Article

2025 3rd International Conference on the Sociology of the Global Economy, Education, Arts and Humanities (GEEAH 2025)

The Intrinsic Relationship between per Capita Disposable Income and the Participation Rate in Urban-Rural Residents' Basic Pension Insurance: Exploration and Insights

Yitong Fu [1,*]

[1] University of International Business and Economics, Beijing, China
* Correspondence: Yitong Fu, University of International Business and Economics, Beijing, China

Abstract: Against the backdrop of an increasingly aging population, China has implemented the Urban-Rural Residents' Basic Pension Insurance system to alleviate the economic pressures faced by the elderly. This study aims to investigate the impact of per capita disposable income on the participation rate in this pension insurance program. By constructing a linear logarithmic polynomial model and conducting an empirical analysis using panel data from 31 provinces, municipalities, and autonomous regions in China from 2017 to 2022, the study reveals a significant negative correlation between per capita disposable income and the participation rate in the Urban-Rural Residents' Basic Pension Insurance. This suggests that individuals with higher disposable income may opt for alternative pension schemes, such as commercial insurance, thereby lowering their participation in the basic system. The findings provide a reference for policymakers, suggesting further optimization of fiscal subsidies and tax incentive policies to enhance pension insurance coverage among low-income groups and promote the balanced development of the social security system.

Keywords: per capita disposable income; Urban-Rural Residents' Basic Pension Insurance participation; time-fixed effects; linear logarithmic polynomial model

Received: 16 March 2025
Revised: 22 March 2025
Accepted: 14 April 2025
Published: 29 May 2025

1. Introduction

This paper focuses on the impact of per capita disposable income on participation in the Urban-Rural Residents' Basic Pension Insurance.

As the country with the world's largest population, China is experiencing a particularly pronounced trend of population aging. According to data from the National Bureau of Statistics, by the end of 2023, the population aged 60 and above reached 296.97 million, accounting for 21.1% of the total population, while those aged 65 and above numbered 216.76 million, representing 15.4%. In the future, with declining birth rates and increasing life expectancy, these figures are expected to rise steadily. This demographic shift poses significant challenges to the social security system, labor market, family structures, and economic development patterns.

To address the challenges brought about by population aging, the Chinese government has introduced the Urban-Rural Residents' Basic Pension Insurance and Basic Medical Insurance systems. These initiatives aim to reduce the economic burden on the elderly and ensure that their basic living and healthcare needs are adequately met. The Urban-Rural Residents' Basic Pension Insurance, as a key component of the social security framework, plays a vital role in mitigating the pressures of an aging population.

Therefore, studying the relationship between per capita disposable income and participation in this pension insurance program is of great significance for crafting rational policies, optimizing the social security system, and improving the quality of life for the elderly. This topic seeks to uncover the intrinsic connection between these two factors through in-depth analysis, offering valuable insights and recommendations for policymakers.

2. Literature Review

Sena highlighted sustainability issues in China's Urban-Rural Residents' Pension Insurance system [1]. She noted that with intensifying population aging and slowing economic growth, the pension insurance system faces numerous challenges. Through an analysis of its revenue sources, expenditure patterns, and administrative efficiency, she proposed several improvement measures, including increasing fiscal investment, expanding insurance coverage, and optimizing investment strategies. These suggestions are critical for ensuring the long-term stability of the pension insurance system.

Yang adopted a systems theory approach to examine the challenges facing China's multi-tiered pension insurance system for urban employees [2]. She pointed out that deficiencies in system design, fund management, and policy implementation have led to various issues. To optimize this system, she recommended enhancing institutional integration, improving fund utilization efficiency, and refining supporting policies. Implementing these measures could boost the overall effectiveness of the pension insurance system.

Jia explored the impact of changes in urban residents' pension insurance coverage on consumption structure upgrades in the context of new-type urbanization [3]. Their empirical analysis revealed that higher pension insurance coverage boosts consumer confidence, thereby facilitating upgrades in consumption structures. This finding has significant implications for advancing the new-type urbanization process and improving residents' quality of life.

Ni, using data from the China Health and Retirement Longitudinal Study (CHARLS), investigated the influence of pension levels on elderly consumption behavior [4]. They found that higher pension levels increase consumption expenditure among the elderly, with particularly notable effects among low-income groups. These results offer valuable insights for refining the pension insurance system and promoting the well-being of the elderly population.

Xu identified risks in the operational management of the Urban-Rural Residents' Basic Pension Insurance system [5]. She noted that challenges such as low levels of informatization, insufficient staffing, and inadequate oversight mechanisms contribute to management risks. To address these issues, she proposed strengthening informatization efforts, optimizing staff allocation, and improving regulatory mechanisms. Implementing these strategies could enhance the management efficiency and service quality of the pension insurance system.

3. Theoretical Model

3.1. Selection of Target Variables and Influence Mechanism

This study uses the participation rate in the Urban-Rural Residents' Basic Pension Insurance as the dependent variable and per capita disposable income as the explanatory variable (i.e., the target variable). The research hypothesizes that there is a negative correlation between per capita disposable income and the participation rate in the Urban-Rural Residents' Basic Pension Insurance, and that this effect is closely tied to the level of per capita disposable income. Specifically, as per capita disposable income increases, residents tend to exhibit stronger risk awareness and greater financial capacity, making them more inclined to purchase commercial insurance products with higher coverage, thereby reducing their reliance on the Urban-Rural Residents' Basic Pension Insurance.

Due to the difficulty in obtaining specific data on individual insurance purchases, this study uses the publicly available participation rate in the Urban-Rural Residents' Basic Pension Insurance as a proxy variable to analyze the impact of per capita disposable income.

3.2. Omitted Variables and Control Variables

This study includes local fiscal education expenditure, local fiscal healthcare expenditure, and the old-age dependency ratio as control variables:

1) Local Fiscal Education Expenditure: This variable reflects the local government's emphasis on education and its fiscal capacity. Regions with higher education expenditure typically have stronger fiscal conditions and relatively higher per capita disposable income. Additionally, residents with higher education levels often demonstrate greater risk awareness and are more inclined to invest in insurance.
2) Local Fiscal Healthcare Expenditure: This variable has a dual effect on the participation rate in pension insurance. On one hand, higher healthcare expenditure indicates better local healthcare standards, which can reduce residents' medical burdens and leave them with more disposable income to invest in pension insurance. On the other hand, improved medical security may reduce residents' overall reliance on public welfare systems for retirement, potentially diminishing their demand for basic pension insurance.
3) Old-Age Dependency Ratio: This variable also exhibits a dual effect. A higher old-age dependency ratio suggests an increased demand for pension insurance; however, an excessively high proportion of elderly individuals can increase family burdens, thereby reducing household disposable income and suppressing the demand for pension insurance.

In summary, these three variables are closely associated with both per capita disposable income and pension insurance participation rates. Failure to control for them could lead to omitted variable bias. Therefore, this study incorporates these variables into the model to account for their influence on the target variable.

3.3. Summary and Hypotheses

Based on the above, the explanatory variable in this study is per capita disposable income (unit: yuan), while the control variables include local fiscal education expenditure (unit: billion yuan), local fiscal healthcare expenditure (unit: billion yuan), and the old-age dependency ratio (unit: %). The dependent variable is the participation rate in the Urban-Rural Residents' Basic Pension Insurance.

Drawing on theoretical analysis and existing research, this study proposes the following hypotheses:

1) There is a negative correlation between per capita disposable income and the participation rate in the Urban-Rural Residents' Basic Pension Insurance.
2) There may be a correlation between local fiscal education expenditure and the participation rate in the Urban-Rural Residents' Basic Pension Insurance, but the direction of this effect remains uncertain and requires empirical verification.
3) There is a positive correlation between local fiscal healthcare expenditure and the participation rate in the Urban-Rural Residents' Basic Pension Insurance, although it may exert both promoting and inhibiting effects.
4) There is a positive correlation between the old-age dependency ratio and the participation rate in the Urban-Rural Residents' Basic Pension Insurance, despite its dual effects.

4. Data Description

4.1. Data Collection and Preparation

This study focuses on 31 provinces, municipalities, and autonomous regions in mainland China, excluding data from Taiwan, Hong Kong, and Macao. The analysis is based on annual panel data from 2017 to 2022. The sample size is calculated as follows: number of entities (n) = 31, time span (T) = 6, total sample size ($n \times T$) = 186. This satisfies the conditions $n \times T > 180$ and $n > T$, which fulfills the fundamental prerequisites for panel data analysis. The variables are detailed in Table 1, with all data sourced from the China Statistical Yearbook.

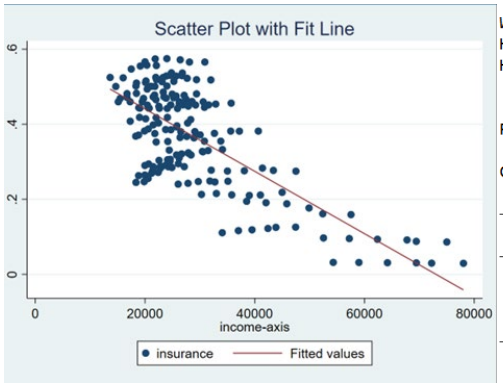
Table 1. Variable Details.

Variable Type	Variable Name	Definition	Data Source
\	year	year	\
\	province	province	\
Dependent Variable	insurance	Ratio of Urban-Rural Residents' Basic Pension Insurance participants to regional population	China Statistical Yearbook
Control Variable 1	income	Per capita disposable income by region (yuan)	China Statistical Yearbook
Control Variable 2	education	Local fiscal education expenditure (billion yuan)	China Statistical Yearbook
Control Variable 3	health	Local fiscal healthcare expenditure (billion yuan)	China Statistical Yearbook
Control Variable 4	ratio	Old-age dependency ratio (%)	China Statistical Yearbook

4.2. Descriptive Data Analysis

4.2.1. Participation Rate in Basic Pension Insurance and Per Capita Disposable Income

As shown in Figure 1, within the per capita disposable income range of 20,000 to 40,000 yuan, the participation rate in the Urban-Rural Residents' Basic Pension Insurance is relatively high but exhibits significant dispersion and variability. This study suggests the following possible reasons: Among individuals with an annual disposable income of around 20,000 yuan, some may engage in repetitive manual labor, have lower education levels, and possess weaker risk awareness, leading them to invest minimally in insurance. However, within this income group, those with stronger risk awareness but limited risk resilience may increase their investment in pension insurance. For individuals with annual disposable incomes exceeding 40,000 yuan, they typically have greater risk awareness and financial capacity, making them more likely to opt for higher-tier insurance products, resulting in a lower participation rate in basic pension insurance.




```
. estat imtest,white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

    chi2(2) =   4.42
Prob > chi2 = 0.1095

Cameron & Trivedi's decomposition of IM-test
```

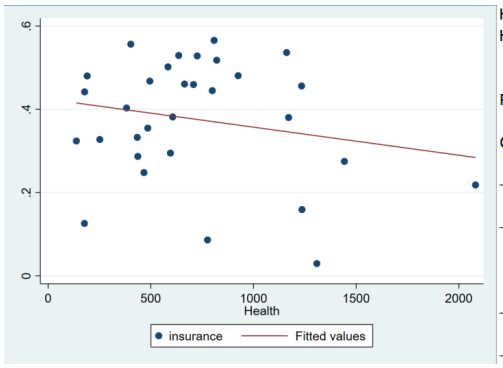
Source	chi2	df	p
Heteroskedasticity	4.42	2	0.1095
Skewness	2.62	1	0.1056
Kurtosis	9.14	1	0.0025
Total	16.18	4	0.0028

Figure 1. Participation Rate in Basic Pension Insurance vs. Per Capita Disposable Income.

Overall, the participation rate in basic pension insurance shows an inverse relationship with per capita disposable income, consistent with this study's hypothesis. However, the bifurcated data distribution suggests the presence of heteroscedasticity. A White Test was conducted with the null hypothesis of homoscedasticity; the results showed a p-value less than 0.01, rejecting the null hypothesis and confirming the presence of heteroscedasticity. To address this, the study employs heteroscedasticity-robust standard errors and, considering potential spatial correlation among sampling units, further adjusts using clustered standard errors.

4.2.2. Participation Rate in Basic Pension Insurance and Local Fiscal Healthcare Expenditure

Figure 2 indicates a negative correlation between local fiscal healthcare expenditure and the participation rate in the Urban-Rural Residents' Basic Pension Insurance. This may stem from the substitution effect within the dual mechanism: Increased public healthcare spending reduces residents' medical burdens, potentially decreasing their reliance on and demand for pension insurance. This substitution effect appears to outweigh the income effect, resulting in a decline in the participation rate in pension insurance, which suggests that increased public healthcare spending reduces the need for basic pension insurance. This finding contradicts intuitive expectations, possibly due to the lack of a clear pattern in the distribution of the scatter plot data, necessitating further validation through regression analysis.




```
. estat imtest,white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

    chi2(2) =   3.32
Prob > chi2 = 0.1906

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	3.32	2	0.1906
Skewness	19.06	1	0.0000
Kurtosis	3.63	1	0.0567
Total	26.01	4	0.0000

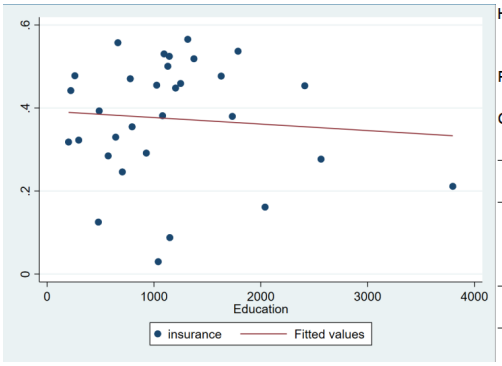
Figure 2. Participation Rate in Basic Pension Insurance vs. Local Fiscal Healthcare Expenditure.

Additionally, the degree of dispersion in the plot varies with changes in healthcare expenditure, suggesting potential heteroscedasticity. A White Test was performed with

the null hypothesis of homoscedasticity; the results returned a p-value less than 0.00001, rejecting the null hypothesis and confirming heteroscedasticity. Consequently, this study uses heteroscedasticity-robust standard errors to enhance model robustness.

4.2.3. Participation Rate in Basic Pension Insurance and Local Fiscal Education Expenditure

As depicted in Figure 3, there is a negative correlation between the participation rate in the Urban-Rural Residents' Basic Pension Insurance and local fiscal education expenditure. This may be due to the fact that higher education expenditure often reflects higher education levels, leading to greater financial literacy and risk awareness, which drives residents to invest in higher-tier insurance products instead of relying on basic pension insurance. A possible explanation is that higher education expenditure often reflects greater cultural literacy and education levels among residents. Those with higher education tend to possess stronger financial literacy and risk awareness, preferring higher-tier, non-basic insurance products to meet diverse protection needs. As a result, demand for basic pension insurance decreases, reducing its participation rate.




```
. estat imtest,white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

    chi2(2) =   3.32
Prob > chi2 = 0.1906

Cameron & Trivedi's decomposition of IM-test
```

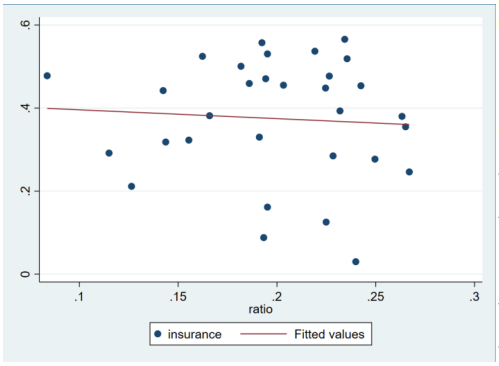
Source	chi2	df	p
Heteroskedasticity	3.32	2	0.1906
Skewness	19.06	1	0.0000
Kurtosis	3.63	1	0.0567
Total	26.01	4	0.0000

Figure 3. Participation Rate in Basic Pension Insurance vs. Local Fiscal Education Expenditure.

The scatter plot shows varying dispersion with changes in education expenditure, suggesting the presence of heteroscedasticity. A White Test confirmed this, with a p-value less than 0.00001, well below conventional significance levels, rejecting the null hypothesis of homoscedasticity. Thus, heteroscedasticity-robust standard errors are used to improve the robustness of the model estimates.

4.2.4. Participation Rate in Basic Pension Insurance and Old-Age Dependency Ratio

Due to the unclear trend in the overall scatter plot, this study analyzes cross-sectional data from 2021. As shown in Figure 4, the participation rate in basic pension insurance exhibits a negative correlation with the old-age dependency ratio, contrary to intuitive expectations. This may be due to the dual effect of the old-age dependency ratio, where the negative effect of increased family burdens outweighs the positive effect of heightened demand for pension insurance, leading to reduced participation in basic pension insurance. Specifically, a higher old-age dependency ratio significantly increases demand for commercial life insurance, and this growth exceeds the suppression of demand for commercial insurance due to reduce per capita disposable income. Consequently, residents favor commercial life insurance over basic pension insurance. This result may also reflect suboptimal data representation in the scatter plot, requiring further validation through regression analysis.




```
. estat imtest,white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

    chi2(2) =   6.81
Prob > chi2 = 0.0333

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	6.81	2	0.0333
Skewness	22.87	1	0.0000
Kurtosis	3.92	1	0.0478
Total	33.59	4	0.0000

Figure 4. Participation Rate in Basic Pension Insurance vs. Old-Age Dependency Ratio.

The plot's dispersion varies with the old-age dependency ratio, suggesting heteroscedasticity. A White Test yielded a p-value less than 0.0001, rejecting the null hypothesis of homoscedasticity. Thus, heteroscedasticity-robust standard errors are applied to enhance the robustness of the model estimates.

4.3. Descriptive Statistics of Variables

The descriptive statistics in Table 2 show that the standard deviations and ranges of the variables are relatively large, indicating a broad data distribution with good representativeness. However, skewness values suggest that the variables exhibit varying degrees of left or right skew, indicating insufficient symmetry in the data distribution. Further analysis of kurtosis reveals that, except for the "participation rate in basic pension insurance" and "old-age dependency ratio", the kurtosis of other variables exceeds the normal distribution value of 3. Notably, "per capita disposable income by region" has a significantly higher kurtosis, suggesting a sharper distribution that deviates from normality. This departure from a normal distribution may affect the internal validity of this study and warrants attention in subsequent analyses.

Table 2. Descriptive Statistics of Variables.

Variable	N	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
Participation rate in basic pension insurance	186	0.3637	0.1383	0.0295	0.5748	-0.6193	2.6022
Per capita disposable income by region	186	29264.49	12249.59	13639.20	78026.60	1.9314	6.7565
Local fiscal education expenditure (billion yuan)	186	1071.465	692.09	122.21	3871.14	1.3975	5.6667
Local fiscal healthcare expenditure (billion yuan)	186	571.412	349.4	93.80	2081.25	1.3069	5.5326
Old-age dependency ratio (%)	186	0.1788	0.04627	0.0804	0.2877	0.0086	2.3098

5. Model Specification and Empirical Analysis

5.1. Selection of Regression Equation Function Form-Multiple Linear Regression

Initially, this study conducts a pooled cross-sectional regression analysis. In the regression model, per capita disposable income is selected as the primary explanatory variable, while local fiscal education expenditure, local fiscal healthcare expenditure, and the

old-age dependency ratio are included as control variables for a multiple linear regression analysis. The regression results are presented in Table 3:

Table 3. Multiple Linear Regression Results.

	1	2	3	4	5
VARIABLES	insurance	insurance	insurance	insurance	insurance
income	-8.31e-06***	-9.28e-06***	-9.22e-06***	-8.98e-06***	-9.71e-06***
	(5.65e-07)	(5.57e-07)	(5.53e-07)	(5.69e-07)	(5.40e-07)
ratio		0.791***			0.552***
		(0.147)			(0.147)
Health			0.000104***		0.000219***
			(1.94e-05)		(5.98e-05)
Education				4.00e-05***	-7.13e-05**
				(1.01e-05)	(2.95e-05)
Constant	0.607***	0.494***	0.574***	0.584***	0.501***
	(0.0179)	(0.0268)	(0.0178)	(0.0182)	(0.0256)
Summary statistics					
Observations	186	186	186	186	186
R-squared	0.541	0.603	0.603	0.577	0.651

Standard errors in parentheses.
*** p < 0.01, ** p < 0.05, * p < 0.1.

1) Simple Linear Regression (Column 1)

The model in Column (1) presents the results of a simple linear regression with per capita disposable income. The findings show that for each additional unit increase in income (yuan), the participation rate in basic pension insurance decreases by an average of 8.31×10^{-6}, with the slope coefficient significant at the 1% level ($p < 0.01$) and the intercept also significant. This suggests that as income rises, people may invest more in alternative insurance products or prefer higher-tier plans, which reduces the reliance on basic pension insurance. The R^2 value of 0.541 indicates that per capita disposable income explains 54.1% of the variation in the participation rate. To mitigate potential omitted variable bias, additional control variables are introduced step by step to examine their effect on the income coefficient.

2) Adding Old-Age Dependency Ratio (Column 2)

The model in Column (2) incorporates the old-age dependency ratio as a control variable alongside per capita disposable income. The slope coefficient for income decreases slightly from -8.31×10^{-6} to -9.28×10^{-6}, and the R^2 value rises to 0.603, suggesting improved explanatory power. The coefficient for the old-age dependency ratio is 0.791 and significant at the 1% level ($p < 0.01$), indicating that an increased old-age dependency ratio positively affects the participation rate. Thus, this variable is retained in subsequent models.

3) Adding Local Fiscal Healthcare Expenditure (Column 3)

The model in Column (3) further includes local fiscal healthcare expenditure. The income coefficient drops to -9.22×10^{-6}, with the R^2 value remaining at 0.603. The healthcare expenditure coefficient is 0.000104 and significant at the 1% level ($p < 0.01$), showing a positive impact on the participation rate. This variable is also retained as a control.

4) Adding Local Fiscal Education Expenditure (Column 4)

The model in Column (4) adds local fiscal education expenditure. The income coefficient continues to decline, and the R^2 value increases, reflecting enhanced explanatory power. However, the education expenditure coefficient is positive (0.00004), contrary to the expected negative relationship, though significant at the 1% level ($p < 0.01$). This discrepancy suggests a possible misspecification of the model, which is addressed in further analyses.

5) Full Variable Regression (Column 5)

Column (5) includes all control variables, yielding an R^2 value of 0.651, indicating strong explanatory power. The coefficients for income, healthcare expenditure, and the old-age dependency ratio remain significant at the 1% level, and their signs are consistent with the study's hypotheses, supporting the proposed relationships. However, the education expenditure coefficient becomes insignificant, suggesting its effect on the participation rate is unclear. Given the model's explanatory power and variable significance, this study opts to abandon the multiple linear models in favor of exploring nonlinear models for further optimization.

5.2. Selection of Regression Equation Function Form-Nonlinear Regression

Model selection and result analysis are as follows:

1) Linear-Log Model (Model 2)

Column (2) presents the linear-log model, with explanatory variables (income, healthcare expenditure, and old-age dependency ratio) logged, while the dependent variable remains linear. The coefficients for income, healthcare expenditure, and the old-age dependency ratio are significant at the 1% level, consistent with the hypotheses proposed in the study. Education expenditure is significant at the 5% level. The R^2 value of 0.672 indicates improved fit compared to the linear model (Model 1), though room for enhancement remains.

2) Log-Linear Model (Model 3)

Column (3) shows the log-linear model, where the dependent variable is logged, and explanatory variables remain linear. The coefficients for income and the old-age dependency ratio are significant at the 1% level, with an R^2 value of 0.777 — the highest among the models. However, the coefficients for healthcare and education expenditure are insignificant, suggesting that these variables do not have a significant effect in the log-linear specification. However, healthcare and education expenditure coefficients are insignificant, leading to the rejection of this model due to the lack of significance in key control variables.

3) Double-Log Model (Model 4)

Column (4) presents the double-log model, where all variables are logged. The coefficients for income, healthcare expenditure, and the old-age dependency ratio are significant at the 1% level, but education expenditure is insignificant. The R^2 value of 0.703 is higher than Model 2 but offers limited improvement. Thus, Model 2 is chosen over Model 4 as the base model (Table 4).

Table 4. Nonlinear Regression Results.

	1	**2**	**3**	**4**
VARIABLES	insurance	insurance	lninsurance	lninsurance
income	-9.71e-06***		-4.72e-05***	
	(5.40e-07)		(1.89e-06)	
ratio	0.552***		2.106***	
	(0.147)		(0.516)	
Health	0.000219***		0.000333	
	(5.98e-05)		(0.000209)	
Education	-7.13e-05**		2.91e-05	
	(0.0256)		(0.000103)	
lnincome		-0.351***		-1.600***
		(0.0190)		(0.0793)
lnratio		0.0879***		0.312***
		(0.0259)		(0.108)
lnHealth		0.122***		0.324***

		(0.0279)		(0.116)
lnEducation		-0.0588**		-0.0817
		(0.0252)		(0.105)
Observations	186	186	186	186
R-squared	0.651	0.672	0.777	0.703

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

6. Consideration of Interaction Terms

Introducing interaction terms could add complexity to model interpretation. With multiple interactions, results may become more intricate and less intuitive, complicating policy analysis, as interpreting them would require a deeper economic understanding. Interaction effects imply assigning weights to each variable, which may be difficult to derive directly from the data. Without strong theoretical justification, adding interaction terms may not be necessary.

7. Polynomial Regression Attempt

Building on the double-log model, this study explores polynomial regression. After several regressions, the squared term of logged education expenditure ($lnEducation^2$) outperforms other polynomial forms, suggesting its better fit in capturing the non-linear effects of education expenditure on participation. In the final linear-log polynomial model, the p-value for lnEducation is less than 0.34, showing improved significance. Thus, this study adopts the linear-log polynomial model for analysis.

7.1. Pooled Cross-Sectional Analysis

7.1.1. Model Overview

The adopted model is a linear-log polynomial regression (Table 5), with 186 observations. The F-statistic is 44.89, with degrees of freedom (4,30) and a p-value of 0.0000, indicating overall statistical significance. The R^2 value of 0.6709 indicates that the explanatory variables explain 67.1% of the variation in the dependent variable. The root mean squared error (Root MSE) of 0.08027 reflects the average error between predicted and actual values.

Table 5. Nonlinear Regression Results with Polynomial Term.

VARIABLES	insurance
lnincome	-0.350***
	(0.0191)
lnratio	0.0843***
	(0.0265)
lnHealth	0.120***
	(0.0294)
lnEducation2	-0.00429**
	(0.00201)
Constant	3.546***
	(0.247)
Observations	186
R-squared	0.671

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

7.1.2. Variable Coefficients and Statistical Significance

1) lnincome (Log of Per Capita Disposable Income)

-Coefficient: -0.3498679

-Standard Error: 0.0317449
-t-value: -11.02
-P > |t|: 0.000
-95% CI: [-0.4146996, -0.2850362]

A significant negative relationship indicates that higher per capita disposable income reduces the participation rate in basic pension insurance. Higher income levels reduce demand for basic pension insurance, possibly as higher-income groups prefer advanced insurance products.

2) lnratio (Log of Old-Age Dependency Ratio)
-Coefficient: 0.0843391
-Standard Error: 0.0636632
-t-value: 1.32
-P > |t|: 0.195
-95% CI: [-0.0456785, 0.2143567]

The old-age dependency ratio shows a positive but statistically insignificant relationship with the participation rate. While theory suggests an increased ratio should boost demand, other factors (e.g., family support or commercial insurance substitution) may offset this effect. Reduced significance may indicate multicollinearity, requiring further investigation.

3) lnhealth (Log of Fiscal Healthcare Expenditure)
-Coefficient: 0.1199996
-Standard Error: 0.0297331
-t-value: 4.04
-P > |t|: 0.000
-95% CI: [0.0592765, 0.1807226]

A significant positive relationship exists, suggesting that improved healthcare coverage reduces medical burdens, increasing demand for pension insurance.

4) lneducation² (Squared Log of Fiscal Education Expenditure)
-Coefficient: -0.0042874
-Standard Error: 0.0020629
-t-value: -2.08
-P > |t|: 0.046
-95% CI: [-0.0085005, -0.0000744]

The squared log of education expenditure shows a negative relationship, significant at the 5% level, suggesting that higher education levels may drive residents to prefer non-basic insurance products, thus reducing demand for basic pension insurance.

5) Constant (_cons)
-Coefficient: 3.54632
-Standard Error: 0.4273865
-t-value: 8.30
-P > |t|: 0.000
-95% CI: [2.67348, 4.41916]

7.2. Panel Regression Analysis

Given that factors such as regional history, culture, industrial structure, and fiscal expenditure vary across regions but remain relatively stable over short time periods, this study introduces individual fixed effects to control for potential omitted variable bias. Macroeconomic factors like interest rates and inflation vary over time but not across regions, so time fixed effects are introduced to control for their impact. Based on this, the study employs dummy variables and individual centering methods for fixed effects estimation, using clustered standard errors for panel data regression analysis.

7.2.1. Individual Fixed Effects Estimation with Dummy Variables

$$F(4, 30) = 895.39$$
$$Prob > F = 0.0000$$

The results show significant individual fixed effects (Table 6) with a high fit, but the coefficient for income is insignificant. This could suggest that income does not have a strong or consistent relationship with the dependent variable across different regions, potentially due to regional differences or omitted factors not accounted for. Further exploration with time fixed effects reveals significant effects and a significant income coefficient at the 1% level (Table 7), prompting the adoption of time fixed effects.

Table 6. Regression Results with Individual Fixed Effects.

VARIABLES	insurance
lnincome	0.045
	(0.034)
Constant	0.009
	(0.393)
Observations	186
R-squared	0.985
province FE	Yes
Time FE	No

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

Table 7. Regression Results with Time Fixed Effects.

VARIABLES	insurance
lnincome	-0.391***
	(0.028)
Constant	3.876***
	(0.411)
Observations	186
R-squared	0.759
province FE	No
Time FE	Yes

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

A combined time and individual fixed effects model show

$$F(5, 30) = 46.98$$
$$Prob > F = 0.0000$$

that the lnincome coefficient is not statistically significant, despite a high R^2 value and good fit. Therefore, the time fixed effects model is selected as its better controls for time-varying factors, leading to more reliable estimates (Table 8).

Table 8. Regression Results with Time and Individual Fixed Effects.

VARIABLES	Insurance
lnincome	-0.123
	(0.091)
Constant	1.815*
	(1.005)
Observations	186
R-squared	0.986
province FE	Yes

Time FE	Yes

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

7.2.2. Fixed Effects Regression Using Stata's Panel Data Commands

At the preliminary stage, individual centering was attempted (Table 9). However, the results were unsatisfactory: the lnincome coefficient was insignificant, other control variables showed low significance, and the R^2 value was low, indicating poor model fit. Thus, individual centering was abandoned in favor of alternative methods to improve fit and significance.

Table 9. Fixed Effects Regression Results.

VARIABLES	Insurance
lnincome	0.045
	(0.031)
lnratio	0.084**
	(0.038)
lnEducation2	-0.001
	(0.001)
lnHealth	-0.036*
	(0.019)
Constant	0.309
	(0.336)
Observations	186
Number of province	31
R-squared	0.332
province FE	Yes
Time FE	No

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

Next, individual centering with time dummy variables was tested (Table 10). However, the lnincome coefficient remained insignificant, and the R^2 value was low, indicating poor fit. This approach was discarded in favor of a more suitable model.

Table 10. Regression Results with Individual Centering and Time Dummies.

VARIABLES	Insurance
lnincome	-0.123
	(0.083)
Constant	1.972**
	(0.846)
Observations	186
Number of province	31
R-squared	0.356
province FE	Yes
Time FE	Yes

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

In conclusion, after evaluating various model forms, this study adopts the linear-log polynomial regression model with time fixed effects, as it provides the best fit and robustness in controlling for heteroscedasticity and time-varying effects. This choice balances coefficient significance, model fit, and robustness to heteroscedasticity and time effects,

ensuring that the model accounts for both unobserved regional variations and time-varying macroeconomic factors (Table 10). Time fixed effects effectively control for unobserved time-varying factors, enhancing the accuracy and reliability of the estimates.

8. Model Evaluation and Analysis

8.1. Summary of Tables

The tables are summarized in Table 11.

Table 11. Model Summary.

	1	2	4	5	6
VARIABLES	**Insurance**	**Insurance**	**Insurance**	**Ininsurance**	**Ininsurance**
income	-9.71e-06***			-4.72e-05***	
	-5.40E-07			-1.89E-06	
ratio	0.552***			2.106***	
	-0.147			-0.516	
Health	0.000219***			0.000333	
	-5.98E-05			-0.000209	
Education	-7.13e-05**			2.91E-05	
	-0.0256			-0.000103	
lnincome		-0.351***	0.045		-1.600***
		-0.019	-0.031		-0.0793
lnratio		0.0879***	0.084**		0.312***
		-0.0259	-0.038		-0.108
lnHealth		0.122***	-0.001		0.324***
		-0.0279	-0.001		-0.116
lnEducation		-0.0588**	-0.036*		-0.0817
		-0.0252	-0.019		-0.105
Observations	186	186	186	186	186
R-squared	0.651	0.672	0.332	0.777	0.703

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

8.2. Threats to Internal Validity

8.2.1. Omitted Variable Bias

Beyond the target variable (per capita disposable income) and the three control variables (local fiscal education expenditure, local fiscal healthcare expenditure, and old-age dependency ratio) included in the model, other potential factors — such as residents' risk awareness or local policies — may also influence the participation rate in the Urban-Rural Residents' Basic Pension Insurance. These factors are difficult to quantify and measure, making their direct inclusion in the model impractical. Additionally, the final model does not incorporate time fixed effects, leaving it unable to control for unobserved variables that vary over time or across both individuals and time. As a result, omitted variable bias may remain, though other methods could be explored in future studies to mitigate this issue. Future studies could mitigate this by introducing instrumental variables to reduce the impact of omitted variable bias.

8.2.2. Functional Form Misspecification

Throughout the study, we tested multiple linear regression models, three nonlinear regression models, and incorporated individual fixed effects, time fixed effects, and interaction terms. After extensive trials and comparisons, we selected the linear-log polynomial model with time fixed effects, achieving an R^2 value of 0.759 and a satisfactory fit. Other functional forms performed less effectively in terms of fit and variable significance.

Therefore, we believe the likelihood of functional form misspecification in the chosen model is minimal.

8.2.3. Measurement Error in Variables

The data used in this study are sourced from the *China Statistical Yearbook*, provided by the National Bureau of Statistics, a widely recognized and credible source. Consequently, we consider the model to be largely free of measurement errors in the variables.

8.2.4. Sample Selection Bias

The dataset covers the participation rate in the Urban-Rural Residents' Basic Pension Insurance, old-age dependency ratio, per capita disposable income, local fiscal healthcare expenditure, and local fiscal education expenditure across 31 provincial-level administrative regions from 2017 to 2022. Potential sources of sample selection bias include:

Regional Economic Development Disparities: Differences in economic development levels across provinces may lead to variations in pension insurance coverage and per capita income, affecting the analysis results.

Differences in Policy Implementation: Variations in the enforcement of pension policies by local governments could introduce systematic bias during sample selection.

Inconsistencies in Data Collection and Reporting Standards: Although the data come from the same statistical yearbook, inconsistencies in data collection and processing may introduce errors.

If suitable instrumental variables (i.e., variables that affect the participation rate but are not influenced by it) can be identified, the instrumental variable method could be employed to address potential endogeneity issues.

8.2.5. Bidirectional Causality

In this study, the explanatory variable is per capita disposable income, the dependent variable is the participation rate in the Urban-Rural Residents' Basic Pension Insurance, and the control variables include local fiscal education expenditure, local fiscal healthcare expenditure, and the old-age dependency ratio. Bidirectional causality may exist between the following variables:

Local Fiscal Healthcare Expenditure: An increase in the participation rate could boost national tax revenue, thereby increasing local healthcare expenditure, suggesting a reverse positive correlation.

Old-Age Dependency Ratio: A bidirectional causal relationship may also exist between the old-age dependency ratio and the participation rate.

However, this study did not employ instrumental variables for further analysis, leaving this issue unresolved and in need of refinement.

8.2.6. Solutions and Approaches

To address omitted variable bias and bidirectional causality, this study plans to introduce instrumental variables in future research to satisfy the assumptions $E(u|x) = 0$ or $E(u|x,w) = E(u|w)$.

8.3. Optimal Model Selection

Based on the above analysis, the optimal overall regression model selected in this study is presented in Table 12.

Table 12. Parameter Interpretation.

VARIABLES	Explanation
β_1	he coefficient of per capita disposable income (yuan) on the participation rate (%). For every 1% increase in income, the participation rate decreases by β_1% (0.391%).
β_2	The coefficient of the old-age dependency ratio on the participation rate (%). For every 1% increase in the old-age dependency ratio, the participation rate changes by β_2% (0.033%).
β_3	The coefficient of local fiscal healthcare expenditure (billion yuan) on the participation rate (%). For every 1% increase in healthcare expenditure, the participation rate increases by β_3% (0.087%).
β_4	The coefficient of the squared log of local fiscal education expenditure (billion yuan) on the participation rate (%). For every 1% increase in education expenditure, the participation rate decreases by β_4% (0.002%).
Adjusted-R 2	The model explains 92.8% of the variation in the participation rate.

$$\text{insurance} = \alpha_i + \beta_0 + \frac{\beta_1}{\text{nincome}} + \frac{\beta_2}{\text{nratio}} + \frac{\beta_3}{\text{nHealth}} + \frac{\beta_4}{\text{nEducation2}}$$

The standard functional form of the sample regression line is as follows:

$$\text{insurance} = \alpha_i + 3.876 - 0.391 \cdot \text{nincome} + 0.033 \cdot \text{nratio} + 0.087 \cdot \text{nHealth} - 0.002 \cdot \text{InEducation2}$$

The standard errors for each coefficient are:

The standard error for the constant term (3.876) is 0.224.

The standard error of the coefficient (-0.391) for ln(income) is 0.014.

The coefficient for Inratio (0.033) has a standard error of 0.031.

The coefficient for InHealth (0.087) has a standard error of 0.030.

The coefficient for InEducation2 (-0.002) has a standard error of 0.002.

This model strikes a good balance between explanatory power, variable significance, and robustness, making it the most suitable choice for this study.

9. Conclusion

This study utilized panel data from 31 provinces, municipalities, and autonomous regions in China spanning 2017 to 2022 to construct a linear-log polynomial model and conduct an empirical analysis. The following conclusions were drawn:

1) Relationship Between Per Capita Disposable Income and Participation Rate in Urban-Rural Residents' Basic Pension Insurance

There is a significant negative correlation between per capita disposable income and the participation rate in the Urban-Rural Residents' Basic Pension Insurance. Specifically, as per capita disposable income increases, residents' willingness to participate in this pension insurance tends to decrease. This phenomenon may be attributed to higher-income groups' preference for alternative insurance products or investment opportunities that better align with their risk management preferences.

2) Impact of Control Variables
   a) Local Fiscal Healthcare Expenditure: This has a positive effect on the participation rate in the Urban-Rural Residents' Basic Pension Insurance. An increase in healthcare expenditure raises the participation rate, likely because improved medical security reduces residents' healthcare burdens, thereby increasing their demand for pension insurance.
   b) Old-Age Dependency Ratio: The effect on the participation rate is not statistically significant, though it shows a general positive correlation. This suggests that while a higher proportion of elderly individuals may increase demand for pension insurance, this effect is not pronounced, possibly due to substitution by family support or commercial insurance.

c) Local Fiscal Education Expenditure: This exhibits a negative relationship with the participation rate, approaching statistical significance. This could indicate that residents in regions with higher education levels are more inclined to opt for alternative investment or savings options rather than pension insurance.

3) Model Selection and Fit

After multiple regression analyses, the linear-log polynomial model with time fixed effects was selected. This model achieved an R-squared value of 0.759, indicating a good fit to the data.

4) Policy Recommendations

a) Increase Fiscal Support: The government should enhance financial assistance for low-income groups, particularly by increasing spending on education and healthcare. This would not only boost pension insurance coverage among low-income populations but also promote social equity and stability.

b) Strengthen Public Education: Efforts should be made to raise awareness of pension insurance through various channels, such as media campaigns, community lectures, and school programs. Enhancing risk awareness, especially among high-income groups, can underscore the importance of pension insurance and encourage broader participation.

c) Optimize Policy Design: Tailored pension insurance policies should be developed based on regional differences to ensure fairness and effectiveness. For instance, in less economically developed areas, additional fiscal subsidies and incentives could encourage participation, while in developed regions, tax benefits could guide high-income groups toward basic pension insurance.

d) Promote a Multi-Tiered Pension System: Beyond the Urban-Rural Residents' Basic Pension Insurance, the government should actively develop a multi-tiered system, including enterprise annuities and personal savings-based pension insurance. Offering diverse products can meet the needs of different income groups, improving overall coverage and sustainability.

5) Future Outlook

a) Expand Sample Scope: Extending the study to include other regions, such as Hong Kong, Macao, and Taiwan, as well as other countries, would enhance the generalizability and international comparability of the findings. Incorporating more diverse data could provide a comprehensive understanding of pension insurance participation behaviors across varying social and economic contexts.

b) Incorporate Additional Control Variables: Beyond the current variables, factors such as cultural differences and policy enforcement strength could be included. These additions may improve the model's explanatory power, leading to a more thorough and nuanced analysis.

References

1. D. Sena and N. K. Nagwani, "Application of time series based prediction model to forecast per capita disposable income," in *Proc. 2015 IEEE Int. Adv. Comput. Conf. (IACC)*, Jun. 2015, pp. 454–457, doi: 10.1109/IADCC.2015.7154749.
2. G. Yang and L. Zhang, "Relationship between aging population, birth rate and disposable income per capita in the context of COVID-19," *PLoS One*, vol. 18, no. 8, p. e0289781, 2023, doi: 10.1371/journal.pone.0289781.
3. Y. Jia and H. Tang, "Modeling China's per capita disposable income by uncertain statistics," *Res. Square*, preprint, 2023, doi: 10.21203/rs.3.rs-3304016/v1.
4. Y. Ni, "The influence of digital economy on rural per capita disposable income: Evidence from China," *Environ. Dev. Sustain.*, pp. 1–22, 2024, doi: 10.1007/s10668-024-05308-8.

5. Z. Xu and M. Xue, "Research on prediction of per capita disposable income of rural residents in Shandong Province based on GM (1, 1) model," in *Proc. 2022 3rd Int. Conf. Comput. Sci. Manag. Technol. (ICCSMT)*, Nov. 2022, pp. 48–51, doi: 10.1109/ICCSMT58129.2022.00017.