Youth Subjective Life Expectancy and Early Labor Market Choices*

Lucas Finamor†

First version: January 2024
This version: October 2024

Abstract

The literature has documented how subjective life expectancy (SLE) is strongly associated with savings and retirement outcomes for those nearing retirement. This paper assesses whether SLE matters when young individuals make consequential career decisions at the labor market entrance. Exploring survey and administrative data from Chile, I show how individuals aged 18–26 with one standard deviation higher SLE had 6.5%–12.3% higher pension wealth 15 years later. In a pension system based on individual capitalization accounts, contributions made early in the career compound for longer and are therefore valuable. I employ different empirical strategies, including exploiting cross-sectional variation in SLE, longitudinal individual fixed-effects approach, and instrumental variables, exploring health and death of family members. All yield similar results. In a simple theoretical framework, I show how ignoring heterogeneity in life expectancy leads to biased predictions and might induce suboptimal policies.

JEL Codes: D84, J22, J46, H55

*I would like to thank Richard Blundell, Áureo de Paula, Cormac O'Dea, David Sturrock, Eduardo Ferraz, Dalila Figueiredo, and Antonia Paredes-Haz for their comments. I thank the ESRC for support under the Centre for the Microeconomic Analysis of Public Policy (CPP) at IFS, grant number ES/T014334/1. This research uses the Chilean Social Protection Survey (*Encuesta de Protección Social*). I thank the Undersecretary of Social Protection, the intellectual owner of the survey, for the authorization to use the de-identified dataset. All the results from this research are the responsibility of the author and do not implicate the Chilean Undersecretary of Social Protection.

†Sao Paulo School of Economics — FGV and Institute for Fiscal Studies. Contact: lucas.finamor@fgv.br

1 Introduction

Survival probabilities are an important ingredient in any intertemporal dynamic problem, being essential to determine savings, adequate financial planning, labor market decisions, and retirement choices. Since the seminal work of Hamermesh and Hamermesh (1983) and Hamermesh (1985), the literature has explored the properties and importance of individual beliefs about their survival probabilities. Hudomiet et al. (2023) review this body of work, showing how subjective life expectancy correlates with health information (e.g., health assessment, diagnoses), lifestyle behaviors (e.g., smoking, exercising), life events (e.g., parental deaths), and, importantly, with actual mortality patterns.

Building on these findings showing how survey-elicited beliefs had desirable properties, researchers have explored the relation between survival beliefs and important economic activities. Subjective survival beliefs are associated with consumption/savings behavior, retirement, and the purchase of annuities.[1] This literature relies mostly on beliefs elicited from the population over 50 years of age, close to retirement.[2] However, the most impactful decisions in terms of career progression, exposure to risk, and coverage by social insurance are taken much earlier on by workers. Even the first job is very consequential (Altonji et al., 2016, Arellano-Bover, 2020). Therefore, it is still unclear how beliefs on life span play a broader role over the life cycle, particularly at the onset of labor market trajectories. This paper fills this gap by exploring elicited beliefs from very young adults and how they relate to future labor market outcomes, such as labor market participation, employment type, and pension contributions.

I combine a longitudinal household survey with administrative pension data from Chile. The survey brings information on labor market status, health, and beliefs measured in several waves. The survey can be combined with precise and frequent data from the pension administrative system at the individual level, including pension wealth. This yields an extraordinary dataset with individuals' trajectories for over 15 years. I restrict the data to young individuals (18–26 years old) to assess the impact of subjective life expectations at the onset of labor market careers. In pension systems based on individual capitalization accounts, such as the one in Chile, contributions made early in the career can compound for longer, being very consequential for retirement outcomes.

[1] Hurd et al. (2004), Bloom et al. (2006), O'Donnell et al. (2008), Van der Klaauw and Wolpin (2008), Salm (2010), Gan et al. (2015), Wu et al. (2015), Bissonnette et al. (2017), Heimer et al. (2019), Bresser (2021), O'Dea and Sturrock (2023)

[2] This specific age range comes from the dataset used, which is, in most cases, the Health and Retirement Study in the United States and their equivalent in other countries. These studies only survey individuals over 50. The exception is Heimer et al. (2019), which has a sample of the US population aged 28–78.

I start the analysis by assessing the properties of the elicited subjective life expectancy (SLE) measure. I show how survival beliefs correlate, in expected ways, with demographics, health, lifestyle behaviors, and personal traits. Those smoking, not exercising regularly, who have been diagnosed with diseases, who self-report bad health, or those with higher body mass index report lower subjective life expectancies. Beliefs are positively related to own and parental education. However, all these variables explain only a quarter of the SLE variance. It is, therefore, important to understand whether this residual SLE has true content or is just noise. I show how SLE (and residual SLE) predict mortality 15 years ahead, measured in the administrative data. That is, SLE has substantial content over and above all these observed measures. The panel dimension is also useful to show how the beliefs are modestly persistent and respond to new information, such as new disease diagnoses.

After verifying the desirable properties of the SLE measure, even when elicited from a young population, I turn the analysis to the labor market outcomes. I start exploring the cross-sectional variation in SLE, conditional on a rich set of control variables. Using the pension administrative data, I show how individuals who reported higher SLE have a higher attachment to the pension system in all years following the initial report. At the end of my sample period in 2019, 15 years after SLE was first measured, this corresponds to a significant gap in pension contributions. Those with 10 years higher SLE have around five months more pension contributions (3.8% of the mean) and around 255 thousand Chilean pesos higher pension wealth (3.8% of the mean).[3] Ten years of SLE equals the 75th–25th percentile gap, or approximately 0.83 standard deviations. Upon retirement, their pension wealth is used to fund retirement benefits. Therefore, these pension gaps associated with SLE can be read as potential gaps in final pension benefits after retirement.[4]

While this cross-sectional analysis is the main empirical strategy explored in the literature, there are two mains concerns with this approach. The first is with measurement. The elicited beliefs are likely measured with error, which could bias the estimates. The second is with identification and the interpretation of this relationship as causal. If there are unobservables that are correlated with SLE and labor market outcomes, the estimated coefficients cannot be interpreted causally. The richness of my data allows me to explore different strategies that address these concerns.

On the measurement error concern, I leverage the existence of two different measures of life expectancy to use an instrumental variable (IV) approach. The two questions were

[3]One thousand Chilean pesos corresponded to approximately 1.6 USD in 2004, around the first wave of the survey.

[4]The gap in final pensions would still depend on the (endogenous) retirement age, minimum pension floor guaranteed by the government, and future contributions.

placed in different modules within the survey and elicited the same underlying beliefs with different wording and scales. The first question asked individuals to what age they believed they would live, and the second asked the probability of being alive at a target age. I first show how only a small fraction of individuals answer inconsistently to the two questions. With the IV approach, I instrument SLE with the elicited probability of being alive at age 65. This yields larger estimates, consistent with the presence of measurement error. Those with 10 years higher SLE have 7.3% higher pension wealth fifteen years later, in December 2019.

On the second concern, the main specification controls for a rich set of controls, including age-gender-education, region, and parental education fixed effects. Nevertheless, it is possible that unobservables that are correlated with SLE and with the labor market outcomes are driving these correlations. I first show how the coefficients are robust to the inclusion of relevant variables, including those measuring health assessment and diagnosis, numeracy, risk aversion, degree of future-oriented preferences, and the big five measures. Applying the robustness procedures by Oster (2019) and Cinelli and Hazlett (2020), I discuss (and quantify) the conditions under which this result could be overturned by unobservables and discuss how it is unlikely that this is the case, considering the impact of the relevant included variables.

Exploiting the longitudinal dimension of the survey and the fact that SLE was elicited in several waves, I also estimate a specification with individual fixed effects, exploiting solely within-individual variation. This specification is, therefore, robust to any time-invariant unobservable, for example, any relevant personal trait and preference that were not measured directly. The fixed-effect estimates produce similar positive results, with smaller coefficients. This is expected as this approach identifies differences in pension contributions over narrower time windows. When using the panel dimension, I also show how conditioning on the first SLE and future SLE beliefs are positively associated with pension outcomes. This would be consistent with a "revision behavior" — individuals with similar initial beliefs about life expectations revise their pension contributions up or down, depending on future information revealed to them. While this approach can deal with time-invariant unobservables, it is still vulnerable to time-varying unobservables that correlate with the expectations and SLE, such as health shocks that could directly affect the labor market.

Leveraging the fact that the survey also collects information on family members, I also analyze another empirical strategy that uses parental death and average health of family members as instruments for life expectancy beliefs. Those with deceased parents or with family members that report worse health have lower life expectancy. This approach alleviates the concerns with the own-health shocks that were present in the panel approach. These IV estimates yield larger positive coefficients than the cross-sectional analysis, however with large confidence intervals.

In summary, independently of the set of assumptions and empirical strategies used in the previously described exercises, all results point in the same direction: individuals with higher SLE have higher pension wealth 15 years later. This cannot be explained by time-invariant unobservables, and any time-variant explanation would need to produce similar biases despite the shocks being explored: own health shocks or new information being revealed, parental death, or health of family members to rule out the results. Moreover, the richness of the data linkage, show how residual SLE predicts future mortality.

Next, I investigate what are the drivers of these pension gaps by SLE. Individuals with higher SLE might chose differently how much they participate in the labor market, or which type of employment to engage with, leading to different pension contributions. First, I find that individuals who have 10 years higher SLE are 1.1pp more likely to participate in the labor force (1.5% of the mean participation rate). I then turn the analysis to the type of employment. Pension contributions are mandatory for those working formally in the private and public sectors, while this is not the case for the large share of individuals in informal jobs or self-employment. Those with 10 years higher SLE are 3.0pp more likely to work as formal employees (6.3% of the mean). Mediation analysis show that labor participation explains 22.9% of the total effect on pension contributions, while type of employment explains 76.6%. These results indicate that employment type accounts for most of the pension gap found. In short, those expecting to live longer are more likely to work and to be working as formal employees with mandatory pension contributions.

I show, with a theoretical framework, how these novel findings are consequential for policy-making, such as the design of social security and private savings incentives. In the simple framework individuals choose how much to invest for retirement, taking into account beliefs on life span. I show that ignoring heterogeneity in (subjective) life expectancy produces biased predictions. Models that only account for the average life expectancy across individuals tend to overestimate optimal savings for retirement and behavioral changes to social security reforms.

This paper offers several contributions to the literature analyzing subjective life expectancy and economic decisions.[5] First, as most of the literature focuses on individuals near retirement, I show novel evidence that early beliefs on survival are strongly associated

[5]Hurd et al. (2004), Bloom et al. (2006), O'Donnell et al. (2008), Van der Klaauw and Wolpin (2008), Salm (2010), Gan et al. (2015), Wu et al. (2015), Bissonnette et al. (2017), Heimer et al. (2019), Bresser (2021), O'Dea and Sturrock (2023)

with labor market choices and pension contributions. In particular, young individuals with higher SLE are more likely to participate in the labor market and to make pension contributions. Given our knowledge of how consequential early life decisions are for career trajectories, this is a critical finding for our understanding of labor market choices. Second, the richness of the data allows me to explore different empirical strategies that address concerns with measurement error and identification, leveraging information collected across different waves and from family members. Third, I add to both literature on SLE and informal labor markets in developing countries by providing new insights into how SLE is associated with employment type. Namely, I show that individuals with higher SLE are more likely to work formally. Employment type is extremely impactful. It is associated with exposure to risk, eligibility to several government programs, and, more generally, career progression. Lastly, the data structure combining longitudinal survey and administrative data allows me to go beyond self-reported outcomes, thus avoiding any assumptions on how reporting behavior in the outcome and subjective beliefs correlate.

The rest of the paper proceeds as follows. The next section provides a brief description of the institutional setting in Chile and of the data. The third section discusses the life expectancy measures and their properties in the data. The following section presents the econometric strategy; the results are presented in the fifth section. In the sixth section, I present a theoretical framework. A discussion combined with some concluding remarks in the last section concludes the paper.

2 Institutional Setting and Data

2.1 Institutional Setting

In 1980, Chile introduced a fully funded individual capitalization pension system. Each month, employees are required to make pension contributions in their pension account at 10% of their wages, up to a cap. These pension funds are administered by private pension managers and are illiquid. Upon retirement, individuals can choose to annuitize their pension wealth entirely or partially.[6] The normal retirement age is 65 for men and 60 for women. Early retirement is an option if the resulting pension benefit is above an absolute threshold set by the government and higher than a fraction of the individual average wage. While voluntary contributions are allowed, they are very rarely observed.

For most of the sample period, self-employed individuals were not required to make

[6] For more information on options upon retirement, check Bello (2023).

pension contributions. That was introduced in a pension reform in 2008, whose implementation was postponed until 2015. Nevertheless, enforcement is a challenge. Finamor (2024) shows that 2/3 of self-employed individuals with at most high-school degrees are not registered with the tax authority, and less than one quarter had any pension contributions in 12 months. Therefore, working as self-employed or informal worker is one way of avoiding making pension contributions. Even though Chile has lower rates of informal work than neighboring countries in Latin America, there is still a substantial fraction of workers in informal jobs and self-employment activities — around 1/3 of the workforce (Finamor, 2024). The government subsidizes a pension floor that was substantially increased in the pension reform in 2008. Individuals with low or no pension wealth at age 65 are likely to be eligible for this minimum pension, depending on their family income and assets.

2.2 Data

This paper uses two main datasets from Chile: a household survey (*Encuesta de Protección Social*, EPS) and administrative data from the pension system (*Historia Previsional de Afiliados*, HPA). The EPS is a longitudinal survey that has been conducted since 2002, with six waves implemented every 2–4 years. Since the second wave, in 2004, it is nationally representative, covering the adult population in Chile. It contains detailed information on demographics, labor market characteristics, family, health, income, and assets for the interviewed person in the household. Crucially for this project, EPS also has questions on subjective life expectancy.

Every individual interviewed in the EPS can be linked to the administrative data from the pension system (HPA), which contains all monthly contributions towards the pension system between 1981 and 2019. Since 2008, pension wealth in every individual account is available. It is also possible to check pension claims and payments, including those paid to family members upon the death of the primary account holder. I complement these datasets with life tables computed by the Chilean National Statistics Office.

From these datasets, I derive the main sample used in the paper as individuals between 18 and 26 years old in 2004. The age restriction is to capture young individuals before or at the beginning of their labor market careers. The lower limit (18) is determined by the survey, which does not interview individuals younger than 18 as the main respondent.[7] The upper limit of 26 was chosen to yield enough sample size for the analysis, particularly when splitting the sample by gender and to conduct the instrumental variables estimation.

[7]There are 76 observations recorded with age 16 or 17, that were not targeted to be in the main sample.

Table 1: Descriptive statistics

Objective Life Expectancy	Men [75.924]			Women [81.464]		
	N Obs	Prop	SLE (mean)	N Obs	Prop	SLE (mean)
Total	894	-	76.002	856	-	73.659
Educational level						
Primary	105	0.117	73.210	86	0.100	72.849
High-School	497	0.556	75.765	455	0.532	72.473
Vocational	116	0.130	75.759	113	0.132	73.531
College	176	0.197	78.500	202	0.236	76.748
(subtotal)	894			856		
Region						
MR Santiago	370	0.414	76.146	357	0.417	74.364
Other	524	0.586	75.901	499	0.583	73.154
(subtotal)	894			856		
Mother education						
Less than HS	406	0.470	74.776	411	0.495	72.221
HS	374	0.433	77.067	328	0.395	74.534
More than HS	83	0.096	77.880	92	0.111	77.652
(subtotal)	863			831		
Father Education						
Less than HS	369	0.439	75.171	369	0.461	72.033
HS	375	0.446	76.509	337	0.421	74.964
More than HS	96	0.114	78.125	95	0.119	77.168
(subtotal)	840			801		

Notes: The table presents the main descriptive statistics for men (first three columns) and women (last three columns). For each variable and gender, the table presents the number of observations, the proportion within the group, and the mean subjective life expectancy measured in years. The first row presents the total number of observations and the average SLE. The following groups show the statistics for the level of education (primary, high-school, vocational, or college education), the region of residence (metropolitan region of Santiago or other), and parental education (less than high-school, high-school degree, or more than high school). In the table header, the number in brackets show the average objective life expectancy, according to the 2005 life tables.

Table 1 presents some descriptive statistics. The final sample contains 1,750 observations (894 men and 856 women). I defer the discussion on the measure of subjective life expectancy to the next section, but we can see how men, on average, believe they will live longer than women, even though the (objective) life expectancy from the life tables in 2004 are much higher for women (81.5 for women versus 75.9 for men). In terms of education, most individuals in the sample have high school degrees, and more than one-third of the sample have higher degrees. Around 40% of the sample live in the metropolitan region of Santiago. On parental education, a minority of parents have more than high school degrees, with almost equal shares between those with and without high school degrees. In terms of subjective life expectancy, we see positive gradients for their own and parental education.

3 Life Expectancy Measure

Surveyed individuals answered two questions on life expectancy in the EPS. The first question inquired to what age individuals believed they would live.[8] The answers were ages measured in years. A second question asked the chances of living to at least age 65.[9] Respondents were expected to give answers in percentages ranging from 0 to 100. Interestingly, these questions were placed in different modules in the survey, resulting in the answers to both questions being separated by dozens of questions.[10]

I expect the two questions to capture the same underlying subjective beliefs around mortality. Indeed, empirically, the two are highly correlated.[11] In the analysis, I favor the first question for two main reasons. First, it is measured in years, arguably a unit of measure that all respondents are more familiar with than probabilities. Second, it displays more dispersion. As I measure beliefs for a very young age range, most respondents answer 100% for the probability of being alive at age 65. While 47% of individuals answered 100% to the second question, only 4% answered any value above 100 years for the first question. Figure 1c plots the raw data for these two questions. We can see how there is much more dispersion in the first than in the second question. This is probably not the case for the papers analyzing beliefs for those over 50 with target ages beyond that. Additionally, I can use both questions to verify whether individuals answer consistently to both questions. Only 2.8% of individuals answered the two questions with unlikely values (answering 99-100% to the second question and less than 65 years to the first question or answering 0-1% to the second question and more than 65 for the first). I keep them in the sample, but results are

[8]In Spanish, the original wording was *"¿Hasta qué edad cree usted que va a vivir?"*

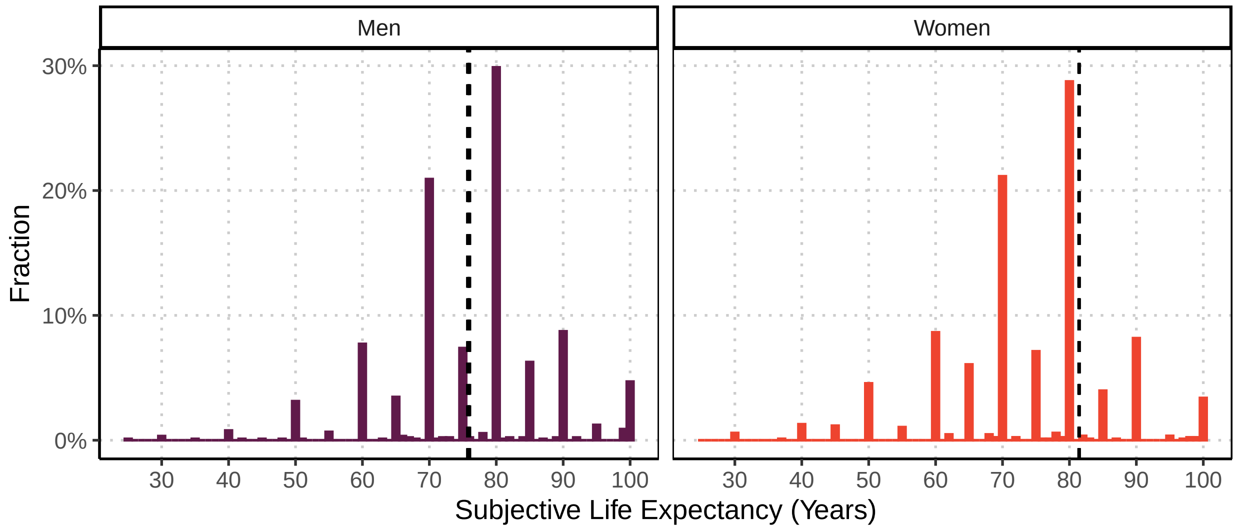
[9]In Spanish, the original wording was *"¿Cuáles son sus posibilidades de vivir hast a los 65 años?*

[10]There were at least 101 questions between the two SLE questions. The exact number depends on each respondent's answers and the exact flow of the questionnaire.

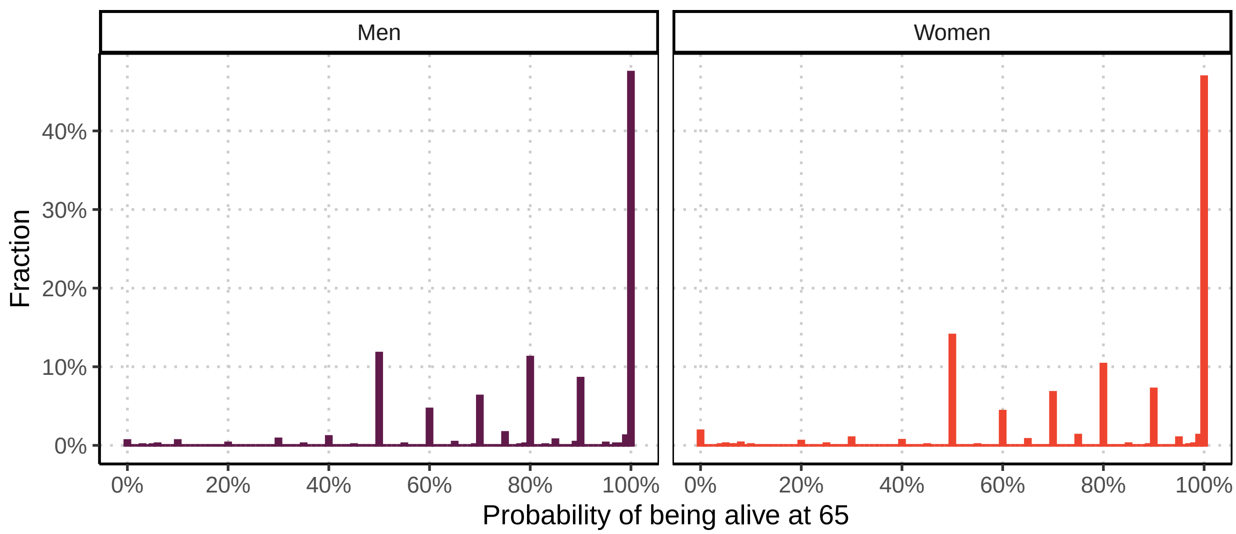
[11]Additionally, the R-square of regressions on future mortality on the subjective life expectancy are almost the same, irrespectively of which variable is used.

robust to removing them.

(a) Subjective Life Expectancy (years)



(b) Probability of being alive at age 65



(c) Distribution of subjective survival beliefs, by gender

Notes: The figure plots the histogram for the raw answers for the two subjective survival beliefs elicited in the survey, separately for men and women. The top panel plots the answer to the question on subjective life expectancy (Question: Up to what age do you believe you will live?) and the bottom panel to the question on probabilities of living up to age 65 (Question: What are the chances of living until 65?). The vertical dashed line shows the average life expectancy from the life tables for this sample.

The drawback of using the question measured in years is that it is unclear which statistic (mean, median, mode, or any other summary statistic) individuals are reporting. This would be more problematic if I were to recover the entire distribution of subjective life expectancy joint with any parametric restrictions. This is not the case. What I explore in

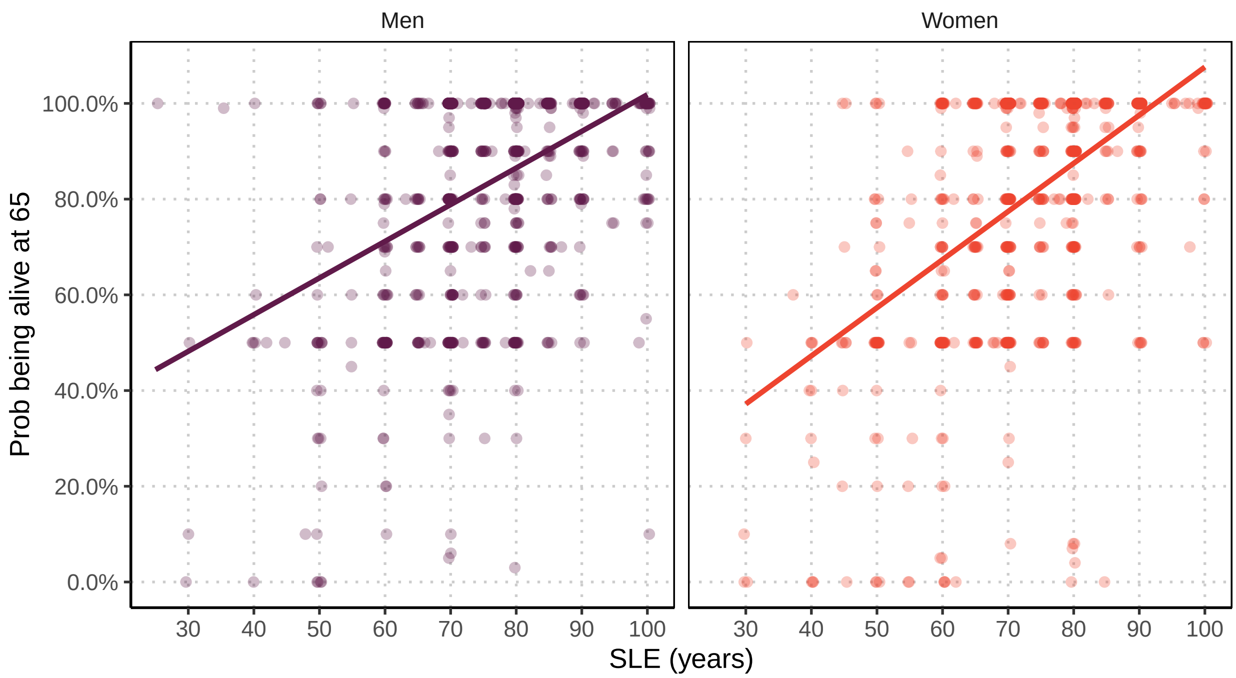


Figure 2: Subjetive life expectancy and probability of being alive at age 65

Notes: For each individual, the figure plots their answer to the two questions regarding life expectancy. On the x-axis is the SLE question measured in years, and on the y-axis is the question on the probability of being alive at age 65, measured in 0-100%. Dots are plotted with transparency, therefore, darker regions represent a higher mass of points. The solid line is the best linear fit of the data.

the analysis is how future labor market behaviors compare with reporting high or low values of subjective life expectancy, using different empirical strategies.

From the raw data in Figure 1c, we can see some rounding behavior. Most individuals answer in multiples of five. Men's answers are closer to the objective life expectancy from the life tables, while women's reports are significantly below. From this figure, the high dispersion of the answers is also evident. The standard deviation is around 12.1 years for men and 12.6 for women. Figure 2 plots the answers to the two questions for each individual, showing their positive correlation. The correlation coefficient is 0.475, which is probably affected by the ceiling at 100% in the probability of living at age 65 question. Figure A.2 plots the raw data by age and education.

4 Empirical Strategy

The empirical strategy contrasts future labor market behavior with initial subjective life expectancy, controlling for the main demographic variables. I run the following OLS regression:

$$Y_i = \beta \text{SLE}_i + \eta_{d(i)} + \nu_{r(i)} + \varphi_{p(i)} + \varepsilon_i, \tag{1}$$

where Y_{it} is the outcome for individual i in time t, regressed on the SLE reported in 2004. The regression controls for demographic-cell fixed effect ($\eta_{d(i)}$), comprised of gender, age, and educational level, for region fixed effects ($\nu_{r(i)}$), and parental education fixed effects ($\varphi_{p(i)}$) which proxy for socioeconomic status.[12] In some specifications, I pool data from more periods and a time fixed effect is also included in these cases. Therefore, β is our coefficient of interest, measuring the correlation between the outcome and subjective beliefs, conditional on our rich set of covariates. If there is no omitted variable that correlates with SLE and Y, we could interpret β as a causal effect of subjective life expectancy.

To ease interpretation, I report the coefficients of changing 10 years of SLE. This is the gap in subjective life expectancy between individuals in the 25th and 75th percentiles of the SLE distribution, conditional on gender. The equivalent 90th–10th gap is 30 years. All the coefficients can be interpreted as an increase of approximately 0.8 standard deviations. Whenever there is more than one observation per individual, standard errors are clustered at the individual level. For cross-section regressions, the standard errors are heteroskedastic robust.

One concern with specification 1 is measurement error. As SLE is self-reported in a survey, it potentially contains measurement error that would attenuate the estimated coefficient of interest. To address this concern, I also use an instrumental variables approach, leveraging the two questions on subjective life expectancy in the survey. Therefore, I instrument SLE using the response on the probability of being alive at age 65 (P^{65}) as an instrument, running a two-stage least squares regression:

$$\begin{aligned} \text{(First Stage) } SLE_i &= \alpha P_i^{65} + \eta_{d(i)} + \nu_{r(i)} + \varphi_{p(i)} + \epsilon_i \\ \text{(Second Stage) } Y_i &= \beta^{IV} SLE_i + \eta_{d(i)} + \nu_{r(i)} + \varphi_{p(i)} + \zeta_i \end{aligned} \tag{2}$$

The assumption for this strategy is that measurement error in both variables are uncorrelated. The fact that the two questions were far apart in the survey supports this assumption. However this would fail if for instance, there is permanent inconsistencies when answering the two questions.

Section 5.3 explores different empirical strategies. Leveraging the longitudinal dimen-

[12]For consistency, education is measured at the initial wave in 2004. While it is possible that some individuals will get more educated over the years, note that the specific control is gender-age-education cell, therefore it compares individuals with the same educational level at the same baseline age.

sion of the data, one of the specifications exploit solely within-individual variation on SLE, running the OLS estimator of equation 3, which includes individual fixed-effects:

$$Y_{it} = \beta^{WI}\mathrm{SLE}_{it} + \eta_i + \varepsilon_{it}. \tag{3}$$

Lastly, two alternative instrumental variables approaches use parental death and health of family members as instruments for the initial beliefs. Let H_i be this measure of whether either of their parents were deceased or the average health self-assesment of family members. I run the following 2SLS estimation:

$$\begin{aligned} \text{(First Stage) } SLE_i &= \gamma H_i + \eta_{d(i)} + \nu_{r(i)} + \varphi_{p(i)} + \epsilon_i \\ \text{(Second Stage) } Y_i &= \beta^{IV2} SLE_i + \eta_{d(i)} + \nu_{r(i)} + \varphi_{p(i)} + \zeta_i. \end{aligned} \tag{4}$$

5 Results

5.1 Health and Mortality

Before analyzing the labor market behavior, I first investigate what in the data is correlated with SLE. From now on, I will use the "net" SLE, which subtracts the objective expectancy from life tables from each individual's answer, as the main variable. Every time I refer to SLE, I will be referring to this net variable unless explicitly stated otherwise. The results are displayed in Table 2. The first four columns introduce the main controls used in all regressions. We can see that gender alone explains 9.2% of the total variation of SLE. The remaining controls (age, education, region, and parental education) explains additional 6.2 percentage points (pp) of the total variation. In the fifth column, I include variables related to health and lifestyle. They all have the expected sign. Smokers have, on average, -1.189 years of SLE. Doing physical activities regularly and assessing your own health status as good are associated with a higher SLE. While having any diagnosis in a given list of diseases and having a higher BMI are associated with lower SLE.[13] Together, they raise the R-square by only 1.7pp. The next column introduces indicators for deceased mother and father, both with negative coefficients. For instance, having a deceased father is associated with a lower SLE by 2.767 years.

In the last three columns (7–9), I add variables related to preferences, skills, and personality traits in order to assess how they relate to SLE. In column 7, I add an indicator

[13]The list of diagnoses includes asthma, pulmonary emphysema, depression, diabetes, hypertension, high blood pressure, heart problems, cancer, arthritis, osteoarthritis, renal diseases, stroke, mental illness, and HIV AIDS.

Table 2: Correlates of SLE

	Outcome: SLE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Smoking					-1.189	-1.162	-1.161	-1.171	-1.191
					(0.607)	(0.605)	(0.606)	(0.605)	(0.606)
Physical Activities					1.765	1.855	1.843	1.837	1.893
					(0.772)	(0.767)	(0.767)	(0.768)	(0.771)
Good Health					3.134	3.017	3.032	3.050	3.057
					(0.969)	(0.966)	(0.967)	(0.959)	(0.958)
BMI					-0.185	-0.183	-0.181	-0.191	-0.184
					(0.083)	(0.084)	(0.084)	(0.084)	(0.084)
Any Diagnosis (Diseases)					-0.768	-0.912	-0.900	-0.918	-0.963
					(1.089)	(1.088)	(1.090)	(1.082)	(1.080)
Deceased Mother						-1.530	-1.500	-1.392	-1.168
						(2.270)	(2.289)	(2.254)	(2.284)
Deceased Father						-2.767	-2.750	-2.720	-2.760
						(1.295)	(1.297)	(1.280)	(1.286)
Risk-Aversion							0.602	0.583	0.587
							(0.633)	(0.632)	(0.632)
Numeracy								0.005	-0.043
								(0.231)	(0.234)
Future-oriented Pref								0.065	0.094
								(0.866)	(0.869)
Agreeableness									0.155
									(0.598)
Conscientiousness									0.256
									(0.494)
Emotional Stability									-0.239
									(0.468)
Extraversion									-1.012
									(0.553)
Openness									-0.687
									(0.494)
Gender	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gender-Age-Educ	-	✓	✓	✓	✓	✓	✓	✓	✓
Region	-	-	✓	✓	✓	✓	✓	✓	✓
Parental Education	-	-	-	✓	✓	✓	✓	✓	✓
Observations	1,750	1,750	1,750	1,750	1,750	1,750	1,750	1,750	1,750
R^2	0.092	0.137	0.146	0.154	0.171	0.175	0.176	0.181	0.187

Notes: The table presents the results from a regression of SLE on selected variables. In each column the corresponded controls are sequentially included. Individual level clustered standard errors are in parentheses.

for the individual being risk-averse. Column 8 introduces a numeracy measure and an indicator for future-oriented preferences, and lastly, 9th column introduces measures of the "Big Five" personality traits. Appendix A shows the precise definitions and construction of these variables.[14] The indicators for risk-aversion and having future-oriented preferences have positive and small associations with SLE, non distinguishable from zero. Surprisingly, the numeracy measure is not correlated with SLE (conditional on remaining controls). Jointly, these variables explain little of the residual variation of SLE. The R-square raises by only 1.2pp. More importantly, we can see that all these variables, including the rich set of baseline controls, do not explain much of the SLE dispersion. The R-squared of this regression is 18.7%. This is not unique to this setting; Hurd and McGarry (2002), Puri and Robinson (2007), and Delavande et al. (2017) report similar findings.

A crucial question is whether these beliefs are predictive of mortality. That is, do they reveal any private information individuals might have about their survival chances. Table 3 below shows the results of an indicator for being deceased in 2019 and the reported SLE in 2004, conditional on the main controls (age-gender-education, region, and parental education fixed effects). As individuals in my sample are only 18–26 in 2004, there is not enough time to observe their mortality. Therefore, I also include in this mortality analysis older individuals aged 27–36, 37–46, and 47–56. We can see that most of the coefficients are of the expected sign, those with higher SLE are less likely to be deceased in 2019. The coefficients are statistically significant for the older samples of men. Panel B adds all the health and lifestyle variables from Table 2 and show that SLE is still predictive of future mortality after conditioning on the main health variables. For women we see much weaker correlations, which are indistinguishable from zero.

One final question is on the persistence of these beliefs. I take advantage of the longitudinal dimension of the survey and explore how future beliefs on life expectancy are correlated with the initial report. The results are presented in Table 4. The first column shows the conditional correlation between the initial SLE reported in 2004 and the one in the next wave, in 2006. One additional year of SLE in 2004 is associated with 0.415 years in 2006, 0.307 in 2009, and 0.189 in 2012. The three coefficients are highly statistically significant. When using the instrumental variable approach, exploiting the probabilistic question as as instrument we obtain larger estimates, which is consistent with the presence of measurement error on the initial SLE measure.

The results in Table 4 show that these beliefs are modestly persistent. One of the reasons that would make the beliefs differ across surveys is when individuals learn more information about their life chances and receive shocks. As the survey also records the

[14] One important caveat is that while all variables from columns 1–7 were measured in the same initial wave (2004), the questions used to derive the numeracy and future-oriented preferences (column 8) were only introduced in the following wave (2006) and the Big Five measures (column 9) were introduced in the next round in 2009.

Table 3: SLE and mortality

	Men				Women			
Age Range	18–26	27–36	37–46	47-56	18–26	27–36	37–46	47-56
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A - Main Specification								
SLE	-0.006	-0.004	-0.017	-0.031	-0.001	0.001	-0.003	-0.007
(s.e.)	(0.004)	(0.004)	(0.007)	(0.010)	(0.004)	(0.002)	(0.004)	(0.007)
Panel B - With Health Controls								
SLE	-0.006	-0.004	-0.015	-0.021	-0.001	0.002	-0.001	-0.004
(s.e.)	(0.004)	(0.003)	(0.007)	(0.010)	(0.004)	(0.002)	(0.004)	(0.007)
Observations	874	1,586	1,614	1,201	820	1,457	1,575	1,008
Mean	0.023	0.021	0.045	0.110	0.013	0.009	0.032	0.063

Notes: The table presents the results from an OLS regression of equation 1 using an indicator for being deceased by December 2019 as the outcome variable. The first four columns are for men, and the last four for women. Panel A is baseline regression and Panel B presents the regression with all health variables included in Table 2. All regressions include the baseline controls (demographic cell, region, and parental education). Heteroskedastic-robust standard errors are included in parenthesis. The mean for each outcome and sample is also shown below the number of observations.

Table 4: SLE in 2004 and subsequent waves

	OLS			IV		
Outcome: SLE in	2006	2009	2012	2006	2009	2012
	(1)	(2)	(3)	(4)	(5)	(6)
SLE in 2004	0.415	0.307	0.189	0.570	0.265	0.249
(s.e.)	(0.055)	(0.052)	(0.045)	(0.133)	(0.107)	(0.098)
Observations	458	458	458	458	458	458

Notes: The table shows the results from a regression of SLE measured in future surveys (2006, 2009, and 2012) on the initial SLE reported in 2004. All regressions included the baseline controls (demographic cell, region, and parental education). The results in the first three columns are from the OLS estimation. The results in columns 4–6 use an IV/2SLS strategy, using the probability of being alive at age 65 as an instrument for SLE in 2004. Heteroskedastic-robust standard errors in parentheses.

diagnosis of diseases, we can see how beliefs vary after a new diagnosis. Results are presented in Table 5. I group diseases according to their severity and type. Group 1 includes new diagnoses for asthma, pulmonary emphysema, diabetes, arthritis, and osteoarthritis. Group 2 includes hypertension, high blood pressure, heart problems, cancer, renal diseases, stroke, and HIV AIDS. Group 3 includes mental illness and depression. We can see that a new diagnosis of diseases in the most severe group (2) is associated with reductions on future SLE by 3.8–4.0 years. We see no difference for group 1 diseases and a reduction of 2.1–2.3 years for group 3.

Table 5: SLE in subsequent surveys and new diagnosis

	Outcome: SLE $_{t+1}$	
	(1)	(2)
New Diagnosis Group 1	0.559	1.015
	(1.692)	(1.567)
New Diagnosis Group 2	-3.984	-3.828
	(1.716)	(1.709)
New Diagnosis Group 3	-2.345	-2.146
	(1.354)	(1.268)
SLE_t		0.326
		(0.025)
Observations	2,507	2,507

Notes: The table shows the results from an OLS regression comparing two SLE reports in two consecutive surveys. The outcome variable is the SLE in the second survey. The explanatory variables are indicators for a new diagnosis of a diseases in each of three groups. Group 1 includes new diagnoses for asthma, pulmonary emphysema, diabetes, arthritis, and osteoarthritis. Group 2 includes hypertension, high blood pressure, heart problems, cancer, renal diseases, stroke, and HIV AIDS. Lastly, group 3 includes mental illness and depression. The second column includes SLE in the first survey as a control. All regressions include the baseline controls (demographic cell, region, and parental education). Standard errors clustered at the individual level are displayed in parentheses.

In summary, individuals answered two questions eliciting beliefs around survival chances that are positively correlated, with the majority of individuals answering both in a consistent way. The beliefs correlate with expected demographics, health and lifestyle behaviors.

Nevertheless, it displays substantial dispersion that cannot be explained by various demographics, health controls, lifestyle behavior, preferences, and skills measures. Importantly, numeracy is not correlated with their reportings after controlling for demographic characteristics. Subjective beliefs of older individuals are associated with future mortality and beliefs display moderate persistence, responding to health shocks.

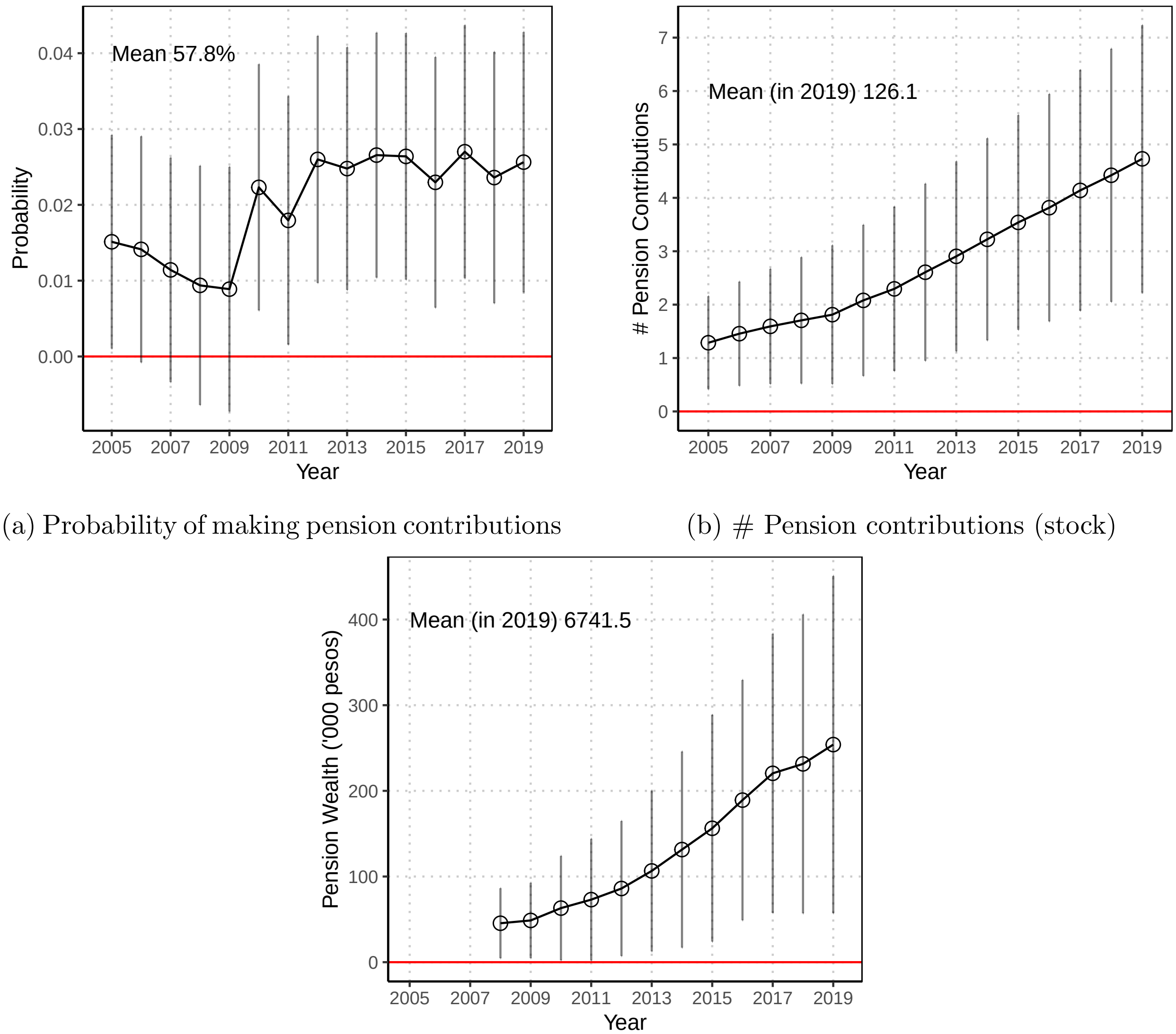
5.2 Labor Market and Pension Contributions

I now explore the richness of the administrative data and compute regression 1 using as outcomes the probability of making pension contributions, the number of contributions (stock), and pension wealth between 2005 and 2019. The results are summarized in figure 3. In the top left plot (3a), we can see the correlation between reporting 10 years higher SLE and the probability of making a pension contribution in any given month. The circles show the β coefficient from the baseline regression, which controls for demographic cells (gender-age-education), region, and parental education fixed-effects. For the first years, the average coefficient is around 1pp and 2.5pp for the remaining sample. That is, individuals who report 10 more years of SLE are more likely to make pension contributions. In terms of magnitude, the overall probability of making a pension contribution is 57.8%. That is, those with 10 years higher SLE are 2–4.5% more likely to contribute. The smaller coefficient in the first years may be attributable to individuals still in schools/universities and navigating the labor market entrance. Nevertheless, I cannot statistically reject that these coefficients are all equal.

In panel A.5b, I show the results for the number of pension contributions individuals have accumulated. Given that those with higher SLE were constantly more likely to make pension contributions, it is unsurprising to see the estimates rising over time. In 2019, those with 10 years more SLE have, on average, 4.7 months more of pension contributions. Which translates to 3.8%, using the average number in December 2019 of 126.1. We can see the same in panel A.5c, where the pension wealth (in thousands of Chilean pesos) is used as the outcome variable.[15] In 2019, those with 10 years higher SLE have a pension wealth 255.1 thousand Chilean pesos higher. Using the mean in December of 2019, that corresponds to 3.8%.

Table 6, panels A and B, present these final results in 2019 for the overall sample and separately for men and women. For each sample (pooled, men, and women) and each variable (rows), the first column presents the outcome mean, and the second column presents the OLS

[15]The pension wealth variable is winsorized at the top 5%, separately for each month.



(a) Probability of making pension contributions

(b) # Pension contributions (stock)

(c) Pension wealth ('000 pesos)

Figure 3: SLE and future pension contributions

Notes: The figure plots the results from the OLS estimation of equation 1. Panel (a) for the binary outcome on the probability of making monthly pension contributions, panel (b) for the number of total pension contributions (stock), and panel (c) pension wealth measured in thousands of Chilean pesos. The regressions are run by pooling all monthly observations but separately for each calendar year. The circles are the estimated coefficient for β, and the solid vertical lines are the 95% confidence intervals. The text in each graph displays the mean for the outcome variable for the entire period (panel a) and for December 2019 (panels b and c). All regressions include the baseline controls (demographic cells, region, parental education, and time fixed effects). Standard errors are clustered at the individual level.

estimation of equation 1. Looking at the pension status on the last date of the sample, in December 2019, there are two summary measures: the number of pension contributions (panel A) and the pension wealth in thousands of Chilean pesos (panel B). Those reporting

10 extra years of SLE have, on average, 4.7 extra months of pension contributions and pension wealth 255.1 thousand pesos higher. Splitting by gender, 10 extra years of SLE is associated with 7.0 more contributions for men and 2.2 for women.

Table 6: SLE and future labor market outcomes

	All			Men			Women		
	Mean	OLS	IV	Mean	OLS	IV	Mean	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. # Pension Contributions (stock) in Dec2019									
SLE	126.099	4.730	5.999	139.685	7.018	6.164	111.862	2.256	5.956
(s.e.)		(1.278)	(2.772)		(1.776)	(4.209)		(1.799)	(3.704)
Num.Obs.		1750	1739		894	888		856	851
Panel B. Pension Wealth ('000 pesos) in Dec2019									
SLE	6741.5	255.1	490.8	7844.6	445.1	796.5	5585.7	47.4	286.5
(s.e.)		(111.5)	(234.3)		(167.3)	(371.5)		(147.4)	(302.1)
Num.Obs.		1750	1739		894	888		856	851
Panel C. Labor Force Participation									
SLE	0.758	0.010	0.011	0.840	0.009	0.007	0.671	0.010	0.013
(s.e.)		(0.005)	(0.011)		(0.005)	(0.013)		(0.008)	(0.016)
Num.Obs.		1750	1739		894	888		856	851
Panel D. Formal Sector									
SLE	0.482	0.015	0.030	0.545	0.021	0.026	0.415	0.008	0.032
(s.e.)		(0.006)	(0.013)		(0.009)	(0.021)		(0.008)	(0.017)
Num.Obs.		1750	1739		894	888		856	851

Notes: The table presents the outcome mean (columns 1, 4, and 7), the OLS estimation of equation 1 (columns 2, 5, and 8), and the IV/2SLS estimation of 2 (columns 3, 6, and 9). The first 3 columns refer to the entire sample, the next 3 columns refer to the sample restricted to men and the final 3 for the sample restricted to women. There are four panels, each with one variable. Respectively, the total number of pension contributions in December 2019, pension wealth in December 2019 in thousands of Chilean pesos, share of periods in the labor force and share of periods in formal employment (in a firm or the public sector). For the OLS and IV columns, the β coefficient is displayed with the estimated standard error (clustered at the individual) level in parenthesis. All regressions include the baseline controls (demographic cell, region, and parental education fixed effects).

As discussed in Section 4, these coefficients are likely attenuated by measurement error in the self-reported life expectancy. To address this concern, I leverage the fact that the survey contains two different questions on life expectancy placed in different modules. I then use the answer to the second question, the probability of being alive at age 65, as an instrument for the SLE variable. The results are presented in the third column for each sub-panel in table 6. Across the board, we see higher coefficients in the IV/2SLS strategy than in the OLS, as would be the case if there is significant measurement error. However, the estimates are more imprecise, partially coming from the ceiling at 100% for the probability question. Table A.2 shows the first stage for these estimates. The F-statistic for the pooled sample is 470.0 For the analysis by gender, the F-statistics are 183.3(men) and 283.0 (women). The IV estimates imply that those with (predicted) 10 years higher SLE have 4.8% higher pension contributions for the pooled sample and, respectively, 4.4% and 5.3% for men and women. In terms of pension wealth, these numbers compound to 7.3% (pooled), 10.2% (men), and 5.1% (women).

Panels C and D explore the survey data on employment to shed some light on the mechanisms that explain these results. To avoid issues with recollection, I only use employment information reported not later than 12 months. For each individual, I compute the share of months that they are participating in the labor force and the share of months they are working formally. Those in the formal sector, that is, working formally in a firm or being a public employee, have mandatory pension contributions. In Panel C, we can see that individuals spend on average 75.8% of months in the labor force (84.0% for men and 67.1% for women). Individuals reporting 10 extra years of SLE have 1.0pp (OLS) and 1.1pp (IV) higher labor force attachment. In Panel D we can see that the average share of months working formally is 48.2%. If the effect was entirely driven by labor force participation we should see coefficients for these regression that were 0.482/0.758=0.636 smaller than the ones in Panel C. We see that exact opposite. Those with 10 extra years of SLE are 1.5pp (OLS) and 3.0pp (IV) more likely to be working formally. That is, even though labor market participation is higher for those with higher SLE, employment type — whether working formally — accounts for a larger share of the documented pension gaps. If we implement a mediation analysis, we would see that labor force participation accounts for 22.9% of the effect on pension contributions, while type of employment accounts for 76.6% of the effect.

In the survey individuals also respond questions related to their knowledge of the pension system. These questions include the normal retirement age for men and women, the value for the mandatory contribution rates and how the pension benefits are computed. I use indicators for correctly answering these questions as outcome variables, using the main specification. Results are presented in Appendix Table A.3. Individuals with higher SLE are more likely to correctly answer the age-threshold for retirement, but not necessarily the contribution rates or the way to compute pension benefits (positive but not statistically significant coefficients).

5.3 Alternative Empirical Strategies

The results in the last section show how SLE measured when individuals were between 18 and 26 years of age are strongly correlated with pension outcomes 15 years later and with labor market outcomes in this entire period. Even though the main specification controls for a rich set of demographics (gender-age-education cell, region, and parental education fixed-effects), it is still possible that there are unobservables that are correlated both with the initial beliefs and the later outcomes. In this section I use several additional empirical strategies to investigate how robust this association is and under which assumptions does it capture a causal effect.

There are four different sets of exercises. The first aims to strengthening the conditional independence assumption by including additional covariates and investigating in which conditions the unobservables would be strong enough to drive the results. The second explores the longitudinal dimension of the data with the subsequent beliefs measured with an within-individual empirical strategy. This approach is interesting as it can deal any time-invariant unobservables. The third and fourth empirical strategies will rely on IV approaches using health of family members and parental deaths as instruments for the initial life expectancy belief.

5.3.1 Conditional Independence Assumption

Under the conditional independence assumption, the estimated coefficient in equation 1 can be interpreted as a causal effect of subjective life expectancy on the outcome variable. For that assumption to hold, there cannot be any unobservables that are correlated with the initial beliefs and the outcome variable. Table 7 shows the coefficient of the OLS estimation of equation 1 with different set of controls. The only significance impact is when adding the gender variable. The coefficient goes from 6.959 to 3.909. This is not surprising, as we saw stark gender difference on beliefs and on labor market participation. The rest of the columns shows how remarkably stable the coefficient is, ranging from 3.909 to 4.730 when a variety of different variables are introduced. Including several health, lifestyle and personal traits measures.

Importantly these variables are relevant for the outcome measure. Together they raise the R-square from 1.6% to 25.6%. With these results we can apply the procedure presented by Oster (2019) to assess how important the unobservables must be to drive these effects. Comparing the point estimates and R-square from the first and last column we conclude that the unobservables need to be 72% as important as all the observed variables to drive

Table 7: Additional control variables

	No controls	+Gender	**Baseline**	+Health	+Parental Death	+Risk-Aversion	+Numeracy & Pref	+Big-Five
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SLE	6.959	3.909	4.730	4.577	4.296	4.234	3.954	3.960
(s.e.)	(1.280)	(1.338)	(1.278)	(1.293)	(1.289)	(1.292)	(1.287)	(1.283)
Obs	1750	1750	1750	1750	1750	1750	1750	1750
R^2	0.016	0.047	0.211	0.222	0.237	0.239	0.246	0.256

Notes: Table presents estimates of the main equation with different sets of control variables. The first column shows the coefficient of SLE, when no additional control is included. Second columns includes gender and the third column all the baseline controls (gender-age-education, region, and parental education fixed effects). The next follows the pattern of Table 2 including sequentially, all health & life-style variables (column 4), parental death (column 5), risk-aversion (column 6), numeracy and future-oriented preferences (column 7), and the Big-Five personality traits (column 8). Standard errors are clustered at the individual level.

the results to zero.[16] Another approach to assess the plausibility of unobservables driving the results is the one proposed by Cinelli and Hazlett (2020). Using their approach, unobservables need to explain at least 8.6% of residual variance of both the outcome variable and the SLE to drive the results to zero. Note how gender, the variables with the most impact in our analysis would not pass this criteria as it only explains 3.1% of the outcome variable. No other observable variable comes close to this threshold.

5.3.2 Longitudinal approach

In this section I exploit the longitudinal dimension of the survey to explore alternative identification strategies. I first use jointly the initial and beliefs elicited in subsequent surveys. Results are presented in Table 9. The first column reproduces the main effect presented in Table 6 — 10 extra years of SLE in 2004 are associated with around 4.7 extra months of pension contribution in December 2019. Column 2 adds the observed SLE in 2006 as a control. The estimated coefficient shows that controlling for the initial SLE in 2004, an extra 10 years of subjective expectations in 2006 are associated with further increases of 2.5 months in pension contribution. The estimates are statistically significant at the 10% level. Similar results are displayed for SLE in 2009 and 2012. The last four columns show the same exercise as the IV approach, using the reported probability of being alive up to 65 as an instrument. Column 5 only reproduces the IV results from Table 6. Columns 6–8

[16]This considers an R^2 maximum of 0.700 for any model explaining the stock of pension contributions. Appendix Figure A.3 presents these value for several values of the maximum R^2. With my rich set of covariates the maximum reached is 0.256.

use as instruments the P^{65} variable in 2004 and 2006/2009/2012. As we saw before, the estimates are overall higher when using the instrument to correct for measurement error. Therefore, this evidence is consistent with individuals acting in ways aligned with revisions of their subjective expectations; those revising their SLE upwardly exhibit more pension contributions.

Table 8: SLE and future revisions

	Outcome: # Pension Contributions (stock) in Dec2019							
	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SLE in 2004	4.730	3.948	4.422	4.301	5.999	3.593	3.095	4.165
	(1.278)	(1.360)	(1.306)	(1.284)	(2.772)	(3.547)	(3.166)	(2.894)
SLE in 2006		2.535				6.285		
		(1.513)				(4.745)		
SLE in 2009			1.815				18.134	
			(1.938)				(6.611)	
SLE in 2012				3.515				12.002
				(2.061)				(6.180)
Observations	1750	1750	1750	1750	1739	1739	1739	1739

Notes: The table shows the results from regressions where the outcome variable is the total number of pension contributions in December 2019. The first four columns are the results of an OLS estimation. The first column only includes the baseline regressor, SLE in 2004. The next three columns include SLE measured in 2006, 2009, and 2012. The last four columns are from an IV/2SLS estimation using the reported probability of living at age 65 in the two surveys as instruments. For instance, in column 6, the instruments are P^{65}_{2004} and P^{65}_{2009}. All regressions include the baseline controls (demographic cell, region, and parental education). Hetoreskedastic-robust standard errors are presented in parentheses.

Another exercise exploring the panel dimension of the data is presented in Table 9. In this exercise, I explore variation in SLE solely coming from time variation, by augmenting the main specification to include individual fixed-effects. This specification addresses the concerns on omitted variables that may correlate with SLE but are permanent characteristics of individuals. For every individual, I consider all SLE reports across the waves and use the cumulative pension contributions at the time of each report as the outcome variable. The first column shows that 10 extra years of within-individuals SLE are associated with 1.4 extra months of cumulative pension contributions. In the 2SLS/IV specification in the second column, I instrument each SLE with the contemporaneous report on the probability of being alive at age 65. We can see that 10 extra years of (predicted) within-individual SLE are associated with an increase of 9.2 months of pension contribution. The distance across two reports is, on average, 31.7 months. These results are similar to the one obtained using the cross-sectional variation. Note that these estimates explore shorter time frame, as it only captures changes in outcomes in between survey, instead of the levels in the entire 15-year period as the other approach.

Table 9: SLE and Pension Contributions — Within-individual variation

Outcome: # Pension Contributions (stock)		
	OLS	IV
	(1)	(2)
SLE	1.377	9.227
	(0.603)	(2.003)
Observations	3,067	3,066

Notes: The table shows the results of a regression where the outcome variable, the stock of pension contributions, is regressed on the SLE variable, individual fixed effects. The second column presents a 2SLS/IV estimation, using the probability of being alive at age 65 as an instrument for SLE. Standard errors clustered at the individual level are in parenthesis.

The results in this section explore how beliefs evolve overtime and are robust to invariant unobservables that are correlated with the beliefs and pension contributions. Examples of concerns that are alleviated with this approach are fixed preferences, skills, and personal traits that were not captured with the questions in the survey. One concern with this approach though is that health shocks may drive the changes in beliefs but also impede individuals to work and make pension contributions. To break this link of shocks potentially affecting the subjective beliefs at the same time as affecting likelihood of work, I turn now to investigate health of family members in an IV approach.

5.3.3 Instrumental Variables — Health and mortality of family members

In this section, I take advantage of the comprehensive data collection in the survey and use information on mortality of parents and general health of other family members in the initial wave. In the first approach I use parental deaths as an instrument for the initial SLE. In the first column of Table 10 we can see the first stage. Having deceased parents decreases the initial SLE by 3.05 years.[17] The F-statistic is 8.11, slightly below the typical threshold of 10. The second column show the IV result for the instrumented SLE on the stock of

[17] The SLE variable is measured in 10 years.

pension contributions in December 2019. Extra 10-years of (predicted) SLE corresponds to 35 extra monthly pension contributions. In order for this instrument to be valid, the only channel through which parental deaths could affect pension contributions is through life expectancy. This assumption would be violated if for instance, through parental deaths, individuals change their behavior in the labor market due to lower family support, lower family income, or due to receiving a large inheritance. With this data, it is difficult to assess the plausibility of these channels, it is worth mentioning though, that some of them would act in the opposite direction of the current findings.

Table 10: Instrumental Variables: Mortality and health status of family members

	Parental Death		Health family members		Health family members	
Stage	First	Second	First	Second	First	Second
	(1)	(2)	(3)	(4)	(5)	(6)
SLE		35.891		25.428		22.208
		(21.311)		(17.087)		(18.699)
Deceased parents	-0.305					
	(0.118)					
Health family members			0.149		0.144	
			(0.049)		(0.053)	
Observations	1,528	1,528	1,303	1,303	1,196	1,196
F-stat	8.110		10.500		9.060	

Notes: The table shows the results from regressions where the outcome variable is the total number of pension contributions in December 2019. The first two columns show, respectively the first and second stage when using parental death as the instrument variable. The third and fourth colum show the results when using as instrument the average health assessment of family members (spouse or parents). The fifth and sixth columns show the same, but removing any cases where any family member reported to be invalid for work. Heteroskedastic-robust standard errors are presented in parentheses.

The next set of results use as instruments the average health status of spouses and parents that were also interviewed.[18] I use their self-reported health status, in the 1–6 scale, as the instrument. The third column shows that 1 extra point in the average health status is associated with an increase of 1.49 years of SLE. The F-statistic is 10.5. The second stage result in the fourth column shows a coefficient of 25.42. One potential violation of the exclusion restriction is if those family members with worse health status require more care from individuals, which might influence their labor market status. To alleviate this concern, In the fifth and sixth columns I repeat this exercise excluding individuals who live with family members that reported to have any disability or to be handicap. The results are very similar. The magnitude of the coefficients estimate with these three instrumental

[18]All family members living in the same place were interviewed.

variables approach are 3–6 times larger than the baseline results with the cross-section OLS approach. However, they are much more imprecise, with F-statistics for the first stage just around the typical thresholds considered.

5.4 Robustness

In this section I present several robustness checks using as benchmark the baseline estimate using the cross-sectional OLS estimation. The first is to understand the role of the initial status in the labor market. In Figure 3, we can see that, using the baseline specification, individuals reporting higher SLE were also more likely to contribute to pensions and had higher stock already in 2005. This is expected, as it has been explored throughout the paper, individuals who expect to live longer have behaviors in the labor market consistent with valuing stable jobs with pension contributions. However, one can be concerned that this initial status may dominate future behavior and is not connected to life expectancy per se. In Figure A.5, I reproduce the main results controlling additionally for the initial labor market status. The results are smaller but have the same overall pattern. Even conditioning on the same initial status in the labor market, those with higher SLE are more likely to contribute to pensions, accumulate more pension wealth, participate in the labor force, and work in formal jobs. I do not include initial status as a control in the main specification, as the contemporaneous choices in the labor market could already be chosen in response to each individual's beliefs and could be argued as part of the main effect.

Table A.4 also shows how the results do not depend on the chosen age range. The estimated coefficients are similar for a variety of different age ranges. Table A.5 shows how we obtain similar results when using the probability of being alive at age 65 as the main variable. To facilitate the comparison, for the probability at age 65, I compute the coefficient associated with increases of 30pp in this probability, as this corresponds to the 75th-25th percentile gap. The overall pattern is very similar across the two distinct measures. Another concern could be with the chosen specification, where SLE enters linearly. I use the non-parametric bins-regression proposed by Cattaneo et al. (2024) to assess this hypothesis. The results are presented in Figure A.4. We can see that the results from the linear specification, represented by the solid black curve, are within the 95% confidence intervals for the pooled sample and separately for men and women. Lastly, Appendix B investigates the role of non-reporting in the SLE variable, not finding any consistent pattern between missing and the observed outcomes.

6 Theoretical Framework

In this section, I develop a simple theoretical framework that shows how ignoring heterogeneity in (subjective) life expectancy can lead to biased conclusions in life-cycle models. In this model, individuals live in a younger and an older stage of life. The instantaneous utility in both periods is u, a concave, increasing, and differentiable function. In the younger stage, an individual receives an exogenous income Y and sets the amount of savings a she can consume in the older stage. The interest rate and the policy environment are such that, to get a stream of income a in the older stage, the agent needs to invest τa in the younger stage. This can encompass private investment for retirement, including working and making social security contributions. There are two states of the world in the older stage. The good state, where individuals live "longer", is realized with probability π. To capture a longer life span in a simple two-period model, I assume that the good state generates H times more utility than the bad state, where $H > 1$.[19] Finally, the relative value of the older stage of life with respect to the younger stage is β. Therefore, an individual faces the following problem:

$$\max_{a\in[0,Y/\tau]} \left\{ u(Y - \tau a) + \beta\big(\pi H u(a) + (1-\pi)u(a)\big) \right\} \tag{5}$$

Since the optimal amount of savings depends on how likely an agent believes she will live longer, we can define the $a^*(\pi)$ as

$$a^*(\pi) = \arg\max_{a} \left\{ u(Y - \tau a) + \beta\big(\pi H u(a) + (1-\pi)u(a)\big) \right\}. \tag{6}$$

That is, this function records, for any π, the optimal level of investments across the two periods.

Figure 4 left plot shows the $a^*(\pi)$ function for a given choice of parameters.[20] We can see that individuals with higher π choose higher values of a. This is consistent with the evidence presented in this paper — young individuals with higher SLE are more likely to make pension contributions.

We often do not observe individual π and solve the model for the average $\mathbb{E}[\pi]$ in this economy. For example, in the case of life expectancy, it is common to obtain $\mathbb{E}[\pi]$ directly from life tables for some sub-groups, for example, men and women. The rationale of

[19]Note, for instance, that if we assume that individuals live for two periods in the good state, that is equivalent of assuming $H = 1 + \beta$, if there are no savings between these two periods.

[20]For this example, $Y = 1$, $\tau = 1$, $\beta = 0.97$, $H = 1 + \beta + \beta^2$, $u(c) = c^{(1-\sigma)}/(1-\sigma)$, $\sigma = 3$, and $\pi \sim U[0.001, 0.999]$.

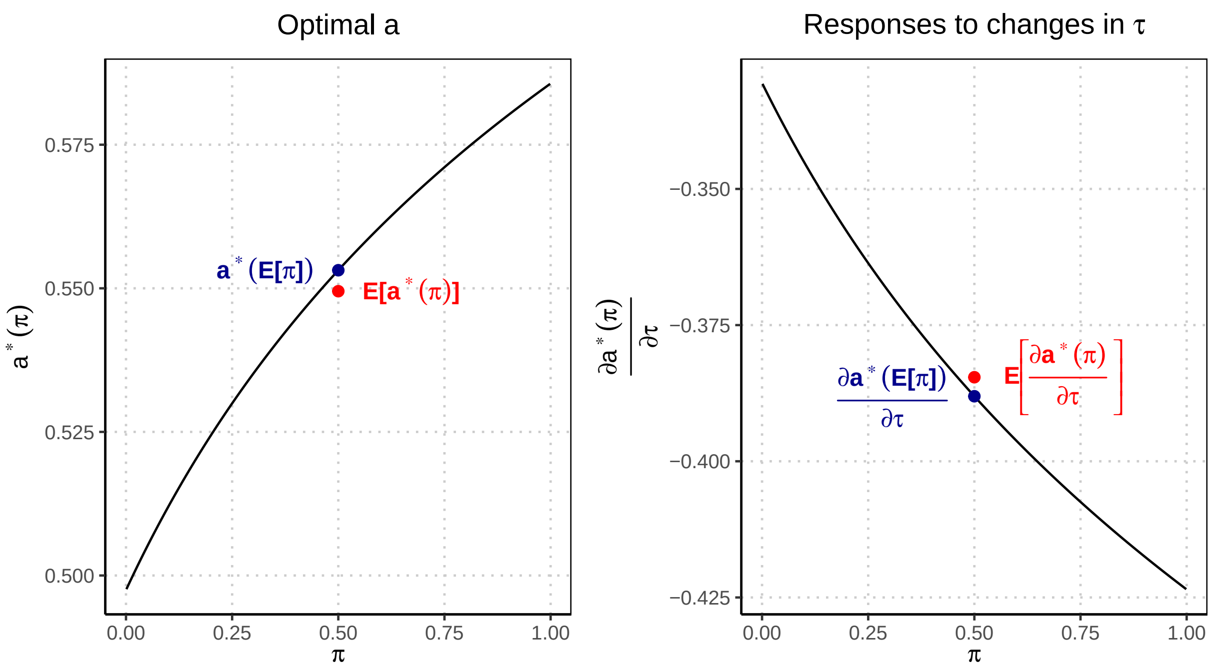


Figure 4: Theoretical Framework

Notes: The plot on the left shows the optimal choices of a as a function of the probabilities π (solid line). The blue point shows the optimal choice for the average value of π. The red point shows the average optimal choice across all values of π. The plot on the right has the same structure, plotting the changes in optimal response to changes in τ.

using $\mathbb{E}[\pi]$ is to assume that individuals have rational expectations and *do not have private information over own survival chances.* The blue circle shows the prediction of a correctly specified model that uses only $\mathbb{E}[\pi]$, ignoring heterogeneity over π. That is, it just computes the optimal behavior function $a^*(.)$ at the average π. In contrast, the red circle shows the true average optimal choice over the distribution of π. Given the concavity of the $a^*(\pi)$ function and Jensen's inequality, we have that $a^*(\mathbb{E}[\pi]) \geq \mathbb{E}[a^*(\pi)]$, where expectations are taken over the distribution of π.[21] This exercise shows that models that ignore heterogeneity in subjective life expectancy and assume a representative agent with average life expectancy always overestimate the optimal average savings for retirement. This result reiterates the importance of the empirical findings, showing how individuals with different beliefs have different labor market and social security choices.

This theoretical framework can also be used to assess policy changes, particularly around social security. We can interpret τ as a reduced-form parameter relating investments

[21]Concavity of $a^*(\pi)$ does not depend on the specific parameters chosen. It inherits the concavity of $u(c)$. Appendix C shows how this holds for a large class of utility functions.

for retirement to pension benefits. Social security reforms that decrease benefits or strengthen the requirements to access public pensions can be seen as increases in τ. Using the structure above, we can assess the changes in the optimal choice of investment from changes in τ: $\frac{\partial a^*(\pi)}{\partial \tau}$. The right plot from Figure 4 plots the values of this derivative for each value of π. When τ increases, all individuals reduce their investments for period 2, as transferring resources across the two periods is more costly. Individuals with higher π exhibit larger changes, reducing a^* to a greater extent. Similarly to the above-mentioned arguments, we can contrast the predictions of a model that ignores heterogeneity in π. A model considering the response only on the average $\mathbb{E}[\pi]$ would overestimate the behavioral changes to policies that enhance τ. This could affect policy-making, as several policies could be decided upon the prediction of models that have biased predictions of individuals' responses to the policy changes.

In summary, this simple two-period model shows how individuals with different life expectancy beliefs adopt different strategies relating to retirement. This is consistent with this paper's findings and, more broadly, with the literature analyzing workers closer to retirement. Moreover, this model highlights how ignoring heterogeneity in life-span beliefs can be consequential. "Naive" models, even if correctly specified and with the true parameters, would produce biased results. Such models would overestimate the optimal savings rates and behavioral responses to policy changes. This may lead to the adoption of suboptimal policies, as they wrongly predict the optimal response of individuals to the policy environment.

7 Discussion and Conclusion

While most of the literature focuses on subjective life expectations of individuals near retirement, in this paper, I explore longitudinal survey data and administrative data from Chile to investigate SLE of young individuals. I show how even if these beliefs are elicited from a young population, they still present good properties: they are internally consistent, correlate with expected behavior, are modestly persistent, respond to new information, and predict future mortality.

The above results showed that individuals who reported beliefs of longer life spans are also more likely to be employed in formal jobs in the private and public sectors, with mandatory pension contributions. After 15 years of the initial report, the 75th-25th gap in SLE translates to individuals having 5.1%–10.2% higher pension wealth. These results are consistent with individuals making labor market choices that depend on how beneficial they are in the future. Individuals who believe they will live longer value making tax-advantaged pension contributions more. The flip side is that individuals with shorter expectations will value more jobs that do not require those pension contributions.

In order to interpret the estimated coefficients as causal links between subjective expectations and life choices, one needs to assume no relevant omitted variable. What the above exercises could show us is that these associations seem to be very stable, even when we include a rich set of covariates that are very relevant to the outcomes being measured. Using the approaches suggested by Oster (2019) and Cinelli and Hazlett (2020), we saw how strong unobservables must be to drive these results. This seems unlikely analyzing the importance of very relevant observables such as gender and education.

Nevertheless, we saw that alternative empirical strategies exploring solely within-individual variation in SLE, or using family members' health and parental death as instruments produce results that are in the same direction of the cross-sectional results. These approaches relax the conditional independence assumption but are each based in different hypothesis. One challenge is, for instance, if reporting beliefs of longer life spans is also correlated with other beliefs that are important for determining behavior. For example, Puri and Robinson (2007) show how SLE can be correlated with general optimism. While I do not have any optimism variable to investigate this in my sample, all the results presented here seem consistent with the large number of papers reviewed by Hudomiet et al. (2023), where these beliefs indeed predict mortality. So, it seems these variables capture agents' private information over their survival chances. Additionally, the general optimism hypothesis would be difficult to reconcile with some of my results, for instance, the large drops in SLE following a new diagnosis and the within-individual variation results.

From a measurement perspective alone, it is interesting that SLE measured at young ages correlates so well with future labor market choices. On its own, this fact already advocates for the broader use of these measures in economics and policy. Given how consequential early labor market choices are for career progression and overall welfare, these results ask for new research to further explore how these beliefs are formed, their accuracy, and their consequences.

REFERENCES

Altonji, J. G., Kahn, L. B. and Speer, J. D. (2016). Cashier or consultant? entry labor market conditions, field of study, and career success, *Journal of Labor Economics* **34**(S1): S361–S401.

Arellano-Bover, J. (2020). The Effect of Labor Market Conditions at Entry on Workers'

Long-Term Skills, *The Review of Economics and Statistics* pp. 1–45.
URL: *https://doi.org/10.1162/rest_a_01008*

Bello, P. (2023). Gender-based price discrimination in the annuity market: Evidence from chile, *European Economic Review* **151**: 104356.

Bissonnette, L., Hurd, M. D. and Michaud, P.-C. (2017). Individual survival curves comparing subjective and observed mortality risks, *Health economics* **26**(12): e285–e303.

Bloom, D. E., Canning, D., Moore, M. and Song, Y. (2006). The effect of subjective survival probabilities on retirement and wealth in the united states, *Technical report*, National Bureau of Economic Research.

Bresser, J. d. (2021). Evaluating the accuracy of counterfactuals the role of heterogeneous expectations in life cycle models.

Cattaneo, M. D., Crump, R. K., Farrell, M. H. and Feng, Y. (2024). On binscatter, *American Economic Review* **114**(5): 1488–1514.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**(1): 39–67.

Delavande, A., Lee, J. and Menon, S. (2017). Eliciting survival expectations of the elderly in low-income countries: Evidence from india, *Demography* **54**(2): 673–699.

Finamor, L. (2024). Labor market informality, risk, and public insurance.

Gan, L., Gong, G., Hurd, M. and McFadden, D. (2015). Subjective mortality risk and bequests, *Journal of Econometrics* **188**(2): 514–525.

Gosling, S. D., Rentfrow, P. J. and Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains, *Journal of Research in personality* **37**(6): 504–528.

Hamermesh, D. S. (1985). Expectations, life expectancy, and economic behavior, *The Quarterly Journal of Economics* **100**(2): 389–408.

Hamermesh, D. S. and Hamermesh, F. W. (1983). Does perception of life expectancy reflect health knowledge?, *American Journal of Public Health* **73**(8): 911–914.

Heimer, R. Z., Myrseth, K. O. R. and Schoenle, R. S. (2019). Yolo: Mortality beliefs and household finance puzzles, *The Journal of Finance* **74**(6): 2957–2996.

Hudomiet, P., Hurd, M. D. and Rohwedder, S. (2023). Mortality and health expectations, *Handbook of Economic Expectations* pp. 225–259.

Hurd, M. D. and McGarry, K. (2002). The predictive validity of subjective probabilities of survival, *The Economic Journal* **112**(482): 966–985.

Hurd, M. D., Smith, J. P. and Zissimopoulos, J. M. (2004). The effects of subjective survival on retirement and social security claiming, *Journal of Applied Econometrics* **19**(6): 761–775.

Ministerio del Trabajo y Prevision Social (2002-2020). Encuesta de Proteccion Social.

O'Dea, C. and Sturrock, D. (2023). Survival pessimism and the demand for annuities, *The Review of Economics and Statistics* **105**(2): 442–457.

O'Donnell, O., Teppa, F., Van Doorslaer, E. et al. (2008). Can subjective survival expectations explain retirement behaviour, *DNB WP188* .

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* **37**(2): 187–204.

Puri, M. and Robinson, D. T. (2007). Optimism and economic choice, *Journal of financial economics* **86**(1): 71–99.

Salm, M. (2010). Subjective mortality expectations and consumption and saving behaviours among the elderly, *Canadian Journal of Economics/Revue canadienne d'économique* **43**(3): 1040–1057.

Superintendencia de Pensiones, Chile (1980–2019). Historia Previsional de Afiliados Activos, Pensionados Y Fallecidos.

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

Wu, S., Stevens, R. and Thorp, S. (2015). Cohort and target age effects on subjective survival probabilities: Implications for models of the retirement phase, *Journal of Economic Dynamics and Control* **55**: 39–56.

Online Appendices

A Definition of variables

Risk-Aversion — The indicator for risk-aversion is 1 if individuals answered that preferred a fixed wage X over a lottery that pays $2X$ with probability 0.5 and $.75X$ with probability 0.5.

Numeracy — The numeracy variable is constructed summing whether individuals responded correctly 6 questions on basic math and financial concepts. The range of the variable goes from 0 (no correct answer) to 6 (all correct answers). There are only integer values with no partial correct answer. These questions were first introduced in the 2006 wave.

Future-oriented Preferences — This is an indicator equal to 0 if individuals answered that they only consider the next months or equal 1 if they considered any period longer than that (next year, next years, next 5-10 years, more than 10 years) to the question of how far in the future they consider when planning their investments and expenses. This question was introduced in the 2009 wave.

Big-Five measures — The big-five measures are constructed using the 10 questions on the big-five personality traits using the *ten item personality measure (TIPI)* proposed by Gosling et al. (2003). The dimensions are agreeableness, conscientiousness, emotional stability, extraversion, and openness. They each range from 1–7.

B Missing Subjective Life Expectancy

While the proportion of missing values for the subjective life expectancy variable is small (9%), in this section, I investigate who are the individuals not answering this question.

Columns 1 and 2 of Table A.1 presents the proportion of missing overall (first column) and by gender and education (column 2). There is no gender difference and a negative gradient in education. Those with more education are more likely to respond this question. Columns 3 and 4 show that answering or not this questions has no consistent relationship with our main outcome variables: the total number of pension contributions (column 3) and pension wealth (column 4) in December of 2019. The coefficients are not statistically significant and go in opposite directions.

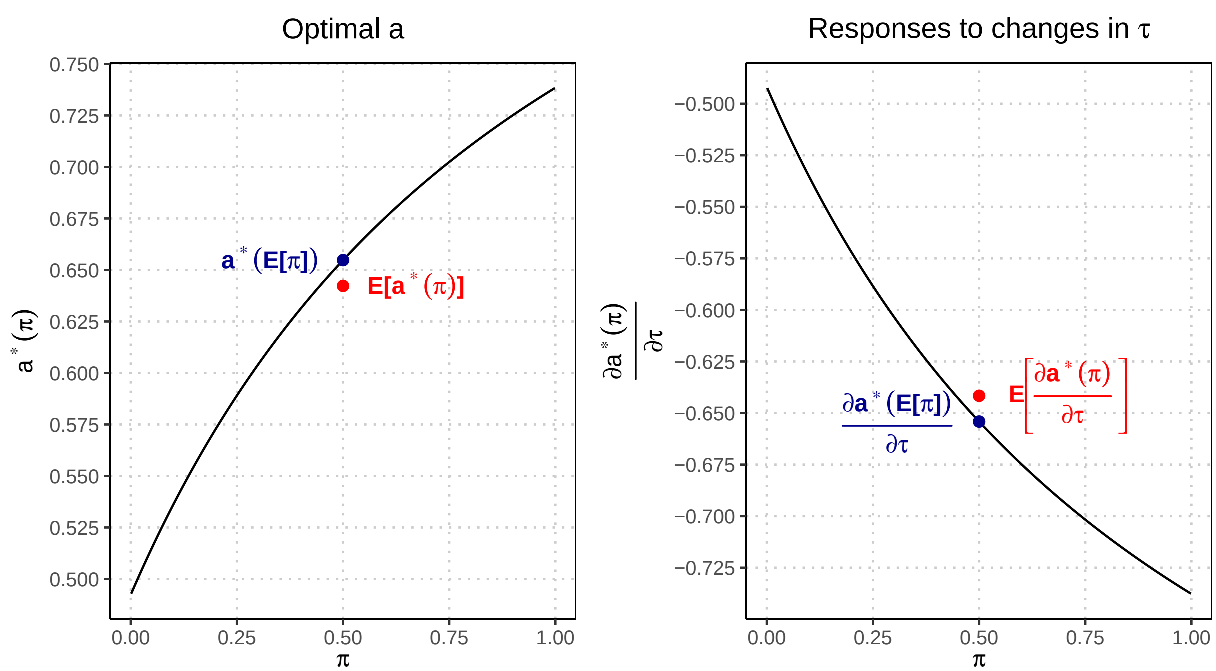
Table A.1: Missing values for SLE

Outcome:	Missing SLE (1)	Missing SLE (2)	# Pension Contrib (3)	Pension Wealth (4)
Constant	0.091	0.132		
	(0.007)	(0.024)		
Women		0.000		
		(0.013)		
High-School		-0.032		
		(0.025)		
Vocational		-0.041		
		(0.029)		
College		-0.089		
		(0.025)		
Missing SLE			1.191	-247.557
			(5.675)	(395.823)
Main Controls	-	-	✓	✓
Observations	1,925	1,925	1,925	1,925

Notes: The table shows the results of an OLS regression using an indicator for not reporting the SLE as the outcome variable. In the first column, only a constant is included. The second column adds an indicator for women and each educational level. In this column, men without high school degrees are the omitted category. Heteroskedastic-robust standard errors are presented in parentheses.

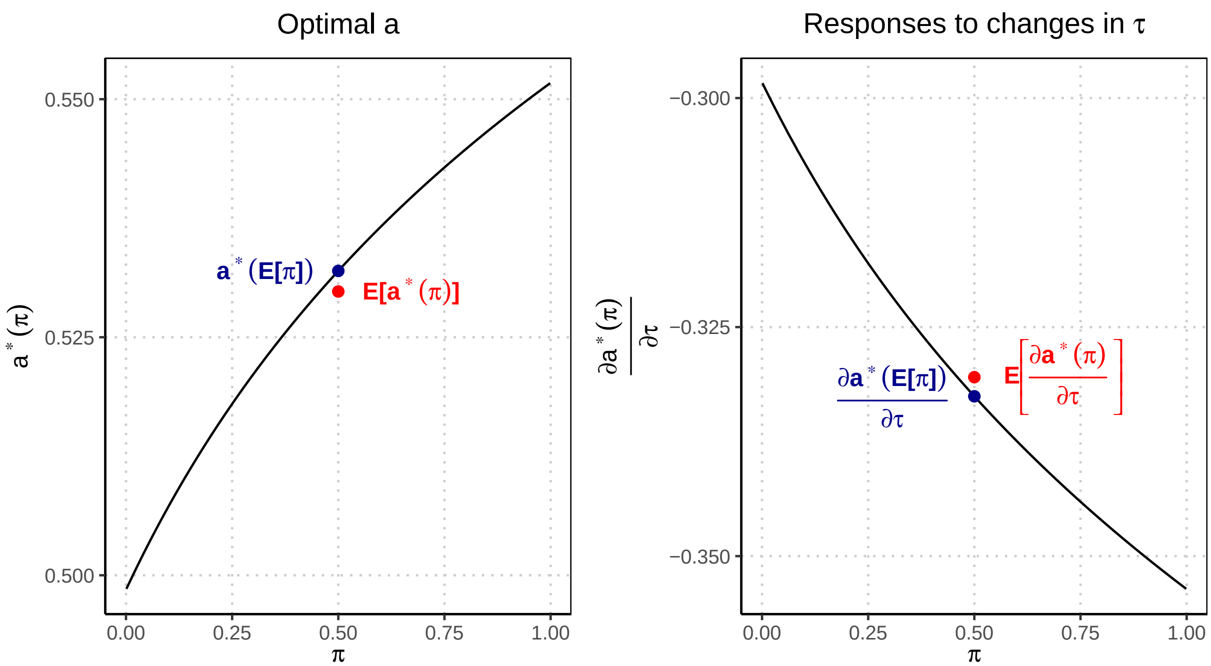
C Theoretical Framework — Other utility functions

The figure below shows the same results from section 6 with different utility functions.

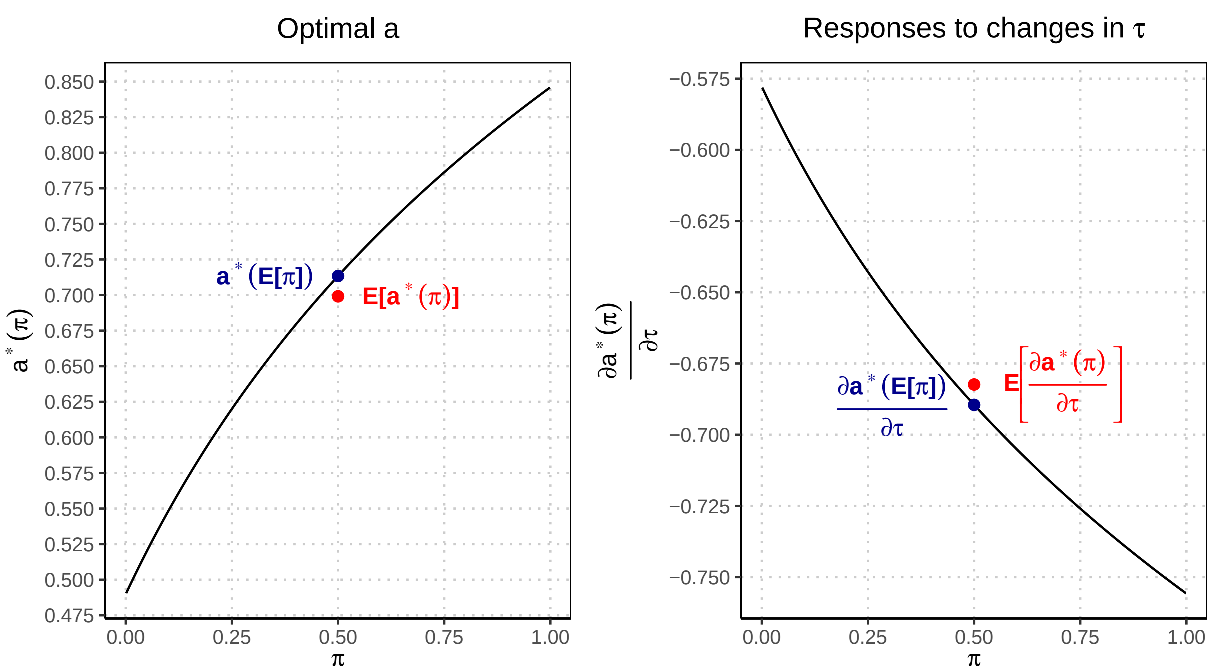


(a) Log-utility, $u(c) = log(c)$

Figure A.1: Theoretical framework with different utility functions



(b) CRRA, $u(c) = c^{(}1-\sigma)/(1-\sigma)$, $\sigma = 5$



(c) CARA, $u(c) = 1 - \exp(-\alpha c)$, $\alpha = 1.5$

Figure A.1: Theoretical framework with different utility functions (cont.)

D Additional Figures and Tables

Table A.2: Instrumental variables — first stage

	Outcome: SLE		
	All	Men	Women
	(1)	(2)	(3)
Prob Living 65	0.249	0.236	0.260
(s.e.)	(0.014)	(0.020)	(0.019)
Observations	1,739	888	851
F-stat	470.0	183.3	283.0

Notes: The table presents the results of the first stage IV/2SLS regression from equation 2. The SLE variable is being instrumented by the P^{65} variable. The F-statistic is displayed in the last row. The first column is for the entire sample, and the next two are for men and women. Standard errors are clustered at the individual level.

Table A.3: Knowledge of the pension system

	OLS				IV			
Outcome:	Ret Age Men	Ret Age Women	Contrib Rate	Pension Formula	Ret Age Men	Ret Age Women	Contrib Rate	Pension Formula
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SLE	0.013	0.011	0.005	0.011	0.036	0.059	-0.005	0.008
	(0.010)	(0.009)	(0.010)	(0.006)	(0.020)	(0.021)	(0.021)	(0.011)
Observations	1,750	1,750	1,750	1,750	1,739	1,739	1,739	1,739
Constant	0.709	0.686	0.359	0.078	0.709	0.686	0.359	0.078

Notes: The table presents the results of the OLS regression of equation 1 (columns 1-4) and the IV regression (columns 5-8). The dependent variable are indicators for correct answers to the questions about the pension system. The questions are: the retirement age for men (columns 1 and 5), for women (columns 2 and 6), the mandatory contribution rates (columns 3 and 7), and the pension formula (columns 4 and 8). Heteroskedastic-robust standard errors are in parentheses.

Table A.4: Robustness — age range

	Outcome: # Pension Contributions in December 2019 (stock)						
Age range	18–26	18–24	18–22	18–28	18–30	20–26	22–26
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SLE	4.730	4.048	2.866	3.937	3.306	5.106	5.465
	(1.278)	(1.418)	(1.854)	(1.180)	(1.127)	(1.348)	(1.509)
Observations	1,750	1,249	669	2,308	2,944	1,618	1,324

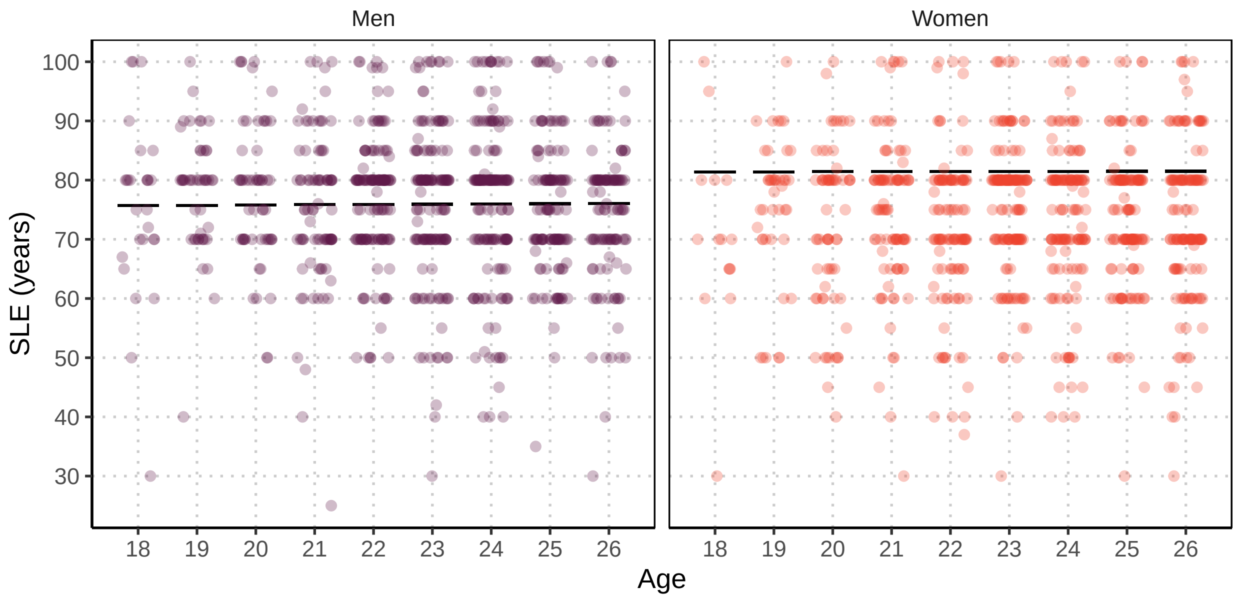
Notes: The table shows the results of the OLS estimation of equation 1 using the total number of pension contributions (stock) in December 2019 as the outcome variable. Each column uses a sample with a different age range, indicated in the top header. All regressions include the baseline controls (demographic cells, region, and parental education). Standard errors are heteroskedastic-robust.

Table A.5: Future labor market outcomes — SLE in years and Prob at age 65

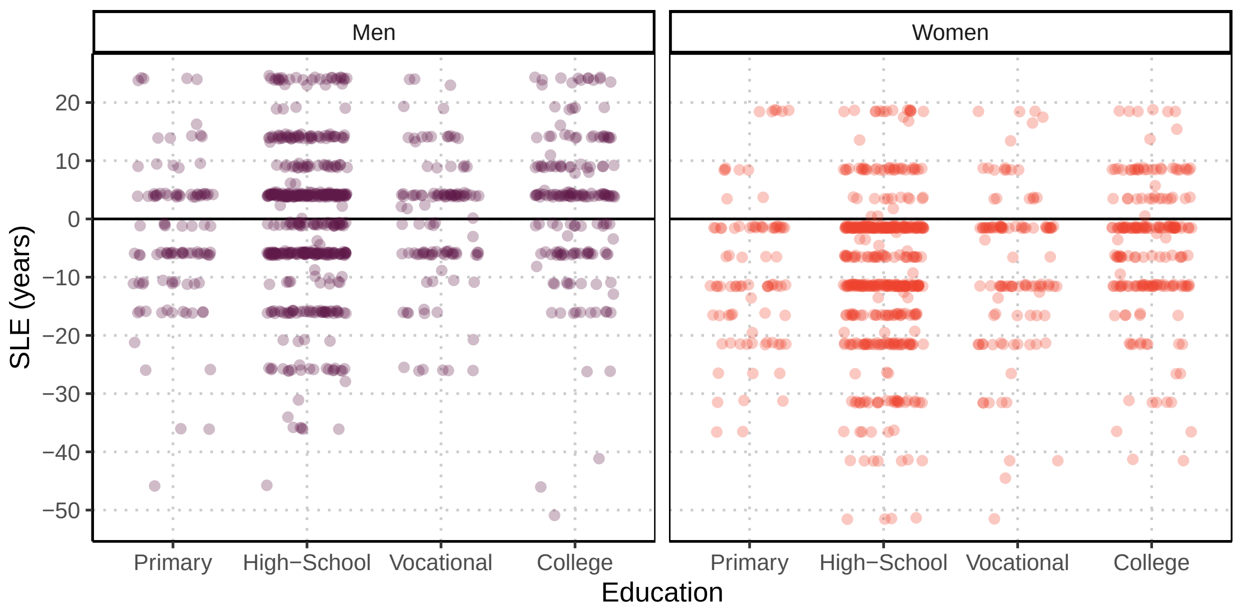
Outcome:	# Pension Contrib		Pension Wealth	
Variable	SLE	P65	SLE	P65
	(1)	(2)	(3)	(4)
Estimate	4.730	3.862	255.078	350.157
(s.e.)	(1.278)	(1.981)	(111.470)	(164.274)
Num.Obs.	1750	1890	1750	1890

Notes: The table presents the results of the OLS regression of equation 1 using the SLE variable (columns 1 and 3) or the P^{65} variable (columns 2 and 4) as the regressor. To make the two sets of results more comparable, I multiply the P^{65} coefficient by 30 to represent the 75th-25th gap. Therefore, it can be read as the coefficient associated with increasing P^{65} by 30 percentage points. In the first two columns the outcome variable is the total number of pension contributions in December 2019, and in the last two columns pension wealth in December 2019 in thousands of Chilean pesos. All regressions include baseline controls (demographic cell, region, parental education, and time fixed effects). heteroskedastic-robust standard errors are in parenthesis.

Figure A.2: SLE by age and education



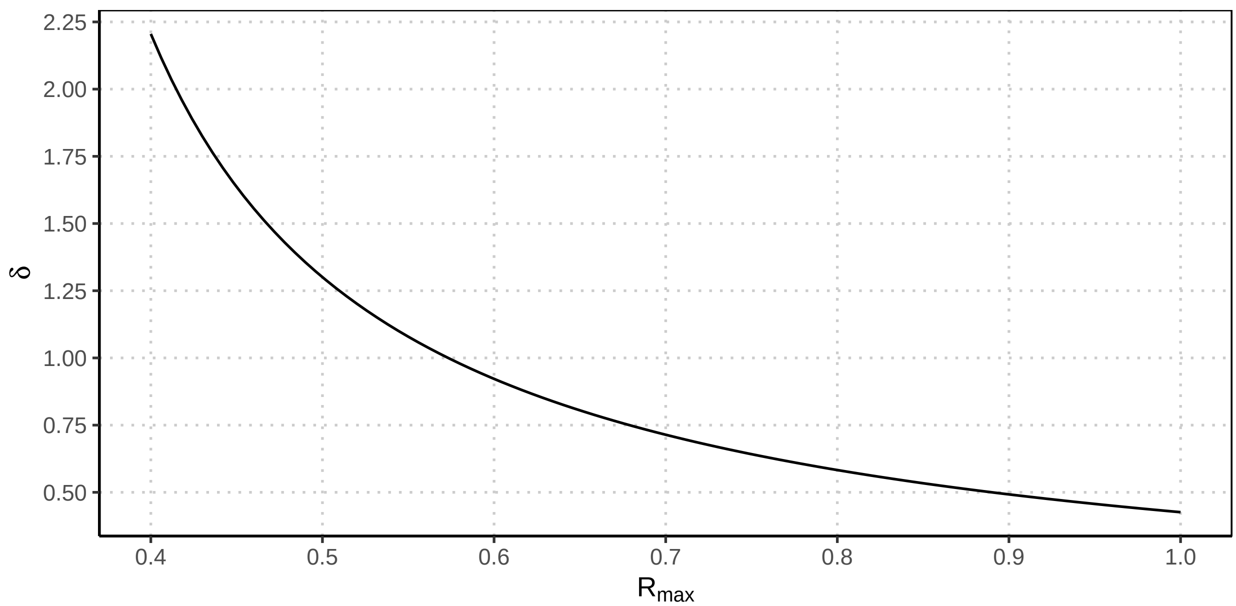
(a) SLE by age



(b) SLE by education

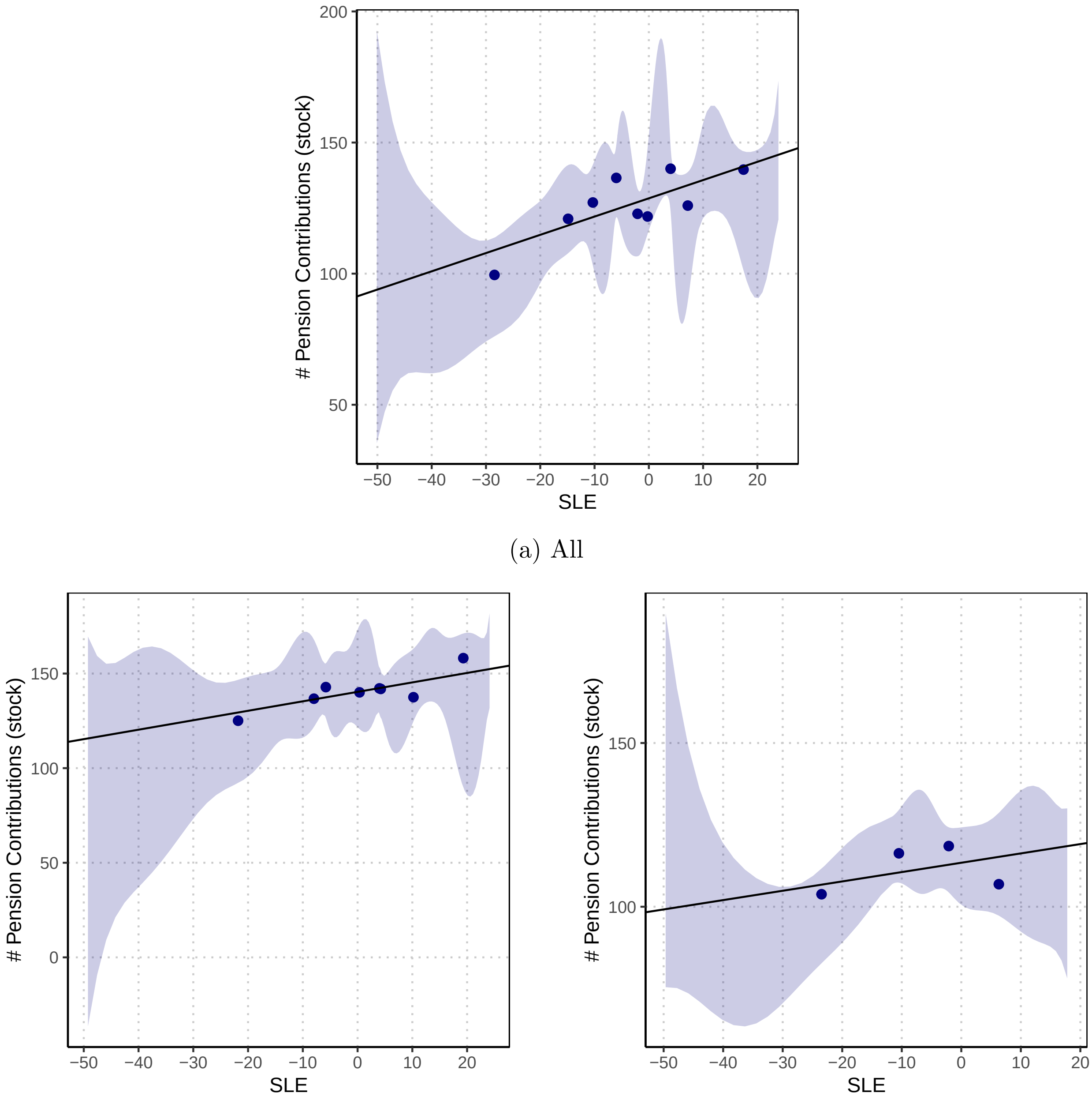
Notes: The figure plots the subjective life expectation by age (in the top plot) and by educational level (in the bottom plot), separately for men and women. Each dot displays each respondent's (raw) answer to the question "Up to what age do you believe you will live?". The solid black on panel (a) lines show the average expected mortality from the life table. To improve on the visualization, dots are spread (jittered) over the x-axis, but not over the y-axis. Additionally, they are plotted with transparency, therefore darker regions represent more mass of points.

Figure A.3: Conditional Independence Assumption — Oster (2019)



Notes: The figure shows the results for the procedure proposed by Oster (2019) to assess how strong the unobservables must be (as a ratio of the explanatory power of the observables) to drive the results of the OLS estimation of equation 1. The curve plots these ratio (δ in the y-axis) as a function of the maximum R-squared that can be obtained. These results used as outcome variable the stock of pension contributions at December of 2019.

Figure A.4: Robustness — non-linear

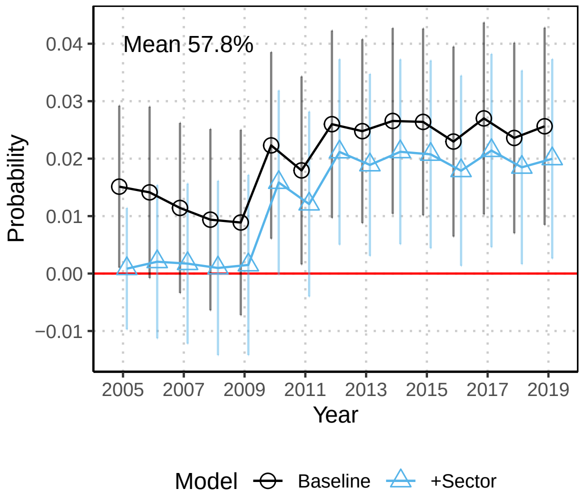


(a) All

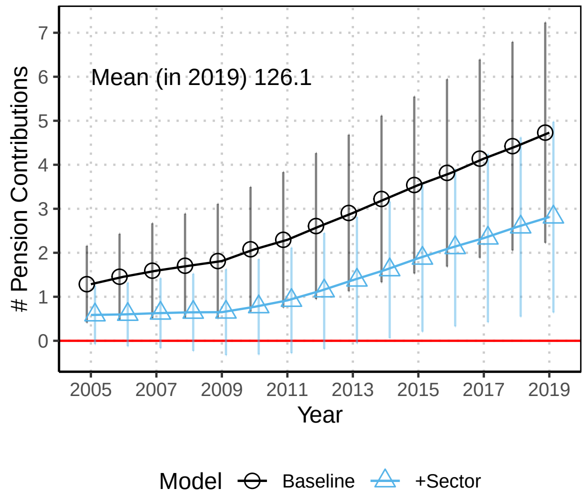
(b) Men

(c) Women

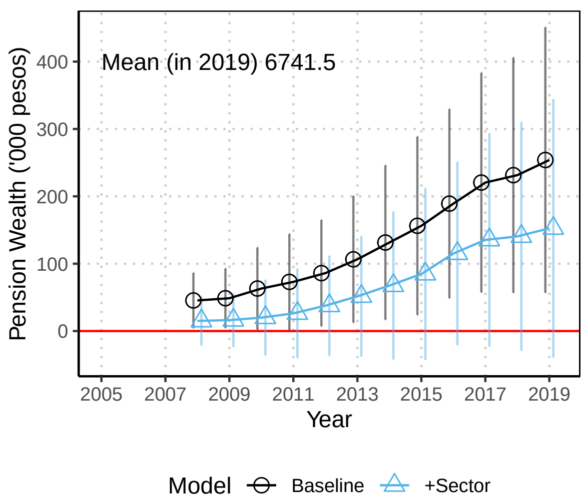
Notes: The figures show the results of applying a non-parametric binscatter regression proposed by Cattaneo et al. (2024). The dots are the resulting bins when using their algorithm to compute the optimal number and position of bins. The shaded area shows the 95% confidence interval for the relation between SLE (in the x-axis) and the total number of pension contributions in December 2019 (in the y-axis). I included the baseline controls (demographic cell, region, and parental education). The solid black curve shows the results from the linear specification from equation 1.



(a) Probability of making pension contributions



(b) # Pension contributions (stock)



(c) Pension wealth ('000 pesos)

Figure A.5: SLE and future pension contributions with additional controls

Notes: The figure plots the results from the OLS estimation of equation 1. Panel (a) for the binary outcome on the probability of making monthly pension contributions, panel (b) for the number of total pension contributions (stock), and panel (c) pension wealth measured in thousands of Chilean pesos. The regressions are run by pooling all monthly observations but separately for each calendar year. The circles/triangles are the estimated coefficient for β, and the solid vertical lines are the 95% confidence intervals. The text in each graph displays the mean for the outcome variable, for the entire period (panel a), and for December 2019 (panels b and c). The color and shape indicate the controls included. The baseline controls (black circles) include demographic cells, region, parental education, and time fixed effects. The blue triangles include, additionally, fixed effects for the initial status in the labor market in 2004 (out of the labor force, unemployment, formal, informal, self-employment, public employment, or any other employment). Standard errors are clustered at the individual level.