Subjective Life Expectancy and Public Pension Participation: Evidence from the Urban and Rural Residents Pension in China

August 2023

Abstract

To provide financial assistance to retirees in the era of population aging, enrolling farmers and informal sector workers in public pensions is essential yet challenging. This paper explores the role of subjective life expectancy in the decision to participate and select contribution level in the Urban and Rural Residents Pension in China. Combining individual fixed effects with instrumental variables for subjective life expectancy that are selected by Lasso regressions from a pool of short-term health shocks, the two-stage least squares estimates show that pension participation and contribution level do not respond to changes in subjective life expectancy. Sensitivity analysis reveals that the results are robust to relaxations of the exclusion restriction.

1 Introduction

As population aging in China continues to deepen, enhancing old-age support becomes an increasingly urgent call. The financial challenge of elderly life planning lies in the the sudden decline in income upon retirement and the uncertainty about one's longevity. A reliable public pension system serves as a social security net for elderly when private financial market is underdeveloped and conventional family-based old-age support is undermined by changing family structure. The Urban and Rural Residents Pension, hereafter URRP, is a nation-wide public pension scheme in China that targets the rural population, informal sector workers and the unemployed. It is considered an ambitious move by the Chinese government to expand social security coverage in preparation for population aging (Shen & Williamson, 2010). The outline of the 14th Five-Year Plan (2021-2025) for National Economic and Social Development specifies the goal to increase public pension participation rate to 95% by 2025.

The standard life cycle model highlights the role of individual life expectancy in old-age financial planning. It predicts that people with shorter life expectancy will discount future more heavily and thus have lower saving motives. Even without heterogeneous time preference, the URRP participation rules may makes it less attractive for people with shorter life expectancy simply because they get less total expected payment. If people select into pension programs based on individual life expectancy, this means people with shorter life expectancy who are characterized with worse socio-economic status are less likely to benefit from this income-smoothing tool. This could exacerbate existing inequality in mortality, access to health care and information. As population life expectancy continues to grow while inequality in adult mortality remains large, it is important to understand the role of individual life expectancy in pension participation.

This paper investigates whether one's subject belief about their life expectancy affects participation in the URRP. I exploit the China Health and

Retirement Longitudinal Study (CHARLS) survey data that collects subjective life expectancy information by asking for a 5-tier Likert scale response to the question: "How likely do you think you can .live to 75 years old?" at each survey round in two/three years intervals. It also collects information on individuals' annual contributions to the URRP, which can vary from year to year under the flexible terms of the pension. This allowed the introduction of individual fixed effects that control for time-invariant factors that confound the relationship between subjective life expectancy and pension contribution, as well as measurement error stemmed from differential framing of the Likert scale highlighted by Delavande et al. (2011). Moreover, I use individual-specific health shocks as instrumental variables for subjective life expectancy. The instrumented variables are systematically selected from a wide range of health indicators measured during the surveys using the Lasso regression to prevent weak instrument issues.

Empirical evidence shows that subjective life expectancy is highly predictive of actual survival (Hurd & McGarry, 2002). Moreover, subjective life expectancy reflects meaningful private information about one's health beyond the prediction of a life table mortality rate based on demographic characteristics and its role in shaping individual forward-looking economic decisions has been well-studied (Gan, Gong, Hurd, & Mcfadden, 2015; Hurd & McGarry, 1995, 2002). Subjective life expectancy affects consumption and saving decisions (De Nardi, French, & Jones, 2009; Gan, Gong, Hurd, & Mcfadden, 2015; Salm, 2010) as well as retirement timing (Hurd, Smith, & Zissimopoulos, 2004; Solinge & Henkens, 2010). Several studies (Beauchamp & Wagner, 2020; Delavande, Perry, & Michigan, 2006; Hurd, Smith, & Zissimopoulos, 2004) have found that people with shorter subjective longevity choose smaller annuity by claiming benefits earlier in the US Old-Age Social Security program. Adverse selection based on individual mortality is also found in private annuity market (Finkelstein & Poterba, 2002, 2004). More broadly, exposure to specific mortality risk factors can change people's forward-looking decisions. For example, recognizing oneself as a carrier of a life-shortening disease reduces human capital investments (Oster, Shoulson, & Dorsey, 2013). Pessimistic longevity prospects due to potential HIV exposure contributes to more risky sexual behaviors (Kerwin, 2022).

Despite the hypothetical relevance of subjective life expectancy, empirical evidence on the relationship between subjective life expectancy and public pension participation is scarce. This paper is intended to fill this gap. One major empirical challenge of studying the effect of subjective expectations with observational data in reduced forms is separating it from unobservable characteristics that are also affecting the outcome of interest (Delavande, Giné, & McKenzie, 2011; Klaauw, 2012). Earlier studies of subjective life expectancy relied on simple ordinary least squares regression, duration model or limited dependent variable models (Hurd, Smith, & Zissimopoulos, 2004; Solinge & Henkens, 2010). Studies that are aimed at estimating causal effects often rely solely on instrumental variables (Beauchamp & Wagner, 2020; Bloom, Canning, Moore, & Song, 2006; Brinch, Fredriksen, & Vestad, 2018; Delavande, Perry, & Michigan, 2006; Khan, Rutledge, & Wu, 2014). The longitudinal household survey data allows me to enhance the instrumental variable design with individual fixed effects.The rich information about one's health allows me to lasso-select the most relevant IV. In addition, I test the sensitivity of the estimated effect to violations of the exclusion restriction using the Conley (2012) method.

I find that the Lasso-selected short-term health shocks are strong predictors of within-individual variations in subjective life expectancy. Nevertheless, for people whose subjective life expectancy is affected by these health shocks, their pension participation decisions are not found to respond to updates in their subjective life expectancy. Sensitivity analysis under relaxed exclusion restrictions shows that the estimated coefficients of subjective life expectancy tend to become negative, contradicting theoretical prediction.

In the following of this paper, I first introduce the Resident Pension in

China in section 2. Then I show key descriptive statistics of the data in section 3. Section 4 explains the identification strategy and the Lasso method in detail. Section 5 shows the regression results. Section 6 discusses possible violation of the exclusion restriction of the instrumental variable and shows sensitivity analysis results.

2 The Urban and Rural Residents Pension in China

The URRP started separately as the Urban Residents Pension and the New Rural Residents Pension with almost identical terms but were offered to urban and rural residents separately when they started at different times. After the introduction in 2009, it gradually rolled out across China over 5 years. The URRP complements the Basic Pension Program of the Firms and the Public Employee Pension, which cover the private and public sector employees, respectively. Compared to the RP, employee pensions are mandatory, designed as pay-as-you-go, require employer contribution, and offer higher benefit levels. People who are not covered by either of the employee pensions are eligible to participate in RP, including farmers and the unemployed. While employers are technically required to provide rural migrant workers with the Basic Pension Program of the Firms, participation is low. On the employer's side, law enforcement on migrant workers' rights is weak. On the migrant worker's side, willingness to participate is low because they move frequently, and the benefits are not easily transferable across employers in different administrative regions. Self-employed people can participate either in a special version of the Basic Pension Program of the Firms or the RP.

Although the detailed participation terms of the URRP vary slightly across provinces and are adjusted across time, the overall principles are coherent. Before age 60, participants make annual contributions at a chosen level among 10 options. The lowest annual contribution option is 100 RMB, which is equivalent to 1/50 of per capital rural resident income in 2009[1] and is generally considered highly affordable. The government pays an additional 30 RMB matched contribution to the participant's personal account for the lowest contribution level. The higher the personal contribution, the higher the matched government contribution. The transaction cost of participating in the program is relatively low. In rural areas, contributions are often collected by the village leaders. In the urban areas, contributions are paid in local social security offices or designated bank branches. More recently, contributions can be paid online using mobile software applications. For an eligible person, contributions from year to year could vary when this person switches from non-participation to participation, misses or skips contribution after enrolling, adjusts annual contribution level, or pays make-up contributions for the skipped years.

Upon turning 60 years old, the participants start to receive monthly pension payments, the level of which equal to a base payment plus 1/139 of the personal account accumulation, which includes personal contribution and government subsidy. The payments last until death and any remainder of the personal contribution can be inherited. Unlike the Old Age Social Security in the US, payments cannot be deferred at the participant's will. People who were already 60 years old when the RP started in the community are not allowed to make contributions but are eligible to receive monthly base payments. For people who join the program before age age 45, they must make at least 15 years of contribution. For people who joined the program after age 45, they are allowed to pay make-up contributions for up to 15 years.

[1] The annual contribution level options include 100, 200, 300,400, 500, 700, 1000, 1200, 1500 RMB per capita. Per capital rural resident income in 2009 was 5153.2 RMB according to the China Statistical Yearbook.

3 Data

The China Health and Retirement Longitudinal Survey (CHARLS) is a randomized survey targeting individuals above 45 years old and their spouses. It has been carried out every two or three years since 2011 over 150 county-level units around China, covering around 20 thousand individuals each round. The sample households are randomized using multistage sampling at the national, county, community, and household levels. A detailed description of the sampling methods is provided on the official ISSS website.[2]

The population of interest in this study is people that are eligible for making contributions to the RP, i.e. people who live in communities where the RP is offered, younger than the pension receiving age of 60, and are not enrolled in other public pensions. To identify this sub-sample in the CHARLS data I combined self-reported eligibility with several criteria to correct for mis-reporting.In the 2018 and 2015 surveys, respondents are directly asked to answer questions about their enroll-eligibility for the RP. While the RP is widely recognized given its magnitude, it is still possible some people are not fully aware of detailed eligibility information. Solely relying on self-reported eligibility could lead to false exclusion of eligible people who are unaware of their eligibility and false inclusion of ineligible people who think they are eligible. To fully exclude ineligible people, I cross-check with their reported participation status in other public pensions and make sure people who were already enrolled in other public pensions are excluded. It is harder to recover eligible people who are not aware, to the extent awareness is correlated with subjective life expectancy, selection bias would be introduced. Fortunately, all the respondents who answered “not eligible” were followed-up with a question to choose from a group of legitimate reasons and people who did not respond to this follow-up question or reported “I don’t know” to the RP eligibility question all happened to be above 59 years old and therefore are

[2]http://charls.pku.edu.cn/index/en.html

not part of the population of interest, eliminating the selection bias concern. Among the enroll-eligible sample, I further drop observations of enrollees that are receiving payments or above the pension-receiving age to identify contribution-eligible people. The 2013 and 2011 surveys did not ask for individual pension enroll-eligibility information, but information from the 2011 community survey can show whether a rural community has carried out the New Rural Residents' Pension by then. Therefore for 2013 and 2011, I can only identify enroll-eligible people who lived in rural communities that have the New Rural Residents' Pension since 2011. The selected eligible people are potential contributors to the URRP if they are not enrolled in other public pensions and are not above 59 years old or receiving payments from the RP. The final sample consists of 18,308 observations of eligible contributors of the RP, among which 13,915 are from individuals observed more than once.

Figure 1 shows the annual contribution levels of contribution-eligible individuals in the sample. The number of eligible individuals increased from 2011 to 2015 due to the expansion of the URRP and started to decrease in 2018 as more people in the sample enter their 60s. The participation rate increased over the years due to the government's continuous effort to expand enrollment. The majority of participants choose the lowest contribution level (100 yuan), while the proportion of people who choose higher contribution levels increases over time. For instance, among the 5694 individuals that are eligible (and observed) in both 2015 and 2018, 39.5% increased contribution level, 7.6% reduced contribution and the others remained the same level.

[Figure 1 here]

To elicit subjective life expectancy, respondents are asked about their perceived likelihood to live to a certain age . The particular age is 10-15 years above the respondent's age. The sample of interest is below 60 years old and therefore the relevant question is on the probability to live to 75 years old. The exact question is "Suppose there are 5 steps, where the lowest step represents the smallest chance and the highest step represents the highest chance, on what step do you think is your chance in reaching the age of 75?" As shown in Figure 2, answers to this question are dispersed and with the middle option being the most popular answer. Compared to non-eligible individuals who are already enrolled in other more generous public pension programs, less eligible individuals report "almost certain" to live to 75 years old and more believe it is "almost impossible." The probability captured by the survey question is essentially a conditional probability, i.e. $Prob(survival\,time \geq 75|survival\,time \geq age\,at\,survey)$. Abstracting away from individual heterogeneity, if we are willing to assume that subjective mortality hazard is a constant λ over time and that the subjective survival rate is of the form $e^{-\lambda \cdot age}$, then this conditional probability is equal to $e^{-\lambda \cdot (75 - age\,at\,survey)}$, an increasing function of age at the time of the survey. On top of this, if we also consider individual heterogeneity and changes in health across time, the subjective distribution of mortality will be different for each individual and will change in each round with new information about one's health, making the relationship between age and subjective life expectancy ambiguous.

[Figure 2 here]

Table 1 compares observations that made zero contribution to observations that made a positive contribution to the RP in terms of the variables that will be used in the model. Participants are slightly older, more likely to have health insurance, spend less on outpatient health care. They do not statistically differ in subjective life expectancy, gender, number of cigarettes smoked per day, inpatient health care expenditure, or household income.

[Table 1 here]

4 Empirical framework

4.1 Identification strategy

The goal of this study is to find out whether subjective life expectancy affects pension contribution decisions. The major empirical challenge is that subjective life expectancy can be affected by various unobserved factors that could also be linked to pension contribution. To address this issue, I combine the panel data method with instrumental variables. The panel structure of the data allows me to control for individual fixed effects, which accounts for time-invariant individual characteristics that both determine pension contribution and correlated to life expectancy, such as long-term health status, pessimism/optimism, socioeconomic status, etc. Individual fixed effects also help to iron out individual differences in their understanding of the association of probabilities and the Likert scale (Delavande, Giné, & McKenzie, 2011), which could be correlated to unobserved heterogeneities. Moreover, I use short-term health shocks as instrumental variables for within-individual variation in subjective life expectancy. The key identification assumption is that within individuals, short-term health shocks are a plausible instrument variable for subjective life expectancy in the sense that it effectively shifts subjective life expectancy and is not correlated with unobserved time-variant factors that affect pension contribution, conditional on a group of control variables. The model estimated is:

$$Y_{it} = \beta_0 + \beta_1 LE_{it} + \beta_2 X_{it} + \delta_i + \epsilon_{it} \tag{1}$$

$$LE_{it} = \alpha_0 + \alpha_1 H_{it} + \alpha_2 X_{it} + \delta_i + e_{it} \tag{2}$$

where i indexes individual and t indexes survey year. Y_{it} is the outcome variable. I explore the effect on different margins of pension contribution and heterogeneity across subgroups. On the extensive margin, it is a dummy variable for contributing or not to the RP. On the intensive margin, it is the amount of annual contribution. One of the phenomena in URRP participation discussed by Williamson et al. (2017) is that an overwhelming majority of people only choose the lowest contribution level, which appears irrational given how low the threshold is and the obvious benefit of choosing a higher contribution level. To explore the role of subjective life expectancy in this irrationality, I define a dummy variable indicating whether a contributor is contributing higher than the minimum level of 100 yuan/year as another outcome variable. LE_{it} denotes subjective life expectancy measured in the 5-tier Likert scale. H_{it} is selected short-term health shocks discussed in the following sub-section.

X_{it} is a vector of control variables, including household income, number of cigarettes smoked per day, having health insurance, and age. Household income represents the amount of financial resources to devote to future planning. It can also be an important factor for short-term health and life expectancy as it determines the quantity and quality of food available and access to health care. I also control for the number of cigarettes smoked per day. It is well-known that smoking is related to a higher risk of lung cancer. Smoking can also be correlated to pension contribution in two other ways: first, it is a sign of impulsiveness (Khwaja, Silverman, & Sloan, 2007),which is related to lower incentive to invest in pension; second, smoking can crowd out pension contribution if they are both in the mental account of "extra money." A dummy variable for having health insurance is also included as the New Rural Cooperative Health Insurance is bundled with the URRP during the promotion in some rural areas. To the extent that having health insurance updates people's belief for longevity, omitting it would confound the regression. I also include a polynomial of age as age is likely to have a non-linear relationship to the subjective life expectancy variable because of the way the latter is defined, as discussed in the last section. It is also relevant to pension contribution through time preference-related motivations. The trade-off between consuming today and making contributions will change non-linearly along with age as one approaches the time of future reward. All financial variables, including household income, health care cost, and pension contribution are transformed using the inverse hyperbolic sine transformation (Bellemare & Wichman, 2020) to reduce the impact of outliers.

4.2 Selecting instrumental variables using Lasso

To capture key aspects in short-term health shocks that are related to one's subjective life expectancy sways, I exploit the rich health information collected in the CHARLS data. In each round of CHARLS survey, respondents answered questions about their disabilities, diagnosed chronicle diseases, recent episodes of heart attacks, stroke, accidents and different types of pains, etc. Such information provides 39 potential instrumental variables, most of which are dummy variables and categorical variables. One of the key criteria for valid instrument variables is joint significance in the regression of subjective life expectancy, conditional on individual fixed effects and control variables. Since many of the indicators may not vary much across time and/or fail to influence subjective life expectancy, including all of them would result in many weak instrument issues discussed in Stock & Yogo (2005). I use the rigorous Lasso regression proposed by Belloni et al. (2012) to identify the most influential ones as the instrumental variables for subjective life expectancy. By adding a penalty term to the objective function, Lasso regression shrinks coefficients of variables with low relevance to zero, keeping only strong predictors. The feature of the rigorous Lasso compared to other Lasso method is that it selects penalty parameters using a data-driven method, instead of cross-validation. Moreover, it allows non-Gaussian and heteroskedastic errors, which are exactly needed in this context as the subjective life expectancy is a categorical variable, and errors are likely to be clustered at the individual level in the panel data. To force the inclusion of individual fixed effects and control variables in the LASSO regression, I use the partialling-out method described in Chernfozhukov et al. (Chernozhukov, Hansen, & Spindler, 2019). This method regresses subjective life expectancy and all instrument variable candidates on the individual fixed effects and the control variables, then uses the residuals from these regressions in places of the corresponding variables in the Lasso regression. In the next step, I use the Lasso-selected instrumental variables to conduct a standard two-stage least squares regression to estimate the effect of subjective life expectancy on pension contributions.

5 Estimation results

This section first summarizes the Lasso instrumental variables selection results based on different model specifications. Then move on to discusses the two-stage least squares regression results using the selected set of instrumental variables.

5.1 Lasso IV selection

Table 2 shows the Lasso regression results based on different versions of the subjective life expectancy and health shocks variables. The first column used the original variables. The second column used variables that have partialled out individual fixed effects, so they represent variations within individuals. The third column further included control variables X_{it}. Without conditioning on the individual fixed effects and the control variables, 21 health shock variables (column 1) out of the 39 candidates are selected by the Lasso method to be highly explantory for longevity beliefs. Among them, having chronic lung disease, having had a stroke and having headache stand out to be the top factors. After introducing the individiual fixed effects, a substantial variation in personal health status is absorbed and 6 health shocks (column 2) still remain highly relevant to updates in subjective life expectancy, including having heart problems, having headache, having keen pain, having lower back pain and having poor far and close distance eyesight.

If we further control for the covariates, the selected predictor for subjective life expectancy reduce to one: close distance eyesight.

The last two rows of Table 2 report the joint significance test of the selected variables. The sup-score statistic, explained in Belloni et al. (2012) is similar in spirit to Anderson & Rubin (Anderson & Rubin, 1949) and Staiger & Stock (Staiger & Stock, 1997) but is more suited for the context where there are very many instruments. Based on the p-value in column 3, the null hypothesis that the coefficients of all the candidate instrumental variables are jointly zero is rejected at the 1% confidence level for all specifications.

In the subsequent 2SLS regression, I use the close distance eyesight selected by the third Lasso regression as the instrumental variable for life expectancy. According to the American Optometric Association, close distance eyesight problem is one of the most common age-related health issues for people between mid-40s to 60 (*Adult Vision*, n.d.). Presbyopia, the most prevalent among close distance eyesight issues for this age group, is caused by reduced flexibility of lens in the eyes and develops progressively as one gets older. A study found the prevalence rate of 67.3% among people above 40 years old in rural northern China (Lu et al., 2011). Similar numbers are found in other developing countries such as Nigeria, Ghana, and India (Patel & West, 2007). In the CHARLS survey, the respondents are asked to report how well do they see things close-up, with glasses if need, such as reading a newspaper. The respondents choose from a 5-scale measurement: 1 - excellent; 2 - very good; 3 – good; 4 – fair; 5 – poor. Among the contribution-eligible sample, 46.7% report having fair eyesight, and 20.9 % report having poor eyesight. Access to quality optical care is poor in rural China, nearly half of the affected people report having no eyesight correction in the northern China study (Lu et al., 2011). Since close distance eyesight problems like presbyopia are developed naturally as one gets older, it can be interpreted by individuals as a sign that the body functionality is going down and therefore leads to a pessimistic prospect of one's quality of future life and longevity. Also, since close distance eyesight problems are often not caused by outside factors, it is less likely to be correlated to unobserved confounders such as economic situations, unhealthy habits and negative life shocks. In addition, there is no obvious reason why having eyesight issues would make people more optimistic about life expectancy. To the extent that close distance eyesight meets the relevance, exclusion, and monotonicity criterion, the subsequent 2SLS estimator is a Local Average Treatment Effect that applies to people whose subjective life expectancy is affected by close distance eyesight problems.

[Table 2 here]

5.2 Two-stage least squares regression results

Table 3 reports the first and second stages of the 2SLS regression. The top panel shows that close distance eyesight problems and health care expenditure are negatively related to subjective life expectancy. This is in line with the fact that eyesight problems are not fully treated to ease away longevity concerns among the affected people. The last row reports Kleibergen-Paap statistics (Kleibergen & Paap, 2006) for the significance of the instrument variable, which is robust to heteroskedasticity and clustering. In the case where there is only one instrumental variable for the one endogenous variable, it is equivalent to the Montiel Olea & Pflueger (Olea & Pflueger, 2013) efficient F-statistic and can be compared to the Stock & Yogo (2005) critical values (Andrews, Stock, & Sun, 2019). The F-statistic of 22.23 for columns 1 and 2 indicates that close distance eyesight is a strong instrumental variable for subjective life expectancy for the eligible people as a whole. When it is applied to URRP participants only, the F-statistic is only 6.64 (column 3), implying that the 2SLS estimator may not be reliable for the third model.

The bottom panel of Table 3 shows the results of the second stage regression. The estimated coefficients of subjective life expectancy in column

1 and column 2 represents its effect on the extensive and intensive margin of pension participation, respectively. The standard errors clustered at the household level since the survey sampling was randomized at the household level. The estimate in column 1 means that one level increase in the likelihood of living to age 75 results in 3.4% increase in the probability of participating in the URRP, although the estimate is not statistically distinguishable from zero. The estimate in column 2 0.258 translates to the elasticity of 0.81 at the sample mean, which is not statistically significant with the standard error of 1.726.[3] The coefficient in column 3 that measures the probability of a pension participant make a higher level of contribution than the default due to a increase in subjective life expectancy is also no statistically different than zero. The overall null effect means that the target population of the URRP do not seem to make pension participation decisions based on their perceived longevity.

Moreover, people with health insurance is 19.3% more likely to participation in URRP. The coefficient of 1.06 in column 2 means that people with health insurance contribute 183% more than those who don't, implying that administrative efforts to promote the URRP in the means of bundling it with the New Rural Cooperative Health Insurance plays an important role.

[Table 3 here]

6 Sensitivity analysis

Although the nature of close distance eyesight problems decides that it cannot be caused by life changes that may also affect pension participation, I cannot fully rule out the possibility that it can affect pension participation in other potential channels. For example, people with poor eyesight may have compromised ability to learn information about the pension. A recent study (Bai,

[3] Calculation using the formula in Bellemare and Wichman (2020) with the standard error computed using the delta method.

Chi, Liu, Tang, & Xu, 2021) shows that the lack of information about the detailed terms and potential benefits of the pension is an important factor behind the lack of participation. If there exist such alternative mechanisms, the 2SLS would be biased. To address this issue, I explore how would alternative exclusion assumptions change the point estimate of the 2SLS estimator using the method elaborated by Conley et al. [@-conleyPlausiblyExogenous2012]. When the instrumental variable only affects the outcome variable through the endogenous variable, the coefficient of the instrumental variable, denoted as γ, in the regression of the outcome variable on both the endogenous variable and the instrumental variable should be zero. To relax this exclusion assumption, I assume that γ deviates from zero. That is, close distance eyesight problems can affect pension participation from other channels. It is arguably plausible to further assume that this alternative mechanism is negative, that in this other channel close distance eyesight problems cannot encourage pension participation. Depending on the magnitude of the negative γ, I calculate the upper and lower bounds for the true causal effect. As reported in Table 4, when γ deviates from 0 to -1, the estimated interval of the causal effect becomes more and more negative. Contradicting the prediction of a standard life cycle model.

[Table 4 here]

7 Conclusion

The study of the relevance of subjective beliefs in forward-looking economic decisions has gained increasing interest (Attanasio, 2009). Beyond the challenge to correctly measure subjective beliefs, learning about the determinants of subjective beliefs and identifying the causal effect of subjective beliefs on various outcomes are relatively fresh areas.

In this study, I examine the effect of subjective life expectancy on voluntary pension participation and contribution decisions in China among informal sector workers. In the search for a strong instrumental variable, I explore how subjective beliefs are shaped. In the case of subjective life expectancy, the Lasso regressions show that long-term individual-specific characteristics explain a large portion of the variation in the subjective life expectancy. In the meantime, people indeed respond to fluctuations in health. The machine learning method allows us to see that the strongest shifters can come from unexpected areas, such as eyesight problems. But if we consider the age-specific characteristics of the population of interest and the context where access to good health care is limited, it is not hard to understand why health issues commonly considered as not life-threatening can lead to pessimistic thinking about longevity.

On the contrary to theoretical predictions, I do not find statistically significant pension participation response to changes in subjective life expectancy. I evaluate this result in the following four aspects: First, the 2SLS estimator is a LATE that only applies to a subgroup of people whose subjected life expectancy is altered by changes in the selected instrumental variable. Second, the subjective life expectancy measurement in the CHARLS data is a conditional probability. It only reflects one side of the first moment of the subjective mortality distribution. Recent studies have started to bring under the light the second moment of subjective beliefs, that is, how certain people are about their beliefs (Hoel, Michelson, Norton, & Manyong, 2021). This calls for more elaborated measurements of the subjective life expectancy in large-scale surveys such as the CHARLS and the Health and Retirement Survey in the US. Third, social security net that is designed to bundle different programs can act as a behavioral nudge to encourage participation in programs that requires more forward-looking decision making function. Fourth, for elderly people and people approaching their elderly age, life expectancy and health may not be merely relevant to individual decisions. Since most elderly people in developing countries still rely on family for old-age care, inter-generational interactions can also be affected by longevity prospects.

In the case of pension participation, people with poor health and compromised self-care ability need to rely on children for old age care. This could be accompanied by a higher incentive to participate in pension to help lift the family financial burden. To sum, instead of confirming the theoretical prediction, the null finding leads to more detailed questions about subjective life expectancy and forward-looking economic decisions that are subject to future exploration.

References

Adult Vision: 41 to 60 Years of Age. (n.d.). https://www.aoa.org/healthy-eyes/eye-health-for-life/adult-vision-41-to-60-years-of-age?sso=y.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, *20*(1), 46–63.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in IV regression: Theory and practice. *Annual Review of Economics.*

Attanasio, O. P. (2009). Expectations and perceptions in developing countries: Their measurement and their use. *The American Economic Review*, *99*(2), 87–92.

Bai, C.-E., Chi, W., Liu, T. X., Tang, C., & Xu, J. (2021). Boosting pension enrollment and household consumption by example: A field experiment on information provision. *Journal of Development Economics*, *150*, 102622. https://doi.org/10.1016/j.jdeveco.2020.102622

Beauchamp, A., & Wagner, M. (2020). Is there adverse selection in the U.S. social security system? *Economics Letters*, *189*, 108995. https://doi.org/10.1016/j.econlet.2020.108995

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics*, *82*(1), 50–61. https://doi.org/10.1111/obes.12325

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*(6), 2369–2429.

Bloom, D. E., Canning, D., Moore, M., & Song, Y. (2006). *The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States.* NBER. Retrieved from http://www.nber.org/papers/w12688

Brinch, C. N., Fredriksen, D., & Vestad, O. L. (2018). Life Expectancy and Claiming Behavior in a Flexible Pension System. *Scandinavian Journal of Economics*, *120*(4), 979–1010. https://doi.org/10.1111/sjoe.12271

Chernozhukov, V., Hansen, C., & Spindler, M. (2019). *High-dimensional metrics in R*. http://r-forge.r-project.org.

De Nardi, M., French, E., & Jones, J. B. (2009). Life expectancy and old age savings. *American Economic Review*, *99*, 110–115. https://doi.org/10.1257/aer.99.2.110

Delavande, A., Giné, X., & McKenzie, D. (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, *94*(2), 151–163. https://doi.org/10.1016/j.jdeveco.2010.01.008

Delavande, A., Perry, M., & Michigan, R. J. W. (2006). *Probabilistic Thinking and Social Security Claiming*. Retrieved from http://www.mrrc.isr.umich.edu/

Finkelstein, A., & Poterba, J. (2002). Selection effects in the United Kingdom individual annuities market. *Economic Journal*, *112*(476), 28–50. https://doi.org/10.1111/1468-0297.0j672

Finkelstein, A., & Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the U.K. Annuity market. *Journal of Political Economy*, *112*(1), 183–208. https://doi.org/10.1086/379936

Gan, L., Gong, G., Hurd, M., & Mcfadden, D. (2015). Subjective mortality risk and bequests. *Journal of Econometrics*, *188*, 514–525. https://doi.org/10.1016/j.jeconom.2015.03.015

Hoel, J., Michelson, H., Norton, B., & Manyong, V. (2021). *Misattribution and uncertainty about beliefs prevent learning*.

Hurd, M. D., & McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *Journal of Human Resources*, *30*(SpecialIssue), S268–S292. https://doi.org/10.2307/146285

Hurd, M. D., & McGarry, K. (2002). The predictive validity subjective probabilities of survival. *Economic Journal*, *112*(482), 966–985. https://doi.org/10.1111/1468-0297.00065

Hurd, M. D., Smith, J. P., & Zissimopoulos, J. M. (2004). The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics*, *19*(6), 761–775. https://doi.org/10.1002/jae.752

Kerwin, J. (2022). *Scared Straight or Scared to Death? Fatalism in Response to Disease Risks*. https://doi.org/10.2139/ssrn.2797493

Khan, M., Rutledge, M. S., & Wu, A. Y. (2014). How Do Subjective Longevity Expectations Influence Retirement Plans? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2376923

Khwaja, A., Silverman, D., & Sloan, F. (2007). Time preference, time discounting, and smoking decisions. *Journal of Health Economics*, *26*(5), 927–949. https://doi.org/10.1016/j.jhealeco.2007.02.004

Klaauw, W. van der. (2012). On the use of expectations data in estimating structural dynamic choice models. *Journal of Labor Economics*, *30*(3), 521–554. https://doi.org/10.1086/664929

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126.

Lu, Q., He, W., Murthy, G. V. S., He, X., Congdon, N., Zhang, L., … Yang, J. (2011). Presbyopia and Near-Vision Impairment in Rural Northern China. *Investigative Opthalmology & Visual Science*, *52*(5), 2300. https://doi.org/10.1167/iovs.10-6569

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Oster, E., Shoulson, I., & Dorsey, E. R. (2013). Limited Life Expectancy, Human Capital and Health Investments †. *American Economic Review*, *103*(5), 1977–2002. https://doi.org/10.1257/aer.103.5.1977

Patel, I., & West, S. K. (2007). Presbyopia: Prevalence, impact, and interventions. *Community Eye Health*, *20*(63), 40–41.

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. https://doi.org/10.1111/j.1540-5982.2010.01605.x

Shen, C., & Williamson, J. B. (2010). China's new rural pension scheme: can it be improved? *International Journal of Sociology and Social Policy*, *30*(5-6), 239–250. https://doi.org/10.1108/01443331011054217

Solinge, H. van, & Henkens, K. (2010). Living longer, working longer? The impact of subjective life expectancy on retirement intentions and behaviour. *The European Journal of Public Health*, *20*(1), 47–51. https://doi.org/10.1093/eurpub/ckp118

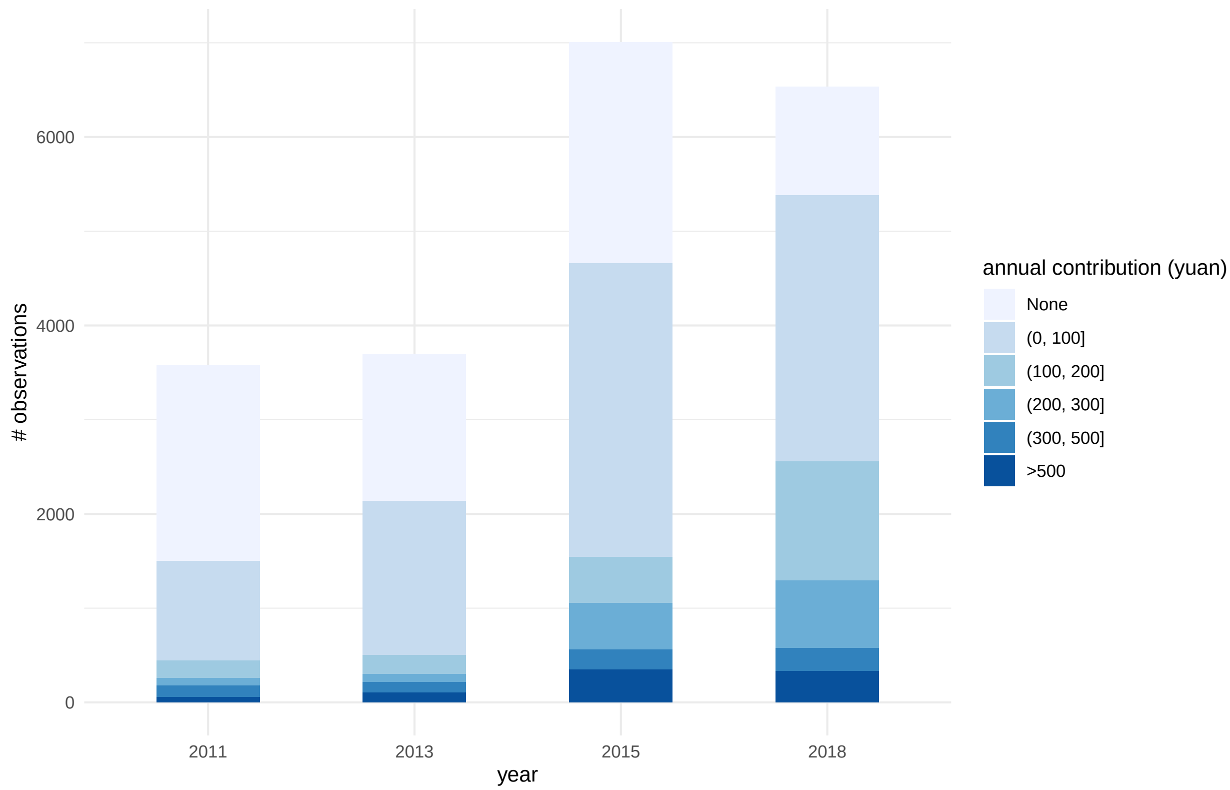
Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica: Journal of the Econometric Society*, 557–586.

Stock, J., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. K. Andrews (Ed.), *Identification and inference for econometric models* (pp. 80–108). New York: Cambridge University Press; Cambridge University Press. Retrieved from http://www.economics.harvard.edu/faculty/stock/files/TestingWeakInstr_Stock%5C%2BYogo.pdf

Williamson, J. B., Fang, L., & Calvo, E. (2017). Rural pension reform in China: A critical analysis. *Journal of Aging Studies*, *41*(April 2017), 67–74. https://doi.org/10.1016/j.jaging.2017.04.003

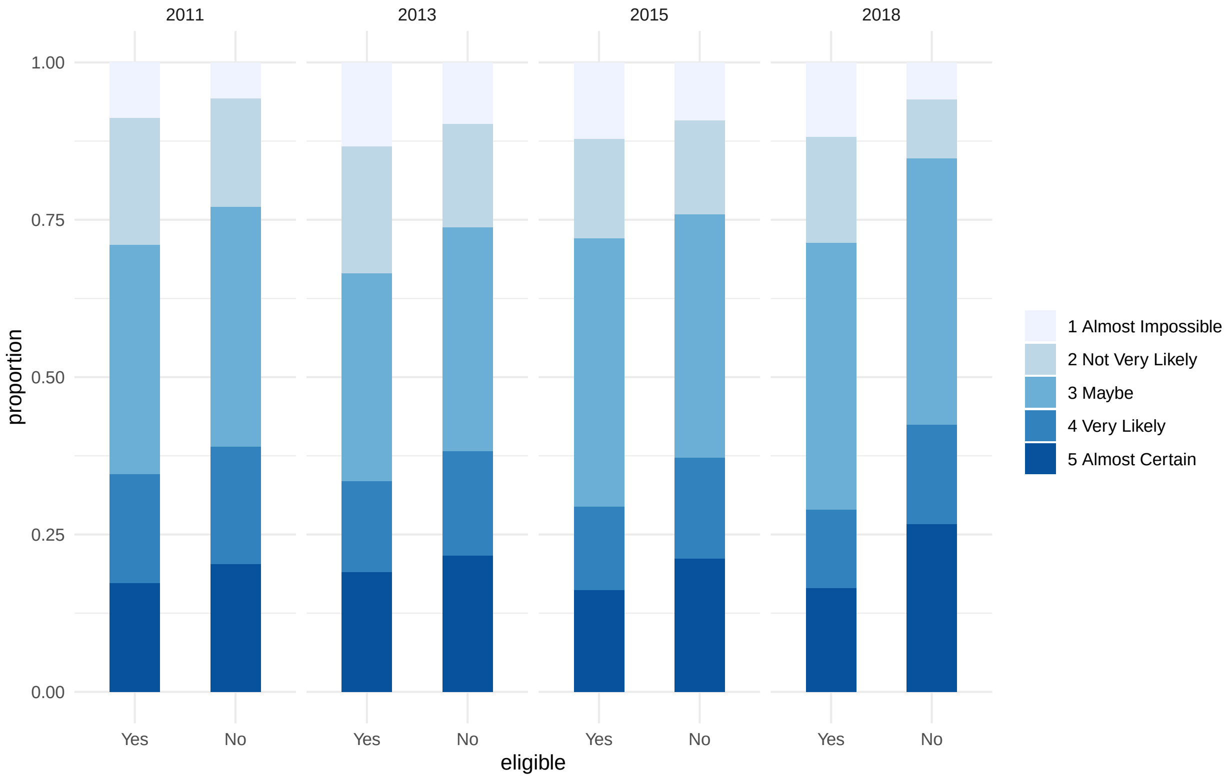
Tables and graphs

Figure 1. Eligible individuals and annual contribution



Notes: Number of individuals eligible for RP contribution by CHARLS survey round, including people younger than 60 years old who are not enrolled in other public pensions. The color gets deeper by annual contribution level. The top color blocks represent individual that did not contribute.

Figure 2. Subjective life expectancy by RP contribution eligibility and year



Notes: Proportions of people under 60 years old by their answer to the question "How likely do you think you can live to 75 years old?", divided by RP contribution eligibility and survey year.

Table 1: Descriptive statistics by RP contribution participation

	No (N=5025)	Yes (N=13283)	p value
Subjective probability of living to age 75			0.085
- 1 Almost Impossible	431 (11.1%)	1294 (12.0%)	
- 2 Not Very Likely	649 (16.7%)	1909 (17.7%)	
- 3 Maybe	1555 (40.1%)	4323 (40.2%)	
- 4 Very Likely	574 (14.8%)	1446 (13.4%)	
- 5 Almost Certain	668 (17.2%)	1785 (16.6%)	
Male			0.074
- Mean(SD)	0.471 (0.499)	0.456 (0.498)	
- Count	5023	13276	
- Range	0.000 - 1.000	0.000 - 1.000	
Age			< 0.001
- Mean(SD)	51.688 (5.103)	52.123 (4.747)	
- Count	5025	13283	
- Range	19.000 - 60.000	22.000 - 60.000	
Have health insurance (0/1)			< 0.001
- Mean(SD)	0.901 (0.299)	0.980 (0.141)	
- Count	4958	13179	
- Range	0.000 - 1.000	0.000 - 1.000	
Number of cigarettes smoked per day			0.991
- Mean(SD)	5.745 (10.883)	5.706 (11.018)	
- Count	4569	12333	
- Range	0.000 - 100.000	0.000 - 120.000	
Outpatient health care expenditure (yuan)			< 0.001
- Mean(SD)	194.653 (1146.218)	193.352 (1390.440)	
- Count	5025	13282	
- Range	0.000 - 40800.000	0.000 - 120000.000	
Inpatient health care expenditure(yuan)			0.016
- Mean(SD)	639.119 (5157.595)	613.505 (4510.574)	
- Count	5020	13272	
- Range	0.000 - 230000.000	0.000 - 200000.000	
Household income (yuan)			0.034
- Mean(SD)	119703.661 (2235754.582)	27312.588 (1066540.249)	
- Count	4987	13237	
- Range	-249000.000 - 50010500.000	-49985500.000 - 50001500.000	

Table 2: Instrument variable selected by the Lasso regression

	(1)	(2)	(3)
Have physical disability (0/1)	-0.149		
Diagnosed with hypertension (0/1)	-0.085		
Diagnosed with dyslipidemia	-0.079		
Diagnosed with diabetes (0/1)	-0.211		
Diagnosed with chronic lung disease (0/1)	-0.298		
Diagnosed with liver disease (0/1)	-0.186		
Heart problems (0 No/1 Diagnosed/2 Had recent heart attack)	-0.145	-0.165	
Stroke (0 No/1 Had/2 Had again recently)	-0.264		
Diagnosed with kidney disease (0/1)	-0.210		
Diagnosed with stomach problem (0/1)	-0.206		
Diagnosed with arthritis or rheumatism (0/1)	-0.063		
Recent accidental fall (0/1)	-0.106		
Eyesight - far (1 Excellent - 5 Poor)	-0.136	-0.051	
Eyesight - close (1 Excellent - 5 Poor)	-0.059	-0.039	-0.056
Have headache (0/1)	-0.266	-0.082	
Have shoulder pain (0/1)	-0.025		
Having arm pain	-0.058		
Have back pain (0/1)	-0.079		
Have lower back pain (0/1)	-0.158	-0.052	
Have leg pain (0/1)	-0.060		
Have knee pain (0/1)	-0.012	-0.072	
Individual fixed effects	No	Yes	Yes
Control variables	No	No	Yes
N	8871	8871	8105
Sup score statistic for joint significance	24.0	2.0	1.6
p-value	0.000	0.000	0.000

Coefficients and test statistics from rigorous Lasso regressions (Belloni et al., 2012) of subjective life expectancy on 39 health shocks. Penalty parameters are selected based on a data-driven procedure and OLS after Lasso (i.e. post-Lasso) is applied. The regression is conducted using the hdm R package developed by Victor Chernozhukov, Christian Hansen and Martin Spindler. Only variables whose coefficients are not shrank to zero are included. The individual fixed effects and control variables are included using the partial-out method (Chernozhukov et al., 2019). The sup score measures the joint significance of the Lasso model.

Table 3: 2SLS regression of RP pension contribution

	(1) Contribute (0/1)	(2) Contribution level	(3) Contribute high (0/1)
First stage			
Eyesight - close (1 Excellent - 5 Poor)		-0.0646***	-0.0399*
		(0.016)	(0.019)
Age		0.210	0.618
		(0.395)	(0.544)
Age^2		-0.00453	-0.0119
		(0.008)	(0.011)
Age^3		0.0000301	0.0000732
		(0.000)	(0.000)
Have health insurance (0/1)		0.00276	-0.100
		(0.088)	(0.140)
Number of cigars smoked per day		-0.000555	0.00221
		(0.003)	(0.003)
Household income		-0.00241	-0.00336
		(0.003)	(0.003)
Second stage			
Subjective probability of living to age 75	0.0340	0.258	-0.0388
1 Almost impossible - 5 Almost certain	(0.094)	(0.547)	(0.183)
Age	-0.00808	0.0451	0.0984
	(0.128)	(0.760)	(0.259)
Age^2	0.00115	0.00592	-0.000480
	(0.003)	(0.015)	(0.005)
Age^3	-0.00000967	-0.0000540	-0.000000732
	(0.000)	(0.000)	(0.000)
Have health insurance (0/1)	0.193***	1.064***	-0.136**
	(0.035)	(0.210)	(0.052)
Number of cigars smoked per day	-0.0000427	-0.00171	0.00118
	(0.001)	(0.005)	(0.001)
Household income	-0.000244	0.000701	0.00321*
	(0.001)	(0.006)	(0.001)
Individual FE	Yes	Yes	Yes
N	12899	12899	9552
Kleibergen-Paap Wald rk F statistic	16.45	16.45	4.325

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Cluster-robust standard error at the household level in parentheses. The Kleibergen-Paap (2006) Wald rk F statistic reflects the significance of the instrument variable in the first stage and is robust to non-i.i.d errors. Contribution level, household income and health care expenditure are hyperbolic sine transformed (Bellemare Wichman, 2020). Dependent variable in column 3 is a dummy variable that shows whether the contribution level is above the minimum level of 100 RMB per year, and therefore only applies to contributors.

Table 4: Intervals

	Contribute		Contribution level		Contribute high	
γ range	Lower	Upper	Lower	Upper	Lower	Upper
$[-0.1, 0]$	-1.892	0.119	-1.918	0.725	-2.255	0.094
$[-0.2, -0.1]$	-3.811	-1.082	-3.729	-0.691	-4.489	-1.099
$[-0.3, -0.2]$	-5.733	-2.201	-5.604	-1.919	-6.727	-2.206
$[-0.4, -0.3]$	-7.655	-3.318	-7.501	-3.082	-8.965	-3.311
$[-0.5, -0.4]$	-9.578	-4.433	-9.409	-4.223	-11.203	-4.415
$[-0.6, -0.5]$	-11.502	-5.549	-11.321	-5.354	-13.442	-5.518
$[-0.7, -0.6]$	-13.425	-6.664	-13.237	-6.479	-15.681	-6.622
$[-0.8, -0.7]$	-15.348	-7.779	-15.155	-7.602	-17.920	-7.725
$[-0.9, -0.8]$	-17.272	-8.894	-17.074	-8.722	-20.158	-8.828
$[-1.0, -0.9]$	-19.195	-10.009	-18.994	-9.842	-22.397	-9.931