Bundling Pension and Health Insurance to Mitigate Adverse Selection: Evidence from Land-Expropriated Farmers in China

Danli Shen [*] Xiang Zhang[†] Xiaoxin Zhang[‡]

Nov 6, 2025[§]

Comments welcome
Click here for the latest version

Abstract

Starting in 2011, Zhejiang Province allowed land-expropriated farmers to voluntarily upgrade from low-premium, low-benefit pension and health insurance programs to high-premium, high-benefit employee schemes. Using administrative data from City C, we document evidence of adverse selection in both insurance programs. Leveraging an instrumental variable design, we estimate demand, average cost, and marginal cost curves to quantify the resulting welfare losses. We then develop a random-coefficient logit model showing that bundling pension and health insurance can achieve a Pareto improvement by enhancing risk pooling and mitigating adverse selection: healthier individuals favor pension insurance, while less healthy individuals prefer health insurance.

[*]Zhejiang University. Y21522011t@zju.edu.cn
[†]Zhejiang University. xiangzhang@zju.edu.cn
[‡]University of Southern California. xiaoxin@usc.edu
[§]Xiaoxin thanks Pablo Kurlat, Thomas Chaney, and Benjamin Handel for their guidance and support. We thank Marianne Andries, Daniel Gottlieb, Dong Woo Hahm, Wenda Ma, Adam Solomon, Teck Yong Tan, and Zhaoji Tang for their thoughts and comments.

1 Introduction

China's rapid economic growth has displaced tens of millions of farmers through land expropriation. To compensate for the loss of livelihoods, Zhejiang Province established pension and health insurance programs for land-expropriated farmers in 2003. Between 2011 and 2020, eligible farmers could pay an additional contribution to upgrade to the higher-premium, higher-benefit employee pension and health insurance programs. Because participation in the upgrade was voluntary, this policy is vulnerable to adverse selection.

Using administrative data from City C (2010–2020), we document adverse selection in both pension and health insurance programs. Life-table estimates show that, at retirement age–50 for women and 60 for men–life expectancy is 6.5 years longer for men and 7.8 years longer for women who upgraded, indicating that individuals with greater longevity expectations are more likely to select into the employee pension. To distinguish this from moral hazard or benefit-induced treatment effects, we compute predicted life expectancy and predicted medical expenditures based on pre-transfer medical spending and compare these measures between upgraded and non-upgraded groups.

Regression results show that each additional expected year of life increases the probability of upgrading to the employee pension by 0.5 percent. Similarly, each additional expected year of life raises the probability of upgrading to employee health insurance by 0.5 percent, consistent with longer-lived individuals joining to accumulate larger balances in their medical accounts. Moreover, higher expected medical spending also predicts upgrading: a 1,000-yuan increase in expected annual medical expenses raises the probability of switching to employee health insurance by 0.1 percent. These patterns provide clear evidence of adverse selection in both insurance programs. The results are robust to alternative specifications, including three-year averages of pre-transfer medical expenses and visits, total expected costs under employee pension or health schemes, and controls for income and liquidity constraints.

After documenting the presence of adverse selection, we quantify the associated welfare loss by estimating demand and cost curves for both pension and health insurance. Premiums for the employee pension and health insurance depend on years of work: the initial farmer pension contribution counts as six years toward the employee system, and individuals must contribute for at least 15 years to qualify. Consequently, the lump-sum payment required at retirement to upgrade declines with each additional year of work. Health insurance premium is determined by years worked after 2006, when employee health was established.

To address the endogeneity arising from farmers working longer to lower their insurance premiums at retirement, we exploit exogenous variation from land expropriation, which is determined by the government at the village level and unlikely to be correlated with individual characteristics. Specifically, years of work at the time of expropriation, years of work after 2006 at the time of expropriation, and age at expropriation serve as instruments:

all of them are strongly correlated with premiums but plausibly exogenous to individual upgrading decisions.

The cross-elasticity estimate of the pension demand curve suggests that employee health insurance is a complement to employee pension. After obtaining employee health coverage, individuals can afford treatments for serious illnesses, which increases their confidence in living longer and thereby raises the perceived value of the employee pension. In contrast, employee pension acts as a substitute for employee health insurance: once individuals secure a higher pension, they expect greater post-retirement cash flows, worry less about future medical expenses, and consequently place a lower value on purchasing employee health insurance.

The demand and cost curves indicate that the welfare loss at the current price level, relative to the efficient allocation, amounts to approximately 372,000 yuan (about \$50,000) for pension insurance and 96,000 yuan for health insurance. Given that the average pension premium is about 108,000 yuan and the health insurance premium is 33,000 yuan, these welfare losses are economically substantial.

For pensions, retirees receive approximately 23,000 yuan annually after retirement, implying that they break even after roughly five years. However, men live an additional 25.9 years and women 38.4 years after retirement, indicating that the employee pension system naturally generates a large welfare imbalance even in the absence of adverse selection. Similarly, under the employee health insurance scheme, participants receive around 3,000 yuan per year in their individual medical accounts. If a retiree lives 30 years, this alone accumulates to nearly 90,000 yuan, far exceeding the total premium paid. Considering the higher reimbursement rate of the employee plan, the fiscal burden of the employee health system is even greater.

The final part of the paper examines whether bundling can mitigate adverse selection and improve welfare. To do so, I estimate a random coefficient logit model in which the deterministic utility is quasilinear in premiums, allowing me to easily recover each individuals willingness to pay for pension and health insurance, as well as the degree of complementarity or substitutability between them. This framework enables the simulation of individual enrollment decisions and welfare outcomes under alternative pricing regimes.

When pension and health costs are negatively correlated, offering a discount for the bundled option can improve efficiency by attracting low-cost participants into both programs, thereby increasing risk pooling and producer surplus. Consumer surplus also rises, as individuals receive a price reduction for an existing choice. However, in my empirical setting, pension and health costs are positively correlated: longer-lived individuals tend to incur higher costs in both employee pension and employee health insurance programs due to the individual medical account component. Under such conditions, individuals who are costly in one program also tend to be costly in the other, so bundling the two schemes works against the principle of risk pooling and may reduce efficiency. Instead, separating the programs and offering targeted discounts or vouchers (e.g., 500 yuan) for each individual program can increase total welfare without redistributive losses, demonstrating that a Pareto improvement is attainable: overall welfare can be improved without making any group worse off.

This paper contributes to several strands of the literature.

First, we extend the information asymmetry models in insurance markets (e.g., Akerlof, 1970; Rothschild and Stiglitz, 1976; Stiglitz, 1977; Wilson, 1977; Miyazaki, 1977; Riley, 1979; Engers and Fernandez, 1987) from a single- to a dual-insurance framework that jointly considers pension and health insurance, the two most critical forms of old-age coverage. While Einav et al. (2010) develop empirical tools to detect adverse selection and quantify welfare costs in single-product markets, we generalize their framework to capture interactions across programs. In addition, we replace their linear demand specification with a logit demand model to ensure that predicted demand probabilities remain within the $[0, 1]$ interval. In doing so, we also connect to work emphasizing heterogeneity in behavioral responses and insurance demand (Spinnewijn, 2017; Wagner, 2022). Unlike Handel et al. (2015), who analyze private markets with endogenous pricing, we study a public setting where prices are regulated by the government.

Second, we study whether bundling pension and health insurance can mitigate adverse selection and improve welfare through improved risk pooling. Prior work has analyzed bundling in other contexts, for example, microfinance and health insurance in India (Banerjee et al., 2014), household-level health insurance bundling in Vietnam (Nguyen, 2022), and spousal bundling in U.S. long-term care insurance (Solomon, 2024). The closest related study is Murtaugh et al. (2001), who use actuarial models to assess bundling annuities and long-term care insurance. Their setting differs in three ways: (i) the empirical correlation between longevity and disability risk is ambiguous; (ii) the long-term care insurance market is relatively small, even if risk pooling is achieved; and (iii) their analysis abstracts from demand responses to bundling and price changes. By contrast, we analyze public pension and health insurance, the two primary pillars of old-age protection worldwide, where opposite selection patterns enable risk pooling and where demand responses under alternative bundling and pricing regimes can be explicitly evaluated. While bundling has been extensively studied in the industrial organization literature (e.g., Stigler, 1968; Adams and Yellen, 1976; Berry et al., 1995), that literature primarily focuses on consumer goods in imperfectly competitive markets, emphasizing complementarity from the demand side without risk pooling on the cost side. To our knowledge, this is the first paper to examine bundling of pension and health insurance in the context of social insurancet

Third, we contribute to the empirical literature on adverse selection (Chiappori and Salanie, 2000), which has focused primarily on private annuity and health insurance markets (e.g., Cutler and Zeckhauser, 1998; Cardon and Hendel, 2001; Finkelstein and Poterba, 2002, 2004; Dror et al., 2005; Wang et al., 2006; McCarthy and Mitchell, 2010; Finkelstein and Poterba, 2014; Yao et al., 2017; Fischer et al., 2023). In contrast, we examine adverse selection in a public insurance system, exploiting a natural experiment in which land-expropriated farmers could voluntarily upgrade to higher-tier pension and health coverage. Moreover, whereas most existing studies rely on post-enrollment outcomes, raising concerns about moral hazard and benefit-induced treatment effects, we use pre-transfer medical expenditures to capture ex ante risk, offering a cleaner test of the theoretical prediction that individuals with higher expected costs are more likely to select into more generous coverage.

The rest of the paper is organized as follows. Section 2 describes the institutional setting and data. Section 3 presents the tests for adverse selection. Section 4 estimates demand and cost curves and quantifies the associated welfare losses. Section 5 conducts counterfactual analyses. Section 6 concludes.

2 Institutional Setting and Data

Institutional Setting. In 2003, Zhejiang Province pioneered a basic living security system for land-expropriated farmers, consisting of a farmer pension and a farmer health insurance scheme. Under this system, affected farmers were required to make a lump-sum contribution to the municipal social insurance institution,[1] after which, upon reaching the statutory retirement age, participants became eligible for lifelong monthly pension benefits set slightly above the local minimum living standard. Enrollment in the farmer pension was automatically accompanied by enrollment in farmer health insurance.

In rural China, land is collectively owned by the village and allocated to individual families. When the government requires land for a project, it may be expropriated. Families affected by expropriation receive monetary compensation as well as a quota that allows them to make a lump-sum contribution and enroll in the farmer pension and farmer health insurance programs. Because pension benefits are only received after retirement, families typically nominate older members to participate in the program.

In response to concerns about the low benefit levels, Zhejiang Province introduced an upgrade option in 2011. Farmers could make an additional lump-sum payment to transfer into the higher-premium, higher-benefit employee pension and employee health insurance system.

To illustrate the relative premiums and benefits, consider the arrangements in 2020. Following land expropriation,[2] farmers could make a lump-sum payment of 20,000 yuan

[1] This contribution was financed out of the land compensation package.

[2] In 2020, the standard land compensation was 62,000 yuan per mu. One mu is roughly 1.6 times the size of a basketball court. The compensation quota was determined by dividing the total land expropriated by the average farmland per capita in the village, which varied across villages and could be either less than or greater than one mu.

(about 3,000 USD) to enroll in the farmer pension. For context, the average annual rural income in City C was about 40,000 yuan in 2020. Upon retirement, at age 50 for women and 60 for men, participants would receive an annual pension of roughly 11,000 yuan, implying an annual return of about 50%. Women would break even if they lived beyond age 52, while men would do so if they lived beyond age 62.

Upgrading to the employee pension required an additional payment of 108,000 yuan (128,000 minus the original 20,000). In return, retirees received an annual pension of about 25,000 yuan, corresponding to an annual return of roughly 20%. Break-even occurred after approximately five to six years of benefit receipt. Thus, compared to the farmer pension, the employee pension involved both higher contributions and higher benefits.

Health insurance was closely tied to pension enrollment. Until 2020, a semi-bundling rule applied: access to employee health insurance was conditional on enrollment in the employee pension, and farmers who remained in the farmer pension could not separately transfer into employee health insurance. This restriction was lifted in 2020, allowing individuals to enroll in the two programs independently.

Farmers enrolled in the farmer pension automatically obtained farmer health insurance at zero additional premium. By contrast, in 2020 the employee health insurance required a lump-sum premium of about 33,000 yuan. The difference in benefits was substantial: reimbursement rates were around 50% under farmer health insurance versus 80% under employee health insurance. Farmers made this one-time payment at retirement and enjoyed the higher reimbursement rate for life.

Data. The data come from City Cs social insurance records and cover all land-expropriated farmers who entered the system between 2010 and 2020, a total of 161,326 individuals. For each person, the records include basic demographic information (such as gender, age, and village) as well as detailed administrative histories from both the pension and health insurance systems. Specifically, I observe pension type, enrollment, transfer, retirement, death, and benefit amounts, together with a complete contribution history that can be traced even prior to 2010. On the health insurance side, the data report insurance type, enrollment and transfer dates, medical expenditures, and disease diagnoses, again accompanied by a complete contribution history. In other words, the dataset provides both the payment history and the claims history for every farmer across pension and health insurance programs.

This richness makes the data particularly well-suited for studying adverse selection.

Summary Statistics. Because the health status of land-expropriated farmers prior to transfer eligibility affects their decision to transfer, we use medical expenditure data from the three years before retirement to compare those who transferred with those who were eligible but did not transfer.

Table 1. Summary Statistics: Pension vs. Health Insurance (Pre-2020)

	Total	Pension Insurance			Health Insurance		
	All	Farmer	Employee	*p-val*	Farmer	Employee	*p-val*
Outpatient Clinic Expenses (K Yuan)	2.10 (2.20)	0.92 (1.46)	2.36 (2.25)	0.000***	2.06 (2.23)	2.58 (2.24)	0.000***
Hospital Expenses (K Yuan)	1.90 (10.80)	2.96 (17.01)	1.50 (9.89)	0.001***	1.32 (8.75)	1.64 (10.65)	0.048**
Prescribed Disease Expenses (K Yuan)	0.10 (1.50)	0.18 (2.28)	0.09 (1.41)	0.289	0.09 (1.52)	0.08 (1.32)	0.474
Outpatient Clinic Visits	12.80 (10.60)	8.25 (9.29)	14.53 (11.43)	0.000***	13.12 (11.34)	15.56 (11.38)	0.000***
Hospitalization Visits	0.15 (0.55)	0.17 (0.51)	0.13 (0.43)	0.001***	0.12 (0.35)	0.14 (0.49)	0.028**
Prescribed Disease Visits	0.18 (1.50)	0.19 (2.00)	0.16 (1.40)	0.536	0.16 (1.95)	0.15 (2.50)	0.700
Male	0.48 (0.50)	0.73 (0.44)	0.40 (0.49)	0.000***	0.26 (0.44)	0.51 (0.50)	0.000***
Premium (Employee Pension / Health)		74.89 (20.77)	43.55 (28.43)	0.000***	29.49 (4.37)	21.81 (8.97)	0.000***
Observations	15,095	3,458	11,451		4,817	6,634	

Means with standard deviations in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1 presents summary statistics comparing land-expropriated farmers who remained in the low-premium farmer pension and health insurance programs with those who voluntarily upgraded to the high-premium employee schemes before 2020. Because of the semi-bundled design of the programs, for health insurance we restrict the comparison to individuals who upgraded to the employee pension, distinguishing between those who also upgraded to employee health and those who remained in farmer health.

Approximately 76% of eligible farmers (11,451 out of 15,095) transferred to the employee pension program. Among those who upgraded to the employee pension, about 58% (6,634 out of 11,451) also transferred to the employee health insurance program.

Individuals in the employee pension program have significantly higher outpatient expenditures (2.36 vs. 0.92 thousand yuan, $p < 0.01$) and more outpatient visits (14.5 vs. 8.3, $p < 0.01$), while exhibiting lower hospitalization expenses (1.50 vs. 2.96 thousand yuan, $p < 0.01$). These patterns suggest that individuals who upgraded are more health-conscious and rely less on inpatient care, consistent with selection on healthier types.

Analogous comparisons for health insurance show that upgraders again exhibit higher outpatient spending and visit frequencies, with statistically significant differences at the 1% level, while their hospitalization costs are slightly higher but economically small. Differences in prescribed-disease expenditures are insignificant, suggesting that chronic conditions alone do not drive the decision to upgrade.

Men are disproportionately represented in the employee pension program (73% vs. 40%, $p < 0.01$), while women are more likely to remain in the farmer scheme, suggesting that upgrading to the employee pension is a better deal for women since they retire at age 50, ten years earlier than men. The average premium for those who upgraded to the employee pension is 43,550 yuan, compared with 74,890 yuan for those who remained in the farmer pension, indicating that individuals facing lower premiums are more likely to upgrade. A similar pattern holds for health insurance, with average premiums of 21,810 yuan versus 29,490 yuan, respectively. Overall, the table reveals systematic differences in both demographic and health characteristics, consistent with adverse selection in both programs.

Table 2. Summary Statistics: Pension vs. Health Insurance (2020)

	Total	Pension Insurance			Health Insurance		
	All	Farmer	Employee	*p-val*	Farmer	Employee	*p-val*
Outpatient Clinic Expenses (K Yuan)	1.53 (2.35)	1.65 (1.88)	1.35 (2.90)	0.0110**	1.54 (1.85)	1.51 (2.86)	0.7860
Hospital Expenses (K Yuan)	1.35 (12.04)	1.70 (15.00)	0.84 (5.16)	0.0830*	1.39 (10.08)	1.31 (14.15)	0.9030
Prescribed Disease Expenses (K Yuan)	0.05 (0.93)	0.06 (1.12)	0.04 (0.56)	0.6000	0.06 (1.14)	0.05 (0.58)	0.8090
Outpatient Clinic Visits	14.7 (10.6)	15.9 (10.68)	13.1 (9.63)	0.0000***	15.0 (10.58)	14.6 (10.06)	0.4470
Hospitalization Visits	0.16 (0.47)	0.16 (0.41)	0.15 (0.65)	0.8260	0.16 (0.46)	0.15 (0.59)	0.4570
Prescribed Disease Visits	0.17 (1.53)	0.17 (1.99)	0.18 (1.56)	0.8770	0.17 (1.99)	0.18 (1.59)	0.8500
Male	0.42 (0.49)	0.39 (0.49)	0.48 (0.50)	0.0002***	0.40 (0.49)	0.47 (0.50)	0.0076***
Premium (Employee Pension / Health)		105.8 (7.69)	74.1 (32.43)	0.0000***	32.75 (2.31)	30.78 (6.79)	0.0000***
Observations	1,791	1,066	725		1,003	788	

Means with standard deviations in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 reports summary statistics for land-expropriated farmers in 2020, comparing those who remained in the low-premium farmer pension and health insurance programs with those who upgraded to the high-premium employee schemes.

Approximately 40% of eligible farmers (725 out of 1,791) transferred to the employee pension program. About 79% (788 out of 1,003) also transferred to the employee health insurance program.

Upgraders to the employee pension program exhibit lower average outpatient and hospitalization expenditures but also fewer clinic visits, suggesting improved health utilization efficiency rather than higher overall spending. Men are more likely to participate in the employee pension and health programs (48% vs. 39% and 47% vs. 40%, respectively; both $p < 0.01$). Average premiums differ substantially across groups, 74,100 yuan versus 105,800 yuan for pension insurance and 30,780 yuan versus 32,750 yuan for health insurance, indicating that those facing lower contribution requirements are more likely to upgrade. Overall, the 2020 data suggest that while differences in basic health indicators narrow after program expansion, systematic selection by gender and premium levels remains evident.

3 Testing for Adverse Selection

This section presents our test for adverse selection in the employee pension and health insurance.

Life Table. We construct life tables[3] to obtain the retirement-age life expectancy for individuals who transferred to the employee pension and those who remained in the farmer

[3]Compiling a life table typically requires micro-level data on millions of individuals. Our dataset, by contrast, contains roughly 160,000 observations. To address the limited sample size within each age group, we construct a 10-year period life table by pooling individual records from City Cs pension registry between 2011 and 2020. Mortality rates and life expectancy are then calculated by five-year age intervals, which yields more stable and accurate age-specific estimates.

pension. If individuals who transferred exhibit longer life expectancy at retirement than those who stayed, this would suggest the presence of adverse selection.

Figure 1. Retirement Age Life Expectancy by Group

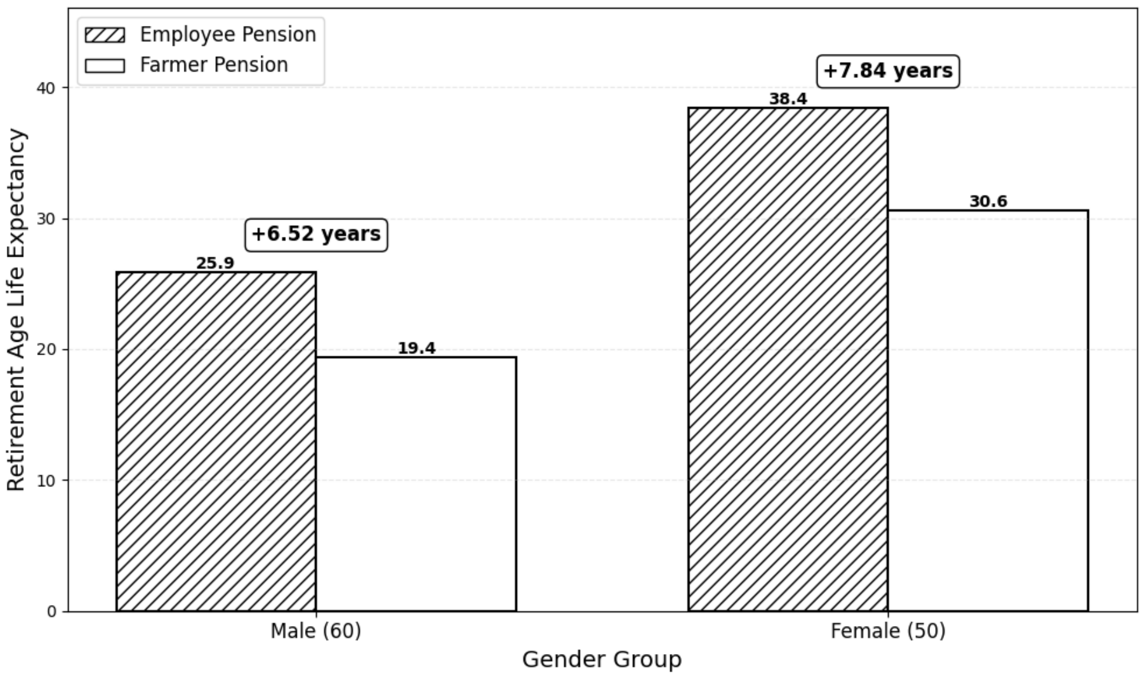


Figure 1 shows the calculated retirement age life expectancy by group. First, among individuals who transferred to the employee pension, average retirement-age life expectancy is 25.89 years for males and 38.39 years for females. Since it takes only about five years to break even, the pension appears to be operating at a substantial lossthis pattern is similar for the farmer pension as well. Second, for males, retirement-age life expectancy is 6.52 years longer among those who transferred than among those who did not, while for females it is 7.84 years longer. This finding suggests that individuals who transfer to the employee pension tend to live longer at retirement, indicating adverse selection in the transfer process.

To address potential concerns regarding moral hazard and the treatment effects of higher pension benefits, we use pre transfer medical expenses to predict individual's life expectancy and medical cost and conduct they the positive correlation test.

Cost Model. This section describes how I predict individual life expectancy and medical expenditure, which can then be used to test for adverse selection and be used to calculate the expected lifetime cost of each insurance participant.

To examine the relationship between pre-retirement medical utilization and post-retirement mortality risk, I estimate a stratified Cox proportional hazards model. The time scale is attained age. Individuals enter the risk set at the age when they become eligible for retirement, and are followed until death or censoring. Let T_i denote the time to death for individual i, and let $\mathcal{X}_i$ denote a vector of pre-retirement medical spending variables.

The hazard function is specified as:

$$h_i(t) = h_{0,s(i)}(t) \exp\left(X\beta + \gamma_{y(i)}\right),$$

where $h_{0,s(i)}(t)$ is an unspecified baseline hazard stratified by gender $s(i)$, allowing men and women to have different baseline risks. The covariates capture the average outpatient, inpatient, and prescribed disease medical expenditures and visits over the three years preceding retirement, which mitigates the influence of one-time medical shocks that are unlikely to predict post-retirement mortality. $\gamma_{y(i)}$ represents retire-year fixed effects. Standard errors are clustered at the individual level to account for serial correlation and heteroskedasticity.

Table 3. Stratified Cox Model: Medical Spending and Mortality

	Hazard Ratio	Robust SE
Outpatient spending (K yuan)	1.111***	(0.020)
Inpatient spending (K yuan)	1.003	(0.004)
Prescribed disease spending (K yuan)	1.014	(0.015)
Outpatient visits	0.985***	(0.003)
Inpatient visits	1.367***	(0.068)
Prescribed disease visits	0.993	(0.019)
Retire-year FE	Yes	
Catastrophic disease FE	Yes	
Observations	62,779	
Deaths	1,970	

Cox proportional hazards model with attained age as the time scale. Baseline hazard stratified by gender. Robust standard errors clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results in Table 3 show a strong and statistically significant relationship between pre-retirement medical utilization and post-retirement mortality risk. Higher outpatient spending prior to retirement is associated with a substantially higher hazard of death, with a one-thousand-yuan increase in outpatient expenses linked to an approximately 11.1% increase in mortality risk. Inpatient and prescribed-disease expenditures, by contrast, show small and statistically insignificant effects, suggesting that routine medical spending is a more salient predictor of underlying health status than occasional or disease-specific costs. Consistent with this pattern, the frequency of outpatient visits is negatively associated with mortality, implying that individuals who seek regular outpatient care tend to live longer. In contrast, inpatient visits are strongly and positively associated with mortality, consistent with the interpretation that hospitalization episodes reflect serious or terminal health shocks. Overall, these results indicate that pre-retirement medical utilization captures substantial heterogeneity in latent health risk that predicts life expectancy after retirement.

To predict post-retirement medical expenditure, I use the average medical costs in the three years prior to retirement. This pre-retirement average summarizes each individuals underlying health risk and medical consumption patterns and provides a strong predictor of future medical expenditure.

Figure 2. Comparison of Total Medical Expenses Pre- and Post-Retirement

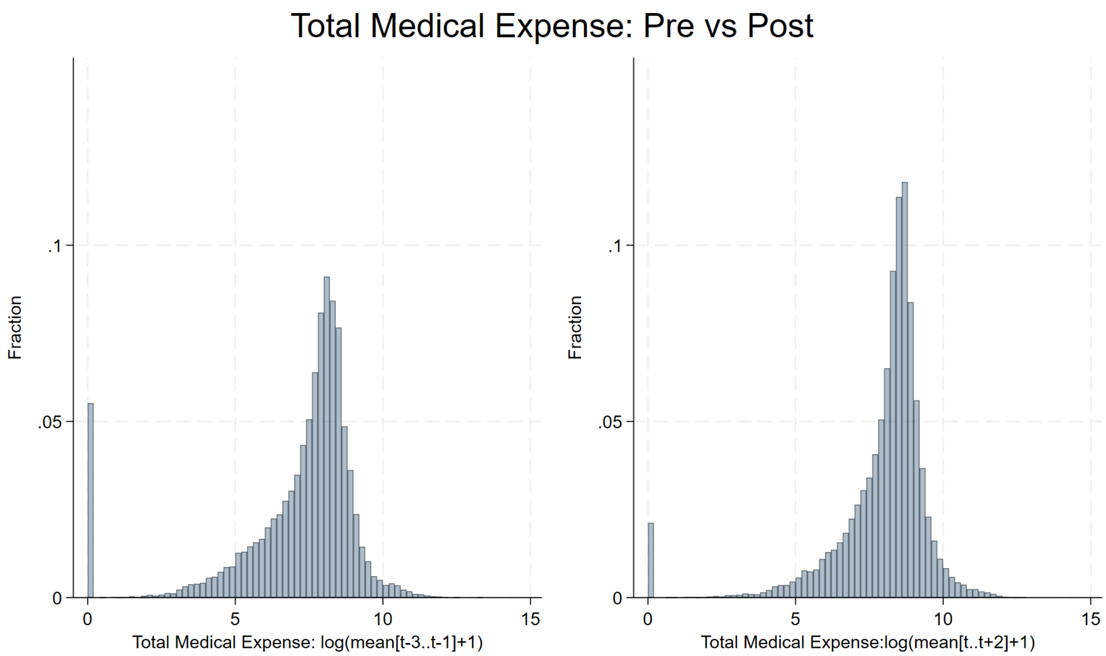
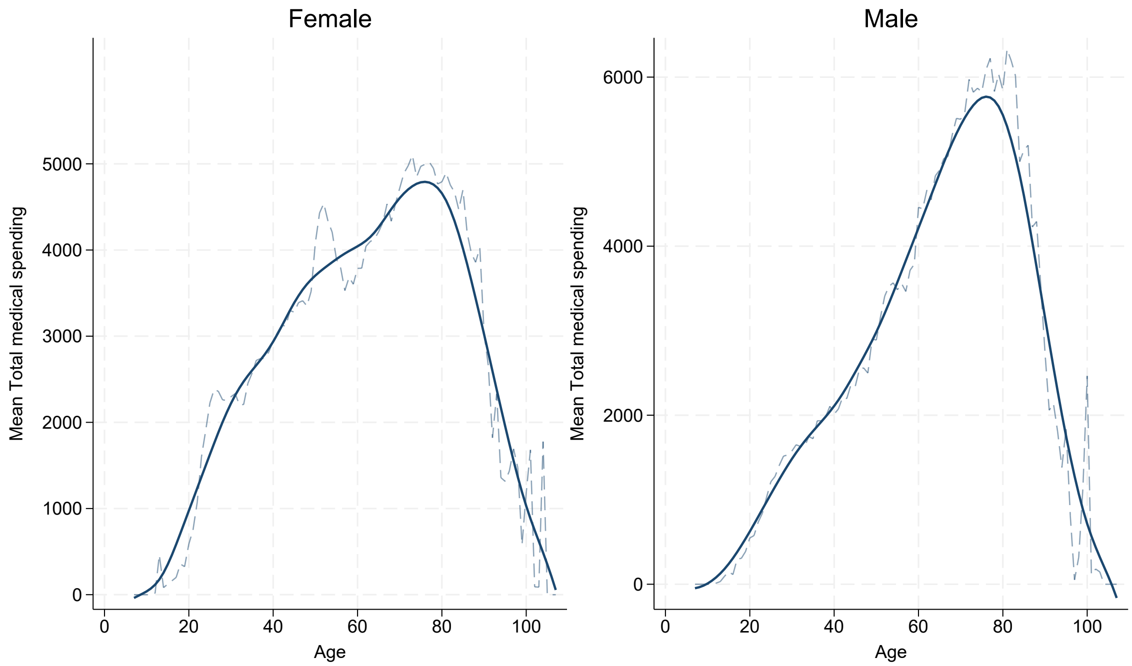


Figure 2 shows that pre-retirement medical expenditures are strongly predictive of post-retirement expenditures. The slight rightward shift indicates potential moral hazard after retirement, as well as increased medical spending associated with aging.

Figure 3 plots the evolution of average total medical expenses by age and gender. Medical spending follows an inverted-U pattern for both men and women: it rises steadily through middle age, peaks around the mid-60s to 70s, and then declines sharply after age 80 as the population thins out. Men consistently exhibit higher medical expenditures across nearly all ages, reflecting both higher utilization intensity and potentially greater incidence of severe conditions. Womens medical spending, while lower on average, increases more gradually and declines later in life, consistent with their longer life expectancy. The smoother age profile among women also suggests fewer extreme medical shocks compared with men.

Using the predicted life expectancy from the Cox model, together with each individ-

Figure 3. Average Total Medical Expenses by Age and Gender



Each panel plots the average total medical spending (yuan) by age, separately for females (left) and males (right). The gray line shows the age-specific mean, and the blue line represents the lowess-smoothed fit. Medical spending includes outpatient, inpatient, and prescribed-disease expenses.

uals average pre-retirement medical expenditure and the estimated age profile of medical spending, I compute the expected lifetime medical cost for each person.

Employee Pension. We estimate a logit model in which the decision to transfer into the employee pension is regressed on predicted post-retirement life expectancy, controlling for pension premiums, gender, and retirement-year fixed effects. We restrict the sample to men retiring at age 60 and women retiring at age 50, because individuals retiring later effectively face a different pension product. A positive coefficient on predicted life expectancy is consistent with adverse selection.

$$\log\left(\frac{\Pr(D_i = 1)}{1 - \Pr(D_i = 1)}\right) = \beta_0 + \beta_1 X_i + \lambda_{it} + \epsilon_i$$

Where D_i indicates whether farmer i transfers to the employee pension, and X_i includes predicted post-retirement life expectancy, pension premium, and gender. The model also includes retirement-year fixed effects λ_{it} to capture end-of-year revisions to benefit formulas. Although we include retire-year fixed effects, note that the data are not panel, since each individual makes only one transfer decision at the time of retirement.

Table 8 presents the logit regression results for employee pension transfer decisions.[4]

Table 4. Employee Pension Take-up and Predicted Longevity

	Pre-2020		2020	
	(1)	(2)	(3)	(4)
Life Expectancy	0.020***	0.005***	-0.004*	0.014*
	(0.001)	(0.008)	(0.002)	(0.008)
Premium		-0.012***		-0.009***
		(0.000)		(0.001)
Male		-0.088***		0.206***
		(0.016)		(0.077)
Retire-year FE	No	Yes		
Observations	14,909		1,791	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Regression results are presented in Table A2. Before 2020, predicted longevity is a strong and statistically significant predictor of upgrading to the employee pension: each additional expected year of life increases the probability of transferring by about 0.5 percentage points. This pattern is consistent with adverse selection, as individuals who expect to live longer are more likely to select into the high-premium, high-benefit employee pension system. The negative coefficient on premium further indicates that individuals facing higher contribution requirements are less likely to upgrade.

For 2020, the coefficients on predicted life expectancy remain positive but are less precisely estimated: each additional expected year of life increases the probability of transferring by about 1.4 percentage points. The weaker statistical significance likely reflects the smaller sample size in 2020. Interestingly, the coefficient on *Male* switches sign from negative in the pre-2020 sample to positive in 2020, suggesting that earlier cohorts of female farmerswho retire ten years earlierhad stronger incentives to upgrade initially, whereas later male cohorts began catching up as the policy matured. Overall, these results confirm that expected longevity remains an important determinant of pension upgrading decisions, consistent with the presence of adverse selection in the employee pension program.

Employee health. To evaluate the benefit of transferring, lets do some simple math.

The average annual medical expense is about 3,000 yuan. If the reimbursement rate rises by 30%, the annual out-of-pocket savings would be around 900 yuan. Adding the

[4]The sample is divided into two periods: before 2020 and the year 2020. Prior to 2020, a semi-bundling policy required individuals to enroll in the employee pension to access employee health insurance. This requirement was lifted in 2020, allowing independent enrollment. However, results from 2020 should be interpreted with caution, as many retirement records are missing due to COVID-related delays in provincial processing. We are currently working to obtain those records. For now, the analysis focuses on the pre-2020 period.

difference in deductibles, lets say the total annual savings is about 1,000 yuan. Over 20 years, which is close to the average life expectancy for men after retirement, this adds up to only about 20,000 yuan. Since the premium exceeds 30,000 yuan, this alone does not seem like a very good deal.

However, individuals who transfer also receive an individual medical account, which provides an annual contribution of about 3,000 yuan. This account balance is both inheritable and continuously rolling over time. For men, who on average live more than 20 years after retirement, this amounts to roughly 60,000 yuanwell above the total premium paid. For women, who typically live more than 30 years after retirement, the accumulated benefit can reach about 90,000 yuan (3,000 30).

Recall the average annual rural income in City C was about 40,000 yuan in 2020, so these amounts are far from trivial.

So transferring appears to be a very good deal, especially for individuals who live longer. Based on this, we expect that those in higher expected life expectancy would be more likely to transfer to Employee Health.

Table 5. Employee Health Take-up and Predicted Life Expectancy

	Pre-2020		2020	
	(1)	(2)	(3)	(4)
Life Expectancy	-0.025***	0.005*	-0.005**	0.007
	(0.001)	(0.003)	(0.002)	(0.007)
Medical Exp.	-0.000	0.001**	-0.001*	-0.000
	(0.000)	(0.000)	(0.001)	(0.001)
Premium		-0.021***		-0.025***
		(0.001)		(0.004)
Male		0.360***		0.123*
		(0.025)		(0.073)
Retire-year FE	No	Yes		
Observations	11,451		1,791	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5 presents the results for employee health insurance take-up. Before 2020, when enrollment was semi-bundled with the employee pension, the sample is restricted to individuals who had already upgraded to the employee pension. Within this group, individuals with higher predicted life expectancy are slightly more likely to transfer to the employee health program: each additional expected year of life increases the probability of transferring by about 0.5 percentage points. This pattern is consistent with the idea that longer-lived individuals expect to accumulate larger balances in their individual medical accounts. Medical expenditures are also positively associated with take-up, suggesting that individuals who spend more on healthcare are more inclined to enroll, possibly reflecting higher expected utilization or stronger perceived benefits from coverage. As expected, higher premiums significantly reduce participation, while men are roughly 36 percentage points more likely than women to upgrade.

For 2020, the estimated effects of predicted life expectancy and medical expenditures are small and statistically insignificant, likely due to the smaller sample size and the stabilization of participation once the bundling constraint was lifted. Premiums remain a strong negative predictor of take-up, and the gender gap narrows substantially to about 12 percentage points. Taken together, these results suggest that before 2020, adverse selection into employee health insurance was primarily driven by differences in life expectancy and medical spending, whereas by 2020, such selection effects appear much weaker or largely absent.

Robustness. First, I estimate alternative specifications that replace predicted life expectancy and medical expenditures with three-year averages of pre-transfer medical expenses and visit frequencies. The results, reported in the Appendix, remain robust to these alternative measures.

Second, I calculate each individuals expected incremental cost from upgrading from the farmer to the employee pension or health insurance schemes. These results, also presented in the Appendix, are consistent with the main findings.

Table 6. Income Regression Result (Pre-2020, Workers)

	Pension			Health		
	(1)	(2)	(3)	(4)	(5)	(6)
Life Expectancy	0.004***	-0.000	-0.000	-0.051***	0.003	0.003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)	(0.006)
Medical Exp.				-0.004***	0.001	0.001
				(0.001)	(0.001)	(0.001)
Income Level			-0.008			0.129*
			(0.041)			(0.067)
Premium		-0.002***	-0.002***		0.022***	0.022***
		(0.000)	(0.000)		(0.000)	(0.000)
Male		-0.046**	-0.046**		0.517***	0.519***
		(0.020)	(0.020)		(0.039)	(0.040)
Retire-year FE	No	Yes	Yes	No	Yes	Yes
Observations	3,200	3,200	3,200	3,085	3,085	3,085

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Third, I examine whether the observed adverse selection may partly reflect income effects, specifically, whether wealthier individuals are simply more able to afford the higher premiums and also tend to be healthier or more cautious. To address this, I focus on land-expropriated farmers with prior work experience and use their pension contribution records to approximate income. [5]

Table 6 reports the results of the income robustness analysis using the subsample of land-expropriated farmers with prior employment experience. Columns (1)–(3) examine pension take-up, while Columns (4)–(6) analyze health insurance take-up. Across specifications, controlling for income does not meaningfully alter the estimated effects of predicted life expectancy or premiums. For pension insurance, the coefficient on life expectancy remains small and statistically insignificant once income is included, while the premium continues to have a significant negative effect. For health insurance, income exhibits a weakly positive coefficient in Column (6), but the magnitude is small and does not substantially change other coefficients. Overall, these results suggest that the observed adverse selection patterns are not primarily driven by income differences among workers, reinforcing the interpretation that health and longevity expectations are the main determinants of upgrading behavior.

Table 7. Income Regression Result (Pre-2020, Non-Workers)

	Pension		Health	
	(1)	(2)	(3)	(4)
Life Expectancy	0.020***	0.007***	-0.019***	0.002
	(0.001)	(0.002)	(0.001)	(0.004)
Medical Exp.			0.002**	0.005***
			(0.001)	(0.001)
Premium		-0.014***		0.028***
		(0.000)		(0.001)
Male		-0.099***		0.263***
		(0.020)		(0.040)
Retire-year FE	No	Yes	No	Yes
Observations	11,709		8,366	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7 reports the regression results for non-workers, that is, land-expropriated farmers without prior formal employment records. The results reveal strong evidence of adverse selection in both the pension and health insurance programs. For the pension system,

[5] Although firms have an incentive to underreport wages to reduce contribution obligations, these data still provide a reasonable proxy for individual earnings.

predicted life expectancy is a highly significant predictor of upgrading: each additional expected year of life increases the probability of transferring to the employee pension by about 0.7 percentage points after controlling for premiums and gender. This indicates that individuals expecting to live longer are more likely to select into the high-premium, high-benefit employee pension scheme. In contrast, for the health insurance system, medical expenditures are a significant positive predictor of take-up, suggesting that individuals with higher expected healthcare costs are more likely to upgrade to the employee health program. Overall, these patterns confirm pronounced adverse selection among non-workers: longevity-based selection in the pension system and health-cost-based selection in the health insurance system.

The absence of adverse selection after controlling for premiums in the worker group may simply reflect the endogeneity of the premium variable: individuals who expect to live longer or remain healthier tend to work longer and thus face lower effective premiums at retirement.

Last I test for the role of liquidity constraints. In 2016, the Agricultural Bank of C City introduced the Harvest Pension Loan. Land-expropriated farmers in C City who qualify for a lump-sum payment to transfer to the employee pension and are under 70 years old can borrow up to 50,000 yuan to cover the employee pension premium.[6] The loan has a maximum term of five years. We use a regression discontinuity design, comparing individuals just below and just above age 70, to test whether this policy affects their transfer decisions.

Table 8. Liquidity Constraint Regression Result

	Pension		Health	
	Pre-2016	Post-2016	Pre-2016	Post-2016
Conventional	-0.079***	-0.134***	-0.053	0.012
	(0.021)	(0.037)	(0.067)	(0.030)
Bias-corrected	-0.037*	-0.092**	-0.098	0.047
	(0.021)	(0.037)	(0.067)	(0.030)
Robust	-0.037	-0.092*	-0.098	0.047
	(0.028)	(0.052)	(0.121)	(0.034)
Controls	Yes	Yes	Yes	Yes
$\Pr(D_p = 1)$	0.5	0.5	0.2	0.2
Number Obs	71695	21419	3893	16055

Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

For the pension outcome, the regression coefficients are negative in both periods, with the post-2016 effect approximately twice as large as the pre-2016 effect. Taking the robust estimates as an example, after 2016, individuals just above age 70 are 2.3= $(0.092 * 0.5 * (10.5) * 100)$ percentage points less likely to transfer to the employee pension than those just below 70. Before 2016, this difference was only 0.925= $(0.037 * 0.5 * (10.5) * 100)$ percentage

[6] Those under 65 receive a credit loan, while those over 65 must provide a family guarantee.

points and statistically insignificant. This pattern suggests that the Harvest Pension Loan, introduced in 2016, effectively reduced liquidity constraints and enabled more farmers under age 70 to transfer to the employee pension.

As the loan program applies only to the employee pension, it is expected that the regression coefficients for employee health insurance are statistically insignificant, which is consistent with the results.

4 Welfare Analysis

Thus far, we have presented evidence of adverse selection in both pension and health insurance. We now turn to the welfare analysis: quantifying the welfare loss resulting from adverse selection.

Demand and cost estimation. We assume demand follows a logit specification, since when demand depends on two prices, it can easily take values outside the $[0, 1]$ range under a linear model. We specify the average cost as a function of prices rather than quantities, and later recover the standard cost curve, as it is easier to find exogenous prices. Assuming that both the logit demand and average cost curves are linear in prices, we estimate the demand and average cost curves for employee pension and employee health insurance as follows.

$$\begin{aligned}
\log\left(\frac{D_i^p}{1-D_i^p}\right) &= \alpha^p + \beta_p^p p_i^p + \beta_h^p p_i^h + \epsilon_i^p \\
\log\left(\frac{D_i^h}{1-D_i^h}\right) &= \alpha^h + \beta_p^h p_i^p + \beta_h^h p_i^h + \epsilon_i^h \\
c_i^p &= \gamma^p + \delta_p^p p_i^p + \delta_h^p p_i^h + u_i^p \\
c_i^h &= \gamma^h + \delta_p^h p_i^p + \delta_h^h p_i^h + u_i^h
\end{aligned}$$

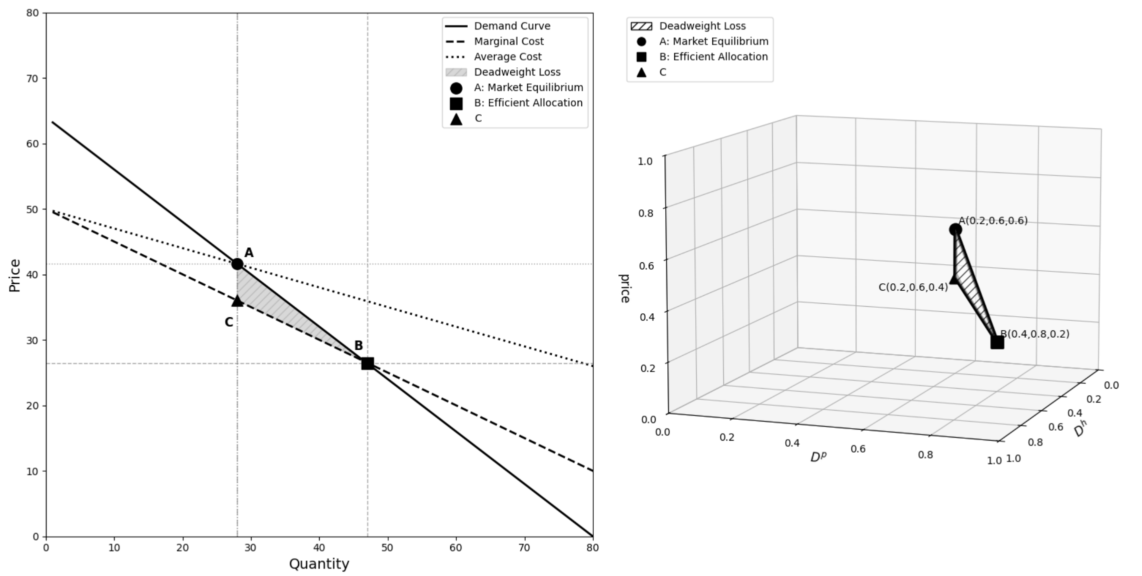
where, $D_i^p(D_i^h)$ is a dummy variable equal to 1 if farmer i chooses to transfer to the employee pension (employee health); $c_i^p(c_i^h)$ is the incremental cost to the government from covering farmer i under the employee pension (employee health) instead of the farmer pension (farmer health); and $p_i^p(p_i^h)$ denotes the incremental payment that farmer i is required to pay to transfer to the employee pension (employee health). The demand and cost of employee pension and health insurance depend on the price of the other insurance, reflecting both complementarity (or substitution) effects and the impact of bundling. In all regressions, standard errors are adjusted to account for an arbitrary variance-covariance matrix within each village. The demand equations are estimated on the entire sample whenever applicable, while the (average) cost equations are estimated on the sample of farmers who (endogenously) chose to transfer.

Given the demand and average cost curves, we can derive the marginal cost curves:

$$MC^p = c_i^p + \frac{\delta_p^p}{\beta_p^p(1 - D_i^p)}$$
$$MC^h = c_i^h + \frac{\delta_h^h}{\beta_h^h(1 - D_i^h)}$$

Graphical Representation With linear demand and cost curves, we can present a graphical analysis of adverse selection and the resulting welfare loss.

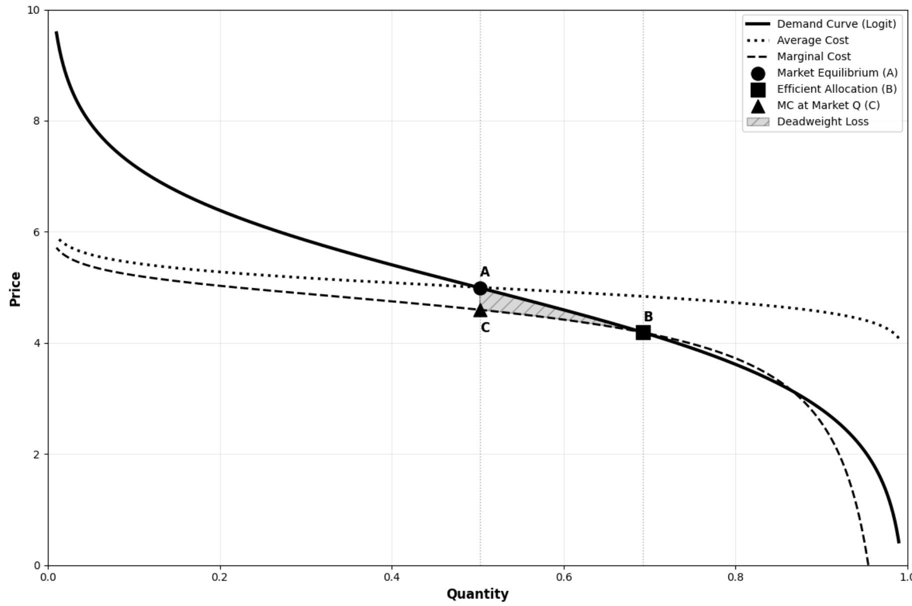
Figure 4. Efficiency Cost of Adverse Selection (Pension)



In the standard single-insurance setting as the left subgraph of Figure 5, adverse selection is indicated if the marginal cost curve slopes downwardbecause individuals with the highest willingness to pay also tend to have the highest expected costs; for example, the first people to enroll in health insurance are often the sickest. To measure deadweight loss, point A is the competitive equilibrium where price equals average cost, point B is the efficient allocation where price equals marginal cost, and the triangle ABC represents the deadweight loss.

In my setting, there are two insurances. Using the two demand and average cost equations, I solve for the market equilibrium: in the pension graph, this is point A, with pension quantity 0.6, health insurance quantity 0.2, and pension price 0.6. Using the two demand and marginal cost equations, I solve for the efficient allocation: point B, with pension quantity 0.8, health insurance quantity 0.4, and pension price 0.2. Point C represents the marginal cost of pension insurance at the market equilibrium quantities (pension 0.6, health insurance 0.2). The area ABC is the deadweight loss for pension insurance. The analysis for health insurance is analogous, except that the price and cost axes refer to health insurance instead of pension.

Figure 5. Efficiency Cost of Adverse Selection (Pension)



Under the logit demand case, the welfare analysis parallels the linear case: we compute the market equilibrium and the efficient allocation and then compare them to obtain the deadweight loss. Because the logit specification may yield two points where price equals marginal cost, we choose the lower one, which corresponds to the larger deadweight loss.

Identification Strategy To conduct the welfare analysis described above, we estimate the demand and average cost curves for both the employee pension and employee health insurance. A central challenge in this exercise is identifying exogenous variation in their respective premiums.

An individuals employee pension premium depends on their years of employment. Specifically, their initial lump-sum payment from the farmer pension is credited as six years of participation in the employee pension system. To qualify for pension benefits after retirement, individuals must accumulate at least 15 years of contributions. Therefore, those who have never worked must make additional payments equivalent to nine years of contributions, while each additional year of work reduces this required payment by one year. A similar rule applies to the employee health insurance program. The initial payment is credited as ten years of participation, and eligibility for post-retirement health benefits requires at least 20 years of contributions. Since the employee health program was established in 2006, only years of formal employment after 2006 count toward this contribution requirement.

Based on years worked, we can calculate each individuals premium. However, individuals may endogenously choose to work longer after land expropriation in order to lower their premium at retirement. To address this, I use years worked at the time of land expropriation and age at expropriation as instruments. They are highly correlated with the pension and health premiums. Both of which are considered exogenous. Because land expropriation decisions are made by the government at the village level and are unlikely to be correlated with individual characteristics.

Estimation Result. We estimate the demand and cost curves for the unbundled setting using data from 2020, as the semi-bundled structure prior to 2020 makes it difficult to identify complementarity and substitution effects directly. The corresponding results for the pre-2020 period are reported in the Appendix.

Table 9. Demand and Cost Regressions Results

Panel A: Baseline

	dp	cp	dh	ch
p^p	-0.094***	0.054**	-0.069***	0.018
	(0.006)	(0.023)	(0.005)	(0.011)
p^h	-0.010	-0.163**	-0.064***	0.036
	(0.014)	(0.066)	(0.016)	(0.027)
male	0.140	-80.919***	0.067	-29.944***
	(0.127)	(1.004)	(0.113)	(0.523)
Constant	8.960***	531.652***	8.455***	124.361***
	(0.674)	(2.354)	(0.969)	(1.105)
Observations	1791	1791	1791	1791
R-squared		0.800		0.677

Panel B: Instrumental Variable

	dp	cp	dh	ch
p^p	-0.259**	0.055	-0.045***	0.014*
	(0.127)	(0.046)	(0.006)	(0.008)
p^h	-0.420	-0.387**	-0.003	0.017
	(0.385)	(0.185)	(0.038)	(0.056)
male	-0.755	-80.938***	0.217*	-29.960***
	(0.676)	(0.966)	(0.120)	(0.524)
Constant	38.348	538.707***	4.256***	125.339***
	(23.931)	(4.781)	(1.319)	(1.591)
Observations	1791	1791	1791	1791
R-squared		0.800		0.676

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 9 Panel B reports the instrumental variable estimates of the demand and cost equations for employee pension and health insurance. As expected, the own-price elasticities of demand are negative and statistically significant: higher premiums reduce take-up in both programs, consistent with standard demand behavior. The cross-price elasticity of pension demand with respect to the health premium is negative, suggesting that employee health insurance is a complement to the employee pension. After obtaining employee health coverage, individuals can better afford treatment for severe illnesses, which enhances their confidence in longevity and raises the perceived value of pension benefits. In contrast, the coefficient on the pension price in the health demand equation is negative but small, indicating that employee pension and employee health act as mild substitutes: those with higher pensions expect greater post-retirement income and thus feel less need for additional health coverage.

On the cost side, the own-price coefficients are positive, while the corresponding demand coefficients are negative. This inverse relationship between demand and cost implies the presence of adverse selection: individuals who are more likely to purchase the insurance (i.e., those first join the insurance) tend to be higher-cost participants.

$$\log\left(\frac{D^p}{1-D^p}\right) = 38.023 - 0.259\, p^p - 0.420\, p^h$$

$$\log\left(\frac{D^h}{1-D^h}\right) = 4.349 - 0.045\, p^p - 0.003\, p^h$$

$$c^p = 503.864 + 0.055\, p^p - 0.387\, p^h$$

$$c^h = 112.442 + 0.014\, p^p + 0.017\, p^h$$

$$MC^p = c^p + \frac{0.055}{-0.259\,(1-D^p)}$$

$$MC^h = c^h + \frac{0.017}{-0.003\,(1-D^h)}$$

Using the average male proportion of 0.43, we calculate the population-level demand and cost curves and recover the corresponding marginal cost curve.

Both the competitive equilibrium and the efficient allocation feature demand levels close to zero, as the implied costs are extremely high. Recall that it takes only about five years for participants to break even, while the average individual can live more than twenty years after retirement. A similar logic applies to the individual medical accounts in the employee health system. Therefore, the welfare loss between the competitive equilibrium and the efficient allocation would be negligibly small. Moreover, since the current setting does not reflect a market equilibrium, our analysis instead compares the efficient allocation with the current average price, and measures the welfare loss as the area enclosed by the demand curve and the marginal cost curve at that price level.

Figure 6. Efficiency Cost under No Bundling

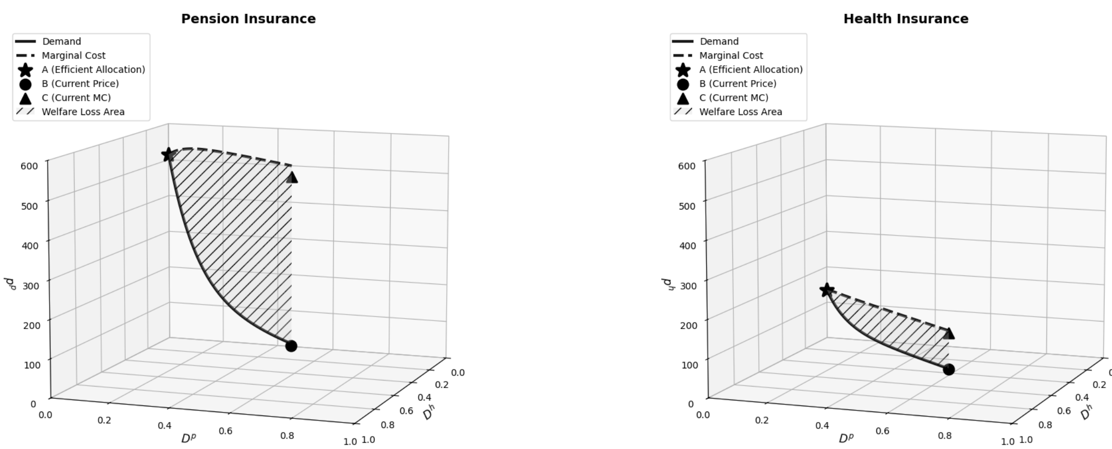


Figure 6 indicates that the welfare loss at the current price level, relative to the efficient allocation, amounts to approximately 372,000 yuan (about \$50,000) for pension insurance and 96,000 yuan for health insurance. Given that the average pension premium is about 108,000 yuan and the health insurance premium is 33,000 yuan, these welfare losses are economically substantial.

For pensions, retirees receive approximately 23,000 yuan annually after retirement, implying that they break even after roughly five years. However, men live an additional 25.9 years and women 38.4 years after retirement, indicating that the employee pension system naturally generates a large welfare imbalance even in the absence of adverse selection. Similarly, under the employee health insurance scheme, participants receive around 3,000 yuan per year in their individual medical accounts. If a retiree lives 30 years, this alone accumulates to nearly 90,000 yuan, far exceeding the total premium paid. Considering the higher reimbursement rate of the employee plan, the fiscal burden of the employee health system is even greater.

Note that this welfare loss is primarily driven by prices being set far below the corresponding costs, rather than by adverse selection. Even in the absence of adverse selection, the welfare loss would remain substantial under the current pricing structuret

5 Counterfactual

This section examines whether bundling can mitigate adverse selection. To do so, I estimate a mixed logit model to recover each individuals willingness to pay for employee pension and health insurance, as well as the complementarity between them. This approach allows me to simulate individual choices under alternative pricing and bundling regimes and, combined with the estimated cost structure, conduct counterfactual welfare analysis.

Figure 7. Efficiency Cost of Adverse Selection (Pension)

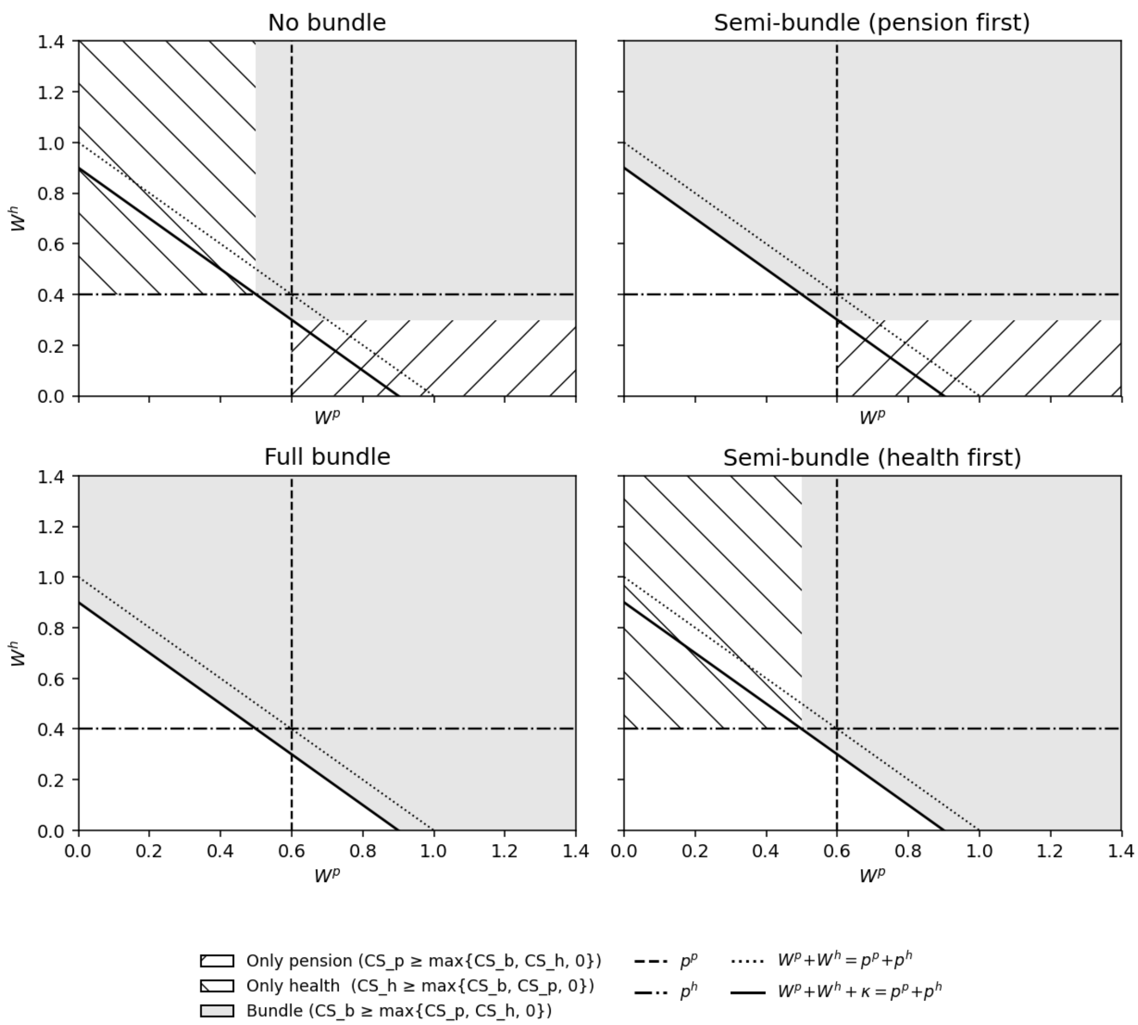


Figure 7 shows the choice sets under different bundling structures. The x-axis shows willingness to pay for pension, and the y-axis shows willingness to pay for health insurance. The pension price is set at 0.6, and the health insurance price at 0.4. Under no bundling, anyone with pension WTP above 0.6 buys pension, and anyone with health WTP above 0.4 buys health insurance. If a complementarity effect exists, individuals value the bundle more than the simple sum of the two parts. This effect is reflected in a leftward shift of the bundle threshold curve, which I denote by κ. Individuals whose bundle WTP is above the sum of the two prices will buy both insurance. Finally, if both single products and the bundle are attractive, people choose whichever option gives the highest consumer surplus. For example, someone in the top left region would consider both health insurance and the bundle, but since health insurance yields higher consumer surplus, they would choose health insurance.

Under semi-bundling, some individuals keep their original choices. But notice two important shifts: A group switches from buying only health to buying the bundle, which leaves them with lower consumer surplus. Another group switches from buying only health to buying nothing, also reducing their consumer surplus.

We can also illustrate the choice set under full bundling or the other semi-bundling scenarios. In fact, the choice set can be illustrated under any bundling or pricing structure. With this choice-set framework, once we recover the joint distribution of willingness to pay (WTP) and the complementarity parameter κ, we can derive the choice set under any bundling or pricing structure.

Random Coefficient Logit Model. How do we recover the joint distribution of willingness to pay and kappa?

I estimate a randome coefficient logit model. Before 2020, enrollment in employee health insurance was conditional on participation in the employee pension, creating a semi-bundled structure:

$$o \in \Omega_{SB} = \{0, P, PH\},$$

where 0 denotes remaining in the farmer pension and health programs, P denotes upgrading to the employee pension only, and PH denotes upgrading to both employee pension and employee health insurance.

In 2020, this restriction was lifted, allowing individuals to purchase employee health insurance independently:

$$o \in \Omega_{UB} = \{0, P, H, PH\},$$

where H denotes upgrading to employee health insurance only.

The indirect utility that individual i derives from alternative o in regime $r \in \{SB, UB\}$ is

$$U_{ior} = V_{ior} + \varepsilon_{ior}, \qquad \varepsilon_{ior} \overset{i.i.d.}{\sim} \text{Type-I EV}.$$

The deterministic component is assumed to be quasilinear, which conveniently leads a welfare analysis based on total surplus:

$$V_{ior} = \begin{cases} 0, & o = 0 \quad \text{(outside option)}, \\ W_{Pi} - p^p_{ir}, & o = P, \\ W_{Hi} - p^h_{ir}, & o = H, \\ W_{Pi} + W_{Hi} + \kappa_i - (p^p_{ir} + p^h_{ir}), & o = PH, \end{cases}$$

where p^p_{ir} and p^h_{ir} are the pension and health insurance premiums faced by individual i under regime r. κ_i captures the complementarity (or substitutability) in utility between pension and health insurance.

The random coefficients are defined as:

$$\begin{aligned} \ln W_{Pi} &= b^p_0 + b^p_1 X_i + \sigma_p \eta_{1i}, \\ \ln W_{Hi} &= b^h_0 + b^h_1 Z_i + \sigma_h \eta_{2i}, \\ \kappa_i &= b^k_0 + b^k_1 M_i + \sigma_k \eta_{3i}, \end{aligned}$$

$$(\eta_{1i}, \eta_{2i}, \eta_{3i})' \sim \mathcal{N}(0, \Sigma), \qquad \Sigma = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where X_i, Z_i, and M_i denote observed individual characteristics (e.g., gender, income, pre-transfer medical costs, or predicted life expectancy and total medical costs).

Conditional on η_i, the choice probability follows the multinomial logit form:

$$P_{ior}(\eta_i) = \frac{\exp\big(V_{ior}(\eta_i)\big)}{\sum_{j \in \Omega_r} \exp\big(V_{ijr}(\eta_i)\big)}.$$

The unconditional probability integrates over unobserved heterogeneity:

$$P_{ior} = \int P_{ior}(\eta_i)\, \phi(\eta_i; \Sigma)\, d\eta_i.$$

Because this integral has no closed form, it is approximated by simulation using S Halton draws $\{\eta_i^{(s)}\}_{s=1}^S$:

$$\widehat{P}_{ior} = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp\big(V_{ior}(\eta_i^{(s)})\big)}{\sum_{j \in \Omega_r} \exp\big(V_{ijr}(\eta_i^{(s)})\big)}.$$

The parameters $\theta = \{b^{\cdot}_{\cdot}, \sigma_{\cdot}, \rho, \Sigma\}$ are estimated by maximizing the simulated log-likelihood:

$$\ell(\theta) = \sum_i \ln \widehat{P}_{i,o,r}.$$

These estimates form the basis for counterfactual simulations of alternative pricing and bundling regimes, allowing welfare comparisons under observed and hypothetical policy environments.

Estimation Results. Table 11 presents the estimated parameters from the random-coefficient logit model. The mean utility parameters $b_0^p = 2.14$ and $b_0^h = 0.81$ indicate that, on average, individuals derive positive utility from enrolling in both the employee pension and employee health insurance programs relative to remaining in the farmer schemes. The slope coefficients $b_1^p = 0.15$ and $b_1^h = 0.66$ measure how these preferences vary with predicted individual contribution costs (c_z^p and c_z^h), respectively. The positive and sizable magnitudes of these coefficients suggest that individuals facing higher predicted pension or medical costs are substantially more likely to value and select into the corresponding insurance programs, consistent with the presence of adverse selection based on expected benefit utilization.

Table 10. Random Coefficient Logit Estimation Result

Parameter	Estimate	Interpretation
Mean utility parameters		
b_0^p	2.1425	Mean preference for pension insurance
b_1^p	0.1498	Effect of pension contribution total (cp_z)
b_0^h	0.8132	Mean preference for health insurance
b_1^h	0.6553	Effect of health contribution total (ch_z)
Variance and correlation parameters		
$\ln \sigma_p$	-6.9078	Log SD of pension utility ($\sigma_p = \exp(-6.9078) \approx 0.0010$)
$\ln \sigma_h$	-6.6469	Log SD of health utility ($\sigma_h = \exp(-6.6469) \approx 0.0013$)
α_ρ	0.6833	Fisher transform for corr.
$\rho = \tanh(\alpha_\rho)$	0.594	Corr. between pension and health utilities
$\ln \sigma_k$	-6.9078	Log SD of bundling utility ($\sigma_k = \exp(-6.9078) \approx 0.0010$)
Bundling component		
b_0^k	1.0026	Mean bundling synergy
b_1^k	0.9682	Heterogeneity in bundling synergy

The estimated correlation coefficient $\rho = 0.594$ indicates a moderate positive correlation between unobserved preferences for pension and health insurance. This implies that individuals who value one form of coverage tend also to value the otherconsistent with complementarities in perceived protection or long-term security. The very small estimated standard deviations ($\sigma_p \approx 0.0010$, $\sigma_h \approx 0.0013$) suggest limited unobserved heterogeneity in preferences for either standalone program after accounting for observable characteristics.

The bundling component further reveals strong complementarity between pension and health insurance when purchased jointly. The positive mean bundling term $b_0^k = 1.00$ indicates that combining the two programs generates additional utility beyond the sum of their individual values. Moreover, the large heterogeneity parameter $b_1^k = 0.97$ shows that this synergistic effect is much stronger among individuals with higher predicted medical costs or longer life expectancythose who stand to benefit most from integrated old-age protection.

Combining the structural estimation with predicted individual life expectancy and medical costs, we find that the consumer surplus is approximately 21.82 K yuan per capita, the producer surplus is 203.32 K yuan per capita, and the total surplus is 181.49 K yuan per capita. The positive consumer surplus (CS) indicates that, on average, individuals derive moderate utility gains from participating in the pension and health insurance programs. In contrast, the negative producer surplus (PS) suggests that the government incurs a fiscal loss under the current regime and contribution rates. Consequently, the total surplus (TS), defined as the sum of CS and PS, is negative, implying that although individuals value participation, the programs cost structure outweighs its welfare benefits under existing prices, highlighting the inherent trade-off between broad coverage and fiscal sustainability.

Counterfactual. To assess whether Pareto improvement can be achieved, we exploit the positive correlation between pension and health insurance costs observed in this setting. This correlation arises because the employee health insurance program includes an individual medical account that provides an annual deposit of approximately 3,000 yuan, so individuals with longer life expectancy tend to incur both higher pension and health costs. We therefore consider a counterfactual anti-bundling design in which the government provides a 500-yuan voucher specifically for purchasing health insurance, but not for the bundled option or the pension program.

Under this alternative policy, the simulated consumer surplus (CS) is 21.83 K yuan per capita, producer surplus (PS) is -203.07 K yuan per capita, and total surplus (TS) is -181.24 K yuan per capita, each slightly higher than in the baseline case. This suggests that offering targeted incentives for individual program participation can raise aggregate welfare without harming any group of participants, achieving a Pareto improvement.

Table 11. Health Subsidy Counterfactual

	Consumer Surplus	Provider Surplus (K yuan per capita)	Total Surplus
Baseline	21.82	-203.32	-181.49
Health Subsidy (500 yuan)	21.83	-203.07	-181.24
Δ (Subsidy - Baseline)	+0.01	+0.25	+0.25

This result illustrates that overall welfare can be improved without harming any group of participants by exploiting the underlying positive cost correlation between pension and health insurance. Because this correlation implies that individuals who are high-cost in one program also tend to be high-cost in the other, bundling the two schemes counteracts the intended benefits of risk pooling. Hence, separating the programs and allowing their prices to adjust independently can achieve higher total welfare without introducing redistributive losses.

6 Conclusion

This paper investigates adverse selection in social insurance using a unique setting from rural China, where land-expropriated farmers were required to enroll in basic pension and health programs but could voluntarily upgrade to higher-premium employee schemes. Using administrative data from City C (2010-2020), we document evidence of adverse selection in both pension and health insurance. We then quantify the welfare costs of adverse selection by estimating demand and cost curves, exploiting exogenous variation induced by land expropriation. Finally, we develop a random-coefficient logit model to assess whether bundling can mitigate these welfare losses. Our results show that when pension and health costs are positively correlated, offering separate vouchers for individual program, rather than a bundled discount, can achieve a Pareto improvement. Taken together, the findings highlight that social insurance design should explicitly account for cross-market interactions and the potential welfare implications of bundling or unbundling in mitigating adverse selection.

For future work, I plan to extend the choice set from two to three coverage options for each insurance. An estimated 45 million workers laid off during Chinas 1997 state-owned enterprise reform are now reaching retirement age. Under the current social insurance system, they face three pension choices, no enrollment, resident pension, or employee pension, and three parallel health insurance choices. We ask the same questions in this multi-insurance setting, but the demandcost framework becomes more complex when individuals face three options. Still, this reflects real-world environments where people often choose among multiple alternatives, making it an important and interesting extension. Second, the idea of bundling to enhance risk pooling, as well as the demandcost and choice-set framework, can also be applied to private markets, if there is suitable setting and data.

References

Adams, William James and Janet L Yellen, "Commodity bundling and the burden of monopoly," *The quarterly journal of economics*, 1976, *90* (3), 475–498.

Akerlof, George A, "The market for lemons: quality uncertainty and the market mechanism," *Market Failure or Success*, 1970, *66.*

Banerjee, Abhijit, Esther Duflo, and Richard Hornbeck, "Bundling health insurance and microfinance in India: There cannot be adverse selection if there is no demand," *American Economic Review*, 2014, *104* (5), 291–297.

Berry, S, J Levinsohn, and A Pakes, "Automobile Prices in Market Equilibriu Econometrica, 63," 1995.

Cardon, James H and Igal Hendel, "Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey," *RAND Journal of Economics*, 2001, pp. 408–427.

Chiappori, Pierre-André and Bernard Salanie, "Testing for asymmetric information in insurance markets," *Journal of political Economy*, 2000, *108* (1), 56–78.

Cutler, David M and Richard J Zeckhauser, "Adverse selection in health insurance," in "Forum for Health Economics & Policy," Vol. 1 De Gruyter 1998.

Dror, David M, Elmer S Soriano, Marilyn E Lorenzo, Jesus N Sarol Jr, Rosebelle S Azcuna, and Ruth Koren, "Field based evidence of enhanced healthcare utilization among persons insured by micro health insurance units in Philippines," *Health policy*, 2005, *73* (3), 263–271.

Einav, Liran, Amy Finkelstein, and Mark R Cullen, "Estimating welfare in insurance markets using variation in prices," *The quarterly journal of economics*, 2010, *125* (3), 877–921.

Engers, Maxim and Luis Fernandez, "Market equilibrium with hidden knowledge and self-selection," *Econometrica: Journal of the Econometric Society*, 1987, pp. 425–439.

Finkelstein, Amy and James Poterba, "Selection effects in the United Kingdom individual annuities market," *The Economic Journal*, 2002, *112* (476), 28–50.

_ and _ , "Adverse selection in insurance markets: Policyholder evidence from the UK annuity market," *Journal of political economy*, 2004, *112* (1), 183–208.

_ and _ , "Testing for asymmetric information using unused observables in insurance markets: Evidence from the UK annuity market," *Journal of Risk and Insurance*, 2014, *81* (4), 709–734.

Fischer, Torben, Markus Frölich, and Andreas Landmann, "Adverse selection in low-income health insurance markets: Evidence from an rct in Pakistan," *American Economic Journal: Applied Economics*, 2023, *15* (3), 313–340.

Handel, Ben, Igal Hendel, and Michael D Whinston, "Equilibria in health exchanges: Adverse selection versus reclassification risk," *Econometrica*, 2015, *83* (4), 1261–1313.

McCarthy, David and Olivia S Mitchell, *International adverse selection in life insurance and annuities*, Springer, 2010.

Miyazaki, Hajime, "The rat race and internal labor markets," *The Bell Journal of Economics*, 1977, pp. 394–418.

Murtaugh, Christopher M, Brenda C Spillman, and Mark J Warshawsky, "In sickness and in health: An annuity approach to financing long-term care and retirement income," *Journal of Risk and insurance*, 2001, pp. 225–253.

Nguyen, Anh, "Household bundling of health insurance," 2022.

Riley, John G, "Informational equilibrium," *Econometrica: Journal of the Econometric Society*, 1979, pp. 331–359.

Rothschild, Michael and Joseph Stiglitz, "Equilibrium in Competitive Insurance Market," *Quarterly Journal of Economics*, 1976, *90.*

Solomon, Adam, "Bundling in Insurance Markets: Theory and an Application to Long-term Care.," 2024.

Spinnewijn, Johannes, "Heterogeneity, demand for insurance, and adverse selection," *American Economic Journal: Economic Policy*, 2017, *9* (1), 308–343.

Stigler, George J, "A note on block booking," *The organization of industry*, 1968, *165.*

Stiglitz, Joseph E, "Monopoly, non-linear pricing and imperfect information: the insurance market," *The Review of Economic Studies*, 1977, *44* (3), 407–430.

Wagner, Katherine RH, "Adaptation and adverse selection in markets for natural disaster insurance," *American Economic Journal: Economic Policy*, 2022, *14* (3), 380–421.

Wang, Hong, Licheng Zhang, Winnie Yip, and William Hsiao, "Adverse selection in a voluntary Rural Mutual Health Care health insurance scheme in China," *Social science & medicine*, 2006, *63* (5), 1236–1245.

Wilson, Charles, "A model of insurance markets with incomplete information," *Journal of Economic theory*, 1977, *16* (2), 167–207.

Yao, Yi, Joan T Schmit, and Justin R Sydnor, "The role of pregnancy in micro health insurance: Evidence of adverse selection from Pakistan," *Journal of Risk and Insurance*, 2017, *84* (4), 1073–1102.

Appendix A: The Cox Model Estimation

Consider the Cox proportional hazards model:

$$h(t \mid X_i) = h_0(t) \exp(X_i' \beta), \tag{1}$$

where $h_0(t)$ is the baseline hazard function and β measures the proportional effect of covariates.

Step 1. Partial Likelihood for β. Let $t_1 < t_2 < \cdots < t_K$ denote the distinct event times, and let D_k be the set of individuals who fail at time t_k, and R_k the corresponding *risk set* (those still under observation right before t_k). The Cox partial likelihood is

$$L_p(\beta) = \prod_{k=1}^{K} \prod_{i \in D_k} \frac{\exp(X_i' \beta)}{\sum_{j \in R_k} \exp(X_j' \beta)}. \tag{2}$$

Maximizing $\log L_p(\beta)$ yields the estimator $\hat{\beta}$.

Step 2. Breslow Estimator of the Baseline Hazard. Given $\hat{\beta}$, the baseline hazard can be estimated nonparametrically. At each distinct event time t_k, the estimated baseline hazard is

$$\widehat{h}_0(t_k) = \frac{d_k}{\displaystyle\sum_{j \in R_k} \exp(X_j' \hat{\beta})}, \tag{3}$$

where $d_k = |D_k|$ is the number of failures at time t_k. This expression is known as the Breslow (1972) estimator.

Step 3. Cumulative Baseline Hazard and Survival. The estimated cumulative baseline hazard function is

$$\widehat{H}_0(t) = \sum_{t_k \le t} \frac{d_k}{\displaystyle\sum_{j \in R_k} \exp(X_j' \hat{\beta})}. \tag{4}$$

The corresponding baseline survival function is then

$$\widehat{S}_0(t) = \exp[-\widehat{H}_0(t)]. \tag{5}$$

For any individual with covariates X_i, the estimated cumulative hazard and survival functions are

$$\widehat{H}_i(t) = \widehat{H}_{0,s(i)}(t) \exp(X_i^\top \widehat{\beta}), \qquad \widehat{S}_i(t) = \exp\big(-\widehat{H}_i(t)\big). \tag{6}$$

We report expected remaining lifetime as

$$\widehat{LE}_i \;=\; \int_0^{T_{\max}} \widehat{S}_i(t)\,dt, \tag{7}$$

numerically integrating up to a large terminal age (e.g., $T_{\max} = 100$).

Appendix B: Welfare Analysis under Logit Demand

Setup. For $k \in \{p, h\}$, demand is logit:

$$\log\left(\frac{D^p}{1-D^p}\right) = \alpha^p + \beta_p^p p^p + \beta_h^p p^h, \quad \log\left(\frac{D^h}{1-D^h}\right) = \alpha^h + \beta_p^h p^p + \beta_h^h p^h,$$

average cost is linear in prices:

$$c^p = \gamma^p + \delta_p^p p^p + \delta_h^p p^h, \qquad c^h = \gamma^h + \delta_p^h p^p + \delta_h^h p^h,$$

and marginal cost satisfies (using the given formulae)

$$MC^p = c^p + \frac{\delta_p^p}{\beta_p^p(1-D^p)}, \qquad MC^h = c^h + \frac{\delta_h^h}{\beta_h^h(1-D^h)}.$$

Market equilibrium (AC pricing). Prices solve $p^p = c^p,\ p^h = c^h$, yielding the closed forms

$$p_{\text{eq}}^p = \frac{\gamma^p(1-\delta_h^h) + \delta_h^p \gamma^h}{(1-\delta_p^p)(1-\delta_h^h) - \delta_h^p \delta_p^h},$$

$$p_{\text{eq}}^h = \frac{\gamma^h(1-\delta_p^p) + \delta_p^h \gamma^p}{(1-\delta_p^p)(1-\delta_h^h) - \delta_h^p \delta_p^h}.$$

Equilibrium demands then are

$$D_{\text{eq}}^p = \frac{\exp(\alpha^p + \beta_p^p p_{\text{eq}}^p + \beta_h^p p_{\text{eq}}^h)}{1+\exp(\alpha^p + \beta_p^p p_{\text{eq}}^p + \beta_h^p p_{\text{eq}}^h)}, \quad D_{\text{eq}}^h = \frac{\exp(\alpha^h + \beta_p^h p_{\text{eq}}^p + \beta_h^h p_{\text{eq}}^h)}{1+\exp(\alpha^h + \beta_p^h p_{\text{eq}}^p + \beta_h^h p_{\text{eq}}^h)}.$$

Efficient allocation (MC pricing). Efficiency requires $p^k = MC^k$ for $k \in \{p, h\}$, which implies the fixed-point system

$$p_{\text{eff}}^p = \gamma^p + \delta_p^p p_{\text{eff}}^p + \delta_h^p p_{\text{eff}}^h + \frac{\delta_p^p}{\beta_p^p\big(1-D_{\text{eff}}^p\big)},$$

$$p_{\text{eff}}^h = \gamma^h + \delta_p^h p_{\text{eff}}^p + \delta_h^h p_{\text{eff}}^h + \frac{\delta_h^h}{\beta_h^h\big(1-D_{\text{eff}}^h\big)},$$

with

$$D_{\text{eff}}^p = \frac{\exp(\alpha^p + \beta_p^p p_{\text{eff}}^p + \beta_h^p p_{\text{eff}}^h)}{1+\exp(\alpha^p + \beta_p^p p_{\text{eff}}^p + \beta_h^p p_{\text{eff}}^h)},$$

$$D_{\text{eff}}^h = \frac{\exp(\alpha^h + \beta_p^h p_{\text{eff}}^p + \beta_h^h p_{\text{eff}}^h)}{1+\exp(\alpha^h + \beta_p^h p_{\text{eff}}^p + \beta_h^h p_{\text{eff}}^h)}.$$

This system generally has no closed form and is solved numerically.

Deadweight loss (definitions). In general, the exact DWL equals the area between inverse demand and marginal cost as we move from the market equilibrium to the efficient allocation. A convenient first-order approximation uses local linearization.

(i) Component-wise Harberger triangle (holding the other price fixed at p_{eq}^{-k}): for $k \in \{p, h\}$,

$$\mathrm{DWL}^k \;\approx\; \frac{1}{2}\,(P^k_{\mathrm{eq}} - MC^k_{\mathrm{eq}})^2\Big(-\frac{\partial D^k}{\partial p^k}\Big)\Big|_{\mathrm{eq}}.$$

Under logit, $\dfrac{\partial D^k}{\partial p^k} = \beta^k_k D^k(1-D^k)$ and $P^k_{\mathrm{eq}} - MC^k_{\mathrm{eq}} = -\dfrac{\delta^k_k}{\beta^k_k(1-D^k_{\mathrm{eq}})}$, so

$$\boxed{\mathrm{DWL}^p \approx \frac{(\delta^p_p)^2}{2\,(-\beta^p_p)}\cdot\frac{D^p_{\mathrm{eq}}}{1-D^p_{\mathrm{eq}}}} \qquad \boxed{\mathrm{DWL}^h \approx \frac{(\delta^h_h)^2}{2\,(-\beta^h_h)}\cdot\frac{D^h_{\mathrm{eq}}}{1-D^h_{\mathrm{eq}}}}$$

and $\mathrm{DWL}_{\mathrm{total}} \approx \mathrm{DWL}^p + \mathrm{DWL}^h$.

(ii) Joint (matrix) Harberger approximation (letting both prices adjust): Let $p = (p^p, p^h)^\top$, $D = (D^p, D^h)^\top$, and J be the Jacobian

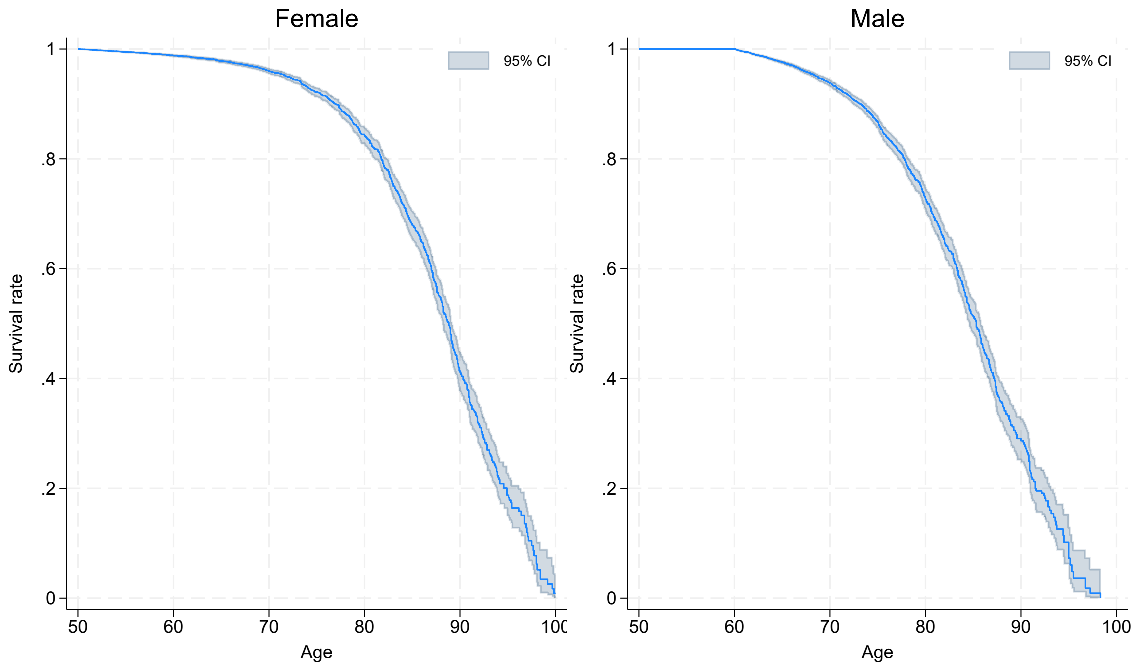
$$J \;=\; \frac{\partial D}{\partial p} \;=\; \begin{bmatrix} \beta^p_p D^p(1-D^p) & \beta^p_h D^p(1-D^p) \\ \beta^h_p D^h(1-D^h) & \beta^h_h D^h(1-D^h) \end{bmatrix}\Bigg|_{\mathrm{eq}}.$$

Define the wedge vector at equilibrium $w := p_{\mathrm{eq}} - MC_{\mathrm{eq}}$ with $MC_{\mathrm{eq}} = (MC^p_{\mathrm{eq}}, MC^h_{\mathrm{eq}})^\top$. A multivariate Harberger approximation is

$$\boxed{\mathrm{DWL}_{\mathrm{joint}} \;\approx\; \tfrac{1}{2}\, w^\top\,(-J)\,w}$$

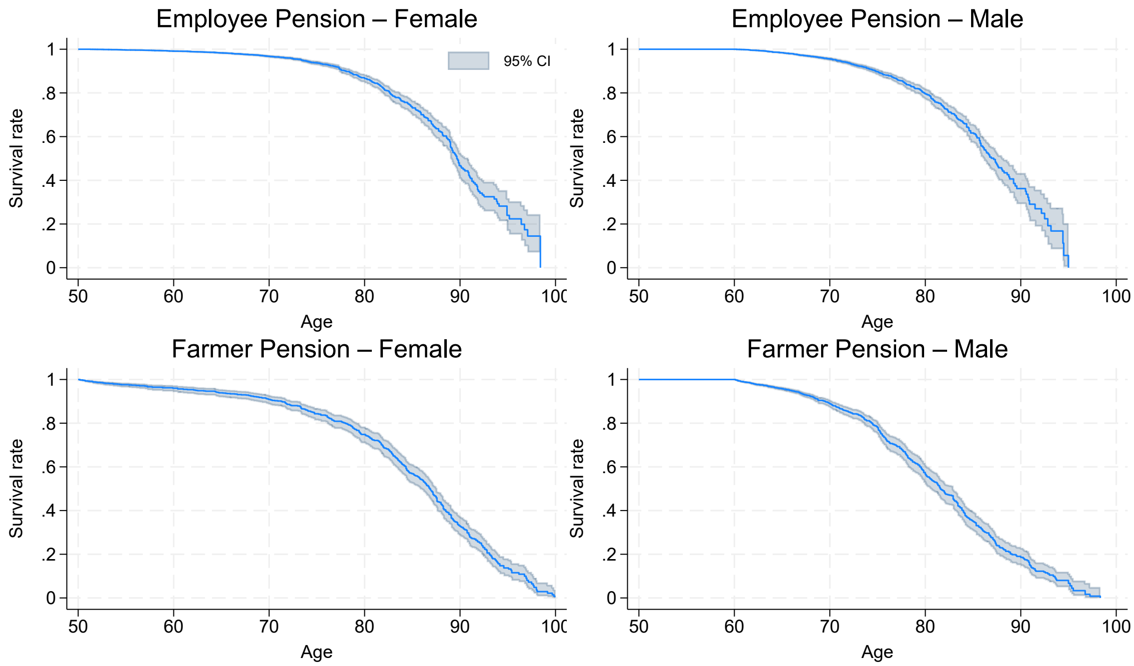
which internalizes cross-price effects via J.

Appendix Figure A1. KaplanMeier Survival Curves by Gender



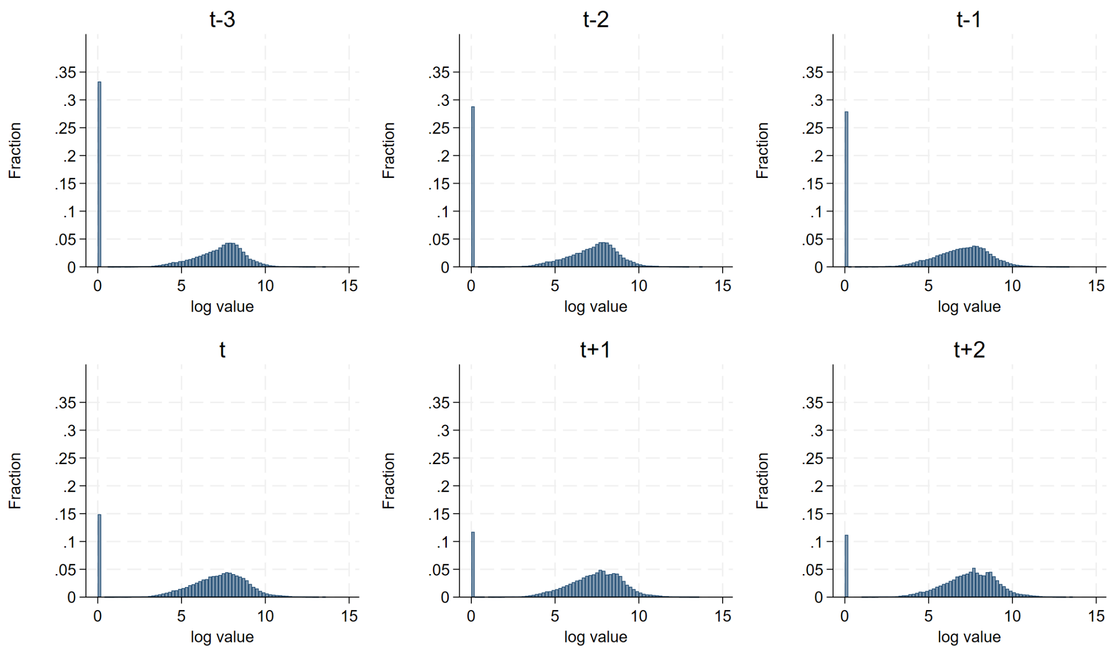
Each panel plots the KaplanMeier survival function $\widehat{S}(t)$ by gender, where t denotes age at exit (in years). The curves are estimated nonparametrically using observed failure times (death) and censoring, based on individuals aged 50 and above. The shaded bands represent the pointwise 95% confidence intervals constructed from Greenwoods variance estimator. The survival probability for females declines more gradually with age, implying longer life expectancy, while male survival falls faster, indicating higher mortality risk across all ages.

Appendix Figure A2. KaplanMeier Survival Curves by Gender and Pension Type



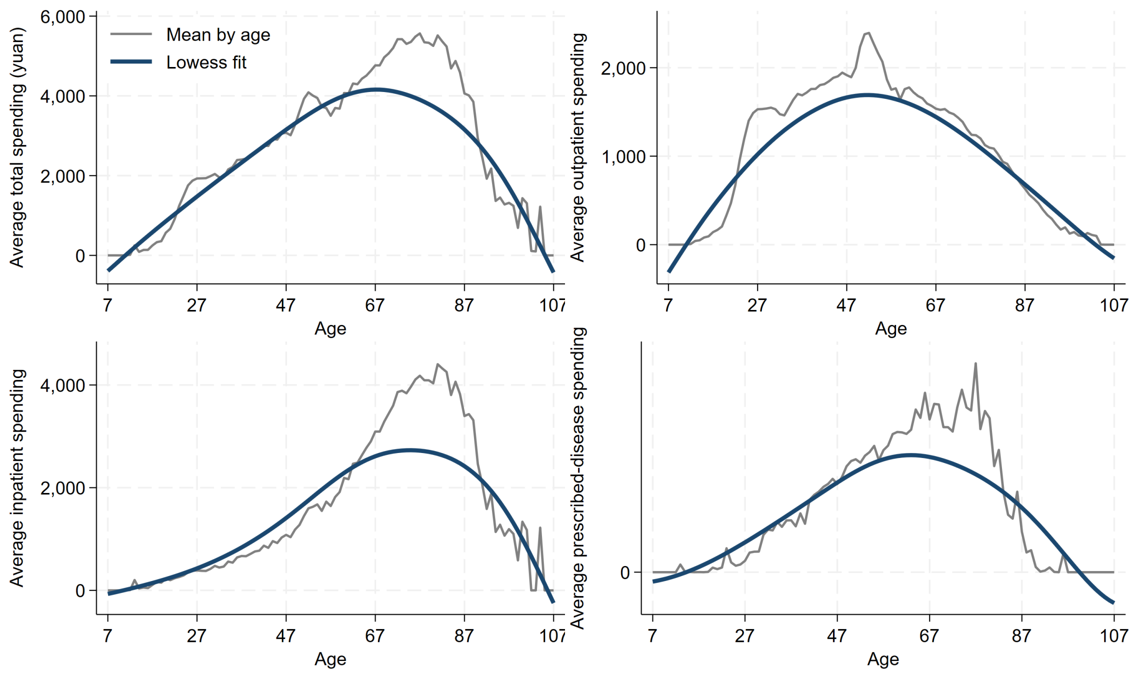
Each panel presents the KaplanMeier survival function $\widehat{S}(t)$ by age (t) for individuals aged 50 and above, estimated separately by pension type (Employee vs. Farmer) and gender (Female vs. Male). The curves are estimated nonparametrically using observed failure times (death) and right-censored observations. The shaded areas represent the pointwise 95% confidence intervals based on Greenwoods variance estimator. The survival probabilities for females are consistently higher and decline more gradually with age, indicating longer life expectancy, while male survival curves drop more steeply. Comparing across pension types, participants in the employee pension scheme exhibit higher survival rates than those in the farmer pension, reflecting differences in baseline health, income, and access to healthcare.

Appendix Figure A3. Comparison of Total Medical Expenses Pre- and Post-Retirement



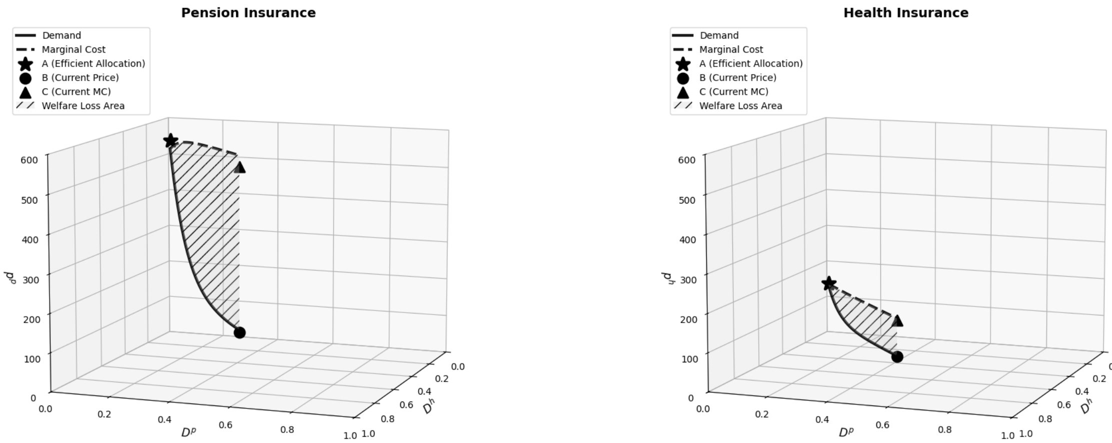
Each panel shows the distribution of total medical expenses (log value) at different years relative to retirement, where $t = 0$ denotes the retirement year. The sample includes individuals continuously observed in all six periods ($t - 3$ to $t + 2$). Each histogram is normalized to represent the fraction of individuals within each log-value bin. The spike near zero corresponds to individuals with no medical expenses in that year.

Appendix Figure A4. Average Medical Spending by Age



Each panel plots the mean and lowess-smoothed average spending by age. Panels (a)–(d) correspond to total, outpatient, inpatient, and prescribed-disease spending, respectively. The total spending in panel (a) equals the sum of the three components shown in panels (b)–(d).

Appendix Figure A5. Efficiency Cost under No Bundling (Baseline Estimates)



This figure compares the estimated demand and marginal cost curves for pension and health insurance. When the instrumental variable (IV) is not used, the estimated demand curve is biased downward because price endogeneity causes healthier individuals to appear less responsive to price changes. As a result, the implied area between the demand and marginal cost curves, representing welfare loss at current price level comparing to the efficient allocation, is underestimated. Using the IV-corrected estimates therefore yields a more accurate and typically larger welfare loss, reflecting the true efficiency cost of selection.

Appendix Table A1. Employee Pension Take-up and Medical Cost

	Pre-2020		2020	
Outpatient spending (K yuan)	0.213***	0.043***	-0.059*	-0.063*
	(0.010)	(0.008)	(0.032)	(0.033)
Inpatient spending (K yuan)	-0.006***	-0.002***	-0.003*	-0.004**
	(0.001)	(0.001)	(0.002)	(0.002)
Prescribed disease spending (K yuan)	-0.018*	-0.009***	0.005	0.004
	(0.009)	(0.003)	(0.016)	(0.015)
Outpatient visits	-0.010***	-0.002***	0.002	0.003
	(0.001)	(0.001)	(0.005)	(0.005)
Inpatient visits	-0.067***	-0.014	0.050*	0.074**
	(0.016)	(0.010)	(0.029)	(0.035)
Prescribed disease visits	0.015***	0.003*	-0.002	0.002
	(0.004)	(0.002)	(0.011)	(0.010)
Premium		-0.010***		-0.009***
		(0.000)		(0.001)
Male		-0.139***		0.046*
		(0.006)		(0.026)
Retire-year FE	No	Yes		
Observations	15,092		1,791	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Before 2020, higher outpatient spending is positively associated with pension upgrading, suggesting that individuals who are more health-conscious or proactive in healthcare are more likely to transfer. In contrast, higher inpatient or prescribed-disease spending reduces the probability of transfer, indicating that less healthy individuals are less inclined to upgradeconsistent with adverse selection based on health risk. After 2020, the coefficients become smaller and weaker in significance, reflecting diminished selection once the majority of eligible farmers had already upgraded.

Appendix Table A2. Employee Pension Take-up and Predicted Total Cost

	Pre-2020		2020	
Total Cost (K yuan)	0.001***	0.000***	-0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.001)
Premium		-0.012***		-0.014***
		(0.000)		(0.001)
Male		-0.125***		0.109
		(0.011)		(0.069)
Retire-year FE	No	Yes		
Observations	14,909		1,791	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Predicted total cost combines expected life expectancy and projected medical expenses to approximate each individuals long-term financial burden. Before 2020, individuals with higher predicted total costs are significantly more likely to upgrade, consistent with adverse selection based on anticipated longevity and utilization. The negative premium coefficients suggest that affordability constraints remain important, while gender differences persist, with women more likely to upgrade. By 2020, these effects become statistically insignificant, indicating that selection dynamics had largely stabilized once participation rates plateaued.

Appendix Table A3. Employee Health Take-up and Medical Cost

	Pre-2020		2020	
Outpatient spending (K yuan)	0.027***	0.157***	-0.010	-0.006
	(0.005)	(0.006)	(0.018)	(0.013)
Inpatient spending (K yuan)	-0.003**	-0.003*	-0.001	-0.002
	(0.001)	(0.001)	(0.003)	(0.003)
Prescribed disease spending (K yuan)	-0.003	-0.007	0.018	0.018
	(0.007)	(0.013)	(0.016)	(0.016)
Outpatient visits	0.001*	-0.005***	0.001	0.001
	(0.001)	(0.001)	(0.003)	(0.002)
Inpatient visits	0.031*	-0.041**	-0.003	0.003
	(0.018)	(0.018)	(0.049)	(0.048)
Prescribed disease visits	-0.002	0.006	-0.005	-0.005
	(0.004)	(0.005)	(0.011)	(0.010)
Premium (health)		0.024***		-0.025***
		(0.000)		(0.004)
Male		0.289***		0.059**
		(0.007)		(0.024)
Retire-year FE	No	Yes		
Observations	11,463		1,791	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.
Before 2020, when enrollment was semi-bundled with the employee pension, individuals with higher outpatient spending are significantly more likely to transfer to employee health insurance, suggesting selection on health awareness or expected benefit utilization. In contrast, higher inpatient spending is associated with lower take-up, indicating that less healthy individuals are less inclined to upgradeconsistent with adverse selection. After 2020, these associations become small and statistically insignificant, implying that selection weakened once the bundling constraint was removed and participation stabilized. Premium remains an important determinant: the positive sign before 2020 reflects policy bundling incentives, whereas the negative coefficient in 2020 reflects standard price sensitivity under voluntary participation.

Appendix Table A4. Employee Health Take-up and Predicted Total Cost

	Pre-2020		2020	
Total Cost (K yuan)	-0.007***	0.009***	-0.002*	-0.000
	(0.000)	(0.002)	(0.001)	(0.001)
Premium		-0.019***		-0.025***
		(0.000)		(0.004)
Male		-0.461***		0.047
		(0.027)		(0.043)
Retire-year FE	No	Yes		
Observations	11,451		1,791	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Predicted total cost combines expected medical expenditures and life expectancy to proxy for individuals long-term expected health burden. Before 2020, total cost is negatively associated with employee health take-up, implying that healthier or longer-living individuals were more likely to upgrade, consistent with adverse selection. After controlling for retire-year fixed effects, however, the coefficient becomes positive but small, reflecting the influence of policy design and semi-bundling incentives rather than pure risk selection. In 2020, these effects vanish, and the negative premium coefficient again highlights affordability as the primary determinant of participation once the system transitioned to fully voluntary enrollment.

Appendix Table A5. Income Regression Result (Pre 2020, Worker)

	Pension		Health	
Outpatient spending (K yuan)	0.027***	0.157***	-0.010	-0.006
	(0.005)	(0.006)	(0.018)	(0.013)
Inpatient spending (K yuan)	-0.003**	-0.003*	-0.001	-0.002
	(0.001)	(0.001)	(0.003)	(0.003)
Prescribed disease spending (K yuan)	-0.003	-0.007	0.018	0.018
	(0.007)	(0.013)	(0.016)	(0.016)
Outpatient visits	0.001*	-0.005***	0.001	0.001
	(0.001)	(0.001)	(0.003)	(0.002)
Inpatient visits	0.031*	-0.041**	-0.003	0.003
	(0.018)	(0.018)	(0.049)	(0.048)
Prescribed disease visits	-0.002	0.006	-0.005	-0.005
	(0.004)	(0.005)	(0.011)	(0.010)
Income Level	1.278*	0.961	-1.360	0.514
	(0.696)	(0.625)	(1.279)	(1.109)
Premium		-0.002***		0.022***
		(0.000)		(0.000)
Male		-0.106*		0.569***
		(0.056)		(0.022)
Retire-year FE	No	Yes	No	Yes
Observations	3,200		3,085	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.
This table reports results for workers with income data derived from pension contribution records. Controlling for income level does not materially change the coefficients on medical spending variables, suggesting that income differences do not explain the observed selection patterns. Higher income is weakly associated with greater likelihood of upgrading to the employee pension, but not significantly related to employee health take-up. Premium remains a strong predictor in both regressions: higher pension premiums reduce take-up, while higher health premiums increase take-up prior to 2020 due to policy bundling effects. Overall, the presence or absence of the income control does not alter the direction or significance of key coefficients, indicating that adverse selection is not primarily driven by income differences among workers.

Appendix Table A6. Income Regression Result (Pre 2020, Worker)

	Pension		Health	
Total Cost (K yuan)	-0.007***	0.009***	-0.002*	-0.000
	(0.000)	(0.002)	(0.001)	(0.001)
Income Level	1.278*	0.961	-1.360	0.514
	(0.696)	(0.625)	(1.279)	(1.109)
Premium		-0.002***		0.022***
		(0.000)		(0.000)
Male		-0.106*		0.569***
		(0.056)		(0.022)
Retire-year FE	No	Yes	No	Yes
Observations	3,200		3,085	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.
This table examines whether income differences explain pension and health take-up among workers before 2020, using total expected lifetime cost as the main explanatory variable. For pension participation, total cost has a small but significant positive coefficient once retire-year fixed effects are included, while for health insurance, the effect is weak and statistically insignificant. Adding income level as a control does not substantially alter the coefficients on total cost or premium, indicating that income heterogeneity among workers does not drive the observed selection. Higher income slightly increases the probability of pension upgrading but shows no meaningful relationship with health insurance take-up. Premium remains a key factornegative for pension and positive for health due to the semi-bundled policy incentivesconfirming that affordability and policy design, rather than income per se, primarily shape workers enrollment decisions.

Appendix Table A7. Income Regression Result (Pre 2020, Non-Worker)

	Pension		Health	
Outpatient spending (K yuan)	0.027***	0.157***	-0.010	-0.006
	(0.005)	(0.006)	(0.018)	(0.013)
Inpatient spending (K yuan)	-0.003**	-0.003*	-0.001	-0.002
	(0.001)	(0.001)	(0.003)	(0.003)
Prescribed disease spending (K yuan)	-0.003	-0.007	0.018	0.018
	(0.007)	(0.013)	(0.016)	(0.016)
Outpatient visits	0.001*	-0.005***	0.001	0.001
	(0.001)	(0.001)	(0.003)	(0.002)
Inpatient visits	0.031*	-0.041**	-0.003	0.003
	(0.018)	(0.018)	(0.049)	(0.048)
Prescribed disease visits	-0.002	0.006	-0.005	-0.005
	(0.004)	(0.005)	(0.011)	(0.010)
Income Level	1.278*	0.961	-1.360	0.514
	(0.696)	(0.625)	(1.279)	(1.109)
Premium		-0.002***		0.022***
		(0.000)		(0.000)
Male		-0.106*		0.569***
		(0.056)		(0.022)
Retire-year FE	No	Yes	No	Yes
Observations	3,200		3,085	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.
This table reports the results for non-workers, who lack employment-based income records and must self-finance their contributions. For this group, outpatient spending is positively associated with pension take-up, while higher inpatient spending significantly reduces the likelihood of upgrading, indicating that healthier individuals are more likely to select into the employee pension-consistent with adverse selection. In the health insurance regressions, most coefficients are small and statistically insignificant, suggesting weaker selection once medical conditions are directly controlled. Including the income proxy does not substantially alter these results, confirming that selection patterns are not primarily driven by income differences. As in previous specifications, higher premiums discourage pension enrollment but appear positively related to health take-up due to semi-bundled policy incentives before 2020. Overall, adverse selection is more pronounced among non-workers, particularly in the pension program, where healthier and more forward-looking individuals are more likely to upgrade.

Appendix Table A8. Income Regression Result (Pre 2020, Non-Worker)

	Pension		Health	
Total Cost (K yuan)	-0.007***	0.009***	-0.002*	-0.000
	(0.000)	(0.002)	(0.001)	(0.001)
Income Level	1.278*	0.961	-1.360	0.514
	(0.696)	(0.625)	(1.279)	(1.109)
Premium		-0.002***		0.022***
		(0.000)		(0.000)
Male		-0.106*		0.569***
		(0.056)		(0.022)
Retire-year FE	No	Yes	No	Yes
Observations	3,200		3,085	

Average marginal effects reported.
Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.
This table presents regression results for non-working individuals, who do not have employer-based contributions and must pay the full premium to upgrade. For the pension program, total cost is positively associated with take-up once retire-year fixed effects are included, implying that individuals with higher predicted lifetime costscorresponding to longer life expectancyare more likely to transfer to the employee pension, consistent with adverse selection. In contrast, for health insurance, total cost has a small and statistically insignificant effect, suggesting limited selection once medical risk is already reflected in the cost measure. Income level is not a significant predictor in either regression, indicating that affordability constraints play a limited role for this group. Premium continues to be a strong determinant: higher pension premiums reduce enrollment, while higher health premiums increase take-up due to the semi-bundled design prior to 2020. Overall, these results confirm that adverse selection is most pronounced among non-workers in the pension program, driven by differences in expected longevity rather than income.