual utility is the same, in this case, 83.3%. Note that even though this worse-off individual does not receive or leave any bequest, the optimal tax rate from his/her perspective is not 100% because with a positive bequest elasticity bequests would drop to zero and the revenue loss would have to be compensated with a rise in the labor income tax rate.

Overall, these empirical calibrations are consistent with the findings of Farhi and

Werning (2013). With an utilitarian SWF they find a negative optimal tax rate.[14] Under a Rawlsian (maxmin) criterion, they obtain a positive tax rate. The responsibility and compensation, with a more intermediate preference for redistribution, is not evaluated in Farhi and Werning (2013) and in our calibrations it leads to a positive tax rate that is closer to the rates observed in current legislation.

Variants of the benchmark case

Table 2 presents the full social optimal tax rate with linear utility calibrated under different values of the benchmark parameters used in table 1.

Table 2: Variants of the full social optimum

	Utilitarian	Resp. & compens.	Rawlsian
Benchmark	**-582.3%**	**48.1%**	**83.3%**
$e_B = 0$	-485.4%	57.7%	100.0%
$e_B = 0.3$	-619.5%	44.4%	76.9%
$e_B = 0.7$	-724.7%	33.9%	58.8%
$e_B = 1$	-776.0%	28.8%	49.9%
$e_B = 3$	-921.3%	14.4%	24.9%
$e_B = 5$	-969.8%	9.6%	16.6%
$e_B = 30$	-1047.9%	1.8%	3.1%
$e_L = 0.1$	-1310.3%	46.4%	83.3%
$e_L = 0.3$	-339.6%	49.7%	83.3%
$e_L = 0.5$	-145.5%	53.0%	83.3%
$\nu = 0.7$	-435.9%	58.7%	83.3%
$\nu = 0.2$	-192.0%	76.3%	83.3%
$\nu = 0$	-94.4%	83.3%	83.3%

Note: Own calculations using SCF 2010.
Benchmark parameters: $e_b = 0.2$, $e_L = 0.2$, $\tau_L = 30\%$, $\nu = 1$, $R/G = 1.8$.

The first panel shows the full social optimal tax rate under different bequest elasticities, e_b.[15] Estimations by Kopczuk and Slemrod (2001) find this elasticity to be around 0.2 and PS13 consider that a value of 1 is implausibly high. However some theoretical models

[14] In Farhi and Werning (2013) this result holds only when the utility of both parents and children is included in the SWF, and the optimal tax rate is zero when only the utility of parents is considered. In PS13 each generation is both a bequest leaver and receiver.

[15] Note that the elasticities e_b and e_L are defined with respect to the net-of-tax rates $(1 - \tau_B)$ and $(1 - \tau_L)$ and therefore take positive values.

such as Chamley (1986) and Judd (1985) are derived under a setup where the elasticity of bequests is infinite. I therefore consider higher elasticities as well.

Higher bequest elasticities reduce the optimal tax rate on bequests. Under the utilitarian criterion the negative tax increases in absolute value. Under the responsibility and compensation criterion and under the Rawlsian criterion the tax rate decreases with the bequest elasticity and it converges to 0% as the elasticity increases. Note that under the Rawlsian criterion with an elasticity $e_B = 0$ the optimal tax rate is 100%, since the social planner only cares about the worst-off individual and there are no efficiency costs from taxing bequests due to the zero elasticity. However, so long as the elasticity of bequests is larger than zero, the optimal tax is smaller than 100%.

The second panel of table 2 shows the effect on the optimal tax rate of different labor supply elasticities to labor income taxes, e_L. We observe that higher labor elasticities increase the optimal tax rate on bequests. The intuition for this result is that the higher the elasticity of labor supply, the larger the efficiency loss from taxing labor income. Hence, to satisfy the government's budget constraint for a given labor income tax rate, a higher tax rate on bequests is needed. Under the utilitarian criterion the optimal subsidy decreases sharply as e_L increases because the large subsidy for the top bequest receivers is now more costly to finance. Under the responsibility and compensation criterion the sensitivity of the optimal tax rate to changes in e_L is moderate, and this result holds across different values of e_B. Under the Rawlsian criterion the optimal tax rate is unaffected by changes in e_L. Actually, under this criterion the only parameter that affects the optimal tax rate is the elasticity of bequest, as discussed above, because the distributional parameters of bequest received and bequest left are equal to zero and the optimal tax formula (3) is reduced to $\tau_B = \frac{1}{1+e_B}$.

The third panel shows the sensitivity of the optimal tax rate to bequest motives. As the share of accidental bequests increases (lower ν) the optimal tax rate under the utilitarian and responsability and compensation criteria increases. This is because taxation of accidental bequests does not impact the utility of the donors since the after-tax bequests left b do not enter their utility function.

Note that for the three social welfare criteria, when bequest motives are fully accidental ($\nu = 0$), the optimal tax rate remains under 100%. This result differs from previous models, like Cremer and Pestieau (2011), in which fully accidental bequest motives are taxed at a 100% rate. The reason is that the flexibility of the model of PS13 allows for the unconventional case where bequest motives are fully accidental but the bequest elasticity is positive. However, if the bequest elasticity is zero the optimal tax rate becomes 100% under the three criteria.

Final remarks

From these calibrations we conclude that the main determinant of the optimal tax rate is the assumed social welfare criterion. Positive full social optimal tax rates under PS13's framework appear only when wealthier individuals are weighted less in the SWF. A second determinant of the optimal tax rate are the ratios between the distributional parameters of bequest received or left and of labor income. These ratios capture the government's social preferences regarding the trade-off between labor and bequest taxes, which depends on the welfare weights and the underlying distribution of wealth and income.

A more concave individual utility function increases the subsidy under the utilitarian criterion because individuals with low labor income (who therefore prefer high taxes on labor) are weighted more, since they have higher marginal utility of consumption. Under the responsibility and compensation criterion, however, a higher concavity of the utility function has non-monotonic effects on the optimal tax on bequests. This is because individuals with low income can have high wealth, so the ratio between the distributional parameters of bequest left and labor income that determine the optimal tax rate on bequests are also non-monotonic as the concavity of the utility increases. From these results we conclude that, conditional on the distribution of the data, the concavity of the individual utility does not have a qualitatively significant impact on the optimal inheritance tax rate.

Finally, we observe that criteria with an intermediate preference for redistribution, such as responsibility and compensation, are the most sensitive to variations of the benchmark parameters such as the elasticities of bequests and labor income and the share of accidental bequests.

5 Introducing heterogeneity in wealth and labor income

The calibration approach of PS13 exploits heterogeneity in bequests received, ordering individuals by the amount of bequest received and calculating the optimal tax rate from the perspective of each percentile. In doing so, the large share of individuals who did not receive any bequest, about 70%, are randomly assigned to each of the first 70 percentiles. These individuals differ in accumulated wealth (future bequests left) and in labor income, but since they are ordered randomly, the average value of wealth and labor income becomes approximately the same for each of the first 70 percentiles and so do the two corresponding distributional parameters and the resulting optimal tax rate. This leads PS13 to conclude that the optimal tax rate by percentile is constant for the first 70 percentiles (see figures 1a and 1b).

In this section, I further exploit individual heterogeneity by sub-ordering individuals by their wealth and labor income. This avoids the random assignment of non-receivers across the 70 first percentiles and offers a more realistic description of the different optimal tax rates from the perspective of each percentile and about the drivers of the optimal tax across the population of non-receivers. This leads to an optimal tax rate that varies significantly among the non-receivers.

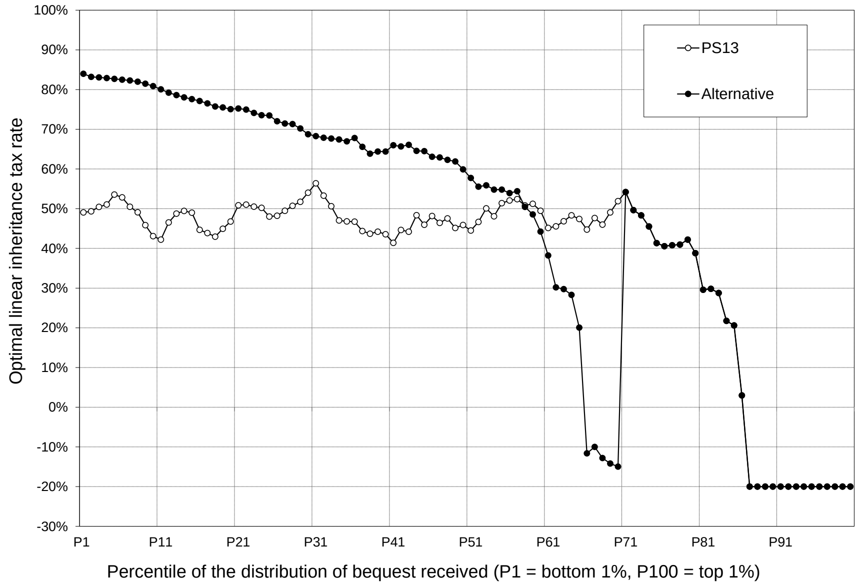
In a way, this approach makes each percentile more representative of the different individuals of the population, incorporating the heterogeneity present in all the variables of PS13's model. Also, this calibration approach is consistent with the assumptions of the model, which explicitly includes heterogeneous wealth and wages, and emphasizes the connection between these variables (e.g. individuals accumulate wealth through labor income, which is likely to be bequeathed) and between their taxes (which must fulfill the government's budgetary needs).

An alternative approach for ordering individuals is to use their total budget (bequest received plus income) or their total budget extended (adding wealth). These two measures have the advantage of capturing individual heterogeneity jointly for bequests, income and wealth leading to a more realistic distribution of the optimal tax rate across the population. However, this approach makes it harder to learn about the drivers of the optimal tax rate by percentile because individuals with a similar budget might have very different bequests, income and wealth. The results from this alternative approach are reported in the appendix figures A2 and A3.

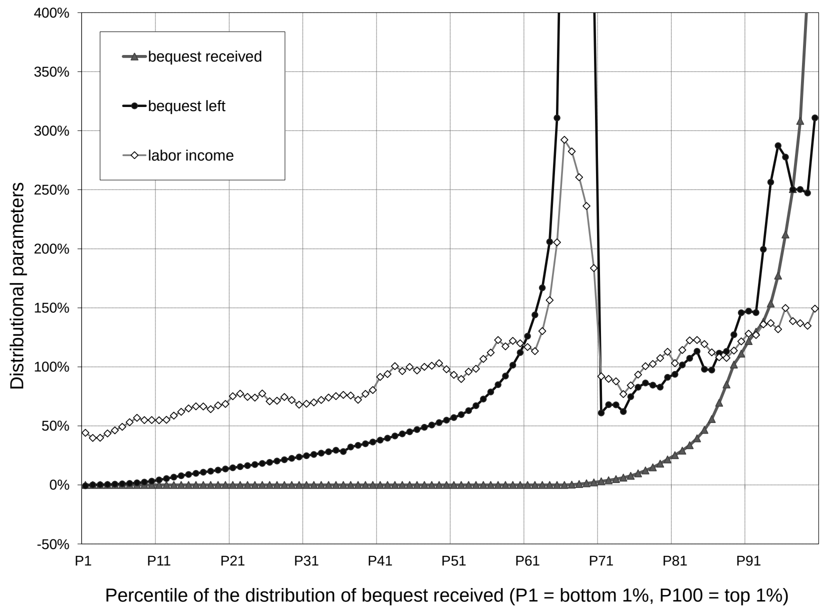
The methodology followed to calculate the new distributional parameters is the same as in PS13, that is, giving uniform social welfare weights g_{ti} to all individuals within each percentile. The distributional parameters $\bar{b}^{\text{received}}$, $\bar{b}^{\text{left}}$ and $\bar{y}_L$ are then the average of bequest left, bequest received, and labor income for each percentile relative to population averages. The change with respect to PS13's calibration is that the individuals included in each percentile are now different, as a result of the different ordering.

Figures 2b and 2a show the optimal tax rate and the distributional parameters resulting from sub-ordering by wealth. Compared to the original calibrations of PS13 we observe that the optimal tax rate is not constant for the first 70 percentiles, and neither are the distributional parameters of bequest left, which by construction increases monotonically for the first 70 percentiles, and labor income. Now the optimal tax rate decreases for the first 70 percentiles, as the individuals' wealth rises. It starts with an optimal tax rate of 83.3% for the bottom 1% (coinciding with the Rawlsian full social optimum) and turns negative, about -14%, for percentiles 66 to 70. This evolution reflects the intuitive idea that those individuals who did not receive any inheritance but have accumulated wealth (which they will probably bequeath) might prefer a low or even negative inheritance tax rate. On the other hand, individuals from the bottom percentiles who own no wealth but earn labor income prefer a tax on inheritances that collects as much as possible (only bounded by the elasticity of bequests), since the remaining financial needs of the government will have to be covered by a rise in labor income taxes.

The results from sub-ordering individual observations by labor income are presented in figures 3a and 3b. In this case the distributional parameter that increases monotonically until percentile 70 is labor income. The distributional parameter of bequest left also tends to increase, but it oscillates more, causing the optimal tax rate to behave more erratically. This shows that the behavior of the distributional parameter of bequest left dominates the



(a) Optimal inheritance tax rate by percentile of bequest received sub-ordering by wealth. Compared to PS13.

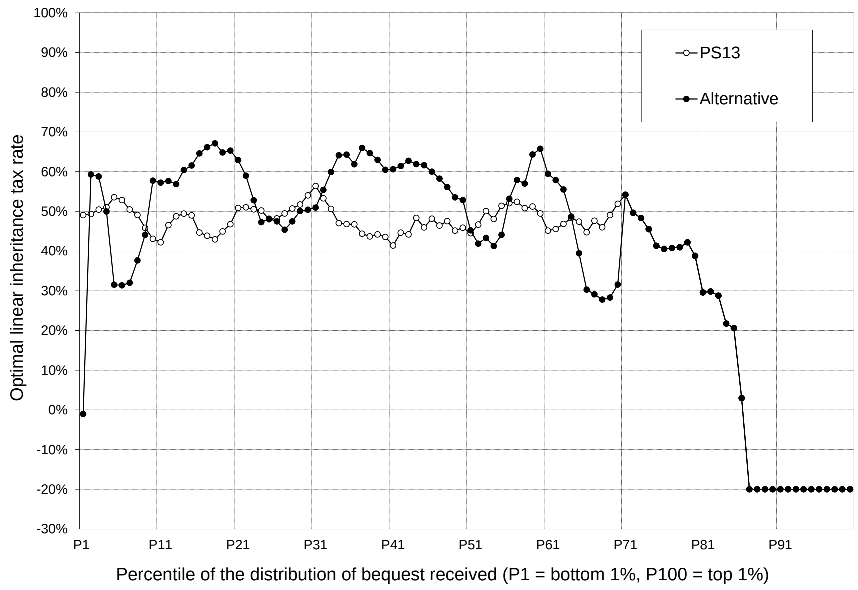


(b) Distributional parameters by percentile of bequest received sub-ordering by wealth.

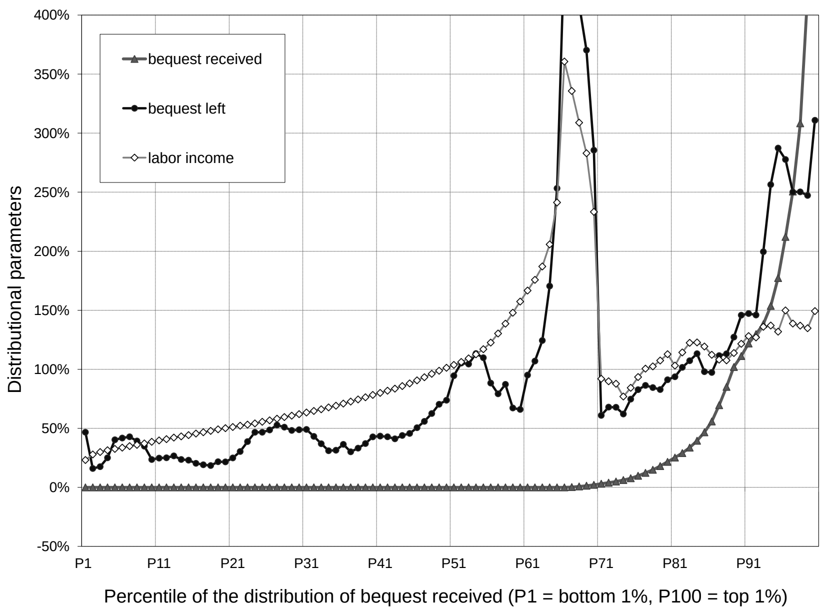
Figure 2: Optimal tax and distributional parameters sub-ordering by wealth.

effect of the distributional parameter of labor income, as we observed when calibrating the different full social optima.

Unlike the case where individuals were sub-ordered by bequest left, now there are no



(a) Optimal inheritance tax rate by percentile of bequest received sub-ordering by labor income. Compared to PS13.



(b) Distributional parameters sub-ordering by labor income.

Figure 3: Optimal tax and distributional parameters sub-ordering by labor income.

percentiles within the first 70 that would prefer a negative inheritance tax. The reason is again that the main driver of that result is the distributional parameter of bequest left but its effect is now more diluted among different percentiles due to sub-ordering by labor income. The only exemption to this is the first percentile, which has a negative tax rate caused by individuals who have accumulated wealth despite not earning labor income (through prizes or reducing their reported income using capital losses). These individuals are willing to take a very high tax on labor income as long as the tax rate on bequests is reduced.

6 Conclusion

This paper shows the crucial role of the assumed social welfare function —SWF— and of individual heterogeneity for the derivation of the optimal inheritance tax rate, which can range from negative to positive and large. Inheritance taxation is particularly sensitive to the choice of SWF due to the positive externalities that arise from joy of giving motives and how they interact with the heterogenous distribution of bequests, which are highly concentrated at the top of the distribution.

Under a utilitarian criterion the optimal inheritance tax rate is always negative. On the other hand, under social welfare criteria that favor redistribution the tax rate becomes positive. For example, under the responsibility and compensation criterion, which weights out of the SWF the 30% of individuals who received positive bequests, the optimal tax rate is about 50%. Under this criterion, the elasticity of bequests to taxation and the share of accidental bequests become relevant determinants of the optimal tax rate. Under a Rawlsian criterion, the optimal tax rate rises to 83.3%, bounded only by the elasticity of bequests to taxation. These findings match and explain the different results obtained by previous literature, and provide an empirical illustration.

In their paper, PS13 opt for calibrating the optimal tax rate from the perspective of each percentile of the distribution of bequest received. This approach leads the authors to conclude that the optimal tax rate by percentile remains fairly constant for the first 70 percentiles (those who do not receive any bequests). However, extending this methodology to also account for heterogeneity in wealth and in labor income, the optimal tax rate obtained for the same 70 percentiles is not constant, varying from 83% for percentile 1 to a negative tax rate of -14% for percentile 70. This new approach offers a richer description of the heterogeneous individuals of the population, in line with the assumptions of PS13's model, which considers the interrelation between bequest received, bequest left, and labor income.

These two findings are crucial for the design of optimal inheritance taxes. Policy makers must account for the effect of different SWFs and the utilitarian criterion is not a neutral benchmark. Models that assume utilitarian SWF lead to inheritance subsidies, but relatively small modifications of the SWF can lead to more realistic tax rates. In addition, the percentile calibrations show a large variation on the optimal tax rate from the individual point of view. This helps explain the public debate around taxation of inheritances given the large variation in preferences that we find.

References

Atkinson, A. and Stiglitz, J. E. (1976). The Design of Tax Structure: Direct Versus Indirect Taxation. *Journal of Public Economics*, 6(1-2):55–75.

Brunner, J. K. and Pech, S. (2012a). Optimal taxation of bequests in a model with initial wealth. *The Scandinavian Journal of Economics*, 114(4):1368–1392.

Brunner, J. K. and Pech, S. (2012b). Optimal taxation of wealth transfers when bequests are motivated by joy of giving. *The BE Journal of Economic Analysis & Policy*, 12(1).

Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica*, 54(3):607–22.

Chetty, R. (2006). A new method of estimating risk aversion. *American Economic Review*, 96(5):1821–1834.

Cremer, H. and Pestieau, P. (2011). The Tax Treatment of Intergenerational Wealth Transfers. *CESifo Economic Studies*, 57(2):365–401.

Diamond, P. A. (1965). National Debt in a Neoclassical Growth Model. *The American Economic Review*, 55(5):1126–1150.

Elinder, M., Erixson, O., and Ohlsson, H. (2012). The Impact of Inheritances on Heirs' Labor and Capital Income. *The B.E. Journal of Economic Analysis & Policy*, 12(1):1–37.

Farhi, E. and Werning, I. (2010). Progressive Estate Taxation. *The Quarterly Journal of Economics*, 125(2):635–673.

Farhi, E. and Werning, I. (2013). Estate taxation with altruism heterogeneity. *American Economic Review*, 103(3):489–95.

Fleurbaey, M. and Maniquet, F. (2018). Optimal income taxation theory and principles of fairness. *Journal of Economic Literature*, 56(3):1029–79.

Judd, K. L. (1985). Redistributive Taxation in a Simple Perfect Foresight Model. *Journal of Public Economics*, 28(1):59–83.

Kindermann, F., Mayr, L., and Sachs, D. (2018). Inheritance taxation and wealth effects on the labor supply of heirs. NBER Working Paper No. 25081.

Kopczuk, W. (2013). Incentive Effects of Inheritances and Optimal Estate Taxation. *The American Economic Review*, 103(3):472–477.

Kopczuk, W. and Lupton, J. P. (2007). To Leave or Not to Leave: The Distribution of Bequest Motives. *Review of Economic Studies*, 74(1):207–235.

Kopczuk, W. and Slemrod, J. (2001). *Rethinking Estate and Gift Taxation*, chapter The Impact of the Estate Tax on Wealth Accumulation and Avoidance Behavior, pages 299–349. Washington, D.C.: Brookings Institution Press.

Piketty, T. and Saez, E. (2013). A Theory of Optimal Inheritance Taxation. *Econometrica*, 81(5):1851–1886.

Saez, E. and Stantcheva, S. (2016). Generalized Social Marginal Welfare Weights for Optimal Tax Theory. *American Economic Review*, 106(1):24–45.

Straub, L. and Werning, I. (2020). Positive long-run capital taxation: Chamley-judd revisited. *American Economic Review*, forthcoming.

A Appendix

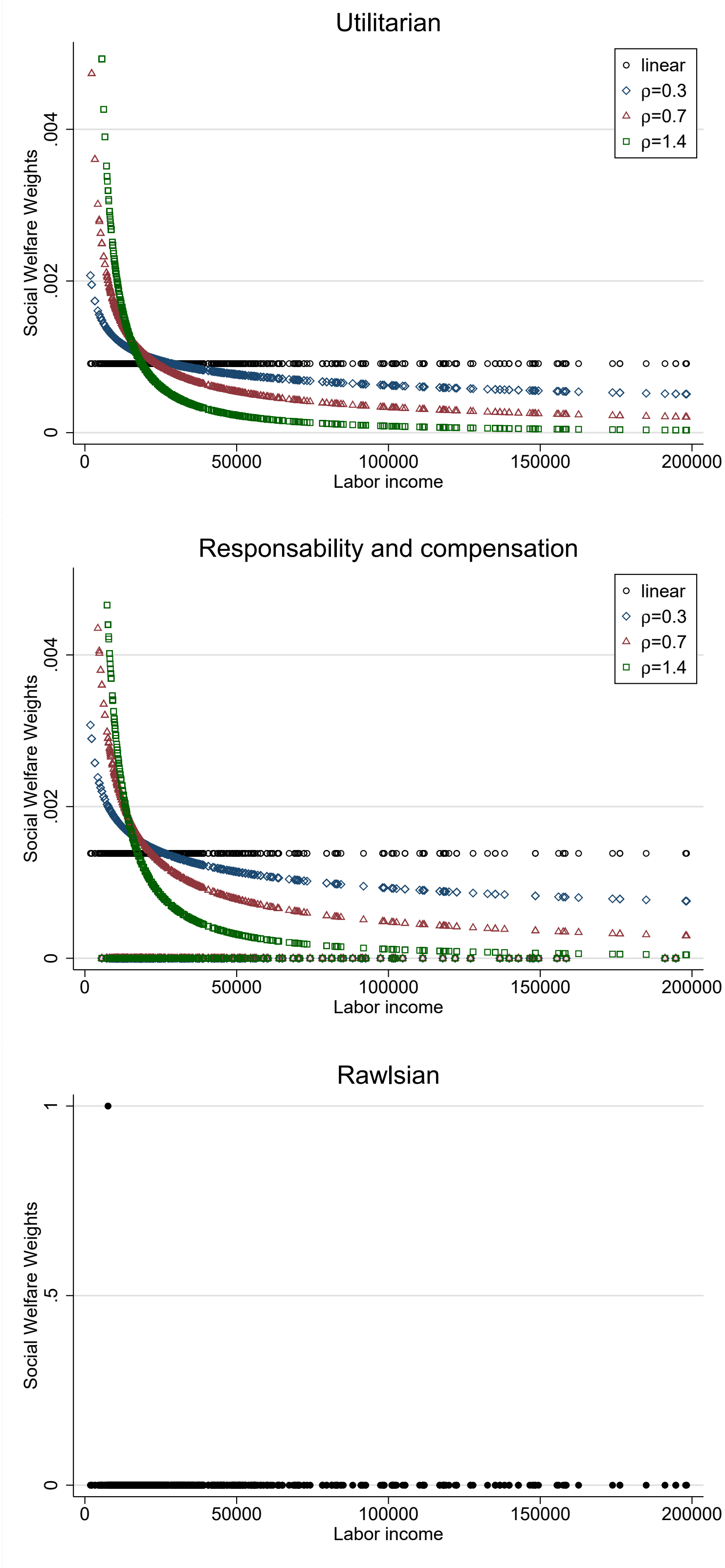
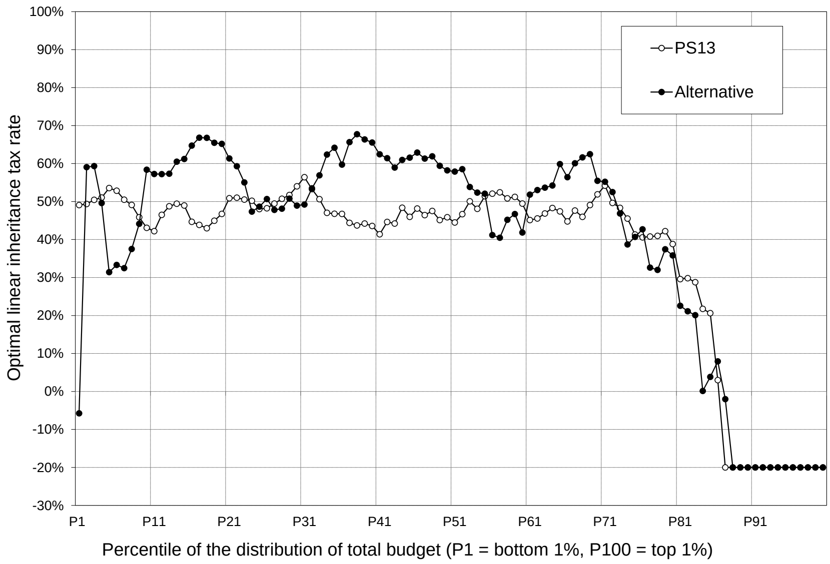
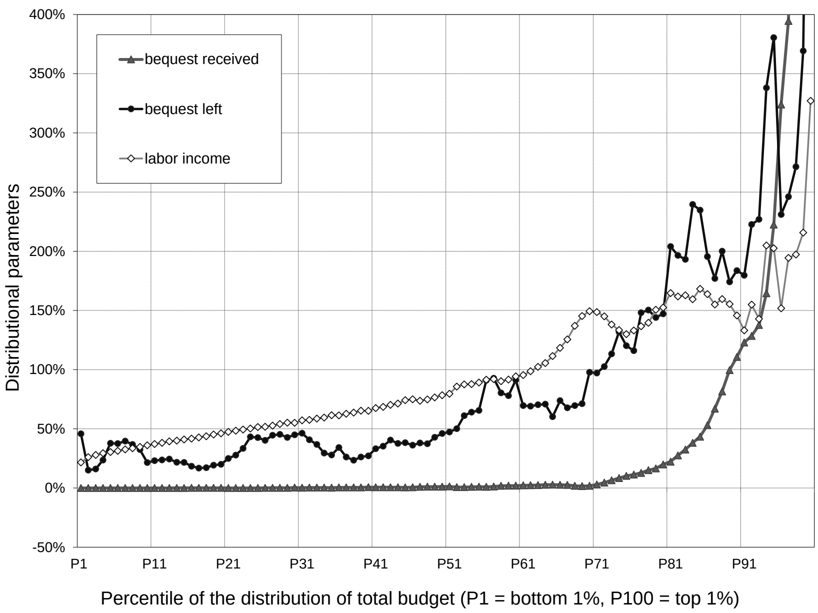


Figure A1: Social welfare weights for each individual under different SWF and concavities of the individual utility function. Under a Rawlsian criterion the welfare weights are independent of the individual utility assumed. Figures are truncated at 0.005 and 200000.

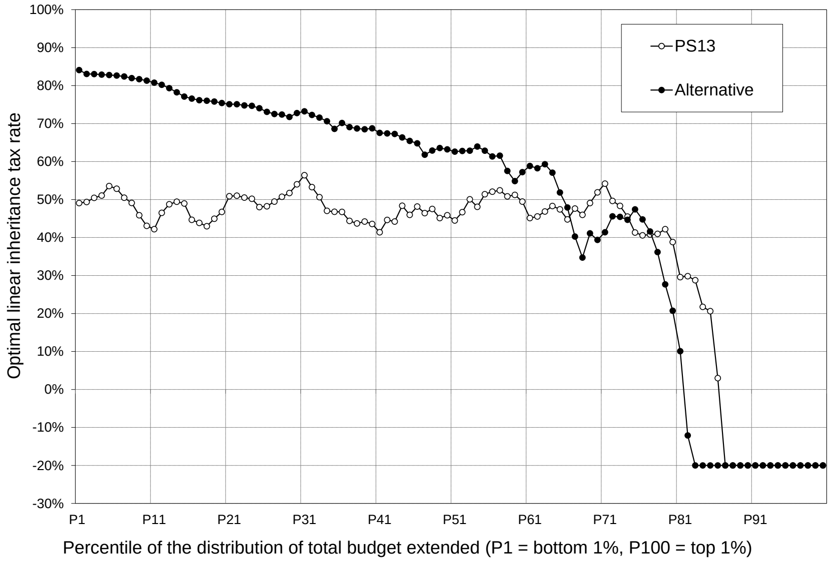


(a) Optimal inheritance tax rate by percentile of total budget. Compared to PS13.

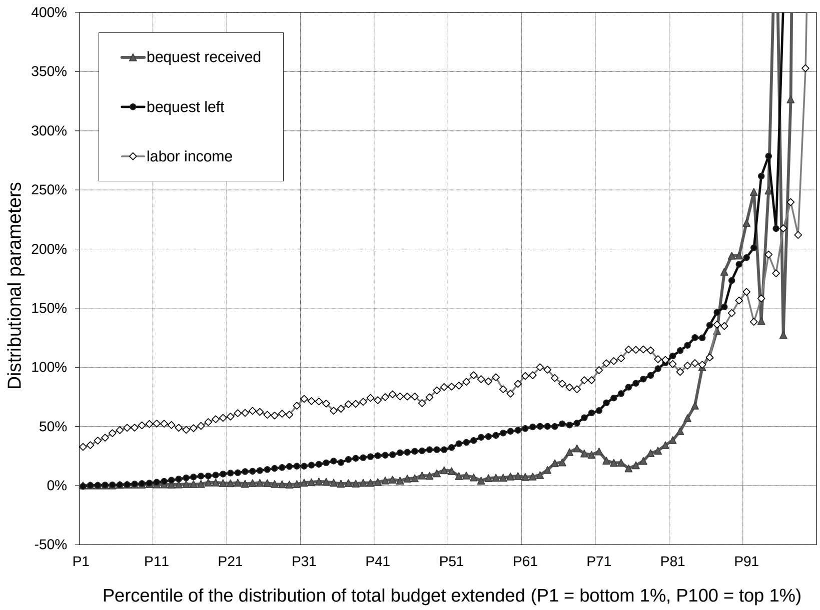


(b) Distributional parameters by percentile of total budget.

Figure A2: Optimal tax and distributional parameters sub-ordering by total budget (inheritance received + labor income).



(a) Optimal inheritance tax rate by percentile of total budget extended. Compared to PS13.



(b) Distributional parameters by percentile of total budget extended.

Figure A3: Optimal tax and distributional parameters sub-ordering by total budget extended (inheritance received + labor income + wealth).

Chapter IV

Are Children's Socio-Emotional Skills Shaped by Parental Health Shocks?

Are Children's Socio-Emotional Skills Shaped by Parental Health Shocks?

Esteban García-Miralles
University of Copenhagen,
Dep. of Economics and CEBI

Miriam Gensowski
University of Copenhagen,
Dep. of Economics and CEBI
and IZA

Abstract

Child skills are shaped by parental investments. When parents experience a health shock, their investments and therefore their children's skills may be affected. This paper estimates causal effects of severe parental health shocks on child socio-emotional skills. Drawing on a large-scale survey linked to hospital records, we find that socio-emotional skills of 11-16 year-olds are robust to parental health shocks, with the exception of significant but very small reductions in Conscientiousness. We study short-run effects with a child-fixed effects model, and dynamics around the shocks with event studies. A sibling comparison suggests some long-run build-up of effects of early shocks.

JEL Classification: J24, I10, I21.

Keywords: Big Five personality traits, development of personality traits, parental health shocks, socio-emotional skills, non-cognitive skills, skill formation

Acknowledgements: We appreciate generous funding from the Novo Nordisk Foundation (grant no. NNF17OC0026542) and from the Danish National Research Foundation through its grant (DNRF-134) to the Center for Economic Behavior and Inequality (CEBI) at the University of Copenhagen. We are also very grateful for discussions with and comments from Aline Bütikofer, Gordon Dahl, Titus Galama, Mette Gørtz, Torben Heien Nielsen, Jon Skinner, Stefanie Schurer, Eddy van Doorslaer, and all participants at the Novo workshop on "Behavioral responses to health innovations and the consequences for socioeconomic outcomes" at the University of Copenhagen.

1 Introduction

Socio-emotional skills, often measured with personality traits, are important determinants of life outcomes. Conscientiousness, for example, demonstrably affects educational performance and attainment (Poropat, 2009), as well as productivity and earnings (Cubel et al., 2016; Fletcher, 2013; Gensowski, 2018; Heineck and Anger, 2010; Mueller and Plug, 2006; Nyhus and Pons, 2005). Several traits also influence healthy living and health outcomes (Roberts et al., 2014). Agreeableness has been linked to economic preferences, such as reciprocity and altruism (Becker et al., 2012), or prosociality (Hilbig et al., 2014). Neuroticism, the reverse of Emotional Stability, is associated with mental health problems and lack of emotional wellbeing (Widinger, 2011), as it reflects the ability to bounce back from negative experiences or to dwell on the past. Overall, these traits are essential building blocks to a healthy and happy life. Importantly, even early childhood personality traits predict major life outcomes (Moffitt et al., 2011).

The formation of all skills, including socio-emotional skills, depends to a large extent on parents (Cunha and Heckman, 2007). Parents invest actively and passively, using their knowledge, resources, energy and time. Therefore parental health could influence the skill acquisition of their children. But whether ill health necessarily decreases child skills is ambiguous. A sick parent might need to be at the hospital, leading to a reduction in the quantity of time available for the child. At the same time, if the illness reduces a parent's work hours, it could also increase the time spent with their children. The quality of time can also be affected, presumably negatively, by a parental illness. Illness can also affect parental financial resources or other labor market outcomes. Children's socio-emotional skills can also be directly affected by the exposure to parental weaknesses, stress, and lack of control over life events. These experiences may even translate to long-run changes in socio-emotional skills if children react to them by adopting different views of their own social roles in a contextual model of personality (Roberts et al., 2006). It is thus conceivable that long-run effects of even temporary health shocks to parents translate to larger differences in socio-emotional skills over time. Alternatively, parents may be able to compensate for lower investments during a health shock, and thus mitigate long-run effects through life-cycle investments (Bharadwaj et al., 2017). Again, there are several channels through which parental health shocks can influence child skills, without a clear expectation for whether ill health would have larger long-run or short-run effects. Despite all these relevant channels through which parental health shocks potentially affect child socio-emotional skills, there is surprisingly little empirical evidence.

This paper estimates causal effects of parental health shocks on children's socio-emotional skills. We contribute to the existing knowledge on this topic in four ways: Firstly, we obtain causal effects using three different estimation strategies that address the main identification challenges of selection and reverse causality. Secondly, we estimate the short-run effects of shocks and provide complementary evidence on long-run effects. Thirdly, our data offers an ideal setting to observe any detrimental effects of parental shocks on child skills: it uses severe shocks, and observes productive socio-emotional skills of children (not only socio-emotional malfunctioning) at an age where one expects most malleability and influence of parents. Finally, we test for heterogeneous effects of these shocks by sex of the parent and child, and family socio-economic status (SES).

The usual identification challenges when estimating the effects of parental health shocks on child skills are threefold. First, parent and child outcomes are correlated due to underlying genes and a shared environment, leading to sample selection—parental shocks are not randomly distributed among children. Second, there is a measurement problem that carries the risk of picking up reverse causality: if the child's socio-emotional skills are reported by the parents, it is possible that parents who are ill score their children lower than they would otherwise (so that their reports do not correctly reflect the child's skills). Third, another type of reverse causality can occur if socio-emotional problems of the child cause worse parental health self-reports, or objectively worse parental health.

We overcome these identification difficulties by using fixed effects estimations that draw on third-party reported, objective parental health shocks that are unlikely to be influenced by child socio-emotional problems, and socio-emotional skill measures that are self-reported by the children. Specifically, we employ three separate empirical strategies that identify the short-run effects of the shocks with great precision (controlling for child fixed effects), the dynamics before and after the shock (event studies), and long-run effects of the shocks (sibling-pair comparisons, or parent fixed effects).

We construct a unique dataset by combining administrative records on parental health with a large-scale survey on children socio-emotional skills. We exploit a population-wide sample of rich administrative data that includes third-party records of parental health in the Danish population. Parents' health shocks are observed as diagnoses for hospitalizations due to cardiovascular shocks, cancer, mental health problems, and also include parental deaths. This is merged to a validated survey panel of socio-emotional skill outcomes for the children that was distributed in all public schools for the period 2015-2018.

Our paper adds to a short list of previous studies that have found mixed results of changes in parental self-reported health on child personality or problem behavior, ranging from negative effects (Mühlenweg et al., 2016; Cuadros-Menaca et al., 2018) to no effects (Le and Nguyen, 2017). Our empirical setting should be in the best position to identify any effect, because we consider objective health shocks that are arguably more severe than changes in self-reported health. So if there are dosage effects (more severe shocks generate greater responses in children), one would expect the effects to be larger in our study. Furthermore, children in our sample are slightly older than children in the existing studies, so we observe them at a time when socio-emotional skills fully develop (McCrae and Costa Jr., 1996), yet where they still depend greatly on parental investments. This would also tend to increase the estimated effect sizes.

Our findings show that socio-emotional skills of children (aged 11-16) are only weakly affected in the immediate aftermath of severe parental shocks, up to 3 years later. Conscientiousness, one of the most important traits, is reduced by .05% of a standard deviation from losing a parent, and .02% of a standard deviation from the health shocks considered jointly. There are no significant effects on Agreeableness, Emotional Stability, or Academic Self-Concept from these two events. The fact that children's socio-emotional skills are so robust to shocks to their parents' health is surprising, given that we are studying the stability of traits at a time of their lives during which we expect both most malleability and the greatest influence from parents, and using severe, objective health shocks.

With 95% confidence, we can rule out effects larger than 4% of a standard deviation for the parental health shocks considered jointly, or 10% of a standard deviation for parental deaths. This is much smaller than the SES gaps we find for the different traits (.12-.29% of a standard deviation for parents having college education, for example). We perform a back-of-the-envelope calculation that extrapolates these effects to adulthood and links them to reported wage returns to personality traits. From this exercise, we can exclude that a parent passing away (the most extreme shock) would have more harmful effects on yearly earnings than a reduction of 0.41%, or reduce educational attainment by more than 0.002 of a standard deviation.

We test whether the effects of shocks are larger among boys or girls, whether it matters that they happen to the father or the mother, and whether children of single mothers or low-income mothers are more vulnerable. Generally, there are no consistent patterns that would suggest a specific at-risk group.

We complement our analysis of the effects of parental shocks on children in the short run with a strategy that compares siblings, allowing us to identify the long-run effects from experiencing a shock earlier in life, while controlling for parental fixed effects. These long-run analyses, which must be interpreted with caution due to the small sample size used for the estimation and the different interpretation of the estimates as within-family timing effects, point to the existence of long run effects on Conscientiousness from shocks that occur earlier in the child's life.

2 Existing Literature

The existing literature that studies the effect of parental health on children focuses mostly on how childrens' health and educational outcomes are affected (see, for example, Currie and Moretti, 2007; Bhalotra and Rawlings, 2011; Kristiansen, 2020). Some research has shown associations between parental health and child educational outcomes in the US (Andrews and Logan, 2010, using the ECLS-K; or Johnson and Reynolds, 2013 using the NLSY), while many papers use data from developing countries (such as Senne, 2014; Dhanaraj, 2016; Alam, 2015) or transition countries (Bratti and Mendola, 2014).

Yet educational attainment is an outcome that is the result of investments and skill formation throughout the child's life. Socio-emotional skills, often referred to as non-cognitive skills, are essential building blocks to further educational attainment (see, e.g., Cunha and Heckman, 2007; Almlund et al., 2011; Lundberg, 2013, 2019). These skills also have a direct impact on later outcomes in life, such as income and health (Almlund et al., 2011; Fletcher, 2013; Gensowski, 2018; Heineck and Anger, 2010; Mueller and Plug, 2006; Roberts et al., 2014; Spengler et al., 2016).

There is only very little evidence on how parental health shocks affect child socio-emotional skills. As far as we are aware, there are only two peer-reviewed studies analyzing the effect of parental health shocks on child socio-emotional functioning on the Strengths and Difficulties Questionnaire (SDQ) measure.[1]

[1]There is also some evidence presented in Cuadros-Menaca et al. (2018), who use an Indonesian panel data from the IFLS, with two personality traits of Conscientiousness and Neuroticism observed in the last wave at

Mühlenweg et al. (2016) use the mother-child sample from the German Socio-economic Panel to study determinants of child socio-emotional skills at age 6, with 639 observations. Their identification strategy is to control for initial child characteristics (including prenatal conditions), and to interpret major changes in self-reported parental health as a health shock. The measure of child socio-emotional skills is the SDQ reported by the mother. This measure focuses mainly on the malfunctioning end of socio-emotional skills rather than on productive traits, as it has been widely used for psychopathological screening (Becker et al., 2006). Mühlenweg et al. (2016) find rather large effects of *maternal* health shocks (no effects of paternal shocks). When the mother's self-reported health decreases, or her number of nights at the hospital increase, the child displays .4-.9 standard deviations more socio-emotional difficulties (a combination of emotional symptoms, conduct problems, hyperactivity and inattention, and peer-relationship problems). These effects are not only statistically significant but also rather large. Yet, there is a remaining risk that they reflect not only the true effect of the shock but also initial differences between families where the mother experiences worsening health versus families where her health stays constant, when these differences are not captured by observable initial child characteristics.

Le and Nguyen (2017) do not risk this selection problem, as they employ a child fixed estimation. They exploit the Australian LSAC panel data, with children between ages 4 and 13, whose socio-emotional skills are also assessed with the SDQ. Their measure of parental health is based on self-reported answers including health status, mental health episodes and other self-reported symptoms. They demonstrate how large negative effects of shocks from OLS regressions disappear when using child fixed effects. Instead, they find only "little detrimental effects of poor parental health on cognitive and non-cognitive skills." Among all tested relationships, the only significant effect on child behavior was serious paternal mental health problems that increased the probability of hyperactivity. In a heterogeneity analysis, it appeared that single mothers' mental health also influenced the child SDQ.

Using parent-reported measures of children's socio-emotional skills can introduce a first problem of reverse causality in measurement: parents whose health suddenly declined may consequently evaluate their child's skills as less favorable—simply because they experienced a health shock, not because their child's skills have actually changed. Le and Nguyen (2017) address this problem by using teacher-reported information on the children's socio-emotional skills. They show that using parent-reported information may over-estimate the effects of parental health shocks. In

around age 24. Using sibling fixed effects, they find no effects of parental health deterioration on these two traits.

this paper, we also avoid this risk of reverse causality by using child-reported information on their own socio-emotional skills.

A second risk of reverse causality comes from the use of parent self-reported measures of their own health. If a child is experiencing socio-emotional problems, this might have an effect in the self-reported health measures of the parents, who might report worse conditions because of their child's behavior even if their actual health is unaffected. We address this by using health shocks that are third-party reported by medical professionals and that can be objectively measured. A third risk of reverse causality would occur if child's socio-emotional skills (or their problematic behavior) actually affect parental health. While we cannot solve this problem with econometric techniques, we argue that it is unlikely for a child's personality traits to influence the objective measures of severe health shocks we consider, such as a cancer diagnosis or a heart attack.

We contribute to this emerging literature and the larger question of how the family environment shapes child skills by providing evidence from self-reported child personality traits and objective health measures for the parents. The health shocks are not simply defined as changes in self-reported health status from one survey wave to the next, but objective medical diagnoses of rather severe shocks. We also provide initial evidence for long-run effects of parental health shocks on child skills. It is unclear *ex ante* whether one should expect the effects of shocks to be attenuated or amplified in the longer run. On the one hand, parental shocks may have initial effects on the child skills that fade out over time. Some health shocks may only work as a temporary disruption of family life, if either the parent gets well again (such as after a heart attack that is successfully treated), or because parents and the child adapt to a new organization of life at home with the illness. For example, Cobb-Clark and Schurer (2013) and Elkins et al. (2017) did not find, in the context of personality traits, consistent effects of common family- or health-related shocks. Also, the "adaptation level" view from psychology has long suggested, both theoretically and empirically, that an individuals' happiness reverts to a baseline level and is not affected by shocks such as lottery wins in the long run (Brickman et al., 1978). To the extent that the parents' well-being bounces back, evidence on maternal life satisfaction indicates that children's socio-emotional skills should benefit as well (Berger and Spiess, 2011).

On the other hand, not all health shocks are temporary—and there are several mechanisms that could transform parental health shocks to persistent long-run changes in the child's socio-emotional skills. For one, even though happiness may bounce back to pre-shock levels, it may do so via a socialization process in which children "grow" with the challenges—and in doing so, alter their socio-emotional skills. In this view of the contextual model of personality, even short-run shocks can have permanent effects on child skills via socialization (Roberts et al., 2006). Also, life cycle skill formation is characterized by sensitive and even critical periods. Therefore, even though parents may actively invest in their child's socio-emotional skills after a health episode has passed to remediate negative initial effects, their efforts may be hampered: The shock may have occurred during such a sensitive or critical period in their child's life that remediation is costly or ineffective (Cunha and Heckman, 2007). If socio-emotional skills of a child were harmed from a temporary shock, this disadvantage may be aggravated later because of missed subsequent self-productivity and dynamic complementarity. Thus, the effects of a parental health shock on child skills may also accumulate over time.

3 Data and Samples of Analysis

We construct a unique dataset by combining several administrative registers for the entire population of Denmark with a nation-wide panel survey of children in public schools. The registers include third-party reported information on health, as well as information on education, socioeconomic variables, and family linkages, allowing us to match children to their siblings and parents. This provides us with a panel of observations that follows the children, their siblings and their parents for potentially their entire lifespan.

3.1 Parental Shocks

Health shocks are identified in the National Patient Registry, which covers hospitalizations from both private and public hospitals. It contains information on the exact date of admission, the duration of the hospitalization, and detailed diagnoses following the International Classification of Diseases and Related Health Problems (ICD-10 system).

We consider three types of health shocks: **Cardiovascular** shocks, including myocardial infarction of the heart or brain; **Cancer** diagnoses, including malignant cancers of any type; and **Mental health** episodes that require hospitalization.[2] We also aggregate the three aforemen-

[2]The specific ICD-10 diagnoses that define each health shock are the following. Cardiovascular: I20-I24, I6. Cancer: C00-C97, D00-D09. Mental Health: F00-F99. Of the latter, around half are related to substance abuse, mostly alcohol. We found no differential effects between substance-abuse related and other mental health hospitalizations.

tioned health shocks into a variable called **Any Health Shock**. We use the first occurrence of each health shock by restricting them to shocks that have not been preceded by the same type of diagnosis in the previous 5 years.

Mortality shocks are identified using administrative registers that contain information on the exact date of the event. There is of course a large number of deaths that are preceded by a health shock. We try to address this with further restrictions that depend on each sample of analysis. They are laid out in Section 3.3.

Parental background is measured with information from the administrative registers. In addition to parental gender, we use information on mother's income and her cohabitation status with the child's father. The child's parents in the registers are defined as the biological parents or legal parents in case of adoptions. For some children, the registers do not list the personal identifier of both mother or father, we include them as long as we have information on at least one parent. For the heterogeneity analyses, we focus on maternal characteristics (as there are very few children without a maternal personal identifier, less than 0.2%). We split the sample into mothers whose household per-capita disposable income is in the bottom quartile vs the top three. Disposable income is a variable provided by Statistics Denmark, taking into account each person's household income and size. Next, we observe whether a mother is cohabiting with, or married to, the biological father of the child. If she is not (either living alone or with another partner), she is classified as "single" for our heterogeneity analyses.

3.2 Child Personality

We obtain our measures of the outcome of interest, child socio-emotional skills, from four waves of a nation-wide survey of public school children, the "Danish Well-being Survey" (DWS)[3]. This survey was introduced in 2015, and until 2018 it was mandatory for all Danish public schools to administer this self-report survey. The survey therefore approaches representativeness at the national level and is less prone to sample selection problems than small voluntary samples.[4] Public schools ("Folkeskole") cover grades 0-9, and we use the survey version given to older students, grade 4-9 (about age 11-16). The three traits of Conscientiousness, Agreeableness,

[3]For general information, see https://emu.dk/grundskole/undervisningsmiljo/trivselsmaling.

[4]It was typically administered during a regular school class in the school's computer room, led by a designated teacher. Schools had to upload the data according to certain standards, which included that all questionnaires should be linked to the students' national identification number. We are therefore able to combine the survey data with data described above on parental health shocks.

and Emotional Stability, as well as Academic Self-Concept, can be measured with selected items (questions) from the survey, as shown by Andersen et al. (2015, 2020). Not only do the items have good internal consistency, but they also correlate well with the relevant items from the Big Five Inventory (John and Srivastava, 1999), as demonstrated in a validation study with a separate data collection in Andersen et al. (2020). The survey remained the same throughout the period, there was only a re-ordering of questions between 2015 and 2016. Thus, we have an unbalanced panel structure, for which we construct the following four scores that measure the otherwise unobserved personality traits:

Conscientiousness, or how *responsible, and careful* one behaves, and one's tendency to *finish work*, is measured with the items "I can complete tasks and projects that I've committed to," "During class, I can concentrate well," "If interrupted during class, I can quickly concentrate again" (Cronbach's α measure of reliability in the full DWS sample, pooled over ages: $\alpha = .69$).[5] **Agreeableness**, reflecting *cooperation and empathy*, draws on "I try to understand my friends' feelings when they are sad or upset," and "I am good at collaborating with others" ($\alpha = .40$). Neuroticism (the reverse of **Emotional Stability**) reflects *vulnerability to stress.* We use the items "I often feel lonely," "My fellow students accept me for who I am," and "I always feel safe at school" ($\alpha = .70$). **Academic Self-Concept** is assessed by "I am doing well academically in school" and "I am making good academic progress in school" ($\alpha = .80$). This trait is not part of the Big Five, but it is predictive of future academic progress (Gensowski et al., 2020).

To measure personality traits, we generate four scores for each individual by first standardizing all items individually to mean zero and standard deviation one, by child's gender, grade, and calendar year, and second, forming the simple average and re-standardizing them. Using these standardized dependent variables means that the estimated coefficients can be interpreted as effects in terms of percentages of a standard deviation. The standardization helps us identify the effects of parental health shocks that are not influenced by other mechanisms that may be happening simultaneously. First, it is well documented in the literature that personality traits display typical developmental maturation patterns, which are changes in traits that appear consistently with age (see, for example van den Akker et al., 2014; Soto, 2016). Adolescence, in particular, is a time during which there are distinct decreases (dips) in Conscientiousness and Agreeableness (Soto et al., 2011) and academic self-esteem (Gensowski et al., 2020). In

[5]The corresponding Cronbach's alphas for the sub-sample of respondents who experience a parental health shock are equivalent or higher: Conscientiousness $\alpha = .70$, Agreeableness $\alpha = .43$, Emotional Stability $\alpha = .71$, Academic Self-Concept $\alpha = .81$.

the context of our analyses, we worry that by comparing personality traits measured after a shock to those measured before the shock, we confound the effect of the shock with spurious age-related differences that reflect overall maturation patterns. The standardization by school grade avoids picking up these spurious effects. Sex effects are also present (Soto et al., 2011), thus standardizing by sex (together with grade) removes differential developments over time by sex. The standardization by year takes out survey-wave specific effects (such as, for example, the re-ordering of items from 2015 to 2016).

3.3 Samples of Analysis

We report the descriptive statistics for the full sample of respondents to the DWS, compared to the two samples of analysis that we introduce below, in Table 1.

Short-run analyses. For the short run analyses, we exploit the panel dimension of the well-being survey data available from 2015 to 2018. Each year there were about 260,000 survey responses.[6] This amounts to 1,026,664 child-year observations from 457,227 children for whom we observe the four socio-emotional skills of interest.

We obtain individual-level variation within each child by restricting the sample of analysis to children who experience a parental shock in between any two DWS waves. We observe both the exact date of the survey and of the shock, so there is a very low probability of assigning the timing of the shock wrong. We only consider health shocks where the parent who experienced the health shock survived at least one year. Otherwise, the shock is considered a mortality shock and assigned to the year where the death occurred. Note that some health shocks might be preceded by symptoms that could affect the child in anticipation. While this is likely the case for mental health, the occurrence of a stroke or a cancer diagnosis is likely to come unexpectedly (Fadlon and Nielsen, 2020). Our analysis allows us to test for these anticipation effects, which could be different depending on the type of shock.

Our sample of analysis contains 10,904 unique children who experience a parental shock and 33,249 child-year observations. We identify 1,253 deaths and 9,679 health shocks of which 3,076 are cardiovascular shocks, 4,074 are cancer shocks, and 2,644 are mental health episodes.[7] These

[6] Precise numbers are 2015: 242,380, 2016: 268,047, 2017: 265,935, 2018: 250,302.

[7] The sum of the disaggregated health shocks is greater than the number of aggregated health shocks because if a child experiences different types of parental health shocks, such as a paternal cancer and a maternal cardiovascular shock, these shocks will both be considered separately for the disaggregated definitions, but when using the

Table 1: Descriptive Statistics

	DWS Sample		Shocked Short-run	Shocked Long-run
	Mean	S.D.	Difference	Difference
Conscientiousness	0.000	1.000	-0.040^{***}	-0.025
Agreeableness	0.000	1.000	-0.002	-0.033^{**}
Emot.Stability	0.000	1.000	0.008	0.005
Acad.Self-Concept	0.000	1.000	-0.029^{***}	-0.023
Age	13.531	1.739	0.117^{***}	0.000
Female	0.491	0.500	0.006^{**}	0.004
Parents College	0.501	0.500	-0.031^{***}	-0.010
Mother Income Lowest Quar.	0.239	0.426	0.054^{***}	0.119^{***}
Single Mother	0.306	0.461	0.069^{***}	0.039^{***}
Cohort Mother	1972.6	5.049	-1.177^{***}	0.141^{*}
Observations	1 026 664		33 249	3 772

Note: Showing mean and standard deviations (S.D.) for the entire sample of children responding to the DWS 2015-2018 (DWS Sample) and the sub-samples of children who experienced a short-run shock (that occurred in between DWS waves) or a long-run shock (that occurred to sibling pairs before they reach age 15). The columns denoted "Difference" report t-tests of means for each shocked subsample, comparing to the full DWS Sample. Note that the long-run sample compares only children who are 15 years old in both samples. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

are on average 418 deaths per year and 3,226 combined health shocks per year.

In comparison to the full sample of DWS respondents, this sub-sample of children who experience a shock between any of the DWS waves scores less favorably on some socio-emotional skills, and is different in terms of parental background—see the third column of Table 1. We will discuss this further below.

Long-run analysis For the long-run analysis, we study the effect of the timing of parental shocks on children's socio-emotional skills measured at age 15,[8] with a parents fixed effects strategy. Hence, we keep all children who answered the DWS at age 15 (166,665 children) and focus on those who experienced a parental shock before age 15 (32,732 children).

We further restrict the sample to siblings (pairs or triplets) who have experienced the same parental shock at different ages (hence excluding twins). To avoid further reducing the sample size, we consider all health shocks together and do not impose a survival period, therefore we identify the compounded effect of both the health shocks and any potential death that followed them. Importantly, we ensure that the health shock experienced by the siblings is the same, either cardiovascular, cancer, or mental health. The resulting sample of analysis contains 3,772

aggregated definition, only the earliest of those shocks will be included.

[8] We choose this age because it is the latest age with full sample size.

children.

The final column of Table 1 shows that this sub-sample differs from the full DWS sample in having less favourable socio-emotional skills, but it differs less than the short-run sample. This is explained by the less strict restriction of having experienced a parental shock over a much longer period of time. The differences are also less significant due to the small sample size.

4 Evidence for the Effects of Parental Health Shocks on Child Socio-emotional Skills

Children whose parents suffer a health shock have, on average, significantly less favorable socio-emotional skills than children of parents who do not, in terms of Conscientiousness and Academic Self-Concept (as shown in Table 1). A naive comparison of these two groups of children would lead us to conclude that parental health shocks produce large and significant differences in some socio-emotional traits in children. Yet, this comparison is flawed because parents who suffer from severe health shocks are different *ex ante*, and are likely to have children that differ *ex ante* as well, so that one cannot attribute differences in skills to the shocks. The naive comparison in Table 1 conflates the causal effect of a parental health shock with selection "into" the shocks.

4.1 Evidence on Short Run Effects

We exploit the panel dimension of the data on socio-emotional skills of the child and parent health, and employ two strategies to obtain causal effect estimates of the effect of parental health shocks on child socio-emotional skills.

4.1.1 OLS with Child Fixed Effects

The first strategy uses child-level fixed effects, identifying the effect of a parental health shock from within-child variation. Intuitively, this compares a child after a shock to him- or herself before the shock. The estimation model is:

$$Y_{it} = \alpha + \beta D_{it} + \phi_i + \epsilon_{it} \quad \text{for } t \in 2015, 2018 \tag{1}$$

where Y_{it} is child i's standardized trait at time t; D_it is an indicator variable that takes 1 from time t and onward if a parental shock took place between $t-1$ and t, and ϕ_i is an individual fixed effect. Under the assumption of no time-covarying unobservables, the parameter β identifies the causal effect of a parental shock on children's socio-emotional skills in the short-run. This strategy is comparable to Le and Nguyen (2017). It is a short-run measure in our setting because skills are observed until at most 3 years after the shock. Note also that β is not time-varying in the specification of Eq. (1), therefore capturing the average of the effects of the shock throughout the short-run post-shock period. (We allow for dynamics in our second strategy in the section below.)

Table 2: The Short Run Effect of Parental Shocks. Child Fixed Effects Estimates

	(1) Conscientiousn.	(2) Agreeablen.	(3) Emot.Stability	(4) Acad.Self-Concept	(5) # Shocks
Death	-0.049^{*}	-0.014	0.014	-0.028	1,253
	(0.03)	(0.03)	(0.03)	(0.03)	
Any Health Shock	-0.022^{**}	-0.001	-0.008	-0.004	9,679
	(0.01)	(0.01)	(0.01)	(0.01)	
Cardiovascular	-0.016	-0.034^{*}	0.013	-0.020	3,076
	(0.02)	(0.02)	(0.02)	(0.02)	
Cancer	-0.026^{*}	0.009	-0.011	0.014	4,074
	(0.01)	(0.01)	(0.01)	(0.01)	
Mental Health	-0.014	0.033^{*}	-0.013	-0.008	2,644
	(0.02)	(0.02)	(0.02)	(0.02)	

Note: Each cell reports the β coefficient of interest from estimating Eq. (1) separately for each personality trait of the children and for each type of parental shock. Each β coefficient identifies the causal effect of experiencing a given parental shock on the children's skills, which are standardized by child's sex, grade, and calendar year to have mean zero and standard deviation 1. Standard errors in parentheses clustered at the child level. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

Table 2 summarizes the results from estimating Eq. (1). The most severe shock, arguably, is that of a parent passing away. A parental death has a small significant effect on the child's Conscientiousness, reducing it by .05 of a standard deviation in the period following the death (one to three years after). This socio-emotional skill is similarly decreased (by over .02 of a standard deviation) from the summary measure of any health shock, particularly by cancer. On the one hand, this finding is important because Conscientiousness is regarded as a "super trait"—it is associated with many productive outcomes in terms of education, the labor market, health, and others. On the other hand, the effects are objectively quite small.

At the same time, we also notice that the other three traits we measure in the DWS are *not* significantly reduced by the loss of a parent or an indicator for any of the three severe health shocks considered: Agreeableness, Emotional Stability, and Academic Self-Concept are not significantly affected. This finding points to children being remarkably robust to even drastic shocks to their parents' health. One would especially have expected Emotional Stability to react to these challenging events, but it is clearly not - some point estimates are even positive. Note that these findings hold despite a substantial sample size for each test, and that the shocks being considered are quite severe. They can be interpreted causally under the assumption that within children's repeated measurements, the shock does not coincide with other unobservable events.

We can exclude, with a 95% confidence bound, harmful effects for Emotional Stability of more than .03 of a standard deviation from Any Health Shock and .04 from parental mortality; and for Agreeableness and Academic Self-Concept we can exclude reductions of more than .02 and .08. Conscientiousness can be decreased by up to .10 of a standard deviation from parental death and .04 from Any Health Shock.

Note that there are two lessons from separating out Any Health Shock into its components of cardiovascular shocks, cancer and mental health diagnoses: First, Conscientiousness is significantly reduced from a cancer diagnosis, which seems to drive the overall finding. Second, the null finding for Agreeableness hides both a harmful effect of a cardiovascular shock (which reduces Agreeableness by .03 of a standard deviation) together with a *beneficial* effect from the parent having a mental health episode.

The overall conclusion we draw from Table 2 is that of relative robustness of children's socio-emotional skills, despite some moderately negative effects on Conscientiousness, and possibly Agreeableness. The reason is that even those significant effect sizes are relatively small—especially in comparison to other effect sizes that are known for personality traits. In our sample, for example, the gender gap in (standardized) Agreeableness is 38% of a standard deviation (higher for females), and females score on average 29% of a standard deviation lower on Emotional Stability (see Table S.1). Children of parents with at least some post-secondary education score 29% of a standard deviation higher on Conscientiousness than children of less educated parents. (The corresponding gaps in Agreeableness are 16%, Emotional Stability 12%, and Academic Self-Concept 27%). From the literature, the evidence on the effects of schooling and other interventions on personality traits also show that these effects are of a different magnitude. For example, increasing schooling from 12 to 13 or more years increases Self-esteem by more than 50% of a standard deviation (Heckman et al., 2006). Randomized interventions have been reported to boost socio-emotional skills by up to 57% of standard deviations (see summary in Almlund et al., 2011). In comparison to these findings, it seems that children's socio-emotional skills are only weakly affected by severe parental health shocks.

Since the detrimental effects of parental health shocks on child socio-emotional skills are rather small, they would also have small effects on other life outcomes of the child, such as earnings or education, if we were to do a simple extrapolation exercise. The summary presented in Almlund et al. (2011), for example, shows that the effect of Conscientiousness on years of schooling is up to .18 of a standard deviation, and of Emotional Stability .09. Thus, taking the short-run effects of Table 2, we could exclude greater reductions in education than .007 of a standard deviation in schooling from any parental health shock on Conscientiousness, and by .003 from Emotional Stability (because the lower bound is so small with the point estimate being positive). Almlund et al. (2011) also present estimates of the effects of standardized personality traits on earnings, where Conscientiousness increases log earnings by .041 and Emotional Stability by .036. Therefore, if the short-term effects of Any Health Shock in Table 2 persisted throughout the children's adult working lives, their annual earnings would decrease by no more than .098% (Conscientiousness) or .108% (Emotional Stability). Even from parental death would we not expect more detrimental effects than a reduction of education by 0.018 of a standard deviation via Conscientiousness, if the short-run effects of Table 2 were extrapolated to the longer term, and we would exclude larger wage effects than 0.41% from the mortality shock's effect on Conscientiousness.

4.1.2 Heterogeneity by Child and Parent Gender

It is possible that the overall results in Table 2, which pool both the sex of the child and of the parent, hide important heterogeneities. Fathers and mothers may differentially affect children's acquisition of the different socio-emotional skills. Mühlenweg et al. (2016) found mothers' health to be significantly more important for child skills (with no effect of fathers), and Le and Nguyen (2017) remarked on specifically paternal mental health being important. Additionally, there is a literature discussing the greater vulnerability of boys relative to girls in terms of family disadvantage (Autor et al., 2019; Brenøe and Lundberg, 2018), or showing that mothers' investments are more reactive to their own mental health status for their daughters than their sons (Baranov et al., 2020). Therefore, it is important to also split the sample of children by sex to test whether boys are affected more by parental health shocks than girls.

Table 3 shows effects on boys, and Table 4 on girls, of health shocks split by whether they occur to the mother or father.[9] We also report the interaction coefficient between health shock and child sex in Table S.2.

These heterogeneity analyses show that there are a few significant differences between shocks coming from the mother vs the father, and that boys are not generally affected more negatively than girls.

For boys, the harmful effect of a parental death on Conscientiousness is entirely driven by losing their father, as the point estimate of losing a mother is insignificant and positive. Similarly, the effects of health shocks are larger if they happened to boys' fathers (the interaction terms for the difference to mothers are all negative in column 3, without being statistically significant). The reduction in Agreeableness from a cardiovascular shock to their parents is equally important between the parents, while the positive reaction to a mental health diagnosis stems from mothers. Agreeableness is one of two cases where in boys, pooling parents masks two significant effects: Firstly, losing their mother significantly *increases* Agreeableness by .13 of a standard deviation (one of the largest point estimates)—while losing a father reduces it insignificantly. This *positive* effect of a severe shock on Agreeableness for boys could not be seen in Table 2. Similarly, Academic Self-Concept of boys is reduced following the mental health diagnosis of their *father*, but not their mother (which has a positive point estimate even, combining to a near-zero pooled effect in Table 2).

Girls' Conscientiousness seems to suffer more from Any Health Shock arising to their fathers than their mothers—similarly to boys, although the magnitude of the effects are larger (.04 and .09 reduction from Any and Cancer shock, vs .03 in boys). Indeed, their Conscientiousness is only decreased from their *fathers* having cancer, not their mothers (a statistically significant difference). We can also observe that the positive effect of a mental health diagnosis for Agreeableness, which was observed in Table 2, stems mostly from the effect of a *paternal* diagnosis on girls, and possibly from a maternal diagnosis on boys. In girls, the difference between the parents is significant. The point estimates even have opposing signs. Girls' Academic Self-Concept

[9] Since these regressions include a direct interaction tests of the effect of the shock by parental gender (every column called "Diff."), we drop the few children who experience a shock from both the mother and father. The share of excluded children ranges from 0.34% (cancer) to 1.16% (any health shock).

is unaffected by shocks to either the mother or the father.

Table 3 and Table 4 do not test directly whether boys are more vulnerable in terms of socio-emotional skills than girls. We include an interaction term explicitly in Table S.2. Overall, there are very few significant interaction terms for all of the health shocks, suggesting no greater vulnerability to boys of parental health shocks. Out of 32 tests, 3 are statistically significant: boys decrease more in Agreeableness from Any Health Shock or a Mental Health diagnosis to the father, but increase more than girls in Agreeableness from a Mental Health diagnosis to the mother. Bereavement from the father does not have a significant interaction term either, although all point estimates are negative, which would point to a greater susceptibility to this type of loss for boys. Yet bereavement from the mother has a significantly positive interaction terms for boys in terms of Agreeableness.

Table 3: The Short Run Effect of Parental Shocks by Parental Gender. Effect on Boys. Child Fixed Effects Estimates

	Conscientiousness			Agreeableness			Emot.Stability			Acad.Self-Concept			# Shocks	
	(1) Father	(2) Mother	(3) Diff.	(4) Father	(5) Mother	(6) Diff.	(7) Father	(8) Mother	(9) Diff.	(10) Father	(11) Mother	(12) Diff.	(13) Father	(14) Mother
Death	−0.081*	0.029	−0.111	−0.081	0.131*	−0.212**	0.018	−0.036	0.054	−0.064	0.062	−0.126	416	220
	(0.04)	(0.07)	(0.08)	(0.05)	(0.07)	(0.09)	(0.05)	(0.07)	(0.08)	(0.05)	(0.07)	(0.08)		
Any Health Shock	−0.029	−0.012	−0.017	−0.029	−0.005	−0.024	−0.026	0.010	−0.035	−0.022	0.024	−0.046*	2,451	2,399
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)		
Cardiovascular	−0.027	−0.017	−0.010	−0.042	−0.048	0.006	−0.021	0.041	−0.062	−0.002	−0.042	0.040	1,098	488
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.05)		
Cancer	−0.033	−0.025	−0.008	−0.005	−0.023	0.018	−0.010	0.001	−0.012	0.014	0.035	−0.022	856	1,177
	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)		
Mental Health	−0.027	0.010	−0.037	0.007	0.053	−0.047	−0.059	0.013	−0.072	−0.084*	0.052	−0.137**	532	768
	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)	(0.06)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.06)		

Note: This table reports the results for the sub-sample of boysonly, distinguishing the parental shocks by whether they are experienced by the father or the mother. Each cell from columns "Father" and "Mother" reports the β coefficient from Eq. (1) that identifies the causal effect of experiencing a given parental shock on the children's socio-emotional skills, which are standardized by child's, grade and calendar year to have mean zero and standard deviation 1. Columns "Diff." report the coefficient on the interaction term between the indicator for the respective shock and the gender of the shocked parent, estimated over the sample of boys, who experience a parental shock to either the mother or the father. Standard errors in parentheses clustered at the child level. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

Table 4: The Short Run Effect of Parental Shocks by Parental Gender. Effect on Girls. Child Fixed Effects Estimates

	Conscientiousness			Agreeableness			Emot.Stability			Acad.Self-Concept			# Shocks	
	(1) Father	(2) Mother	(3) Diff.	(4) Father	(5) Mother	(6) Diff.	(7) Father	(8) Mother	(9) Diff.	(10) Father	(11) Mother	(12) Diff.	(13) Father	(14) Mother
Death	−0.071	−0.039	−0.032	0.014	−0.096	0.110	0.027	0.016	0.011	−0.015	−0.083	0.068	392	221
	(0.04)	(0.07)	(0.08)	(0.05)	(0.07)	(0.09)	(0.05)	(0.06)	(0.08)	(0.05)	(0.07)	(0.08)		
Any Health Shock	−0.039**	−0.003	−0.036	0.029	−0.004	0.032	0.009	−0.027	0.036	−0.011	−0.010	−0.001	2,377	2,337
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)		
Cardiovascular	0.010	−0.030	0.040	−0.022	−0.011	−0.011	0.027	0.042	−0.015	−0.020	−0.022	0.002	1,021	444
	(0.03)	(0.04)	(0.05)	(0.03)	(0.05)	(0.06)	(0.03)	(0.05)	(0.06)	(0.03)	(0.04)	(0.05)		
Cancer	−0.085***	0.017	−0.102***	0.038	0.028	0.010	0.016	−0.038	0.054	0.006	0.001	0.004	800	1,225
	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)		
Mental Health	−0.049	−0.013	−0.036	0.114***	−0.049	0.163***	−0.004	−0.030	0.026	−0.011	−0.018	0.007	603	724
	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)		

Note: This table reports the results for the sub-sample of girlsonly, distinguishing the parental shocks by whether they are experienced by the father or the mother. Each cell from columns "Father" and "Mother" reports the β coefficient from Eq. (1) that identifies the causal effect of experiencing a given parental shock on the children's socio-emotional skills, which are standardized by child's, grade and calendar year to have mean zero and standard deviation 1. Columns "Diff." report the coefficient on the interaction term between the indicator for the respective shock and the gender of the shocked parent, estimated over the sample of girls, who experience a parental shock to either the mother or the father. Standard errors in parentheses clustered at the child level. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

4.1.3 Heterogeneity by SES

Parental health shocks are expected to affect their children's socio-emotional skills at least partially because the shocks are thought to reduce parental resources, thereby leading parents to reduce their subsequent investments in children. Many of the mechanisms we discussed earlier for how parental shocks translate to child skills involve parental resources. Therefore, children of low-resource parents could be disproportionally affected by such shocks. It is important to emphasize that low-resource parents could be more *likely* ex ante to *experience* a shock, but that we are interested in the causal effect of a shock on child skills conditional on belonging to an at-risk group.

At-risk groups in terms of resources could be cash-strapped parents, or parents who have a tighter time constraint than others. Therefore, we analyze whether the shocks have greater effects on the children of low-income parents, or of single mothers.

As described earlier, we define low-income families on the basis of the household's disposable per-capita income, which takes household income and the size of the family (number of children and adults) into account. We take this measure for the mother as a marker for the child's relevant financial resources, even if the child is not living at the mother's household. Table 5 contrasts mothers in the bottom quartile of disposable income in 2014 to the other three quartiles. Unlike our previous results, the income split shows differential effects of *health shocks* vs *bereavement*. Losing a father or mother has no detrimental effect on any socio-emotional skill of children in the bottom income quartile; yet it significantly decreases Conscientiousness of children whose mother's income was in the top 3 quartiles. *Health shocks*, on the other hand, display some negative effects for the poorest children, in terms of Conscientiousness, Agreeableness, and Academic Self-Concept. The magnitudes are reductions of .03-.04, larger than we had seen for the full sample, but in the same ballpark. For non-disadvantaged children (income quartiles 2-4), parental health shocks aggregated reduce Conscientiousness, as we have found before (but by less than their disadvantaged peers). The *increased* Agreeableness in the wake of a parental *mental* health diagnosis that was already detected in Table 2 stems exclusively from these non-disadvantaged children.

Single mothers are defined in our analysis as mothers who, in 2014, do not live with the biological father of the child in question. These mothers may, therefore, not be actually living alone with their child (but with a new partner)—but "broken families" are traditionally associated with worse outcomes; either because of the aforementioned time constraints of truly single parents, or the challenges of bringing in new parent-figures and co-parenting with a physically distant biological father. In that sense, our definition would designate children who may have more disadvantages than children who consistently lived with their biological mother and father.

As Table 6 shows, for these children of single mothers, losing one of the two parents[10] has *less detrimental* effects than for children living with both parents. Note that there are no statistically significant effects of bereavement for children of single mothers. This could reflect that these children are less close to their biological father, and therefore see their investments less reduced than children who interact closely with their father.

Health shocks have mixed effects on children of single mothers; it would be a stretch to conclude that the children from separated families or from single mothers experience greater detrimental effects than the children from stable homes. The results on Conscientiousness look very similar between the two groups. Cardiovascular shocks reduce Agreeableness of children living with both parents, Academic Self-Concept of children to single mothers, and *increase* Emotional Stability of children of single mothers. A Mental Health diagnosis increases Agreeableness of the disadvantaged group here—in contrast to the result from Table 5, where the increase happened in the top three income quartiles.

We conclude that health shocks do not consistently affect children of single mothers vs children living with both parents differentially. Yet health shocks that happen to *low-income* parents have stronger effects on their children than when they happen to non-disadvantaged parents. Bereavement tends to affect the more advantaged group more. One interpretation of this finding could relate to the importance of parental quality in skill formation. Losing a parent who was very efficient at investing in their children's skills would have a greater effect than losing an absent parent or one whose quality of time or investment was lower (to the extent that disposable income and quality of parental investments can be associated).

[10] Recall that the definition of single mother is based on current residence with both biological parents, thus a child to a single mother is not necessarily one with literally only one parent; their father is most likely alive, but living apart.

Table 5: The Short Run Effect of Parental Shocks by Disposable Income Quartiles. Child Fixed Effects Estimates

	Conscientiousness		Agreeableness		Emot.Stability		Acad.Self-Concept		# Shocks	
	(1) Top Q2-Q4	(2) Bottom Q1	(3) Top Q2-Q4	(4) Bottom Q1	(5) Top Q2-Q4	(6) Bottom Q1	(7) Top Q2-Q4	(8) Bottom Q1	(9) Top Q2-Q4	(10) Bottom Q1
Death	-0.084^{***}	0.015	-0.024	0.003	0.015	0.017	-0.047	0.014	784	456
	(0.031)	(0.048)	(0.035)	(0.051)	(0.033)	(0.050)	(0.032)	(0.049)		
p-value of difference		0.0843		0.665		0.976		0.303		
Any Health Shock	-0.018^{*}	-0.031^{*}	0.017	-0.040^{*}	-0.011	0.001	0.009	-0.034^{*}	6,881	2,813
	(0.011)	(0.018)	(0.012)	(0.020)	(0.011)	(0.019)	(0.011)	(0.019)		
p-value of difference		0.534		0.0159		0.591		0.0503		
Cardiovascular	-0.015	-0.015	-0.022	-0.054	0.001	0.038	-0.013	-0.025	2,082	956
	(0.020)	(0.031)	(0.021)	(0.035)	(0.021)	(0.032)	(0.021)	(0.032)		
p-value of difference		0.999		0.440		0.331		0.751		
Cancer	-0.021	-0.050	0.024	-0.043	-0.006	-0.023	0.024	-0.024	3,244	813
	(0.015)	(0.032)	(0.016)	(0.037)	(0.015)	(0.035)	(0.015)	(0.034)		
p-value of difference		0.414		0.0963		0.650		0.200		
Mental Health	-0.011	-0.020	0.059^{**}	-0.009	-0.019	-0.006	0.019	-0.049	1,585	1,034
	(0.023)	(0.031)	(0.025)	(0.033)	(0.023)	(0.033)	(0.024)	(0.034)		
p-value of difference		0.828		0.101		0.744		0.0983		

Note: Each cell reports the β coefficient from Eq. (1) that identifies the causal effect of experiencing a given parental shock on the children's socio-emotional skills, in the respective sub-sample by quartile of disposable income of the mother. The child skills are standardized by child's sex, grade and calendar year, to have mean zero and standard deviation 1. Standard errors in parentheses clustered at the child level. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

Table 6: The Short Run Effect of Parental Shocks by Single Mother Status. Child Fixed Effects Estimates

	Conscientiousness		Agreeableness		Emot.Stability		Acad.Self-Concept		# Shocks	
	(1) Both Par.	(2) Single Moth.	(3) Both Par.	(4) Single Moth.	(5) Both Par.	(6) Single Moth.	(7) Both Par.	(8) Single Moth.	(9) Both Par.	(10) Single Moth.
Death	-0.111^{***}	0.019	-0.031	0.005	-0.027	0.059	-0.061^{*}	0.008	648	607
	(0.036)	(0.038)	(0.040)	(0.042)	(0.036)	(0.041)	(0.036)	(0.040)		
p-value of difference		0.311		0.323		0.840		0.880		
Any Health Shock	-0.022^{**}	-0.020	-0.005	0.008	-0.017	0.008	0.002	-0.015	6,204	3,583
	(0.011)	(0.016)	(0.012)	(0.018)	(0.012)	(0.017)	(0.012)	(0.017)		
p-value of difference		0.292		0.330		0.914		0.783		
Cardiovascular	-0.021	-0.006	-0.047^{**}	-0.006	-0.016	0.073^{**}	-0.002	-0.055^{*}	2,014	1,065
	(0.020)	(0.030)	(0.022)	(0.034)	(0.021)	(0.031)	(0.021)	(0.031)		
p-value of difference		0.296		0.311		0.772		0.642		
Cancer	-0.024	-0.029	0.012	0.001	-0.012	-0.009	0.017	0.007	2,877	1,205
	(0.015)	(0.026)	(0.017)	(0.029)	(0.016)	(0.028)	(0.016)	(0.027)		
p-value of difference		0.292		0.314		0.956		0.839		
Mental Health	-0.019	-0.010	0.017	0.050^{*}	-0.019	-0.006	-0.025	0.011	1,351	1,295
	(0.026)	(0.026)	(0.027)	(0.029)	(0.026)	(0.028)	(0.026)	(0.029)		
p-value of difference		0.295		0.361		0.963		0.890		

Note: Each cell reports the β coefficient from Eq. (1) that identifies the causal effect of experiencing a given parental shock on the children's socio-emotional skills, in the respective sub-sample by whether or not the mother lives with the biological father of the child ("Both Par.") or alone/with a new partner ("Single Moth."). The child skills are standardized by child's sex, grade and calendar year, to have mean zero and standard deviation 1. Standard errors in parentheses clustered at the child level. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

4.1.4 Event Studies

Child fixed effects are a credible identification strategy, but they do not give insight into the dynamics of how the effects of shocks play out in the child's skills over time. Fixed effects could also hide potentially interesting anticipation effects. This is particularly salient in our context, where some diagnoses, such as mental health diagnoses, are likely to occur after the family has already experienced the effects of symptoms.

Event study regressions can provide insight into such dynamics. Due to the sharp occurrence of the shocks, as defined by the date of diagnosis, we can estimate the following event study regressions:

$$Y_{it} = \alpha + \sum_{t \neq -1} \beta_t \cdot t + \varepsilon_{it} \tag{2}$$

where Y_{it} is child i's standardized trait at time relative to shock t (with $t = 0$ already affected). Since we observe up to 4 waves of the well-being surveys for each child, we can identify parameters up to three years after the shock (periods 0, 1 and 2) for children who experienced the shock right after they took the first survey in 2015, and up to 2 years before the shock (periods -1 and -2). Note that we do not include individual fixed effects in this event-study model, since a potential linear trend would not be identified (as pointed out by Borusyak and Jaravel, 2017). We perform this event study on the same sample of children as the OLS estimates presented above.

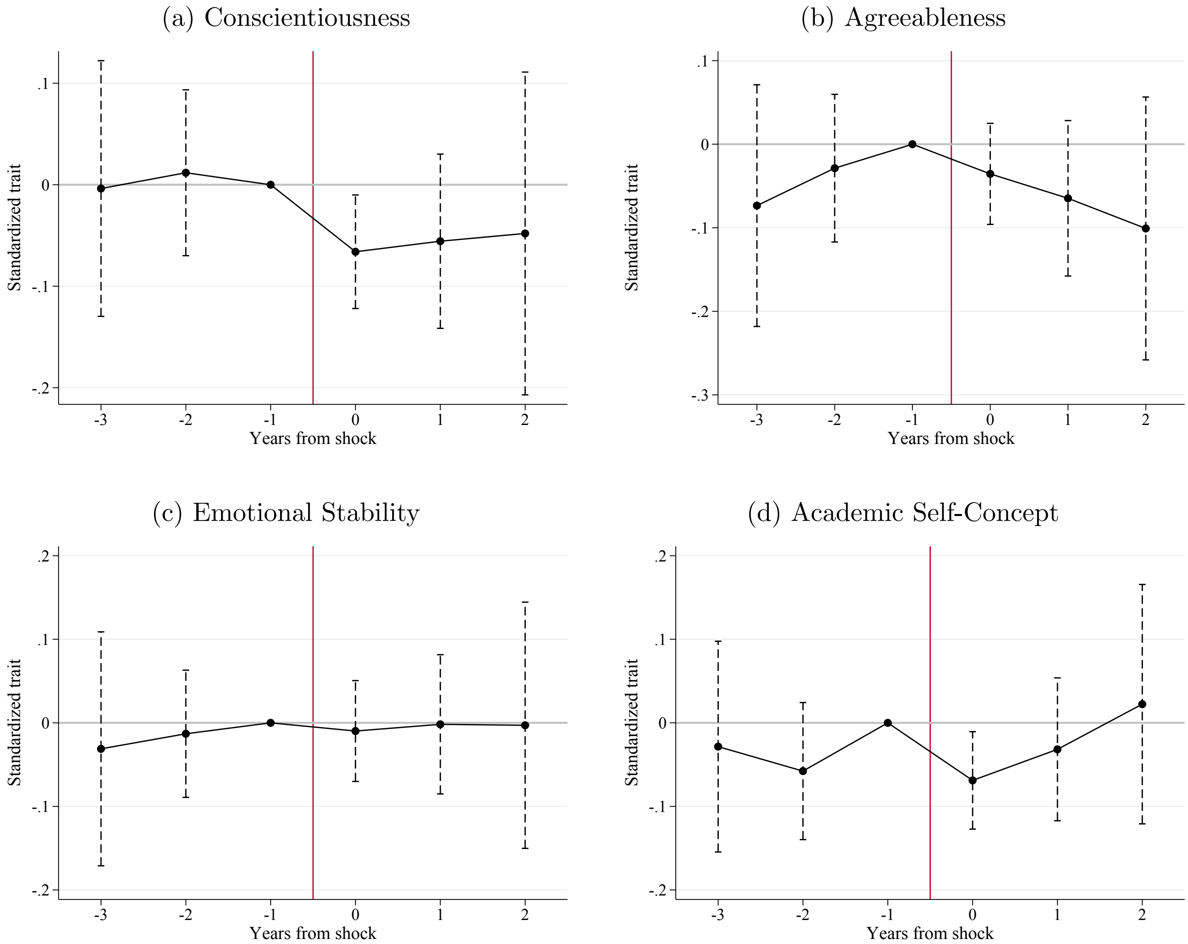
Figure 1 plots the β_t coefficients estimated from Eq. (2) that show the dynamics of the different socio-emotional skills around the *death* of a parent, and Fig. 2 around the combined parental health shocks.[11] Clearly, despite the possibility of important *anticipation* effects, there are no statistically significant dynamics. Consider the case of Conscientiousness, where we had identified significant negative effects from a parental death or the aggregate health shocks (Table 2): in the periods leading up to the shock (periods -2 and -1), children have no statistically significant differences in Conscientiousness, if anything the time trend before the health shock looks like it *increases* leading up to the shock. Thus, nothing suggests that children's Conscientiousness picks up pre-diagnosis effects of the parental health shock.

As for dynamics after the shock, coefficients tend to be more negative, particularly in period

[11] See Appendix Figs. S.1 to S.4 for results disaggregated by type of health shock and child's gender.

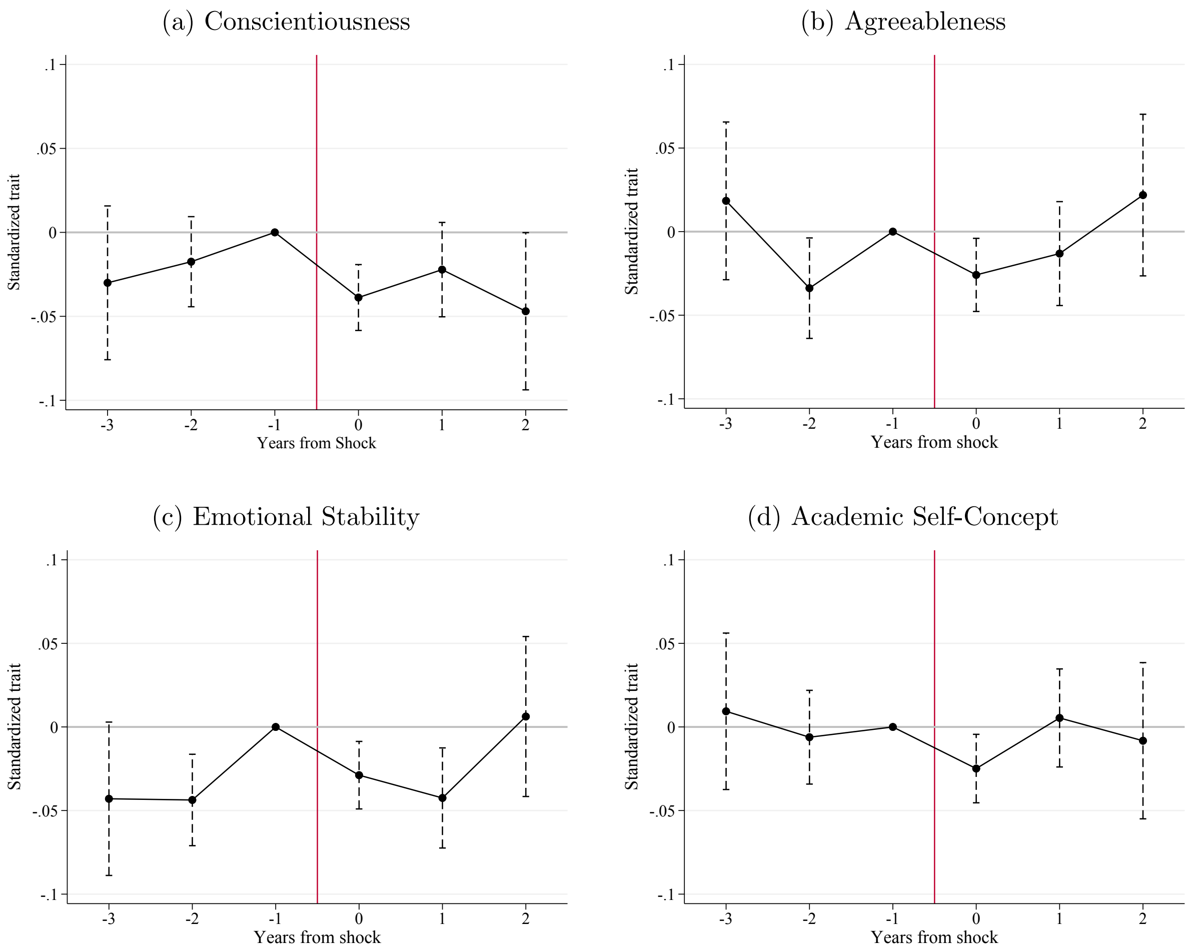
0, immediately after the shock. For the case of Conscientiousness, the coefficient for $t = 0$ is negative and statistically significant while coefficients in periods 1 and 2 remain negative but not significantly different from zero. Despite the increased noise in these event study estimations, particularly in periods further away from period -1 where the number of observations decreases, the dynamics of the traits are consistent with the estimations from the pooled child-fixed effects strategy presented in Table 2.

Figure 1: Event Study: Death



Note: These figures show the β_t coefficients estimated from Eq. (2) describing the dynamics of each socio-emotional skill around the time of Death, which is indicated with the vertical red line between -1 and 0. See Appendix Figs. S.1 to S.4 for the results from dissagregating the health shocks, and from a split by child's gender. The confidence intervals of each coefficient at the 95% level are calculated from standard errors clustered at the individual level.

Figure 2: Event Study: Any Health Shock



Note: These figures show the β_t coefficients estimated from Eq. (2) describing the dynamics of each socio-emotional skill around the time of Any Health Shock, which is indicated with the vertical red line between -1 and 0. See Appendix Figs. S.1 to S.4 for the results from dissagregating the health shocks, and from a split by child's gender. The confidence intervals of each coefficient at the 95% level are calculated from standard errors clustered at the individual level.

4.2 Evidence on Long Run Effects

The previous sections have painted a picture of relative robustness of children's socio-emotional skills, with small negative effects of parental health shocks on Conscientiousness.

Yet as discussed earlier, one could surmise that the effects of parental shocks do not materialize immediately, but accumulate over time into long-run effects. Under this hypothesis, children who have experienced a shock a longer time ago would have less advantageous socio-emotional skills later in life.

Both the pooled child-fixed effects and the event study design are limited to the study of shocks that occurred during the 4 years during which we observe the socio-emotional skills in the DWS, at ages 10 to 16. While the event study lets us explore dynamics to some extent, the latest we observe child outcomes is three years after the parental health shock. To explore the long-run effects of parental shocks that affect children from earlier ages, we employ an empirical strategy that identifies these effects comparing siblings. This strategy has also been employed by Laird et al. (2020) on Danish data to study the effect of divorce on educational attainment, and by Chen et al. (2009) to study the effect of a parental death on educational attainment in Taiwan. Specifically, we estimate the following model over a sample of sibling pairs who experienced the same parental health shock at different ages from 0 to 14:

$$Y_{ipa} = \alpha + \sum_{s=1}^{13} \beta_s \cdot I(AgeShock_i = s) + \phi_p + \gamma X_i + \epsilon_{ipa} \tag{3}$$

where Y_{ipa} is the standardized trait of child i, born to parent p, measured at age a (in our case, 15 years); $AgeShock_i$ is an indicator for child i experiencing a shock at age s; ϕ_p is a parent fixed effect; and X_i is a vector of controls, including birth order and gender of the child. The β_s parameters identify the causal effect of experiencing a parental shock at a given age with respect to experiencing it at a baseline age (here, age 14). With this strategy we consider all shocks a child can experience from age 0 to age 14. We are, however, restricted to analyzing sibling pairs who have lived through the same parental shock and who have both completed the DWS at age 15. Therefore, with four waves of the survey, the sample of siblings considered can be born at most four years apart, and the gap in a given shock occurring between the two can also be at most four years.[12]

[12] This leads to larger confidence intervals of the coefficients for shocks experienced at ages further from the baseline age of 14, as they are the compounded effect from smaller gaps in the shocks experienced by siblings.

Figure 3 implements this parent fixed effects strategy for the measure of Any Health Shock. We plot the estimated effect of experiencing a shock at a given age relative to experiencing it at age 14, by comparing siblings of different ages. Children's traits are robust to experiencing shocks earlier in live, with the exception of Conscientiousness, which decreases as shocks occur earlier. The non-linear effects of early shocks on Conscientiousness are, in general, significant and their size is up to 1 standard deviation for shocks experienced during the first years of the children's lives. Note that the large standard errors are due to the small size of the sample and the estimation with parent fixed effects.

Since these flexible age-by-age estimates of Eq. (3) look rather linear, we also estimate the following linear specification to gauge the magnitude of the effect

$$Y_{ipa} = \alpha + \beta \cdot AgeShock_i + \phi_p + \gamma X_i + \epsilon_{ipa} \tag{4}$$

This specification only differs from Eq. (3) in that the age at which each child i experiences the shock enters linearly, with the linear effect given by the parameter β.

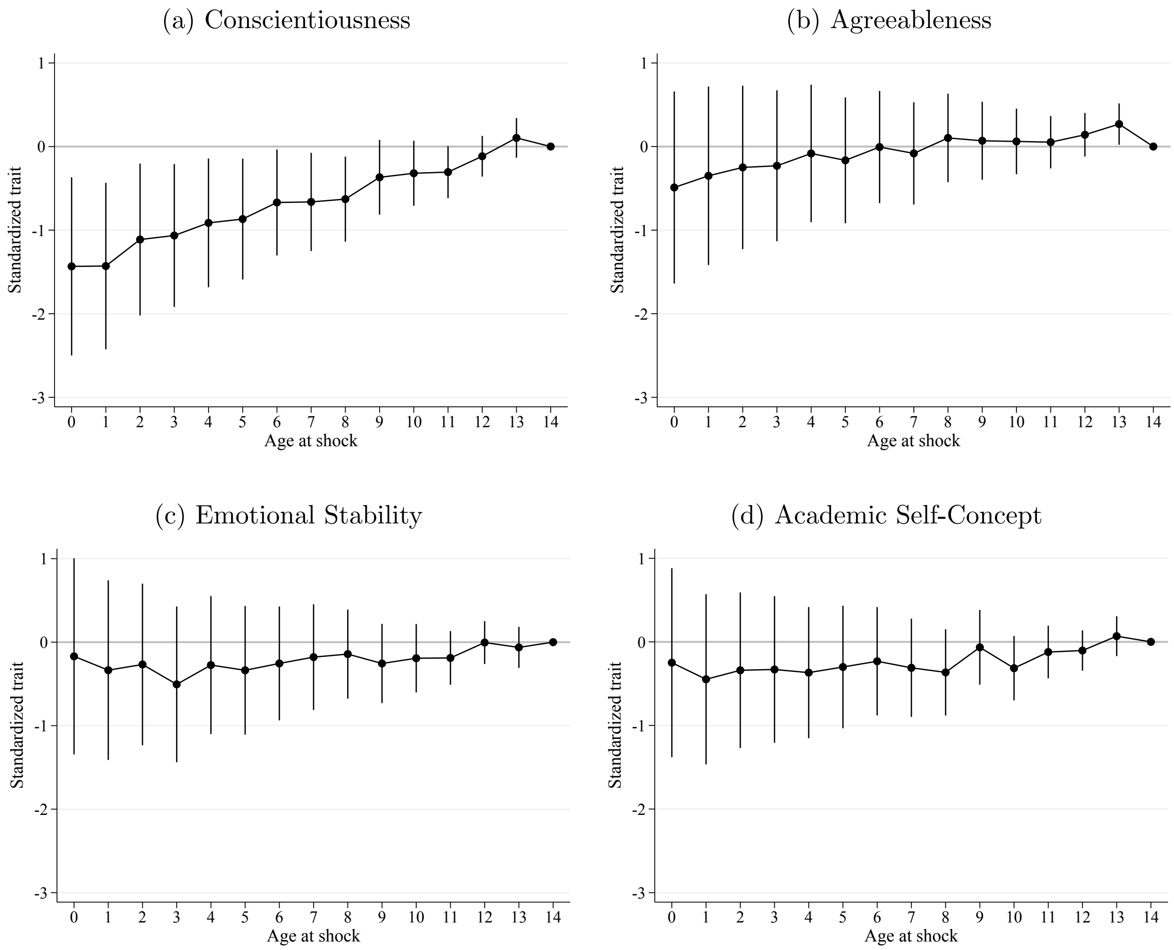
Table 7: The Long Run Effect of Parental Shock. Linear Estimates from Parent Fixed Effects

	(1) Conscientiousness	(2) Agreeableness	(3) Emot.Stability	(4) Acad.Self-Concept
Age at shock	0.110***	0.0382	0.0233	0.0378
	(0.037)	(0.041)	(0.040)	(0.038)
Observations	3,772	3,772	3,772	3,772

Note: This table reports the β coefficient estimated for Eq. (4) for each socio-emotional skill, which is standardized by child's sex, grade, and calendar year to have mean zero and standard deviation 1, and is measured at age 15. The coefficients identify the linear effect of experiencing a parental health shock one year later, closer to the baseline age of 14. Robust standard errors clustered at the parental level are reported in parentheses. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$

Table 7 shows results consistent with the more flexible estimation presented in Fig. 3: that experiencing a shock earlier, rather than later, has significantly more harmful effects on Conscientiousness by age 15. Experiencing a parental health shock one year later increases Conscientiousness by 0.11 of a standard deviation. The other traits also have positive estimates for experiencing the shock later in life, but are smaller and not statistically significant. These results would be in line with an accumulation of disadvantage, which could stem from the dynamic complementarity in skill formation over the life cycle, or scarring and socialization (as discussed

Figure 3: The Long Run Effect of Experiencing a Parental Shock at Different Ages



Note: These graphs report the β_s coefficients from Eq. (3), where each coefficient identifies the causal effect of experiencing a parental health shock at a given age relative to experiencing the same shock at age 14. Identification comes from comparing siblings who both experienced the same shock but at different ages. We report confidence intervals at the 95% level from clustered standard errors at the parental level.

in Section 2).

Perhaps surprisingly, the long-run effect of experiencing a shock one year later is larger than the short-run effect of experiencing the shock as estimated in Table 2. Note however that the two estimates are not directly comparable: the short-run results in Table 2 use a child's pre-shock socio-emotional scores as a counterfactual to identify the effect of the shock immediately after, while the long-run effects from Table 7 use the sibling's score measured at the same ages (15) as a counterfactual to identify the effect of having experienced a shock at different ages. Also, the short-run sample considers shocks that can occur in between ages 10-15 while the long-run shocks considers shocks between ages 0-14. The large standard errors of the long-run estimates also make us interpret the long-run results with caution.

One explanation under which both results would be reconciled, if we engaged in the thought experiment of considering them strictly comparable, is the case where there is no effect on socio-emotional skills in the period immediately after the shock, but one emerges some time after. This could be seen as an "incubation" period. Under this hypothesis, our short-run strategy would not capture the effects from the post-incubation period, while the long-run strategy, which evaluates the traits at a later age of 15, would. If this is the case, the long-run strategy should not find an effect for shocks experienced right before the socio-emotional skills are measured, and we indeed see a flat or less pronounced slope for shocks between ages 12 and 14.

With these caveats, we see the long-run results as complementary of the main analysis. It is very difficult to obtain very early pre-shock measures of socio-emotional skills for a large sample of children, and then have long-run follow-up data. Therefore, even though we interpret the results with caution, they provide important suggestive evidence for the importance of the timing of early shocks — this calls for further research in this area to further explore long-run dynamics of children's personality formation and to speak directly to the literature on life cycle skill formation.

5 Testing Non-Response in the DWS as a Function of Shocks

An important challenge for our analysis is the potential of selective non-response from the sample of children who experience a parental shock. We define two types of non-response. First, non-participation: children who experience a parental shock might be less likely to participate in the survey (e.g. not attending school when the survey was distributed). Second, partial-response: children who participate in the DWS after experiencing a parental shock might be less likely to answer the specific questions that we use to construct their measured socio-emotional skills.

This challenge is common to most other studies, but our access to register data for the entire population of children in Denmark gives us the unique opportunity to quantify the degree of non-participation in the survey following the parental shocks. We also test the degree of partial-response among participants.

Non-participation in the DWS. We test whether children are less likely to participate in the DWS after they experience a parental shock compared to the years before the shock.

We use the full population of children in Denmark who were enrolled in schools that collected student responses to the survey in a given year, for a given age group.[13] We then focus on children who experienced a parental shock during the period in which the DWS was collected, and who participated in the DWS the year before the shock. This is a necessary restriction as we need to observe the date when the DWS was taken to assign the timing of the shock correctly. This is the same restriction we applied in our main analysis.

The resulting sample of analysis therefore contains only observations from children who attend a school where the survey was distributed, and who are observed at an age and year where they should have participated. By construction, all children in this sample took the survey a year before the parental shock occurred ($t = -1$). Therefore, we test whether the parental shock increases the likelihood of not participating in th DWS the years after ($t = \{0, 1, 2\}$) against the likelihood of not participating in the DWS the years before ($t = \{-2, -3\}$). Specifically, we estimate the following regression

$$Y_{it} = \alpha + \beta \cdot Post_{it} + \delta \cdot D_{-1} + \gamma X_i + \epsilon_{it} \tag{5}$$

[13]Note that we do not observe the class or grade of children who do not participate in the DWS, which is why we use age instead.

where Y_{it} is an indicator for not participating in the DWS, $Post_{it}$ is an indicator for time after the parental shock $t = \{0, 1, 2\}$ and D_{-1} is a dummy variable for the period just before the shock, that we exclude because by construction all children participate in the DWS in that period. X_i is a vector of controls composed by children's age interacted with gender.

We report the results of the non-participation test in column (1) of Table 8. We see that children who experience a parental shock have a slightly higher probability of not participating in the DWS the years after. The effects are small and only significant for mortality shocks (.032 of a standard deviation) and cancer shocks (.016 of a standard deviation).

The increased probability of non-participation in the DWS after a parental shock could bias our results if the non-participant children are a selected subsample, such as those affected the most by the parental shock. To be reassured that our results are robust to this possible bias, we replicate our analysis imputing the traits of the missing respondents with the least favorable and the most favorable outcomes from the observed distribution of children who participated in the DWS. This test is inspired by Lee (2009). The results of this exercise are reported in Appendix Table S.3. We observe that for the least favorable imputation, corresponding to the assumption that all non-participant children would have scored the worst outcomes (10^{th} percentile of the distribution) only a few coefficients are significantly different from zero, and all point estimates are below -0.075 of a standard deviation (in absolute terms), which are still fairly small effects.

Non-participation is very small in the DWS thanks to the way the DWS was distributed, reaching almost all children from the schools where it was distributed. However, note that non-participation is more likely to occur in smaller, voluntary surveys particularly if the respondents are the potentially shocked parents. Unfortunately, testing for selective non-participation in these cases is often unfeasible.

Partial response in the DWS. The second type of non-response would occur if children who experience a parental shock are less likely to answer the questions we use to construct the socio-emotional skills, and are therefore excluded from the analysis.

To test partial response we consider the full sample of children who participated in the DWS, and construct an indicator variable for when a child did not answer one or more of the questions used to construct the socio-emotional skills and therefore misses one or more trait. We then apply the same empirical strategy as for the short-run analysis, and estimate Eq. (1) for the partial-response dummy variable. The results are reported in column (3) of Table 8 and we find no evidence of a greater likelihood of missing traits (partial response) after experiencing a parental health shock. (For bereavement, the point estimate is 0.023 but it is not statistically significant from zero.)

Table 8: Non-Response as Function of Parental Shocks

	(1) Non-participation	(2) # Shocks	(3) Partial Response	(4) # Shocks
Death	0.032**	1,598	−0.023	1,368
	(0.01)		(0.014)	
Any Health Shock	0.002	11,720	−0.006	10,515
	(0.00)		(0.005)	
Cardiovascular	−0.007	3,777	0.002	3,371
	(0.01)		(0.009)	
Cancer	0.016**	4,773	−0.012	4,356
	(0.01)		(0.007)	
Mental Health	0.003	3,335	−0.006	2,922
	(0.01)		(0.01)	

Note: This table reports the results from two different tests of non-response for the different types of parental shocks. Each cell of column (1) reports the β coefficient from Eq. (5) estimated for each parental shock, capturing the increased likelihood of not participating in the DWS after experiencing a parental shock. Column (3) reports the β coefficients resulting from estimating Equation (1) for an outcome variable that takes one if a children did not answer one or more questions used to construct the socio-emotional traits. Columns (3) and (4) report the number of shocks that are considered in each estimation respectively. The number of shocks is larger in the test of non-participation because we also include shocked children who did not participate in the DWS. Robust standard errors clustered at the individual level. $^{*}(p < 0.10)$,$^{**}(p < 0.05)$,$^{***}(p < 0.01)$.

6 Discussion and Conclusion

We have presented causal evidence for the effects—or absence of effects—of parental health shocks on child socio-emotional skills. Our short run analyses (child fixed effects and event studies) suggest relative robustness of 11-16 year-old children's socio-emotional skills against even the severe parental health shocks considered. One trait, Conscientiousness, was consistently lowered in the wake of parental health events or bereavement, but the magnitude was small. Testing heterogeneity by SES, we did not find specific groups that were more or less at risk of seeing decreases in their socio-emotional skills following parental health shocks. Even in the complementary long-run analysis, three of the four skills tested are unaffected.

Was this to be expected? On the one hand, yes; considering the studies of Cobb-Clark and Schurer (2013) and Elkins et al. (2017). For adults, Cobb-Clark and Schurer (2013) convincingly showed that the personality trait of Locus of Control is invariant to life events including a range of types: family formation/dissolution, fertility, labor market shocks, retirement, and health shocks. Elkins et al. (2017) observe adolescents into adulthood for an eight-year span, and do not find any personality trait to respond systematically to the majority of common one-off family-, income-, and health-related shocks.

On the other hand, this was in no way to be expected for the sample of children in the age range 11-16 we have considered. Cobb-Clark and Schurer (2013) also show that personality changes are concentrated among the *young* (even if they are not related to the shocks tested). Childhood and early adolescence is the time in one's life during which personality traits are potentially the most malleable (see, for ex. Roberts and DelVecchio, 2000). The "plaster theory" contends that personality becomes fixed by the age of 30 only (McCrae and Costa Jr., 1996). Childhood is also the time during which parents still exert a considerable influence—thus leaving the door open for the largest spill-overs. We study the period in life during which one would expect the largest potential effects of parental shocks on socio-emotional skills. Furthermore, we consider more severe shocks than others in the literature who have found no effects once selection was taken care of (Le and Nguyen, 2017). Moreover, Kristiansen (2020) uses the same data as we do, and finds significant effects of parental health shocks on children's educational performance and attainment.

Note however, that comparing siblings who experience the same parental shock at different ages did provide suggestive evidence for significant long-run decreases in child's Conscientiousness. While we interpret these estimates with caution due to the large standard errors, we think they provide novel evidence on the formation of children's socio-emotional skills, potentially reflecting accumulation or incubation dynamics following early shocks. Providing this type of causal estimates demands large datasets over a long period of time, with information on both parents' health and children's socio-emotional skills, making our dataset ideally suited for the task. Still, we think further research is needed to better understand the long-run dynamics of parental shocks on children's personality.

We have tested whether parent health shapes child socio-emotional skills causally. We thereby contribute to the literature on life cycle skill formation, because child skills are shaped by parents' investments in terms of time and resources, and both of these are possibly affected by shocks to parents' health. A large literature documents how early childhood experiences drive long-run outcomes. Many of these experiences are intertwined and correlated with other parental characteristics. Therefore, it is important that we find that parental health shocks *in themselves* do not generate large differences in child skills, at least in the short-run.

References

Alam, S. A. (2015). Parental health shocks, child labor and educational outcomes: Evidence from Tanzania. *Journal of Health Economics 44*, 161–175.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 4*, Volume 4, Chapter 1, pp. 1–181. Elsevier B.V.

Andersen, S. C., M. Gensowski, S. Ludeke, and O. P. John (2020). A Stable Relationship Between Personality and Academic Performance from Childhood through Adolescence. An original study and replication in hundred-thousand-person samples. *Journal of Personality 01*(01), forthcoming.

Andersen, S. C., M. Gensowski, S. Ludeke, J. H. Pedersen, L. V. Beuchert-Pedersen, J. Niclasen, R. Piatek, and M. K. Thomsen (2015). Evaluering af den nationale trivselsmåling for folkeskoler–og forslag til justeringer.

Andrews, R. J. and T. D. Logan (2010). Health, Children's Own Health, and Test Score Gaps. *The American Economic Review: Papers and Proceedings 100*(2), 195–199.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. *American Economic Journal: Applied Economics 11*(3), 338–381.

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020). Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial. *American Economic Review 110*(3), 824–859.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The Relationship Between Economic Preferences and Psychological Personality Measures. *Annual Review of Economics 4*, 453–478.

Becker, A., H. C. Steinhausen, G. Baldursson, S. Dalsgaard, M. J. Lorenzo, S. J. Ralston, M. Döpfner, A. Rothenberger, and ADORE Study Group (2006). Psychopathological screening of children with ADHD: Strengths and difficulties questionnaire in a pan-European study. *European Child and Adolescent Psychiatry 15*(SUPPL. 1), 56–62.

Berger, E. M. and C. K. Spiess (2011). Maternal Life Satisfaction and Child Outcomes: Are They Related? *Journal of Economic Psychology 32*(1), 142–158.

Bhalotra, S. and S. B. Rawlings (2011). Intergenerational persistence in health in developing countries: The penalty of gender inequality? *Journal of Public Economics 95*(3-4), 286–299.

Bharadwaj, P., J. Eberhard, and C. Neilson (2017). Health at Birth, Parental Investments, and Academic Outcomes. *Journal of Labor Economics*, 695616.

Borusyak, K. and X. Jaravel (2017). Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume. *SSRN Electronic Journal*, 1–25.

Bratti, M. and M. Mendola (2014). Parental health and child schooling. *Journal of Health Economics 35*, 94–108.

Brenøe, A. A. and S. Lundberg (2018). Gender gaps in the effects of childhood family environment: Do they persist into adulthood? *European Economic Review 109*, 42–62.

Brickman, P., D. Coates, and R. Janoff-Bulman (1978). Lottery Winners and Accident Victims: Is Happiness Relative? *Journal of Personality and Social Psychology 36*(8), 917–927.

Chen, S. H., Y. C. Chen, and J. T. Liu (2009). The impact of unexpected maternal death on education: First evidence from three national administrative data links. *American Economic Review 99*(2), 149–153.

Cobb-Clark, D. A. and S. Schurer (2013). Two Economists' Musings on the Stability of Locus of Control. *The Economic Journal 123*(August), F358–F400.

Cuadros-Menaca, A., A. Gaduh, and G. Zamarro (2018). *The Effect of Parental Health Shocks on Non-Cognitive Skills Formation in a Developing Country.* Ph. D. thesis, Cali-Colombia.

Cubel, M., A. Nuevo-Chiquero, S. Sanchez-Pages, and M. Vidal-Fernandez (2016). Do Personality Traits Affect Productivity? Evidence from the Lab. *The Economic Journal 126*(May), 654–681.

Cunha, F. and J. J. Heckman (2007). The Technology of Skill Formation. *The American Economic Review: Papers and Proceedings 97*(2), 31–47.

Currie, J. and E. Moretti (2007). Biology as Destiny? Short- and Long- Run Determinants of Intergenerational Transmission of Birth Weight. *Journal of Labor Economics 25*(2), 231–264.

Dhanaraj, S. (2016). Effects of parental health shocks on children's schooling: Evidence from Andhra Pradesh, India. *International Journal of Educational Development 49*, 115–125.

Elkins, R. K., S. C. Kassenboehmer, and S. Schurer (2017). The stability of personality traits in adolescence and young adulthood. *Journal of Economic Psychology 60*, 37–52.

Fadlon, I. and T. H. Nielsen (2020). Family Labor Supply Responses to Severe Health Shocks: Evidence from Danish Administrative Records. *American Economic Journal: Applied Economics (forthcoming).*

Fletcher, J. M. (2013). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization 89*, 122–135.

Gensowski, M. (2018). Personality, IQ, and Lifetime Earnings. *Labour Economics 51*, 170–183.

Gensowski, M., S. G. Ludeke, O. P. John, and S. C. Andersen (2020). The emerging gender confidence gap: A nation-wide study of self-assessed and objective academic performance in children and adolescents. *University of Copenhagen, mimeo*.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Heineck, G. and S. Anger (2010). The returns to cognitive abilities and personality traits in Germany. *Labour Economics 17*, 535–546.

Hilbig, B. E., A. Gloeckner, and I. Zettler (2014). Personality and prosocial behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology 107*(3), 529–539.

John, O. P. and S. Srivastava (1999). The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In *Handbook of Personality: Theory and Research*, Number 510.

Johnson, E. and C. L. Reynolds (2013). The effect of household hospitalizations on the educational attainment of youth. *Economics of Education Review 37*, 165–182.

Kristiansen, I. L. (2020). Short and Long-Term Consequences of Severe Parental Health Shocks. CEBI working paper series 10-20, University of Copenhagen.

Laird, J., N. F. Nielsen, and T. H. Nielsen (2020, April). Differential Effects of the Timing of Divorce on Children's outcomes: Evidence from Denmark. CEBI working paper series 20-11, University of Copenhagen.

Le, H. T. and H. T. Nguyen (2017). Parental health and children's cognitive and noncognitive development: New evidence from the longitudinal survey of Australian children. *Health Economics* (26), 1767–1788.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies 76*(3), 1071–1102.

Lundberg, S. (2013). The College Type: Personality and Educational Inequality. *Journal of Labor Economics 31*(3), 421–441.

Lundberg, S. (2019). Non-Cognitive Skills as Human Capital. In C. R. Hulten and V. A. Ramey (Eds.), *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, pp. 219–243. University of Chicago Press.

McCrae, R. R. and P. T. Costa Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives*, pp. 51–87. New York: Guilford Press.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, and A. Caspi (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences 108*(7), 2693–8.

Mueller, G. and E. Plug (2006). Estimating the effects of personality on male and female earnings. *Industrial and Labor Relations Review 60*(1), 3–22.

Mühlenweg, A. M., F. G. Westermaier, and B. Morefield (2016). Parental health and child behavior: evidence from parental health shocks. *Review of Economics of the Household 14*, 577–598.

Nyhus, E. K. and E. Pons (2005). The effects of personality on earnings. *Journal of Economic Psychology 26*, 363–384.

Poropat, A. E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin 135*(2), 322–338.

Roberts, B. W. and W. F. DelVecchio (2000). The Rank-Order Consistency of Personality Traits From Childhood to Old Age: A Quantitative Review of Longitudinal Studies. *Psychological Bulletin 126*(1), 3–25.

Roberts, B. W., C. Lejuez, R. F. Krueger, J. M. Richards, and P. L. Hill (2014). What is Conscientiousness and How can it be Assessed? *Developmental Psycholgy 50*(5), 1315–1330.

Roberts, B. W., K. E. Walton, and W. Viechtbauer (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin 132*(1), 1–25.

Senne, J.-N. (2014). Death and schooling decisions over the short and long run in rural Madagascar. *Journal Of Population Economics 27*, 497–528.

Soto, C. J. (2016). The Little Six Personality Dimensions From Early Childhood to Early Adulthood: Mean-Level Age and Gender Differences in Parents' Reports. *Journal of Personality 84*(4), 409–422.

Soto, C. J., O. P. John, S. D. Gosling, and J. Potter (2011). Age Differences in Personality Traits from 10 to 65: Big Five Domains and Facets in a Large Cross-Sectional Sample. *Journal of Personality and Social Psychology 100*(2), 330–348.

Spengler, M., B. W. Roberts, O. Lüdtke, R. Martin, and M. Brunner (2016). The Kind of Student You Were in Elementary School Predicts Mortality. *Journal of Personality 84*(4), 547–553.

van den Akker, A. L., M. Dekovic, J. Asscher, and P. Prinzie (2014). Mean-Level Personality Development Across Childhood and Adolescence: A Temporary Defiance of the Maturity Principle and Bidirectional Associations With Parenting. *Journal of Personality and Social Psychology 107*(4), 736–750.

Widinger, T. A. (2011). Personality and Psychopathology. *World Psychiatry 10*, 103–106.

S Appendix

Table S.1: Association of Socio-Emotional Skills with Demographic Characteristics

	(1) Conscient.	(2) Agreeableness	(3) Emot.Stability	(4) Acad.Self-Concept
Female	-0.035^{***}	0.389^{***}	-0.292^{***}	-0.023^{***}
	(0.003)	(0.002)	(0.003)	(0.003)
Parents College	0.289^{***}	0.160^{***}	0.121^{***}	0.269^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Mother Income Lowest Quart.	-0.241^{***}	-0.151^{***}	-0.155^{***}	-0.216^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Single Mother	-0.246^{***}	-0.133^{***}	-0.172^{***}	-0.210^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1026664	1026664	1026664	1026664

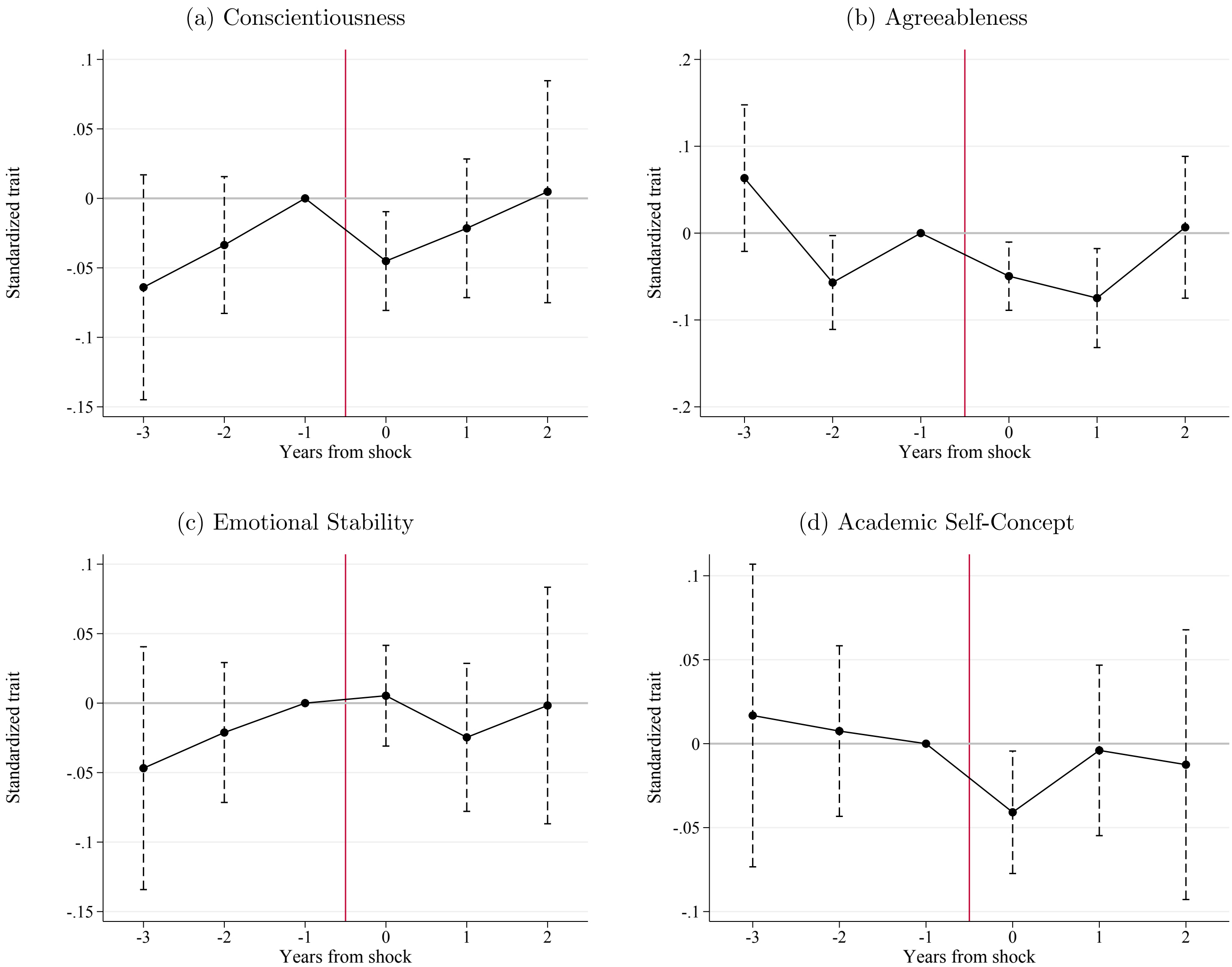
Note: This table shows the differences in socio-emotional skills by socio-demographic characteristics for the full DWS. Each cell reports the β coefficient from estimating the equation $Y_{it} = \alpha + \beta D_i + \epsilon_{it}$ where D_i is a variable that takes 1 if the child's gender is female, or their parents have college education, or their mothers' income is in the lowest quartile or if, sequentially, the mother is a single mother. Socio-emotional skills are standardized by child's gender, grade, and calendar year except for the estimation of the gender gap, where we do not standardize by gender. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

Table S.2: The Short Run Effect of Parental Shocks by Parent and Child Sex

	(1) Conscientiousness	(2) Agreeableness	(3) Emot.Stability	(4) Acad.Self-Concept
Death Father	-0.071^{*}	0.014	0.027	-0.015
	(0.04)	(0.05)	(0.05)	(0.05)
Death Father × Male	-0.010	-0.095	-0.009	-0.050
	(0.06)	(0.07)	(0.07)	(0.07)
Death Mother	-0.039	-0.096	0.016	-0.083
	(0.07)	(0.07)	(0.06)	(0.07)
Death Mother × Male	0.068	0.227^{**}	-0.052	0.145
	(0.10)	(0.10)	(0.09)	(0.09)
Health Shock Father	-0.039^{**}	0.029	0.009	-0.011
	(0.02)	(0.02)	(0.02)	(0.02)
Health Shock Father × Male	0.010	-0.058^{**}	-0.034	-0.011
	(0.03)	(0.03)	(0.03)	(0.03)
Health Shock Mother	-0.003	-0.004	-0.027	-0.010
	(0.02)	(0.02)	(0.02)	(0.02)
Health Shock Mother × Male	-0.009	-0.001	0.037	0.034
	(0.03)	(0.03)	(0.03)	(0.03)
Cardiovascular Father	0.010	-0.022	0.027	-0.020
	(0.03)	(0.03)	(0.03)	(0.03)
Cardiovascular Father × Male	-0.037	-0.020	-0.048	0.018
	(0.04)	(0.04)	(0.04)	(0.04)
Cardiovascular Mother	-0.030	-0.011	0.042	-0.022
	(0.04)	(0.05)	(0.05)	(0.04)
Cardiovascular Mother × Male	0.013	-0.037	-0.000	-0.020
	(0.06)	(0.06)	(0.06)	(0.06)
Cancer Father	-0.085^{***}	0.038	0.016	0.006
	(0.03)	(0.03)	(0.03)	(0.03)
Cancer Father × Male	0.052	-0.043	-0.027	0.008
	(0.04)	(0.05)	(0.04)	(0.04)
Cancer Mother	0.017	0.028	-0.038	0.001
	(0.02)	(0.03)	(0.03)	(0.03)
Cancer Mother × Male	-0.042	-0.051	0.039	0.034
	(0.04)	(0.04)	(0.04)	(0.04)
Mental Health Father	-0.049	0.114^{***}	-0.004	-0.011
	(0.04)	(0.04)	(0.04)	(0.04)
Mental Health Father × Male	0.022	-0.107^{*}	-0.055	-0.074
	(0.06)	(0.06)	(0.06)	(0.06)
Mental Health Mother	-0.013	-0.049	-0.030	-0.018
	(0.03)	(0.04)	(0.04)	(0.04)
Mental Health Mother × Male	0.024	0.102^{**}	0.043	0.070
	(0.05)	(0.05)	(0.05)	(0.05)

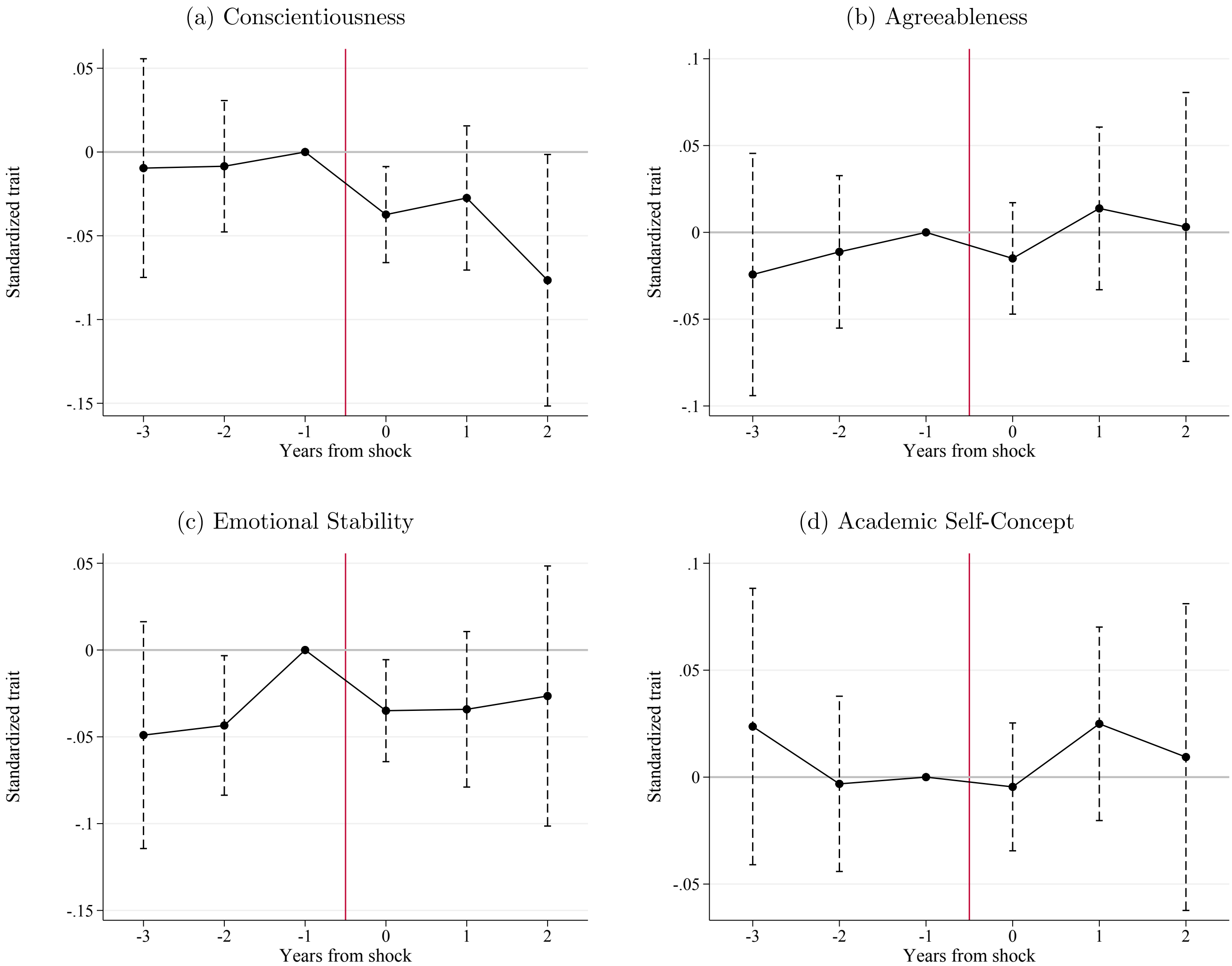
Note: This table presents the main results for the short-run effects estimated from Eq. (1) over the pooled sample of boys and girls but adding an interaction term if the child is male and experienced the a parental shock (such as "Death Father × Male." Parental shocks are also disaggregated by parental gender. This table therefore subsumes both Table 3 and Table 4 offering a statistical test for whether boys and girls are significantly affected differently by each type of parental shock. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.

Figure S.1: Event Study: Cardiovascular Shock



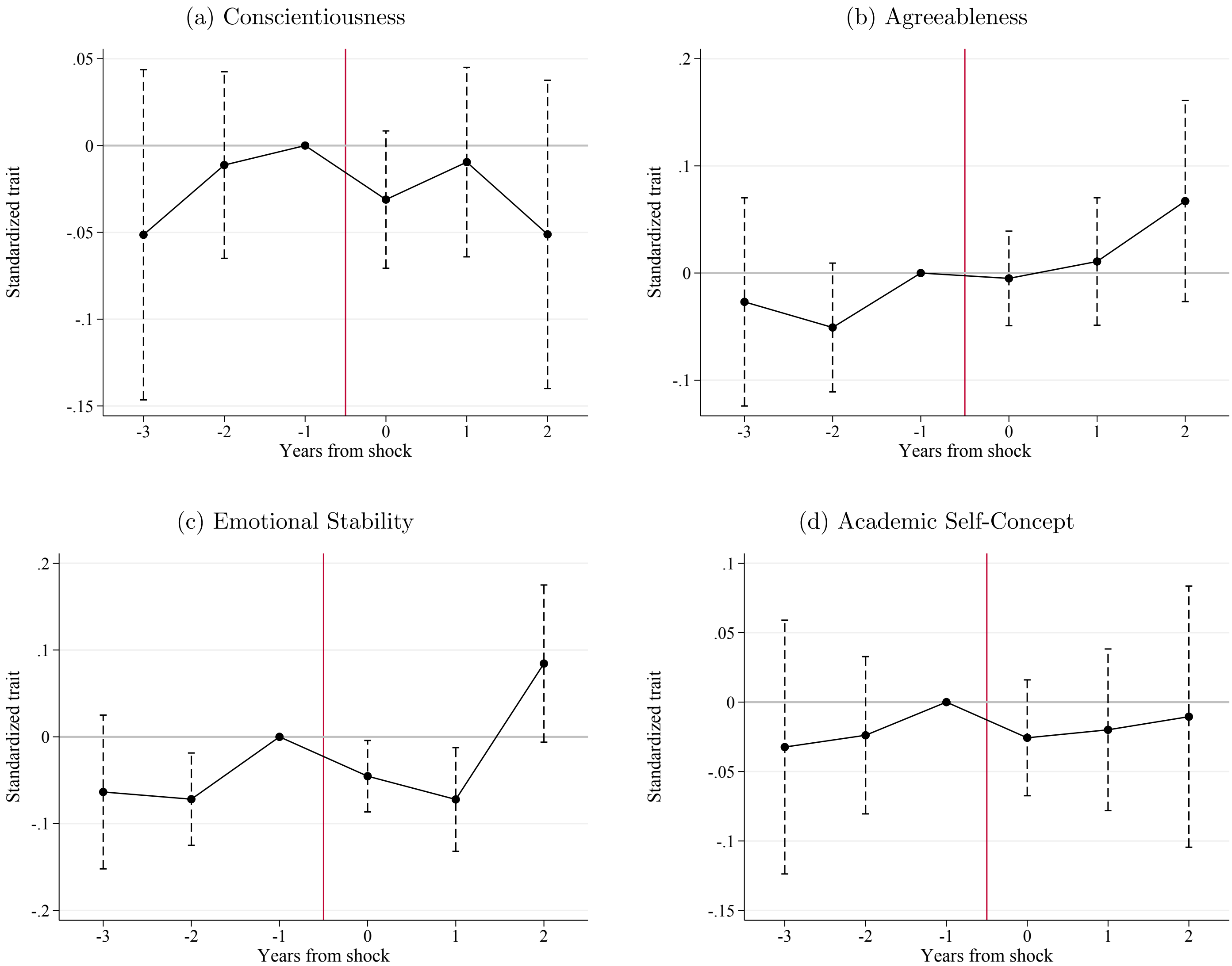
Note: See notes to Fig. 1 for further notes.

Figure S.2: Event Study: Cancer



Note: See notes to Fig. 1 for further notes.

Figure S.3: Event Study: Mental Health Diagnosis

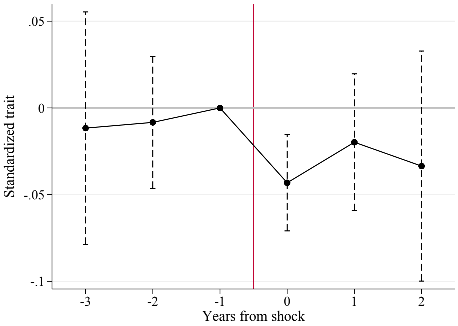


Note: See notes to Fig. 1 for further notes.

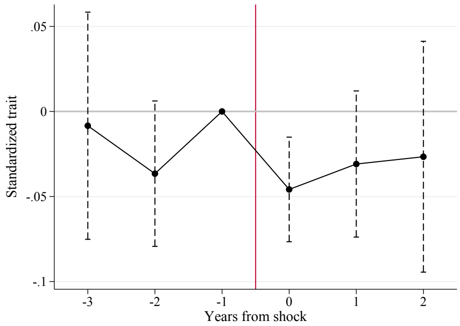
Figure S.4: Event Study: Any Health Shock, by Child Gender

A. Boys

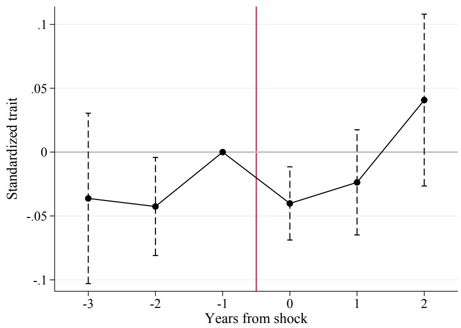
(a) Conscientiousness



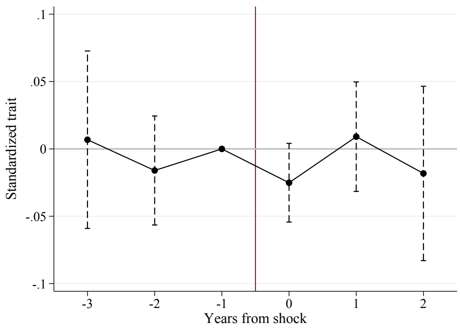
(b) Agreeableness



(c) Emotional Stability

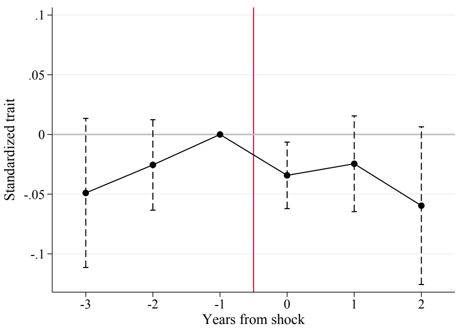


(d) Academic Self-Concept

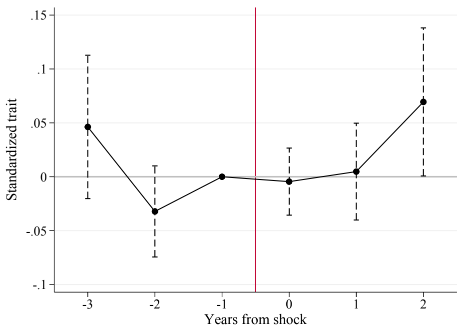


B. Girls

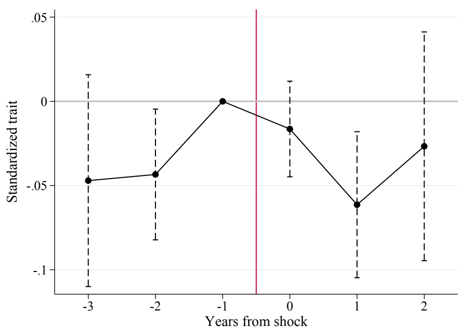
(e) Conscientiousness



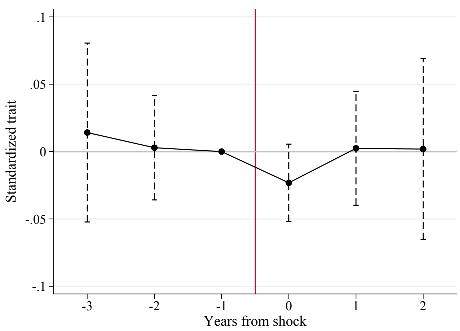
(f) Agreeableness



(g) Emotional Stability



(h) Academic Self-Concept



Note: See notes to Fig. 2 for further notes. Here replicating the strategy simply split by child sex.

Table S.3: Individual Fixed Effects, Short Run

	(1) Conscientiousness	(2) Agreeableness	(3) Emot. Stability	(4) A. Self-Concept	(5) # Shocks
A. Baseline					
Death	-0.054	0.026	0.022	0.002	$1,436$
	(0.04)	(0.05)	(0.04)	(0.04)	
Any Health Shock	-0.018	0.001	0.025^{*}	-0.003	$10,699$
	(0.01)	(0.02)	(0.02)	(0.02)	
Cardiovascular	-0.008	-0.004	0.035	-0.035	$3,436$
	(0.03)	(0.03)	(0.03)	(0.03)	
Cancer	-0.026	-0.015	0.022	0.004	$4,452$
	(0.02)	(0.02)	(0.02)	(0.02)	
Mental Health	-0.010	0.054^{*}	0.037	0.026	$2,960$
	(0.03)	(0.03)	(0.03)	(0.03)	
B. Lower bound					
Death	-0.073^{*}	-0.043	-0.055	-0.037	$1,531$
	(0.04)	(0.04)	(0.04)	(0.04)	
Any Health Shock	-0.027^{*}	-0.030^{**}	-0.003	-0.020	$11,221$
	(0.01)	(0.02)	(0.02)	(0.01)	
Cardiovascular	-0.034	-0.057^{**}	-0.030	-0.077^{***}	$3,610$
	(0.03)	(0.03)	(0.03)	(0.03)	
Cancer	-0.029	-0.031	0.010	-0.006	$4,620$
	(0.02)	(0.02)	(0.02)	(0.02)	
Mental Health	-0.022	0.001	0.004	0.007	$3,145$
	(0.03)	(0.03)	(0.03)	(0.03)	
C. Upper bound					
Death	0.073^{*}	0.110^{**}	0.090^{**}	0.123^{***}	$1,531$
	(0.04)	(0.05)	(0.04)	(0.05)	
Any Health Shock	0.022	0.020	0.045^{***}	0.033^{**}	$11,221$
	(0.02)	(0.02)	(0.01)	(0.02)	
Cardiovascular	0.078^{***}	0.060^{**}	0.080^{***}	0.045	$3,610$
	(0.03)	(0.03)	(0.03)	(0.03)	
Cancer	-0.000	-0.001	0.039^{*}	0.026	$4,620$
	(0.02)	(0.02)	(0.02)	(0.02)	
Mental Health	0.014	0.039	0.040	0.047	$3,145$
	(0.03)	(0.03)	(0.03)	(0.03)	

Note: This table presents the result of the bounding exercise. Panel A presents the baseline estimates where children who do not participate in the DWS are not included. Note that this panel is equivalent to Table 2 except that we have excluded the observations from the year just before the parental shock, since by definition all children from this period participate in the DWS and including them in the estimation would bias the bounding exercise by adding one entire year of observations to the pre-shock period that will not be imputed. Instead, we follow the same strategy as we used to quantify the degree of non-participation and exclude the year before the shock. Panel B presents the estimates from a sample where for all children who did not participate in the DWS but who should have participated (based on the school they attend, their age and the calendar year), their traits have been imputed with the 10^{th} percentile of the observed distribution of children who participated. Panel C reports the results from imputing the most favorable outcomes to the non-participant children, based on the 90^{th} percentile of the observed distribution. See notes from Table 2 for more details. $^{*}(p < 0.10)$,** $(p < 0.05)$,*** $(p < 0.01)$.