DISCUSSION PAPER SERIES

DP14769

SOCIAL SECURITY, LABOR SUPPLY AND HEALTH OF OLDER WORKERS: QUASI-EXPERIMENTAL EVIDENCE FROM A LARGE REFORM

Itay Saporta-Eksten, Ity Shurtz and Sarit Weisburd

LABOUR ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

SOCIAL SECURITY, LABOR SUPPLY AND HEALTH OF OLDER WORKERS: QUASI-EXPERIMENTAL EVIDENCE FROM A LARGE REFORM

Itay Saporta-Eksten, Ity Shurtz and Sarit Weisburd

Discussion Paper DP14769
Published 16 May 2020
Submitted 14 May 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SOCIAL SECURITY, LABOR SUPPLY AND HEALTH OF OLDER WORKERS: QUASI-EXPERIMENTAL EVIDENCE FROM A LARGE REFORM

Abstract

We study the effects of public pension systems on the retirement timing of older workers and, in turn, the health consequences of delaying retirement by those workers. Causal inference relies on a social security reform in Israel that shifted payments from husbands to their (non-working) wives, thereby substantially reducing the implied tax on the husband's employment while keeping overall household wealth constant. Using administrative social security data, we estimate extensive-margin labor supply elasticities w.r.t. the average net-of-tax rate of about 0.43 for men over 65. Using the reform to instrument for employment, we find that working an additional full year at old age decreases longevity. This mortality effect occurs after age 75 and is driven by workers holding blue-collar jobs. Finally, we evaluate the effect of the reform on earnings. The results imply a small value for an additional year of life, suggesting that workers underestimate the health cost of employment at older ages.

JEL Classification: J10, J26, J22, H31

Keywords: Labor Supply, Social Security, tax reform, health, Mortality

Itay Saporta-Eksten - itaysap@tauex.tau.ac.il
Tel Aviv University, University College London and CEPR

Ity Shurtz - shurtz@bgu.ac.il
Ben-Gurion University of the Negev

Sarit Weisburd - saritw@post.tau.ac.il
Tel Aviv University and CEPR

Social Security, Labor Supply and Health of Older Workers: Quasi-Experimental Evidence from a Large Reform

Itay Saporta-Eksten (Tel Aviv University, IZA and CEPR)
Ity Shurtz (Ben-Gurion University of the Negev)
Sarit Weisburd (Tel Aviv University and CEPR)

May 2020

Abstract We study the effects of public pension systems on the retirement timing of older workers and, in turn, the health consequences of delaying retirement by those workers. Causal inference relies on a social security reform in Israel that shifted payments from husbands to their (non-working) wives, thereby substantially reducing the implied tax on the husband's employment while keeping overall household wealth constant. Using administrative social security data, we estimate extensive-margin labor supply elasticities w.r.t. the average net-of-tax rate of about 0.43 for men over 65. Using the reform to instrument for employment, we find that working an additional full year at old age decreases longevity. This mortality effect occurs after age 75 and is driven by workers holding blue-collar jobs. Finally, we evaluate the effect of the reform on earnings. The results imply a small value for an additional year of life, suggesting that workers underestimate the health cost of employment at older ages.

Keywords: labor supply, social security, tax reform, health, mortality

Acknowledgements: We are very grateful to Daniel Gottlieb, Miriam Shmelzer, Gabriella Heilbronn, and members of the National Insurance Institute of Israel staff for their help in conducting this study. Research funding from the Maurice Falk Institute for Economic Research in Israel, The Pinhas Sapir Center, and the Israel Science Foundation is gratefully acknowledged.

1. Introduction

Facing aging populations, almost half of all OECD countries implemented, in recent years, reforms to their public pension systems that encourage work among older workers.[1] Such policies, however, can potentially take a toll on older workers' health and longevity. Evidence on the effects of public pension systems on the employment of older worker and on the consequences of work at older ages on health is therefore key to informing a better design of such systems. In this paper, we use administrative data to draw causal inference on these two issues by leveraging a social security reform in Israel that changed the implied tax on delayed retirement while holding benefit generosity constant for a well-defined segment of the population—housewife households (hereinafter the "Housewives Reform" or the "Reform").

Concretely, we first study how the implied tax on delayed retirement, often induced by the existence of an earnings test in public pension systems, affects employment. Namely, we measure the extensive margin labor supply elasticity with respect to the average net-of-tax rate of older workers. While this is an important statistic in the context of public pension systems policy, the evidence pertaining to its size is mixed and inconclusive. The Reform provides a particularly favorable setting to estimate this statistic due to the existence of both a within-cohort unaffected group (non-housewife households) and a sharp eligibility age cutoff. Thus, we are able to implement a triple difference framework as well as a regression discontinuity framework. We then use the Reform to estimate the effect of employment at older ages on longevity. Despite the importance of the relationship between work and health and its increasing relevance for older workers, existing evidence on this issue is scant. Because the Reform created a change in employment incentives with almost no effect on wealth, it provides an opportunity to estimate this relationship. We therefore use eligibility to the Reform as an instrument for the additional employment of elderly workers to study this question.

Our analysis uses the Housewives Reform that was implemented in 1996. The Reform shifted public pension payments from husbands to their (non-working) wives, thereby reducing the implied tax on the husband's employment while keeping overall benefit generosity roughly constant. The Reform implied a 10,000 NIS (about $3,125) reduction in the implied annual tax on employment for husbands married to an eligible housewife—a housewife born after January 1st, 1931.[2] This reduction was large, roughly amounting to a 17% increase in the average net-of-tax rate.[3]

[1] According to an OECD report (OECD, 2015), between 2013 and 2015 almost all OECD countries implemented some reforms to their public pension systems, with almost half of the countries conducting reforms that target work incentives.

[2] All NIS and dollar figures are reported in real 1996 terms.

[3] Because the reform shifted payments from husbands to their wives, a change in household behavior may arise due to a shift in spouses' respective bargaining power (see Blundell, French and Tetlow, 2016 and Chiappori and

We apply a triple difference (DDD) specification to compare cumulative retirement rates of the full population of husbands in Israel married to housewives born in 1931 and 1932 (treatment group) and those married to housewives born in 1930 (comparison group) pre- and post-Reform implementation in 1996. The identifying assumption underlying this analysis is that a husband married to a housewife born in 1931 or 1932 would have followed a similar retirement trend to a husband married to a housewife born sometime in the previous calendar year absent the introduction of the Reform. To account for potential differences across these cohorts, we apply a third difference using husbands married to non-housewives that were born in the same years. We find that the Reform reduced retirement rates of treated husbands by about 5.4 percentage points at the year of its implementation (1996). Once husbands turn 70 years old there is no longer a tax disincentive to work, and indeed as husbands age, the differences in cumulative retirement rates between treatment and comparison husbands disappear. Our estimates imply that the elderly exhibit moderate to high extensive-margin labor supply elasticities of about 0.43.

Because the Housewives Reform targeted a very specific segment of the population (housewife households where the wife was born after 1931) with two similar comparison groups, it naturally lent itself to a DDD specification. However, it is also true that the closest comparison group is households with housewives born just a little bit too early to be included in the Reform. We therefore complement the DDD analysis with a regression discontinuity design (RDD), where we use the sharp cutoff in birth date of the housewives (January 1st 1931). Focusing on the effect of the Reform on retirement rates on impact, we find RDD estimates, which are close in magnitude to the DDD estimates. The RDD estimates corroborate the DDD results by showing that the response we measure arises sharply around the eligibility threshold of the Reform. We run two types of placebo tests for this analysis. First, we show that the retirement rates of husbands in housewife households do not show any shift in retirement behavior around their wives' birthday cutoff in 1994, i.e. prior to the 1996 change in legislation. Second, we show that this decrease in retirement rates around the birthdate cutoff (January 1st 1931) does not hold for non-housewife households.

Next, we use this exogenous shift in employment at older ages to examine the impact of work on longevity. We assess nonparametrically how the Reform changes the probability that affected husbands remain alive past any age between 65 and 85. Our analysis reveals that the Reform impacts survival probabilities specifically at older ages. While we find a positive association between employment and survival in our sample when applying OLS, when we instrument for employment using the Reform, we find that the effect of employment on longevity is *negative* and occurs years after retirement. Specifically, we find no effect of delaying retirement on mortality between the ages 65 and 74. However, years after the delay in retirement, at the age range of 75-85, we document a decline in the survival of affected husbands.

Mazzocco, 2017 for recent surveys). We discuss this issue in Section 5 and demonstrate that in our setting it is unlikely to be quantitatively important.

Overall, we find that working an additional year decreases longevity by 9 to 12 months. We further show that this effect is concentrated among blue-collar workers, who are more likely to be performing manual, physical tasks.

Finally, we estimate the effect of the Reform on earnings, finding that the Reform increased after-tax earnings (net of benefits lost) by about 18,900 NIS. Combining this result with the effect of the Reform on survival we recover the value that treated individuals attribute to an additional year of life. Such a calculation implicitly assumes that individuals internalize the health effect. Even when taking into account that the earnings increase was immediate while the health consequences occurred later in life, we find a very small value for an additional year of life, in the order of 126,000 NIS ($39,375). The small implied value calls into question the notion that when making employment decisions around retirement, workers fully understand the cost of employment in terms of their health and longevity and supports the view that workers tend to underestimate these costs.

This paper is related to the large literature on employment incentives created by the social security system. The pioneering work by Krueger and Pischke (1992) studies the "notch generation's" employment response to the large reduction in social security benefits. While they find a limited effect, recently, Gelber, Isen, and Song (2016), using administrative data and an RDD design, find that benefit reduction led to a substantial increase in labor supply that reflects a large income effect.[4] Fetter and Lockwood (2018) analyze the Old Age Assistance Program in the U.S. between 1930 and 1960 and find that it had a large negative impact on labor force participation of men aged 65-74.

Another body of literature studies how changes in the earnings test, which creates an implied tax on delayed retirement, impact extensive-margin labor supply.[5] Taking a structural approach, French (2005) shows that eliminating the earnings test in the U.S. system (canceling the implied tax on delayed retirement) has a larger impact on the timing of retirement than reducing benefits or delaying the benefits eligibility age. Song & Manchester (2007) study the implications of the U.S. earnings test removal of 2000 and find inconclusive evidence on labor force participation while Friedberg & Webb (2009) report an increase in employment following the elimination of the earnings test in 2000. Baker and Benjamin (1999) and Disney and Smith (2002) report small or no changes in participation for elimination of the tests in Canada and the UK, respectively.

[4] Additional body of work studies the defined early retirement age (ERA) and full retirement age (FRA). See works by Mastrobuoni (2009), Blau and Goodstein (2010), Behaghel and Blau (2012) and Manoli and Weber (2016). See also the volume edited by Gruber and Wise (2004) and Coile and Gruber (2007) for international micro evidence using the option value approach outlined in Stock and Wise (1990).

[5] The earnings test implies that earning above the earnings threshold means forgoing at least some retirment benefits.

In a recent study, Gelber et al. (2017) develop a method to estimate the elasticity of participation with respect to the average net-of-tax rate using the kink in the budget set induced by the earnings test. They find larger elasticities than those previously reported in this literature. Leveraging the richness of the Housewives Reform, we estimate extensive margin labor supply elasticities as well. However, we rely on a very different identification approach, taking advantage of the sharp discontinuity in tax rates implied by date of birth within the treated population (housewife households) as well as the existence of a non-treated population within birth cohort (non-housewife households). Remarkably, the magnitudes of the elasticities we find in our quasi-experimental setting are very similar to those found by Gelber et al. (2017).[6]

What is the relationship between employment and health outcomes? Research shows that job-loss has negative health implications (see e.g. Sullivan & Von Wachter (2009)), and that higher unemployment rates are associated with mortality (Gerdtham & Ruhm, 2006). However, these studies do not separate the role of employment from other aspects of job loss. Particularly, they do not distinguish between exogenous separations (layoffs), and endogenous employment decisions, which are much more relevant in the context of retirement choices. A small strand of literature uses policy changes, and eligibility cutoffs to specifically study how retirement age affects health and mortality, finding somewhat contradicting effects for short- versus long-term effects.[7] At the (very) short-term there is some evidence that early retirement has a negative effect on health. Fitzpatrick and Moore (2018) find a discontinuous increase in male mortality around the Early Eligibility Age in the United States. Kuhn, Wuellrich, and Zweimüller (2010) find that early retirement increases male deaths at early ages, close to retirement (before age 67). They show that the adverse effects are likely to be focused on involuntary rather than voluntary job losses. Using discontinuities in eligibility age in Germany, Giesecke (2019) demonstrates heterogeneity in the short-term impact of retirement on mortality, with decreasing mortality for low-earning manual worker, and increasing effects for high-earners.

Recent research draws a different picture for the long-term effects of retirement on health. Hallberg et al. (2014) apply a difference-in-differences strategy to a military pension reform in Sweden, and find that early retirement (before age 55) increased health conditions and reduced mortality in the long-term (ages 56-70). Analyzing a Dutch reform, Bloemen et al. (2017) find that early retirement reduced the probability

[6] Another related strand of research examines the implications of the earnings test for intensive margin labor supply decision. See works by Burtless and Moffitt (1985), Baker and Benjamin (1999), Friedberg (2000), Disney and Smith (2002), Gruber and Orszag (2003), Haider and Loughran (2008), Engelhardt and Kumar (2014), Gelber, Jones, and Sacks (2020) and Gelber et al. (2020).

[7] An older literature studies how measured and self-reported health indicators change around retirement ages (see for example Bound, J., and Waidmann, 2007; Neuman, 2008; Coe, N. B., and Zamarro, 2011). Most of these papers find positive effects of retirement on health indicators, and none report a negative effect. In a more recent study, Shai (2018) uses a change in the full retirement age for men in Israel to show that employment at older ages has a negative impact on health indicators while employed.

of men dying within 5 years by 2.6 percentage points.[8] Applying an instrumental variables approach using a Norwegian reform to early retirement age, Hernaes et al (2013) find neither a positive nor negative effect of retirement age on mortality between ages 67 and 77. Our work provides support for these results using a clean quasi-experimental source of variation in employment, keeping wealth almost constant, which is quite rare in this literature.

The paper proceeds as follows. In section 2 we provide a brief review of the social security system in Israel and the Housewives Reform. Section 3 presents the data. Section 4 provides analysis of the impact of the implied tax of delayed retirement on labor supply. Section 5 reviews our results with respect to the effect of employment on longevity. Section 6 concludes.

2. The Social Security System in Israel and the Housewives Reform

Israel has a universal pay-as-you-go social security pension system, where contributions are withheld from the worker's salary up to a contribution cap. Each worker then receives his/her retirement benefits from the system at the end of his/her working life. Eligibility for retirement benefits depends on an individual's age and the *eligibility period*. There are two different age requirements. The first is *retirement age* – individuals who reach this age qualify for retirement benefits subject to an *earnings test*. During the relevant time period of our study, the *earnings test* was set annually at roughly 20% above minimum wage. At the time of the Reform, social security benefits were phased out dollar for dollar for earnings above the *earnings test.*[9] The second age requirement is *eligibility age* – an age above which individuals are eligible to collect their retirement benefits regardless of their earnings. It is important to note that individuals that reach the retirement age and delay the receipt of their benefits receive a delayed retirement credit – a 5% increase in their social security benefits for each year of delayed retirement. The retirement age in the relevant time period for the Housewives Reform was 65 and 60 for men and women respectively, and the eligibility age was 70 and 65 for men and women respectively.[10] The *eligibility period* is the total periods of social security coverage that an individual must accrue in order to qualify for retirement benefits. Unlike social security systems in many other developed countries that require accrual of employment periods in order to receive

[8] An exception is Snyder & Evans (2006), who find that while the notch generation in the U.S. received lower social security payments they had lower mortality rates than similar individuals who were born one quarter earlier and received higher benefits. They propose that this outcome may have been driven by the fact that the notch cohort was 5 percent more likely to work between the ages of 68-70 than the slightly older cohort. Thus, suggesting that the impact of the decrease in wealth on mortality was more than offset by the increase in employment.

[9] Starting in 1999 the phase out rate was decreased from 100% to 60%.

[10] The retirement age is 65 for men who were born before March 1939 and 60 for women who were born before June 1944 - the relevant time for the Housewives Reform. For younger men this age gradually increased to 67 and for younger women it was gradually increased to 62. The eligibility age for men has not changed over the years. For women however it was gradually increased to age 67.

retirement benefits, to be covered by the Israeli social security system one only needs to be an Israeli resident.[11]

A *housewife* is defined in social security law as a married woman whose spouse is insured by the social security retirement benefits program and who does not have sufficient working history. Prior to 1996 a housewife was ineligible for social security retirement benefits. Instead, her spouse would receive a supplemental payment to his social security retirement benefits to account for his wife as a “dependent”, as long as she was over 45, not working, and not eligible for benefits. This means that prior to 1996, while a man who did not work and was married to a woman who was eligible for social security retirement benefits would receive a social security pension at eligibility age, a woman in an otherwise similar situation would not, causing gender discrimination.

The discrimination between married men and married women received substantial public criticism and eventually the Israeli parliament changed the social security law in 1996 with the aim of eliminating this discrimination against women. The change was applied, however, only to women born on January 1st 1931 or later (the Housewives Reform).[12] Thus, the Reform created a sharp difference in the benefits schedule of married couples with housewives born pre- vs. post- January 1st 1931. Under the old regime, the couple would have lost all benefits if the husband was employed (and earns a salary above the breakeven point in which the benefits are taxed), while under the new regime, a significant portion of the benefits is paid unconditional on husband employment, once his wife reaches 65. For husbands, this Reform is equivalent to a large decrease in the implied tax on delayed retirement.

In order to analyze the impact of the Reform, it is important to understand the timing and publicity of the announcement. To shed light on this issue, we searched news articles in two out of the three large newspapers in Israel (in terms of circulation), for the term “housewives” or “social security” from January 1st 1994 (two years before the Reform) through December 31st 1996. Table A1 summarizes the time-line of the legislation as recovered from these news articles. While the idea for the law was first raised in 1994, the first draft only appeared in the summer of 1995, and the final law was drafted and signed in the last quarter of 1995. Throughout this later period there were multiple mentions of the law in the press, including detailed articles explaining the change. This suggests that the public did have access to information about the nature of the Reform, but that this information was made available late in 1995, hence the behavioral responses are expected primarily in 1996.

[11] While baseline benefits do not depend on employment, benefits increase linearly in the number of employment years up to a cap (but are not a function of earnings).
[12] In 2013, the law was changed again to include women who were born before January 1931.

3. Data and Sampling

Our analysis draws on administrative data from the National Insurance Institute of Israel ("NII")—Israel's Social Security Administration. These data are collected by the NII from various sources (including the IRS and the Ministry of Interior affairs) for internal use. The data contain information on employment history and earnings from 1984 and onward, and social security benefits starting from 2003. They also contain demographic information such as country of origin, ethnicity, gender, date of birth, marital status and the birth date of each child. Importantly, these data can link spouses, allowing us to differentiate between husbands who are married to housewives and non-housewives and to determine whether the wife's birth-date results in the household being impacted by the legislative change. Furthermore, the data provide the date of death of each individual in our sample up to 2015, thus we can create an indicator for survival and use it to measure the impact of employment on longevity in the empirical analysis.[13]

In order to execute the main triple difference analysis of the Housewives Reform, we created a dataset of all women who were born between January 1st, 1930 and December 31st, 1932, who were married in 1996, excluding self-employed, kibbutz members and new immigrants.[14] We trace each woman's spouse and obtain information about both partners' age, employment history, earnings, and social security benefits. As we are interested in retirement behavior, we restrict the sample to households in which husbands are (still) employed in 1994 unless stated otherwise. This sample consists of 9,080 households.

According to administrative records, the entire population of households with wives born 1930-1932 consists of 30,641 households. Appendix Table A2, compares the characteristics of the entire population with the characteristics of households included in the main sample. Our sample is slightly more Jewish, and almost by construction, husbands tend to be somewhat younger. Naturally, most individuals that are not in our sample (i.e. retired by 1994) are not working in 1993, hence average income (including 0's) for the population is much lower. Appendix Figure A1 provides the full distribution of retirement age for men in both the entire population and our sample. It is apparent from the figure that bunching occurs at age 65, the NII retirement age in that period. Appendix Figure A1 also demonstrates that conditioning on working in 1994 implies a higher median retirement age for men in our sample of 70 compared to 65 in the population.[15]

Next, we create a housewife indicator that takes the value 1 if we classify a wife as housewife and 0 otherwise. The social security housewife status flag is available only if both spouses survive past the year

---

[13] See Appendix 1 for a description of the administrative databases combined for this analysis.
[14] We exclude from the analysis women that immigrated to Israel in 1989 or later as they are subject to a different set of rules concerning old age pensions.
[15] As expected, the sample restriction is less related to retirement age of working wives, and indeed median retirement age is 63 for non-housewives in our sample compared to 61 for the entire population.

2003. The 6,816 households (out of 9,080) that belong to this group have an administrative coding for the housewife indicator based on their NII status. In order to complete the assignment of the housewife indicator for the remaining households we proceed in two steps. The first step is based only on the wives' work history. According to the NII definition, any wife that has worked 60 months or more in the ten years leading up to retirement age or accumulated at least 144 months of work in her lifetime is not a housewife. Because our data begins in 1984, we can use the first criterion to classify 658 households where the wife worked at least 60 months as non-housewife households.[16] However, for the remaining 1,606 households, we are unable to use the NII criteria to differentiate between housewife and non-housewife households because we do not have their lifetime work history. Hence, in the second step, we apply a machine learning approach to identify housewives among the remaining households. We rely on husband and wife's employment and earnings history as well as a rich set of household characteristics and leverage the households with the NII housewife flag to train the data. This procedure produces very accurate results.[17]

Assigning a housewife indicator to the entire sample is important because it alleviates concerns regarding sample selection bias. Concretely, as we noted above, the administrative housewife flag is only available starting in 2003. Exclusion of households for whom this classification is unavailable would imply sample selection based on survival past that year, which may be affected by the Reform.

Table 1 provides some descriptive statistics of the Housewives Reform sample. The DDD identification strategy that we employ in Section 4.3 compares retirement decisions of husbands married to housewives born in 1930 - the comparison group - with those of husbands married to housewives born in 1931 and 1932 - the treatment group - while differencing out birth cohort effects using husbands married to non-housewives within each group. Therefore, columns (1) - (4) summarize the characteristics of housewife (or "HW") households and non-housewife (or "non-HW") households for the different cohorts. For time varying characteristics such as income, the table shows pre-reform (1993) values. Additionally, for each of the observable characteristics, we compare the differences-in-differences between HW and non-HW households over the cohorts, in column (5) of the table. The treatment and comparison groups look well balanced on observables with no significant differences between them in each of the characteristics after the removal of cohort effects using non-Housewife households. Notably, there are no significant differences in husband ages, alleviating concerns that the results that we report below are driven by such differences. Additionally, we examine whether there is a difference in how these characteristics are jointly

[16] We were able to verify the accuracy of this step using the group of households for whom we have both employment history information and the social security flag.

[17] Appendix 2 provides a full description of the machine learning process, and discusses classification errors.

associated with the housewife indicator across our cohorts using an F test and find no significant differences across the two groups (p-value 0.88).[18]

4. The Impact of (Implied) Income Tax on Retirement

4.1. Conceptual framework

We introduce a static labor supply model to illustrate how the Housewives Reform is related to a broader set of typical social security reforms and to tie it directly to the impact of tax on labor supply on the extensive margin at old age. We then use the model to derive some predictions about the impact of the Housewives Reform on labor force participation.

Consider a worker who is working from age 0 until age R and lives for another $T - R$ years after retirement. Suppose that there is no choice of labor supply on the intensive margin, and that the wage rate for each year of work is w. The worker is eligible for social security retirement benefits b per year starting at age R^0. The benefits, however, are subject to an earnings test, implying that for each additional year that the worker stays employed (and earns above the test cutoff) after R^0, the worker loses a portion τ of her annual retirement benefits. This setup captures the fact that delayed retirement schemes are not actuarially fair (or at least not perceived as such by some workers). For illustrative purposes, suppose that the earnings test threshold is zero. The worker's lifetime earnings as a function of retirement age are illustrated in Panel A of Figure 1.[19] If $\tau = 0$ (dashed blue line), the system is actuarially fair, i.e., for each forgone dollar due to delayed retirement, the worker receives an extra dollar after retirement.[20] In many social security systems (including the Israeli system), however, a forgone dollar in delayed retirement is compensated by less than an extra dollar post-retirement, introducing the kink in the budget constraint, namely, an implied tax on delayed retirement (τ^A , black solid line).

The effect of the Housewives Reform on the life-time budget constraint is very similar to a *decrease* in the implied tax on delayed retirement. Pre-Reform, delayed retirement is associated with forgoing both the husband *and* the dependent's benefits. However, post-Reform, delayed retirement is associated with forgoing only the husband's benefits while the dependent's benefits are paid regardless of husband's employment status. In Figure 1, this corresponds to a reduction in τ (τ^B, dotted red line). The main takeaway

[18] We run a regression of the housewife indicator on country of origin, nationality of both spouses, husband age, and husband income in 1993 and their interaction with a dummy for the 1931-1932 cohort. We then apply an F test for the joint significance of the interaction terms.
[19] We consider the benefits change to have a first order effect when forming predictions since our empirical specification is focused on agents who are affected by the policy after paying most or all of their life-time taxes. This allows us to abstract away from effects that could arise due to the impact of the changes in benefits on tax collection prior to retirement age.
[20] In this simple model this is also equivalent to a system without an earnings test.

from this budget set analysis is that *the Reform can be interpreted as reduction in the implied tax on delayed retirement* when the worker is older than R^0, *holding the generosity of benefits constant*. This feature of the Reform—creating a pure tax change—is quite unique among social security reforms analyzed in the literature, where it is usually the case that generosity of the benefits is changed (either directly or through change in eligibility age) at the same time that the tax on employment is changing.[21] Panel B of Figure 1 demonstrates the effect of another reform typically analyzed in the literature—a reduction in social security benefits (red dotted line). Such a reform indeed reduces the implied tax rate on delayed retirement, but at the same time decreases total benefits distributed. The figure demonstrates that analyzing such a reform captures both the wealth effect and effect of the tax change. Thus, the Housewives Reform provides an opportunity to study the implications of taxes on the employment decisions of older workers.

Panel A of Figure 2 also provides a framework for interpreting the elasticities we recover. The change in policy is equivalent to a change in the after tax earnings when employed. To the extent that intensive margin responses are small (i.e. that w is not affected by the Reform), the change in slope captures a change in the average net-of-tax rate paid to an employed worker.[22]

To form a prediction about the effect of this policy change on the timing of retirement, consider a worker who derives utility from total life-time goods consumption and from life-time leisure. The worker maximizes utility under the life-time budget constraint described above (Panel A of Figure 1). Note that this formulation requires assuming perfect capital markets. If workers are heterogeneous in their disutility from work (or in wages) then pre-Reform some workers are working less than R^0 years, some bunch at R^0, and some work more then R^0. The first order and unambiguous prediction of the model is an (average) increase in labor supply for workers who pre-Reform choose to work exactly R^0 years (this is illustrated using indifference curves in Panel A of Figure 2). Workers who pre-Reform retired before age R^0 are expected to be unaffected by the policy change. The effect of the policy on workers who pre-Reform have chosen to work more than R^0, assuming that the labor supply curve is upward sloping, is delaying retirement. Alternatively, a tax reduction may induce earlier retirement if the income effect dominates the substitution effect in the labor supply response to wage changes.[23] A final point, which is important to highlight using this framework, is that from the point of view of the household, the Reform had little effect on life-time earnings *other than* through changes in employment.

---

[21] One exception is a reform that eliminates the earnings test.

[22] In this simple model, total tax paid is $b\tau$, hence the average net-of-tax rate is given by $\left(1-\frac{\tau b}{w}\right)$.

[23] The Reform might also incentivize early retirement for individuals who retired at R^0 due to credit constraints and whose wives gained access to the housewife's benefit before they retired. Pre-Reform these husbands weren't able to borrow against their future retirement benefits, post-Reform they could stop working earlier and claim their wives benefits.

The magnitude of the income effect and the (lack of) effect of the reform on life-time earnings are important for the validity of the Reform as an instrument for employment, when studying the impact of employment on health outcomes. We address these points in detail in section 5, as part of the discussion about the monotonicity assumption and exclusion restriction of the instrument.

4.2. The impact of the Housewives Reform on benefits

In this section, we show the effect of the Reform on the household's social security retirement benefits using data from 2003-2007 when almost all workers in our sample were eligible for old age benefits regardless of their employment status.[24] The first four columns of Table 2 show how benefits are allocated between husband and wife for the 1930 and 1931-1932 cohorts for housewife and non-housewife households. Column (5) reports the differences-in-differences between HW and non-HW households over cohorts. The first row of the table shows the wives' retirement benefits. Housewives that belong to the 1930 cohort (column (1)) received essentially zero retirement benefits while housewives that belong to the 1931-1932 cohorts (column (3)) received on average over 10,000 NIS, illustrating how, following the Reform, housewives became recipients of retirement benefits.

In the second row, the table displays the effect of the Reform on Housewives' spouses. Spouses in HW households in the 1930 cohort (column (1)) received on average about 24,000 NIS while those in HW households in the 1931-1932 cohorts (column (3)) received about 16,000 NIS on average. This difference arises because the Reform canceled the supplemental "dependent" payment for housewives' husbands. These numbers illustrate that the Reform caused a sharp change in the incentives to retire, substantially reducing the penalty on employment for the HW households in the 1931-1932 cohorts. Notably, while wives in HW households in the 1931-1932 cohorts receive a much larger share of household benefits than wives in HW households in the 1930 cohort, the total benefits collected at the household level are, on average, only slightly higher for the 1931-1932 cohorts. In other words, the main impact of the Reform was not changing the overall benefit level, but rather shifting payments from husbands to wives. Overall, Table 2 establishes that the Reform corresponds to an almost pure change in the implied tax on delayed retirement. Appendix Figure A2 provides a graphical representation of the Reform. While the message of the figure is similar to that of Table 2, the figure illustrates the sharp cutoff in benefits collection for households with a housewife born pre- vs. post-January 1st, 1931.

Note that while the average difference in benefits of wives in HW and non-HW households in the 1931-1932 cohorts is much smaller than the same difference in the 1930 cohort (due to the Reform), benefits

[24] Social security benefits data is only available starting in 2003. Therefore, in creating this table, we must condition on survival. In this context, however, the potential selection issues are negligible as there is hardly any variation in social security benefits in our institutional setting as discussed above.

of the non-HW group are somewhat larger. The source of this gap is the difference in employment histories between wives in the two types of households. Social security retirement benefits do not depend on earnings histories; they do depend, however, on employment histories. By construction, housewives do not receive credit for employment history, explaining the differences between wives' benefits in columns (3) and (4) of Table 2.

Appendix Table A3 provides information on private pensions for individuals in our sample between 2003 and 2007. This table is important because it illustrates that social security benefits represent a sizeable fraction of pension income to retired households in this period. Specifically, social security benefits make up over 30 percent of total pension benefits for non-housewife households, and over 40 percent for housewife households. Notably, columns (1) and (3) show no significant difference in the private pensions of either husbands or their wives when comparing the 1931-1932 and 1930 cohorts within housewife households.

4.3. The response to the Housewives Reform – a DDD approach

The ideal experiment to study the effects of the Housewives Reform would involve a random assignment of housewife-households to a "treatment group" that is subject to the Reform's new rule and a "control group" that remains under the old, pre-Reform, rule. The environment we study lends itself to a standard triple difference (DDD) design that closely approximates such a thought experiment. Concretely, treated housewife households are those with a wife born on January 1st, 1931 or later and the comparison group includes housewife households with wives that were born before this date. The identifying assumption is then that absent the Reform, time trends would have been similar for those two groups. However, one might be skeptical about the common trend assumption when comparing different birth cohorts to conduct the analysis. To address that, we invoke the DDD approach, much in the spirit of Gruber (1994), where we use non-HW households, to correct for potentially different trends across cohorts. We also show, in Section 4.4, that an RDD approach, which addresses this issue by comparing housewife households around the cutoff date of January 1st, delivers very similar results.

Our first step is to run a differences-in-differences analysis to examine how the Reform affected retirement rates of the "treatment" households—those with housewives born in 1931-32—relative to non-HW households of the same birth cohort. We then run a similar DID analysis with the "comparison" households—those with wives that were born in 1930. In the second step, we combine these two analyses to one DDD framework. Thus, the DDD framework provides estimates of the effect of the Reform on retirement rates in the treatment group relative to the comparison group while differencing out any birth cohort effects using non-HW households.

Figure 3 illustrates graphically the results from our DDD approach outlined above.[25] The figure shows, side by side, the retirement trends of husbands married to wives born in the 1931-1932 cohorts (panel A) and husbands married to 1930-born wives (panel B). In each panel, cumulative retirement rates of husbands (conditional on working in 1994) in the period 1994-2005 are graphed for both husbands married to housewives, and to non-housewives. Starting at the first year of the Reform, there is an apparent divergence between husbands married to housewives and husbands married to non-housewives in the 1931-1932 cohort, with no similar divergence in the 1930 cohort. This divergence is due to a lower retirement rate of husbands married to housewives from the 1931-1932 cohort (relative to those married to non-housewives). This slower retirement is consistent with an increase in labor supply caused by a reduction in the implied tax on employment for this cohort.

Columns (1) and (3) of Table 3 report the regressions results for the effect of the reform on retirement by year, which map to the lines in Figure 3. Each row in column (1) reports the difference between housewife- and non-housewife households in the 1930 cohort (this corresponds to the difference between the two lines in the right panel of Figure 3). Column (3) reports the same numbers for the 1931-1932 cohorts. Column (5) reports the DDD estimates, i.e. the difference between columns (3) and (1). Columns (2), (4) and (6) report the same estimates as is in columns (1), (3) and (5), respectively, relative to retirement in 1995. As the table shows, the effect of the Reform on retirement is statistically significant and economically large in the first year of the Reform.[26] Using the estimates from column (6), conditional on working in 1994, husbands married to 1931-1932 cohort housewives are 5.4 percentage points (s.e. 1.7) less likely to retire in 1996 (i.e. the first year of the Reform) compared to husbands married to wives that were born in 1930. As the years pass, the effect vanishes. This is not surprising, given that over time more husbands reach the eligibility age of 70, where they are not subject to an earnings test anymore (and therefore see no tax on delayed retirement).

These estimates can be used to recover an extensive margin elasticity of labor supply with respect to the average net-of-tax rate.[27] We define this elasticity as the percentage change in employment divided by the percentage change in the net of tax rate

---

[25] In our baseline specification, we define the last year before retirement to be the last year for which we see the individual working for at least 6 months and earning on average at least the monthly minimum wage. We show that our results are not sensitive to the retirement definition in our discussion of *Robustness and Alternative Designs*.

[26] While the household's eligibility occurs only when the wife reaches age 65, which occurs at different points for different households according to the exact date of birth of the wife, it is expected that the effect occurs upon the legislation change due to the irreversible nature of retirement.

[27] An alternative approach would have been to estimate the elasticity of employment w.r.t to the implied tax by instrumenting for the tax change using the reform (see for example Gruber and Saez (2002)). However, we do not observe the benefits collected at the household level around the reform (only for later years). Thus, instead of imputing these earlier benefits, we estimate the effect of the reform on employment directly, and recover average elasticities using average social security benefits.

$$\frac{dProb(E)/Prob(E)}{d(1-TR)/(1-TR)}.$$

The numerator of the elasticity is the percentage change in employment due to the reform. In Column (6) of Table 3, we find that the Reform increased employment on impact by 5.4 percentage points. To recover the numerator, we divide that by the counterfactual probability of employment absent the reform, which is 0.73. In recovering the denominator, we must make some assumptions about the way workers perceive the delayed retirement credit. We assume that workers are myopic or alternatively that they do not fully understand the delayed retirement credit system, whereby they do not realize that forgone benefits are replaced by delayed retirement credit.

More formally, as discussed in Section 4.1, the net of tax rate $1 - TR = 1 - \frac{\tau b}{w}$. This implies that $\frac{d(1-TR)}{(1-TR)} = \frac{\tau_0 b - \tau_1 b}{w - \tau_0 b}$. The difference in the numerator $(\tau_0 b - \tau_1 b)$ is the difference in the implied tax on employment with and without the Reform, and can be read as the difference between retirement benefits of husbands to housewives in Columns (1) and (3) of Table 2. To compute after-tax earnings (w), we use earnings conditional on employment from Column (3) of Table 1 and subtract 16% - the average income tax rate in 1996.[28] The assumption that workers are myopic places an upper bound on the percent change in average net-of-tax rate, and thereby a lower bound on the elasticity. These assumptions imply a 17 percent decline in the net-of-tax rate, which implies an elasticity of 0.43, with a standard error of 0.14. Our estimated elasticity is on the high side of estimates of extensive margin elasticities as reported in quasi-experimental studies (See e.g. Table 1 in Chetty et al (2012)), however, it is comparable with the recent estimates reported by Gelber et al. (2017).[29]

So far we have assumed that workers are myopic or that they have poor knowledge of the delayed retirement credit. If workers fully understand the delayed retirement credit and they are not fully myopic, the elasticity calculation is sensitive to the discount factor, which incorporates their perception of the mortality rate, and to the extent of present bias. For example, assuming hyperbolic discounting, with a 5% discount factor, and a present bias parameter of 0.5 (the midpoint between myopia and no present bias), will increase our elasticity estimate to 0.72. Alternatively, assuming a 10% discount factor would imply a smaller adjustment to an elasticity of 0.6.

---

[28] Note that employee's income tax payments in Israel are typically calculated individually and withheld by the employer.

[29] In a recent paper, Manoli and Weber (2016b) estimate participation semi-elasticities w.r.t. financial incentives of between 0.1 and 0.3 applying bunching methods to Austrian data in the context of employer provided severance payments. They find that elasticities are most significant for financial incentives that have a time horizon of 6 to 9 months.

Robustness tests and alternative designs.

We check the robustness of our results to sample and outcome definitions. Columns (1) and (2) of Table 4 report the findings for the baseline DDD specification when restricting the control sample of non-HW households to include only households where the wife has less attachment to the labor force. To do so, we require that the wife was employed for less than 60 months in the ten years leading up to age 65, making the non-HW households more similar to the HW households. Reassuringly, the results are almost identical to the ones reported for the full sample. In columns (3) - (6) we evaluate the sensitivity of the results to different definitions of the retirement year. In columns (3) and (4) we report the results with an employment definition that requires only 3 months of employment in a given year, while in columns (5) and (6) we require a monthly income above the earnings test threshold in order to consider someone employed. The results are again very similar to our baseline results.

Since we define retirement based on the last year of employment, our empirical analysis is implicitly based on a "traditional" notion of uninterrupted employment that ends upon retirement. While this may be quite plausible for the population we study, we also examine whether our results are sensitive to this issue. To do so, we analyze actual employment, rather than retirement, in a given year by defining the independent variable as an employment dummy that takes the value of 1 if the husband is employed in a given year and zero otherwise, maintaining the employment definition as in columns (5) and (6). We then use this variable to analyze whether the Reform increased employment. Note that this reduces the number of observations because we include only those who work in 1994 according to this new definition. The results of this robustness exercise, reported in columns (7) and (8) of Table 4, are qualitatively similar to the retirement estimation results, with the opposite sign, though statistical significance is lost in the last specification reported in column (8).

Cumulative effect on employment.

So far, we have focused on the year-by-year response of retirement to the tax decline induced by the Reform. Table 5 reports the effect of the Reform on cumulative husband's employment.[30] This is useful for two purposes. First, the cumulative effect summarizes the total effect of the Reform on employment. Second, in Section 5, we explore the effect of delayed retirement on long run health outcomes and survival. Naturally, survival is affected by the entire history of employment, hence neglecting the cumulative effect (for example by associating the entire health effect with the 1996 employment effect) would result in an over-estimate of the effect of employment on health. Our approach compares the difference in cumulative employment of husbands married to housewives and non-housewives in the 1931-1932 cohorts to the

[30] We calculate the husband's cumulative employment during the 5 years beginning in 1996. We experimented with extending this period to as long as 10 years and found no qualitative difference in the results.

equivalent difference for husbands married to housewives in the 1930 cohort.[31] Column (1) of Table 5 reports the effect, showing that, overall, the Reform increased cumulative employment (or delayed retirement) by 0.327 years of work. With household level controls, the estimate decreases to 0.285 years of work, as shown in column (2) of the table. We will use these results when we study the effect of employment on health. In other words, these estimates are the first stage results for the analysis of the effect of employment on health that we perform in Section 5.

Appendix Table A4 explores heterogeneity in the cumulative employment effect. In the first two columns, we repeat the specification from Table 5 for ease of comparison. In columns (3) - (4), we restrict the sample to households with non-native spouses that account for about 83% of the sample. In columns (5) - (6) we restrict the sample to households with Jewish spouses, comprising 96% of the sample. The estimates for these two groups are very close to those in Table 5, demonstrating that our results are not driven by outlier groups in the population. Next, we explore heterogeneity over husbands' ages. The husbands that were most impacted by the Reform were those in the age range 65-70 in 1996. It is precisely in this period that workers face the implied tax on employment. For the 1931 cohort, this means an age difference of 0-5 years between husbands and wives and for the 1932 cohort, it implies an age difference of 1-6 years. We therefore examine the response of a subsample of husbands with an age difference of 1-5 years between them and their wives, which allows us to keep the husbands' age distribution comparable across cohorts.[32] The results, reported in Columns (7) and (8) of the table, show that, as expected, the employment response of this group is much more pronounced than that of the entire sample.

4.4. The Response to the Housewives Reform – an RDD Approach

In this section, we complement the DDD analysis with evidence from a regression discontinuity design.[33] This approach exploits the sharp age-based rule within a regression discontinuity design framework. To illustrate how this would work, consider two housewives: one that was born on January 1st 1931 and another that was born on December 31st 1930. Assuming that the wives' exact date of birth is uncorrelated with their other characteristics and particularly their husbands' retirement decision, comparing the retirement patterns of their husbands resembles the ideal experiment that examines the effect of the Reform on husbands' retirement, that we described in the previous section.

[31] This approach is consistent with the results in Tables 3 and 4 showing that there are no pre-trends in those differences.
[32] The age difference of 1-5 years keeps approximately 42% of our sample. The results are quite similar when restricting the sample to an age difference of 0-5 years, which includes 52% of the sample.
[33] When conducting the RDD analysis we add the 1929 cohort in order to allow a symmetric 24 months of birth window around the January 1st 1931 cutoff.

More formally, let τ indicate the wife's date of birth in terms of days elapsed since January 1st 1931. For example, if a wife was born on December 20th 1930, $\tau = -12$; if she was born on January 10th 1931, $\tau = 9$. Let the treatment indicator, D, equal 1 if the wife was born in January 1st 1931 or later, and 0 otherwise. Consider the following model relating the husband's timing of retirement (y) with the wife's date of birth in terms of τ and the treatment indicator:

$$y = \alpha + \beta D + f(\tau) + \epsilon. \tag{1}$$

$f(\tau)$, is a completely flexible control function, and it is continuous at $\tau = 0$. The parameter of interest in this model is β that measures the causal effect of the Reform on y. Intuitively, given that $f(\tau)$ absorbs any continuous relationship between a wife's date of birth and her husband's retirement decision, the coefficient β estimates the discontinuous relations between the Reform and the husband's retirement decision. We estimate such a model using standard regression discontinuity design methods (see Lee and Lemieux (2010) for a survey).

Motivated by the DDD results, the RDD analysis aims to examine how likely a husband that was employed in 1994 is to retire by 1996—the first year of the Reform—as a function of his wife's date of birth. Figure 4 displays the results of the RDD analysis for HW households. The figure shows the retirement probabilities by quarter of birth of the wife, illustrating that there is a sharp drop in retirement probabilities for affected husbands. The corresponding estimates are reported in columns (1) - (4) of Table 6. Column (1) shows a statistically significant drop of 8.8 percentage points in retirement rates of husbands of housewives using a specification with a linear polynomial and no household level controls. The result is almost unaffected by the inclusion of household level controls, as column (2) demonstrates. The results are also very similar when we repeat the analysis using a quadratic polynomial, as columns (3) and (4) of the table show.[34] Appendix Figure A4 repeats the specification from column 1 of Table 6 for all bandwidths in the range of 6 to 24 months for both uniform and Epanechnikov kernels, demonstrating that the results are very stable across bandwidths and kernels.

Next, we look for any indication that observable characteristics (that are determined pre-Reform) change sharply around the January 1931 threshold. If this were the case, it could raise the concern that selection could be affecting our results. In panels A - C of Figure 5 we examine the behavior of the age-gap between husbands and wives, the log of husband's earnings in 1993 (i.e. pre-Reform), and the share of immigrants among husbands, respectively. All three variables appear to trend quite smoothly around the January 1931 threshold. In panel D we report the husband's predicted probability to retire by 1996 using a model that includes 3rd order polynomials in the first two variables (age-gap and husband's log monthly

[34] In Appendix Figure A3 we report the RDD results for all years in the period 1996-2001. Compared to the DDD, the RDD approach shows somewhat higher effect on impact (1996) and a steeper gradient for the decline of the effect. Having said that, the standard errors are large and overall, the dynamic patters are similar.

earnings), and the husband's immigration status. As Panel D of Figure 6 illustrates, the predicted values generated by this model also appear to trend smoothly around the threshold. Table 7 provides the corresponding estimates. As the table indicates, there are no statistically significant discontinuities in the three observables we analyze, as well as in the predicted values of the probability to retire.[35] Overall, this analysis shows no indication that our results are an artifact of sample selection.

Our setting provides an opportunity to examine the validity of these results using two placebo tests. The first test relies on the fact that information about the Reform only became available towards the end of 1995. Therefore, while there may have been some response towards the end of 1995, the Reform should not have affected the retirement decisions of households in 1994. However, if the drop in retirement rates in 1996 arose because spouses married to housewives born January 1st 1931 or later tend to retire later regardless of the Reform, we would expect this to also manifest in the 1994 retirement decision. Thus, we conduct a test in which we replicate the RDD analysis using retirement in 1994 as the outcome variable. Note that in order to do this, we must change the sample so that it includes the husbands who worked in 1993. For this altered sample we first repeat the main analysis with respect to retirement in 1996 and we find that our main result hold.[36] We then run the analysis for retirement in 1994. Panel A of Figure 6 displays the results of this exercise. As one might expect, retirement rates are lower in 1994 (conditional on work in 1993), yet they are still substantial. Retirement rates trend smoothly around the treatment threshold, with no indication that husbands married to housewives born January 1st, 1931 or later have a tendency to retire less before the Reform.[37]

The second placebo test takes advantage of the non-HW group. Panel B of Figure 6 illustrates the results of a placebo exercise using the non-HW households. To focus on households that resemble HW households, we restrict the sample to include households where the wife has less attachment to the labor force, using the same criterion we applied in the previous section (columns (1)-(2) of Table 4), namely, non-HW households in which the wife was employed for less than 60 months in the ten years leading up to age 65. As the figure illustrates, this group does not exhibit a similar pattern of retirement around the January 1931 threshold as they were unaffected by the Reform. These results corroborate the interpretation

[35] Appendix Figure A5 complements this analysis, by showing that there is no discontinuity in the density around the cutoff. Our birth date data is discrete at the monthly level, hence we cannot conduct a formal McCrary test (McCrary, 2008). Panel A of the figure shows the distribution of birth months without any controls. As missing birth months are recorded as April, there are noticeable spikes in number of records in each April, as well as slightly more records in January. Panel B, shows that controlling for 12 monthly dummies, the density is very smooth across the cutoff.

[36] See appendix Figure A6.

[37] We repeat the placebo exercise using retirement in 1995 as the outcome variable. This placebo exercise is not as clean because we can expect some effect on retirement towards the end of 1995 (see Appendix Table A1 for the timing of the reform). Nevertheless, the results as reported in Figure A7, show a negative, small and statistically insignificant result.

of the results as stemming from the Housewife Reform. They clearly show that the delay in retirement that we documented in the DDD analysis arises sharply around the January 1931 threshold and only in the case of the HW group.

5. The Effect of Employment on Health

So far, we have established a causal link between the decline in taxation and delaying retirement. We turn now to the second question that we have posed in this paper – what is the effect of extended employment on health? To address this question, given our setting, we analyze a model of the form

$$\text{(2)} \qquad Survival = \alpha + \beta_1 \cdot Employment + \beta_2 \cdot HW + \beta_3 \cdot \text{born\_1931} + X \cdot \gamma + \epsilon$$

$Survival$ is our main health outcome. It is defined as a dummy variable that equals 1 if a husband survives past a given age threshold. We define a set of 21 such outcome variables spanning the age range of 65-85. In the analysis that follows, $Employment$ is defined as in Table 5 – the number of years a husband worked in the five year period starting in 1996. HW is a dummy variable that equals 1 if the husband belongs to a HW household, born_1931 is an indicator for households whereby the wife was born in January 1st 1931 or later, and X is a vector of household characteristics. Naïvely analyzing this model using OLS for example, is likely to provide biased estimates of β_1, the effect of employment on survival, because of underlying unobserved factors that affect both the employment decision and the survival of the individual. For example, Individuals with a health condition, unobserved by the econometrician, may tend to work less and have a lower likelihood to survive longer, generating a positive association between employment and longevity. Here, we aim to study this relationship using the Housewives Reform as an exogenous source of variation in employment. Namely, we estimate the model in Equation (2) using the Housewives Reform and, specifically, the interaction term between HW and born_1931 as an instrument for $employment$. Below, we will use this setting to analyze the causal effect of employment of elderly workers on their entire survival path in the age range of 65-85.[38]

Before proceeding to the results of this analysis, a discussion about the validity and interpretation of this instrument is warranted. Is it reasonable to assume that the exclusion restriction holds? Namely, that the Reform affected health only through its effect on employment. One obvious alternative channel is that the Reform affected household *resources* not through employment. However, Table 2 and Appendix Table A3 show that overall household income from private and public pensions post-retirement remained very

[38] We only observe time of death up to 2015. This limits the ability to analyze survival of younger husbands in later ages. We therefore analyze survival up to age 85.

similar for households with retired husbands affected by the Reform compared to households with unaffected husbands. Specifically, the Reform increased total annual income from public and private pensions by 1,057 NIS (Summing over column 5 in Tables 2 and A3), over an annual flow of 57,404 NIS of pension benefits (Summing over column 1 in Tables 2 and A3), implying an overall change of 1.8% in the annual income from public and private pensions. Even ignoring other sources of annual income (which would make this number even smaller), this change in pension income is an order of magnitude smaller than the 17% change in the net-of-tax rate implied by the reform.

It is important to note that while there was little change in the housewives households' benefits for the 1931-1932 cohorts post-retirement, the Reform could have increased income through two other channels: First, delaying retirement may increase income as workers accumulate more years of labor earnings. Second, the Reform affected net income of the employed that *did not* change their employment behavior by increasing their average net-of-tax rate. This implies that on average those treated by the Reform earned higher incomes until they reached age 70. We argue that these income increases are not a source of concern to our identification strategy for two reasons. First, they are likely to have only a small impact on lifetime earnings. Second, even if the Reform did increase household resources, we would expect the increased resources to increase survival of the treated husbands (for example by providing additional funds for heating or medical expenses). Yet, the results we report below indicate that the Reform caused a decrease in survival of the treated husbands. Therefore, to the extent that increased resources affect the results, they attenuate them towards zero.

Another concern regarding the instrument estimation is that, in theory (as illustrated in Figure 2 panel (B)), for some individuals, the income effect caused by the Reform may be stronger than the substitution effect and the sign of the labor supply response to the reform could be reversed, violating the monotonicity assumption (Imbens and Angrist, 1994). While we cannot completely rule that out, we highlight that income effects are likely to be small in our setting, due to the small size of the change in discounted income as implied by the reform. The present value of the change in retirement benefits due to the reform for a representative husband in our sample (by age 90, discount factor 5%) amounts to about 36,800 NIS. This number incorporates both the extra income due to eligibility of housewives for benefits while the husband still works, and the small increase in total household benefits post-husband's retirement. Dividing this amount by 691,000 NIS, the total retirement and pension benefits available for workers absent the reform (based on 1930 housewives households), we find that the overall increase in income due to the reform amounted to 5%. Importantly, this calculation does not account for income sources that are unaffected by the reform, hence 5% is an upper bound on the Reform's income effect. To estimate the expected labor supply response to this income effect, one needs to take a stand on the magnitude of the income elasticity. If we use an income elasticity of -0.2 and multiply that by the 5% change in discounted income, this would suggest that the average person in our sample would decrease employment by 0.02 years after the reform.[39]

Finally, in a non-unitary framework, the shift of social security payments from the husband to the housewife for the post-retirement years, could shift bargaining power within the household towards the wife. The literature discusses two classes of bargaining models. The first approach models threat points as resorting to noncooperative behavior in the context of public consumption goods (e.g. Lundberg and Pollak, 1993). The second approach stresses divorce as a relevant threat point (see for example Voena, 2015, as well as detailed discussion in Chiappori and Mazzocco, 2017). Our setting involves elderly households, with almost zero divorce rates, hence at least the channel highlighted in this second approach is muted. Additionally, by construction, housewives do not work, hence this shift would not affect the household budget constraint through a change in the wife's labor supply. However, it could effectively operate in a manner that is equivalent to a *negative* income effect, and as such offset the positive income effect discussed thus far. While we cannot directly estimate the magnitude of such an effect in our data, we can revise the income effect calculations to take into account different scenarios for the magnitude of the bargaining power effect. We do that by assuming that only a fraction of the 10,070 NIS of benefits paid to the wife post-reform (column (5) of Table 2) is now considered by the husband when making his retirement decision. Appropriately adjusting for discounting, and repeating the calculation above implies that when setting the fraction to 75%, the 0.02 years income effect calculated above drops to zero.

To summarize, given the large estimated increase in employment of 0.327 years that we attribute to the reform it seems unlikely that either the income effect or the bargaining-adjusted effect would reverse the direction of the labor supply response in our context.[40] Having said that, if monotonicity is violated, the results in the next section maintain their reduced form interpretation for the impact of the reform on longevity.

Finally, a condition for the validity of the Reform as an instrument is the existence of a first stage. We reported the first stage results, indicating a statistically significant increase of close to four months in employment, as part of our extensive discussion in Section 4 about the effect of the Reform on employment. Notably, the F-stat for the effect of the Reform on employment is 6.2 (column 1, Table 5), which is below

[39] This -0.2 income elasticity is quite large relative to existing results in the literature. In a recent study, Cesarini et al. (2017) report an income elasticity of -0.17 and much lower elasticities in the range of -0.04 for older individuals when analyzing responses to lotteries in Sweden. McClelland and Mok (2012) survey estimates of income elasticities and conclude that they generally fall in the range of -0.11 and 0.

[40] We further verify that such an effect cannot be driving the magnitude of the elasticity reported in Section 4.3. Repeating this calculation when the husband considers an even smaller fraction of 50% of his wife's benefits in making his retirement decision, implies that the adjusted income effect drops to -0.019, which is still very small.

the standard threshold often discussed in the literature.[41] Given that we have one endogenous variable with one instrument, our IV estimates are median-unbiased, however inference could be problematic. To that end, we calculate confidence intervals using bootstrap, which, despite having known shortcomings in this context, are thought to be relatively reliable (Davidson & Mackinnon, 2014). Additionally, while we report the bootstrapped confidence intervals below, we also calculated the IV standard errors for survival by age regressions using the approach suggested by Chernozhukov & Hansen (2008). The two methods provide very similar confidence intervals.

5.1. The Causal Link between Employment and Health

Here we assess the effect of extended employment on health, taking advantage of the Reform as an exogenous source of variation in employment. As we reported in Table 5, the first stage results show that the Housewives Reform induced, on average, close to four months of additional employment in the five years after 1996. In Figure 7, we report the effect of the Reform on the survival of affected husbands – the reduced form analysis results. To do so, we estimate the model in Equation (2), replacing the $Employment$ variable by the interaction term between HW and a dummy for wife born in 1931 or 1932 – our instrument. The figure displays the coefficient of the interaction term and its 90% bootstrapped confidence interval, for each of the 21 $Survival$ indicators (survival past 65 - survival past 85) as the outcome variables. As the figure shows, in the age range 65-74 the Housewives Reform was not associated with a change in the survival of affected husbands. In subsequent years, however, there is a statistically significant decrease in the likelihood of the affected husbands to survive.[42]

Next, we assess the cumulative effect of employment on survival. We summarize these results in Table 8. According to Figure 7, in the age range 65-74 there appears to be no effect on survival, but, in age range 75-85, employment appears to have taken a toll on survival of those who extended their employment following the reform. Therefore, we report the cumulative results in three parts, corresponding to the three outcome variables in Table 8. The first outcome variable reports the results for cumulative survival in the entire age range of 65-85, which measures the overall effect of the Reform on longevity. In the second and third outcome variables, we split the survival curve to two pieces: the early years and the later years, the age ranges of 65-74 and 75-85, respectively. Every statistic in the table is the sum of the age-by-age survival

[41] E.g. in chapter 4 of Angrist & Pischke (2008) an F-stat of 10, based on Stock, Wright, & Yogo (2002) is regarded as the safe zone.

[42] We are only able to define survival by specific ages (the left hand side variables) for husbands who reached that age by 2015 (the last available year of the mortality data). Recall that the treatment and comparison groups are balanced on age (see Table 1), hence for each point, the estimates represent the treatment effect of the Reform for the relevant husband age cohorts. To ensure that changes in the composition of birth cohorts over survival age are not driving the results, we repeat this exercise in Appendix Figures A8 and A9 using a sample of husbands which we observe until (and including) age 80, and find that the pattern is almost identical.

regression coefficients for the relevant age range. For each statistic, we report its 90% bootstrapped confidence interval.[43]

Table 8 reports the cumulative effects on longevity of husbands. Because we focus on survival between ages 65 and 85, the max longevity we observe is 21 years. Absent the reform, husbands of housewives live about 19 of the 21 years. The first row of Panel A shows results for the entire age range of 65-85. Column (1) reports the sum of reduced form coefficients, namely the sum of coefficients from Figure 7. The reform decreased longevity by 0.35 years, about four months. This result is insignificant at the 10% level (a zero effect is contained in the 90% confidence interval). Adding household level controls, the result remains very similar as shown in column (2). In columns (3) and (4) we report the cumulative results from the OLS estimation of equation (2). Both estimates are statistically significant and positive in the order of four months. Even though they merely capture the correlation between employment and longevity, we note that the positive sign of these estimates is consistent with the typical simultaneity problem of the retirement-health link (see for example Insler (2014) for a literature review). Next, the IV estimation result in column (5) of Table 8, shows a negative effect on longevity – an additional year of employment reduces longevity by 0.9 years. The result, however, is again insignificant at the 10% level. When we add household level controls in column (6) of Table 9, the results stay roughly unchanged.

It is important to put these magnitudes in context. First, the 0.9 years effect of employment at old age, amounts to about a 4.7% (0.9/19) change in number of years lived between 65 and 85. Second, while the decision to work, as well as the returns to work in terms of income, occur on impact (in 1996), the health consequences seem to occur later in life. To account for this issue, we introduce a discounted estimate of the effect of employment at old age on cumulative years survived between 65 and 85, using a 5% discount factor for every year after 65 (see bottom of Panel A of Table 8). Had the survival response occurred on impact, applying the discount factor would have little impact on the results. However, as the table shows, the discounted effect of 0.4 (see Column 5) is much smaller, consistent with the longevity effect happening only late in life.

Based on the results in Figure 7, we would expect to find a significant effect on longevity at older ages (75-85), as opposed to younger ages (65-74), hence we turn now to report cumulative effects separately by age range. Both reduced form (columns (1) and (2)) and IV specifications (columns (5) and (6)) in Table 8 find little effect of employment on longevity at early ages (65-74). When turning to the reduced form estimate for the later age range of 75-85, we find a decrease in survival of -0.35, which is very similar to the overall effect, but is now significant at the 10% level. This implies that the Reform caused a decrease of about four months in the longevity of husbands at older ages. Adding household level controls slightly

[43] We used 500 bootstrap replications.

increases the magnitude of the effect. The OLS estimates in columns (3)-(4) increase in size and remain statistically significant. The IV estimate shown in column (5) is -0.93. While the upper bound of the confidence interval is still slightly above zero, it becomes slightly negative when adding controls (column (6)). These results suggest that an additional year of employment significantly decreases longevity past age 75 by one year.[44]

Panel B of Table 8 repeats the analysis for a sample of men, who survived to age 74. Given that we find no trace of an effect at earlier ages, this approach allows us to focus on those affected by the Reform, and thus increase statistical precision of our estimates. Indeed, focusing on this sample, the reduced form results are very similar (columns (1) and (2)), with tighter confidence intervals and a slightly smaller overall effect of employment on longevity in columns (5) and (6) due to a slightly larger first stage estimate of 0.427 (s.e. 0.136) and 0.387 (s.e. 0.132) in these two columns, respectively.

In summary, we find evidence that additional employment of elderly workers reduces life expectancy. The effect arises entirely in later years of life. The point estimates are quite large, indicating that delaying retirement by another year decreases life expectancy by 9 to 12 months, years after the retirement decision. However, while the estimates for older ages are significant and negative, their confidence intervals are quite large, and the estimates do not provide a clear bottom line with respect to the magnitude of the effect.

It is important to highlight that the effects on life expectancy that occur at later ages, such as the ones we find, cannot be attributed to workplace fatalities. Our outcomes are better explained by a reality where the effect of occupation is observed long after the worker is no longer in that specific environment (see for example Moore and Hayward (1990)). Indeed, Shai (2018) documents that increasing the retirement age in Israel in 2004, resulted in workers working later into their 60's, and exhibiting an increased onset of illnesses such as hypertension, diabetes, heart attacks, asthma, cancer, lung disease, and ulcers. While our data does not provide information on cause of death, a possible mechanism for explaining the estimated link between delayed retirement and mortality years later, is that delayed retirement increased the likelihood of illnesses with long-term implications for longevity.

5.2. Heterogeneity

It is well documented that blue collar workers face higher mortality risks than white collar workers (see Moore and Hayward (1990), Johnson, Sorlie, and Backlund (1999), Burnett, Maurer, and Dosemeci

[44] Note that we see very similar patterns applying the RDD approach from section 4.4. Appendix Table A5 reports the results, and Appendix Figure A10 demonstrates the year by year RDD graphs for selected years.

(1997)).[45] Our data provide an opportunity to analyze the long term implications of extended employment on longevity for these different types of workers.

Ideally, we would like to compare the effect of work at older ages on health across different occupations. While we do not have direct information about the individuals' occupation, we use the industry in which the individual was employed in 1993 as a proxy for the individual's likelihood of working in blue- versus white-collar occupations. Specifically, we characterize each industry as blue- or white-collar using the Israeli labor force surveys from the periods 1995-2000. These surveys contain information about the composition of employee occupations in each industry. We define an industry as blue-collar if at least 50% of the industry's employees have blue collar occupations, dropping from the sample 812 husbands for whom we did not have the industry composition information.[46] With blue- and white-collar defined, we run the same analysis as above separately for each of the two groups.

In Figure 8, we report the reduced form analysis results for each of the 21 $Survival$ indicators. As the figure shows, the white collar workers' estimates are all close to zero and statistically insignificant. This group appears to exhibit no change in survival following the reform. The Blue collar workers' estimates show no effect on survival in the age range 65-74 but then begins to decrease in later years.

Panel A of Table 9 combines the age-by-age results to assess the cumulative effect of employment on survival, as in Table 8. The reduced form estimates for the blue-collar group are reported in column (1) of the table (comparable to column (1) of Table 8). The estimate for the entire age range is -0.5 and it is statistically significant with a 90% confidence interval in the range [-0.985, -0.01]. The reduced form estimate for the blue-collar group in the age-range 65-74 is very small and statistically insignificant and the estimate for the age range of 75-85 is a significant -0.51 with a 90% confidence interval of [-0.916, -0.087]. Namely, there is an overall six months decrease in the survival of blue-collar workers that were affected by the reform that arises entirely from the age range 75-85. The reduced form estimates for the white-collar group are reported in column (3) of the table and they are all small and statistically insignificant. The first stage results for the two group are reported at the bottom of the panel. The first stage effect size in the two groups is quite similar to the overall effect, but it is a little more pronounced for the white-collar group. The IV results for the blue- and white-collar group are reported in columns (2) and (4) of the table, respectively. The IV results are overall qualitatively similar to the reduced form results, however, the confidence intervals are very large, and the results are insignificant. Similar to the approach introduced in Table 8, here as well, Panel B reports the results for those who survived to age 74, allowing us to increase statistical precision.

---

[45] Shai (2018) uses education as a proxy for worker type and shows that relative to more educated workers, workers with less than 12 years of education report a general deterioration in their health from working more.

[46] White-collar occupations are defined as the following major categories (equivalent to major categories 1 through 4 in the ISCO-88 occupation classification system): Legislators, senior officials and managers, Professionals, Technicians and associate professionals and Clerks.

While for white-collar workers, estimates remain small and insignificant, for blue-collar workers both reduced form and IV estimates become more precise. The IV results indicate that for blue-collar elderly workers, an extra year of employment results in a 14 month decrease in longevity. Given that absent the reform blue collar husbands of housewives also lived about 19 of the 21 years, the estimates imply a negative 6% effect on number of years lived between 65 and 85.

Putting together these results, it appears that while the response to the Reform, in terms of employment, is quite similar for the blue- and white-collar groups, the effect on health seems to occur only in the blue-collar group. That is to say, we find that the health of blue-collar workers is affected more adversely by the Reform, not because they delay retirement more than white-collar workers do, but because their jobs take a higher toll on their health.

5.3. The Health Effect of the Reform and the Value of Life

The results we report in the previous sections indicate that employment has a large negative effect on health. However, they also raise another question – are workers aware of the significant cost they pay for prolonged employment in terms of their health? Providing an answer to this question is difficult. Here, we offer a two-step approach using the changes induced by the Reform to recover the value of an additional year of life as reflected in our results. First, we estimate the effect of the Reform on household earnings. Second, we recover the effect of the Reform on life expectancy using our estimates for the effect of the Reform on survival. Together, this allows us to elicit the value that workers attach to an extra year of life under the assumption that workers *are aware* of the health consequences of employment at older ages.

Using a similar specification as applied to the effect of the reform on cumulative employment (see Table 5), we can estimate the effect of the reform on labor earnings of husbands. We find that the reform increased after-tax labor earnings by 24,050 NIS (s.e. 13,117). While we do not observe benefits directly for that period, we can assess the benefits loss resulting from work for affected husbands by multiplying average annual benefits for affected husbands (Column 3 of Table 2) by the extra few months of work (0.327 of a year from Column 1 of Table 5). This calculation amounts to an average benefits loss of 5,100 NIS. Thus, overall net labor earnings for the treated group increased by about 18,900 NIS due to the Reform. Our analysis shows that husbands that were affected by the Reform lost about four months of their life due to working more. However, as discussed above, most of the impact was later in life, suggesting that discounting is important. If individuals understand the effect of employment on mortality, we would expect the total effect of the reform on earnings to at least compensate for the loss of life. Define X to be the monetary value of an extra year of life and suppose that X is constant. The change in earnings would compensate for the loss of life if

$$\Delta earnings \geq -\sum_{t=65}^{85} \Delta P_t \frac{X}{(1+\delta)^{t-65}}$$

where ΔP_t are the estimated effects of the Reform on survival at each age (as in Figure 7), and δ is a discount factor, which we set to 5% as above. From this equation we can back out the maximal value of and extra year of life X. Using our -0.15 estimate of the discounted effect of the Reform on number of years survived (Table 8 Column 1) implies a value of about 126,000 NIS ($39,375 in 1996 terms) for an extra year of life assuming that the worker is indifferent to spending his time working instead of consuming leisure. This estimate is well below the "Value of a Life-Year" typically found in the literature.[47] If we were to assume that the change in earnings had to compensate for both the mortality risk and lost leisure this would suggest an even lower value for an extra year of life.

Returning to the question we posed in the beginning of this section, the results indicate that if workers are aware of the health costs of prolonged employment then the value they assign to an additional life-year is very low. Since there is no reason to suspect that this is the case, the results suggest that elderly workers may be jeopardizing their health because they are unaware of the longevity costs of employment.

6. Conclusions

In this paper, we study the effects of public pension systems on employment of older worker and, in turn, the health consequences of older workers' employment. We leverage a social security Reform in Israel that shifted payments from husbands to their (non-working) wives, thereby reducing the implied tax on the husband's employment while keeping overall benefit generosity roughly constant. We estimate extensive-margin labor supply elasticities of about 0.43 for elderly men w.r.t. the average net-of-tax rate. These numbers, which are consistent with other recent evidence on this issue, support the view that the existence of an implied tax on employment substantially affects the retirement timing of older workers.

We then estimate the effect of employment on longevity of older workers using the Reform to instrument for employment. While we find a cross-sectional positive correlation between employment at old age and longevity, IV estimates indicate that working an additional full year at old age decreases longevity by about one year. Importantly, this effect occurs later on in life and appears to be driven by workers in blue-collar occupations. Finally, we estimate the effect of the Reform on earnings, finding that the Reform increased earnings (net of benefits lost) by about 18,900 NIS. Combining this result with the effect of the Reform on survival at older ages, we find that treated individuals attribute a very small value to an additional year of life. This result calls into question the notion that when making employment

[47] Murphy & Topel (2006), for example, calibrate that the value of a life-year at age 70 is over $200,000 in 2004 dollars (see also Hall & Jones (2007) for further discussion).

decisions around retirement, workers fully understand the cost of employment in terms of their health and longevity and supports the view that workers tend to underestimate these costs.

Granted that recent years have seen many developed countries implement changes to their public pension systems that encourage work among older workers, the results in this paper suggest that a better understanding of factors underlying employment decision-making, as well as of the broader impact of employment at old ages, are crucial for policymaking.

Bibliography

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baker, M., & Benjamin, D. (1999). How do retirement tests affect the labour supply of older men? *Journal of Public Economics*, *71*(1), 27–51.

Behaghel, L., & 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.

Blau, D. M., & Goodstein, R. M. (2010). Can Social Security Explain Trends in Labor Force Participation of Older Men in the United States? *Journal of Human Resources*, *45*(2), 328–363. http://doi.org/10.3368/jhr.45.2.328

Blundell, R., French, E., & Tetlow, G. (2016). Retirement incentives and labor supply. In *Handbook of the economics of population aging* (Vol. 1, pp. 457-566). North-Holland.

Bloemen, H., Hochguertel, S., & Zweerink, J. (2017). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. *Health Economics*.

Bound, J., and Waidmann, T. (2007). Estimating the health effects of retirement. *Michigan Retirement Research Center Research Paper No. UM WP*, *168*.

Burnett, C., Maurer, J. D., & Dosemeci, M. (1997). Mortality by occupation, industry, and cause of death; 24 reporting states, 1984-1988.

Burtless, G., & Moffitt, R. A. (1985). The Joint Choice of Retirement Age and Postretirement Hours of Work. *Journal of Labor Economics*, *3*(2), 209–236. Retrieved from http://www.jstor.org/stable/2534983

Cesarini, D., Lindqvist, E., Notowidigdo, M. J., and Östling, R. (2017). The effect of wealth on individual and household labor supply: evidence from Swedish lotteries. *American Economic Review*, *107*(12), 3917-46.

Chernozhukov, V., & Hansen, C. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, *100*(1), 68–71.

Chetty, R., Guren, A., Manoli, D. S., & Weber, A. (2012). Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities ? A Meta- Analysis of Extensive Margin Elasticities. In *NBER Macroeconomics Annual*.

Chiappori, P.-A., & Mazzocco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, *55*(3), 985–1045.

Coe, N. B., and Zamarro, G. (2011). Retirement effects on health in Europe. *Journal of health economics*, *30*(1), 77-86.

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

Davidson, R., & MacKinnon J. G.(2014). Bootstrap confidence sets with weak instruments. *Econometric Reviews* 33, no. 5-6, 651-675.

Disney, R., & Smith, S. (2002). The labour supply effect of the abolition of the earnings rule for older workers in the United Kingdom. *The Economic Journal*, *112*(478), C136--C152.

Eibich, P. (2015). Understanding the effect of retirement on health: mechanisms and heterogeneity. *Journal of Health Economics*, *43*, 1–12.

Engelhardt, G., & Kumar, A. (2014). Taxes and the labor supply of older Americans: recent evidence from the Social Security earnings test. *National Tax Journal*, *67*(2), 443–458.

Fetter, D. K., & Lockwood, L. M. (2018). Government old-age support and labor supply: Evidence from the old age assistance program. *American Economic Review*, *108*(8), 2174-2211.

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

Friedberg, L. (2000). The labor supply effects of the social security earnings test. *Review of Economics and Statistics*, *82*(1), 48–63.

Friedberg, L., & Webb, A. (2009). 1 New Evidence on the Labor Supply Effects of the Social Security

Earnings Test. *Tax Policy and the Economy*, *23*(1), 1–36.
Gelber, A. M., Isen, A., & Song, J. (2016). *The Effect of Pension Income on Elderly Earnings: Evidence from Social Security and Full Population Data*.
Gelber, A. M., Jones, D. and Sacks, D.W., (2020). *Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test.* American Economic Journal: Applied Economics, 12 (1): 1-31.
Gelber, A. M., Jones, D., Sacks, D. W., & Song, J. (2020). The employment effects of the social security earnings test. *Journal of Human Resources*.
Gelber, A. M., Jones, D., Sacks, D. W., & Song, J. (2017). *Using Non-Linear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Social Security Earnings Test* (NBER Working Paper No. 23362).
Gerdtham, U.-G., & Ruhm, C. J. (2006). Deaths rise in good economic times: evidence from the OECD. *Economics & Human Biology*, *4*(3), 298–316.
Giesecke, M. (2019). *The retirement mortality puzzle: Evidence from a regression discontinuity design* (No. 800). Ruhr Economic Papers.
Gruber, J. (1994). State-mandated benefits and employer-provided health insurance. *Journal of Public Economics*, *55*(3), 433–464.
Gruber, J., & Orszag, P. (2003). Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt? *National Tax Journal*, *56*(4), 755–773. http://doi.org/10.3386/w7923
Gruber, J., & Saez, E. (2002). The elasticity of taxable income: evidence and implications. *Journal of Public Economics*, *84*(1), 1–32.
Gruber, J., & Wise, D. A. (Eds.). (2004). *Social Security Programs and Retirement around the World: Micro Estimation, vol. 2.* Chicago: The University of Chicago Press.
Haider, S. J., & Loughran, D. S. (2008). The Effect of the Social Security Earnings Test on Male Labor Supply: New Evidence from Survey and Administrative Data. *Journal of Human Resources*, *43*(1), 57–87. http://doi.org/10.1353/jhr.2008.0031
Hall, R. E., & Jones, C. I. (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics*, *122*(1), 39–72.
Hallberg, D., Johansson, P., & Josephson, M. (2014). *Early retirement and post retirement health*.
Hernaes, E., Markussen, S., Piggott, J., and Vestad, O. L. (2013). Does retirement age impact mortality?. *Journal of health economics*, *32*(3), 586-598.
Imbens, G., and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475.
Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources*, *49*(1), 195–233.
Johnson, N. J., Sorlie, P. D., and Backlund, E. (1999). The impact of specific occupation on mortality in the US National Longitudinal Mortality Study. *Demography*, 36(3), 355-367.
Krueger, A. B., & Pischke, J.-S. (1992). The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation. *Journal of Labor Economics*, *10*(4), 412–437.Kuhn, A., Wuellrich, J.-P., & Zweimüller, J. (2010). Fatal attraction? Access to early retirement and mortality.
Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655–674.
Lee, D. S., & Lemieuxa, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.
Lundberg, S., & Pollak, R. A. (1993). Separate spheres bargaining and the marriage market. *Journal of political Economy*, *101*(6), 988-1010.
Manoli, D. S., & Weber, A. (2016). *The Effects of the Early Retirement Age on Retirement Decisions* (Working Paper Series). Retrieved from http://www.nber.org/papers/w22561
Manoli, D., & Weber, A. (n.d.). Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions. *American Economic Journal: Economic Policy*.
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.

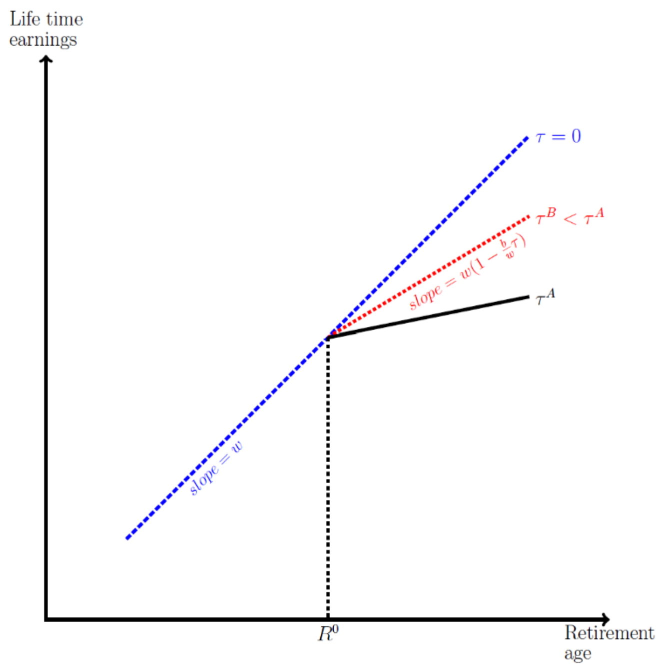
http://doi.org/http://dx.doi.org/10.1016/j.jpubeco.2009.07.009
McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714.
McClelland, Robert, and Shannon Mok. 2012. “A Review of Recent Research on Labor Supply Elasticities.” Congressional Budget Office Working Paper 2012-12.
Moore, D. E., and Hayward, M. D. (1990). Occupational careers and mortality of elderly men. *Demography*, *27*(1), 31-53.
Murphy, K. M., & Topel, R. H. (2006). The value of health and longevity. *Journal of Political Economy*, *114*(5), 871–904.
Neuman, K. (2008). Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research*, *29*(2), 177-201.
OECD. (2015). *Pensions at a Glance*. *Pensions* (Vol. 15). http://doi.org/10.1787/9789264107007-en
Shai, O. (2018). Is retirement good for men’s health? Evidence using a change in the retirement age in Israel. *Journal of health economics*, *57*, 15-30.
Snyder, S. E., & Evans, W. N. (2006). The effect of income on mortality: evidence from the social security notch. *The Review of Economics and Statistics*, *88*(3), 482–495.
Song, J. G., & Manchester, J. (2007). New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000. *Journal of Public Economics*, *91*(3), 669–700.
Stock, J. H., & Wise, D. A. (1990). Pensions, the Option Value of Work, and Retirement. *Econometrica*, *58*(5), 1151–1180. Retrieved from http://www.jstor.org/stable/2938304
Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529.
Sullivan, D., & Von Wachter, T. (2009). Job displacement and mortality: An analysis using administrative data. *The Quarterly Journal of Economics*, *124*(3), 1265–1306.
Voena, A. (2015). Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples?. *American Economic Review*, *105*(8), 2295-2332.

Hebrew Publications

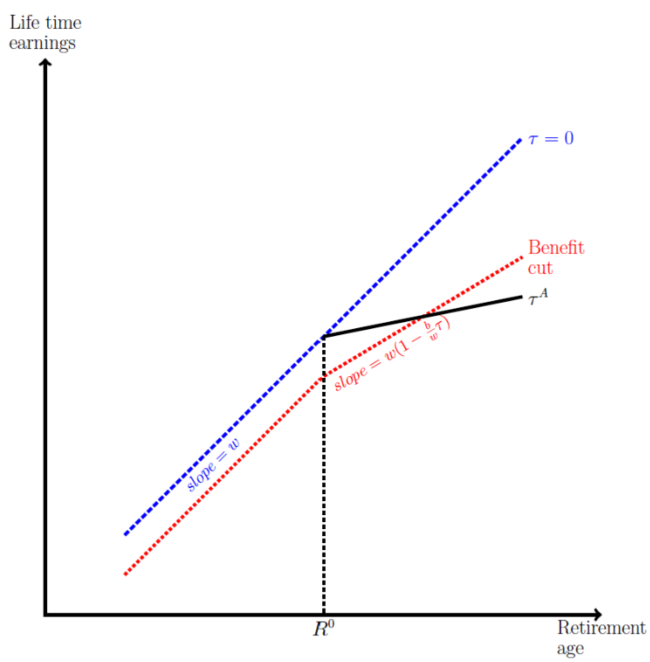
National Insurance Institute of Israel (NII), August 2016. “Monthly Statistical Report.

Figure 1. Life Time Budget Constraint Effects of Social Security Reforms

A. The Housewives Reform



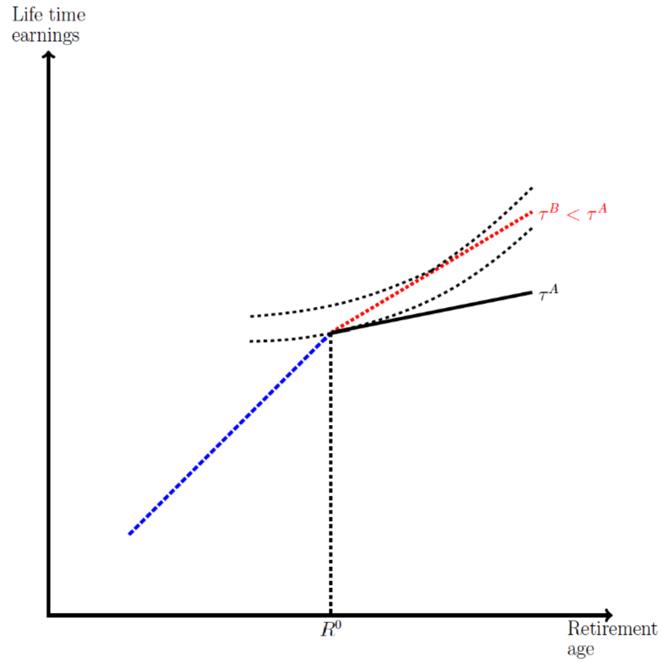
B. A benefit reducing social security reforms



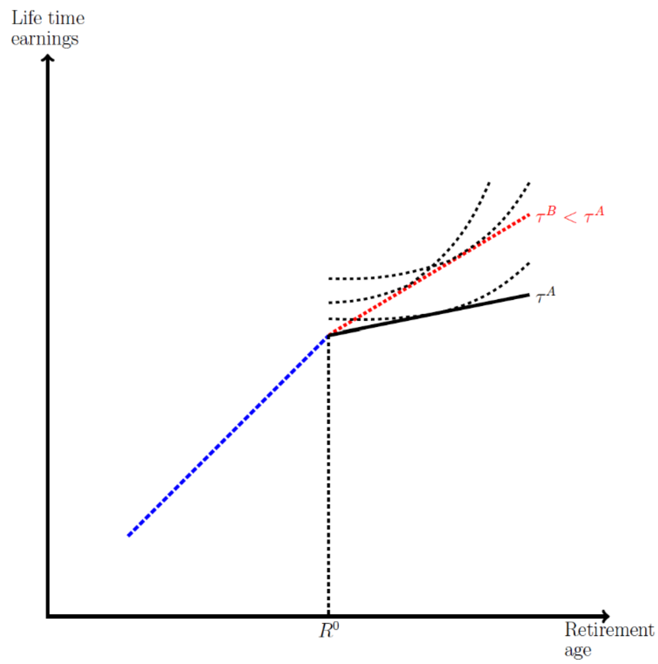
Note: Authors illustration. In panel A, the (blue) dashed line, denoted $\tau = 0$, represents actuarially fair social security system. The (black) solid line, denoted τ^A, represents an earnings test—a tax on delayed retirement. The (red) dotted line, denoted τ^B, illustrates the effect of the Housewives Reform on the budget line, a decrease in the implied tax on delayed retirement. In panel B, the (red) dotted line, denoted Benefit cut, shows the impact of a reduction in social security benefits on life-time earnings, illustrating that such a reform causes a decrease in the implied tax on delayed retirement as well as a decrease in overall life-time earnings.

Figure 2. Expected Behavioral Responses to the Housewives Reform

A. Retire at R^0



B. Retire after R^0

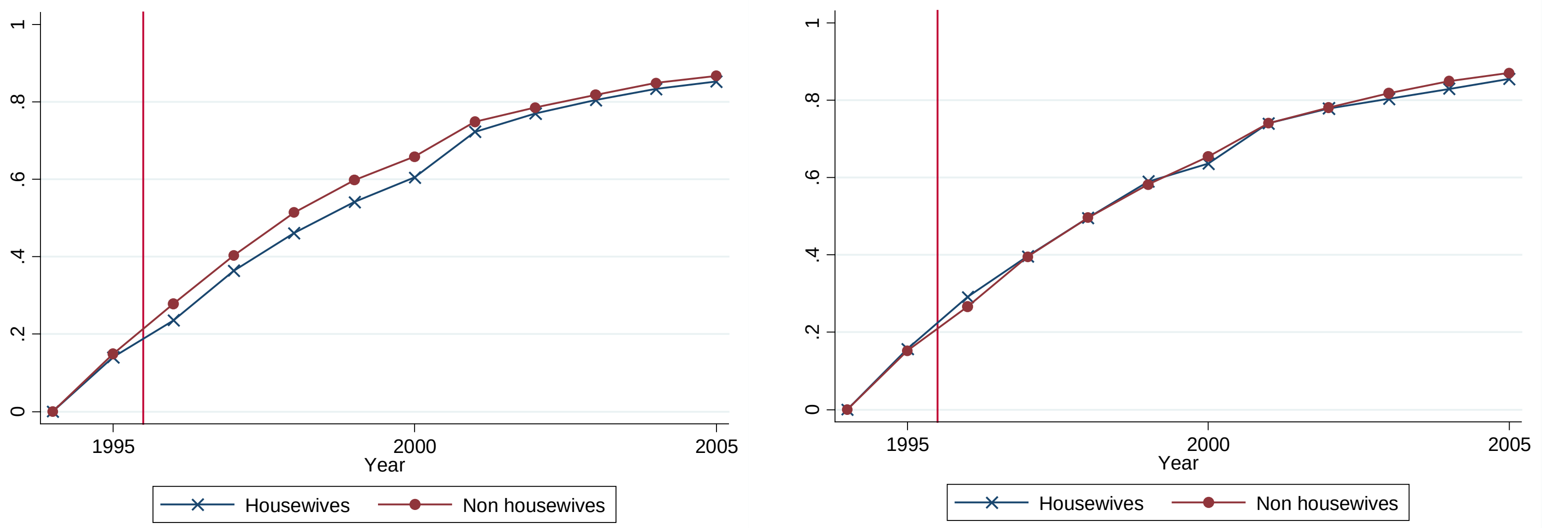


Note: Authors illustration. Panel A of the figure illustrates the response to the Reform by an individual whose intended timing of retirement pre-Reform was R^0. Panel B illustrates the response of an individual whose intended date of retirement was later than R^0.

Figure 3. DDD Results: Husbands Affected by the Reform Retire Later

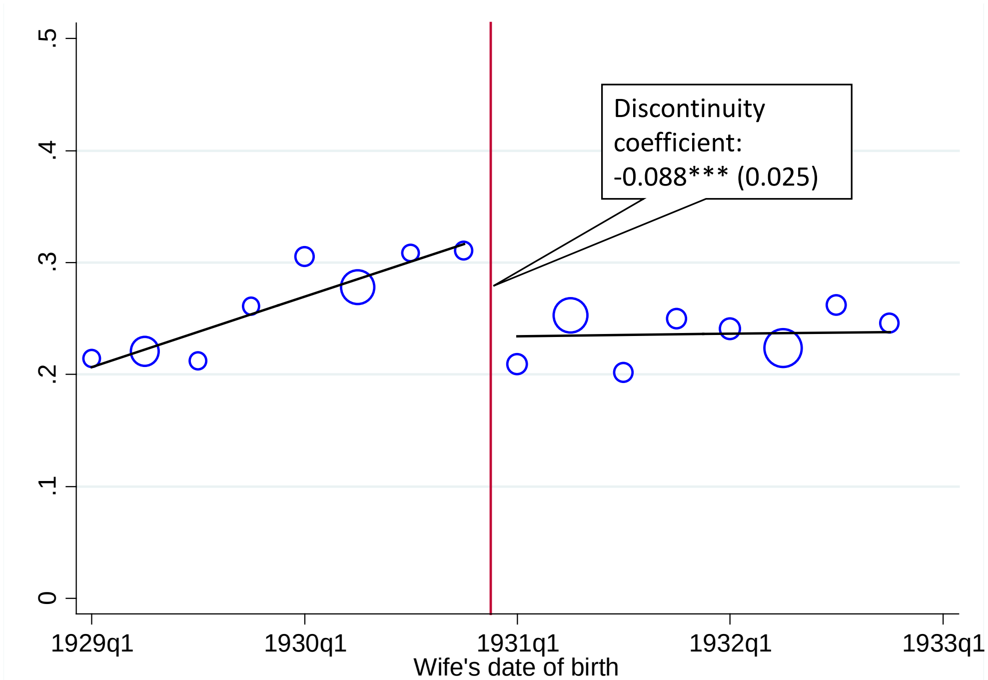
A. 1931-1932 cohort: eligible for the Reform

B. 1930 cohort: non-eligible for the reform



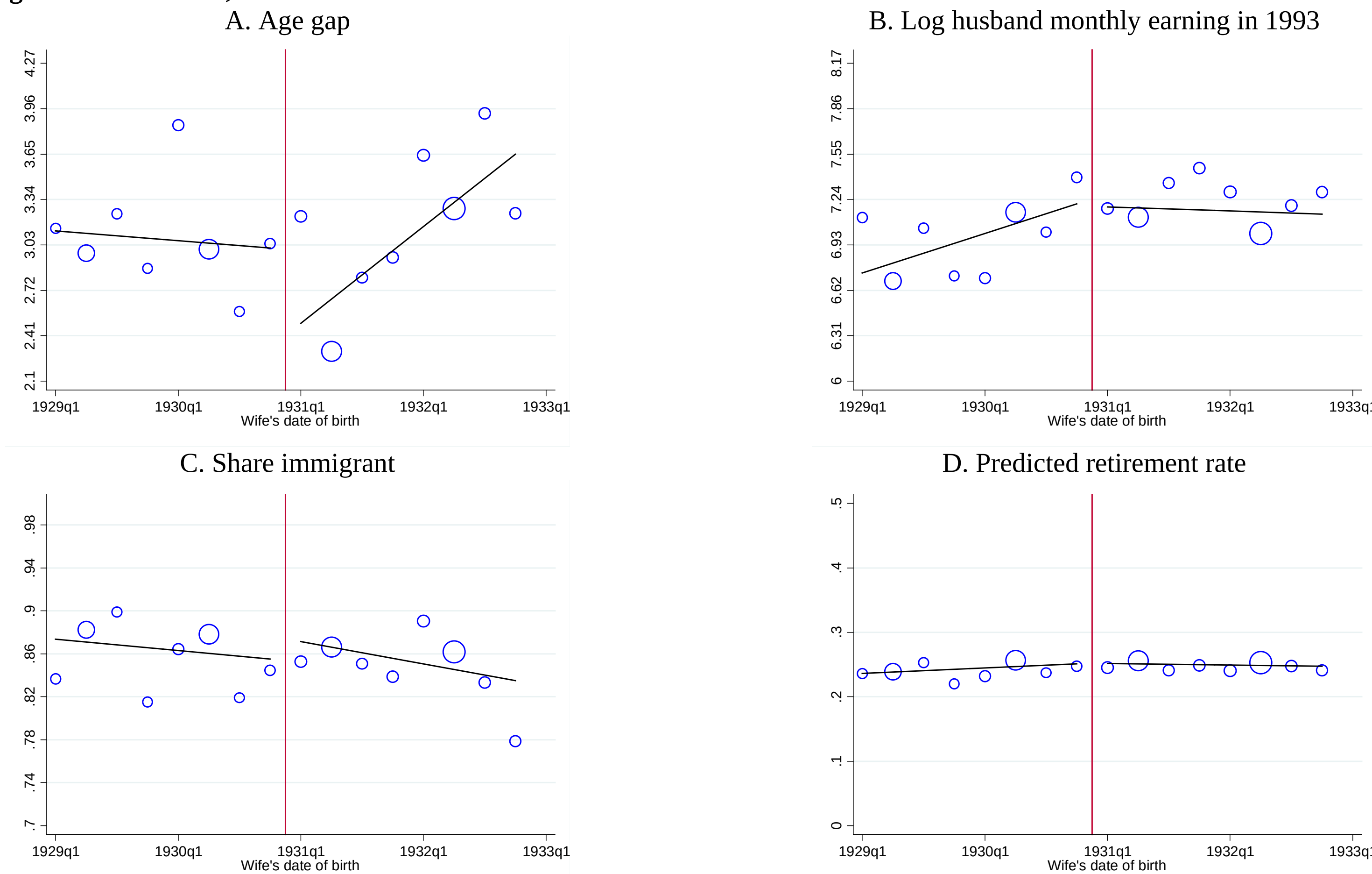
Note: Results from differences-in-differences regressions using the sample of couples that were married in 1996, conditioning on husband's employment in 1994. The differences-in-differences regressions were conducted separately for the 1930 and 1931/32 cohorts to illustrate the patterns for all 4 groups involved in the analysis. The results reported in the text are from the DDD regression which takes the difference between the differences between the two graphs.

Figure 4. **RDD, Retirement Rate by 1996, by Wife's Birth Quarter**



Note: This figure shows retirement rates by 1996 of husbands married to wives born 1929 to 1932, conditional on working in 1994. Circle size is proportional to the number of observations in the cell. Straight lines represent best linear fit on each side of the cutoff.

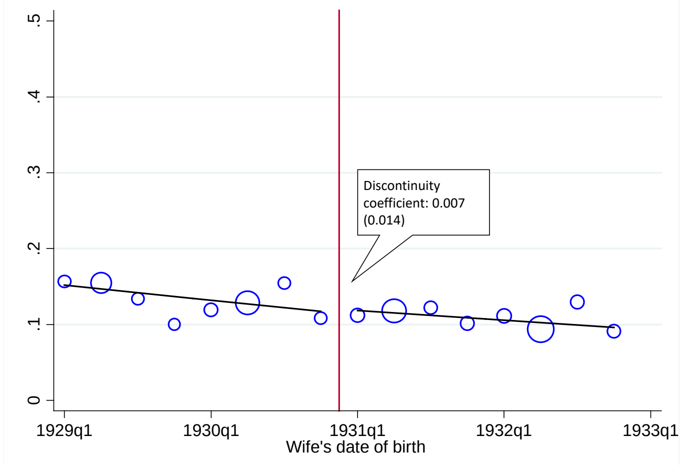
Figure 5. RDD, Selection on Observables



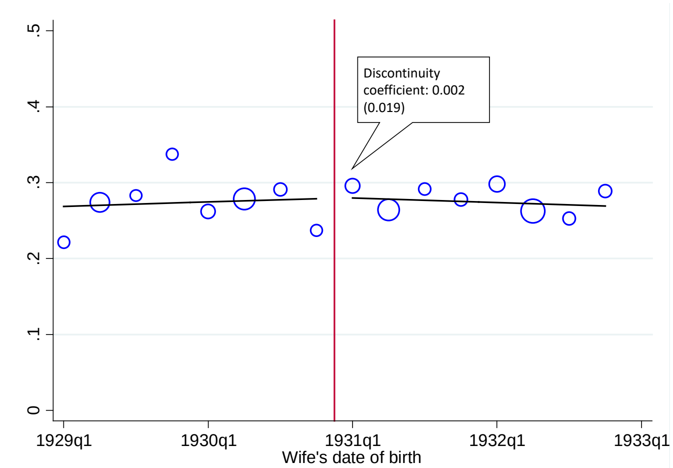
Note: Panels (A)-(C) of this figure display RDD analysis for the covariates age-gap between husband and wife, husband's log monthly earnings in 1993 (replacing log no-earnings with 0), and whether the husband is an immigrant. Panel (D) shows the predicted probabilities from a model that includes all three covariates, as well as 3rd degree polynomials of the first two.

Figure 6. RDD, Retirement Placebo Tests

A. Placebo 1: Housewives, Retirement by 1994 (conditional on work in 1993)

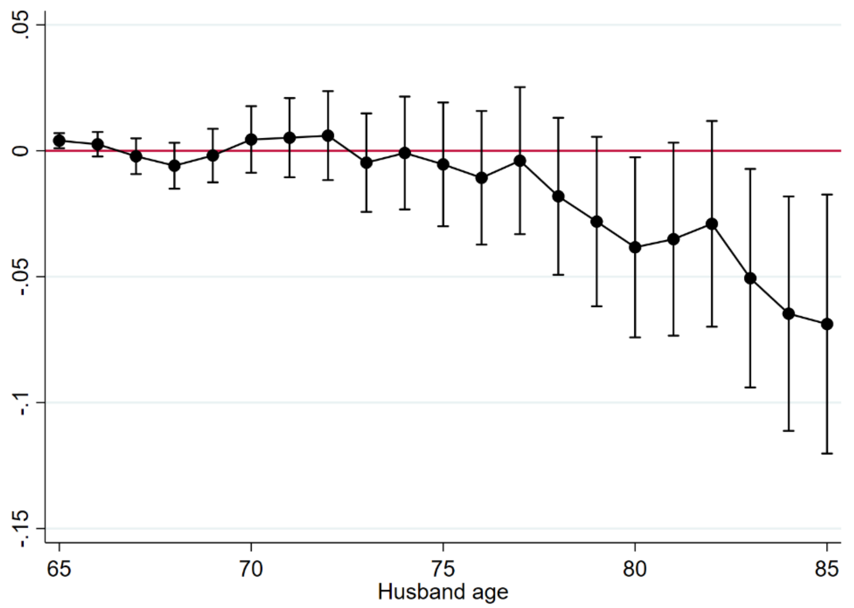


B. Placebo 2: Non-Housewives, Retirement by 1996 (conditional on work in 1994)



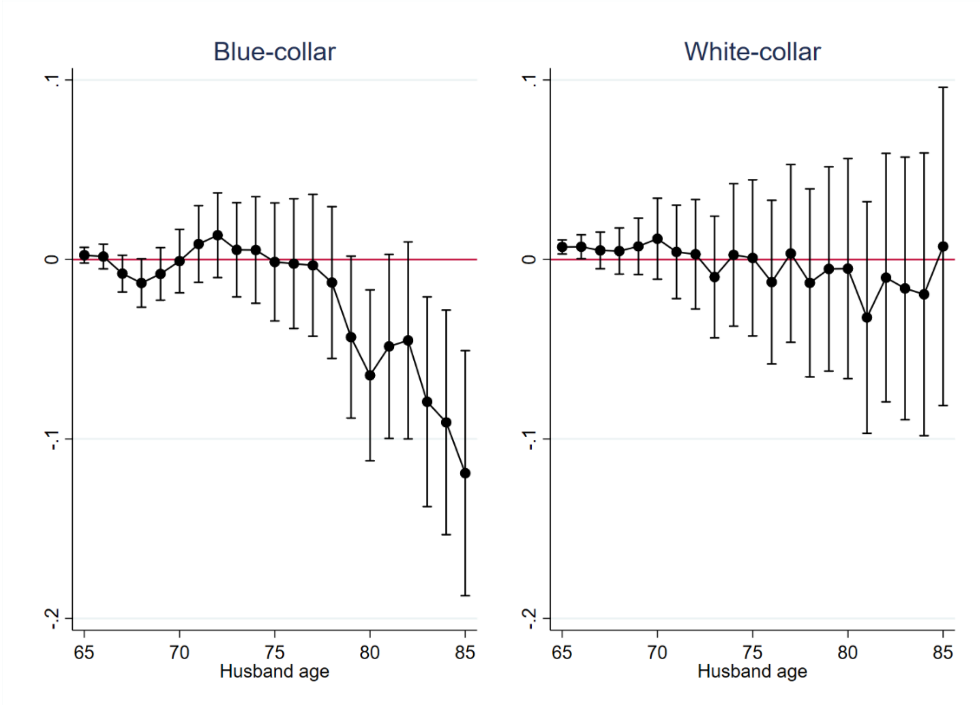
Note: Panel A of this figure shows retirement rates in 1994 of husbands married to wives born 1929 to 1932, conditional on working in 1993. Panel B, displays retirement rates in 1996 of husbands to non-Housewives who were not employed 60 months or more in the ten years before age 65. Circle size is proportional to the number of observations in the cell. Straight lines represent best linear fit on each side of the cutoff.

Figure 7: The Effect of the Housewives Reform on Survival



Note: This figure displays the reduced form results for survival by age. The figure displays the coefficient of the instrument – the interaction term HW× Wife born Jan. 1st 1931 or later and its 90% bootstrap confidence interval, for each of the 21 Survival indicators (survival past 65 - survival past 85) as the outcome variables.

Figure 8: The Effect of the Housewives Reform on Survival by Occupation



Note: This figure displays the age-by-age reduced form results, by industry type. The figure displays the coefficient of the instrument – the interaction term the interaction term HW× Wife born Jan. 1st 1931 or later and its 90% confidence interval, for each of the 21 Survival indicators (survival past 65 - survival past 85) as the outcome variables.

Table 1. Descriptive Statistics

	1930 Cohort		1931-1932 Cohort		Diff in Diff
	(1)	(2)	(3)	(4)	(5)
	HW	Non-Hw	HW	Non-Hw	1931-32 vs 1930
Wife's characteristics					
Immigrant flag	0.83	0.798	0.823	0.795	-0.004
	(0.376)	(0.401)	(0.382)	(0.404)	(0.021)
Jewish	0.895	0.989	0.879	0.986	-0.014
	(0.307)	(0.106)	(0.327)	(0.116)	(0.01)
Immigration year	1951.1	1953.1	1951.6	1953.5	0.184
	(9.1)	(11.7)	(9.6)	(11.9)	(0.655)
Husband's characteristics					
Husband's age in 1993	66.1	65.4	64.5	64	-0.2
	(4)	(4.6)	(4.1)	(3.9)	(0.216)
Immigrant flag	0.863	0.831	0.854	0.817	0.005
	(0.344)	(0.375)	(0.354)	(0.387)	(0.02)
Jewish	0.893	0.987	0.879	0.986	-0.013
	(0.309)	(0.113)	(0.326)	(0.116)	(0.01)
Immigration year	1950.6	1952.2	1950.9	1952.7	-0.275
	(10.5)	(12.4)	(9.8)	(12.2)	(0.668)
Average Income in 1993	76,306	77,956	76,024	85,399	-7725
	(119,709)	(96,082.9)	(105,770.1)	(102,325.1)	(5,430.1)
Average Income in 1993|income>0	89,627	87,672	88,300	94,629	-8284.1
	(125,061)	(97,626.3)	(109,134.4)	(103,579.5)	(5,953.3)
Observations	693	1,949	1,640	4,798	
(% HW within cohort)	(26.2)		(25.5)		

Notes: Descriptive statistics for the sample of couples where wife was born between January 1930 and December 1932, conditioning on husband's employment in 1994. Columns (1) and (2) for wife's birth cohort of 1930, and columns (3) and (4) for birth cohorts 1931-32. Column (5) shows the differences-in-differences for each characteristic (first taking the difference between HW and non-HW within cohort, and then taking the difference of the difference between cohorts). All amounts are in NIS and deflated to 1996.

Table 2. Social Security Benefits by Cohort

	1930 Cohort		1931-1932 Cohorts		Diff in Diff
	(1)	(2)	(3)	(4)	(5)
	HW	Non-Hw	HW	Non-Hw	1931-32 vs 1930
Average benefits 2003-2007					
Wife	54.9	12,939.7	10,238	13,052.3	10,070.4***
	(702.6)	(3,214.1)	(669.4)	(3,372.6)	(172.9)
Husband	24,235.6	15,506.4	15,669.1	15,589.7	-8,649.9***
	(2,843.4)	(3,719.4)	(3175)	(3,526.4)	(211.7)
Total	24,290.6	28,446	25,907.1	28,642	1,420.5***
	(2,831.9)	(5,237.8)	(3,186.1)	(4,484.3)	(263.4)
Observations	521	1,104	1,265	2,948	

Note: Calculated for the sample of households with married wives born in 1930 or 1931-32, conditioning on husband's employment in 1994 and survival until 2007. Columns (1) and (2) report benefits for the 1930 cohort, and columns (3) and (4) for the 1931-32 cohorts. Column (5) shows the differences-in-differences for each row (first taking the difference between HW and non-HW within cohort, and then taking the difference of the difference between them). All amounts are in NIS and deflated to 1996. Average benefits are calculated for the years 2003-2007.

Table 3. DDD Results for the Effect of the Reform on Retirement by Year

	Differences in Differences Estimates				DDD	
	1930 cohort		1931-1932 cohorts			
Coefficient	(1)	(2)	(3)	(4)	(5)	(6)
Year=1995 X HW	0.005		-0.009		-0.014	
	(0.016)		(0.01)		(0.019)	
Year=1996 X HW	0.025	0.02	-0.042***	-0.033***	-0.067***	-0.054***
	(0.02)	(0.015)	(0.012)	(0.009)	(0.023)	(0.017)
Year=1997 X HW	0.001	-0.004	-0.039***	-0.03**	-0.041	-0.027
	(0.022)	(0.019)	(0.014)	(0.012)	(0.026)	(0.022)
Year=1998 X HW	0	-0.005	-0.053***	-0.044***	-0.053**	-0.039
	(0.022)	(0.021)	(0.014)	(0.013)	(0.026)	(0.025)
Year=1999 X HW	0.008	0.003	-0.056***	-0.047***	-0.064**	-0.05*
	(0.022)	(0.022)	(0.014)	(0.014)	(0.026)	(0.026)
Year=2000 X HW	-0.018	-0.023	-0.054***	-0.045***	-0.035	-0.021
	(0.021)	(0.022)	(0.014)	(0.014)	(0.025)	(0.026)
Observations	18,494	18,494	45,066	-0.033***	63,560	63,560

Note: Sample comprised of married couples with wives born between 1930 and 1932, conditioning on husband's employment in 1994. Columns (1) and (2) show differences-in-differences results for households with wives born in 1930 (non-eligible to the reform). Columns (3) and (4) show differences-in-differences results for the 1931-1932 cohorts (eligible to the reform). Columns (5) and (6) show DDD results for the eligible vs. non-eligible groups. Standard errors are clustered at the individual level.

Table 4. Robustness Tests for the DDD Estimation

	DDD 1930, 1931-1932							
	Nearly housewives		Less restrictive		More restrictive		Actual work	
Coefficient	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year=1995 X HW	-0.007		-0.01		-0.027		0.044	
	(0.021)		(0.017)		(0.024)		(0.029)	
Year=1996 X HW	-0.059**	-0.052***	-0.056**	-0.046**	-0.065**	-0.038**	0.077**	0.033
	(0.026)	(0.019)	(0.023)	(0.018)	(0.026)	(0.016)	(0.031)	(0.025)
Year=1997 X HW	-0.048*	-0.041	-0.043*	-0.033	-0.048*	-0.02	0.049	0.005
	(0.028)	(0.025)	(0.025)	(0.023)	(0.026)	(0.021)	(0.031)	(0.03)
Year=1998 X HW	-0.058**	-0.051*	-0.045*	-0.036	-0.054**	-0.027	0.061**	0.017
	(0.029)	(0.027)	(0.026)	(0.025)	(0.026)	(0.024)	(0.03)	(0.032)
Year=1999 X HW	-0.067**	-0.06**	-0.057**	-0.047*	-0.036	-0.009	0.044	0.001
	(0.029)	(0.029)	(0.026)	(0.026)	(0.025)	(0.025)	(0.027)	(0.033)
Year=2000 X HW	-0.034	-0.027	-0.04	-0.03	-0.018	0.009	0.015	-0.028
	(0.028)	(0.029)	(0.026)	(0.026)	(0.024)	(0.026)	(0.026)	(0.033)
Observations	41,573	41,573	63,560	63,560	63,560	63,560	48,048	48,048

Note: Columns (1) and (2) show results for a control sample that only includes wives who were not employed 60 months or more in the ten years before age 65. Columns (3) - (6) show results for alternative definitions of employment. “Less restrictive” (columns (3) and (4)) requires only 3 months of employment per year while our main specification requires a minimum of 6 months of work to be considered employed in that year. “More restrictive” (columns (5) and (6)) requires monthly income to be above the earnings test threshold while our main specification only requires earning at least minimum wage. Columns (7) and (8) use actual work at the particular year (rather than a retirement definition), maintain the work definition from columns (5)-(6). Requiring actual work in 1994 decreases the sample size for this group. Standard errors are clustered at the individual level.

Table 5. Cumulative Effect of the Reform on Employment

Sample	Cumulative number of extra years worked	
	DDD sample	
	(1)	(2)
HW× Wife born Jan. 1st 1931 or later	0.327**	0.285**
	(0.131)	(0.127)
HH level controls	No	Yes
Observations	9,080	9,080

Note: Analysis of the impact of the Reform on husbands' cumulative years of work after 1995. All regressions include a constant, HW dummy, and a dummy for 1931-32 cohort. The controls in column (2) include dummies for Jewish, and for immigrant status of both husband and wife, as well as a 3rd degree polynomial of husband log monthly earnings in 1993 (log no-earnings with 0), and a 3rd degree polynomial of the husband-wife age difference. Standard errors are calculated using Huber-White heteroscedasticity correction.

Table 6. RDD Results for the Effect of the Reform on Retirement by 1996

Polynomial degree	One		Two	
	(1)	(2)	(3)	(4)
Wife born Jan. 1st 1931 or later	-0.088***	-0.091***	-0.07**	-0.072**
	(0.025)	(0.020)	(0.030)	(0.030)
Household Controls	No	Yes	No	Yes
Observations	2,894	2,894	2,894	2,894

Note: Analysis of retirement of housewife husbands in 1996, conditioning on husband's employment in 1994. Polynomials are allowed to differ on two sides of the 1930 quarter 1 cutoff. Household controls include dummies for Jewish, and for immigrant status of both husband and wife, as well as a 3rd degree polynomial of husband log monthly earnings in 1993 (replacing log no-earnings with 0), and a 3rd degree polynomial of the husband-wife age difference. Following Lee & Card (2008), standard errors are clustered by wife's month of birth (the running variable for wife date of birth is discrete at the monthly level).

Table 7. RDD Selection on Observables Tests

	Age Gap		Log husband's earnings in 1993		Husband immigrant		Predicted values	
Polynomial degree	One	Two	One	Two	One	Two	One	Two
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wife born Jan. 1st 1931 or later	-0.618	-0.554	-0.082	-0.288	0.022	0.015	0.002	0.001
	(0.424)	(0.619)	(0.173)	(0.378)	(0.03)	(0.045)	(0.011)	(0.013)
Household Controls	No	No	No	No	No	No	No	No
Observations	2,894	2,894	2,894	2,894	2,894	2,894	2,894	2,894

Note: RDD analysis for the covariates age-gap between husband and wife, log husband's monthly earnings in 1993 (replacing log no-earnings with 0), and whether the husband is an immigrant. Columns (7) and (8) show the predicted probabilities from a model that flexibly includes all three covariates. Following Lee & Card (2008), standard errors are clustered by wife's month of birth (the running variable for wife date of birth is discrete at the monthly level).

Table 8. The Effect of Employment on Life-Expectancy: Difference-in-Differences Estimates

		Reduced form		OLS		IV	
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample							
	Number of years survived between age 65 and 85	-0.346	-0.372	0.351	0.339	-0.911	-1.033
		[-0.746,0.037]	[-0.733,0.008]	[0.32,0.387]	[0.305,0.375]	[-3.941,0.174]	[-4.058,0.197]
outcome variable	Number of years survived between age 65 and 74	0.007	0.012	0.094	0.09	0.021	0.043
		[-0.102,0.099]	[-0.095,0.102]	[0.084,0.105]	[0.081,0.101]	[-0.492,0.348]	[-0.537,0.461]
	Number of years survived between age 75 and 85	-0.353	-0.384	0.258	0.249	-0.932	-1.076
		[-0.684,-0.023]	[-0.686,-0.05]	[0.23,0.289]	[0.221,0.279]	[-3.314,0.029]	[-4.032,-0.027]
Discounted number of years survived between age 65 and 85							
		-0.15	-0.158	0.194	0.187	-0.398	-0.443
		[-0.37,0.045]	[-0.361,0.038]	[0.177,0.213]	[0.169,0.206]	[-1.956,0.186]	[-1.946,0.249]
	Observations	9,080	9,080	9,080	9,080	9,080	9,080
Panel B: Sample of survivors to age 74							
	Number of years survived between age 75 and 85	-0.354	-0.367	0.099	0.102	-0.777	-0.843
		[-0.602,-0.069]	[-0.617,-0.096]	[0.077,0.124]	[0.08,0.128]	[-1.819,-0.18]	[-2.005,-0.205]
	Observations	7,978	7,978	7,978	7,978	7,978	7,978
	HH level controls	No	Yes	No	Yes	No	Yes

Note: This table provides a summary of the age-by-age estimates of the survival analysis in Equation (2). Every statistic in the table is the sum of the age-by-age survival regression coefficients for the relevant age range. For each statistic, we report in square brackets its 90% bootstrap confidence interval (with 500 replications).

Table 9. The Effect of Employment on Life-Expectancy: Blue vs. White Collar Occupations

		Blue		White	
		Reduced form	IV	Reduced form	IV
		(1)	(2)	(3)	(4)
Panel A: Full sample					
	Number of years survived between age 65 and 85	-0.504	-1.533	-0.06	-0.097
		[-0.985,-0.01]	[-9.289,2.364]	[-0.64,0.599]	[-1.702,1.426]
outcome variable	Number of years survived between age 65 and 74	0.007	0.024	0.043	0.092
		[-0.128,0.163]	[-1.045,0.962]	[-0.112,0.199]	[-0.355,0.676]
	Number of years survived between age 75 and 85	-0.51	-1.557	-0.103	-0.189
		[-0.916,-0.087]	[-7.719,1.814]	[-0.567,0.516]	[-1.702,0.963]
	First stage		0.289*		0.464**
			(0.176)		(0.216)
Discounted number of years survived between age 65 and 85					
		-0.218	-0.675	-0.01	-0.008
		[-0.468,0.043]	[-4.003,1.047]	[-0.311,0.328]	[-0.823,0.84]
	Observations	4721	4721	3547	3547
Panel B: Sample of survivors to age 74					
	Number of years survived between age 75 and 85	-0.501	-1.151	-0.197	-0.307
		[-0.888,-0.103]	[-4.455,-0.116]	[-0.693,0.236]	[-1.72,0.359]
	First stage		0.408**		0.582**
			(0.182)		(0.228)
	Observations	4162	4162	3090	3090

Note: This table provides a summary of the age-by-age estimates of the survival analysis in Equation (2). Every statistic in the table is the sum of the age-by-age survival regression coefficients for the relevant age range. For each statistic, we report in square brackets its 90% confidence intervals which we calculated using bootstrapping (with 500 replications). The number of observations does not sum up to 9,080 because for 812 husbands of the sample we did not have the industry composition information and could not attribute the type of industry.