Social Insurance for Entrepreneurs*

Youssef Benzarti† Jarkko Harju‡ Tuomas Matikka§
Ella Mattinen¶ Alisa Tazhitdinova†

October 14, 2025

Abstract

We estimate entrepreneurs' preferences for social insurance (SI) coverage and the extent of adverse selection and moral hazard in social insurance for business owners in Finland. Our results show that, when given a choice over SI contributions, owners prefer contributions and coverage levels that are on average 16% lower than the standard mandate. To identify adverse selection and moral hazard responses, we use positive correlation tests and a reform that allowed a subset of business owners to freely choose their SI contributions and coverage. We find that both moral hazard and adverse selection responses are limited.

Keywords: entrepreneurs; social insurance; moral hazard, adverse selection.

JEL: H55, J32, L26.

*We thank seminar participants at Pomona College, Tampere University, University of Oslo, UC San Diego, VATT, the SustAgeable Consortium Meeting, 2023 NTA conference, 2024 STAX meetings, 2024 IIPF congress, and 2025 Spring NBER Public Economics meetings for insightful comments.

†University of California, Santa Barbara and NBER

‡Tampere University, FIT, VATT and CESifo

§VATT, FIT and CESifo

¶Tampere University, FIT and VATT

1 Introduction

Self-employment, independent contracting, and small business ownership have been steadily increasing over recent decades in developed countries (Boeri et al., 2020; Goetz et al., 2025). The majority of small business owners face substantially greater income volatility than wage earners and may therefore stand to benefit from access to social insurance programs such as sick leave pay and disability benefits (Audoly, 2022; Boeri et al., 2020). Nevertheless, entrepreneurs are rarely fully covered under traditional social insurance schemes. While most OECD countries extend some form of social insurance to the self-employed, such coverage is typically voluntary or limited to protection against extreme poverty.

Designing social insurance for business owners is challenging. They are often better informed about the risk profiles of their businesses and may possess a greater capacity to exploit social insurance benefits than regular wage earners. Consequently, social insurance programs targeting business owners can be susceptible to adverse selection, and may entail higher costs due to potential moral hazard responses, where a higher level of social insurance benefits would lead to increased benefit claiming. However, the magnitude of adverse selection and moral hazard among business owners has not been previously examined, largely because of the absence of suitable data and identifying variation. Moreover, it remains unclear whether entrepreneurs themselves value more extensive social insurance coverage or, alternatively, prefer to allocate their scarce financial resources toward investments in their businesses.

In this paper, we exploit a unique institutional setting in Finland to provide novel evidence on small business owners' willingness to pay for social insurance (SI) and the extent of adverse selection and moral hazard in this market. In Finland, certain business owners can relatively freely choose their level of SI coverage, thereby determining both the size of their SI contributions and the corresponding pension, sick leave, and parental leave benefits they are entitled to. Before 2011, this flexibility was limited to business owners holding more than 50% of the shares of their firm, while those with smaller ownership shares were

mandatorily enrolled in the traditional SI scheme similar to wage earners, with contributions and benefits fixed according to earnings and no option to deviate from this regulation. A policy reform in 2011 lowered the ownership threshold from 50% to 30%, granting owners with a 31–50% ownership the ability to determine their SI contributions.

First, we apply a standard positive correlation test framework (e.g. Chiappori and Salanie, 2000a) to examine the relationship between coverage choices and benefit claims, revealing how risks are associated with coverage levels among business owners who have discretion over their SI choices. Second, we analyze how SI contributions evolve before benefit claims, capturing potential anticipatory responses in SI coverage. These approaches conflate both adverse selection and moral hazard responses, and provide informative evidence on the overall association between coverage levels and benefit claims. To separately identify moral hazard, we leverage the exogenous policy change described above, which allows us to estimate entrepreneurs' preferences for social insurance and its impact on claiming behavior. Underlying our difference-in-differences approach is the assumption that the control and treatment groups evolve similarly. We test this assumption by showing that the trends in SI contributions and benefit claims follow parallel trends in the treatment (business owners with 31–50% ownership share) and control group (business owners with 10-30% and 51–70% ownership).

We further complement our analysis with surveys of entrepreneurs conducted by Statistics Finland. The surveys include detailed information on attitudes toward the SI system, knowledge about social benefits, and self-assessed health status and retirement plans. Using unique identifiers, we are able to link the survey responses with administrative records, enabling us to reveal, for example, how retirement plans and perceived health are linked to actual SI coverage choices. Overall, our analysis draws on exceptionally rich administrative data sources covering social insurance contributions, benefit claims, and comprehensive individual and business tax records.

Our first result shows that business owners with discretion over their SI coverage

choose substantially lower levels of coverage than comparable owners whose contribution levels are mandated. In particular, those with higher total income choose significantly lower SI coverage rates than other business owners under the standard SI mandate with similar income levels. Second, following the 2011 reform, business owners with ownership shares between 31 and 50% – who thereby became eligible to choose their social insurance contributions – reduced both their contributions and coverage by approximately 16% compared with otherwise similar owners lacking such discretion (those with a 10–30% ownership share). This finding confirms our descriptive evidence that those owners who are allowed flexible contributions tend to contribute substantially less.

Beyond underlying preferences, SI coverage choices may also be influenced by perceived or private information about risks, potentially generating adverse selection or moral hazard (Chetty and Finkelstein, 2013; Chiappori and Salanie, 2000b). However, in our setting, we find that SI contributions are only weakly correlated with key outcomes determining benefit usage, such as retirement age, life expectancy, sick leave or parental leave benefit claims. This evidence suggests a limited selection of SI contribution by risks. To assess the magnitude of the observed correlations, for pension benefits we estimate that the cost of selection is only 7.2 cents per additional euro of SI contributions. The corresponding estimates for sick leave and parental leave benefits are even smaller – 0.7 and 0.6 cents, respectively. Consistent with these results, we find no evidence of anticipatory adjustments in SI contributions in the years preceding predictable claim events such as sickness, childbirth, or retirement.

Our analysis of the correlations between SI coverage and benefit claims suggests that both moral hazard and adverse selection are limited in this setting. However, this pattern may be affected by offsetting forces. Advantageous selection – where lower-risk individuals insure themselves more than higher-risk individuals – could mask underlying moral hazard effects. To disentangle potential moral hazard responses from adverse selection, we exploit the 2011 reform using an event-study design. Specifically, we compare the SI

claims of individuals affected by the reform (those with 30–50% ownership shares) to those of unaffected individuals (with 10–30% or 50–70% ownership shares) before and after the reform. This design leverages the fact that assignment to treatment and control groups is fixed over time. By comparing pre- and post-reform claims within these groups, we effectively observe the same individuals' behavior over time. As a result, the adverse selection channel is eliminated, and differences in claims identify moral hazard responses, provided that individual risk profiles remain stable or evolve similarly over time across the treatment and control groups due to, for example, inherent risk-type volatility or broader institutional and demographic changes.

Our event-study estimates indicate that moral hazard responses are negligible. Despite substantial variation in SI contribution levels induced by the reform, we find no evidence that this variation is linked with benefit usage. We find that lower SI coverage choices do not translate into fewer claims, nor do higher contributions result in substantially larger claims in the future. This points to the fact that the observed correlations primarily reflect selection in SI choices rather than moral hazard effects.

We leverage entrepreneurial surveys to better understand other potential mechanisms behind SI choices and the absence of sorting by risk types. First, we find that limited trust in the social insurance system, liquidity constraints, private savings and expectations of continued work during retirement are among the most frequently reported reasons for opting for lower SI coverage. Nevertheless, by leveraging the unique opportunity to combine survey responses with register data, we find that the actual coverage choices are remarkably similar across various stated reasons for choosing low SI coverage. Consistent with the survey answers on private savings crowding-out SI contributions, we find a gradual increase in personal capital income among those with discretion in selecting their SI coverage after the reform, suggesting that other types of savings seem to at least partly substitute lower SI coverage.

Furthermore, survey evidence indicates that entrepreneurs' perceived health status

is not associated with their SI coverage choices, further supporting our findings on the lack of adverse selection on sick leave benefits. Also, we find that stated knowledge about the SI system is not significantly associated with differences in actual SI choices. However, we find a small association with SI coverage and stated plans on working after retirement: those with no plans on continuing working past the retirement age because of health reasons have slightly larger SI contributions compared to those who plan to work when retired. This finding aligns with our evidence on the small negative association between SI contributions and retirement age.

In sum, our findings indicate that both moral hazard responses and adverse selection are negligible in the social insurance market for business owners in Finland. The absence of significant adverse selection implies that imposing a stricter social insurance mandate on business owners may not be warranted. Indeed, such mandates could be distortionary, as they may impose unnecessary liquidity constraints on younger and growing firms (Benzarti et al., 2020). At the same time, the lack of detectable moral hazard responses shows that entrepreneurs do not systematically exploit the social insurance system. This implies that offering business owners social insurance coverage comes with low efficiency costs.

To the best of our knowledge, this is the first study to estimate moral hazard and adverse selection responses to social insurance for entrepreneurs. While the existing literature has examined the extent of adverse selection and moral hazard in a range of settings – such as health insurance markets, unemployment insurance, and disability programs (e.g., Finkelstein and McGarry, 2006; Einav et al., 2010; Finkelstein and Poterba, 2004; Landais et al., 2021) – none of these studies focus on the self-employed or small business owners. This omission is notable, as entrepreneurs face distinctive and often more volatile risks, making the design of social insurance policies particularly consequential for them. Moreover, our study contributes to the literature on the adverse selection and moral hazard effects of publicly-mandated social insurance schemes in general.

By focusing on entrepreneurs, our study fills this gap in two important ways.

First, we provide direct evidence on whether the theoretical concerns about moral hazard and adverse selection – often assumed to be particularly acute in this group – materialize in practice. Second, we offer novel insights into how entrepreneurs respond to policy-induced variation in social insurance coverage, thereby contributing new empirical evidence to inform the design of more efficient and equitable social insurance systems for this economically important segment of the labor market.

2 Institutional Setting

2.1 The Finnish Social Insurance System

Overview. The Finnish social insurance system is financed through taxes and social insurance contributions that are based on individual earnings. These contributions are paid by both employees and employers so that approximately one-third is nominally paid by employees and two-thirds by employers. Although social insurance contributions are deductible from taxable income, they are otherwise separate from income taxation. The system provides coverage for individuals who reside permanently in Finland, as well as for those working in the country on a temporary basis.

The Finnish system offers a broad range of benefits, including old-age and disability pensions, sickness and unemployment benefits, parental allowances, and rehabilitation support. The level of these benefits is earnings-related and thus directly linked to prior social insurance contributions. In addition to earnings-related benefits, flat-rate minimum benefits are available for individuals with low or no previous earnings.[1] Beyond these social insurance benefits, publicly provided health care is universally available to all residents in Finland.

[1]Furthermore, a means-tested, last-resort income support is available for households whose income from work, benefits, or assets does not cover essential needs such as food and housing.

TyEL and YEL Contribution Schemes. Wage earners in Finland receive social insurance benefits through the TyEL insurance scheme ("Työnantajan eläkevakuutus" in Finnish). Under TyEL, social insurance contributions are standardized: individuals have no discretion over either the level of contributions or the scope of coverage. Contributions are determined by a government-mandated rate and are assessed on the basis of earnings. In 2016, the total contribution rate under TyEL – including both employee and employer components – was 23.6% of earnings.[2]

Self-employed individuals and owners of partnerships and privately held corporations are covered by social insurance through one of two schemes. The YEL insurance scheme ("Yrittäjän eläkelaki") applies to all self-employed individuals, and to owners of unincorporated partnership firms who, alone or together with family members, own at least a specified share of their firm.[3]

In addition, owners of privately held corporations who hold a leading position in the firm (e.g., CEO or chair of the board) and own a specified minimum share of the company are also classified as YEL entrepreneurs. All these conditions are binding, and entrepreneurs cannot opt out of the YEL insurance scheme. Entrepreneurs who do not meet these criteria are automatically covered by the TyEL insurance scheme, that is, the same insurance program that applies to wage earners.

The main difference between the TyEL and YEL schemes is that while the TyEL-covered owners have no discretion over their social insurance contributions, YEL-covered owners can very freely set their contribution level. To do this, YEL owners are required to report an income level to an insurance company administering their contributions. Social

[2]Contribution rates vary slightly over time and are set annually. Rates also differ by age; for example, in 2016 the TyEL contribution rate for individuals aged 53 or older was 25.2%, compared to 23.6% for younger individuals.

[3]To qualify for YEL, entrepreneurs must also meet several conditions: they must be between 18 and 67 years of age, their business must have been operating for at least four months, and the annual value of their work contribution for the firm must exceed 7557 euros (in 2016). The Self-employed Persons' Pension Act (HE 1272/2006) is available online here (in Finnish): https://www.finlex.fi/fi/laki/alkup/2006/20061272 (accessed May 24, 2019). More information in English is available here: https://www.ilmarinen.fi/en/self-employed-person/self-employed-persons-pension-insurance/yel-contributions (accessed May 24, 2019).

insurance contributions are then paid based on this reported income. The reported income must fall within a specific range (in 2016, between 7557 and 171,625 euros), and the owners can adjust their reported incomes at any point in time.[4] For simplicity, we refer to the reported income as the *social insurance (SI) income*, and to their overall earnings as their *total income (TI)*. The overall earnings include both wage income and dividend income from the firm.

YEL regulations stipulate that the reported SI income should correspond to the wage that the business owner would pay an external employee performing similar tasks within the firm. The underlying policy rationale is to ensure that social insurance coverage provides adequate protection against earnings losses in an adverse event such as sickness, and adequate pension when retired. In practice, however, this rule is not strictly enforced, primarily because the appropriate counterfactual wage for an external employee is often very hard to approximate. As a result, entrepreneurs effectively self-report their SI income without automatic adjustments when their actual earnings fluctuate. This contrasts with the TyEL scheme, where contributions and coverage levels are automatically linked to the individual's actual earnings. The contribution rate applied to the SI income under YEL is broadly comparable to that under TyEL – 23.7% versus 23.6% in 2016.[5] Importantly, there are no differences in income taxation between YEL- and TyEL-covered individuals.

In summary, YEL owners have substantially more discretion over their mandatory social insurance contributions than TyEL participants. Since future benefit entitlements are directly linked to the level of contributions, any adjustment in self-reported SI income by YEL entrepreneurs effectively alters the benefits to which they will be entitled. In contrast, for TyEL-covered individuals, both contributions and benefit entitlements are automatically determined based on actual labor earnings, leaving no scope for individual discretion. For

[4]To update the social insurance income, an entrepreneur would have to contact the pension company and fill out minimal paperwork. There are no restrictions on how frequently such changes can be made.

[5]YEL contribution rates, as well as the minimum and maximum SI income thresholds, are adjusted annually. Rates also vary by age, with higher contributions for older individuals. For example, in 2016, the YEL rate for individuals aged 53 to 62 was 25.1%. Lower contribution rates apply to new entrepreneurs: 19.6% for those aged 53 to 62, and 18.4% for younger individuals.

this reason, we refer to all entrepreneurs to whom YEL rules apply as *unconstrained* owners, and all entrepreneurs to whom TyEL rules apply as *constrained* owners in our analysis.

Ownership Share Thresholds and the 2011 YEL Reform. Eligibility for the YEL insurance scheme requires that a business owner, either individually or together with family members, holds a minimum ownership share in the firm. Prior to 2011, this ownership threshold was set at 50%. A policy reform in 2011 lowered this threshold to 30%, thereby extending YEL eligibility – and the associated discretion over social insurance contributions – to owners with a 30–50% ownership share.[6] Aside from this reform, there were no other significant changes to the social insurance system during the period we study. In our empirical analysis, we will compare the outcomes of those owners who switched from TyEL to YEL (ownership shares 30–50%) to those who remained YEL owners (51–70%) and to those remained TyEL owners (10–29%) both before and after the reform, discussed in more detail below.[7]

Benefit Entitlements. Our analysis focuses on three key categories of earnings-related benefits available to entrepreneurs: retirement, sickness, and parental leave benefits.

Old-age pension benefits and *disability pension benefits* are determined based on the total accumulation of SI income (or earnings for wage earners) over an individual's working career, whether earned as an employee or entrepreneur. Contributions made in each year of employment are weighted approximately equally in the pension formula, with an accrual rate of about 1.5% per year. Consequently, the level of pension benefits increases with both the number of years worked and the amount of SI contributions, regardless of

[6]More information on the reform (HE 135/2010) is available here (in Finnish): https://www.finlex.fi/fi/esitykset/he/2010/20100135 (accessed May 24, 2019).

[7]Entrepreneurs with ownership shares between 30–50% were permitted to opt out of switching to the YEL scheme for a transitional period of up to three years following the 2011 reform. However, in practice, we observe that most eligible owners transitioned to YEL coverage shortly after the reform was implemented. Additionally, newly eligible YEL participants in this ownership range were granted a 22% discount on their annual SI contributions for a maximum of three years, similarly as all new business owners who enter the YEL system. Importantly, during this discount period, benefit entitlements were calculated based on the full reported SI income, rather than the discounted contribution amounts.

when in the working career those contributions are made.[8] An exception to this rule was the increased accrual rates for individuals who continued working past the minimum retirement age of 63, which was in place until 2017. Before 2017, the annual accrual rate was increased to 4.5% for individuals aged 63–68.

Pension benefits are not subject to a cap, and claiming old-age pension does not prevent individuals from continuing to work or operate a business. As a result, entrepreneurs may continue to actively manage their firms while receiving pension benefits. If the earnings-based pension is very low, it is supplemented with a guaranteed pension to provide a minimum basic income to retired individuals (approximately €770 per month in 2016). Full old-age pension benefits can be claimed once an individual turns 63, while reduced partial pension benefits can be claimed starting at age 60 in our examination period. By law, individuals are required to begin collecting the old-age pension benefit by the age of 68.

Pension benefits are likely the most salient component of social insurance for entrepreneurs. First, most entrepreneurs will eventually consider retirement or claiming pension benefits, making this form of insurance particularly relevant over the life cycle. Second, the official Finnish name of the YEL program, *Yrittäjien eläkevakuutus*, translates to "Pension Insurance for Entrepreneurs," even though the scheme also covers other benefits such as sickness and parental leave. This nomenclature likely reinforces entrepreneurs' awareness that YEL contributions directly affect their future pension entitlements. Given this, a central focus of our analysis is on examining the relationship between YEL contributions and pension claiming behavior.

Parental leave benefits can be claimed by either parent or jointly, for a maximum duration of six months. Typically, parental leave begins when the child is approximately three months old, following the conclusion of maternity leave, and extends until the child

[8]For example, if a business owner reports an SI income of €10,000 for 15 years and €30,000 for 20 years, the approximate annual pension would be 0.015×(15×€10,000 + 20×€30,000)=€11,250. The exact pension amount may differ slightly due to specific provisions tied to the statutory retirement age, periodic regulatory adjustments and an annual pension index that is based in on inflation and wage development in Finland.

reaches nine months of age.[9] Until 2015, the parental leave benefit was approximately 70% of the individual's SI income over the preceding 12 months, for SI incomes up to €32,892. For SI incomes above this threshold, the replacement rate declined. For example, in 2010, an individual with an SI income of €50,000 would have received a parental allowance of €99.60 per working day.

sick leave benefits can be claimed when individuals experience a long-term health issue, subject to eligibility based on a medical review. The replacement rate for sick leave follows the same rules as those for parental leave benefits, but it reduces more rapidly for larger incomes compared to parental leave.[10] For TyEL owners, there is a nine-day waiting period at the start of a sickness spell, after which sickness benefits can be claimed for up to a maximum of 300 working days. In contrast, YEL owners face only a one-day waiting period before becoming eligible. Given this distinction, we focus our analysis on sickness benefit spells that extend beyond the first nine days. In the results section, we further discuss how the difference in waiting periods between the TyEL and YEL insurance schemes may affect the interpretation of sick leave claims following the 2011 reform.

Unemployment benefits Entrepreneurs are also entitled to unemployment benefits under certain conditions. However, we exclude unemployment benefits from our analysis for several reasons that would otherwise complicate it. First, eligibility for earnings-related unemployment benefits requires membership in an unemployment fund for at least six months prior. Membership in these funds is voluntary and involves a small annual fee. While over 70% of wage earners in Finland belong to unemployment funds, the membership rate among

[9] In addition to parental leave, parents are eligible for maternity leave and home care leave. Maternity leave is usually claimed starting one month prior to birth and can last up to four months. Maternity leave benefits also depend on an individual's SI income. However, since nearly all mothers claim this benefit, we exclude it from our analysis. Home care leave, by contrast, is typically claimed after parental leave and is available until the child turns three years old. Home care leave provides lower replacement rates than maternity or parental leave, and in many municipalities, the benefit is independent of SI income. For these reasons, we exclude this benefit from our analysis. Maternity and parental leave benefits can overlap by at most three weeks.

[10] The current benefit formulas can be found at the Finnish Social Insurance Institution's website, https://www.kela.fi/annual-income-calculation.

entrepreneurs is substantially lower.[11] Consequently, for the vast majority of entrepreneurs, unemployment benefits do not depend on their selected SI income, implying that choices of SI income based on unemployment risk are largely irrelevant. Moreover, our data do not allow us to identify unemployment fund members, preventing a focused analysis on entrepreneurs eligible for earnings-related unemployment benefits. Second, strict eligibility requirements further limit benefit receipt for entrepreneurs. Claimants must register as unemployed job seekers, actively search for full-time work, and be ready to accept a new job when offered. These conditions effectively prevent entrepreneurs from continuing to run their firms while claiming unemployment benefits. To qualify for unemployment benefits, they generally must cease their business activities and exit the YEL insurance scheme. For these reasons, we do not examine unemployment benefits in this paper.

3 Data and Descriptive Statistics

3.1 Data and Sample Restrictions

We use multiple administrative data sources in this study: (1) social insurance contribution records from Finland's two largest pension providers administering the TyEL and YEL systems, (2) individual-level annual benefit claims data, and (3) comprehensive individual and corporate tax return data encompassing all entrepreneurs in Finland, capturing both firm and owner-level outcomes along with key demographic and economic characteristics. To enrich our analysis, we also incorporate survey data collected by Statistics Finland. We can link the survey data with the administrative data at the individual-level.

SI Contribution Data. YEL contribution-level data are obtained from the two largest Finnish pension companies, which together cover approximately 70% of all entrepreneurs

[11] According to a report by the Finnish Financial Supervisory Authority (FSA, 2010), approximately 27,000 entrepreneurs out of over 200,000 were fund members in 2010. Survey data from Statistics Finland further confirms that only about 10% of surveyed entrepreneurs hold such memberships.

insured under the YEL scheme. The data span the years 2006–2014 and include information on the reported income base used to calculate social insurance coverage and benefits, the amount of annual social insurance contributions, and relevant contract dates.

Benefit Claims Data. We observe all earnings-related benefits claimed by TyEL- and YEL-covered business owners. The data provide the annual amount of benefits received (in euros). We combine these data with information on individuals' SI contributions and benefit formulas to infer the number of days claimed. Retirement dates are identified from pension benefit claims.

Our analysis focuses on the following claims measures: for sick leave, we record the number of sick days claimed each year for TyEL owners, and for YEL owners we record $\max(0, \text{sick days} - 10)$ to account for the shorter waiting period under the YEL scheme, as discussed earlier. For parental leave benefits, we record the number of days claimed per year for both TyEL and YEL owners. For pension benefits, we record whether an individual retired in a given year and measure the annual pension income of retired individuals.

Tax Return Data. We use comprehensive tax return data from the Finnish Tax Administration covering the entire population of business owners and their firms over the period 2006–2016. This data set contains detailed information on the financial statements and tax records of all Finnish businesses, as well as granular data on individual total income and its composition. Crucially, the dataset allows us to observe SI contributions made by business owners under the TyEL pension scheme.

Survey Data. We complement our administrative data with information from two distinct surveys conducted by Statistics Finland. The first survey, *Itsensätyöllistäjät Suomessa, (2013)*, was targeted at entrepreneurs and the self-employed, and yielded 1573 responses. It collected data on working conditions, perceived business potential, income levels, and experiences with the social insurance system for entrepreneurs. The second survey for business

owners, *Yrittäjyys Suomessa (2017)*, included 1072 respondents, with a 13.4% response rate; 94.8% of whom were covered under the YEL scheme. This survey gathered detailed information on business activities, health and well-being, absences due to family reasons or illness, awareness and attitudes toward the social insurance system, and retirement planning.

In this paper, we focus on survey responses related to pensions, social insurance, and subjective well-being. We have a unique opportunity to link survey data with administrative records using individual-level identifiers, enabling us to connect entrepreneurs' perceptions of the social insurance system with their actual coverage decisions. This integrated approach offers valuable insights into the mechanisms underlying the patterns observed in the administrative data.

Sample Restrictions. In our baseline analysis, we focus on owners of privately held corporations, excluding sole proprietors and owners of partnership firms. This restriction enables a more comparable assessment of outcomes and choices between TyEL and YEL owners. By definition, all TyEL owners are incorporated, and thus incorporated TyEL and YEL owners are presumably more similar in terms of business activities than sole proprietors or partnership owners, who often operate single-person enterprises. Throughout the paper, we use the terms business owner, YEL/TyEL owner, entrepreneur, and firm owner interchangeably to refer to owners of privately held corporations. Focusing on incorporated business owners also allows us to examine SI coverage choices and their implications among individuals more likely to pursue growth-oriented and higher-risk business strategies that involve greater cognitive demands, as opposed to self-employed individuals who more commonly engage in routine, non-cognitive tasks (see, e.g., Levine and Rubinstein 2017; Lazear 2004; Evans and Leighton 1989). However, due to the relatively small sample sizes in the survey data, we include all types of business owners (privately held corporations, partnership firms, and sole proprietors) when analyzing the relationship between the survey evidence and SI coverage choices.

Second, our baseline sample comprises all YEL-insured entrepreneurs whose total income (labor and capital income) falls between 1.2 times the regulated minimum YEL contribution and the YEL maximum contribution divided by 1.2. We use this sample to examine both the overall choice of SI coverage levels and the relationship between claims and coverage. These income restrictions ensure that each entrepreneur has scope to either decrease or increase their contributions by at least 20% relative to their true total income. In addition, we exclude individuals who own multiple firms from the sample. However, our results are robust to including them, as only 7.3% of business owners in our data simultaneously own more than one firm.

3.2 Descriptive Statistics and Social Insurance Choices

Descriptive Statistics. Table 1 reports descriptive statistics covering the years 2008–2010 – the years preceding the YEL reform of 2011. Column (1) presents the statistics for our baseline sample including all business owners (both YEL and TyEL). Column (2) includes the owners of privately held corporations who transitioned from TyEL to YEL system following the 2011 reform (labeled as Treated in the table), and columns (3) and (4) for those who were YEL owners both before and after 2011 (Always Unrestricted) and those who were TyEL owners before and after the reform (Always Restricted), respectively.

From Table 1 we can observe that the key characteristics, such as labor income, total income, age, education and industry distribution, are, on average, similar across all of the groups. Variables that most clearly differ between the groups are the social insurance contributions and contributions relative to total income, denoted as SI/TI: the always unrestricted owners pay significantly lower social insurance contributions than the other groups. This suggests that a stricter social insurance mandate is associated with clearly higher contribution levels among always-restricted owners and the treated group prior to the reform, compared with the self-reported social insurance income and contributions of the always-unrestricted owners. Additionally, the always unrestricted owners are, on average, more

likely to be men, older and have more sick leave days than other groups, which we take into account in our empirical analysis. Overall, however, the groups appear to be otherwise quite similar, especially in terms of income levels and industry distribution.

Social Insurance Choices. Figure 1(a) illustrates differences in social insurance choices between restricted (TyEL) and unrestricted (YEL) business owners in more detail. The figure shows the average SI income in each 2.5% bin of labor income for both of the groups. For restricted owners, the SI choice closely follows the 45-degree line with labor income, as can be expected based on the SI regulations for this. In contrast, unrestricted owners with higher incomes tend to select substantially lower levels of SI coverage, on average, compared to restricted owners. This highlights the underlying heterogeneity in SI decisions and shows that unrestricted owners insure themselves to a much lesser extent through the public SI insurance scheme.[12]

Figure 1(b) complements this analysis by combining survey responses with administrative register data to examine the motivations behind SI contribution choices among unrestricted owners. This analysis aims at addresses the key questions regarding the drivers of observed SI contribution choices. In the survey, the respondents were asked about their motivations behind their low SI coverage choices (CITE THE SURVEY HERE). Figure 1(b) illustrates the shares of different categories the respondents selected (left-hand side y-axis). Note that in the survey respondents could choose several contributing factors, so the shares of the answers do not sum up to 1.

The survey results reveal a range of contributing factors behind low SI coverage, and no single reason clearly stands out as a most significant explanation. In particular,

[12] Appendix A Figure A.1 presents the distribution of annual reported SI income for unrestricted owners and indicates a clear preference for round numbers, with noticeably higher mass at bins that are multiples of 10 and 24 (i.e. 12 months $\times$ 2) compared to adjacent bins. In addition, Appendix A Figure A.2 presents the share of reported social insurance income within the total income distribution, disaggregated by individual characteristics. The figure shows that heterogeneity in SI income is small with respect to gender (panel a) and parental status (panel d). In contrast, Figure A.2 shows that the share of social insurance income is higher among older owners compared to younger ones (panel b) and among married business owners compared to their unmarried counterparts (panel c).

plans to work during retirement (65% share) or selling the business (40%), private pensions crowding out SI contributions (43%), liquidity constraints (58%), and the lack of trust in the SI system (43%) stand out as the most commonly stated reasons for low SI contributions. In contrast, delays in updating the self-reported SI income (17%) or simultaneous TyEL insurance coverage (5%) were not common explanations for low SI coverage in the survey.

Figure 1(b) includes the average SI income relative to total income (SI/TI) with 95% confidence intervals within each answer category (right-hand side y-axis), calculated from the administrative data. Interestingly, despite the widely differing self-reported explanations for low SI coverage, actual SI choices observed in the administrative data are strikingly similar within each category, even when controlling for observable characteristics (age, gender and total income). Therefore, it seems that none of the stated reasons for low coverage in the survey are clearly associated with actual coverage choices observed in the data.

Furthermore, many of the stated motivations for low SI coverage in the survey are largely unrelated to the specific risks that SI is designed to insure against, such as old-age, disability and sickness. This observation raises a central question: to what extent are observed SI choices shaped by adverse selection or moral hazard? Next, we investigate these mechanisms in greater depth.

4 Social Insurance Coverage and Claims

4.1 Positive Correlation Tests

We implement positive correlation tests (Chiappori and Salanie, 2000b) to study the relationship between individuals' SI coverage choice and SI benefit claims, revealing how risks are associated with the chosen coverage levels. This analysis conflates adverse selection responses with moral hazard responses, since both lead to a positive relationship between one's SI coverage choice and claims. Therefore, this analysis illustrates the general association be-

tween risks and SI coverage without taking a stand on the role of underlying mechanisms explaining the correlations.

We conduct this analysis for all unrestricted owners (YEL) in our baseline sample, and present the results in Figure 2. In the figure, individuals are grouped into 5% bins based on their SI income relative to their total income (SI/TI_i^t), treating each person i in year t as a separate observation. We then calculate the average claims in each bin to evaluate how SI coverage choices are associated with various outcomes: average retirement age (panel a), life expectancy (b)[13], and sick leave days (c) and parental leave days (d) per year. All these analyses include controls for age, total income, gender, municipality and industry.

Figure 2 shows that the estimated slopes between SI choices and claims align with expectations on adverse selection or moral hazard behavior. Figure 2(a) illustrates that individuals with higher SI/TI_i^t are more likely to retire younger, as there is a clear downward slope between mean retirement age and SI contributions. Similarly, Figure 2(b) shows that those who contribute more tend to have longer predicted life expectancy, meaning that those who are more likely to collect pension benefits for a longer time also have, on average, higher SI contributions. Therefore, both of these measures imply that those with a higher SI contribution rate end up collecting more pensions benefits compared to those with smaller SI/TI_i^t.

Figures 2(c) and 2(d) show that higher SI/TI_i^t is associated with more sick leave and parental leave days, respectively. It also appears that the correlation between claims and SI coverage is stronger for sick leave days than for parental leave. For parental leave, the positive correlation is mainly driven by those who have contribution rates close to 1, but the the correlation is much smaller for those with SI/TI_i^t between 0.3–1.

However, the magnitudes of these correlations, collected in Panel A of Table 2 and illustrated in Figure 2, are economically small. We can intuitively illustrate the magnitudes by comparing the benefit levels along the horizontal axis of Figure 2. Based on the estimated

[13]Appendix B describes in detail how we construct the measure of predicted life expectancy using our register data.

slope in Figure 2(a), moving from a low coverage choice of SI/TI_i^t=0.2 to a high coverage of SI/TI_i^t=1.2 is associated with retiring approximately 4 months earlier, which, all other things equal, translate into collecting pension benefits for 0.33 years longer. Similarly, in Figure 2(b) those with SI/TI_i^t=1.2 claim pensions for 10 months longer due to higher their higher life expectancy compared to those with SI/TI_i^t=0.2. In Figures 2(c)–(d), increasing SI/TI_i^t from 0.2 to 1.2 corresponds to a 1.13 and 0.67 increase in annual sick and parental leave pay days, respectively.

To illustrate the magnitude of the costs stemming from this risk-based selection in SI coverage choices, we relate the monetary increase in benefits claimed when moving from a low SI coverage choice to high SI coverage to the consequent increase in SI contributions observed in the data. We collect the numbers for this analysis in Panel B of Table 2. A natural benchmark for this evaluation is the fully mandated SI system, where the SI choices are determined by total income and there is, by definition, no positive correlation between SI/TI_i^t and claims.[14]

First, to evaluate the selection costs for pension benefits, we combine our findings in Figures 2(a) and 2(b): moving from SI/TI_i^t=0.2 to SI/TI_i^t=1.2 implies in total 14 months more pension benefits, stemming from earlier retirement and living longer for those with a high coverage rate compared to those with low SI/TI_i^t. From Panel B of Table 2, we observe that annual SI contributions increase between those with SI/TI_i^t=0.2 and those with SI/TI_i^t=1.2 by €11,409 on average, and the consequent self-reported SI incomes by €47,537. This means that those with higher coverage are also eligible for higher pension benefits. Using this increase in SI coverage and benefit entitlements together with the pension benefit formula briefly discussed in Section 2, we find that benefits increase by approximately €827 annually, due to the 14 months (1.16 years) additional pension benefits for those with higher SI income.[15] Relating this number to the increased SI contributions gives $827/11,409 =$

[14]For simplicity, this benchmark concept rules out the possibility that TI is affected by benefit replacement rates.

[15]Notes in Table 2 present the details of the benefit calculations.

0.072. This number indicates that for pension benefits the cost of selection by risks is 7.2 cents per additional euro of SI contributions.[16]

Using the same approach for sick leave and parental leave give us estimates that are even smaller: 0.7 and 0.6 cents of additional benefits per one euro of SI contributions, respectively. These small numbers stem from the fact that increasing the benefit days by 1.1 (sick leave) and 0.7 (parental leave) per year when moving from SI/TI_i^t=0.2 to SI/TI_i^t=1.2 has only a small effect on annual benefits (€83 and 76), while the increase in SI contributions is the same as above (€11,409). Therefore, the costs of risk-based selection for sick leave and parental leave are an order of magnitude smaller than with pension benefits.

Intuitively, the small selection cost estimates relative to SI contributions are driven by the small correlation between claims and SI choices. This means that benefit usage does not increase very much along with self-reported SI income and contributions, as illustrated in Figure 2. In contrast, the observed increase in SI contributions along with the coverage rate does not drive these estimates. While larger SI income relative to total income does increase the denominator when evaluating benefit costs per SI contributions, it also affects the benefit levels in a similar manner, as larger SI income consequently leads to larger benefit levels, and vice versa. Therefore, assuming larger or smaller values for the increase in SI contributions along the horizontal axis in Figure 2 does not significantly affect the estimates or their implications.

A well-known approach to estimating a lower bound on the extent of adverse selection is to measure the positive correlations between past benefit claims and future insurance choices (see e.g. Chetty and Finkelstein 2013). The use of past SI claims intuitively shuts down the moral hazard channel by sorting individuals on ex-post immutable risks, and therefore reveals to what extent plausibly higher-risk individuals choose higher SI coverage rates. We conduct this test in Figure A.6 in Appendix A, which plots the average claims

[16] Assuming that individuals would increase their SI income equivalently for more than a year does not change this calculation. Increasing the number of years increases both pension benefits and SI contributions with the same multiplier (number of years), which cancel each other out when relating the benefit costs with SI contributions in the above formula.

in 2007–2010 by bins of SI/TI choices in 2011–2014 for all benefit claims in panel (a) and separately for sick leave (b) and parental leave days (c). The figure shows practically no positive relationship between past claims and future SI coverage choices. The weak correlation suggests that no significant adverse selection by realized risks in the past. However, it is important to note that this analysis may not reveal the full extent of adverse selection because risk types may change over time, which would make past claims an imperfect proxy for the current risk type. This is particularly likely in our setting where claims do not exhibit a high degree of correlation over time, as discussed in more detail below in Section 4.2.

In sum, our estimates imply that the costs of selection by risks are small in this setting. This gives us first evidence that the combined effects of adverse selection and moral hazard are likely to be limited among business owners, despite their large discretion over social insurance coverage. Next, we examine intertemporal responses in SI choices to future risks to further illustrate the potential scope of risk-based selection. In Section 5, we attempt to isolate the effect of moral hazard responses using the variation generated by the 2011 reform.

4.2 Anticipation of Future Claims

Another approach to assess the association between SI claims and coverage is to study whether unrestricted owners adjust their SI income in anticipation of predictable claim events, such as sickness, parenthood and retirement. Similarly as above in Figure 2, potential anticipation of future claims includes both adverse selection and moral hazard responses, reflecting the general correlation between claims and SI contributions.

To study the significance of anticipatory responses, we estimate event-study regressions of the following form for all unrestricted owners:

$$SI/TI_i^t = \alpha + \beta_{-2} \cdot t_{-2} + \beta_{-1} \cdot t_{-1} + \beta_0 \cdot t_0 + \beta_1 \cdot t_1 + \beta_2 \cdot t_2 + X_{it} + \varepsilon_{it}, \tag{1}$$

where the coefficients β_j, $j = -2, ..., 2$ reflect the development of insurance contributions relative to total income (SI/TI_i^t) before and after a claiming event. We estimate this specification for sick leave and parental leave, where t_0 is defined as the year we observe an individual receiving sick leave or parental leave benefits. Also, we also estimate the model for retirement decisions using data from up to seven years before retirement. However, once an individual retires, they typically stop making SI contributions, and thus post-event indicators are not available for this outcome. In the estimations we include as controls age, gender, municipality, industry and year fixed effects.

For retirement (Figure 3a), SI income relative to total income remains approximately at the same level the full seven year period before starting to claim pension benefits. This implies that the unrestricted business owners do not increase their reported SI income in the final years before retirement to increase their pension benefits (conditional on age). This pattern is feasible given that all SI contributions during the working career affect future pension benefits with the rate of 1.5%. Therefore, increasing SI income in the years preceding pension claiming has only a small effect on the forthcoming benefit. However, the increased accrual rate of 4.5% for those who continued working after the age of 63 years could in theory lead to increased SI contributions for those who did not claim their pension at this minimum retirement age. Nevertheless, Figure 3(a) shows that this exception had no visible effect on overall SI contributions before starting to claim pension benefits.

The results for sick leave and parental leave benefits are presented in panels (b) and (c) of Figure 3, where the development of SI/TI_i^t is presented relative to year $t-3$ - three years before we observe an individual claiming a benefit. Overall, we find minimal evidence of anticipatory behavior for these benefits. The estimated coefficients are small in magnitude and not statistically different from each other over time before the claiming events.

More specifically, for sick leave and parental leave we observe that SI/TI_i^t remained at the same level for the two years preceding a claiming event. However, at year

t_0 there is a small and statistically insignificant increase in SI income of approximately 1-2 percentage-points, suggesting a minimal average anticipation response. Note that in contrast to pension benefits, SI income from the 12 months prior to claiming sick leave or parental leave define the benefit level, which in principal could make anticipatory responses more relevant for these benefits. However, Figure 3 illustrates that these responses do not appear to be relevant.[17]

A feasible explanation for the lack of significant anticipatory responses for sick leave and parental leave is that the short-run increase in SI contributions need to be sufficiently large for it to have a notable impact on the benefit levels. For example, doubling the average annual SI contributions of the always unrestricted owners (Table 1) from €5580 to €11,160 would increase the sick leave and parental leave benefits by approximately only €25 and €69 per working day, respectively. Given the timely restricted eligibility to parental leave (see Section 2) and the likely uncertainty regarding the length of a sick leave period, these relatively small and uncertain incentives can at least partly explain the lack of anticipation of SI/TI_i^t to future claims.

Overall, Figure 3 shows that individuals do not, on average, "game" the SI system by strategically increasing their SI income in years preceding benefit claims. This holds even when the pensions system included increased accrual rates for those who retained from claiming pension benefits beyond the minimum retirement age of 63. These findings reinforce the earlier conclusion from Figure 2 that risk-based selection is limited in this context, and the combined effect of adverse selection and moral hazard are likely to be small.

[17]Figure A.5 in Appendix A provides additional evidence on the lack of anticipatory responses by documenting the prevalence of active changes in SI income in the years around claiming events. The figure shows that active changes in SI contributions did not increase before claiming benefits, thus showing no signs of significant anticipatory responses.

5 Responses to the 2011 Reform

The standard positive correlation tests and anticipatory response estimates documented in Section 4 conflate adverse selection (AS) and moral hazard (MH), as both of these mechanisms predict a positive association between insurance coverage and claims. In our context, the absence of strong positive correlations between SI coverage and benefit usage shows that the combined effects of AS and MH are likely to be small. However, this aggregate pattern could also be consistent with offsetting forces. Specifically, the presence of substantial advantageous selection, where individuals with lower risks insure themselves more than those with higher risks, could mask the presence of underlying moral hazard effects. Therefore, we cannot fully rule out potential moral hazard responses based on the evidence presented in Section 4.

To address this issue, we utilize the SI reform in 2011 and introduce a framework that directly estimates the extent of moral hazard. This allows us to disentangle behavioral responses to benefit levels from selection effects, and therefore assess more precisely the role of different mechanism in shaping SI choices and benefit claims.

5.1 Methods and Identification

We use an event-study analysis exploiting the 2011 reform to estimate how SI income and benefit claims change for treated business owners relative to the control group. Specifically, we estimate the following equation:

$$Outcome_{it} = \sum_{\substack{\ell=2006 \\ \ell \neq 2010}}^{2014} \beta_\ell \; Treat_i \; 1_{t=\ell} + Treat_i + \eta_t + \delta X_{it} + \varepsilon_{it}, \tag{2}$$

where the subscript i denotes individual business owners and t denotes calendar year. The term δ_t captures year fixed effects, and X_{it} includes a set of individual-level controls (age, gender, municipality, and industry classification of the firm). The coefficients of interest, β_ℓ, measure the differential outcomes for treated individuals relative to the control group in

each event year ℓ. We estimate the model from 2007–2014 and omit the year 2010 from the regression, making it the reference period; all estimated effects are thus interpreted relative to the year preceding the 2011 reform.

The treatment group consists of business owners who hold more than 30% but less than 50% of company shares prior to the 2011 reform. These individual faced a change from the restricted TyEL scheme to the unrestricted YEL system in the reform. We compare their outcomes to two alternative control groups: The first control group – referred to as the *always restricted* – includes individuals owning more than 10% but no more than 30% of company shares. These individuals are continuously enrolled in the TyEL scheme, which does not permit discretion over SI contribution levels during the study period. The second control group – referred to as the *always unrestricted* – consists of owners with 50–70% ownership share, who retain full discretion over SI contributions throughout the period.

Our baseline analysis uses the *always restricted* group as the primary control group. This group remains untreated throughout the sample period and thus plausibly satisfies the identifying assumptions of a standard difference-in-differences (DiD) framework (Tazhitdinova and Vazquez-Bare, 2023). As a robustness check, we replicate the analysis using the *always unrestricted* as the control group, and find that the results are qualitatively very similar as when using the always unrestricted as the control group. Importantly, our event-study design is not subject to the concerns raised in recent critiques of two-way fixed effects estimators (de Chaisemartin and D'Haultfoeuille, 2020; Sun and Abraham, 2021; Callaway et al., 2021; Goodman-Bacon, 2021), since the treatment is assigned uniformly in the same calendar year for all treated units.

Our key identifying assumption is that the treatment and control groups would have followed similar trends in the absence of the reform – commonly known as the parallel trends assumption. We assess the validity of this assumption by comparing the evolution of our main outcome variables for the treatment and control groups prior to the reform. We show support for this assumption for both SI income and benefits claims, reported below in

Figures 4 and 6.[18]

A further potential threat to identification is that ownership shares could have been adjusted in response to the 2011 reform, meaning that business owners could self-select into restricted or unrestricted insurance scheme as a response to the reform. However, we find no empirical evidence of such behavior. Appendix A Figure A.3 shows the shares of the owners with changes in ownership from less than 30% to more than 30%, and from more than 50% ownership to less than 50% in 2007–2015. Overall, such changes are rare – with less than 3% of all owners changing ownership shares across these thresholds – and more importantly, there are no significant changes in these shares around the 2011 reform. This alleviates concerns that owners manipulated their ownership in response to the reform.

5.2 Social Insurance Choice

We begin by more carefully examining the willingness to pay for SI contributions by studying how treated owners adjust their SI income when granted increased flexibility in choosing their coverage level after 2011. Figure 4(a) shows the evolution of SI income relative to total income (SI/TI) for the treatment group compared to both control groups (the always restricted and the always unrestricted). First, the figure shows that the development of SI income is very similar in all three groups before the reform (denoted by the red dashed line in the figure). This strongly supports the parallel trends assumption. Second, after the reform, the SI incomes of both control groups develop very similarly as before the reform, but the treated owners significantly reduced their SI contributions. Also, we observe that the SI income choices of the treatment group converges toward the levels of the always unrestricted

[18]This type of a setup with SI regulations varying with particular ownership shares could in principle allow us to use a regression discontinuity design (RDD) instead of a difference-in-difference approach. However, using a RDD design in this case is challenging. First, the owners are unevenly distributed across the ownership share distribution, as there is a large number of owners with certain ownership shares - such as 25, 33 and 50% for firms with 4, 3 and 2 owners - which makes it challenging to define the bandwidths in RDD estimation. Second, RDD is very data intensive, and the number of owners just around the 30% threshold is not large enough to provide very precise results. Therefore, we believe that a difference-in-differences approach - with similar pre-trends in key outcomes in the treatment and control groups - provides a more robust and accurate results in this case.

owners towards the end of our examination period.[19]

Figure 4(b) displays the corresponding event-study estimates with 95% confidence intervals using both control groups. The results indicate that SI/TI decreased on average by about 13 percentage points in the treatment group relative to both control groups. Given the pre-reform mean of 0.83 in the treatment group, this corresponds to an average drop in SI/TI of approximately 16%. Overall, this result highlights that when given more flexibility in SI choices, the treated business owners clearly reduced their SI contributions.

The average responses in Figure 4 may hide substantial variation in SI choices within the treatment group. To examine potential heterogeneity, we study the SI coverage choices across different risk measures constructed from pre-reform data. Dividing the SI income responses by pre-reform risks can also shed more light on the adverse selection channel by revealing how risks shape future SI coverage choices.

Figure 5 presents results the event-study results on SI income choices by age, life expectancy, and sick leave benefits before the reform. Panel (a) of the figure shows that there is substantial heterogeneity between age groups. We observe that younger owners (below 35 years) reduced their SI contributions by as much as 20% relative to the control group. In contrast, owners above 50 years of age reduced their SI/TI by approximately 5%, and those between 36–50 years by 11%. This result is similar to the observation in Figure A.2, which shows that SI/TI is lower across the income distribution when examining the SI choices among all unrestricted owners.

Figure 5(b) shows that there is much less variation in SI income responses by estimated life expectancy groups. Individuals are divided into three groups according to whether their estimated, residualized life expectancy falls below the 33rd percentile, between the 33rd and 66th percentiles, or above the 66th percentile. Life expectancy is residualized prior to group construction, rather than using the raw data, to account for systematic differences such as the higher life expectancy of younger individuals relative to older ones. Residualiza-

[19] Appendix Figure A.4 plots the evolution of SI income relative to *labor* income, yielding results that closely mirror our baseline approach, which uses total income as the denominator.

tion is performed by regressing life expectancy on gender and age cohort dummies, and their interactions, industry, municipality and income decile dummies before constructing these groups. Treated owners with the highest predicted life expectancy ($\geq P66$) reduced their contributions slightly less than those with lower life expectancy ($\leq P33$ and $P33 - P66$). This finding is broadly similar to the cross-sectional positive correlation results shown in Figure 2(b), where those with the highest life expectancy have higher SI contributions compared to others. Nevertheless, the differences in event-study estimates by life expectancy groups are relatively small.

Finally, Figure 5(c) demonstrates that the largest reductions occur among individuals with no history of sick day claims before the reform, whereas groups with prior sick day claims reduced their contributions slightly less.[20] This finding aligns with the positive correlations in Figure 5(c), where those with higher sick leave benefits days have, on average, higher SI contribution rates.

In sum, our findings highlight a strong and persistent response in SI coverage choices among the treatment group: when given more discretion over SI contributions and coverage, the treated owners significantly reduced their contributions. Also, the SI income of the treated owners converge very close to those observed among the always unrestricted owners, who had the freedom to choose their SI coverage both before and after the reform. These findings imply that the treated owners, given their reductions in SI contributions, are eligible to lower benefits after the reform. Next, we analyze how SI claims evolved after the 2011 reform in both the treatment and control groups, in order to assess the extent of moral hazard responses.

5.3 Measuring Moral Hazard Responses

Conceptual Background and Expected Impacts. As shown above, treated owners significantly reduced their SI contributions relative to the control groups following the 2011

[20]Sickness claim groups are formed using residualization in the same way as for life expectancy, except that zero sickness claims are based on actual observations prior to 2011.

reform. If moral hazard is present – that is, if the level of SI coverage and benefit generosity affect claiming behavior – we would expect the reduction in SI coverage and future benefit entitlements to reduce SI claims, on average, in the treatment group after the reform.

However, interpreting behavioral responses in SI claims as evidence of moral hazard in the event-study framework requires additional assumptions beyond the parallel trends assumption discussed above. First, we assume that, at the individual level, realized SI claims are determined by both the underlying claiming risks and moral hazard, with the latter being influenced by the level of SI coverage and benefit entitlements (SI/TI), but the former is unaffected by SI choices. Within our event-study framework, we further assume that this relationship applies similarly to both the treatment and control groups.

Under this simple framework, we can interpret potential changes in SI claims as evidence of moral hazard as long as we assume that underlying claiming risks unrelated to SI/TI evolve similarly in the treatment and control groups over time. In other words, the event-study estimates on SI claims capture moral hazard responses *unless* the true claiming risks of treatment and control group owners changed differently over time. This assumption thus allows for underlying risks to fluctuate over time, for example, due to inherent volatility, or institutional and demographic reasons, but requires that such changes are, on average, similar across these groups. We consider this assumption to be plausible, at least conditional on observable characteristics such as age, gender and industry. However, this assumption cannot be empirically tested, since we only observe realized SI claims – which combine both underlying risks and moral hazard – rather than the two components separately.

As shown in Figure 4(b), the treatment group's SI income relative to total income declined by approximately 16% on average. This aggregate response, however, hides notable heterogeneity in SI choices as highlighted above and documented above. In addition to the heterogeneity presented in Figure 5, we observe that while some business owners reduced their SI contributions after the 2011 reform, others increased them. If moral hazard responses are significant, we would expect those who reduced their SI/TI to also reduce their SI claims

after the reform, whereas those who increased their SI contributions should increase their claims after 2011. We use this variation in SI choices to further evaluate the significance of potential moral hazard responses.

Claims by SI choices. Our event-study analysis on SI claims in Figure 6 first shows the average number of total days claimed in panel (a), and then splits the sample of treated individuals into three subgroups in panels (b) – (e) based on their observed changes in SI income after the reform: those with SI/TI below 0.8, those with SI/TI between 0.8 and 1.1, and those with SI/TI above 1.1 after the reform. In this analysis, the control group consists of the always-unrestricted owners (ownership share 10–30%), and we control for age, gender, municipality and industry in the estimations.

First, Figure 6 shows that the pre-trends of SI claims are parallel across all five panels – for total average claiming days (panels a and b), pension benefits (panel c), sick leave (panel d) and parental leave day claims (panel e) – as well as across all subgroup comparisons. This illustrates that SI benefit claiming developed very similarly prior to the reform in the treatment and control groups, supporting our baseline event-study assumption.

The results in Figure 6(a) show that total realized SI claims including all benefit types did not differ significantly across groups in the post-reform period, despite the significant average drop in SI/TI in the treatment group. This provides prima facie evidence that the reduction in SI contributions was not followed by a drop in SI claims within the treatment group (on average), which is consistent with no significant moral hazard effects. In addition, when splitting the treatment group by the observed SI/TI responses and analyzing each benefit type separately in Figure 6(b)–(e), we observe no systematic increases or decreases in SI claims for any of the subgroups. This suggests that despite the heterogeneous responses in SI income within the treatment group, there is no evidence of moral hazard responses in SI claims. Table 3 shows the corresponding DiD estimates presented in Figure 6.

This pattern indicates the absence of meaningful moral hazard behavior: despite

substantial variation in SI contributions, there is no corresponding variation in benefit usage. In other words, lower coverage choices do not translate into fewer claims, nor do higher contributions lead to substantially larger claims. This finding indicates that behavioral responses to the reform primarily reflect selection in SI choices rather than moral hazard behavior.

6 Other Potential Mechanisms behind SI Choices

Taken together, our findings above suggest that adverse selection responses are limited and hazard responses are virtually absent among entrepreneurs who have substantial discretion in determining their level of SI coverage and contributions. This implies that individual risks and benefit claims do not constitute a central determinant of SI coverage decisions, even though many entrepreneurs select a relatively low level of insurance coverage. In the following analysis, we examine potential mechanisms that may account for these observed patterns. We consider factors beyond pure risk-based motives — such as informational limitations and lack of knowledge, learning, preferences, and other insurance and savings options — that may shape entrepreneurs' SI decisions and outcomes.

Knowledge and Learning. A key factor influencing SI choices is information. Social insurance systems are typically complex, and individuals may lack sufficient understanding of how the system functions and how their choices affect future claims. We exploit survey evidence to better understand how well business owners know and comprehend the SI system and its benefits. By linking the survey findings to administrative data, we can characterize how self-evaluated knowledge about social insurance is associated with actual SI coverage choices.

Figure 7 presents survey-based evidence on entrepreneurs' knowledge and awareness of the SI system. Panel (a) of Figure 7 illustrates the distribution of responses to the question "Do you know what benefits you are entitled to?", ranging from very well [1] to

not at all [4]. The left-hand side axis displays the distribution of responses illustrated by the shaded bars in the figure, and the right-hand side plots the corresponding average of SI/TI for each response group with 95% confidence intervals, and both with and without controls.

The results indicate that knowledge of the system is concentrated in the middle categories (2 and 3), with lower and relatively similar shares observed at both ends of the distribution (1 and 4). In particular, below 15% of the respondents stated that they know very well the benefits they are entitled to (1). Therefore, limited understanding of SI benefits may help to explain the weak observed association between individual risks and SI claims documented above. However, Figure 7(a) reveals no clear association between respondents' self-reported knowledge about SI and their actual SI coverage choices. Specifically, individuals who report a higher degree of knowledge are not significantly more likely to choose higher SI contributions, and vice versa. This suggests that although limited information may partly explain low SI coverage rates and the lack of adverse selection and moral hazard, it does not appear to be the primary driver of the observed variation in SI contributions.

Panel (b) of Figure 7 focuses on awareness of sick-pay replacement rates, based on a survey question that specifically addressed sick leave benefits: "Are you aware of the amount of sick leave you would receive, if needed?", with pre-defined answer categories "Yes" or "No". A clear majority of respondents (over 70%) report that they are unaware of the amount of sick leave, while more than 25% state that they are aware of the benefit size. This again indicates that a large share of business owners are not particularly well aware of the details of SI benefits. However, despite the differences in self-reported awareness, differences in the actual SI contribution rate (SI/TI) observed in the data across the two groups are very small, indicating that the awareness of benefit regulations is not associated with SI contribution choices. Nevertheless, those who are aware of the sick leave benefit replacement rate contribute slightly more on average than those who are not, but this difference is not statistically significant.

Individuals' own risk perceptions may also diverge from their actual risks, meaning

they are not necessarily aware of their true risk exposure when making SI contribution choices. To explore this channel, we combine survey-based measures of perceived health status with actual SI choices among entrepreneurs in panel (c) of Figure 7. The figure shows the distribution of perceived health status (ranging from 0–10) in the survey alongside with the average SI/TI chosen within each response category.

From Figure 7(c) we observe that the respondents report that they are in relatively good health: fewer than 10% rate their health at 6 or lower, while more than 50% report their health status to be 9 or 10. However, there is no clear correlation between the perceived health status and SI coverage rates, as the average SI/TI is very similar within each health category. This indicates that SI coverage decisions are, on average, similar across individuals with different self-assessed health, which is consistent with our earlier findings of weak correlations between observed sick leave claims and SI coverage choices. Therefore, it seems that the implications of our findings are similar regardless of using a measure of perceived conceptions about health-based risks or the observed SI claims in the administrative data.

In addition to the survey evidence, we use administrative data to examine potential learning effects. Despite the small average correlation between claims and SI coverage choices observed above, it could be that individuals learn about their risks over time, leading to larger (positive) correlations between risks and SI coverage choices. We analyze learning mechanisms and their significance in Figure 8 by studying whether newly treated business owners – those with discretion over their SI income after the 2011 reform – exhibit an increasingly positive relationship between sick-pay benefit claims and SI contributions over time. To do this, we compare the correlations between claims and coverage choices immediately after the reform (2011) with that observed three years later (2014). In this analysis we focus on sick leave benefits, as they are likely to be the most common benefit an individual would apply for during their working life within a fixed three-year window (2011–2014), compared to old-age pensions and parental leave benefits.

Figure 8 demonstrate that the positive correlation between claims and coverage choices (SI/TI) increased among treated individuals over time, suggesting that some gradual learning about risks appears to occur. However, while the correlation slope estimate increases from 1.9 to 3.2 over these years, the pattern also becomes much more non-linear over time. On average, the claimed sick leave benefits increased the most for both those with highest and lowest SI contribution rates, while the claims for those close to the average SI/TI increased much less.

In addition, it is relevant to note that the average increase in the correlation slope over time in Figure 8 would imply a 1.3-day increase in sick leave benefits when comparing those with low (SI/TI_i^t=0.2) and high SI contributions (SI/TI_i^t=1.2), following a similar approach as above in Section 4.1. In terms of sick leave benefits, this would imply an increase in annual SI benefits of less than 100€. This can be considered economically very small when compared to, for example, the associated average increase in annual SI contributions of 11,400€when comparing those with low contributions to those with a high contribution rate, or to the mean SI contributions among the treated business owners (6900€). Therefore, learning about risks over time is likely not to induce a significant impact in monetary terms, at least in a shorter run. Also, the average correlation estimate between SI contributions and sick day benefit claims among all unrestricted owners in Figure 8 (b) is relatively modest (1.13), as discussed in Section 4.1 in more detail. This estimate includes owners with many years of experience with the SI system, indicating that the overall scope of potential learning effects in increasing average risk-based selection over time is likely to be limited.

Retirement Plans. As discussed above, pension benefits and retirement decisions are presumably the most transparent and important parts of the social insurance bundle. Also, our evidence above suggests that retirement benefits (retirement age and life expectancy) have the highest correlation with SI contribution choices (see Section 4.1). An additional consideration regarding pension benefits is whether individuals plan to continue working

after retirement. Among entrepreneurs, working beyond the statutory retirement age is relatively common, as they have considerable flexibility over this decision within their own firms compared to regular wage earners. This may also influence their SI decisions: one can expect that business owners who plan to work beyond retirement age would contribute less during their active working years, since their need for pension benefits is lower when their income is supplemented by entrepreneurial earnings.

In Figure 9, we combine survey responses with registry data on SI/TI choices to assess this hypothesis. The figure displays the distribution of answers to the survey question "Do you plan to continue working after you retire?". The answer categories include yes or no, separated by preferences, financial reasons and business-related possibilities for working after retirement, and preferences or health reasons for not working beyond retirement. The survey results (right-hand side axis of the figure) show that approximately 61% of the respondents stated that they plan to work after retirement, mostly stemming for their preference to work after retirement (38%) and to smaller extent financial reasons (6%) or conditional on other possibilities to continue to run a business past the retirement age (17%). Approximately 39% of respondents stated that they do not plan to work after retirement, mostly because of preferences on retirement (35%) and to smaller extent on health reasons (4%).

The results on SI coverage in Figure 9 show that there is a small correlation between self-reported plans on working after retirement and the chosen SI coverage (right-hand side axis). We observe that SI/TI is, on average, slightly larger among those who plan not to work after retirement, particularly among those who stated that they do not prefer to work after retirement because of heath reasons. In contrast, SI coverage is lower among those who plan to work after retirement, especially among those who plan to work because of financial reasons.

Overall, this association between retirement plans and SI contributions is in line with our findings on the (small) positive correlation between retirement age and SI contributions (see Figure 2), indicating that earlier retirement is associated with higher SI contribu-

tions and retirement benefits. The survey evidence further reveals that this association can be linked with preferences for retirement, as those who prefer not to work after retirement have higher SI contributions compared to others.

Crowd-out and Capital Income. A limitation of our analysis – shared with much of the prior literature on insurance systems, moral hazard and adverse selection – is the inability to observe individuals' complete insurance portfolios, including private insurance policies and precautionary savings against risks such as health shocks and old age. This means that we cannot accurately observe the extent of which business owners complement their mandatory SI contributions with other types of savings or insurance. We can, however, observe taxable returns to various types of savings instruments through capital income flows from individual tax records. To characterize the relevance of the role of other types of savings, we use the 2011 reform to test whether these income sources were increased among treated business owners after they sharply reduced their SI income after the reform.

Figure 10 illustrates the development of personal capital income in the treatment group relative to the two control groups around the 2011 reform, using a similar framework as in Section 5. The figure shows that capital income developed similarly in the treatment and control groups before 2011, but average capital income among the treated business owners gradually increased after the reform compared to the control groups. This suggests that at least part of the observed average reduction in SI contributions after 2011 (Figure 4) was crowded-out by increased savings in other forms. This finding aligns with an earlier study by Benzarti et al. (2020), who use the same reform and data to analyze firm-level outcomes and report that firm-level savings increased in response to the reform among older firms.

Nevertheless, while complementing reduced SI contributions by other forms of savings could at least partly explain the low average levels of SI coverage among unrestricted owners, it does not affect the implications of our findings above. As SI coverage choices are not significantly driven by either adverse selection or moral hazard, the efficiency costs of

the SI system in terms of risk-based selection remain small regardless of potential crowd-out effects. However, these results suggest that the large discretion over mandatory SI contributions gives the unrestricted business owners an opportunity to select their desired combination of various savings and insurance instruments, including the SI contributions. Our evidence indicates that they appear to utilize this opportunity when their statutory mandates for SI contributions are relaxed.

7 Conclusions

This paper examines the social insurance choices of business owners in Finland, where part of owners of privately held businesses have a large discretion over their mandatory social insurance (SI) contributions and coverage. We find that freedom to choose one's SI contributions leads to a significantly lower SI payments and coverage, compared to business owners with a stricter SI mandate. However, we find that SI choices are only weakly associated with claiming behavior (retirement, sick leave and parental leave), indicating that adverse selection and moral hazard are limited in this setting.

We exploit variation from the 2011 SI reform that granted discretion over SI contributions to a larger group of business owners. Using this quasi-experimental setting, we can separately estimate business owners' moral hazard responses. Our results clearly indicate that the owners who were given a greater discretion over SI contributions systematically select lower coverage levels than comparable individuals who are subject to stricter SI mandates. The event-study analysis around the reform shows that moral hazard responses are negligible: we find that differential changes in SI coverage and benefit entitlements at the time of the reform do not translate into differences in benefit claims in the future.

We complement our register data analysis with survey evidence to shed light on other potential mechanisms explaining SI choices. The survey evidence indicates that various factors such as expectations of working beyond retirement, liquidity constraints and reliance

on private savings are common reasons for why business owners choose low SI contributions. Relatedly, we observe increased personal capital income flows following the reductions in SI contributions after the 2011 reform among treated business owners, which is consistent with substitution toward private savings. Also, we find that lack of information and knowledge about the SI benefit system is common among business owners, but linking survey results to administrative data reveals that differences in knowledge about social insurance benefits is not associated with actual SI choices. In addition, we find that self-assessed health status is unrelated to SI coverage choices, which is consistent with our observation of no significant adverse selection between coverage choices and claims.

Overall, our results provide novel evidence suggesting that both adverse selection or moral hazard responses are very limited in a setting where business owners have large discretion over their mandated SI contributions. This implies that freedom to choose SI coverage does not generate notable selection or efficiency costs. Therefore, these findings suggests that a SI system for business owners can include some degree of freedom in choosing a preferred level of SI coverage without significant efficiency costs. Nevertheless, in a welfare state environment with pre-determined minimum social benefits – as in the Finnish case – the minimum SI contributions need to be large enough to ensure the financial sustainability of the social insurance system.

References

Audoly, R. (2022): "Self-employment and Labor Market Risks," Mimeo.

Benzarti, Y., J. Harju, and T. Matikka (2020): "Does Mandating Social Insurance Affect Entrepreneurial Activity?" *American Economic Review: Insights*, 2, 255–268.

Boeri, T., G. Giupponi, A. B. Krueger, and S. Machin (2020): "Solo Self-Employment and Alternative Work Arrangements: A Cross-Country Perspective on the Changing Composition of Jobs," *Journal of Economic Perspectives*, 34, 170–195.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2021): "Difference-in-Differences with a Continuous Treatment," arXiv Working Paper 2107.02637.

Chetty, R. and A. Finkelstein (2013): "Chapter 3 - Social Insurance: Connecting Theory to Data," in *Handbook of public economics, vol. 5*, ed. by A. J. Auerbach, R. Chetty, M. Feldstein, and E. Saez, Elsevier, vol. 5 of *Handbook of Public Economics*, 111–193.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016): "The association between income and life expectancy in the United States, 2001-2014," *Jama*, 315, 1750–1766.

Chiappori, P. and B. Salanie (2000a): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56–78, publisher: The University of Chicago Press.

Chiappori, P.-A. and B. Salanie (2000b): "Testing for asymmetric information in insurance markets," *Journal of Political Economy*, 108, 56–78.

de Chaisemartin, C. and X. D'Haultfoeuille (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–2996.

Einav, L., A. Finkelstein, and M. R. Cullen (2010): "Estimating Welfare in Insurance Markets Using Variation in Prices," *The Quarterly Journal of Economics*, 125, 877–921.

Evans, D. S. and L. S. Leighton (1989): "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, 79, 519–535.

Finkelstein, A. and K. McGarry (2006): "Multiple dimensions of private information: evidence from the long-term care insurance market," *American Economic Review*, 96, 938–958.

Finkelstein, A. and J. Poterba (2004): "Adverse selection in insurance markets: Policyholder evidence from the UK annuity market," *Journal of political economy*, 112, 183–208.

FSA (2010): "Työttömyyskassat 2010," Annual Report the of Financial Supervisory Authority.

Goetz, C., H. R. Hyatt, Z. Kroff, K. Sandusky, and M. Stinson (2025): "Business Owners and the Self-Employed: Thirty-Three Million (and Counting!)," NBER Working Paper 34252.

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225, 254–277.

Landais, C., A. Nekoei, P. Nilsson, D. Seim, and J. Spinnewijn (2021): "Risk-based selection in unemployment insurance: Evidence and implications," *American Economic Review*, 111, 1315–1355.

Lazear, E. P. (2004): "Balanced Skills and Entrepreneurship," *American Economic Review*, 94, 208–211.

Levine, R. and Y. Rubinstein (2017): "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *The Quarterly Journal of Economics*, 132, 963–1018.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225, 175–199.

Tazhitdinova, A. and G. Vazquez-Bare (2023): "Difference-in-Differences with Unequal Baseline Treatment Status," NBER Working Paper 31063.

Tables and Figures

Table 1: Descriptive statistics: Years 2008-2010

	All	Treated	Always Unrestricted	Always Restricted
N	184660	58834	18195	107631
SI contributions	7611.6	6865.4	5578.7	8311.6
	(4894.5)	(4272.2)	(3927.4)	(5190.0)
Labor income	36049.0	33387.3	33918.5	37864.2
	(22372.1)	(19975.1)	(20688.7)	(23671.2)
SI/TI	0.84	0.83	0.71	0.87
	(0.20)	(0.19)	(0.29)	(0.18)
Total income	43358.9	42616.4	43500.1	43740.9
	(24705.8)	(24024.9)	(25079.3)	(24998.7)
No. sick days	1.17	1.33	1.95	0.95
	(11.63)	(11.89)	(15.58)	(10.67)
No. parental leave days	0.98	0.84	0.92	1.07
	(11.94)	(10.92)	(11.08)	(12.61)
Women	0.30	0.25	0.18	0.34
Age	44.4	44.6	47.0	43.9
	(11.5)	(9.4)	(9.5)	(12.7)
Tertiary educated	0.51	0.44	0.46	0.55
Industry distribution				
Manufacturing & Construction	0.19	0.20	0.21	0.18
Services	0.44	0.41	0.38	0.47
Wholesale & Transportation:	0.29	0.30	0.33	0.27
Other	0.09	0.1	0.09	0.08

Notes: Table 1 presents the descriptive statistics in our pre-reform period (i.e. years 2008-2010). All euro values are expressed in 2010 euros, corrected using the Consumer Price Index (CPI). Values are presented in means with standard deviation shown in parathesis. The first two columns use the sample presented in Figure 1 (a). Columns 3, 4 and 5 present the statistics for our treatment group (ownership share between 0.3 and 0.5), and two alternative control groups, always restricted (ownership share between 0.1 and 0.3) and unrestricted (ownership share between 0.5 and 0.7) defined as in Figure 4.

Table 2: Positive Correlation Estimates and the Costs of Risk-based Selection

Panel A: Correlation estimates

	Correlation estimate (standard error)	N
Retirement Age	-0.59 (0.17)	XX
Life Expectancy	0.79 (0.15)	XX
sick leave	1.13 (0.18)	XX
Parental Leave	0.67 (0.18)	XX

Panel B: Selection cost evaluation

	Mean SI Contributions (data)	
	SI/TI=0.2	SI/TI=1.2
SI Contributions	1662	13,071
SI Income	6925	54,462

	Benefit costs due to selection	
Pension	Sick Leave	Parental Leave
827	83	76

	Benefit costs per SI contributions	
Pension	Sick Leave	Parental Leave
0.072	0.007	0.006

Notes: Panel A of the table presents the positive correlation estimates from Figure 2 between the chosen SI coverage (SI/TI) and retirement age, life expectancy, sick leave and parental leave. Panel B presents the numbers used for the evaluation of the costs due to risk-based selection in SI contributions, based on the estimated correlations. Table shows the mean SI contributions and SI income for those with SI/TI=0.2 and SI/TI=1.2 in Figure 2. Based on these numbers and changes in SI claims from SI/TI=0.2 to SI/TI=1.2, the table presents the benefit costs due to selection. Benefit costs for pension are calculated as $0.015 \times 47,537 \times 1,16 = 827$, where 0.0.15 is the accrual rate, 47,537 is the increase in SI income from SI/TI=0.2 to SI/TI=1.2, and 1.13 is the increase in the time collecting pension benefits between SI/TI=0.2 and SI/TI=1.2, stemming from earlier retirement (4 months) and living longer (10 months). For sick leave, the benefit formula is $(0.7 \times 47,537/300 - 0.55 \times ((47537 - 28241)/300)) \times 1.1 = 83$, where 0.7 and 0.5 are replacement rates for income below and above €28,241, and 1.1 is the estimated increase in sick leave leave days. For parental leave, the formula is $(31,99 + (44702 - 13713)/300 \times 0.7 + 0.4 \times (47537 - 44702)/300) \times 0.7 = 76$, where 31,99 is the minimum benefit for those who have income below €13,712, 0.4 is the reduced replacement rate for income above €44,702, and 0.7 is the estimated increase in parental leave days. Benefit costs per additional euro of SI contributions are calculated by dividing the benefit costs with the increase in SI contributions from SI/TI=0.2 to SI/TI=1.2, which is €11,409.

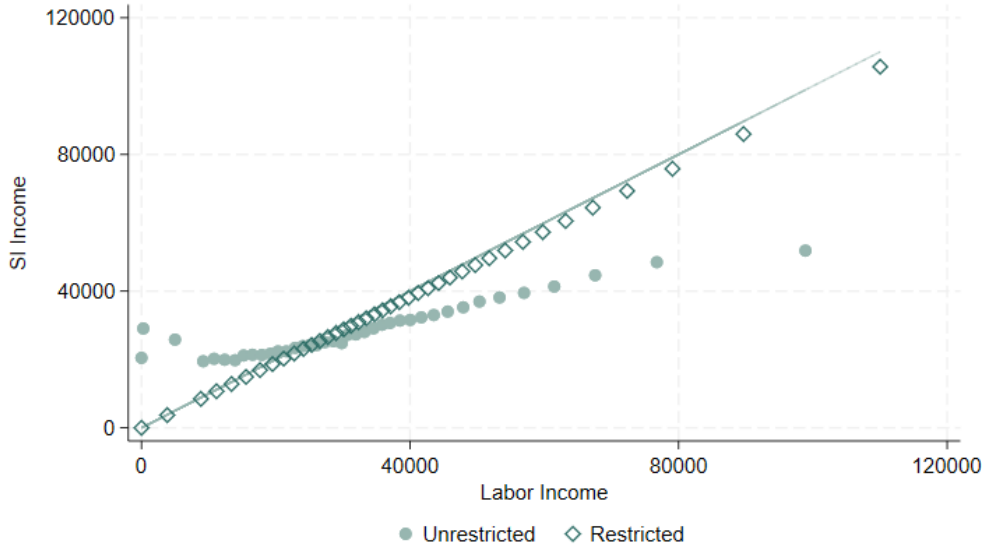
Table 3: Reduced Form Estimates on Claims [TBA]

	Average	$SI/TI < 0.5$	$SI/TI \in [0.5, 0.9]$	$SI/TI > 0.9$
DiD estimates:				
Average SI/TI change				
	( )	( )	( )	( )
Total Days				
	( )	( )	( )	( )
Pension Claiming				
	( )	( )	( )	( )
Sick Leave Days				
	( )	( )	( )	( )
Parental Pay Days				
	( )	( )	( )	( )
N				
R2				
Pre-reform mean				

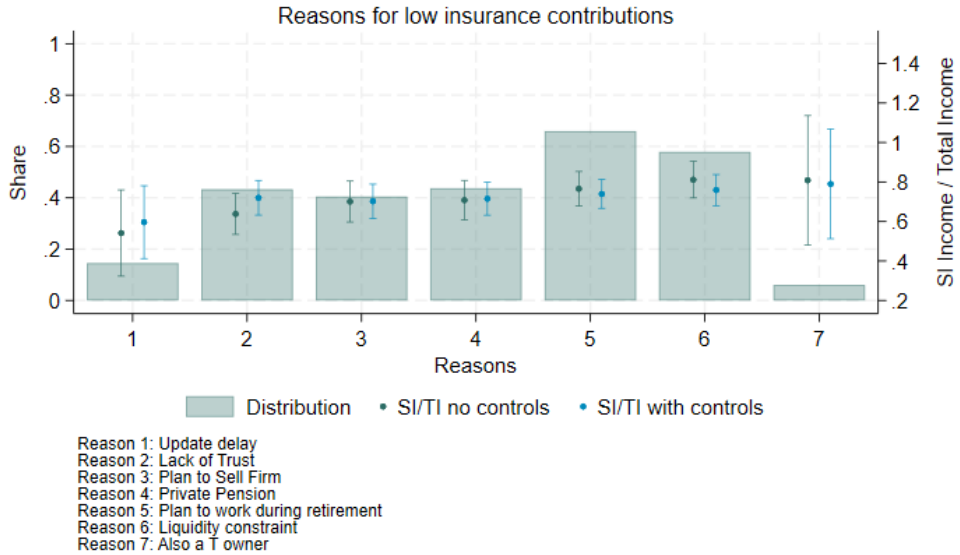
Notes: The table presents the reduced form estimates for moral hazard responses estimating equation $y_{it} = \beta Treat_i \ 1_{t=\ell} + Treat_i + \eta_t + \delta X_{it} + \varepsilon_{it}$. The structure presented follows that in Figure 6. The control group is always unrestricted and the results are pooled difference-in-differences estimates for the pre and post years.

Figure 1: Choice of Social Insurance Coverage, YEL (unrestricted) vs. TyEL (restricted) Owners

(a) SI Income Choice: Restricted vs Unrestricted

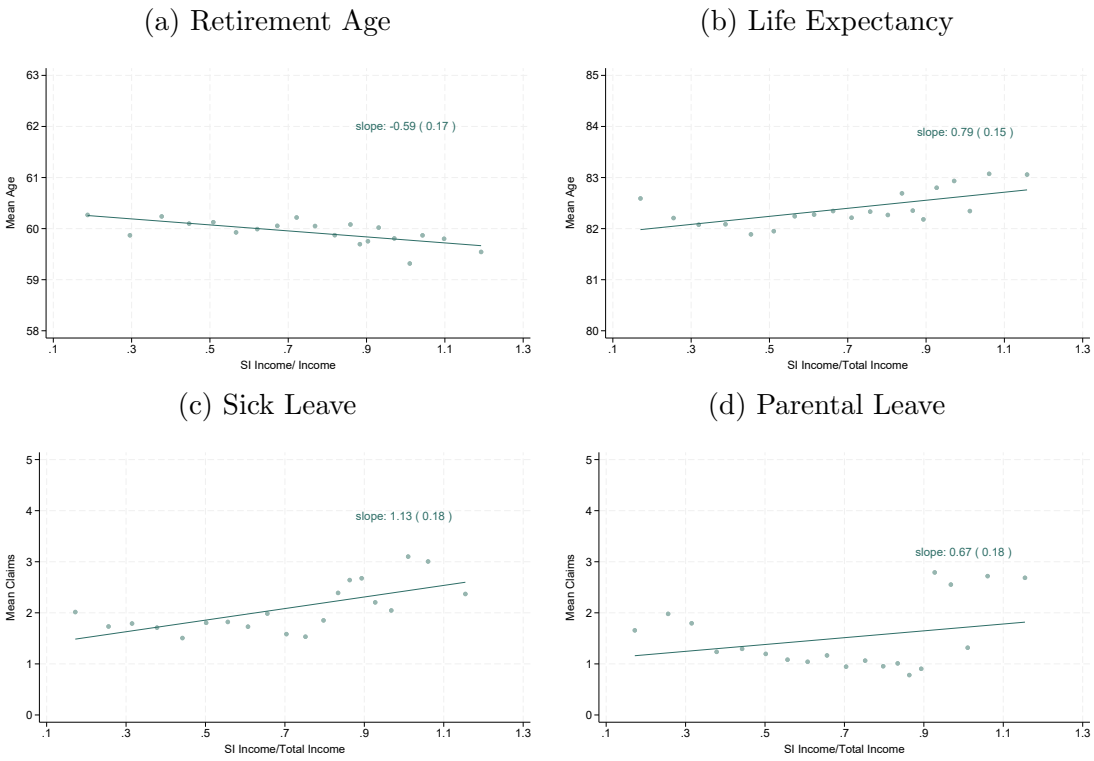


(b) Actual SI Income Choice vs. Stated Reasons for Low Insurance Choice



Notes: Figure (a) shows the social insurance (SI) income relative to labor income (LI) for all restricted (TyEL) and all unrestricted (YEL) owners. Each point contains 2.5% of observations with respect to their group. Figure (b) shows the distribution on the answers to the survey question of what are the reasons for low YEL insurance contributions (left-hand side y-axis). The main answer categories include: 1) update delay on YEL income, 2) plan to sell the firm, 3) private pension savings, 4) plan to work during retirement, 5) lack of trust in the SI system, 6) liquidity constraints in paying SI contributions, and 7) simultaneous TyEL insurance. The figure includes the average social insurance income per labor income with 95% confidence interval within each answer category (right-hand side axis), measured for all survey respondents.

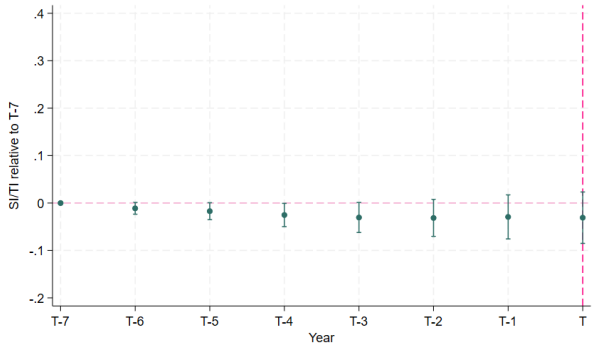
Figure 2: SI Coverage vs. Claims: Positive Correlation Tests



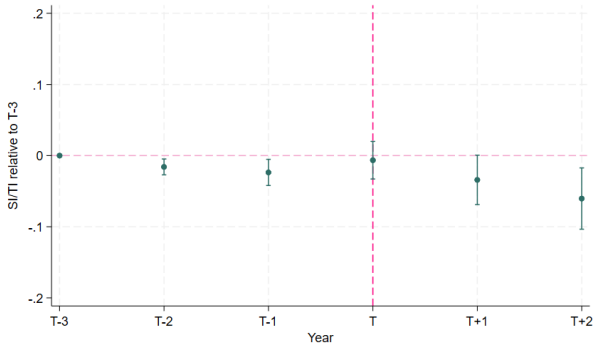
Notes: These figures show the average number of benefit days claimed by the unrestricted business owners in our core sample by bins of SI/LI. Each SI/LI bin contains 5% of person-year observations. The y-axis shows the average number of days claimed (including zeros). The graph in panel (a) plots the average retirement age by SI/LI choice, highlighting the relationship between SI coverage and retirement decisions. Panel (b) plots the approximated life expectancy by SI/LI choice. Appendix B explains in detail how we predict life expectancy using our register data. In panels (c) and (d) the outcomes are sick leave and parental leave days. The fitted lines are estimated from the raw data (i.e. person-year observations), and includes the following controls: age, gender, municipality, industry and total income.

Figure 3: Anticipatory Responses to SI Claims

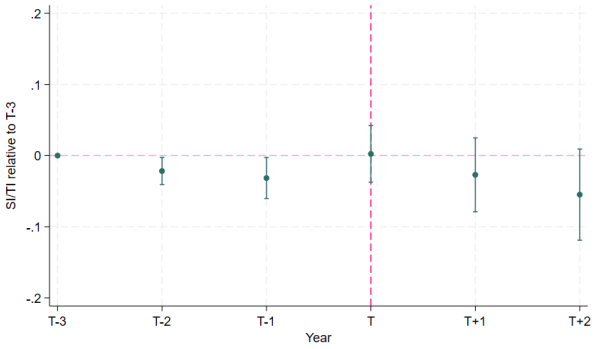
(a) Anticipation of Retirement



(b) Sick Leave



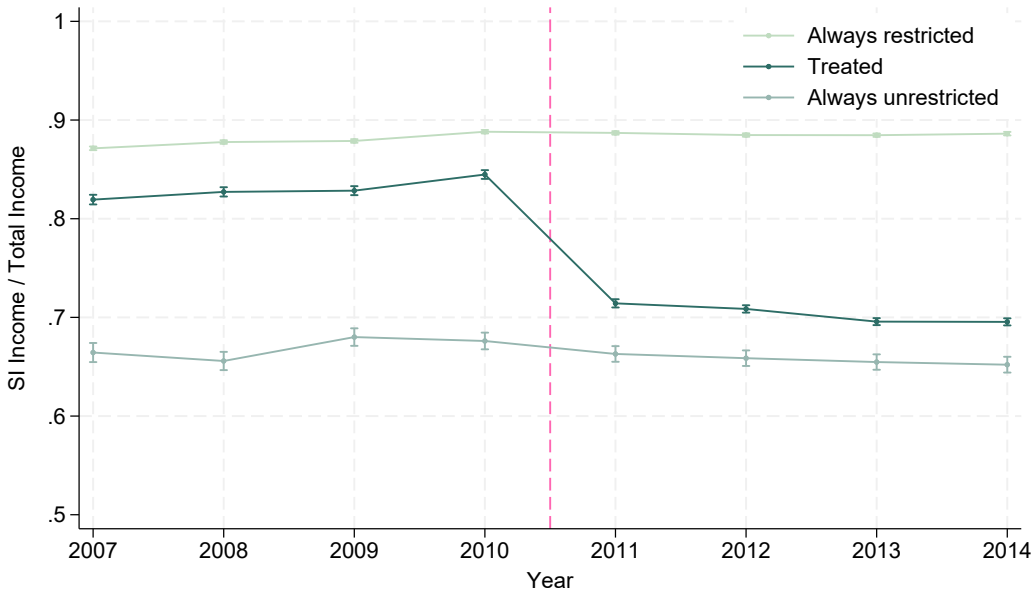
(c) Parental Leave



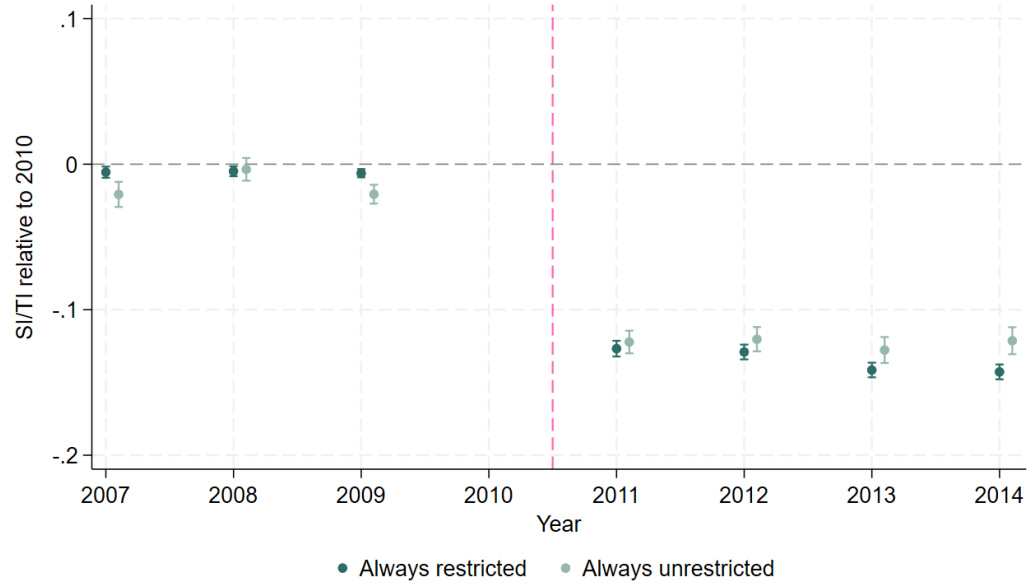
Notes: Figure shows the results of estimating equation (1) on the core sample of unrestricted and restricted entrepreneurs. Year T on the horizontal axis identifies the first year when an individual claims a benefit. Panel (a) studies the development of SI/TI seven years prior to retirement, and panels (b) and (c) three years before and after claiming sick leave and parental leave. The estimation includes as controls age, gender, municipality and industry, and year fixed effects. The 95% confidence intervals are derived from standard errors that are clustered at the individual level.

Figure 4: Choice of Social Insurance Coverage after 2011 Reform

(a) Development of SI Income Around the 2011 Reform



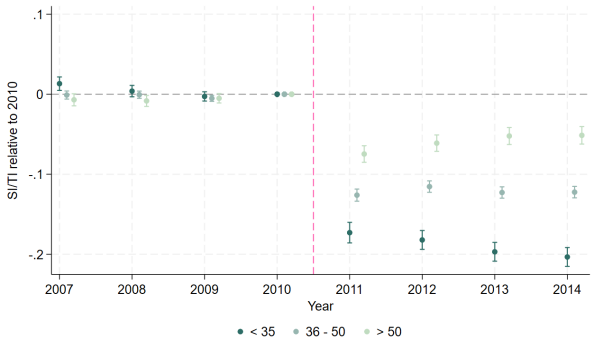
(b) Event-study Estimates



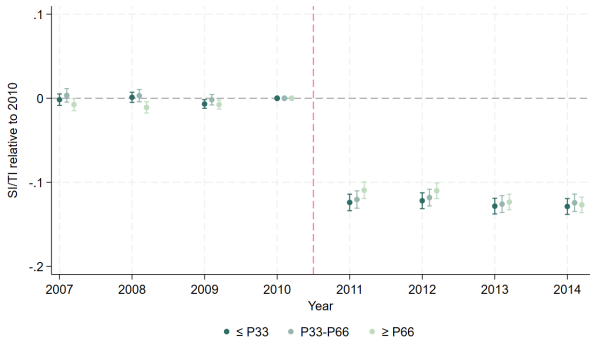
Notes: Panel (a) of the figure shows the results of estimating equation (2), where the outcome variable is SI income relative to labor income. The control groups consists of all always restricted (ownership share between 10% and 30%) and always unrestricted owners (ownership share between 51% and 70%). Panel (b) shows the same results as a difference in differences estimate. Standard errors are clustered at the individual level.

Figure 5: Heterogeneity in Social Insurance Coverage after 2011 Reform

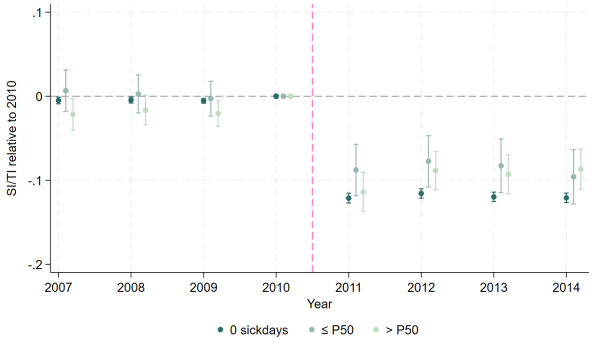
(a) Age



(b) Life Expectancy



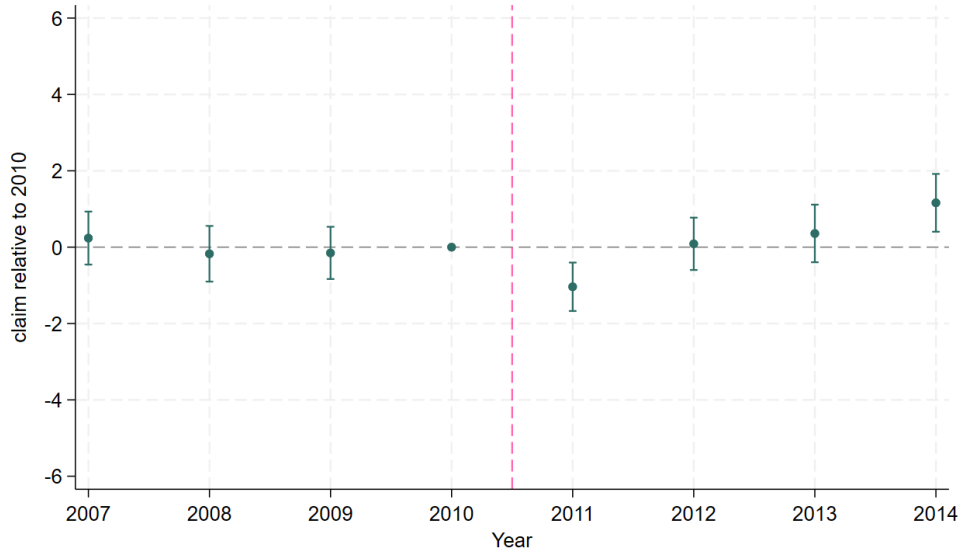
(c) Sick Claims



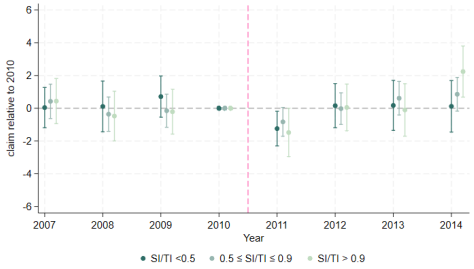
Notes: These figures show the results of estimating equation (2), where the outcome variable is SI income relative to labor income. The control groups consists of all always restricted (ownership share between 10% and 30%). Panel (a) shows the heterogeneity in the results when individuals are separated into three groups based on their age in 2010. Panel (b) illustrates heterogeneity in life expectancy. Individuals are divided into three groups according to whether their estimated, residualized life expectancy falls below the 33rd percentile, between the 33rd and 66th percentiles, or above the 66th percentile. Residualization is performed by regressing life expectancy on gender and age cohort dummies, and their interactions, industry, municipality and income decile dummies before constructing these groups. Lastly, panel (c) shows heterogeneity by number of sick days in the pre period, where one group has zero sickness claims, the middle group has sickness claims below or at the median level and the highest group has above median sickness claims. Sickness claim groups are formed using residualization in the same way as for life expectancy, except that zero sickness claims are based on actual observations prior to 2011. Standard errors are clustered at the individual level.

Figure 6: SI Claims and the 2011 Reform

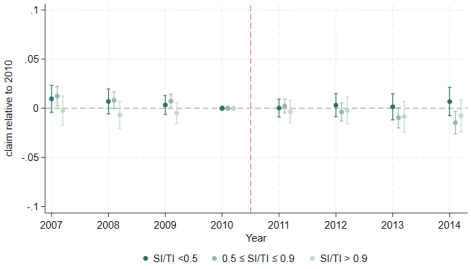
(a) Total Days



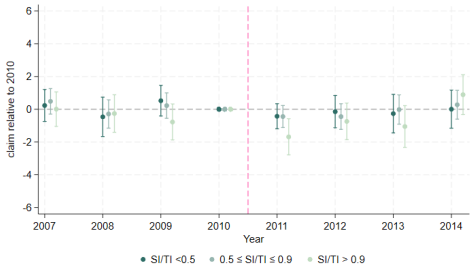
(b) Total Days



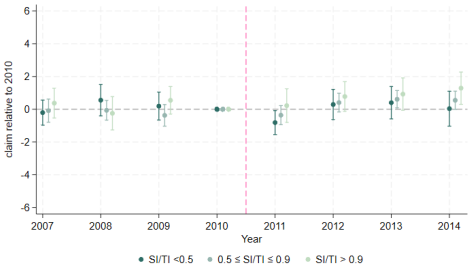
(c) Pension Claiming



(d) Sick Leave Days



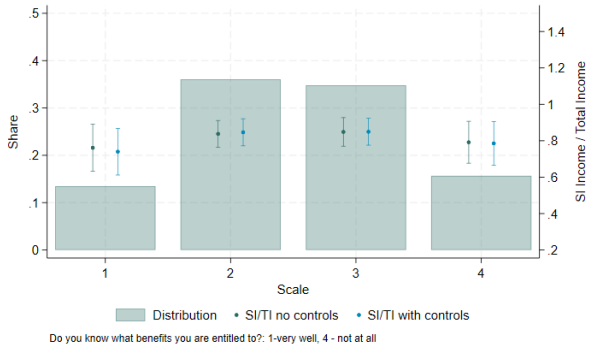
(e) Parental Pay Days



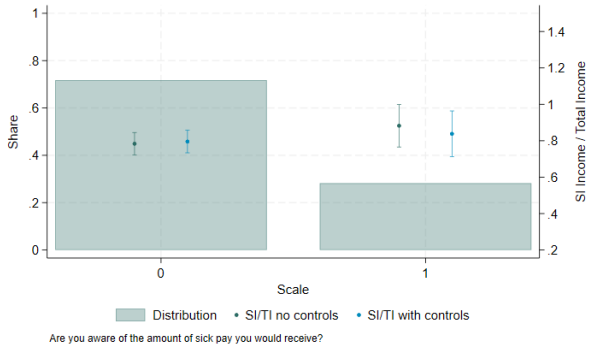
Notes: These figures show the results of estimating (2), where the outcome variable measures the number of days per year an individual claims absence. In each figure, the individuals are separated into three groups based on the average SI/TI in post-reform years (2011-2014): those with SI/TI below 0.5, between 0.5 and 0.9, and above 0.9. The control group consists of all always unrestricted individuals (whose ownership share is between 50 and 70%). In panel (a) the outcome is the total number of absent days, including absences due to sickness and parental leave. In Figures (b), (c) and (d) the outcomes are the likelihood of claiming pension benefits, and the number of sick leave and parental leave days, respectively. Standard errors are clustered at the individual level.

Figure 7: Perceived Knowledge about SI and Health

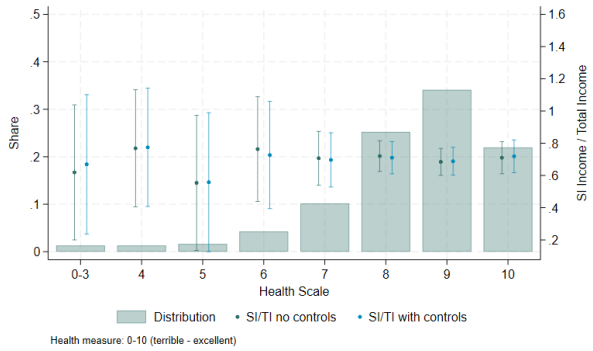
(a) Knowledge about SI Claims



(b) Awareness of Sick Leave Replacement Rates



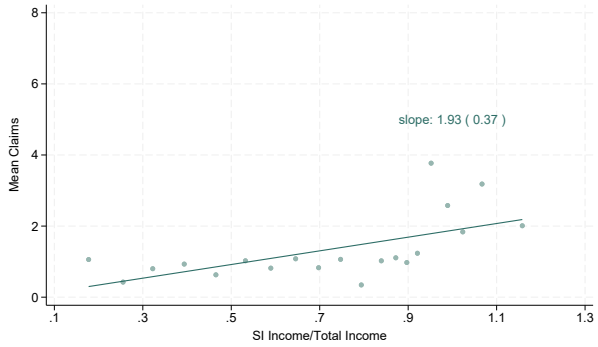
(c) Perceived Health



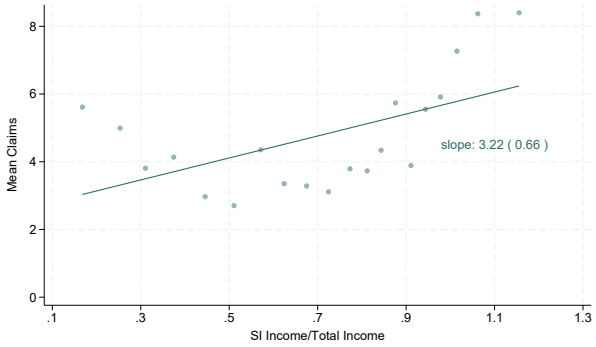
Notes: The figure shows the distributions on the answers to the following survey questions (left-hand side axis): Evaluation of your own perceived health status on the scale from 0 to 10 in panel (a); Do you know what SI benefits you are entitled to, on the scale of 1 (very well) to 4 (not at all) in panel (b), and are you aware of the size of your sick leave benefit (yes or no) in panel (c). The figure includes the average social insurance income per labor income with 95% confidence intervals within each answer category of each question (right-hand side axis), measured for all survey respondents.

Figure 8: Learning over Time

(a) Treated Individuals: 2011

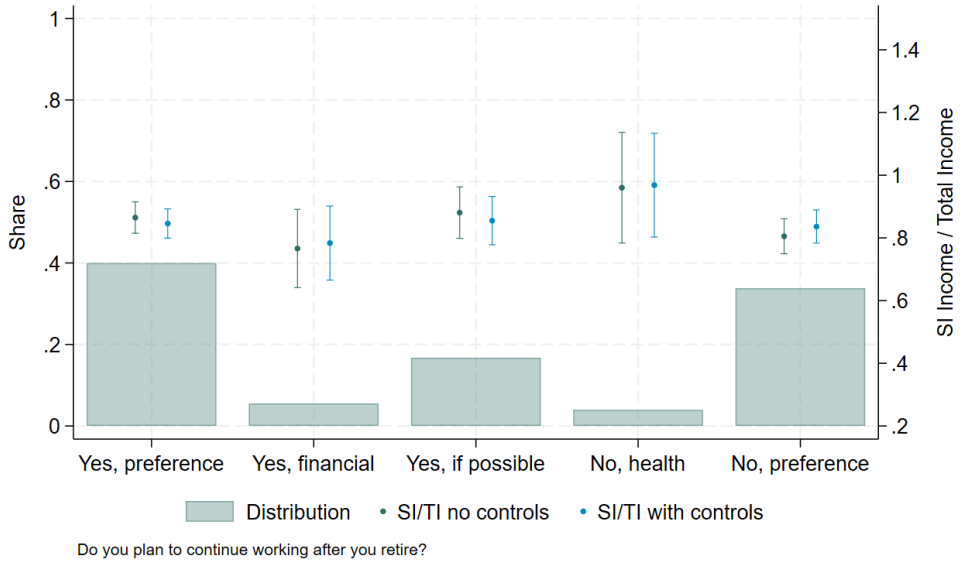


(b) Treated Individuals: 2014



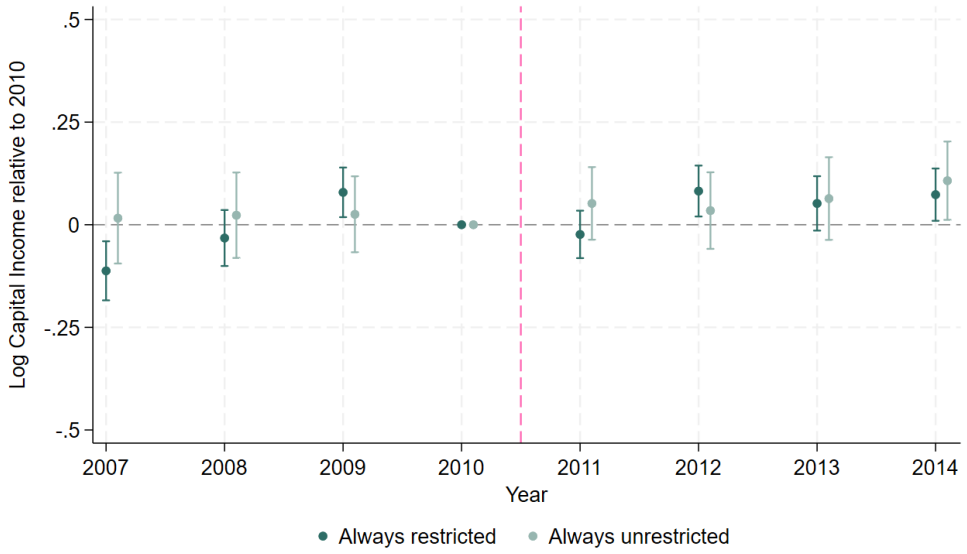
Notes: Figures (a) and (b) show the average number of benefit days claimed (including zeros) by the treated individuals in 2011 and 2014, by bins of SI/TI in 2011-2014. Each SI/TI bin contains 10% of 2011-2014 person-year observations. The fitted line shown is estimated on the raw data (i.e. person-year observations), and includes the following controls: age, gender, municipality and industry. Each SI/TI bin contains 5% of person-year observations. The y-axis shows the average number of parental leave days claimed, conditional on claiming.

Figure 9: Plans on Working After Retirement



Notes: Figure shows the distributions on the answers to the following survey question (left-hand side axis): Plans to continue working after retirement. The figure includes the average social insurance income per labor income with 95% confidence intervals within each answer category of each question (right-hand side axis), measured for all survey respondents. Standard errors are clustered at the individual level.

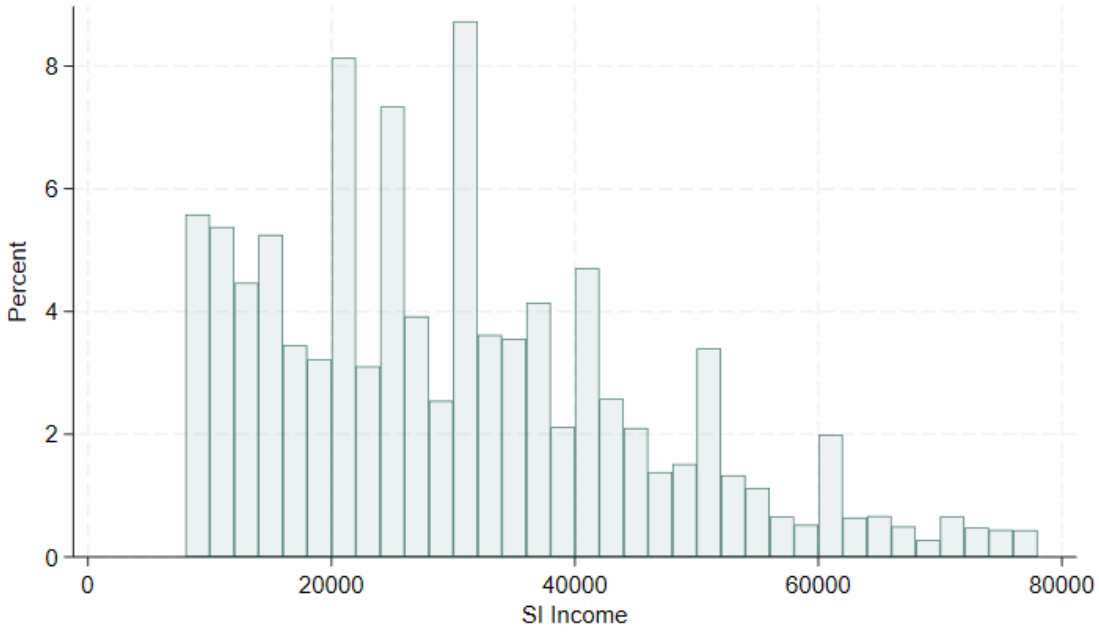
Figure 10: Crowd-out: Capital Income and the 2011 Reform



Notes: Figure shows the results of estimating (2), where the outcome variable measures capital income for individuals in logs. The control group consists of all always restricted individuals (whose ownership share is between 10 and 30%) and the always unrestricted individuals (whose ownership share is between 50-70%). Standard errors are clustered at the individual level.

A Appendix

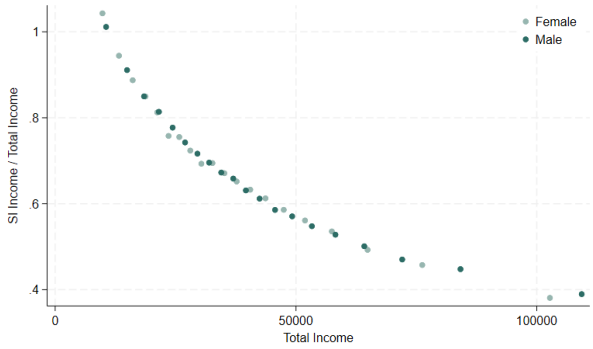
Figure A.1: Distribution of SI Income of YEL owners



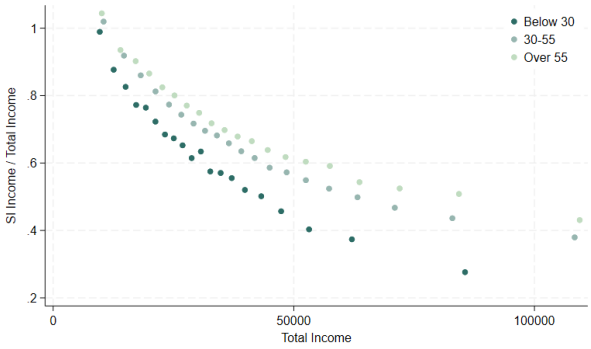
Notes: Figure shows the distribution of SI income in 2010 for the always unrestricted entrepreneurs in our baseline sample.

Figure A.2: Choice of SI Income by Characteristics

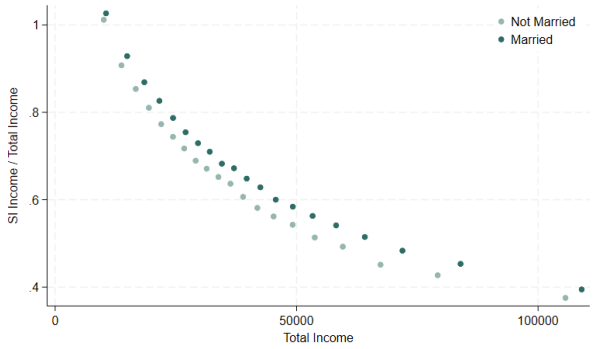
(a) SI Income Choice by Gender



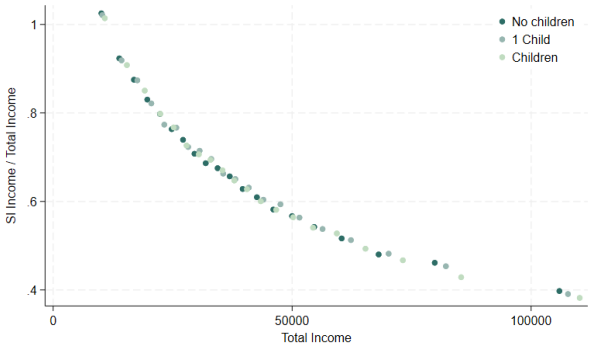
(b) SI Income Choice by Age



(c) SI Income Choice by Marital Status

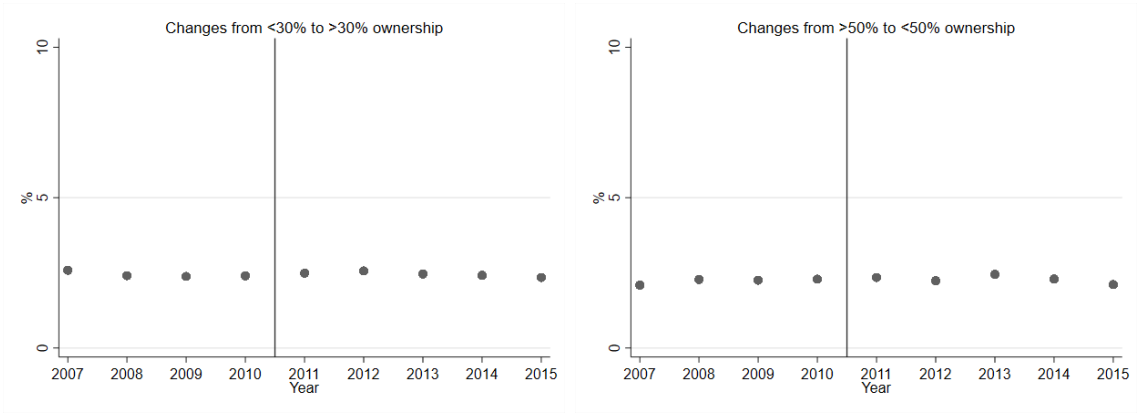


(d) SI Income Choice by Children



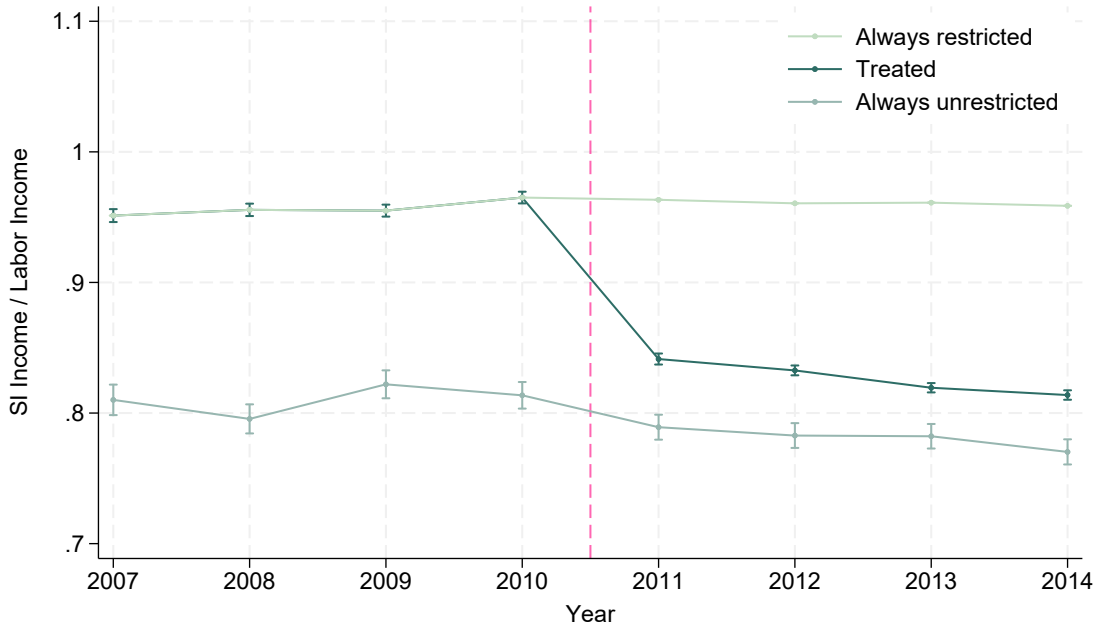
Notes: Figure shows the choice of SI income within each labor income bin by different characteristics: Gender (panel a), age groups (panel b), marital status (panel c) and the number of children (panel d). The figure shows that there are no significant differences in by gender or number of children. In contrast, the SI income is lower within each income group among younger entrepreneurs and those who are married.

Figure A.3: Ownership share changes over time (NEEDS UPDATE)



Notes: These figures show the percent of individuals that switched ownership status from below 30% to above 30%, or and from above 50% to below 50%. In both cases, only approximately 2.5% of business owners changed their ownership status enough to move from treatment to control group or vice versa. Importantly, there is no change in the share of owners jumping across these thresholds at the time of the 2011 reform.

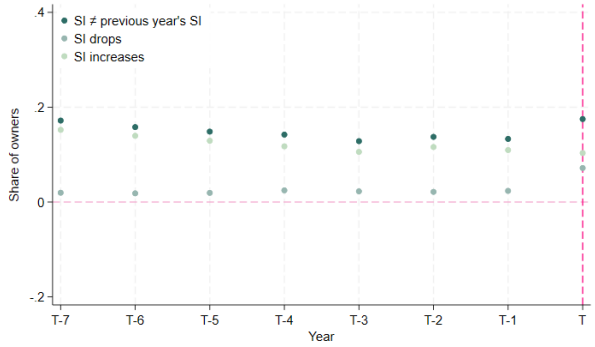
Figure A.4: Changes to Reported SI Income in the 2011 reform: Time Series



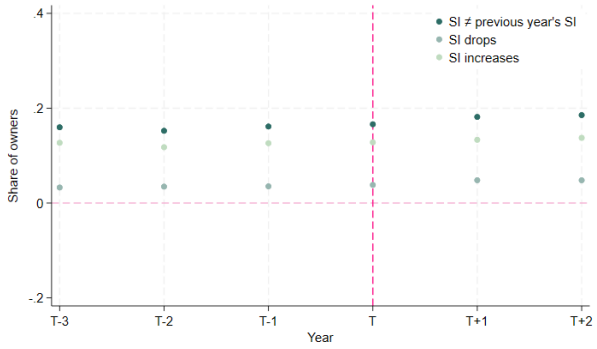
Notes: Figure shows the evolution of SI/LI (social insurance income divided by labor income) over time for treated individuals (ownership share between 30-50%) and the always unrestricted individuals (ownership share between 50-70%).

Figure A.5: Anticipatory Responses to SI Claims: Active Changes in SI Levels

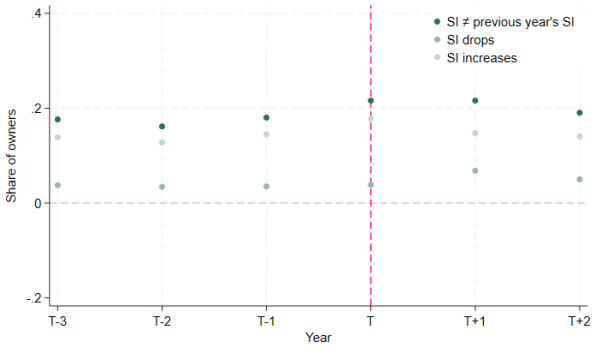
(a) Anticipation of Retirement



(b) Sick Leave



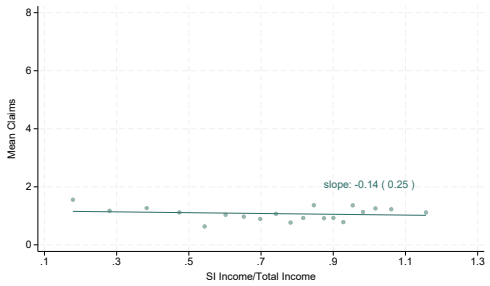
(c) Parental Leave



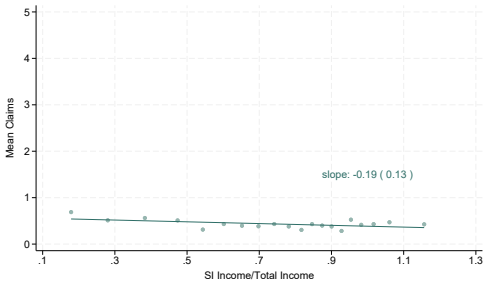
Notes:

Figure A.6: Pre-reform Claims and Post-reform Insurance Choice

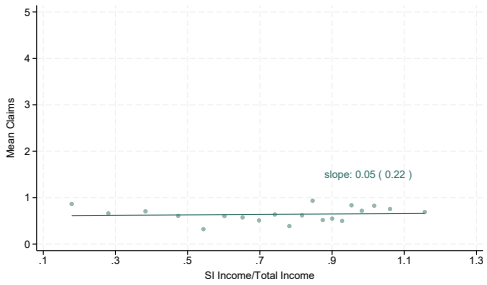
(a) All claims



(b) Sick Leave



(c) Parental Leave



Notes: These figures show the average number of benefit days claimed by the treated individuals (those with ownership share between 30-50%) in 2006-2010 (Panel A) and in 2011-2014 (Panel B) by bins of SI/LI in 2011-2014. Each SI/LI bin contains 5% of 2011-2014 person-year observations. The y-axis shows the average number of days claimed (including zeros): in figures (a) the outcome is the total number of benefit days, and in (b) and (c) the number of sick leave and parental leave days. The fitted line shown is estimated on the raw data (i.e. person-year observations), and includes the following controls: age, gender, municipality and industry.

B Predicting Life Expectancy

Similarly to Chetty et al. (2016), we utilize Gompertz's Law, which posits a log-linear relationship between mortality and age. To utilize this law, we run a maximum likelihood estimation, which follows a Gompertz survival model. We use this model in order to estimate the life expectancy of individuals in our sample.

Using maximum likelihood estimation which follows a Gompertz survival model, we estimate the shape parameter as well as the baseline hazard for men and women separately. Using a large number of explanatory variables, we can create a crude estimate of survival probabilities for each individual separately. The survival function is as follows:

$$S(t) = \exp\left(-\lambda_j \gamma^{-1}\left(e^{\gamma t_j} - 1\right)\right) \tag{3}$$

where the model is implemented by parameterizing $\lambda_j = \exp(x_j\beta)$, implying that the hazard function, $h_0(t) = \exp(\gamma t)$, where γ is an ancillary/shape parameter, which is estimated from the data accounting for differences in life expectancies for each cohort.

Using the individual survival probabilities for each age, we can estimate the life expectancy of each individual. Ideally, we could take the area under the survival curve, however, due to the complexity of the integral, we opt to calculating the life expectancy as follows:

$$\text{Life Expectancy} = Age + \sum_{t=1}^{100} S(t) \tag{4}$$

where $S(t)$ calculates the probability of surviving t years after Age, which is the age of an individual first obsereved in the data. We restrict the sample to those surviving at least until 40. Thus, the life expectancy calculates the probability of surviving each additional year after entering our sample.

We use the model to estimate parameters for characteristics, baseline hazard as well as a shape parameter. In the estimation of these parameters, we utilize full population

data in Finland of individuals between 1987 and 2019. We use a sample who are born between 1920 and 1970. We also exclude individuals who die before the age of 40 and older than 100, so the model will not associate early deaths with individuals born later and later deaths with individuals born earlier. In order to gauge the reliability of the estimated life expectancies, we split the data into 2 different groups: training and validation (70/30). We randomize individuals into these groups to ensure a balanced demographic among them. This split first allows us to fit a Gompertz model onto the training data and once the model fits well, we can test the parameters on the validation data. Wellness of fit is determined by seeing how well the model estimates the death of individuals who have already passed away in the data and how well the model estimates the share of individuals that are alive at each age. An important aspect of the fit for our analysis is also a positive correlation between estimated life expectancy and real life expectancy.

In order to account for differences in the longevity of lives between cohorts, the shape parameter is calculated separately for each cohort. All parameters are estimated separately for men and women, as these groups tend to have dichotomous life expectancies and explanatory variables may affect hazard rates differently for men and women.

The following are used to explain differences in the death hazard rate: education (dummies for each level of education), occupation (white/blue collar higher/lower), region born (4 regions), region of residence (52), income (income rank within 7 bins), unemployment, sick leave, disability pension, number of children, marital status, interaction between income rank and region of residence and interaction between marital status and income rank. The maximum likelihood model then fits a Gompertz distribution on the sample, and we can use the estimated coefficients to estimate different life expectancies for our study sample.

Note, that the model estimates life expectancy based on income, industry and basic demographics of individuals, thus if these individuals have any medical conditions, we are not able to observe this. However, we can proxy health using a variable which tells us the amount of sickness allowance an individual has used yearly.

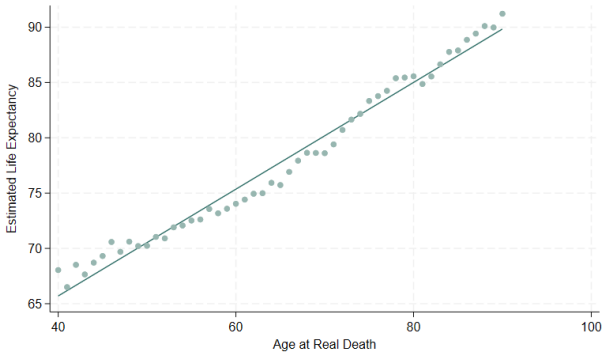


Figure B.1: Life Expectancy and Death

Notes: The figure shows the correlation between estimated and real life expectancies for individuals who have passed away.

The model predicts the individual life expectancies of our sample quite poorly, as our model over predicts life expectancies for most individuals. This is likely because our explanatory variables alone cannot explain early deaths in the sample. However, as seen in Figure B.1, there is a strong positive correlation between estimated life expectancies and real ages at death. This means that our model is able to correctly predict that individuals who actually die earlier have lower life expectancies, which is important for our analysis.