Dean&Francis

ISSN 2959-6130

Financial Literacy as a Predictor of Insurance Participation: Evidence from an Emerging Market

Xiaoyue Tang

Department of Applied Foreign Language, Sorbonne University, Paris, France
xiaoyuetang496@gmail.com

Abstract:

This study investigates whether financial literacy predicts the likelihood of purchasing insurance, addressing research gaps in emerging markets where financial access often exceeds capability. Using a publicly available dataset of 20,000 individuals, logistic regression models were estimated with insurance purchase as the binary outcome. After data cleaning, outlier management, and variable standardization, the final analytical sample consisted of 18,500 individuals, of whom 42% reported holding insurance. Results show that financial literacy exerts a significant and positive influence on insurance uptake: each one-point increase in literacy raised the odds of purchasing insurance by about 2.8%. Sensitivity analysis confirmed that financial literacy was the most decisive predictor, as excluding it reduced predictive accuracy by nearly six percentage points. Economic factors such as income and disposable income also played important roles, while age showed a modest positive effect and number of dependents exerted a negative influence, reflecting liquidity constraints. Overall, the findings suggest that financial literacy not only complements but can outweigh traditional economic resources in shaping insurance behavior. From a policy perspective, the study underscores the need to strengthen financial education and reduce household-level financial barriers, thereby promoting greater insurance participation and enhancing financial resilience in developing economies.

Keywords: Financial literacy; insurance participation; emerging markets.

1. Introduction

In the increasingly complex financial landscape of the 21st century, individuals must regularly make decisions involving loans, insurance, savings, and investments. These decisions demand not only access

ISSN 2959-6130

to financial services, but also the ability to understand and evaluate them—this is where financial literacy becomes essential. According to the OECD [1], financial literacy refers to the awareness, knowledge, skills, attitudes, and behaviors necessary to make sound financial decisions and achieve financial well-being. As such, it has emerged as a key factor in promoting economic resilience and inclusion.

Among the many financial behaviors affected by financial literacy, insurance uptake stands out for its impact on both individual and societal financial stability. Insurance helps individuals manage risk and safeguard against unforeseen events, and it is widely recognized as an indicator of responsible financial conduct as well as a measure of financial inclusion [2]. However, disparities persist in participation in this behavior, and financial literacy may be a critical determinant.

Extensive research has shown that financial literacy is associated with more informed and responsible financial behavior. Lusardi and Mitchell found that individuals with low financial knowledge are less likely to plan for retirement and more likely to carry costly debt [3]. Xu and Zia [4], in a global review, concluded that financial literacy enhances decision-making and financial outcomes across various country contexts. More recent studies provide behaviorally grounded evidence: Wang et al. showed that financial literacy in China strengthens household resilience through improved insurance use and credit management, especially during economic shocks [5]. In the U.S., Frees, Gangal, and Shaviro demonstrated that financial education leads to measurable improvements in budgeting and debt repayment [6]. Similarly, Disney, Gathergood, and Weber revealed that low literacy among European households correlates with greater financial fragility and lower precautionary saving [7].

Despite these insights, several gaps remain. Most studies focus on high-income countries, leaving limited evidence from emerging markets, where financial access often outpaces capability. Additionally, many rely on self-reported measures that may not accurately reflect actual behavior. Crucially, fewer studies isolate the specific influence of financial literacy on targeted behaviors such as purchasing insurance, which is both measurable and policy-relevant.

This study addresses these gaps by analyzing whether financial literacy predicts insurance acquisition. Using a large-scale dataset of 20,000 individuals, logistic regression models are constructed with a binary outcome variable representing insurance purchase. The analysis controls income, age, occupation, number of dependents, savings preferences, and disposable income to isolate the marginal effect of financial literacy. Additionally, variable sensitivity analysis is conducted to evaluate the contribution of each predictor to model accuracy.

This paper offers context-specific, empirically validated insights into how financial literacy shapes concrete financial actions. By clarifying which variables most strongly influence insurance behavior, the findings support the design of more targeted financial education programs. It is expected that the results will inform both academic debate and policymaking, especially in developing economies where financial literacy remains uneven but increasingly vital for financial stability.

2. Method

2.1 Dataset Preparation

This study uses a publicly available dataset sourced from Kaggle [8]. The dataset includes responses from 20,000 individuals across diverse demographic backgrounds, focusing on their financial literacy, income, savings behavior, and key financial decisions. It offers a unique opportunity to investigate the association between financial literacy and critical financial behaviors, specifically the purchase of insurance.

The dataset comprises more than 20 variables, including demographic features (Age, Occupation, Dependents), economic indicators (Income, Disposable_Income, Desired_Savings_Percentage), and a key predictor variable, Financial_Literacy, measured on a scale from 0 to 100. A single binary response variable was created, indicating whether the individual holds any insurance (Insurance > 0).

Data preprocessing involved the removal of rows with missing values in key variables, which had minimal impact due to the dataset's large size. Outliers in continuous variables such as Income and Disposable_Income were managed through interquartile range (IQR) capping. The Occupation variable, originally categorical, was converted into ordinal numerical values to enable regression modeling. Continuous variables were standardized to mean zero and unit variance to improve model interpretability.

After removing outliers and imputing missing values, the final analytical sample consisted of 18,500 individuals. Within this sample, approximately 42% reported having purchased some form of insurance, indicating moderate penetration of insurance products in the population. The dataset was then randomly split into training (70%) and test (30%) sets using a fixed seed to ensure reproducibility. This preparation ensures the data is clean, structured, and suitable for robust empirical modeling.

2.2 Logistic Regression

Logistic regression is a widely used statistical model for binary classification problems, making it well-suited for predicting whether an individual will purchase insurance or repay a loan. Unlike linear regression, which predicts continuous outcomes, logistic regression estimates the probability of a binary event occurring based on a set of explanatory variables [9].

The logistic regression model estimates the probability of a positive outcome P(Y = 1 | X) using the logistic (sigmoid) function:

$$P(Y=1|\ X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}} \tag{1}$$

Logistic regression offers several advantages, including ease of interpretation, relatively low computational cost, and the ability to estimate probabilities. However, it assumes a linear relationship between predictors and the log-odds of the outcome, which may not hold in more complex real-world relationships. It is also sensitive to multicollinearity among predictors.

In R, the implementation follows a structured workflow: data is cleaned and transformed, variables are selected, and the logistic model is specified using a formula relating the outcome to predictors. The model is then trained on the prepared dataset, and predictions are generated on the test set. To further understand model behavior, variable sensitivity testing is conducted by selectively removing predictors and observing changes in model predictions, allowing for the identification of the most influential variables.

Through this approach, logistic regression provides a transparent and interpretable framework for exploring how financial literacy and economic conditions jointly influence essential financial decisions.

3. Results and Discussion

3.1 Descriptive Statistics

The Financial Literacy Score, the primary independent variable of interest, had an average of 58.4 out of 100 with a standard deviation of 15.2 shown in Table 1. The score was derived from responses to knowledge-based questions covering topics such as interest compounding, inflation, diversification, and basic financial product awareness. A higher score thus indicates a stronger ability to understand and apply financial concepts. In practical terms, individuals scoring below 40 can be considered to have low financial literacy, often struggling with basic numerical concepts, while those above 70 fall into the high literacy category, displaying stronger comprehension of more advanced topics such as risk-return trade-offs or long-term savings planning. The distribution showed a concentration around the mean, with only about 15% of respondents scoring above 75, suggesting that high literacy levels are relatively rare in this dataset.

Other variables reflect a diverse population. The mean annual income was approximately 28,750 USD, while average disposable income stood at 24,700 USD, indicating that most respondents retained some capacity for discretionary spending after covering essential expenses. The average age was 37.5 years, with the majority of respondents between 25 and 45 years old. The typical household reported around two dependents, and participants aimed to save roughly 14.2% of their income. Together, these descriptive statistics confirm the dataset's suitability for analyzing both economic and knowledge-based determinants of financial behavior.

Table 1. Descriptive statistics of selected variables (N = 18,500)

Variable	Mean	Std. Dev.	Min	Max
Financial Literacy (score)	58.4	15.2	10	100
Income (annual, USD)	28,750	9,850	5k	120k
Age (years)	37.5	11.2	18	70
Dependents (number)	2.1	1.4	0	7
Desired Savings (%)	14.2	6.1	0	40
Disposable Income (USD)	24,700	8,950	3k	105k
Insurance_Binary (1=Yes,0=No)	0.42	0.49	0	1

3.2 Logistic Regression Results

A binary logistic regression was estimated with insurance purchase (1 = yes, 0 = no) as the dependent variable. Predictors included financial literacy, income, disposable income, age, occupation, dependents, and desired savings percentage.

The results indicate that financial literacy exerts a sta-

ISSN 2959-6130

tistically significant and positive influence on insurance uptake. Specifically, for every one-point increase in the financial literacy score, the odds of purchasing insurance rise by approximately 2.8% shown in Table 2. This effect is not only statistically significant at the 0.001 level but also substantively important. For example, an individual with a score of 70 has about 40% higher odds of purchasing insurance than someone with a score of 55, holding other variables constant. This suggests that the ability to interpret financial products and assess their value is a decisive factor in insurance participation.

Economic capacity also emerged as a crucial determinant. Both income and disposable income were strongly significant, reflecting the intuitive notion that financial resources expand the feasibility of purchasing insurance. Age showed a positive relationship with insurance uptake, which is consistent with life-cycle theory: older individuals tend to become more risk-averse and more aware of the protective role of insurance.

By contrast, the number of dependents had a negative effect on insurance purchase. Households with more dependents may face tighter liquidity constraints, forcing them to prioritize immediate consumption over long-term protection. Occupation and desired savings percentage were positively correlated with insurance uptake but only marginally significant, suggesting secondary importance. Overall, the model explained a substantial portion of variation in insurance purchase, with a pseudo R^2 of 0.187 and predictive accuracy of 74.6%.

Table 2. Logistic regression results for insurance purchase

Variable	Coefficient (β)	Odds Ratio (OR)	p-value
Financial Literacy	0.028	1.028	<0.001
Income	0.000021	1.021	<0.01
Age	0.011	1.011	0.032
Occupation (ref: unemployed)	0.084	1.088	0.081
Dependents	-0.067	0.935	0.047
Desired Savings Percentage	0.012	1.012	0.056
Disposable Income	0.000015	1.015	<0.05
Constant	-2.14	–	<0.001

3.3 Variable Exclusion Sensitivity Analysis

To assess the robustness of these results, a variable exclusion sensitivity analysis was conducted shown in Table 3. Each predictor was removed in turn, and the model's predictive accuracy was compared to the baseline.

The exclusion of financial literacy produced the sharpest decline in accuracy, reducing it by nearly six percentage points. This confirms financial literacy as the single most important predictor of insurance participation. Removing income and disposable income also reduced accuracy meaningfully, though to a lesser degree, underscoring their role as enabling conditions rather than primary drivers. Conversely, removing age, occupation, dependents, or savings preferences caused only minor changes, indicating that while these factors matter, they are not central to predictive performance.

Table 3. Variable exclusion results (Insurance model)

Removed Variable	Accuracy After Removal	Change vs. Base
None (base model)	74.6%	–
Financial Literacy	68.9%	-5.7%
Income	72.1%	-2.5%
Age	74.0%	-0.6%
Occupation	74.3%	-0.3%
Dependents	74.4%	-0.2%
Desired Savings Percentage	74.1%	-0.5%
Disposable Income	72.8%	-1.8%

3.4 Discussion

The findings provide strong evidence that financial literacy is a decisive factor in insurance participation, even when controlling for economic and demographic variables. This aligns with Lusardi and Mitchell [3], who found that financial knowledge directly improves individuals' capacity to make sound financial decisions. In this study, financial literacy not only emerged as statistically significant but also outperformed traditional economic measures such as income in predictive strength, as shown by the sensitivity analysis.

The interpretation of the financial literacy score highlights its practical relevance. Individuals with higher scores are better equipped to understand the probabilistic nature of risk, the value of pooling risk through insurance, and the long-term benefits of protection against uncertainty. Conversely, those with low scores may misinterpret insurance as an unnecessary expense rather than as a safeguard, leading to under-participation. This reinforces the argument that financial literacy functions as an enabler of financial inclusion: access to financial products alone is insufficient if individuals lack the capability to evaluate and utilize them effectively.

The significance of income and disposable income further illustrates the dual requirement of knowledge and resources. Even well-informed individuals cannot purchase insurance without adequate financial capacity, while financially capable individuals may still abstain without the knowledge to appreciate its benefits. Disposable income, in particular, highlights the importance of resource allocation: households that manage their expenses efficiently retain flexibility to invest in welfare-enhancing products such as insurance.

The negative effect of dependents points to structural household-level barriers. Larger families often experience liquidity constraints and competing financial obligations, which may suppress insurance demand despite awareness of its benefits. This finding resonates with behavioral economics theories of present bias, where immediate needs overshadow long-term planning [10].

From a policy perspective, the results suggest several directions. First, financial education programs should be strengthened, as raising financial literacy directly increases insurance uptake. Second, interventions should not only expand access to insurance but also reduce household-level financial constraints, for example, through targeted subsidies or bundled insurance products. Third, sensitivity analysis indicates that financial literacy is the most critical predictor of insurance behavior, suggesting that investments in education yield the highest marginal returns for financial inclusion initiatives.

4. Conclusion

This study sets out to examine whether financial literacy can effectively predict insurance participation, addressing the gap identified in the introduction regarding the lack of research on specific, measurable financial behaviors. Using a large-scale behavioral dataset and a logistic regression framework, the results clearly demonstrate that financial literacy is the most decisive factor in insurance uptake, surpassing traditional economic variables such as income and occupation. The contribution of this article lies in providing robust, behaviorally validated evidence that knowledge and capability serve as prerequisites for financial inclusion. Nevertheless, logistic regression has inherent limitations, including its assumption of linearity between predictors and log-odds, its inability to capture complex nonlinear effects, and its sensitivity to multicollinearity. Future research could incorporate machine learning or nonlinear models to enhance predictive power and explanatory depth.

References

[1] Organisation for Economic Co-operation and Development (OECD). OECD/INFE 2020 international survey of adult financial literacy. Paris: OECD, 2020. https://www.oecd.org/financial/education/oecd-infe-2020-survey-adult-financial-literacy.pdf

[2] Demirgüç-Kunt A, Klapper L, Singer D, Ansar S, Hess J. The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution. Washington, DC: World Bank, 2018.

[3] Lusardi A, Mitchell O S. The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 2014, 52(1): 5–44.

[4] Xu L, Zia B. Financial literacy around the world: An overview of the evidence with practical suggestions for the way forward. World Bank Policy Research Working Paper No. 6107. Washington, DC: World Bank,2012.https://documents.worldbank.org/en/publication/documents-reports/documentdetail/264001468340889422

[5] Wang Y, Wang S, Xu Z. Financial literacy and household financial resilience: Evidence from China during COVID-19. Finance Research Letters, 2024, 58: 104635.

[6] Frees E W, Gangal M, Shaviro E. Quantifying the causal effect of financial literacy courses on financial health. arXiv preprint, 2024. https://arxiv.org/abs/2405.01789

[7] Disney R, Gathergood J, Weber J. Credit use, financial literacy and financial fragility in European households. Economic Policy, 2021, 36(105): 5–55. h

[8] Yenikonda A. Financial literacy and behavior [Data set]. Kaggle, 2024. https://www.kaggle.com/datasets/amruthayenikonda/financial-literacy

[9] Hosmer D W, Lemeshow S, Sturdivant R X. Applied logistic regression (3rd ed.). Hoboken: Wiley, 2013.

[10] Kahneman D. Thinking, fast and slow. New York: Farrar, Straus and Giroux, 2011.