VATT INSTITUTE FOR ECONOMIC RESEARCH

VATT WORKING PAPERS 185

Moral Hazard and Adverse Selection in the Social Insurance Market for Entrepreneurs

Youssef Benzarti
Jarkko Harju
Tuomas Matikka
Ella Mattinen
Alisa Tazhitdinova

VATT WORKING PAPERS 185

VATT WORKING PAPERS 185

Moral Hazard and Adverse Selection in the Social Insurance Market for Entrepreneurs

Youssef Benzarti
Jarkko Harju
Tuomas Matikka
Ella Mattinen
Alisa Tazhitdinova

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Helsinki 2026

VATT WORKING PAPERS 185

Youssef Benzarti, University of California, Santa Barbara and NBER

Jarkko Harju, Tampere University, FIT, VATT and CESifo

Tuomas Matikka, VATT, FIT and CESifo

Ella Mattinen, Tampere University, FIT and VATT

Alisa Tazhitdinova, University of California, Santa Barbara and NBER

VATT Working Papers 185
URN:NBN:fi-fe2026040125050

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Helsinki, April 2026

Moral Hazard and Adverse Selection in the Social Insurance Market for Entrepreneurs*

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

March 26, 2026

Abstract

This paper studies social insurance contribution choices, adverse selection, and moral hazard among Finnish entrepreneurs. We exploit quasi-exogenous variation from a reform that relaxed mandatory contribution requirements for a subset of entrepreneurs, combining administrative registry data with linked survey evidence. Entrepreneurs who gained discretion reduced their contributions by 16%, on average, relative to entrepreneurs subject to a strict mandate. Using this variation, causal tests of anticipatory responses and moral hazard as well as positive correlation tests, we show that moral hazard and adverse selection effects are near zero in this market. Survey responses help illuminate the mechanisms underlying these results.

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

JEL: H55, J32, L26.

*We are grateful to 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. Harju, Matikka and Mattinen gratefully acknowledge funding from the Academy of Finland (No. 346252), and Matikka from the Strategic Research Council (SRC) at the Academy of Finland (No. 345386).

†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 OECD countries (Boeri et al., 2020; Goetz et al., 2025). The majority of small entrepreneurs face substantially greater income volatility than wage earners and may therefore stand to benefit from access to social insurance programs (Audoly, 2022; Boeri et al., 2020). Yet, 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 entrepreneurs is challenging. Entrepreneurs 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 them can be susceptible to higher levels of adverse selection, and may entail higher costs due to potential moral hazard responses, relative to social insurance programs for wage earners. However, little is know about the extent of adverse selection and moral hazard among entrepreneurs, largely because of the absence of suitable data and identifying variation. Moreover, it remains unclear just how much coverage entrepreneurs want, and whether they prefer to allocate their scarce financial resources toward investing in their businesses instead.

In this paper, we exploit a unique institutional setting in Finland to provide novel evidence on small entrepreneurs' willingness to pay for social insurance (SI) and the extent of adverse selection and moral hazard in this market. In Finland, certain entrepreneurs can 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 applied only to entrepreneurs who held more than 50% of their firm's shares. Those with smaller ownership stakes were required to participate in the traditional SI scheme for wage earners, with contributions and benefits determined by their earnings and no option to deviate from these rules. A policy reform in 2011 lowered the ownership threshold from 50% to 30%, granting owners

with an 30–50% ownership share the ability to determine their SI contributions.

Our main analysis relies on rich administrative data sources covering social insurance contributions, benefit claims, and comprehensive individual and business tax records. We complement administrative data with surveys of entrepreneurs conducted by Statistics Finland. The surveys include detailed information on attitudes toward the SI system, knowledge about social insurance benefits, as well as 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. We use these two sources of data to estimate the magnitude of moral hazard and adverse selection using several different empirical tests.

We begin by measuring the extent of SI coverage chosen by entrepreneurs. Following the 2011 reform that expanded contribution flexibility to entrepreneurs holding 30–50% of their firm's shares, newly eligible owners markedly reduced their social insurance payments. Event-study estimates show that these entrepreneurs lowered their SI income relative to total income by roughly 13 percentage points, corresponding to an average 16% decline in contributions compared to otherwise similar owners who remained under the mandated contribution levels. This reduction brought their contribution levels close to those of entrepreneurs who already had full discretion prior to the reform. While some business owners lowered their contributions to the legal minimum, many did not. This shows that entrepreneurs value social insurance but prefer lower coverage levels than is normally mandated in Finland. We also provide evidence based on a survey of Finnish entrepreneurs that is consistent with these findings.

Next, we use three different approaches to estimate the extent and magnitude of moral hazard and adverse selection in this market. First, we estimate moral hazard responses using event-study analyses around the 2011 reform. Intuitively, as entrepreneurs decrease their social insurance coverage, we should observe a decrease in claims if moral hazard responses are prevalent. More 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 shut down, and differences in claims identify moral hazard responses, provided that individual risk profiles either remain stable or evolve similarly over time across the treatment and control groups.

Our event-study estimates indicate that moral hazard responses are negligible: although the 2011 reform induced treated entrepreneurs to reduce their SI contributions by roughly 16%, the corresponding change in benefit usage was economically and statistically insignificant. Estimated differences in total SI claims, pension benefits, sick leave days, and parental leave days are all close to zero (less than 0.2 days or 0.1%) relative to the control groups. This effect holds both on average, and when focusing on claims of individuals with very large (>50%) reform-induced reductions of SI coverage.

Second, we use anticipatory response tests to estimate adverse selection: if adverse selection behavior is present, entrepreneurs should adjust their social insurance coverage in anticipation of known future claims. To estimate such responses, we restrict the sample to individuals who ultimately claim a certain benefit type and we study the level of contributions (rather than claim durations) before and after the claim event. By doing so we shut down the moral hazard margin and isolate adverse selection responses. We find very limited evidence of anticipatory responses in this market: the ratio of SI income to total income remains approximately constant close to sickness and parental leave events, or prior to retirement. Our results thus rule out economically meaningful adverse selection effects in this setting.

Finally, we quantify the overall relationship between coverage levels and claims using positive correlation tests (Chiappori and Salanie, 2000). These tests estimate the combined effect of moral hazard and adverse selection and have the advantage of relying on very few assumptions. We find that SI contributions are only weakly correlated with key determinants of insurance usage. Our results show that an additional euro of con-

tributions only increases pension benefit claims by 9.88 cents, sick leave claims by 1 cent and parental leave benefits 0.62 cents. These positive correlation tests suggest that the joint effect of moral hazard and adverse selection is limited in this setting.

In sum, while allowing entrepreneurs to choose their SI coverage flexibly leads them to select lower contribution – and thus lower coverage – levels, this flexibility does not generate any meaningful adverse selection effects. Entrepreneurs neither sort by underlying risk, resulting in similar SI choices across demographic groups, nor adjust their coverage in anticipation of future claims. We also find negligible moral hazard effects: changes in benefit claims following adjustments in SI contributions are minimal and sometimes move in the opposite direction. Overall, the relationship between coverage and subsequent claims is weak. Thus, aside from the choice of coverage level itself, behavioral responses to flexible SI coverage are limited.

Our results carry important policy implications for the design of social insurance for entrepreneurs, and, possibly, for wage earners as well. The lack of moral hazard responses suggests that SI coverage for entrepreneurs need not be less generous than for wage earners. The lack of strong adverse selection responses suggests that social insurance schemes for entrepreneurs can be flexible, and the rules need not be primarily dictated by selection concerns. At the same time, our results highlight strong preferences for lower levels of coverage, in line with paternalistic motivations for rigid SI rules. Whether these paternalistic motivations are justified remains unclear for entrepreneurs as we are unable to observe the entirety of their financial positions at the moment of retirement. Benzarti et al. (2020) show that rigid rules may be too restrictive, especially for younger firms, preventing them from making high-return investments.

Our analysis, however, is subject to two important caveats. First, our research explores adverse selection and moral hazard responses along the intensive margin. All entrepreneurs in Finland are covered by minimum benefit provisions, and can only choose the level of coverage above this minimum. It is possible and likely that adverse selection and moral hazard are stronger along the extensive margin. Second, our analysis explores risk selection in a bundled setting. While the bundles studied are representative of typical

SI settings – in the majority of countries, SI benefits are bundled at least to some extent, the bundled nature makes sorting along risk dimensions more difficult. Nonetheless, we find little risk-sorting for pension benefits, arguably the most valuable benefit in our setting.

Our work contributes to several literatures. To the best of our knowledge, this is the first study to estimate moral hazard and adverse selection responses in government provided social insurance market for entrepreneurs. Business owners face distinctive and often more volatile risks, making the design of social insurance policies particularly consequential for them. By focusing on entrepreneurs, we provide direct evidence on whether the theoretical concerns about moral hazard and adverse selection – often assumed to be particularly acute in this group – exist in practice.

Furthermore, our study explores a classical social insurance scheme that is mandatory, administered by the government, and provides bundled coverage. In contrast, most of the previous work focused on unbundled insurance markets (e.g., health, long-term care, car, annuity, etc., see Finkelstein and McGarry, 2006; Einav et al., 2010; Finkelstein and Poterba, 2004; McCarthy and Mitchell, 2010) typically sold by private insurers, even in the case of social-insurance policies (e.g. Landais et al., 2021; Cabral et al., 2022; Cabral and Dillender, 2024; Seibold et al., 2025, study supplemental unemployment insurance, worker's compensation, and disability insurance, respectively). These differences in settings are likely to explain key differences in our findings. In contrast to annuity markets – which are not mandatory and attract a highly selected pool of individuals – participation in the pension insurance scheme we study is mandatory (but contribution levels are not). Consequently, lower levels of adverse selection in our setting may be driven by the (smaller) intensive margin incentives compared to the (larger) extensive margin incentives present in non-mandatory settings. Similarly, the bundled nature of the program makes sorting harder, again reducing selection, as compared, for example, to elective unemployment insurance market.

2 Institutional Setting

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, and parental allowances. The levels of these benefits are 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] Finally, publicly provided health care is universally available to all residents in Finland.

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 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.[2] In 2016, the total contribution rate under TyEL – including both the employee and employer components – was 23.6% of earnings.[3]

Self-employed individuals, as well as 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

[1] 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.

[2] The income base for TyEL contributions consists of all wage and salary compensations, including e.g. fringe benefits and bonuses. Typically the TyEL income base closely follows the income tax base for earnings, but some items are excluded from TyEL such as options and travel allowances.

[3] 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.

owners of unincorporated partnership firms who, alone or together with family members, *own at least a specified share of their business.* To qualify for YEL coverage, entrepreneurs must 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 must exceed €7,557 (in 2016). In addition, owners of privately held corporations who hold a leading position in the firm (such as CEO or chair of the board) and own at least the required minimum share are also classified as YEL entrepreneurs. All these conditions are binding, and entrepreneurs cannot opt out of the YEL scheme. Entrepreneurs who do not meet the YEL criteria are automatically covered by the TyEL insurance scheme, the same social insurance system that applies to wage earners.

The main difference between the TyEL and YEL schemes is that while the TyEL-covered owners have little discretion over their social insurance contributions, YEL-covered owners can freely set their contribution level. For TyEL-covered individuals, both contributions and benefit entitlements are automatically determined based on their wages. As a result, TyEL-covered owners can only reduce social insurance contributions by lowering own wages and instead paying out dividends, which generally leads to costly income tax implications.[4] In contrast, YEL-covered owners simply report an income level at which they wish to be covered to the insurance company that administers their contributions.[5] Social insurance contributions are then paid based on this reported income, with no automatic adjustments in response to earnings fluctuations. The reported income must fall within a specific range (in 2016, between €7557 and €171,625), and the owners can adjust their reported incomes at any point in time.[6] The contribution rate applied to the SI income under YEL is broadly comparable to that under TyEL – 23.7% versus

[4]The optimal payout structure involves a mixture of wages and dividends, with wages accounting for a larger share, but neither form strictly dominating the other. Empirically, wages account for approximately 87% of business payouts. Importantly, these income tax incentives are the same for both TyEL and YEL owners.

[5]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, which would ensure that social insurance coverage provides adequate protection against earnings losses in an adverse event. In practice, however, this rule is not enforced, primarily because such a counterfactual wage is hard to approximate.

[6]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.

23.6% in 2016.[7] 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 owners. Since future benefit entitlements are directly linked to the level of contributions, any adjustment in self-reported social insurance 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 determined based on wage earnings, leaving less scope for discretion. For this reason, we refer to all entrepreneurs to whom YEL rules apply as *unrestricted* owners, and all entrepreneurs to whom TyEL rules apply as *restricted* owners in our analysis. We will further refer to the reported income of YEL owners and the income base for social insurance contributions (i.e. wages) of TyEL owners as the *social insurance (SI) income*, and to the overall earnings – wage plus dividend income from the firm – of both groups as their *total income (TI)*. We then measure *SI coverage* for both YEL and TyEL owners as their SI income over their total income (SI/TI).

Ownership Share Thresholds and the 2011 YEL Reform. Eligibility for the YEL insurance scheme requires that an entrepreneur, 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.[8] 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 who remained TyEL owners (10–29%) both before and

[7] YEL contribution rates, as well as the minimum and maximum SI income thresholds, are adjusted annually. Similary to TyEL, 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.

[8] 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).

after the reform, discussed in more detail in Section 4.1.[9]

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

Retirement benefits and *disability benefits* are determined based on the total accumulation of inflation-adjusted 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 when in the working career those contributions are made.[10]

Pension benefits are likely the most salient component of social insurance for entrepreneurs. First, most entrepreneurs will eventually consider retirement. 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. For this reason, pensions are the central focus of our analysis.

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 maternity leave, and lasts until the child reaches nine months of age.[11] The income base for parental leave benefits is defined by the SI income

[9]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 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 to all new entrepreneurs 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.

[10]For example, if an entrepreneur reports an (inflation-adjusted) SI income of €10,000 for 15 years and €30,000 for 20 years, the approximate annual pension would be $0.015\times(15\times$€$10{,}000 + 20\times$€$30{,}000)=$€$11{,}250$.

[11]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 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

12 months prior to claiming the benefit, in contrast to the accumulated income over the working career used for pension benefits. In 2012, the flat minimum benefit for those with income smaller than €9841 was €22.96 per day, and the replacement rate was 70% for SI incomes up to €34,496.[12] For SI incomes exceeding this threshold, the replacement rate declined gradually. As an example, an individual with an SI income of €50,000 would have received a parental allowance of €101.16 per working day in 2012, with an average replacement rate of 60.7%.

Sick leave benefits can be claimed when individuals experience a long-term health issue. Eligibility is contingent on a medical review. In 2007–2014, the replacement rates and other regulations for sick leave followed similar rules as those for parental leave benefits. 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 only face 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.

Entrepreneurs are also entitled to *unemployment benefits* under certain conditions. However, we exclude unemployment benefits from our analysis for several reasons. 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. Unfortunately, our data do not allow us to identify unemployment fund members, preventing a focused analysis of entrepreneurs eligible for earnings-related unemployment benefits. While over 70% of wage earners in Finland belong to unemployment funds, the membership rate among entrepreneurs is substantially lower.[13] As a result, for the vast majority of entrepreneurs, unemployment benefits do not depend on their selected SI income, implying that choices of SI income

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.

[12]In 2007–2014, the benefit regulations remained similar, but the income thresholds and minimum benefits were adjusted annually according to increases in the price index.

[13]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.

based on unemployment risk are largely irrelevant. 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 that we can link together: (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 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 insured under the YEL scheme.[14] 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.

[14]We do not view entrepreneurial selection into pension companies as a significant concern. Although individuals are free to choose the pension company through which they channel their contributions, statutory contribution rates and the resulting pension benefits are fully determined by law and are uniform across companies. Consequently, no financial advantage can be gained from selecting one pension company over another, as firms are prohibited from tailoring contributions or benefit accrual to individual or entrepreneurial characteristics.

Benefit Claims Data. We observe all earnings-related benefits claimed by TyEL- and YEL-covered entrepreneurs. 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 claim 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 entrepreneurs 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 entrepreneurs under the TyEL pension scheme.

Survey Data. We complement the administrative data with information from two distinct surveys conducted by Statistics Finland. The first survey, *Itsensätyöllistäjät Suomessa* (Pärnänen and Sutela, 2014), was conducted in 2013 and targeted entrepreneurs and the self-employed individuals, and yielded 1,573 responses. It collected data on working conditions, perceived business potential, income levels, and experiences with the social insurance system for entrepreneurs. The second survey, *Yrittäjyys Suomessa* (Pärnänen and Sutela, 2017), was conducted in 2017 and included 1,072 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 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, which allows to make sense of 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. All TyEL entrepreneurs 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 entrepreneurs 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).

Second, we exclude individuals who own multiple firms from the sample. However, our results are robust to including them, as only 7.3% of entrepreneurs in our data simultaneously own more than one firm. Third, our baseline sample includes all TyEL and YEL-insured entrepreneurs whose total income (labor and capital income) falls between 1.2 times the minimum YEL contribution and the maximum YEL contribution divided by 1.2. We use this sample to examine both the overall choice of SI coverage level 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.

3.2 Descriptive Statistics and Social Insurance Choices

Descriptive Statistics. Table 1 reports descriptive statistics covering years 2008–2010 – the years preceding the YEL reform of 2011. Column (1) shows these statistics for our baseline sample including all entrepreneurs (both YEL and TyEL). Column (2) includes the owners of privately held corporations who transitioned from the TyEL to the 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 groups. Variables that most clearly differ between the groups are 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 the always restricted owners and the treated group prior to the reform, compared with the self-reported social insurance 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. Otherwise, the groups appear to be quite similar, especially in terms of income levels and industry distribution.

Social Insurance Choices. Figure 1(a) provides a detailed comparison of the social insurance choices of restricted and unrestricted entrepreneurs. The figure depicts the average SI income within each 2.5% bin of labor income for both groups. For restricted owners, SI income aligns closely with the 45-degree line, which is expected given that the legal definition of SI income is tightly linked to reported labor income.[15] Among

[15] As noted in Section 2, some smaller labor income items are excluded from TyEL, meaning that the SI income and reported labor income of restricted owners do not perfectly align with the 45-degree line in the figure.

unrestricted owners, SI incomes exceed those of restricted owners at lower income levels, reflecting the binding minimum SI income requirement under the unrestricted scheme. At higher income levels, by contrast, unrestricted owners select markedly lower SI coverage than restricted owners. Importantly, although their SI income choices are substantially lower, they do not converge to the regulatory minimum.[16]

In addition, Appendix Figure A.2 shows the share of reported social insurance income within the distribution of total income, disaggregated by individual characteristics. The figure shows limited gender heterogeneity in SI income (panel a) and parental status (panel b). In contrast, Figure A.2 shows that the share of social insurance income is higher among older owners compared to younger ones (panel c) and among married entrepreneurs compared to their unmarried counterparts (panel d).

To shed light on the drivers of observed SI income choices, Figure 1(b) combines survey responses about the motivations behind SI contribution choices among unrestricted owners with administrative register data. The survey results reveal a range of contributing factors for low SI coverage, and no single reason clearly stands out as the most significant explanation. (Note that each respondent could pick several reasons, so total shares sum up to more than 1.) In particular, plans to work during retirement (65% of respondents) 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 seldom chosen. Importantly, 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 suggests that moral hazard and adverse selection effects are unlikely to be first order drivers of SI choices.

Figure 1(b) also shows the average SI income relative to total income (SI/TI), estimated using the administrative data. Despite the widely differing self-reported expla-

[16] Additionally, Appendix 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).

nations for low SI coverage, actual SI choices are strikingly similar within each category, even when controlling for observable characteristics (age, gender and total income).

Overall, Figure 1 demonstrates that, when given an opportunity to choose SI coverage level, owners make an active choice that does not default to the legal minimum but falls below the level mandated for wage earners, and that this choice does not appear to be influenced by moral hazard and adverse selection considerations.

4 Moral Hazard and Adverse Selection Estimates

4.1 Empirical Design

To estimate moral hazard responses, we implement an event-study design around the 2011 reform. The key idea is that the reform induced treated entrepreneurs to adjust their contribution levels without changing their underlying risk type. As a result, any corresponding changes in benefit claims capture moral hazard rather than adverse selection. We discuss the identifying assumptions in Section 4.3.

Our event studies estimate the following equation:

$$Outcome_{it} = \sum_{\substack{\ell=2007 \\ \ell \neq 2010}}^{2014} \beta_\ell \; Treat_i \; 1_{t=\ell} + Treat_i + \eta_t + \delta X_{it} + \varepsilon_{it}, \tag{1}$$

where the i identifies individual entrepreneurs 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 this equation for years 2007–2014.

The treatment group consists of entrepreneurs who hold more than 30% but less than 50% of their company shares prior to the 2011 reform.[17] These are the en-

[17]The nature of the reform could, in principle, allow for a regression discontinuity design (RDD) rather than a difference-in-differences approach. In practice, however, implementing an RDD is difficult. First, ownership shares are unevenly distributed, with large spikes at common shares—such as 25%, 33%, and 50% for firms with four, three, and two owners—which complicates bandwidth selection and undermines the smoothness assumptions required for RDD. Second, RDD is data-intensive, and the

trepreneurs who are now able to adjust their contribution levels. 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 allow them to choose SI contribution levels. The second control group – referred to as the *always unrestricted* – consists of owners with 50–70% ownership share, who retain full discretion over their SI contributions. Results are quantitatively and qualitatively similar irrespective of the control group used.[18] Importantly, our event-study design is not subject to the concerns raised in recent critiques of two-way fixed effects estimators because the treatment is assigned uniformly in the same calendar year for all treated units (de Chaisemartin and D'Haultfoeuille, 2020; Sun and Abraham, 2021; Callaway et al., 2021; Goodman-Bacon, 2021).

A potential concern for identification is that ownership shares may have been adjusted in response to the 2011 reform, allowing entrepreneurs to self-select into the restricted or unrestricted insurance schemes. However, we find no empirical evidence supporting this possibility. Appendix Figure A.3 plots the share of owners whose ownership share changed from below 30% to above 30%, and from above 50% to below 50%, over the period 2007–2015. Such changes are generally uncommon—fewer than 3% of owners crossed either threshold—and, crucially, there is no discernible shift in these patterns around the time of the reform. This evidence alleviates concerns that owners strategically manipulated their ownership shares in response to the policy change.

number of owners near the 30% threshold is too small to yield precise estimates. For these reasons, we rely on a difference-in-differences design, which—given the parallel pre-trends in key outcomes between treatment and control groups—provides more robust and reliable estimates in this setting

[18]Note that, on the one hand, the always restricted group is a superior control group because this group remains untreated throughout the sample period and thus satisfies the identifying assumptions of a standard difference-in-differences (DiD) framework which requires treatment and control individuals to experience the same treatment status in the baseline period (Tazhitdinova and Vazquez-Bare, 2023). On the other hand, only the always unrestricted group allows for accurate heterogeneity analysis by SI/TI levels.

4.2 Changes in SI Contributions Following the 2011 Reform

First, we study the extent to which treated owners adjust their SI income, and thus their insurance coverage, once they are allowed to do so after 2011. Figure 2(a) plots the evolution of SI income relative to total income (SI/TI) for the treatment group and both control groups. Pre-reform trends are nearly identical across all three groups, providing strong support for the parallel trends assumption. Following the reform, however, the trajectories diverge sharply: treated owners substantially reduce their contribution levels relative to both control groups. This reduction among the treatment group brings their contribution levels close to those of entrepreneurs who already had full discretion prior to the reform.[19]

Figure 2(b) plots the corresponding event-study estimates with 95% confidence intervals using both control groups. The results indicate that SI/TI decreased on average by approximately 13 percentage points in the treatment group relative to either control group. Given the pre-reform mean of 0.83 in the treatment group, this corresponds to an average decrease in SI/TI of approximately 16%.

In sum, event studies confirm descriptive evidence from Section 3.2: when given more discretion over SI contributions, average owner chooses lower level of coverage than mandated for wage earners or restricted entrepreneurs.

4.3 Moral Hazard (Non-)Response Following the 2011 Reform

In the presence of moral hazard, a reduction in SI coverage should induce a corresponding decline in SI benefit claims. In this section, we use the event-study framework described in Section 4.1 to examine whether the 16% decrease in average SI contributions documented in Section 4.2 translated into a measurable reduction in subsequent claiming behavior.

Interpreting changes in benefit claims as a measure of moral hazard requires two assumptions. First, underlying risk types must evolve similarly over time in the treatment and control groups. Second, changes in moral hazard responses *induced by*

[19] Appendix Figure A.4 shows analogous evidence using SI income relative to *labor* income, yielding results that closely mirror our baseline approach with total income as the denominator.

underlying risk type changes must also evolve similarly over time in the treatment and control groups, or individuals' moral hazard responses must be uncorrelated with their risk types. To a large extent, both of these assumptions are implied by the standard DiD assumption: for our outcome (benefit claims) to follow a similar trend, claimants' risk types and risk-type-induced moral hazard responses must evolve similarly. Since we are comparing similar groups of business owners (recall Table 1), both assumptions are plausible. Indeed, Figures 2 and 3 provide further support for these and the standard DiD assumption, by showing that SI coverage levels and benefit claims evolved similarly prior to the reform. Our analysis also controls for observable characteristics such as age, gender, and industry that are likely to influence one's risk type and/or moral hazard capacity.

To see why these assumptions are necessary, note that realized SI claims reflect two potentially correlated components: an individual's underlying risk type and their moral-hazard behavior. The 2011 reform does not affect underlying risk types and therefore should not influence SI claims through this channel. The first assumption thus ensures that we can account for any exogenous changes in risk types via our control group. By contrast, the reform altered individuals' SI coverage levels which may have affected the moral-hazard component of their claims. If moral hazard responses are correlated with risk types, then the total magnitude of moral hazard responses will further be affected by changes in risk types. Accordingly, we can interpret our event study estimates as reflecting total moral-hazard responses as long as risk types are unrelated to moral-hazard behavior. If moral hazard effects are dependent on one's risk type, then our event studies estimate changes in moral hazard induced by the reform but do not account for moral hazard changes due to changes in risk types over time.[20]

[20]To see this, suppose person i's claims are given by $Claim_t^i = True(Risk\_type_t^i) + MH(Risk\_type_t^i, SI_t^i)$. If the first assumption is satisfied, the event studies estimate $DiD = [MH(Risk\_type_{post}^T, SI_{post}^T) - MH(Risk\_type_{pre}^T, SI_{pre}^T)] - [MH(Risk\_type_{post}^C, SI_{post}^C) - MH(Risk\_type_{pre}^C, SI_{pre}^C)]$, since the risk based component cancels out. If moral hazard effects are uncorrelated with risk types, then $DiD = MH(SI_{post}^T) - MH(SI_{pre}^T)$, since the control group experiences no changes in moral hazard responses because their SI coverage levels do not change. If risk type and moral hazard effects are correlated, then $DiD = [MH(Risk\_type_{post}^T, SI_{post}^T) - MH(Risk\_type_{post}^T, SI_{pre}^T)] - [MH(Risk\_type_{post}^T, SI_{pre}^T) - MH(Risk\_type_{pre}^T, SI_{pre}^T)]$, as long as $MH(Risk\_type_{post}^T, SI_{pre}^T) - MH(Risk\_type_{pre}^T, SI_{pre}^T) = MH(Risk\_type_{post}^C, SI_{pre}^C) - MH(Risk\_type_{pre}^C, SI_{pre}^C)$, i.e. moral hazard responses induced by underlying risk type changes evolve similarly between the two groups.

Figure 3 presents event-study estimates showing how benefit claims changed in the treatment group relative to the always-unrestricted (ownership share 50–70%) control group. We present this evidence separately for three groups of treated individuals: those whose post-reform contributions SI/TI fell below 0.5, those with post-reform SI/TI between 0.5 and 0.9, and those with post-reform SI/TI above 0.9. We do so because the aggregate 16% SI contribution decrease hides notable heterogeneity in SI choices (see Appendix Figure A.1). This breakdown enables us to separately estimate the responses of individuals who experienced a substantial change in contributions and those who experienced little to no change.

Panel (a) of Figure 3 shows changes in the total number of days claimed, across all benefit types, while panels (b) – (d) show estimates for specific benefits types – pension benefits (panel b), sick leave (panel c), and parental leave (panel d). Figure 3 shows no systematic increases or decreases in SI claims for any of the subgroups or benefit types. We see no changes in claims among individuals who dramatically decreased their SI contributions as a result of the 2011 reform, nor among those who left SI contributions unchanged or increased them. Table 2 shows DiD estimates corresponding to the event studies shown in Figure 3.

Our results demonstrate lack of economically meaningful moral hazard behavior: despite substantial changes in SI contributions, there are no corresponding changes in benefit usage. As a result, lower contributions do not translate into fewer claims, nor do higher contributions lead to larger or more numerous claims.

5 Estimates of Adverse Selection

In the presence of adverse selection, individuals sort based on their underlying risk types. When risk types are stable, adverse selection appears as a positive correlation between individuals' claim rates and their chosen coverage levels. When risk types vary but changes are predictable, it also manifests through anticipatory adjustments in coverage prior to foreseeable claim events. In this section, we examine whether unrestricted owners adjust

their SI income in anticipation of childbearing, sickness, and retirement. By restricting attention to individuals who ultimately claim these benefits—and by focusing on the level of contributions rather than claim duration—we shut down the moral hazard margin and isolate adverse selection responses.

For our test to be valid, individuals must be able to foresee a claim event and adjust their contributions in a timely manner. This condition is easily met for both retirement and parental leave. Pension benefits depend on one's full history of contributions, allowing individuals to modify their contribution levels well in advance of retirement. Parental leave benefits, in turn, are based on SI contributions during the 12 months preceding the leave spell. Because leave is typically taken once the child is approximately three months old, expecting parents have ample time to adjust their contributions ahead of future claims. Formally, we estimate regressions of the following form for all unrestricted owners:

$$SI_{it}/TI_{it} = \alpha + \beta_{-2} \cdot t_{-2} + \beta_{-1} \cdot t_{-1} + \beta_0 \cdot t_0 + \beta_1 \cdot t_1 + \beta_2 \cdot t_2 + X_{it} + \varepsilon_{it}, \qquad (2)$$

where the coefficients β_j, $j = -2, ..., 2$ correspond to insurance contributions relative to total income (SI_{it}/TI_{it}) before and after a claiming event occurring at time $t = 0$. 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. We estimate a similar model for retirement decisions using data from up to seven years before retirement. However, once an individual retires, they typically stop SI contributions, and thus post-event indicators are not available for this outcome. Our analysis controls for age, gender, municipality, industry, income and year fixed effects.

Figure 4 presents the results. For retirement (panel a), the ratio of SI income to total income remains approximately constant over the seven-year period preceding the start of pension benefit claims. This pattern implies that unrestricted entrepreneurs do not increase their reported SI income in the final years before retirement in an attempt to raise future pension benefits (conditional on age). Such behavior is largely consistent with the institutional setting, as all SI contributions made during the working career

affect future pension benefits at an accrual rate of 1.5%. However, the higher accrual rate of 4.5% for individuals who continue working beyond the age of 63 should incentivize increased SI contributions among those who postpone claiming their pensions. Nevertheless, Figure 4(a) show that this provision had no visible effect on overall SI contributions prior to the onset of pension benefit claims.

The results for sick leave and parental leave benefits are shown in panels (b) and (c) of Figure 4, where the we plot SI_{it}/TI_{it} relative to year $t-3$, i.e. three years before we observe an individual claiming a benefit. Once again, we see minimal evidence of anticipatory responses. The estimated coefficients are small in magnitude and not statistically different from each other over time. We observe that SI_{it}/TI_{it} remains approximately constant in the years preceding a claiming event.

Appendix Figure A.5 provides additional evidence consistent with the absence of anticipatory behavior. The figure plots the share of owners who increase or decrease their SI contributions around claiming events. In line with Figure 4, which shows no meaningful changes in the *level* of SI contributions, Figure A.5 likewise reveals no discernible shifts in the *propensity* of adjusting one's contributions.

A limitation of our analysis is that we lack a control group and therefore cannot isolate naturally occurring income changes around claim events. Yet any such changes would have to be matched by adjustments in social insurance income to produce the observed null effect, leading to the same qualitative conclusion that SI coverage rates remain stable.

Overall, our results show that individuals do not try to "game" the SI system by strategically increasing their SI income in years preceding benefit claims. The observed pattern is consistent with the absence of adverse selection effects.

6 Joint Identification of Adverse Selection and Moral Hazard Using Positive Correlation Tests

In this section, we quantify the overall relationship between coverage levels and subsequent claims using positive correlation tests (Chiappori and Salanie, 2000). These tests capture the joint effect of moral hazard and adverse selection, as both mechanisms imply a positive association between chosen SI coverage levels and realized claims. Given the evidence presented in Sections 4 and 5, we expect to find little to no relationship between SI coverage levels and subsequent claims. While the test confounds moral hazard and adverse selection effects, a key advantage is that it relies on minimal assumptions, making it a valuable complement to our separate estimates of moral hazard and adverse selection obtained in the preceding sections.

6.1 Estimates of Correlation Between SI Coverage and Claims

We implement positive correlation tests for all unrestricted owners (YEL) in our baseline sample. Individuals are grouped into 5% bins based on their SI income relative to total income SI_{it}/TI_{it}, treating each person–year (i, t) as a separate observation. The structure of our data allows us to accurately link these coverage levels to subsequent claims, thus we can examine how SI coverage choices relate to realized outcomes. Importantly, we control for income in all specifications, as the generosity of certain benefits (sick leave and parental leave) decreases with income. We also control for age, gender, municipality, and industry.

Figure 5 summarizes the relationship between SI choice and several outcomes realized in the future: average retirement age (panel a), predicted life expectancy (panel b)[21], annual sick-leave days claimed (panel c), and annual parental-leave days claimed (panel d). Overall, Figure 5 shows weak but generally positive correlation between SI choices and claims. Panel a shows that individuals with higher SI_{it}/TI_{it} are more likely to retire younger. Similarly, Panel b shows that those who contribute more tend to have

[21] Appendix B provides details on how we construct predicted life expectancy using our register data.

longer predicted life expectancy. Therefore, both of these measures imply that those with a higher SI contribution rate end up collecting slightly more pension benefits compared to those with smaller SI_{it}/TI_{it}.

Panels c and d show that higher SI_{it}/TI_{it} are also associated with more sick leave and parental leave days claimed. Moreover, the correlation between claims and SI coverage is arguably 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 correlation is much smaller for those with SI_{it}/TI_{it} between 0.3 and 1.

Figure 6 provides similar evidence but relies on survey and shows the relationship between entrepreneurs' self-reported health and their SI contribution levels. In the survey, the respondents were asked to evaluate their own perceived health status on a scale from 0 to 10. The results reveal no systematic relationship between self-assessed risk type and coverage choices. However, because the survey samples are very small, the results are noisy.

6.2 Economic Magnitude of the Positive Correlation Tests

The magnitudes of the correlations seen in Figure 5, and summarized in Panel A of Table 3, are economically small. Figure 5 shows that increasing coverage from SI_{it}/TI_{it}=0.2 to SI_{it}/TI_{it}=1.2 is (on average) associated with: (a) claiming pensions 7 months earlier because of earlier retirement; (b) claiming pensions for 9.5 months longer because of higher life expectancy; (c) claiming 1.13 more sick pay days; and (d) claiming 0.67 more days of parental leave. In this section, we describe how we convert these estimates into economic costs (which we summarize in Panel B of Table 3) in order to assess their economic relevance and magnitude.

We start by evaluating the selection costs associated with pension benefits, as these account for the largest share of SI expenditures. Note that the insurance payouts for individuals with $SI_{it}/TI_{it} = 1.2$ are higher than those for individuals with $SI_{it}/TI_{it} = 0.2$ for two main reasons. First, social insurance payouts increase mechanically with one's contributions, which may lead to higher or lower net costs to the social insurance system

depending on whether it is actuarially fair and/or progressive. Importantly, we are not considering this effect, since our focus is solely on moral hazard and adverse selection rather mechanical effects. Second, payouts increase as a result of adverse selection and moral hazard, since as shown in Figures 5(a) and 5(b), individuals with higher SI contributions retire earlier and live longer. Let $N(0.2)$ and $N(1.2)$ denote the number of pension-years while $SI(0.2)$ and $SI(1.2)$ denote the average level of SI income, in euros, for entrepreneurs who choose $SI_{it}/TI_{it} = 0.2$ and $SI_{it}/TI_{it} = 1.2$, respectively. Then the difference in payouts between those who choose $SI_{it}/TI_{it} = 0.2$ and those who choose $SI_{it}/TI_{it} = 1.2$ is given by $\Delta_{payouts} = 0.015 \cdot [N(1.2) \cdot SI(1.2) - N(0.2) \cdot SI(0.2)]$, accounting for the 1.5% increase in pension benefits per €1 of SI income increase. For simplicity, this back-of-the-envelope formula disregards discounting as well as cost-of-living adjustments (pension benefits are adjusted annually and received over many years). Since we are only interested in measuring the second channel, we can rewrite $\Delta_{payouts}$ as:

$$\Delta_{payouts} = \underbrace{0.015 \cdot [N(1.2) - N(0.2)] \cdot SI(1.2)}_{\text{adverse selection}} + \underbrace{0.015 \cdot N(0.2) \cdot [SI(1.2) - SI(0.2)]}_{\text{mechanical increase}}. \quad (3)$$

The first term measures the increase in payouts as a result of risk-based selection, while the second term measures the mechanical increase in payouts because of the benefit formula.

Recall that Figures 5(a) and 5(b) implied a total of $N(1.2) - N(0.2) = 1.38$ years increase in pension claim durations, stemming from both earlier retirement (0.59) and longer life expectancy (0.79) among those with higher coverage compared to individuals with lower SI/TI. Furthermore, from Panel B of Table 3, we observe that annual SI income of entrepreneurs with $SI/TI = 1.2$ is $SI(1.2) =$ €54,462. Altogether, the first term implies that expenditures increase by €1,127 as a result of risk-based selection. This increase can be compared to the corresponding increase in SI contributions $\Delta_{contributions} = 0.24 \cdot [SI(1.2) - SI(0.2)] =$ €11,409.[22] This comparison suggests that, for pension benefits, the cost of risk-based selection amounts to approximately 10 cents per

[22]Note that if the pension formula is flat (as is the case in Finland) and actuarially fair, then $\Delta_{contributions}$ would be equal the mechanical increase in payouts.

additional €1 of SI contributions.[23]

To calculate the economic costs of adverse selection and moral hazard for sick pay and parental leave benefits, we estimate the first term in the following equation:

$$\Delta_{payouts} = \underbrace{[N(1.2) - N(0.2)] \cdot BI(1.2)}_{\text{adverse selection}} + \underbrace{N(0.2) \cdot [BI(1.2) - BI(0.2)]}_{\text{mechanical increase}}, \quad (4)$$

where $BI(0.2)$ and $BI(1.2)$ measure the daily sick pay and parental leave benefit rates when $SI_{it}/TI_{it} = 0.2$ and $SI_{it}/TI_{it} = 1.2$, respectively, while $N(0.2)$ and $N(1.2)$ measure the number of days each of these benefits are claimed. Note that the second term is not zero because of the progressive nature of the benefits: individuals with high SI income qualify for lower replacement rates than individuals with low SI incomes. Figures 5(c) and 5(d) imply $N(1.2) - N(0.2) = 1.13$ days for sick pay and $N(1.2) - N(0.2) = 0.67$ days for parental leave. As an illustration, we calculate daily $BI(1.2) = €106.50$ using the 2012 benefit formula.[24] Formulas for other years follow a similar structure. Our calculations imply a selection cost of €120 for sick pay and €71 for parental leave. Consequently, the costs of risk-based selection for sick leave and parental leave are an order of magnitude smaller than those associated with pension benefits.

In sum, our estimates imply that the costs of selection by risks are small in this setting. The positive correlation tests imply that the combined effects of adverse selection and moral hazard are likely to be limited among entrepreneurs, despite their discretion over social insurance coverage.

6.3 Why Are Costs Small?

Taken together, our results indicate that adverse selection responses are limited and moral hazard responses are virtually absent among entrepreneurs, despite substantial discretion in setting their level of SI coverage. In this section, we examine some plausible

[23] Assuming that individuals increase their SI income equivalently for more than one year does not alter this calculation. Extending the time horizon scales both pension benefits and SI contributions proportionally, so the ratio between the two remains unchanged.

[24] In 2012, the first €34,496 of SI income qualified for the 70% replacement rate, incomes between €34,496 and €53,072 qualified for the 40% replacement rate, and the replacement rate was 25% for incomes above €53,072. The benefits are paid for 300 days per year.

explanations for these findings.

Knowledge of Rules and Incentives. A possible explanation of our findings is that entrepreneurs may lack a clear understanding of the SI system and the relationship between reported SI income and benefit entitlements. Social insurance rules are complex, and the benefits of higher SI coverage accrue mainly in the future, while the costs are immediate.

Figure 7 shows survey-based evidence linking entrepreneurs' knowledge and perceptions of SI benefits to their actual SI contribution choices observed in the administrative data. In the survey the respondents were asked their awareness of the SI benefits they are entitled to on a scale from 1 ("very well") to 4 ("not at all"). Panel (a) shows that only a small share of respondents report knowing "very well" what benefits they are entitled to, suggesting that entrepreneurs have limited understanding of the incentives they face. Furthermore, we see no strong association between self-reported knowledge and SI coverage: individuals reporting higher knowledge do not systematically choose higher or lower SI/TI ratios.

Additionally, the respondents were asked whether they are aware of the amount of sick leave benefit they would be entitled to. Panel (b) shows a similar pattern as above: although most respondents report being unaware of the benefit level, the average SI/TI differs only slightly between those who say they are aware of it and those who are not, and these differences are not statistically significant. Taken together, Figures 7 (a) and (b) suggest that, while informational limitations are common, they are unlikely to be the primary driver of the observed heterogeneity in SI coverage.

Learning from Experience. For adverse selection and moral hazard responses to be feasible, individuals must be aware of the rules and of their own risk types. The previous section showed that, on average, entrepreneurs have poor self-reported understanding of the SI benefit rules. However, as individuals interact with the SI system, they may improve both their understanding of the rules and update their beliefs about their own risk type. If learning is an important mechanism, one would expect a stronger correlation

between coverage levels and claims as time progresses.

Figure 8 compares the correlation patterns between SI coverage choices and all future claims for treated individuals – those newly eligible for the unrestricted status in 2011 – in 2011 versus in 2014. The figure shows a small increase in positive correlation over time, consistent with learning. However, the relationship appears to be U-shaped in 2014, with higher claims observed both very low and high contribution levels. Low contributors may be liquidity constrained and face higher underlying risks, leading to greater claims. Conversely, high contributors may file more claims after learning about the extent to which contributions entitle them to benefits, therefor increasing claims.

Retirement Planning and Long-Term Expectations. Survey evidence reported in Figure 1(b) shows that a variety of motivations unrelated to adverse selection influence SI coverage choices. Many entrepreneurs report liquidity constraints, intentions to rely on future work or business sale for retirement income, and limited trust in the SI system. These factors may reduce the perceived value of SI coverage relative to immediate income needs or alternative savings and insurance arrangements.

Figure 9 sheds light on entrepreneurs' long-term retirement planning and how it relates to SI coverage decisions. The surveys asked whether the respondents planned to continue working after retirement age, and what was their primary motivation behind this plan, including preferences for retirement, and financial and health-related reasons. The figure shows that a substantial share of respondents report intending to continue working beyond the statutory retirement age, mostly because of preferences for working after retirement. However, those who stated financial issues as their primary reason for working after retirement had slightly lower average SI contributions than the others, suggesting that these individuals are relying on smaller public pensions. Relatedly, those who stated that they do not plan to work after the retirement age because of health reasons had slightly larger average SI coverage, compared to those who plan not to work after the retirement age because of preferences for retiring. The differences in SI contributions between these groups, however, remain relatively small in magnitude. Overall, Figure 9 suggests that long-term retirement expectations contribute to heterogeneity in SI choices,

but these expectations do not generate dramatic differences in contributions.

Overall, the evidence indicates that entrepreneurs' SI coverage decisions are only weakly correlated with individual risk exposure, both ex-ante (perceived or expected risks) and ex-post (realized claims). Information gaps and limited learning likely play a role, but they do not fully explain the observed variation. Instead, broader financial constraints, preferences, and alternative planning strategies appear to be more important drivers of low SI coverage and lack of selection responses.

7 Conclusion

This paper examines the social insurance choices of entrepreneurs in Finland and the associated moral hazard and adverse selection costs. We show that entrepreneurs with discretion over their social insurance contributions choose lower levels of coverage as compared to entrepreneurs who face a strict mandate. However, we find that the level of coverage is only weakly correlated with claiming behavior (retirement, sick leave and parental leave) and that entrepreneurs did not change their claiming behavior in response to a large exogenous shock in social insurance contributions. Our results imply small levels of adverse selection and moral hazard in this setting.

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.

Cabral, M., C. Cui, and M. Dworsky (2022): "The demand for insurance and rationale for a mandate: Evidence from workers' compensation insurance," *American Economic Review*, 112, 1621–1668.

Cabral, M. and M. Dillender (2024): "The impact of benefit generosity on workers' compensation claims: Evidence and implications," *American Economic Journal: Applied Economics*, 16, 436–481.

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., 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.-A. and B. Salanie (2000): "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.

McCarthy, D. and O. S. Mitchell (2010): *International Adverse Selection in Life Insurance and Annuities*, Dordrecht: Springer Netherlands, 119–135.

Pärnänen, A. and H. Sutela (2014): "Itsensätyöllistäjät Suomessa 2013," Mimeo, Statistics Finland.

——— (2017): "Yrittäjät Suomessa 2017," Mimeo, Statistics Finland.

Seibold, A., S. Seitz, and S. Siegloch (2025): "Privatizing disability insurance," *Econometrica*, 93, 1697–1737.

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	184,660	58,834	18,195	107,631
SI contributions	7611.6	6865.4	5578.7	8311.6
	(4894.5)	(4272.2)	(3927.4)	(5190.0)
SI income	36,049.0	33,387.3	33,918.5	37,864.2
	(22,372.1)	(19,975.1)	(20,688.7)	(23,671.2)
Total income (TI)	43,358.9	42,616.4	43,500.1	43,740.9
	(24,705.8)	(24,024.9)	(25,079.3)	(24,998.7)
SI/TI	0.84	0.83	0.71	0.87
	(0.20)	(0.19)	(0.29)	(0.18)
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 presents the descriptive statistics in our pre-reform period (years 2008-2010) for individuals whose total income falls between 1.2 times the minimum YEL contribution and the maximum YEL contribution divided by 1.2. All euro values are expressed in 2010 euros, adjusted using the Consumer Price Index (CPI). Values are presented in means with standard deviation shown in parenthesis. The first column presents statistics for all entrepreneurs, the second column for our treatment group (ownership share between 0.3 and 0.5), and the next two columns for the two control groups – the always unrestricted (ownership share between 0.5 and 0.7) and the always restricted (ownership share between 0.1 and 0.3).

Table 2: Effects of the 2011 Reform on SI Coverage Levels and Claims

(1)	(2) Average	(3) $SI/TI < 0.5$	(4) $SI/TI \in [0.5, 0.9]$	(5) $SI/TI > 0.9$
SI/TI	-0.126	-0.343	-0.097	0.043
	(0.004)	(0.008)	(0.005)	(0.005)
N	126,194	30,661	56,739	37,962
R2	0.18	0.39	0.10	0.18
Pre-reform mean	0.842	0.842	0.842	0.842
Total Days Claimed	0.149	-0.212	0.161	0.168
	(0.279)	(.542)	(0.392)	(0.595)
N	153,736	37,601	67,110	46,740
R2	0.02	0.01	0.01	0.03
Pre-reform mean	2.00	2.00	2.00	2.00
Indicator of Pension Claiming	-0.003	0.004	-0.007	-0.005
	(0.003)	(0.005)	(0.004)	(0.006)
N	153,736	37,601	67,110	46,740
R2	0.41	0.43	0.41	0.39
Pre-reform mean	0.047	0.047	0.047	0.047
Sick Leave Days Claimed	-0.235	-0.221	-0.162	-0.676
	(0.223)	(0.404)	(0.326)	(0.449)
N	152,731	37,378	66,737	46,363
R2	0.01	0.01	0.01	0.01
Pre-reform mean	1.20	1.20	1.20	1.20
Parental Pay Days Claimed	0.368	-0.025	0.312	0.821
	(0.173)	(0.367)	(0.225)	(0.400)
N	152,452	37,251	66,558	46,399
R2	0.03	0.02	0.02	0.04
Pre-reform mean	0.82	0.82	0.82	0.82

Notes: Table presents DiD estimates of the effects of the 2011 reform on social insurance coverage choice and claims. Column (1) specifies the outcome y_{it} while columns (2)-(4) report estimates of coefficient β from $y_{it} = \beta Treat_i \ 1_{t>=2011} + Treat_i + \eta_t + \delta X_{it} + \varepsilon_{it}$ estimated using 2007-2014 data. The results thus report average treatment effects for years 2011-2014. Standard errors are clustered at the individual level. Column (2) presents the average effect on all treated individuals, while columns (3) through (5) separate treated individuals 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 used for all estimates is the always unrestricted group (50%–70% ownership share). Equivalent event study estimates of equation (1) are shown in Figures 2 and 3. For more details see Section 4.1.

Table 3: Positive Correlation Estimates and the Costs of Risk-based Selection

Panel A: A 100% increase in SI/TI leads to the following changes

	mean (standard error)	N of observations
Retirement Age (years)	-0.59 (0.17)	28,308
Life Expectancy (years)	0.79 (0.15)	113,325
Sick Leave (days)	1.13 (0.18)	126,791
Parental Leave (days)	0.67 (0.18)	126,935

	SI Income of Entrepreneurs with:	
	SI/TI=0.2	SI/TI=1.2
Average SI Income	€6,925	€54,462
Average SI Contributions	€1,662	€13,071

Panel B: Economic costs due to adverse selection and moral hazard for entrepreneurs with SI/TI=1.2 vs SI/TI=0.2.

Total costs in euros:

Pension$^{\alpha}$	Sick Leave$^{\beta}$	Parental Leave$^{\beta}$
€1,127	€120	€71

Total costs as percent of SI contributions:$^{\gamma}$

Pension	Sick Leave	Parental Leave
9.88%	1%	0.62%

Notes: Panel A of the table presents the positive correlation estimates from Figure 5 between the chosen SI coverage (SI/TI) and retirement age, life expectancy, sick leave and parental leave. Panel A also shows the mean SI income for those with SI/TI=0.2 and SI/TI=1.2 in Figure 5 to which the 24% contribution rate applies.

Panel B presents estimates of selection costs as a result of adverse selection and moral hazard for entrepreneurs with SI/TI=0.2 vs SI/TI=1.2. For detailed explanations, see Section 6.2.

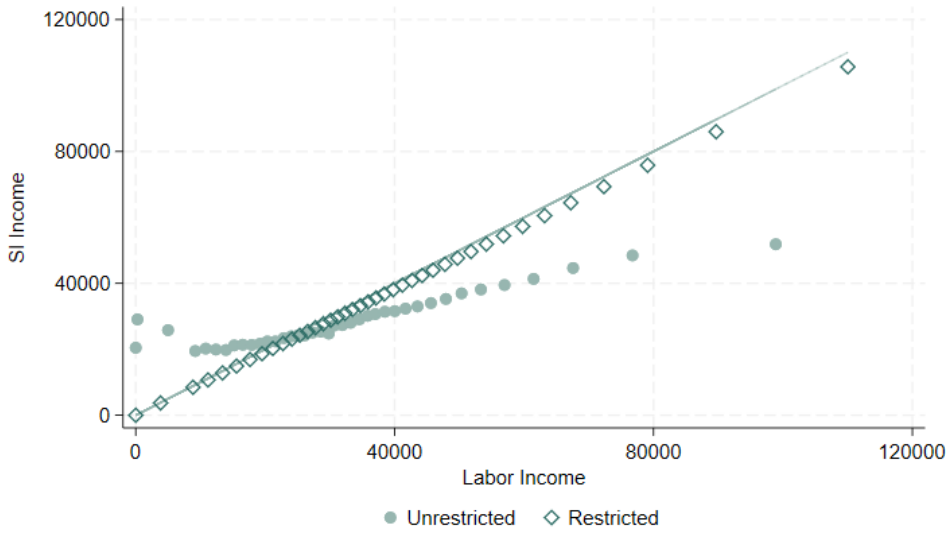
$^{\alpha}$ Selection costs for pension benefits are calculated as $0.015 \times$€$54{,}462 \times (0.59 + 0.79) =$ €1,127, where 0.015 reflects the 1.5% per euro accrual rate, €54,462 is the level of SI income for entrepreneurs with SI/TI=1.2, and $0.59 + 0.79$ is the increase in the time pension benefits are collected, stemming from earlier retirement and longer lifespan, per our estimates from Panel A.

$^{\beta}$ Selection costs for sick pay are calculated as $1.13 * BI(1.2)$ and for parental leave as $0.67 * BI(1.2)$, where 1.13 is the estimated increase in sick leave days per year (see Panel A), and 0.67 is the estimated increase in parental leave days per year (see Panel A), while $BI(1.2)$ reflects the daily benefit rate calculated using the 2012 benefit formula: $BI(1.2) = 1/300 * [0.7 \times 34496 + 0.4 \times (53072 - 34496) + 0.25 \times (54462 - 53072) \times 1.1] =$ €106.5. The above benefit formula reflects 70%, 40% and 25% replacement rates for incomes below €34,496, between €34,496–€53,072, above €53,072, respectively. Sick pay and parental leave benefits are paid for 300 days per year.

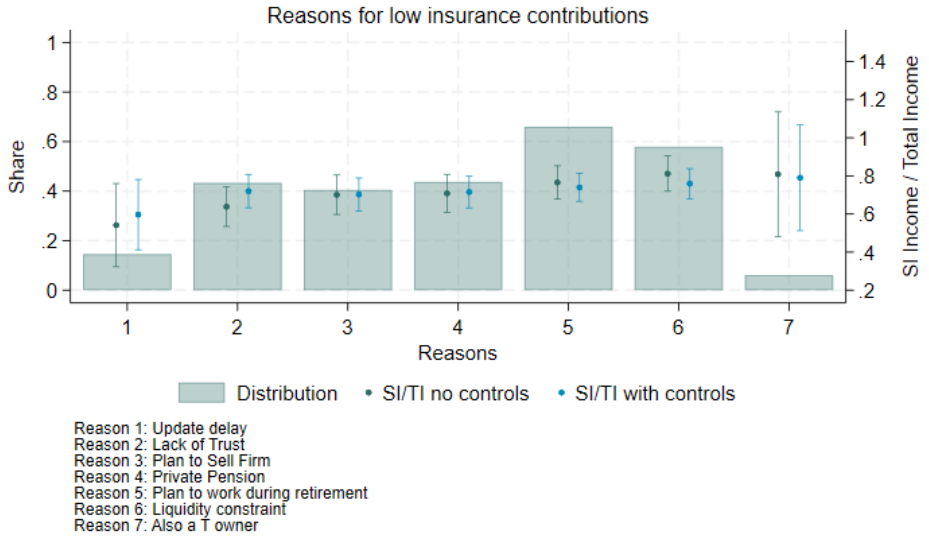
$^{\gamma}$ Total costs as percent of SI contributions are calculated by dividing the total costs in euros by the change in SI contributions of individuals with SI/TI=1.2 versus SI/TI=0.2, which is €11,409.

Figure 1: Choice of Social Insurance Coverage

(a) SI Income Choice: Restricted vs Unrestricted



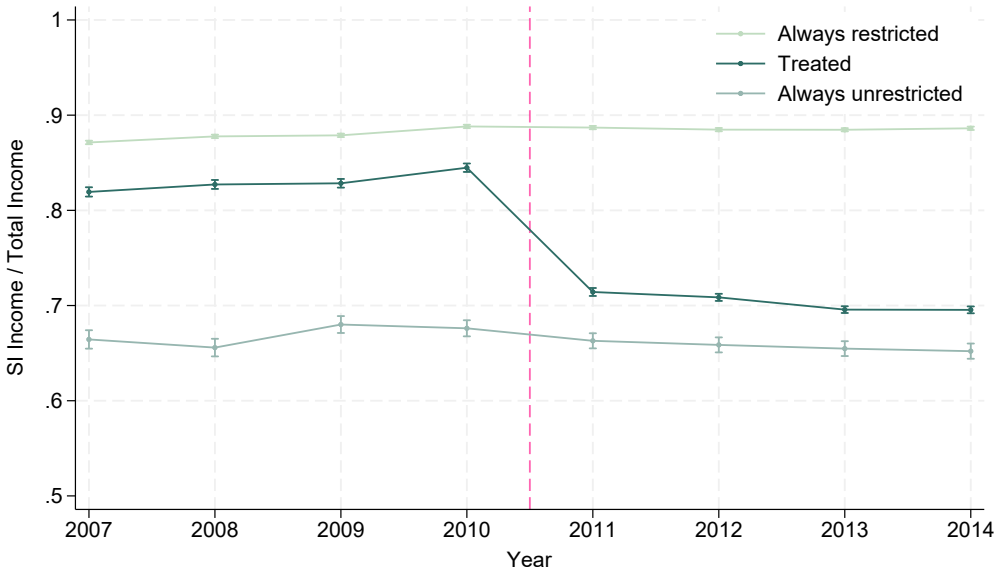
(b) Actual SI Income Choice vs. Stated Reasons for Low Insurance Choice



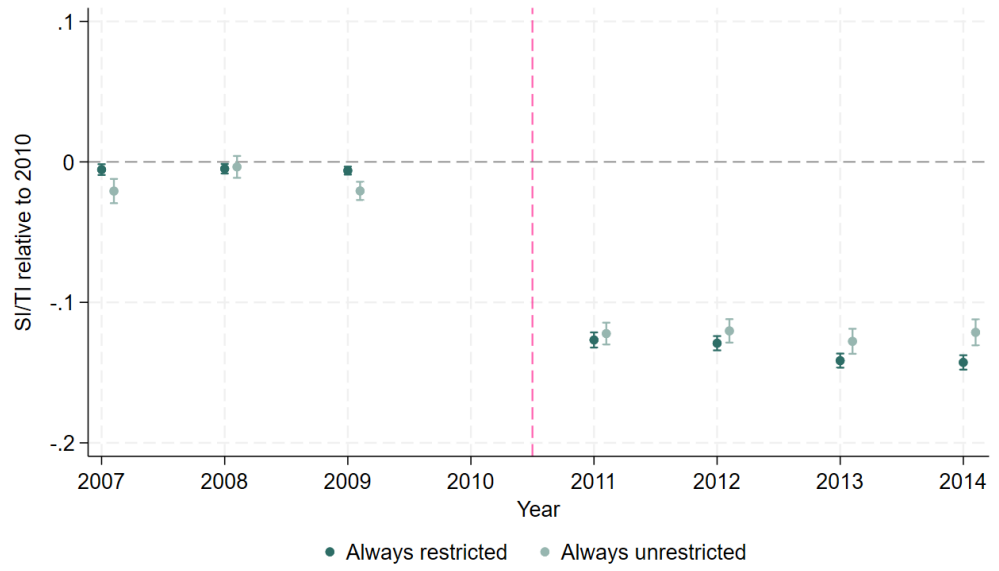
Notes: Figure (a) shows the social insurance (SI) income relative to labor income (LI) for all restricted (TyEL) and all unrestricted (YEL) owners. Each point contains 2.5% of observations of the respective group. Figure (b) shows the distribution of the answers to the survey question about the reasons for low YEL insurance contributions (left-side 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. Respondents could select multiple answer categories simultaneously. The figure also plots the average social insurance income per total income with 95% confidence interval within each answer category (right-side axis) with and without controls, measured for all survey respondents.

Figure 2: Choice of Social Insurance Coverage Following the 2011 Reform

(a) Development of SI Income Around the 2011 Reform

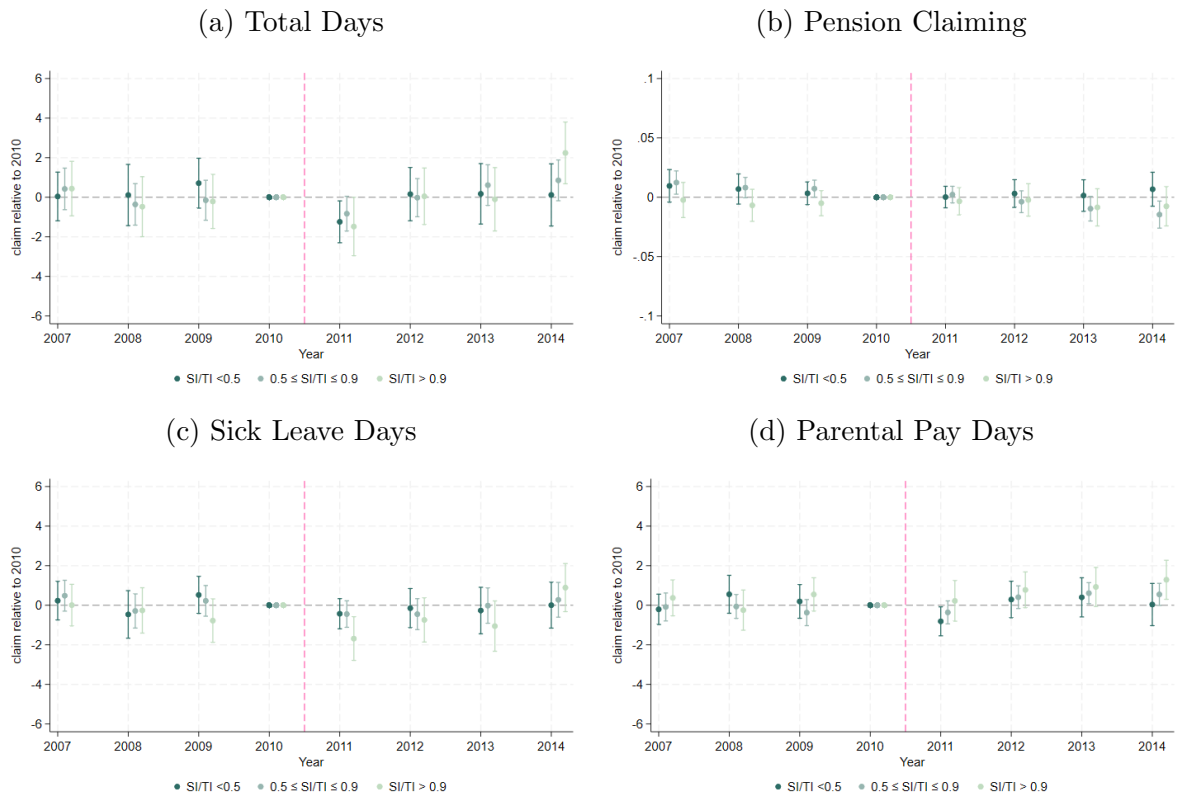


(b) Event-study Estimates



Notes: Figure (a) displays average SI income relative to total income for the treated group, the always restricted owners (ownership shares between 10% and 30%), and the always unrestricted owners (ownership shares between 51% and 70%). Figure (b) presents the estimates from equation (1), where the dependent variable is the logarithm of SI income relative to total income. The control groups are the always restricted or the always unrestricted owners.

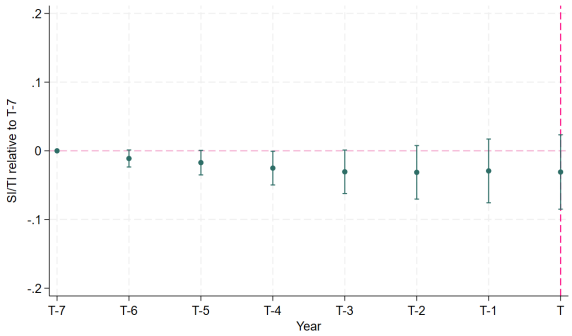
Figure 3: SI Claims Following the 2011 Reform



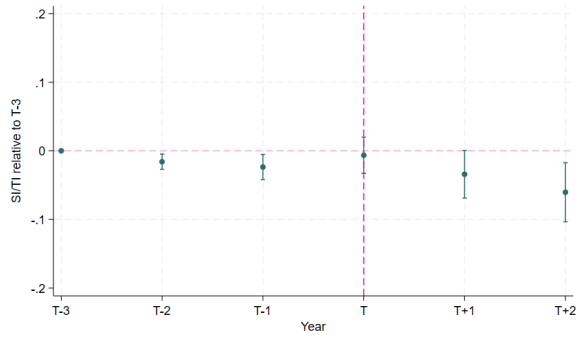
Notes: These figures show the results of estimating equation (1), where the outcome variable represents a measure of SI claims. In each figure, the individuals are separated into three groups based on the average SI/TI in post-reform years (2011-2014): those with SI/TI below 0.5, between 0.5 and 0.9, and above 0.9. The control group consists of the always unrestricted individuals (ownership share is between 50 and 70%). In figure (a), the outcome is the total number of benefit days, including absences due to sickness and parental leave. In figures (b), (c) and (d) the outcomes are the indicator of claiming pension benefits, and the log number of sick leave and parental leave days, respectively. Standard errors are clustered at the individual level. For more details see Section 4.1.

Figure 4: Anticipatory Responses to SI Claims

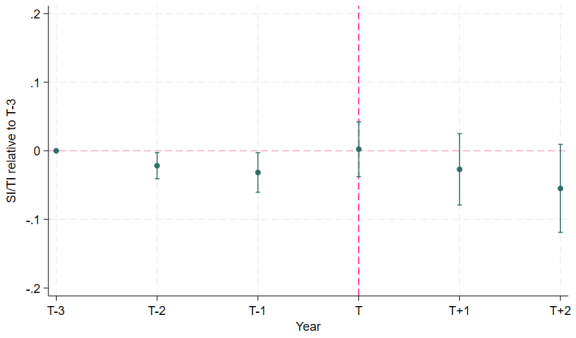
(a) Anticipation of Retirement



(b) Sick Leave

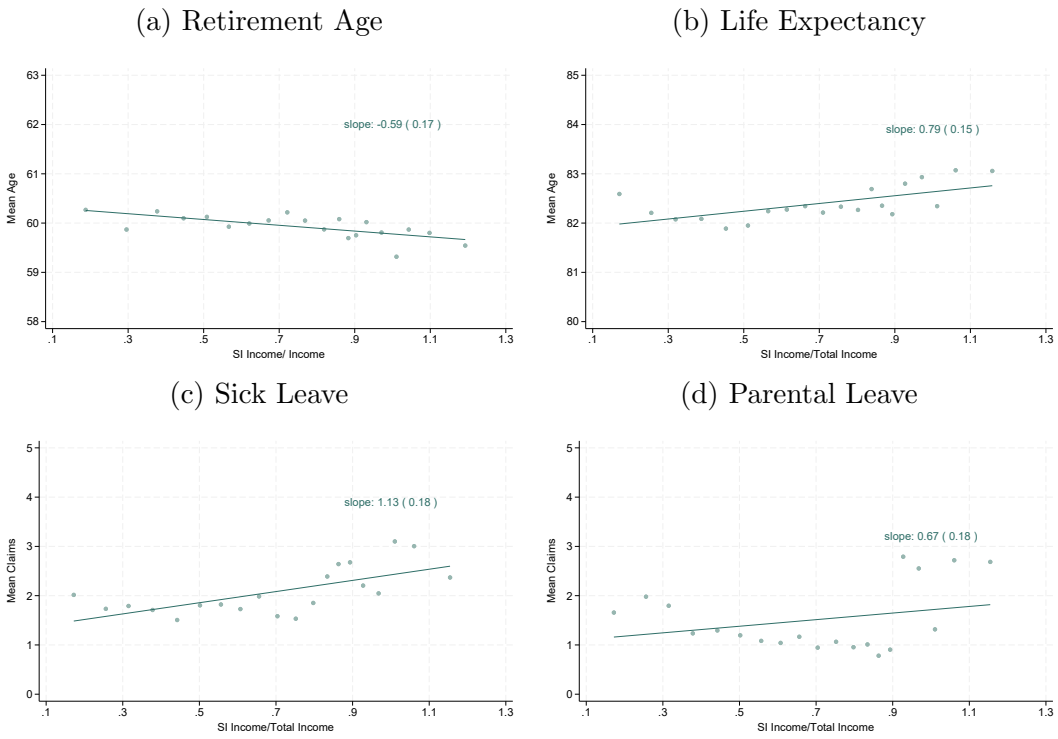


(c) Parental Leave



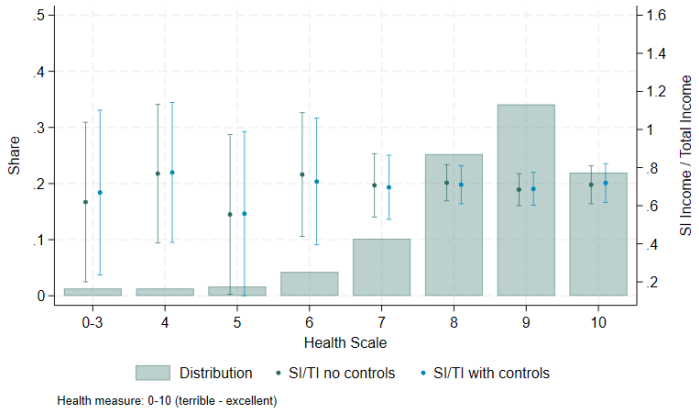
Notes: Figure shows the results of estimating equation (2) on the core sample of unrestricted entrepreneurs. Year T on the horizontal axis identifies the first year when an individual claims a benefit. Panel (a) studies the development of SI/TI seven years prior to retirement, and panels (b) and (c) three years before and after claiming sick leave and parental leave. We control for age, gender, municipality and industry, and include year fixed effects. The 95% confidence intervals are derived from standard errors that are clustered at the individual level.

Figure 5: SI Coverage and Claims: Positive Correlation Tests



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

Figure 6: SI Coverage and Self-Reported Health Status

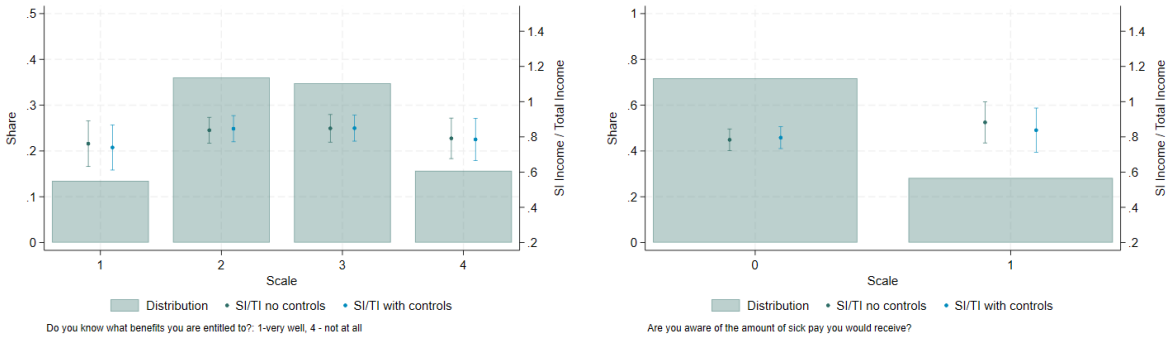


Notes: The figure shows the distributions of the answers to the following survey question (left-hand side axis): Evaluation of your own perceived health status on a scale from 0 to 10. The figure includes the average social insurance income per total income with 95% confidence intervals within each answer category (right-hand side axis) with and without controls, measured for all survey respondents.

Figure 7: Perceived Knowledge of SI Rules

(a) Knowledge about SI Claims

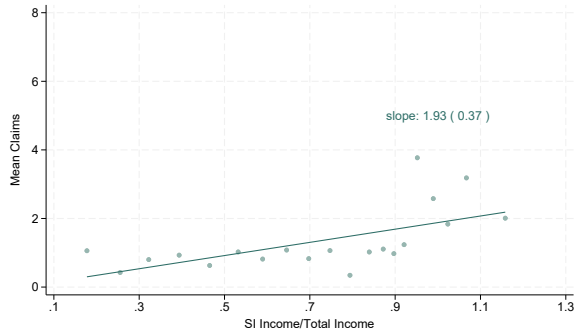
(b) Awareness of Sick Leave Replacement Rates



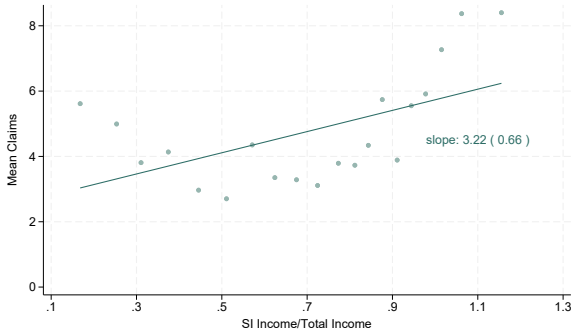
Notes: The figure shows the distributions of answers to survey questions (left-side axis) and corresponding average coverage levels with 95% confidence intervals (right-side axis) within each answer category with and without controls. Figure (a) question: "Do you know what SI benefits you are entitled to, on a scale of 1 (very well) to 4 (not at all)?" Figure (b) question: "Are you aware of the size of your sick leave benefit (yes or no)?"

Figure 8: Learning over Time

(a) Treated Individuals: 2011

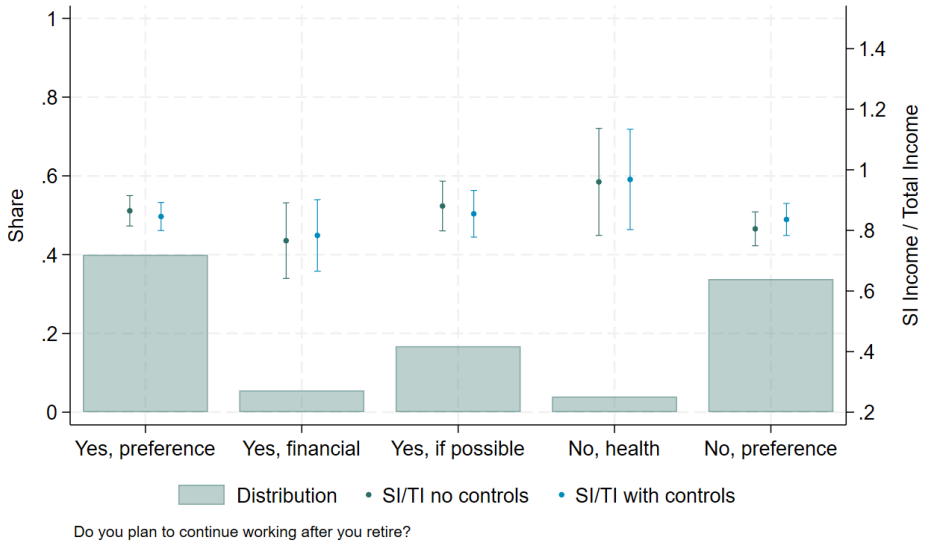


(b) Treated Individuals: 2014



Notes: Figures (a) and (b) show the average number of total benefit days claimed (including zeros) by the treated individuals in 2011 and 2014 by bins of SI/TI in these years, respectively. Each SI/TI bin contains 5% of 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.

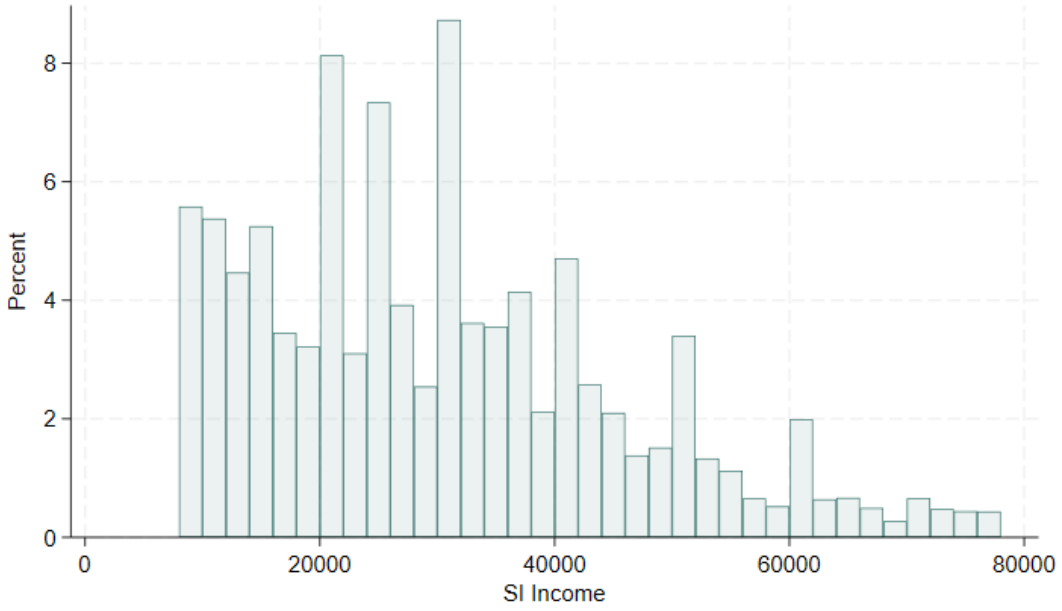
Figure 9: Plans on Working After Retirement



Notes: Figure shows the distributions of the answers to the following survey question (left-hand side axis): "Do you plan to continue working after you retire?" In addition, the respondents were asked to choose one of the following reasons behind their plans: "Yes, because I prefer working after retirement", "Yes, because it is not financially possible to retire", "Yes, if there is enough work", "No, because of health-related issues", and "No, because of preferences for retiring". The figure also includes the average social insurance income per total income with 95% confidence intervals within each answer category with and without controls (right-hand side axis), measured for all survey respondents. Standard errors are clustered at the individual level.

A Appendix

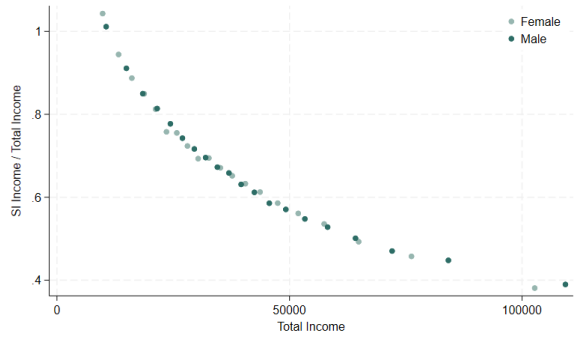
Figure A.1: Distribution of SI Income of YEL owners



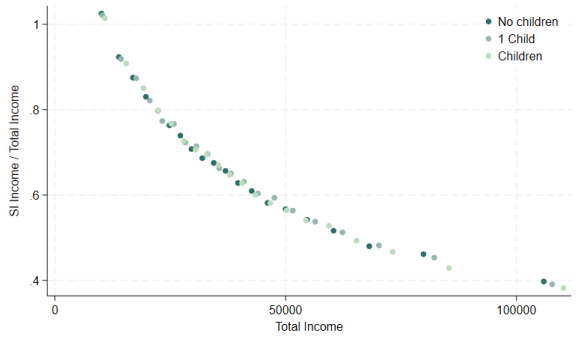
Notes: Figure shows the distribution of SI income in 2010 for the always unrestricted entrepreneurs in our baseline sample.

Figure A.2: Choice of SI Income by Characteristics

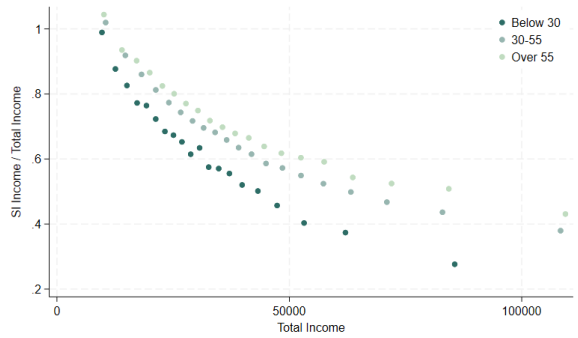
(a) SI Income Choice by Gender



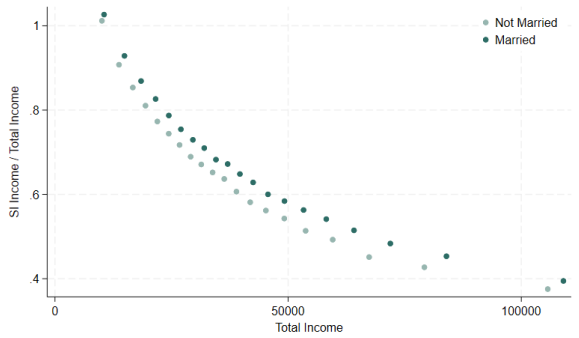
(b) SI Income Choice by Children



(c) SI Income Choice by Age

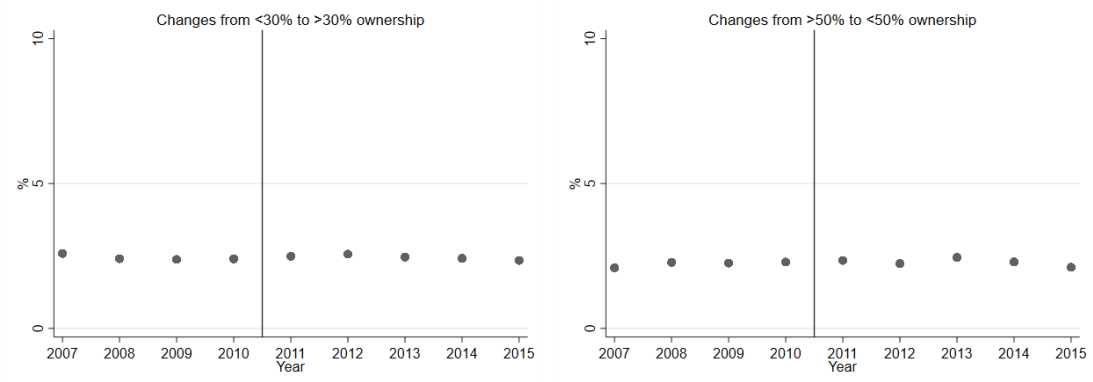


(d) SI Income Choice by Marital Status



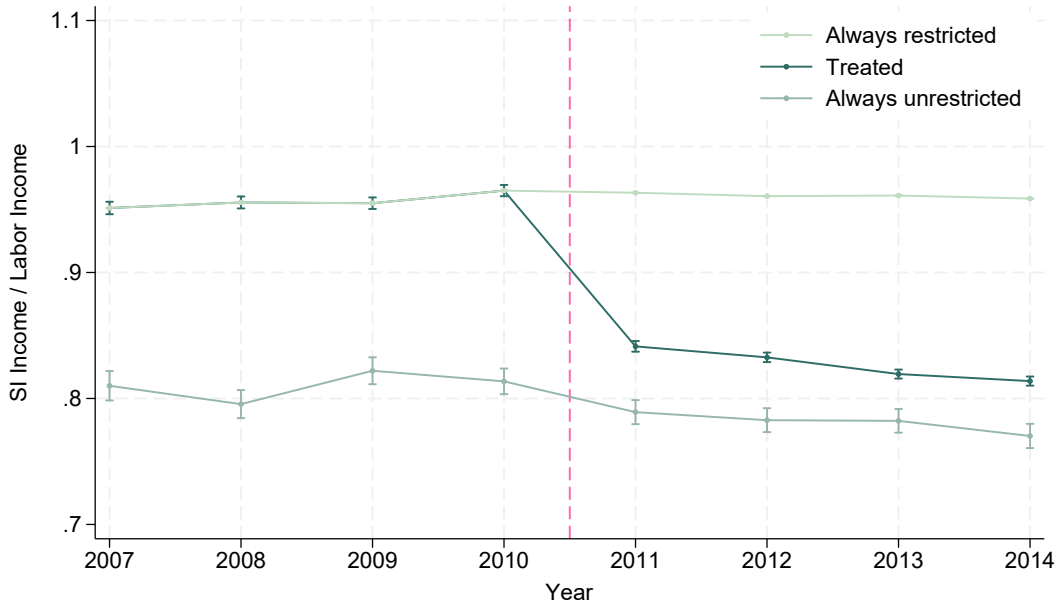
Notes: Figure shows the choice of SI income within each total income bin by different characteristics: Gender (panel a), the number of children (panel b), age groups (panel c) and marital status (panel d). Each bin contains 5% of person-year observations by total income. The figure shows that there are no significant differences in SI income choices by gender or the number of children. In contrast, the SI income is lower within each income group among younger individuals and those who are married.

Figure A.3: Ownership share changes over time



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

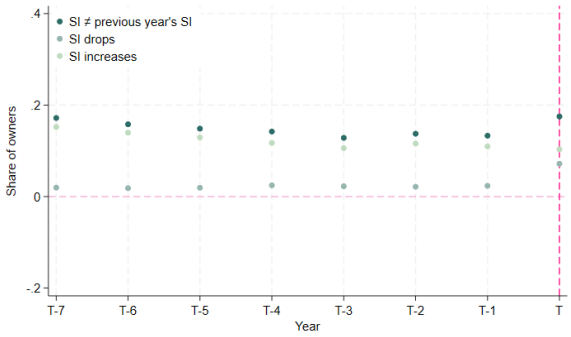
Figure A.4: Choice of Social Insurance Coverage Following the 2011 Reform: SI Income/Labor Income



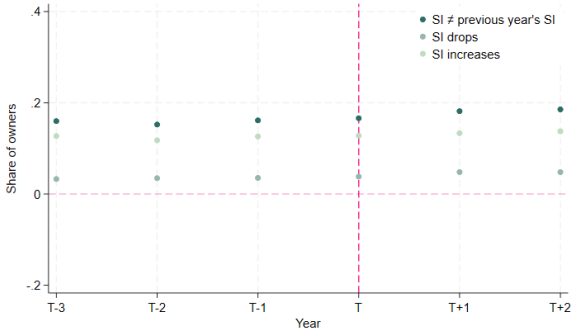
Notes: Figure shows the evolution of SI/LI (social insurance income divided by labor income) over time for treated individuals (ownership share between 30-50%) and the always restricted owners (ownership shares between 10% and 30%), and the always unrestricted owners (ownership shares between 51% and 70%).

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

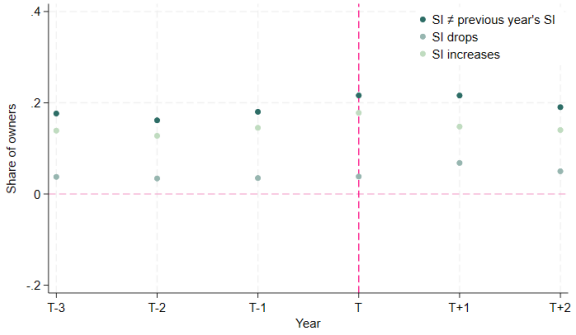
(a) Anticipation of Retirement



(b) Sick Leave



(c) Parental Leave



Notes: The figure shows the shares of unrestricted individuals who change their SI Income levels in the years close to retirement, sickness or having a child. Year T on the horizontal axis identifies the first year when an individual claims a benefit. Panel (a) studies the development of changes in SI Income levels seven years prior to retirement, and panels (b) and (c) three years before and after claiming sick leave or parental leave, respectively.

B Predicting Life Expectancy

Following Chetty et al. (2016), we leverage Gompertz's Law, which implies a log-linear relationship between age and mortality. We estimate a Gompertz survival model via maximum likelihood and use the fitted parameters to compute life expectancy for individuals in our sample.

We estimate a Gompertz survival model via maximum likelihood, separately by sex, recovering both the shape parameter and the baseline hazard for men and women. Using a rich set of explanatory variables, we then construct individual-level survival probability estimates. The survival function is given by:

$$S(t) = \exp\left(-\lambda_j \gamma^{-1}\left(e^{\gamma t_j} - 1\right)\right) \tag{5}$$

we implement the model by parameterizing $\lambda_j = \exp(x_j\beta)$, which implies a Gompertz baseline hazard of the form $h_0(t) = \exp(\gamma t)$. The parameter γ is estimated from the data and accounts for differences in life expectancies for each cohort.

Using each individual's age-specific survival probabilities, we then estimate life expectancy. In principle, this could be obtained by integrating the survival function (i.e., taking the area under the survival curve). However, because the integral is cumbersome in our setting, we instead compute life expectancy as follows:

$$\text{Life Expectancy} = Age + \sum_{t=1}^{100} S(t) \tag{6}$$

where $S(t)$ denotes the probability of surviving t additional years beyond Age, the individual's age when first observed in the data. We restrict the sample to individuals who survive to at least age 40; accordingly, our life-expectancy measure reflects the probability of surviving each subsequent year after entry into the sample.

We use the model to estimate covariate effects, the baseline hazard, and the Gompertz shape parameter. Estimation relies on population-wide Finnish data covering 1987–2019 and cohorts born between 1920 and 1970. We exclude individuals who die before age 40 as well as observations beyond age 100, so the model does not mechanically

link early deaths to later-born cohorts or very late deaths to earlier-born cohorts.

To assess the reliability of the resulting life-expectancy estimates, we split the data into training and validation samples (70/30). Individuals are randomly assigned to each group to maintain comparable demographic composition. We fit the Gompertz model on the training sample and then evaluate its performance on the validation sample. Model fit is assessed by how accurately it predicts mortality among individuals who die during the observation window and how well it matches the observed survival share at each age. For our purposes, an additional key diagnostic is that estimated life expectancy is positively correlated with realized longevity.

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.

We model heterogeneity in the mortality hazard using a rich set of covariates: education (indicator variables for each attainment level), occupation (high/low white- and blue-collar categories), region of birth (four regions), region of residence (52 regions), income (ranked into seven bins), unemployment, sick leave, disability pension receipt, number of children, and marital status. We also include interactions between income rank and region of residence, and between marital status and income rank. We then estimate the model by maximum likelihood under a Gompertz specification and use the fitted coefficients to generate individual-specific life-expectancy estimates for our study sample.

Note that the model infers life expectancy from income, industry, and basic demographic characteristics, and therefore does not directly observe individuals' underlying medical conditions. To partially capture health differences, we use annual sickness allowance uptake as a proxy for health status.

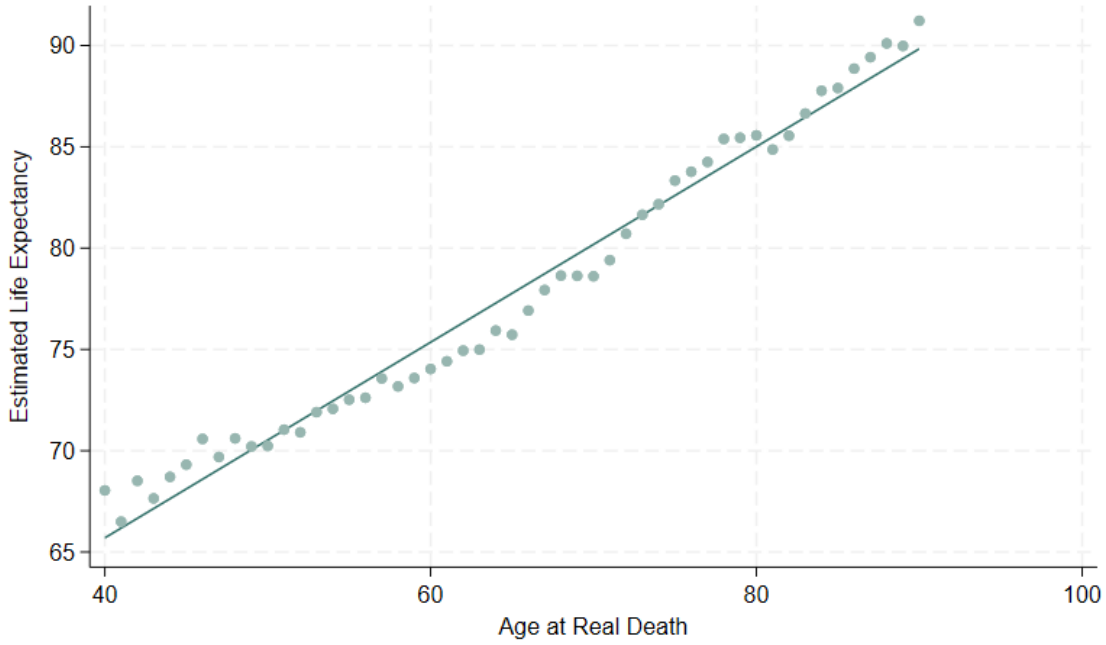


Figure B.1: Life Expectancy and Death

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

Overall, the model performs poorly in predicting levels of individual life expectancy, tending to over-predict longevity for most individuals. This likely reflects the fact that the available covariates cannot fully account for premature mortality in the sample. Nevertheless, as shown in Figure B.1, estimated life expectancy is strongly positively correlated with realized age at death. In other words, while the model misses levels, it preserves relative ranking: individuals who die earlier are assigned lower predicted life expectancy, which is the key requirement for our analysis.