Product Innovation and Demand Matching in China's Retirement Finance: Evidence Across Income Groups

Haoran Zhang
Shanghai University Of Finance And Economic, Shanghai, China

Abstract: China's rapid population aging heightens the need for well-functioning second- and third-pillar arrangements, yet frictions persist between product supply and stratified household demand. We take a demand-side perspective and examine whether-and through which channels-households at different income levels participate in retirement-related financial products (supplementary pension/annuity, commercial health insurance, housing provident fund, and marketable securities). Using micro data from the 2019 China Household Finance Survey (CHFS) and the 2020 China Health and Retirement Longitudinal Study (CHARLS), we estimate baseline logit/probit models with province fixed effects, province-clustered standard errors, and survey p-weights, and implement structural equation modeling (SEM) to assess mediation by financial literacy and risk attitudes; heterogeneity and robustness are evaluated by subgroup interactions and alternative link/measurement choices. The effective household sample is N = 33,835. Participation exhibits a clear post-control income gradient: relative to the lowest-income tercile (T1), T3–T1 average marginal effects (pp) are 3.7 for supplementary pensions, 6.3 for health insurance, 22.7 for the housing provident fund, and 3.8 for marketable securities; baseline T1 participation probabilities (%) are 0.1, 2.3, 2.2, and 0.6, respectively. Model fit is adequate to strong (AUCs 0.768–0.870 across outcomes). We find steeper income–participation slopes in urban areas: the T3–T1 contrast differs by +3.29 pp for securities and +10.91 pp for the housing provident fund. Mediation analysis indicates positive but statistically non-significant indirect effects via literacy and risk: for securities, the shares of the total income effect explained by literacy and risk are 0.1% (95% CI [−7.6, 7.8]) and 6.2% (95% CI [−1.9, 14.4]), respectively. Findings are robust to alternative income definitions and ranking schemes (logs/asinh, winsorization, weighted vs. unweighted tiles) and to probit, complementary log–log, and rare-events corrections under the same fixed-effects/weighting discipline. Building on the evidence, we outline a tiered product-and-policy toolkit: affordability and simple defaults for lower-income households, reliable accumulation and accessibility for middle-income households, and diversified, customizable portfolios for higher-income households.

Keywords: Retirement Finance; Product Innovation; Demand–Supply Matching; Income Heterogeneity; Financial Literacy; Structural Equation Modeling; Risk Attitudes; CHFS/CHARLS Microdata; China; Housing Provident Fund

1. Introduction

1.1 Problem Statement

China's population is aging rapidly, while the breadth and depth of second- and third- pillar arrangements have yet to fully match households' stratified retirement needs. On the supply side, financial institutions primarily offer standardized products across insurance, mutual funds, and bank wealth management. On the demand side, households differ markedly in income, financial literacy, risk attitudes, and liquidity constraints, forming a ladder of needs that ranges from basic protection to steady accumulation and customized allocation. Asynchrony between supply and demand, together with informational and behavioral frictions, leads to recognizable mismatches: products exist but remain out of reach; willingness exists but affordability is limited; intent to allocate exists but tools are missing. It is therefore necessary to revisit the design boundary and transmission channels of retirement finance from the demand side: Which products resonate with which income groups?

Through what behavioral channels are income differences translated into actual participation? What product–service bundles raise the degree of product–demand fit given constraints on affordability, comprehensibility, and sustainability?

1.2 Academic and Practical Relevance

Academic: Prior work often focuses on overall participation or single-product determinants, leaving the link from demand heterogeneity to product innovation and empirical identification only partially connected. This paper develops a mechanism in which household income affects participation through financial literacy and risk attitudes, implements reproducible identification on micro data, and maps heterogeneity and supply attributes to measurable fit.

Practical: For institutions and regulators, the evidence translates into three lines of action: income- and behavior-aware product and entitlement design (thresholds, fees, defaults, exit and protection clauses); distribution and investor-education paths optimized for accessibility, intelligibility, and sustainability; and differentiated incentives and inclusion policies across regions and demographic segments to improve the coverage and stickiness of the second and third pillars.

1.3 Research Questions and Contributions

1.3.1 Research questions

RQ1 (Demand heterogeneity): Do participation rates in supplementary pensions, commercial health insurance, and market-based assets differ significantly across income groups?

RQ2 (Fit): After accounting for demographics and regional differences, what is the extent of product–demand matching under current supply?

RQ3 (Mechanism): To what extent do financial literacy and risk attitudes mediate the income–participation relationship, and how large are the indirect contributions relative to the total effect?

RQ4 (Policy & product): Conditional on the evidence, what tiered product and policy toolkit is most effective for low-, middle-, and high-income households?

Contributions.

Evidence: document participation gaps and segment profiles under income heterogeneity.

Methods: identify the income–behavior–participation chain via logit/probit and SEM, with province fixed effects and sampling weights; quantify direct and indirect effects with bootstrap intervals and interpret mediation cautiously.

Application: propose product and policy bundles targeted at accessibility, intelligibility, and sustainability, balancing defaults with customization for implementable guidance.

1.4 Roadmap and Structure

Problem definition → Literature review → Theory and hypotheses → Data and variables → Identification strategy → Results → Implications and design. As shown in Figure 1, the study proceeds from problem definition to results, then to implications and product design. Accordingly, Section 2 **reviews** the literature and clarifies the gap; Section 3 proposes the conceptual framework and hypotheses; Section 4 introduces data, variables, and descriptive facts; Section 5 details identification; Section 6 reports results; Section 7 provides product and policy implications; Section 8 concludes.

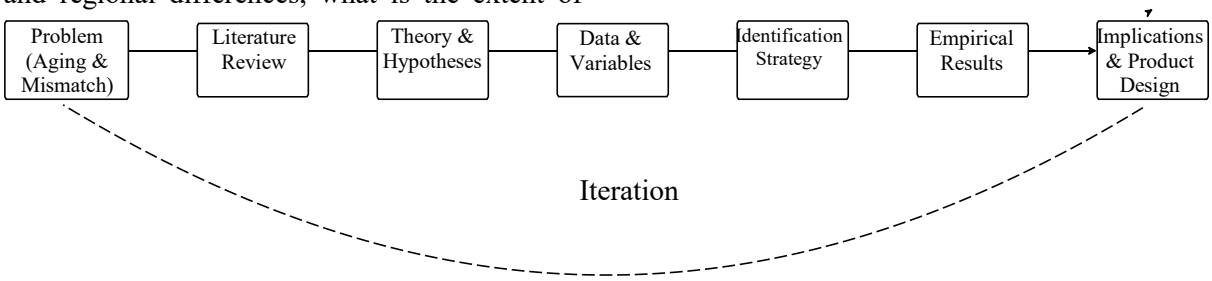


Figure 1. Research Roadmap and Paper Structure

2. Literature Review

2.1 Concepts and Institutional Framework

Retirement finance comprises arrangements and services that safeguard well-being in old age across three phases-*accumulation*, *risk management*, and *payout*. A large strand of Chinese scholarship frames the core task as channeling funded assets in the second and third pillars into capital markets under sound governance, so as to preserve and grow pension wealth while supplying long-term risk capital to the real economy[1]. Building on the OECD

three-pillar architecture, China still exhibits a marked structural imbalance: the first pillar remains dominant while the market-driven pillars lag behind, constraining replacement rates and long-run sustainability[2], with market/institutional frictions documented in micro evidence[3]. Recent sector discussions emphasize personalization and omni-channel delivery within a multi-pillar framework[4]. Official statistics and policy summaries show rapid coverage expansion but persistent imbalance and pressure on the basic pension, motivating deeper reform[5]; international practice offers useful templates for pillar deepening[6]. This imbalance is visible in Figure 2, where Pillar I remains dominant relative to Pillars II and III. **Table 1** summarizes representative studies and clarifies how our setting and methods connect to the literature.

2.2 Product and Service Innovation

On the supply side, innovations cluster along three lines:

Bank-led ecosystems. Major banks have launched senior-finance brands, pension wealth products, and integrated service bundles[4].

Insurance-led integration. Commercial pension insurance, annuities/occupational annuities, and long-term care insurance have expanded under policy support; tax-deferred individual accounts, pension wealth-management products, and dedicated pension annuities form a serialized menu[5,7].

Trusts and funds. Pension trusts and target-date/target-risk funds extend long-horizon vehicles within the third pillar and complement employer-based plans[5].

A continuing theme is FinTech enablement and institution-building-tax incentives, product diversification, multi-agent participation, and risk safeguards-to achieve healthy interaction between product supply and household demand[5,7].

2.3 Demand Heterogeneity

Demand varies with income, financial literacy, and risk attitudes, interacting with urban–rural residence, education, and health constraints. Practice-oriented studies emphasize modular design, simplified defaults, and omni-channel delivery tailored to older consumers[3].

2.4 Methodological Trajectory

The literature ranges from descriptive policy analysis to econometric identification with fixed effects and survey weights, mediation-oriented SEM/GSEM, and-where data permit- machine-learning–assisted prediction and segmentation.

2.5 Gaps and Positioning

A systematic, quantitative assessment of product – demand fit under an explicit "income segmentation + mediation" framework is scarce. We estimate the income → (literacy/risk)→participation pathway on CHFS/CHARLS, map heterogeneity, and translate the evidence into a tiered supply toolkit.

Table 1. Representative Studies on Retirement Finance: Settings, Methods, Data, Findings and Links to this Paper

Author (Year)	Setting /Population	Data (years, N)	Method	Key Findings
Hu (2013)	OECD+China	OECD facts; Chinese policies	Conceptual	Funded pillars with strong governance, supply long-term capital.
Sun (2016)	China	Sector evidence	Review	First-pillar dominance; need stronger market linkage.
Zhou (2022)	China	Survey *N*=373	Factor analysis	Four barrier factors: market / institutional, most binding.
Wang (2025)	China	Practice evidence	Thematic	Personalization, omni-channel delivery, safeguards.
Cheng et al. (2024)	Japan/US	Policy/market records	Comparative	GPIF allocation, revised NISA, advisor role
Li et al. (2025)	China	National statistics	Analytical	Coverage, balances, deficits; tech-enabled pension finance.
Wu (2024)	China	Commentary	Policy	Tax incentives, multi-agent participation, tech platforms.

3. Theoretical Framework and Hypotheses

3.1 Theoretical Foundations

Two strands motivate our framework. Consumer financial behavior links life-cycle optimization with bounded rationality: financial literacy raises the intelligibility and implementability of choices, while risk attitudes govern willingness to accept uncertainty and volatility. Urban

residence and education shape information access and opportunity sets. A financial needs hierarchy characterizes retirement-related demand in three tiers:

protection (supplementary pensions; commercial health/long-term care insurance);

steady accumulation (annuities, target-date/lifecycle, pension wealth-management products);

customization/market exposure (mutual funds, equity/bond allocation, advisory bundles).

As income rises, budget constraints relax, pushing households from protection toward accumulation and customization. This upgrade typically operates through financial literacy and risk attitudes.

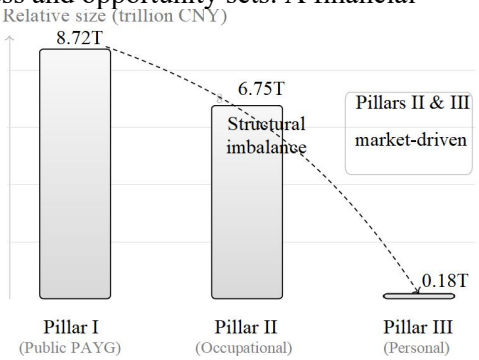


Figure 2. Sizes of the Three Pillars (Trillion CNY)

Notes: nominal, year-end 2023; official sources (MOHURD National Housing Provident Fund Annual Report 2023; MOHRSS Statistical Bulletin 2023; NBS China Statistical Yearbook 2023). See References[8–10].

3.2 Mechanism and Identification

We posit Income → (Financial Literacy / Risk Attitudes) → Participation. Figure 3 illustrates the conceptual pathway and the SEM structure used for mediation analysis. Let FinLit and RiskAtt be standardized indices; define participation as binary indicators for protection, accumulation, and market products. With controls X (age, gender, education, household size, urban, etc.) and province fixed effects λ_{prov}, the empirical system is:

$$\text{FinLit}\&=a_1\,\text{Income}+\delta_1^{\top}X+\lambda_{prov}+\varepsilon_1 \quad (1)$$

$$\text{RiskAtt}\&=a_2\,\text{Income}+\delta_2^{\top}X+\lambda_{prov}+\varepsilon_2 \quad (2)$$

$$\Pr(\text{Participation}=1)\&=F\left(c'\text{Income}+b_1\text{FinLit}+b_2\text{RiskAtt}+\gamma^{\top}X+\lambda_{prov}\right) \quad (3)$$

We expect a_1, $a_2 > 0$ and b_1, $b_2 > 0$. Mediation is assessed via a_1b_1 and a_2b_2 with bootstrap confidence intervals; the indirect-effect share is $(a_1b_1 + a_2b_2)$/(Total effect).

Identification and inference. All regressions use household survey p-weights, province fixed effects, and province-clustered robust standard errors; all margins and predicted probabilities adopt the same weighting and FE specification. Heterogeneity is evaluated by subgroup estimation and interactions (urban/rural, education, age). Robustness spans alternative income measures (log or asinh; winsorized variants), weighted/unweighted tiles, and alternative link functions (logit, probit, complementary log–log, and rare-events/Firth corrections).[1]

3.3 Hypotheses

Income–participation gradient. Household income exhibits a *monotonic* positive gradient in the probability of participating in supplementary pension/annuity, commercial health insurance, housing provident fund, and marketable securities, when income is ranked by terciles/quintiles (weighted and unweighted).

Mediation by literacy and risk. The effect of income on participation operates *indirectly* through higher financial literacy and greater risk tolerance; the total indirect effect and its share are expected to be positive; we assess significance using SEM/GSEM with bootstrap confidence intervals.

Heterogeneity. The income → (FinLit, RiskAtt) → participation pathway is *stronger* for urban and better-educated households and for the "younger-old" (pre-/early-retirement), relative to rural, low-education, and older cohorts. We summarize heterogeneity using tercile-slope AME differences (e.g., T3–T1) and interaction coefficients.

Robustness. Results hold under alternative income measures and ranking schemes (log or asinh; winsorized), with and without survey weights, and across logit/probit/cloglog/rare-events specifications with province fixed effects and clustered SE.

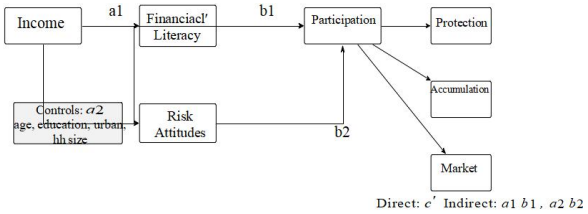


Figure 3. Conceptual Model and SEM Path Diagram

Variable-to-name mapping (e.g., dataset field

names for outcomes and mediators) is provided in Appendix A.

4. Data, Variables, and Measurement

4.1 Data Sources and Samples

The primary dataset is the 2019 China Household Finance Survey (CHFS) at the household level[11]. A harmonized household-level analytic file is used for modeling and descriptive statistics. For robustness and complementary insurance/health information, we construct a parallel 2020 China Health and Retirement Longitudinal Study (CHARLS) individual-level analytic file with binary indicators for social pension and medical insurance and corresponding survey weights[12]. When available, CHFS household p-weights are used for weighted means/SDs and for computing weighted income quantiles; otherwise, unweighted statistics are reported[11,12]. Descriptive tables use the full CHFS household file (N=34,643), whereas regressions use the effective sample (N=33,835) after listwise deletion for covariates and mediator availability.

4.2 Cleaning and Construction

Identifiers are standardized to hhid and pid; survey weights are retained and validity-checked ; merges are performed on hhid–pid with post-merge uniqueness checks. Income is constructed from total household income (total_inc; fallback to a master field if components are missing), per-capita income (pcinc), and several transforms: hyperbolic sine (asinh_*), symmetric logs (lns_*), and natural logs (ln_*). We winsorize income at the 1/99 percentiles and form winsorized logs lntinc_w and lnpinc_w. Income terciles (inc_terc) and quintiles (inc_quint) are based on lnpinc_w; when weights exist, weighted tiles (inc_terc_w, inc_quint_w) are also created[11].

Household-level outcomes are: supplementary pension/annuity (ins_pension_hh), commercial health insurance (ins_health_hh), housing provident fund (has_housing_pf), and marketable securities (sec_any); an extended flag wealth_prod_hh is included if present. Mediators are financial literacy (fin_know_z) and risk attitudes (risk_att_z). Baseline controls comprise age, male, edu_yrs, hhsize, and urban, plus province fixed effects (preferred fields FE_P/FE_P_*; otherwise province or encoded fe_prov). Descriptive outputs are written to outputs/tables/descriptives_chfs.xlsx. A companion plotting script reads these sheets to generate figures (overall; tercile; quintile; urban; age; education) under outputs/figures.

4.3 Master Variable Dictionary

Table 2 Defines all Variables, Coding, and Sources Used in the Analysis.

Table 2. Variable Definitions, Construction, Coding, and Sources[11,12]

Variable (alias)	Level	Definition / construction	Coding / transform	Source
hhid, pid	HH / IND	Standardized identifiers; strings enforced.	Renamed to hhid, pid.	CHFS / CHARLS
weight_hh / wt_hh	HH	Household survey p-weights for weighted summaries	Numeric; ≤ 0 set to	CHFS
		and weighted tiles.	missing.	
INDV_weight	IND	Individual survey p-weights; optional svyset.	e.g., svyset pid [pweight = ...].	CHARLS
total_inc	HH	Total household income (fallback to master field if	Double.	CHFS
		needed).		
pcinc	HH	Per-capita income = total_inc/hhsize if valid.	Double.	CHFS
asinh_tinc, asinh_pinc	HH	Hyperbolic sine transforms of total / per-capita income.	asinh ( ).	CHFS
lns_tinc, lns_pinc	HH	Symmetric logs of total / per-capita income.	sign(x) ln ($\lvert x\rvert$ + 1).	CHFS
lntinc, lnpinc	HH	Natural logs of total / per-capita income.	ln (x + 1).	CHFS
lntinc_w, lnpinc_w	HH	Logs after 1/99 winsorization.	Winsorize then log.	CHFS
inc_terc, inc_quint	HH	Income terciles/quintiles based on lnpinc_w (un-	$nq(3)$/ $nq(5)$ tiles.	CHFS
		weighted).		
inc_terc_w, inc_quint_w	HH	Weighted tiles when weights exist.	aweight approximation.	CHFS
ins_pension_hh	HH	Supplementary pension/annuity participation.	Binary (0/1).	CHFS
ins_health_hh	HH	Commercial health insurance participation.	Binary (0/1).	CHFS
has_housing_pf	HH	Housing provident fund (household has account).	Binary (0/1).	CHFS

sec_any	HH	Marketable securities (funds/stocks/bonds) held.	Binary (0/1).	CHFS
wealth_prod_hh	HH	Extended wealth-product flag (if present).	Binary (0/1).	CHFS
fin_know_z	HH	Standardized index of financial literacy.	*z*-score.	CHFS
risk_att_z	HH	Standardized index of risk attitudes/tolerance.	*z*-score.	CHFS
age, male, edu_yrs, hhsize, urban	HH	Baseline controls.	Numeric / dummies.	CHFS
FE_P / province / fe_prov	HH	Province fixed effects (preferred: FE_P).	Factor dummies.	CHFS
age_band	IND/HH	Age groups: <55 / 55–64 / 65+.	1/2/3 with labels.	CHARLS / CHFS
pension_ins_bin	IND	Any social pension insurance.	Binary (0/1).	CHARLS
medical_ins_bin	IND	Any social medical insurance.	Binary (0/1).	CHARLS
Tiered outcomes	HH	Protection / Accumulation / Market triplet used in figs-	Derived from bina-	CHFS
		ures/tables.	ries above.	

Notes. CHFS is household-level; CHARLS is individual-level[11,12]. An asterisk (*) denotes optional household mirroring- for example, mapping an individual's age band to the household head.

4.4 Descriptive Statistics and Visualization

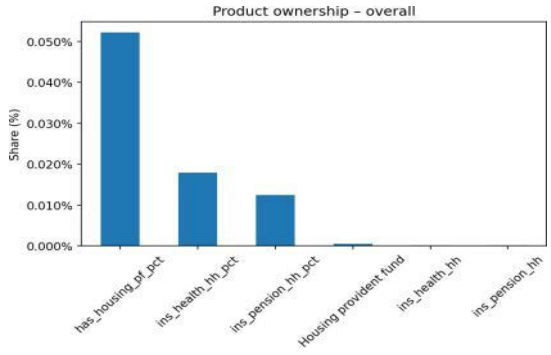
We report overall and income-quantile (tercile/quintile) breakdowns in both unweighted and weighted forms, plus urban, age-band, and education segments. Summary statistics are reported in Table 3 for overall, tercile, and quintile groups. Figure 4 consolidates ownership rates across overall and subgroup breakdowns. The Excel workbook stores precomputed percentages (columns with suffix _pct) where available; otherwise percentages are reconstructed from counts and sample sizes. A companion plotting script generates figures with consistent styling (overall; tercile; quintile; urban; age; education).

Table 3. Descriptive Statistics: Overall and by Income Terciles/Quintiles

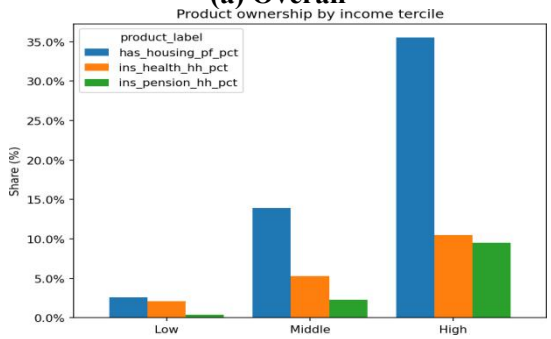
Group	N / N_w	Housing PF (%)	Commercial health (%)	Supplementary pension (%)	Per-capita income	Total income
Panel A. Unweighted						
Overall	34,643	18.1%	6.2%	4.3%	29,238.0	82,272.4
T1	10,817	2.6%	2.1%	0.4%	4,506.5	14,941.7
T2	10,376	13.9%	5.3%	2.3%	17,014.8	56,527.2
T3	12,654	35.5%	10.5%	9.5%	63,936.7	171,375.9
Q1	6,748	1.6%	1.7%	0.2%	2,578.8	8,025.3
Q2	6,077	5.4%	3.1%	0.7%	8,828.3	30,397.1
Q3	6,156	13.4%	5.1%	2.2%	16,594.2	56,150.4
Q4	7,009	24.3%	7.2%	4.8%	28,795.7	84,986.4
Q5	7,857	41.3%	12.5%	12.1%	83,914.1	221,017.1
Panel B. Weighted						
Overall	440,578,921	14.3%	5.5%	3.3%	26,024.9	73,005.9
T1	158,319,066	1.7%	1.9%	0.2%	4,491.4	14,073.6
T2	133,002,686	10.9%	5.2%	1.9%	16,835.0	54,777.2
T3	138,387,911	32.9%	10.0%	8.4%	63,094.2	168,287.8
Q1	99,763,023	1.1%	1.6%	0.1%	2,619.5	7,611.7
Q2	85,582,900	3.5%	2.8%	0.4%	8,765.9	28,882.7
Q3	79,279,050	10.8%	5.1%	1.9%	16,535.4	55,094.0
Q4	82,848,620	22.2%	6.8%	4.2%	28,711.7	84,652.0
Q5	82,236,070	38.6%	12.2%	11.0%	84,883.2	221,196.1

Notes. Percentages are shares within each group. The weighted panel uses survey p-weights where available. "Housing PF" denotes the housing provident fund; incomes are in CNY.

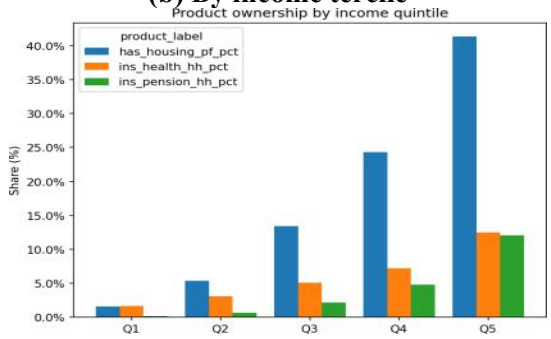
Legend mapping (in-figure variable names): ins_pension_hh = Supplementary pension/annuity; ins_health_hh = Commercial health insurance; has_housing_pf = Housing provident fund; sec_any = Marketable securities.



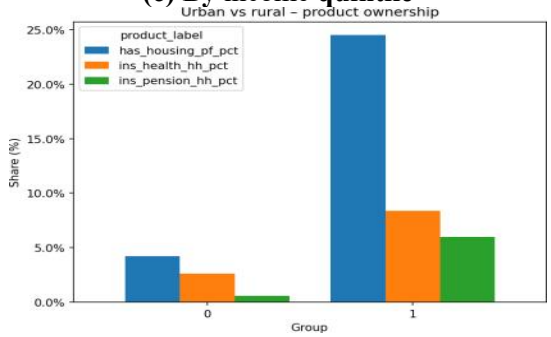
(a) Overall



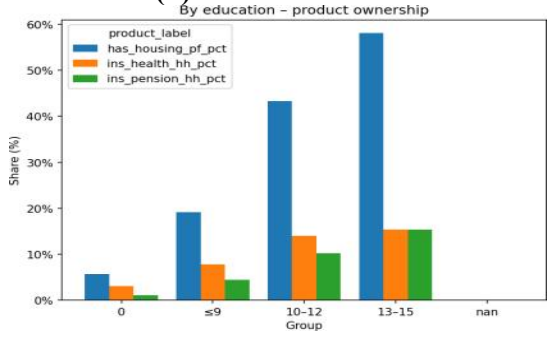
(b) By income tercile



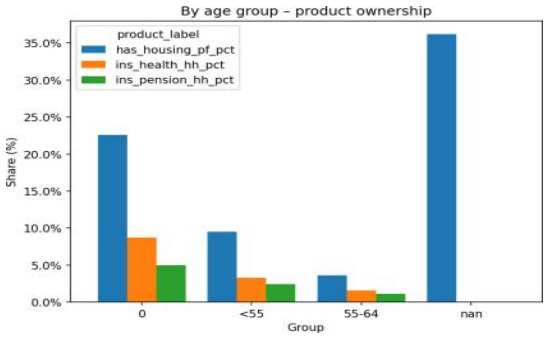
(c) By income quintile



(d) Urban vs. Rural



(e) By education



(f) By age band

Figure 4. Ownership Rates-Consolidated View.

5. Research Design and Identification Strategy

5.1 Empirical Framework and Identification Logic

Our outcomes are household participation in four retirement-finance products (dataset–field mapping in Appendix A): supplementary pension/annuity, commercial health insurance, housing provident fund, and marketable securities. Following Section 4, we estimate binary-choice models with province fixed effects and province-clustered standard errors:

$$\Pr(Y_h=1)=F\left(\alpha+\beta\,\mathrm{Income}_h+\gamma^{\top}X_h+\lambda_{prov}+\varepsilon_h\right) \quad (4)$$

where $F(\cdot)$ is logit (probit as robustness). Income is ranked by terciles (T1–T3; T1 as the reference) based on log per-capita income; continuous log measures are used in robustness. The baseline controls X_h include age, sex (male), education in years, household size, and urban status. We report coefficients together with average marginal effects (AME) and predicted probabilities. Weights and winsorization rules follow Section 4. Full stepwise and main-specification coefficients are provided in Appendix Tables 6–10.

Interpretation discipline. To enhance comparability across outcomes and avoid the scale pitfalls of logit coefficients:

(i) the core baseline is M2 (income terciles + controls + province FE);

(ii) we complement odds-ratio inference with AME and predicted-probability profiles in Section 6;

(iii) mediation is assessed via stepwise inclusion of financial-literacy and risk-attitude indices and via SEM to quantify indirect effects.

What the model answers.

Is there a monotonic post-control income gradient in participation?

How much of the gradient is mediated by financial literacy and risk attitudes?

Does the gradient differ across urban/rural, education, and age groups? These map to Table 4 and Section 6.

Table 4. Core Evidence Dashboard (M2 Logit: Income Terciles + Controls + Province FE)

Pension		Health	Housing PF	Securities
Income T2 vs T1	↑**	↑**	↑***	↑***
Income T3 vs T1	↑***	↑***	↑***	↑***
T1 baseline prob. (%)	0.1	2.3	2.2	0.6
T3–T1 AME (pp)	3.7	6.3	22.7	3.8
N (effective sample)	33,835	33,835	33,835	33,835
Controls included	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
SE clustered	Province	Province	Province	Province
Weights (p-weights)	Yes	Yes	Yes	Yes

Notes. Arrows/stars are from the M2 logit (*p<.10, **p<.05, ***p<.01). "T1 baseline prob."and "T3–T1 AME"come from the same M2 specification using margins with province clustering and pweight. Dataset–field mappings are in Appendix A.

M2 core specification (used across main tables)

Estimator. Logit in the main text (probit/cloglog/rare-events as robustness).

Outcomes. Household participation in four products: supplementary pension/annuity, commercial health insurance, hous- ing provident fund, and marketable securities.

Income rank. Categorical terciles inc_terc with T1 as reference (continuous lnpinc_w/lntinc_w in robustness).

Controls. Age, male, education (years), household size, urban.

Fixed effects & SE. Province fixed effects; province-clustered robust SE.

Weights. Household survey p-weights (applied consistently in estimation and margins).

Reporting. Coefficients, AME (pp), predicted probabilities; unified effective sample size *N*

5.2 Baseline Findings (Narrative)

Across all outcomes, income shows a monotone post-control gradient: T2 and T3 have progressively higher participation than T1. The gradient is strongest for market exposure (sec_any) and the housing provident fund (has_housing_pf), moderate for commercial health insurance (ins_health_hh), and smaller but positive for supplementary pension/an-nuity (ins_pension_hh). Controls behave as expected (education and urban residence positive ; age patterns are product-specific).

Economic magnitudes. The dashboard anchors magnitudes via T1 baseline probabilities and T3–T1 AME (pp). Section 6 visualizes tercile profiles and reports AME with 95% CIs; values in Table 4 are read from margins.

5.3 Mediation and Mechanism Evidence

Stepwise inclusion of *financial literacy* and *risk attitudes* attenuates income coefficients, with the largest proportional shrinkage for marketable securities. A structural equation model indicates positive indirect effects; however, in this sample the bootstrap 95% confidence intervals include zero, so the indirect effects are *not statistically significant*. *Quantitatively (securities):* the literacy channel explains about 0.1% of the total income effect (bootstrap 95% CI [-7.6, 7.8]), and the risk channel about 6.2% (95% CI [-1.9, 14.4]). Outcome-specific SEM tables/figures are in Appendix E (see overview in Appendix Figure 7).

5.4 Heterogeneity and Robustness

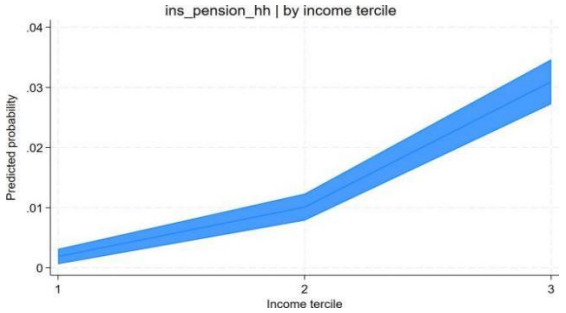
Urban and higher-education groups exhibit steeper income-to-participation slopes. For securities, the tercile contrast (T3–T1) is 3.29 pp larger in urban than rural households ($p <$ .001; from margins with lincom). Results are robust to continuous-income specifications (lnpinc_w/lntinc_w), weighted tiles, probit/LPM alternatives, and narrower outcome definitions. Given low baselines for some protection outcomes, Firth logit, cloglog, and rare-events logit yield the same signs and similar magnitudes (Appendix D). Coefficient details for urban/rural splits are reported in Appendix Tables 11–14.

5.5 Transparency on Weights and Fit

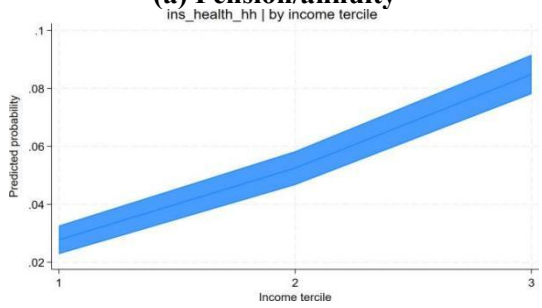
All baseline and margins computations use household p-weights with province-clustered SE. Predicted probabilities and AME are obtained under identical weighting and FE settings. Appendix D reports Pseudo R^2 and AUC by outcome (M2), alongside ROC plots. Model fit metrics (Pseudo R^2 and AUC) are reported in Appendix Table 17.

AME-based from the M2 logit with controls and province fixed effects; axes are harmonized across panels. *Legend mapping:* ins_pension_hh = Supplementary pension/annuity; ins_health_hh

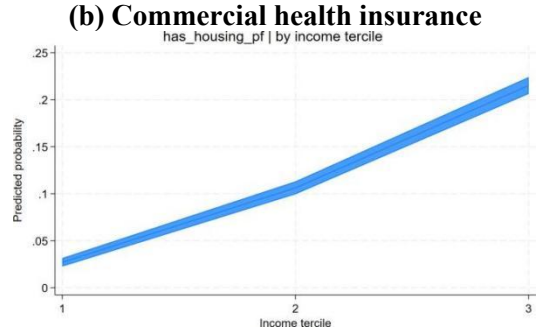
= Commercial health insurance; has_housing_pf = Housing provident fund; sec_any = Marketable securities.



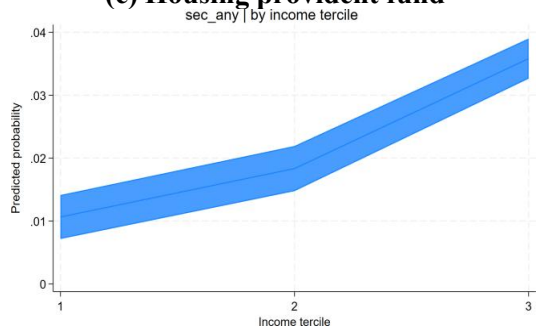
(a) Pension/annuity



(b) Commercial health insurance



(c) Housing provident fund



(d) Marketable securities

Figure 5. Predicted Probabilities by Income Terciles

6. Empirical Results and Discussion

6.1 Participation Patterns

Participation is higher among urban, better-educated, and younger-old groups. Income gradients are monotonic across outcomes, but slopes differ by product: protection vs. accumulation vs. market exposure.

6.2 Baseline Effects and Magnitudes

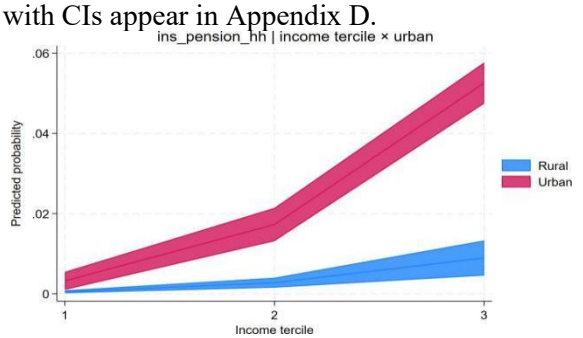
Using the core M2 specification, we find a clear T1→T3 increase across outcomes. The gradient is strongest for marketable securities and the housing provident fund, moderate for commercial health insurance, and smaller for supplementary pension/annuity. Table 4 reports T1 baseline probability (%), T3–T1 AME (pp), and a unified N . Predicted probabilities by income terciles are shown in Figure 5.

6.3 Mediation and Mechanisms

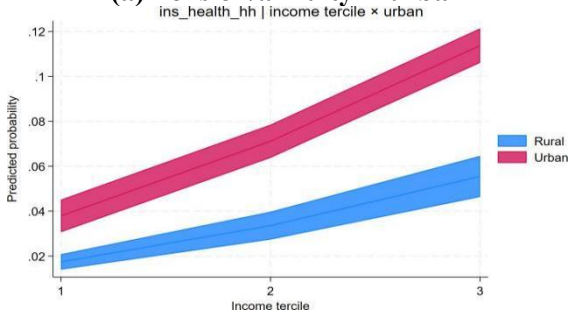
Adding *financial literacy* and *risk attitudes* reduces income coefficients, most visibly for securities. SEM shows positive indirect effects but with bootstrap 95% CIs overlapping zero; thus indirect effects are *not statistically significant*. For securities, literacy explains 0.1% (95% CI [-7.6, 7.8]) and risk explains 6.2% (95% CI [-1.9, 14.4]) of the total income effect. Fit indices and full indirect-effect tables are in Appendix E (Figure 7)

6.4 Heterogeneity Across Urban Status, Education, and Age

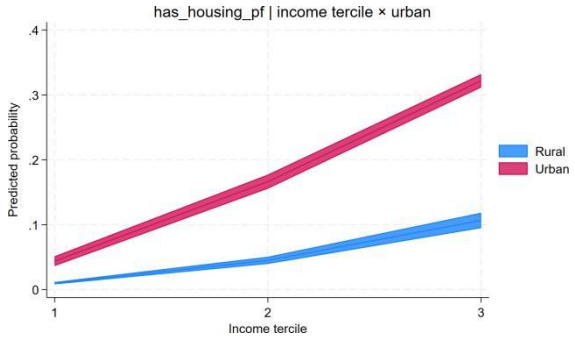
Income–participation slopes are *steeper in urban areas* and among higher-education groups; the younger-old show stronger responses for market exposure. The T3–T1 contrast differs by 3.29 pp for securities ($p < .001$) and by 10.91 pp for housing PF ($p<.01$). Figure **6** shows urban interactions; education/age panels and AME with CIs appear in Appendix D.



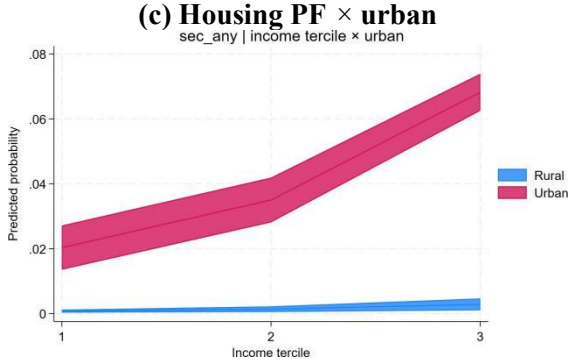
(a) Pension/annuity × urban



(b) Health insurance × urban



(c) Housing PF × urban



(d) Securities × urban

Figure 6. Income–Participation Gradients by Urban/Rural

Slopes are generally steeper in urban subsamples; exact AME and CIs are provided in Appendix figures.

6.5 Robustness and Rare-Outcome Diagnostics

Findings hold under continuous-income transformations (log and asinh), weighted tiles, probit/LPM alternatives, and narrower outcome definitions. Given lower base rates for pension and health insurance, Firth logit, cloglog, and rare-events logit replicate signs and significance. Robustness to alternative income measures and estimators appears in Appendix Tables 15–16.

6.6 Implications for Product and Policy

Protection (lower income). Emphasize affordability and simple defaults; streamline thresholds, fees, and claims.

Accumulation (middle income). Lifecycle/target-date and pension wealth–management products balance growth with liquidity.

Customization/market exposure (higher income). Combine annuities, securities, health/LTC, and trust solutions for diversification and intergenerational planning.

Mediation via literacy and risk tolerance motivates defaults, investor education, and suitability governance to improve product–demand fit.

7. Product Innovation and Policy Recommendations

This section translates the empirical patterns into implementable designs along three axes:

A tiered supply aligned with income-linked needs and participation frictions;

Service and distribution models spanning banking and insurance ecosystems with trust/fund linkages; and

FinTech enablement and governance supporting personalization, suitability, and risk control. We conclude with policy proposals to enhance inclusiveness and long-run sustainability of the second and third pillars.

7.1 Tiered Supply Framework (Aligned with Empirical Segments)

Table 5 summarizes the tiered product framework aligned with empirical segments. Low-income (protection and inclusion). Prioritize high-protection/low-threshold products with premium subsidies, auto-enrollment/defaults, simple clauses, and frictionless claims.

Middle-income (stable accumulation). Emphasize target-date/target-risk strategies, fixed-income-plus allocations, pension wealth-management products, and convenient redemption windows.

High-income (diversification and customization). Offer annuities + commercial health + LTC portfolios, optionally combined with (family) trusts, tax-advantaged wrappers, and cross-account interoperability.

7.2 Services and Channels

Bank-led ecosystems. Integrate pension accounts, payments, advisory, age-friendly UX, health-management modules, community resources.

Table 5. Tiered Product Framework Aligned with Empirical Segments

Segment	Primary objective	Product menu (China-specific)	Design levers (policy-consistent)	Safeguards / KPIs
Low-income	Protection & inclusion	• Personal pension account *CNY·12,000capp. a.*; • Entry-level pension WMPs (bank label); • Exclusive commercial pension insurance	• Low minimums, elder-friendly UX; • Default options (where permitted);	• Coverage in Q1/Q2; • Claim resolution time;

		(pilot cities); • Long-term care insurance (pilots).	• Transparent fee disclosure.	• Suitability pass rate.
Middle-income	Stable accumulation	• Target-date/target-risk funds; • Pension bank WMPs (steady income-plus); • Commercial pension annuities.	• Glide-path defaults; • Contribution nudges/payroll links; • Liquidity windows for life events.	• Net-of-fee return vs. benchmark; • Tenure/persistence; • Cost disclosure compliance.
High-income	Diversification & customization	• Bundles: annuity+health+LTC; • Advisory portfolios (funds/ETFs/bonds); • Pension/family trusts.	• Goal-based IPS & rebalancing; • Cross-account tax coordination; • ESG/constraints on demand.	• After-tax IRR, shortfall risk; • Fiduciary process trail; • Suitability & disclosure KPIs.

Insurance-led integration. Build loops of products + care pathways + senior-living, standardize assessments and SLAs; link claims with medical records (lawfully).

Joint operations. Shared suitability rules, interoperable KYC/AML/consent, unified complaint/rectification channels; modular product architecture.

7.3 FinTech Enablement and Governance

Privacy-preserving analytics. Federated learning/secure aggregation; on-device scoring; tokenized consent.

Personalization and risk control. Recommendations calibrated to risk tolerance and financial literacy; early-warning models with human-in-the-loop; scenario stress.

Model governance. Explainability artifacts, fairness/impact audits, data minimization, challenge processes; versioning and attestations for suitability.

7.4 Policy Recommendations

Enhance inclusiveness in Pillars II/III. Differentiated tax preferences or subsidies; broaden defaults (auto-enroll/escalate) with opt-out; fee caps and standardized key-facts.

National pooling and market linkage. Advance national pooling while expanding funded pillars; align investing with a multi-tier capital market under prudent limits.

Investor education and evaluation. Institutionalize curricula, embed nudges, and run longitudinal impact evaluations with public dashboards.

Implementation roadmap. Pilot zones and A/B defaults; open templates; unified complaint portals; evidence-based KPIs (coverage in lower quantiles, net-of-fee yield, lapse/claim metrics, rectification rate).

8. Conclusion

This paper investigates whether-and through which channels-households at different income levels participate in retirement-related financial products in China, and how well current supply matches stratified demand under the three-pillar system. Using CHFS 2019 with complementary CHARLS 2020, we estimate logit/probit models with province fixed effects and survey weights, examine heterogeneity, and quantify mediation via stepwise regressions and SEM.

Direct answers.

Demand heterogeneity. Participation shows a clear, monotonic income gradient; urban residence, education, and younger-old status amplify market exposure.

Product–demand matching. Protection aligns with low-income needs (when thresholds are low); accumulation products fit middle-income stability; diversified bundles match high-income customization-yet gaps remain in affordability/intelligibility/continuity.

Mechanism chain. Literacy and risk attitudes form channels from income to participation; indirect effects are positive but statistically non-significant in SEM bootstraps here.

Policy relevance. Evidence supports tiered supply with defaults and safeguards for inclusion, reliable accumulation for the middle segment, and customizable portfolios for the upper segment, delivered via integrated channels with consistent suitability and risk control.

Limitations and future research. Cross-sectional identification, measurement of mediators/outcomes, and institutional heterogeneity may limit external validity; panel/event-study designs and richer batteries can sharpen inference. At production scale, privacy-preserving analytics should pair with explainability, fairness testing, and longitudinal dashboards.

Aligning products, services, and governance with income-linked needs and measured frictions

can improve product–demand fit and strengthen the sustainability of China's second and third pillars.

References

[1] Hu, J. (2013). Retirement finance: Conceptual definition and practice issues. Finance & Trade Economics, 34(6), 43–52. [In Chinese].

[2] Sun, X. (2016). Status and development of China's retirement-finance industry. Chinese Journal of Gerontology, 36(12), 3062–3065. https://doi.org/10.3969/j.issn.1005-9202.2016.12.111. [In Chinese].

[3] Zhou, X. (2022). Empirical analysis of key obstacles to the development of China's pension finance. Financial Theory & Teachin, 2022(5), 45–50. [In Chinese].

[4] Wang, Q. (forthcoming). Personalized pension-finance services and market demand. Forthcoming / working paper. [In Chinese].

[5] Li, C., Wu, W., & Lu, Y. (forthcoming). Theoretical logic and practical paths for devel- oping pension finance in China. Forthcoming / working paper. [In Chinese].

[6] Cheng, H., Liu, L., & Pan, L. (2024). How Japan/U.S. capital markets support pension finance: Practices and lessons. China Forex, 2024(20), 68–70. [Commentary; non-academic source]. [In Chinese].

[7] Wu, Q. (2024). Pension finance is the trend: How to build healthy interaction with capital markets? Sohu (repost). [Commentary; non-academic source]. Retrieved from https://www.sohu.com/a/754002378_121117082. Accessed 2025-10-09. [In Chinese].

[8] National Bureau of Statistics of China (NBS). (2023). China Statistical Yearbook 2023. Beijing: China Statistics Press. Available at http://www.stats.gov.cn/sj/ndsj/. [Official yearbook].

[9] Ministry of Human Resources and Social Security (MOHRSS). (2023). Statistical Bulletin on the Development of Human Resources and Social Security. Available at http://www.mohrss.gov.cn. [Official statistical bulletin].

[10] Ministry of Housing and Urban–Rural Development (MOHURD). (2023). National Housing Provident Fund Annual Report (2023). Available at https://www.mohurd.gov.cn. [Official report; housing provident fund].

[11] Gan, L., Yin, Z., Jia, N., Xu, S., & Ma, S. (2013). Data You Need to Know About China. Springer. [Official CHFS data reference].

[12] Zhao, Y., Strauss, J., Yang, G., Giles, J., Hu, P., Lei, X., Park, A., Smith, J. P., & Wang,Y. (2013). China Health and Retirement Longitudinal Study (CHARLS), 2011–2012 National Baseline User' Guide. National School of Development, Peking University.

Appendix A. Data dictionary and variable mapping

Datasets. CHFS 2019 household analytic file; CHARLS 2020 individual analytic file.

Identifiers. hhid, pid; province FE may be encoded to province_id.

Weights. Household p-weight weight_hh (fallbacks wt_hh, weight_household).

Income. total_inc, pcinc; logs lntinc, lnpinc; winsorized logs lntinc_w, lnpinc_w; asinh_tinc, asinh_pinc.

Income ranks. inc_terc, inc_quint; weighted versions inc_terc_w, inc_quint_w. Outcomes. Pension/annuity (ins_pension_hh); Commercial health (ins_health_hh); Housing provident fund (has_housing_pf); Marketable securities (sec_any).

Mediators. fin_know_z, risk_att_z.

Controls. age, male, edu_yrs, hhsize, urban; FE: i.province / i.province_id. Segments. Urban (urban); education bands (by edu_yrs); age bands (<55 / 55–64 / 65+). Outputs.

```
Tables -> outputs/tables/
Figures -> outputs/figures/
```

Mapping Note.

```
If province is string, encode to province_id
and set FE macro to i.province_id.
```

Appendix B. Cleaning and statistical conventions

Missing/specials. Non-positive weights → missing; income winsorized at 1%/99%.

Tiles. Ranks based on lnpinc_w; weighted tiles apply survey p-weights.

Inference. Province FE; province-clustered robust SE; report coefficients and AME.

Reproducibility. Pipeline exports:

```
descriptives_chfs.xlsx reg_*.tex
(regression tables)
med_effects_*.csv (mediation summaries)
figures: ownership, margins, SEM paths
```


Appendix C. Replication notes (files, logs, and paths)

Tables available (examples):

```
\detokenize{reg_stepwise_logit.tex}
\detokenize{reg_main_logit_base.tex}
\detokenize{reg_main_logit_full.tex}
\detokenize{reg_main_probit_base.tex}
\detokenize{reg_main_probit_full.tex}
\detokenize{reg_hetero_urban0_logit.tex}
\detokenize{reg_hetero_urban1_logit.tex}
\detokenize{reg_hetero_urban0_probit.tex}
\detokenize{reg_hetero_urban1_probit.tex}
\detokenize{reg_robust_income_logit.tex}
\detokenize{reg_robust_income_probit.tex}
```

Figures available (examples):

```
outputs/figures/holdings_overall.png
outputs/figures/holdings_by_tercile.png
outputs/figures/predprob_*_by_inc_terc.png
outputs/figures/predprob_*_inc_terc_by_urban.png
outputs/figures/sem_path_*.png
```

Logs.

```
03_main_regs.log
```

Computational environment and run order

Stata 16.0; Python 3.10 (pandas 2.1, matplotlib 3.8); LaTeX (TeX Live 2025). Key Stata user packages: winsor2.

Run order: 01_clean.do → 02_figures.do → 03_main_regs.do.

All logs and intermediate files are listed in Appendix A/C with file-time stamps.

Table 6. Baseline Stepwise Regressions (Logit)

	(1)	→	(2)	→	(3)
Securities (Income Only) Securities + Controls Securities + Mediators					
inc_terc	β		β		β
T2 vs T1		0.936***		0.936***	1.094***
		(0.181)		(0.181)	(0.287)
T3 vs T1	β	2.134***	β	2.134***	2.347***
		(0.190)		(0.190)	(0.335)
age	β		β	0.0143***	0.0191***
				(0.00229)	(0.00548)
male	β		β	-0.319***	-0.306***
				(0.0806)	(0.100)
edu_yrs	β		β	0.190***	0.170***
				(0.0136)	(0.0301)
hhszie	β		β	0.107***	-0.0400
				(0.0314)	(0.0344)
urban	β		β	2.470***	2.184***
				(0.313)	(0.397)
fin_know_z	β		β		-0.109
					(0.178)
risk_att_z	β		β		-0.596***
					(0.194)
Constant	β	-8.427***	β	-8.427***	-6.635***
		(0.412)		(0.412)	(1.678)
Province FE	Yes		Yes		Yes
Clustered SE	Province		Province		Province
Observations	33835		33835		33835

Appendix D. Full tables Moved from the Main Text

D.1Baseline stepwise (logit)

D.2Main specifications (reference)

D.3Heterogeneity: urban vs. rural

Table 7. Main Logit (Base)

	(1)	(2)	(3)
	Pension/ annuity	**Commercial health insurance**	**Housing provident fund**
inc_terc			
T2 vs T1	0.408***	0.372***	0.813***
	(0.090)	(0.067)	(0.089)
T3 vs T1	0.864***	0.785***	1.870***

	(0.099)	(0.083)	(0.099)
age	0.005	0.013***	0.010***
	(0.004)	(0.002)	(0.003)
male	-0.065	-0.228**	-0.180***
	(0.064)	(0.053)	(0.056)
edu_yrs	0.116**	0.114***	0.141***
	(0.010)	(0.009)	(0.009)
hlthzse	0.046*	0.018	0.087***
	(0.022)	(0.020)	(0.020)
urban	1.115***	0.955***	1.756***
	(0.238)	(0.184)	(0.177)
Constant	-5.610***	-6.124***	-6.280***
	(0.307)	(0.241)	(0.245)
Province FE	Yes	Yes	Yes
Clustered SE	Province	Province	Province
Observations	33835	33835	33835

Table 8. Main Logit (Full)

	(1)	(2)	(3)
	Pension/annuity	**Commercial health insurance**	**Housing provident fund**
inc_terc			
T2 vs T1	0.351** (0.101)	0.323** (0.078)	0.742** (0.094)
T3 vs T1	0.734** (0.120)	0.651** (0.099)	1.673** (0.105)
age	0.009* (0.005)	0.016** (0.002)	0.011** (0.003)
male	-0.054 (0.071)	-0.196** (0.058)	-0.151** (0.058)
edu_yrs	0.100** (0.013)	0.101** (0.011)	0.121** (0.011)
hhsize	0.030 (0.023)	0.009 (0.021)	0.073** (0.021)
urban	1.003** (0.257)	0.860** (0.201)	1.673** (0.194)
fin_know_z	0.151** (0.073)	0.118** (0.058)	0.032 (0.056)
risk_att_z	0.096 (0.086)	0.085 (0.069)	0.028 (0.067)
Constant	-6.026** (0.339)	-6.585**** (0.268)	-6.606** (0.270)
Province FE	Yes	Yes	Yes
Clustered SE	Province	Province	Province
Observations	33835	33835	33835

Table 9 Main Probit (Base)

	(1)	(2)	(3)
	Pension/annuity	**Commercial health insurance**	**Housing provident fund**
inc_terc			
T2 vs T1	0.210*** (0.0446)	0.180*** (0.0330)	0.424*** (0.0433)
T3 vs T1	0.426*** (0.0479)	0.373*** (0.0391)	0.934*** (0.0459)
age	0.00255 (0.00229)	0.00679*** (0.00121)	0.00520*** (0.00172)
male	-0.0353 (0.0345)	-0.121*** (0.0296)	-0.0883*** (0.0315)
edu_yrs	0.0600*** (0.00559)	0.0583*** (0.00500)	0.0723*** (0.00488)
hhszie	0.0271** (0.0136)	0.0104 (0.0120)	0.0426*** (0.0110)
urban	0.564*** (0.117)	0.474*** (0.0913)	0.894*** (0.0870)
Constant	-3.014*** (0.160)	-3.160*** (0.130)	-3.334*** (0.132)
Province FE	Yes	Yes	Yes
Clustered SE	Province	Province	Province
Observations	33835	33835	33835

Table 10 Main Probit (Full)

	(1)	(2)	(3)
	Pension/annuity	**Commercial health insurance**	**Housing provident fund**
inc_terc			
T2 vs T1	0.190***	0.161***	0.375***
	(0.0498)	(0.0366)	(0.0457)
T3 vs T1	0.381***	0.329***	0.868***

	(0.0587)	(0.0472)	(0.0520)
age	0.00451	0.00808***	0.00574***
	(0.00257)	(0.00130)	(0.00178)
male	-0.0322	-0.106***	-0.0828***
	(0.0379)	(0.0326)	(0.0325)
edu_yrs	0.0533***	0.0529***	0.0622***
	(0.00684)	(0.00572)	(0.00564)
hhsize	0.0195	0.00385	0.0352***
	(0.0143)	(0.0127)	(0.0116)
urban	0.527***	0.445***	0.868***
	(0.128)	(0.101)	(0.0959)
Constant	-3.271***	-3.464***	-3.503***
	(0.172)	(0.139)	(0.140)
Province FE	Yes	Yes	Yes
Clustered SE	Province	Province	Province
Observations	33835	33835	33835

Table 11. Heterogeneity (Logit): urban=0

	Securities (urban=0)
inc_terc T2 vs T1	1.909
T3 vs T1	(1.241) 3.733***
	(0.983)
age	0.0266
male	(0.0243) -1.629***
edu_yrs	(0.606) 0.200***
	(0.0512)
hhsize	-0.0540
	(0.134)
Constant	-9.053***
	(2.172)
Province FE	Yes
Clustered SE	Province
Observations	5359

Table 12. Heterogeneity (logit): urban=1

	Securities (urban=1)
inc_terc T2 vs T1	0.476**
T3 vs T1	(0.197) 1.177***
	(0.182)
age	0.0135***
male	(0.00233) -0.294***
edu_yrs	(0.0803) 0.202***
	(0.0136)
hhsize	0.111***
	(0.0333)
Constant	-5.822***
	(0.313)
Province FE	Yes
Clustered SE	Province
Observations	21921

Table 13. Heterogeneity (probit): urban=0

	Securities (urban=0)
inc_terc T2 vs T1	0.829
T3 vs T1	(0.507) 1.475***
	(0.377)
age	0.0105
male	(0.00929) -0.667***
edu_yrs	(0.243) 0.0971***
	(0.0243)
hhsize	-0.0236
	(0.0531)
Constant	-2.690***
	(0.707)
Province FE	Yes
Clustered SE	Province
Observations	5359

Table 14. Heterogeneity (probit): urban=1

	Securities (urban=1)
inc_terc T2 vs T1	0.211**
T3 vs T1	(0.0875) 0.518***
	(0.0782)
age	0.00564***
male	(0.00140) -0.129***
edu_yrs	(0.0359) 0.0878***
	(0.00612)
hhsize	0.0346***
	(0.0121)
Constant	-2.510***
	(0.143)
Province FE	Yes
Clustered SE	Province
Observations	21921

D4. Robustness: Income Measures and Methods

Table 15. Robustness (Logit): Alternative Income Measures

	lnpinc_w spec	lintnc_w spec
lnpinc_w	0.470***	
	(0.0549)	
lintnc_w		0.492***
		(0.0538)
age	0.0140***	0.0145***
	(0.00229)	(0.00232)
male	-0.328***	-0.338***
	(0.0809)	(0.0810)
edu_yrs	0.189***	0.186***
	(0.0136)	(0.0137)
hhsize	0.112***	-0.0410
	(0.0316)	(0.0346)
urban	2.582***	2.565***
	(0.315)	(0.315)
Constant	-12.25***	-12.48***
	(0.648)	(0.640)
Province FE	Yes	Yes
Clustered SE	Province	Province
Observations	33835	33835

Table 16 Robustness (Probit): Alternative Income Measures

	lnpinc_w spec	lntinc_w spec
lnpinc_w	0.243***	
	(0.0283)	
lntinc_w		0.257***
		(0.0280)
age	0.00624***	0.00638***
	(0.00120)	(0.00121)
male	-0.168***	-0.173***
	(0.0300)	(0.0301)
edu_yrs	0.102***	0.100***
	(0.00498)	(0.00500)
hhsize	0.0467***	0.00167
	(0.0127)	(0.0138)
urban	1.005***	0.992***
	(0.126)	(0.126)
Constant	-6.364***	-6.396***
	(0.319)	(0.314)
Province FE	Yes	Yes
Clustered SE	Province	Province
Observations	33835	33835

Table 17. Model Fit (M2, logit): Pseudo R2 and AUC by Outcome

Outcome	N	Pseudo R^2	AUC
supp. pension/annuity	33,835	0.232	0.855
commercial health	33,835	0.132	0.768
housing PF	33,835	0.339	0.870
securities	33,835	0.246	0.866

Notes. Same M2 controls/FE/weights as the main text. AUC uses in-sample predictions without weights (approximation when p-weights are used in estimation).

Appendix E. SEM Figures and Mediation Summaries

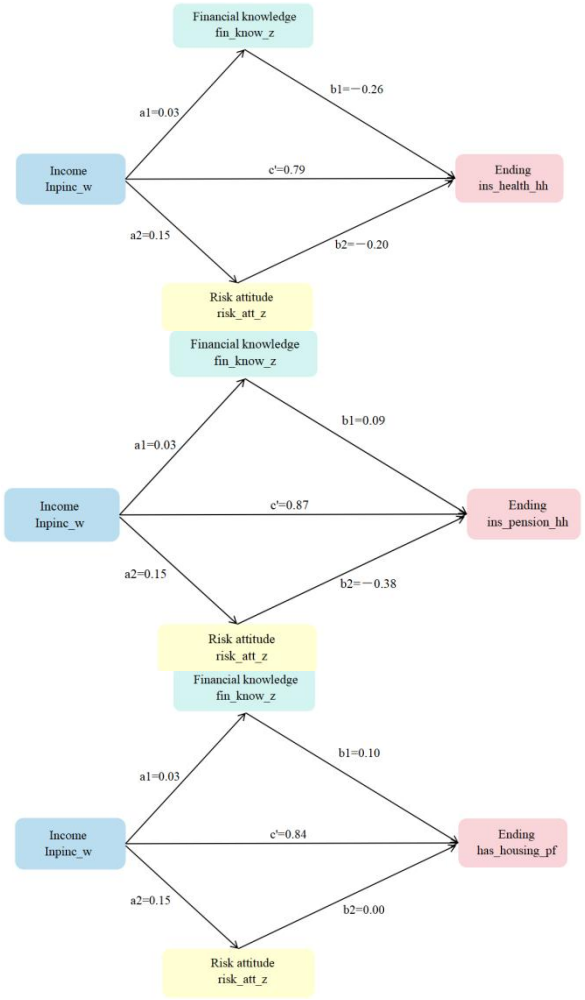


Figure 7. Representative SEM Paths (Three Outcomes). (Top) Pension/Annuity; (Middle) Commercial Health Insurance; (Underside) Housing Provident Fund

Appendix F. Policy and Product notes

Personal pensions (since 2022). Trial framework per MOHRSS Order No. 48; fund menu supervised under CSRC interim provisions.

Exclusive commercial pension insurance (pilot since 2021). Launched by CBIRC in select regions; staged expansion.

Long-term care insurance pilots. NHSA guidance (2020) on LTCI pilots; local implementation varies.

