Understanding Thailand's Credit Market: The Role of Age, Occupation, and Education

By Oliver Keyes-Krysakowski*

This study analyzes Thailand's "Million Baht Village Fund" program, assessing its impact on household borrowing across age, education, and occupation groups. Using a two-stage least squares (2SLS) approach with inverse village size as an instrument, we estimate the program's heterogeneous effects. Findings reveal that younger, more educated, and non-farming households exhibit stronger borrowing responses, highlighting the role of socio-economic characteristics in shaping credit utilization. The analysis extends prior work by Kaboski and Townsend, offering micro-level insights and underscoring the importance of targeted microfinance programs. Our results suggest that policy efforts aimed at these demographic groups can enhance financial inclusion and promote economic growth. Future work will apply causal machine learning methods and quantile regression to further refine the findings.
JEL: G21, O16, O53
Keywords: microfinance, credit market, socio-economic factors, Thailand, demographic analysis

This paper evaluates the short-term impact of Thailand's "Million Baht Village Fund" program on the credit market, focusing specifically on how the program's effects vary across household characteristics such as age, education, and occupation. Utilizing pre- and post-program panel data, we aim to answer a critical economic question: How do different socio-economic groups respond to the availability of microcredit? Understanding these heterogeneous responses is essential for designing more effective microfinance programs and financial policies that maximize economic growth and poverty alleviation.

Microfinance initiatives like the Million Baht Village Fund have been widely adopted as tools to alleviate poverty and stimulate economic development by providing credit access to underserved communities. However, the effectiveness of these programs is not uniform across all demographic groups. Factors such as age, educational attainment, and occupation significantly influence financial behavior and the capacity to utilize credit productively. For instance, younger individuals may have longer investment horizons and a greater willingness to take risks compared to older individuals. Similarly, higher education levels can enhance financial literacy, enabling better assessment of investment opportunities and risks. Occupational differences, such as between farmers and non-farmers,

* Oliver Keyes-Krysakowski: University of Toronto, Department of Economics, Toronto, ON, Canada. Email: oliver.keyeskrysakowski@mail.utoronto.ca.

can also affect credit needs and repayment capabilities. Therefore, exploring the heterogeneity in responses to microcredit availability is crucial for policymakers and financial institutions aiming to optimize the allocation of credit resources.

Our research builds upon the seminal work of Kaboski and Townsend's "The Impact of Credit on Village Economies" (Kaboski and Townsend, 2012*a*), which analyzed the aggregate effects of Thailand's Million Baht Village Fund Program. While their study provided valuable insights into the overall impact of the program on credit, consumption, and income growth, it did not delve into the differential effects across various socio-economic groups. We address this gap by investigating how the impact of the program on credit uptake and utilization differs among households categorized by age, education, and occupation. By doing so, we contribute to the literature by offering a more nuanced understanding of the micro-level dynamics that underlie the aggregate outcomes observed in previous studies.

Additionally, our study contributes to the broader literature on microfinance by addressing a critical gap in existing research: the heterogeneous effects of microcredit at the household level. Most prior work, including studies such as Mahjabeen's (Mahjabeen, 2008) analysis of microfinance in Bangladesh and Awojobi's (Awojobi, 2011) examination of microfinancing in Nigeria, has focused on the broader implications of microfinance on poverty reduction, welfare, and economic growth. However, these studies often overlook the nuanced impacts across different demographic groups. By segmenting our data based on household head characteristics such as age, education, and occupation, our approach provides a more granular understanding of how microcredit can be optimized to foster economic growth. This detailed perspective allows us to identify specific demographic factors that influence the effectiveness of microfinance programs, thereby enhancing the precision of policy interventions.

Moreover, our study extends the discourse on credit constraints and financial inclusion, with significant implications for the banking and credit lending sectors. By identifying which demographic groups are more likely to increase borrowing in response to credit availability, our findings offer valuable insights for traditional financial institutions. Microfinance programs not only increase borrowing among previously underserved populations but also have the potential to introduce new customers to the formal banking system, potentially expanding the customer base for traditional banks. While Vanroose and D'Espallier (2009) demonstrated that microfinance thrives where formal banking services are lacking, our analysis suggests that microfinance programs can complement the banking sector by revealing the specific characteristics of individuals most responsive to credit, thereby aiding banks in effectively targeting potential borrowers. This synergy between microfinance and traditional banking could lead to more inclusive financial systems and sustainable economic development.

Our findings reveal significant differences in how various household groups respond to the availability of village fund credit. Younger and more educated household heads who are non-farmers exhibit a stronger borrowing response compared to their counterparts. Specifically, younger households borrow significantly more new short-term credit than older households when village fund credit becomes available. Household heads with higher levels of education show a greater increase in borrowing, suggesting that education enhances the ability to leverage credit for productive purposes. In terms of occupation, non-farming households demonstrate a more substantial borrowing response than farming households, indicating that occupation type influences credit utilization patterns. However, while this trend points to potentially meaningful differences, the borrowing response among non-farming households is not statistically significant, suggesting that occupation-based credit behaviors may vary across contexts or require further study.

These results have important implications for both microfinance practitioners and policymakers. By identifying the demographic groups that are most responsive to microcredit availability, financial institutions can tailor their products and outreach strategies to better meet the needs of these populations. For policymakers, understanding the heterogeneity in credit responses can inform the design of more targeted and efficient microfinance programs, ultimately contributing to more effective poverty reduction and economic development strategies. Ultimately, our study suggests that microfinance programs like the Million Baht Village Fund can have a more substantial economic impact when targeted towards younger, more educated, and non-farming households.

I. Data

This study utilizes panel survey data derived from the Townsend Thai Project, a comprehensive dataset spanning the years 1997 to 2007 (Kaboski and Townsend, 2012*a*,*b*). The data was collected from rural villages in Thailand, focusing on household-level observations. The primary aim was to capture detailed economic and social information from households before and after the implementation of the Million Baht Village Fund program. This program was a major government initiative launched in 2001 to provide microcredit to rural villages. Each village was allocated a uniform amount of one million baht, regardless of its size, with the goal of fostering economic growth and alleviating poverty through increased access to credit. This uniform allocation introduced exogenous variation in the availability of credit per household, as smaller villages received more funds per household than larger ones. This unique feature creates a quasi-experimental setting that is instrumental for analyzing the program's impact across different household characteristics.

The dataset comprises a stratified, clustered random sample of 960 households, with 15 households selected from each of 64 villages. These villages are distributed across four provinces (changwats) in Thailand: Chachoengsao and Lopburi in the Central region, and Buriram and Sisaket in the Northeast region. The sample includes both rural and semi-urban areas, providing a diverse representation of

Thai households and enabling the examination of regional differences. Data collection was rigorous and consistent, with annual surveys conducted over eleven years—five years prior to the program's implementation (1997–2001) and six years after (2002–2007). The low attrition rate, averaging about three percent annually due to migration, resulted in a balanced panel of 800 households over the seven years following the program's introduction and 655 households for the entire eleven-year period.

The dataset is rich in variables related to household economic activities. It includes detailed information on income, borrowing and saving behaviors, assets, investments, consumption patterns, education levels, occupation, business operations, and household composition. This depth of information allows for a comprehensive analysis of how the Village Fund program influenced household behavior and economic outcomes.

Building on this rich dataset, we examine the specific impact of the Million Baht Village Fund program on household borrowing behavior. In particular, our study focuses on the new short-term credit level, which captures the amount of new loans taken by households with a term of one year or less. This variable provides insight into how households respond to the availability of credit in the immediate term following the program's implementation. Our primary explanatory variable is the current level of village fund credit, which represents the total amount borrowed by each household from the program. This allows us to assess how access to the fund influences further borrowing decisions.

To explore the heterogeneity in borrowing behavior, we decompose the explanatory variable across three key household characteristics: the age of the household head, years of education, and a farming dummy that indicates whether the household head's primary occupation is farming (1) or non-farming (0). By examining these variables, we can better understand how different socio-economic groups respond to village fund credit and whether certain groups—based on age, education, or occupation—are more or less likely to increase their short-term borrowing.

II. Variable Selection

To determine which household characteristics most effectively capture the heterogeneous effects of the Million Baht Village Fund program on short-term credit uptake, we utilized a variety of machine learning models as tools for variable selection. Specifically, we applied a decision tree, random forest, boosting, bagging, LASSO, and Ridge regression, each of which provided insights into the predictive importance of individual socio-economic variables. By employing these models in succession, we systematically evaluated which factors consistently emerged as influential in predicting household borrowing behavior.

Our analysis revealed that age and education were repeatedly highlighted as significant predictors across all models, underscoring their relevance in understanding borrowing responses to microcredit access. These variables not only indicated strong predictive power but also aligned with theoretical expectations;

younger individuals may have a longer investment horizon, while higher educational attainment likely enhances financial literacy and credit utilization efficiency.

Although other variables, such as the number of children, also showed high predictive importance, we ultimately chose occupation—specifically, the farming status of the household head (farmer vs. non-farmer)—as the focal variable. The decision to prioritize occupation was motivated by its potential to provide meaningful insights into the types of credit-driven opportunities available to households. Non-farming households may have more immediate or diverse investment options in non-agricultural sectors, potentially offering quicker or higher returns. Analyzing occupation as a moderating factor allows us to examine how economic opportunities influence borrowing behavior, a crucial aspect of our study's objective.

In the sections that follow, we discuss each model's role in the selection process and how their outcomes informed our choice to focus on age, education, and occupation in analyzing the differential impacts of the village fund program.

A. *Decision Tree*

To identify the household characteristics most predictive of short-term credit uptake in response to the Million Baht Village Fund, we constructed a regression tree with a depth of three, as shown in Figure 1 below. This model captures the primary variables influencing borrowing behavior in a straightforward manner.

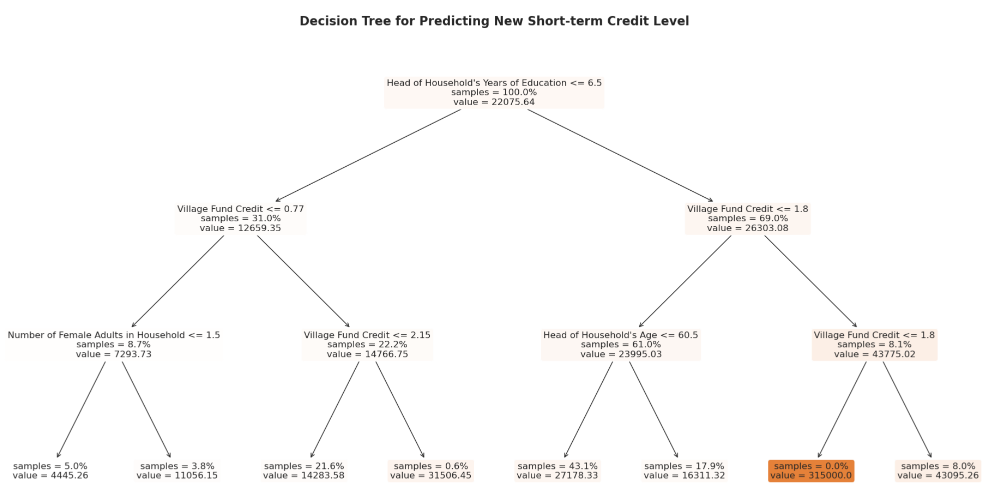

FIGURE 1. DECISION TREE FOR PREDICTING NEW SHORT-TERM CREDIT LEVEL

The first split in the tree is based on the head of household's years of education (≤ 6.5 years), indicating that households with lower educational attainment tend to borrow less new short-term credit, suggesting that education may enhance the ability to utilize credit productively.

Subsequent splits highlight the importance of both age and the amount of village fund credit received. Younger household heads (age ≤ 60.5) are more likely to increase their borrowing in response to the program, aligning with our hypothesis that younger individuals might have longer investment horizons and greater risk tolerance. Village fund credit levels also frequently appear as split criteria, underscoring the program's central role in shaping short-term credit uptake. Additionally, the tree identifies the number of female adults as a splitting factor, suggesting that household composition may subtly influence borrowing behavior and credit needs.

This regression tree analysis highlights education, age, village fund credit levels, and household composition as the key determinants of borrowing behavior, providing a basis for further investigation into how these factors contribute to heterogeneous responses across households.

B. *Random Forest*

To further refine our selection of household characteristics predictive of short-term credit uptake, we employed a random forest model, which aggregates predictions from multiple decision trees to improve accuracy and reduce overfitting. Figure 2 below presents the feature importance values derived from this model, highlighting the most influential variables in predicting borrowing behavior.

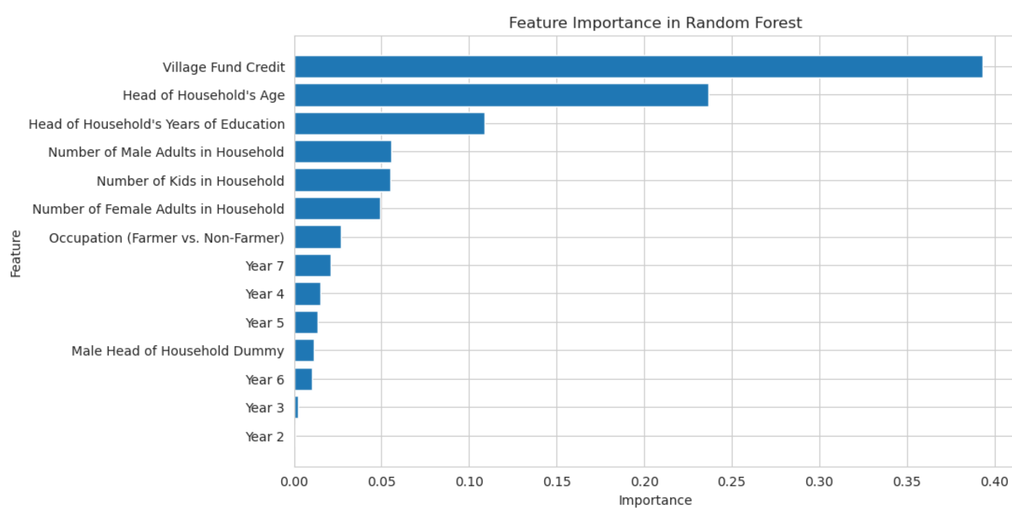

Figure 2. Feature Importance in Random Forest Model

The random forest analysis emphasizes three key variables: village fund credit, head of household's age, and years of education. Village fund credit emerges as the most critical factor, reinforcing its central role in determining short-term borrowing levels among households. The age of the household head follows closely, consistent with our hypothesis that younger individuals may respond more actively to available credit due to longer investment horizons and higher risk tolerance. The years of education of the household head also rank highly, suggesting that financial literacy and decision-making abilities play a crucial role in borrowing behavior.

Interestingly, the model also identifies household composition factors, such as the number of male, female, and total children in the household, as contributing to prediction accuracy, though they are less influential than age, education, and village fund credit. Occupation (farmer vs. non-farmer) appears further down in importance, yet it remains a variable of interest for understanding credit behavior differences across economic activities.

This random forest analysis corroborates our decision to focus on age, education, and village fund credit as the primary variables for examining heterogeneous responses to the microcredit program, with additional attention to household composition factors.

C. *Boosting*

Using a gradient boosting model, we further evaluated feature importance to identify factors influencing short-term credit uptake. Figure 3 below displays the results, with village fund credit, head of household's age, and years of education emerging as the top predictors.

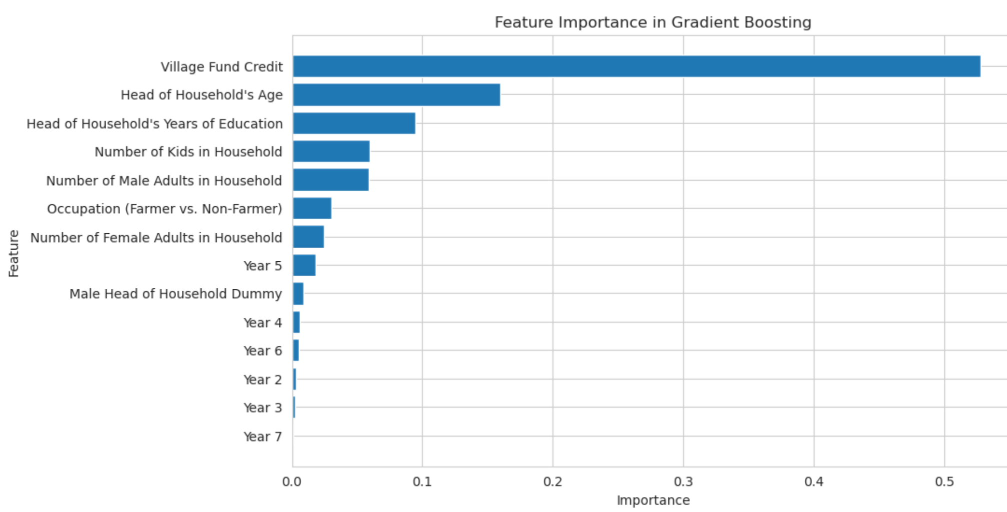

Figure 3. Feature Importance in Gradient Boosting Model

In alignment with previous models, village fund credit holds the highest importance, emphasizing its primary role in driving borrowing behavior. Age and education of the household head again rank prominently, supporting our focus on these variables for understanding demographic variations in response to credit access. Household composition factors, including the number of children and male adults, also contribute to prediction accuracy, though with lower importance.

This boosting analysis reaffirms our focus on village fund credit, age, and education, providing additional confidence in these variables as central to capturing the heterogeneous effects of the program.

D. Bagging

We applied a bagging model to further assess the importance of household characteristics in predicting short-term credit uptake. Figure 4 below shows the feature importance, with village fund credit, age, and education of the household head consistently ranking as the top predictors.

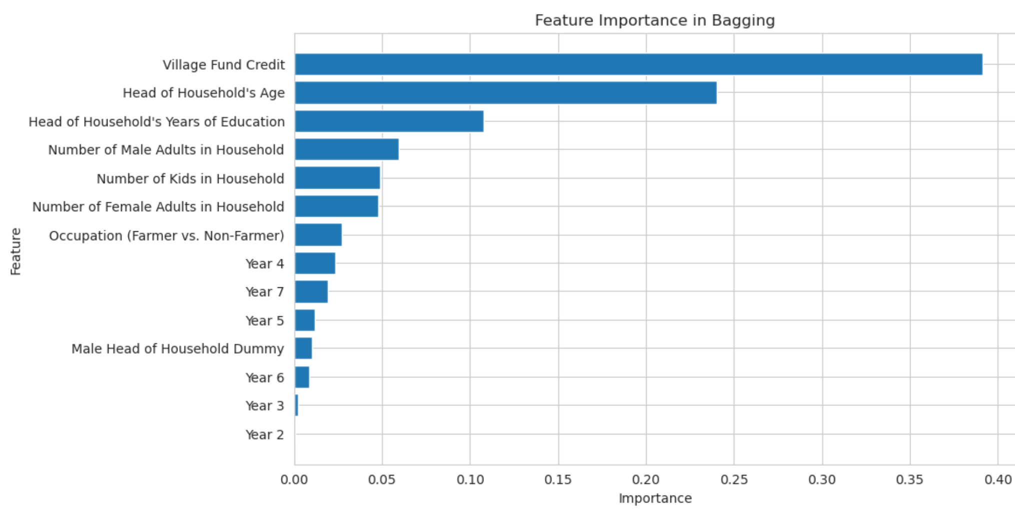

Figure 4. Feature Importance in Bagging Model

As in previous models, village fund credit remains the most influential factor, followed closely by the head of household's age and years of education, reinforcing their significance in determining borrowing behavior. Household composition variables, such as the number of male and female adults and children, also appear but with lower importance.

The bagging results align with our earlier findings, supporting the selection of village fund credit, age, and education as primary variables for analyzing heterogeneous responses to the program.

E. LASSO

To identify key predictors of short-term credit uptake, we applied LASSO regression, which shrinks less influential variables to zero. The cross-validated mean squared error (MSE) and coefficient paths are shown in Appendix (See Figures 17 and 18). Figure 5 below highlights the important features identified by LASSO.

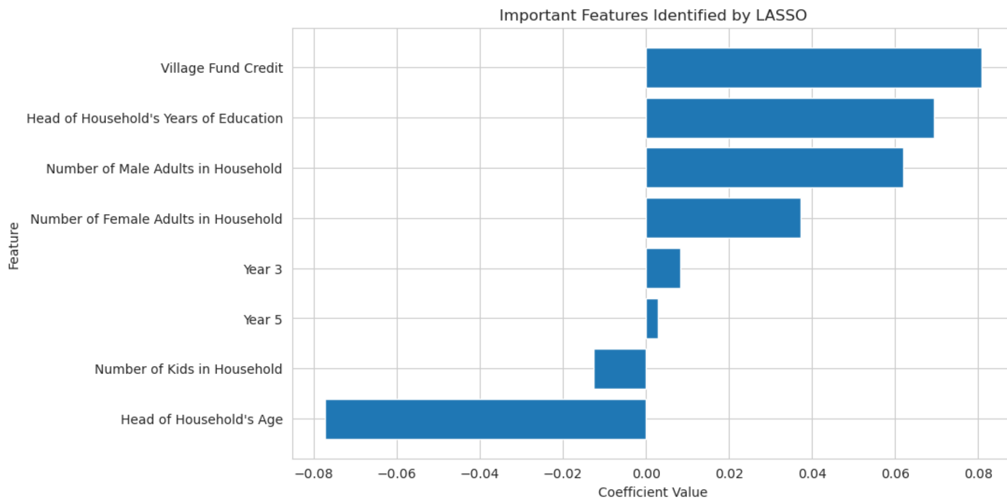

FIGURE 5. IMPORTANT FEATURES IDENTIFIED BY LASSO

Village fund credit emerges as the strongest predictor, underscoring its primary role in shaping borrowing behavior. Education is also significant, likely reflecting the positive impact of financial literacy on decision-making. Household composition—particularly the number of male and female adults and children—also influences credit uptake, suggesting differing financial roles and needs within the household. Additionally, age has a negative coefficient, indicating that younger individuals are more responsive to credit availability. Notably, occupation (farmer vs. non-farmer) does not appear among the top predictors, suggesting it plays a limited role in this model compared to other household characteristics.

These results reinforce the focus on village fund credit, education, age, and household composition as primary variables for understanding the heterogeneous effects of the program.

F. Ridge

To further explore variable significance while addressing multicollinearity, we applied Ridge regression. This technique regularizes coefficients, allowing us to identify influential predictors without reducing any to zero. The cross-validated mean squared error (MSE) and coefficient paths are shown in Appendix (See Figures 19 and 20). Figure 6 below highlights the important features identified by Ridge.

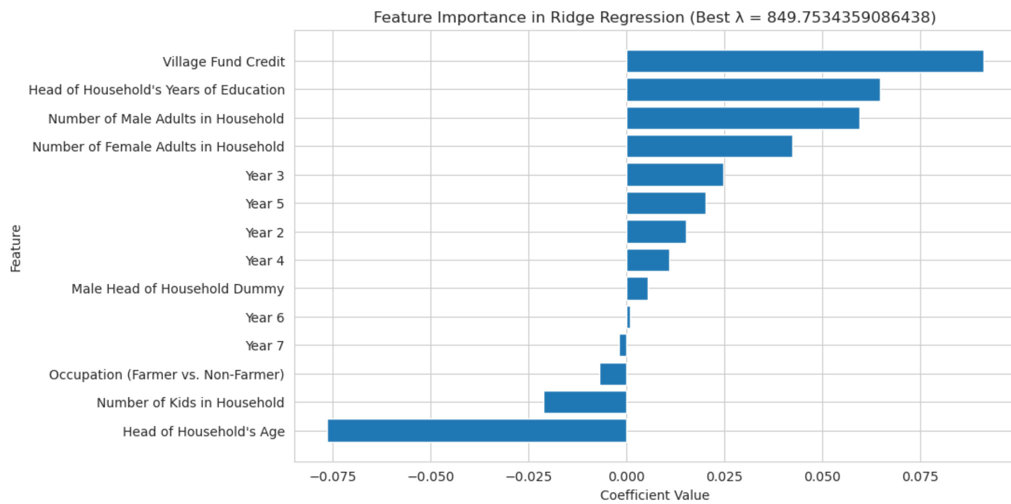

Figure 6. Feature Importance in Ridge Regression (Best λ)

As with previous models, village fund credit and education emerge as primary predictors, reaffirming their strong influence on credit uptake. Household composition, including the number of male and female adults, is also notable, indicating its role in credit decisions. Age has a negative coefficient, aligning with the trend that younger individuals are more likely to utilize credit. Several year dummies (e.g., Years 2, 3, and 5) appear with smaller coefficients, capturing potential temporal effects that could reflect changing economic conditions or shifts in program impact over time. Like we've seen in other model importance plots, the year effects add nuance to the model by accounting for time-specific influences, though they are relatively minor compared to other variables. Occupation and the number of children have minimal impact in this model, suggesting these variables play a limited role relative to others.

These results consistently support the inclusion of village fund credit, education, age, and household composition as key factors in analyzing heterogeneous responses to the village fund program, with year dummies adding essential context. While occupation, particularly farming status, did not emerge as highly predictive, it is still valuable to explore potential influences and further investigate its role in credit utilization. Given these models and the focus of our study, we select village fund credit as the primary variable of interest, examining its effect on new short-term credit with an emphasis on the heterogeneous impacts of education, age, and occupation.

III. Summary Statistics

A. General Summary

Table 1—Summary Statistics for Key Variables

	Count	Mean	Std. Dev.	Min	25%	Median	75%	Max
New Short-Term Credit	725	21831.1	35763.4	0.0	3571.4	10857.1	28321.4	528285.7
Village Fund Credit	725	9326.7	8626.2	0.0	0.0	10000.0	16500.0	50000.0
Age of Household Head	725	53.9	13.0	26.0	43.0	52.6	63.6	90.0
Years of Education of Household Head	722	6.1	3.0	0.0	6.4	6.6	6.6	16.0
Occupation								
Farmer (Count)	3248 (64% of sample)							
Non-Farmer (Count)	1827 (36% of sample)							

The table presents summary statistics for our key variables. The New Short-Term Credit variable, which measures new loans taken with a term of one year or less, averages 21,831 baht, with substantial variation ranging from 0 to 528,286 baht. The Village Fund Credit shows an average of 9,327 baht, with a median of 10,000 baht, and borrowing levels ranging up to 50,000 baht, reflecting diverse household participation in the fund. Household heads have an average age of 53.9 years, ranging from 26 to 90, while years of education average 6.1 years, with a median of 6.6, highlighting relatively medium educational attainment. Additionally, 64% of household heads are engaged in farming, as indicated by the farming dummy. In the subsequent analysis, we will examine the distribution of key characteristics—age, education, and occupation—within the dataset.

B. Age

To analyze the difference between young and old household heads in terms of the impact of village fund credit on new short-term credit, we first need to establish a clear cutoff point to define "young" and "old." To determine this split, we begin by examining the distribution of the age variable using density and histogram plots.

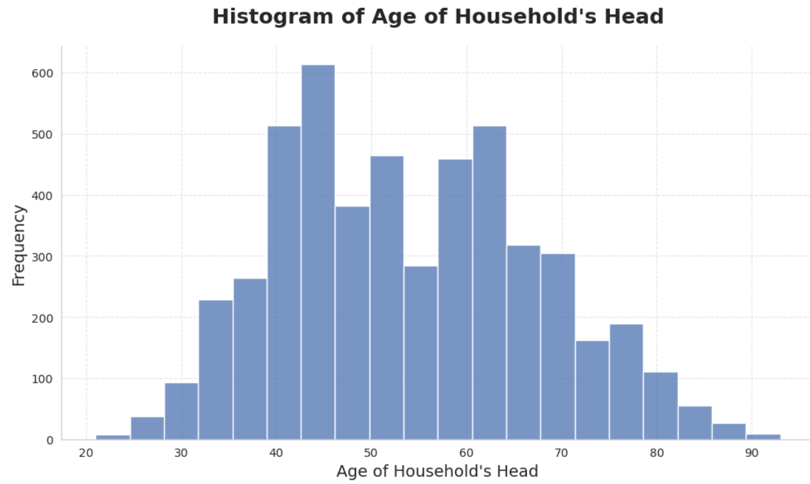

Figure 7. Histogram of Age of Household's Head

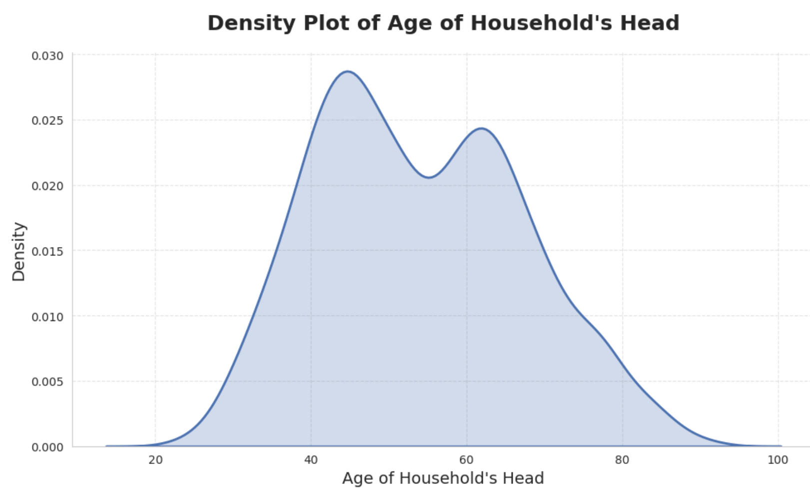

Figure 8. Density Plot of Age of Household's Head

The density and histogram plots of household head ages (Figures 7 and 8) show a bimodal distribution with peaks around 40 and 60, suggesting a natural divide between younger and older groups. This split likely reflects generational transitions in rural Thailand, where younger household heads (around 40) may have taken over from their parents, while older heads (around 60) continue to hold authority. These groups also correspond to different life stages, with younger individuals entering mid-career and older individuals nearing retirement. To define our split, we will now find the minimum point between the two peaks.

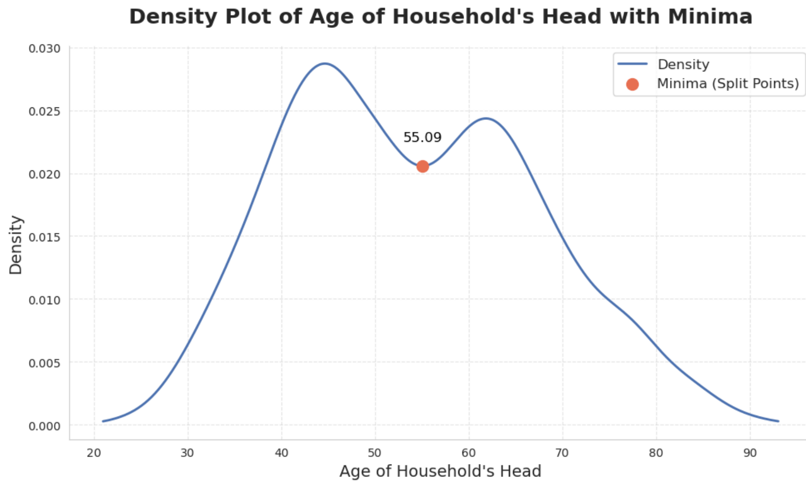

FIGURE 9. DENSITY PLOT OF AGE OF HOUSEHOLD'S HEAD WITH MINIMA

Based on the density plot in Figure 9, the minimum point between the two peaks is around 55, so we will use age 55 as the cutoff between young and old household heads.

TABLE 2—SUMMARY STATISTICS FOR YOUNG AND OLD HOUSEHOLD HEADS

	Count	Mean	Std. Dev.	Min	25%	Median	75%	Max
Old	2245	66.58	7.69	56	61	65	71	93
Young	2792	43.73	6.83	21	39	44	49	55

As shown in Table 2, young household heads have an average age of around 43, ranging from 21 to 55, while old household heads average around 66 years, with ages ranging from 56 to 93.

C. *Education*

The initial summary table shows that years of education for household heads range from 0 to 16. Analyzing individuals for each year would be impractical, so we create a histogram to identify suitable education groups for categorization.

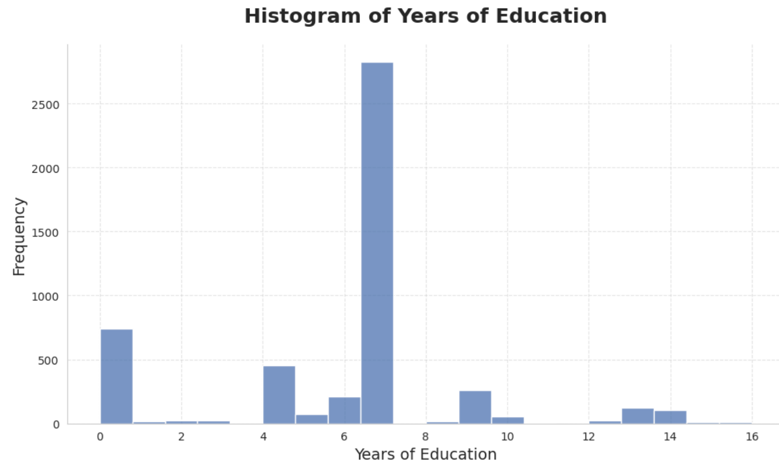

Figure 10. Histogram of Years of Education

Based on the distribution in the histogram (Figure 10), a reasonable approach is to create three education groups that reflect meaningful breaks in the data. The first group, Low Education (0-4 years), captures household heads with little to no formal education, including those who likely did not complete primary school. The second group, Medium Education (5-8 years), typically aligns with the completion of primary education and possibly some secondary education. Finally, the High Education group (9-16 years) includes household heads who completed secondary education, possibly high school, and those with some or full tertiary education. This categorization provides a structured way to analyze education levels in the dataset.

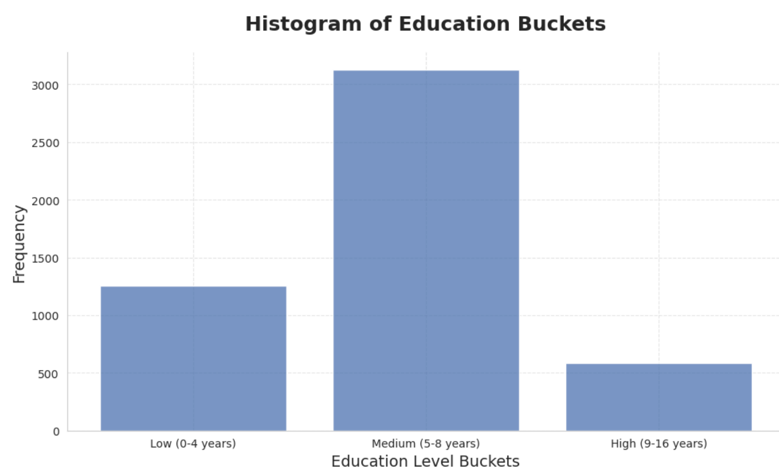

Figure 11. Histogram of Education Level Buckets

TABLE 3—SUMMARY STATISTICS FOR EDUCATION GROUPS

	Count	Mean	Std. Dev.	Min	25%	Median	75%	Max
Low (0-4 years)	1255	1.54	1.90	0.00	0.00	0.00	4.00	4.00
Medium (5-8 years)	3124	6.89	0.40	5.00	7.00	7.00	7.00	8.00
High (9-16 years)	588	11.16	2.22	9.00	9.00	10.00	13.00	16.00

As a result, among the three groups, the majority of household heads fall into the medium education category, followed by the low education group, with the fewest in the high education group.

D. Occupation

Since the dataset includes a dummy variable for the household head's occupation—farmer or non-farmer—we will use this classification for our analysis of occupation. As shown in the initial summary table, 64% of household heads are engaged in farming, a distribution that is similarly reflected in the histogram plot.

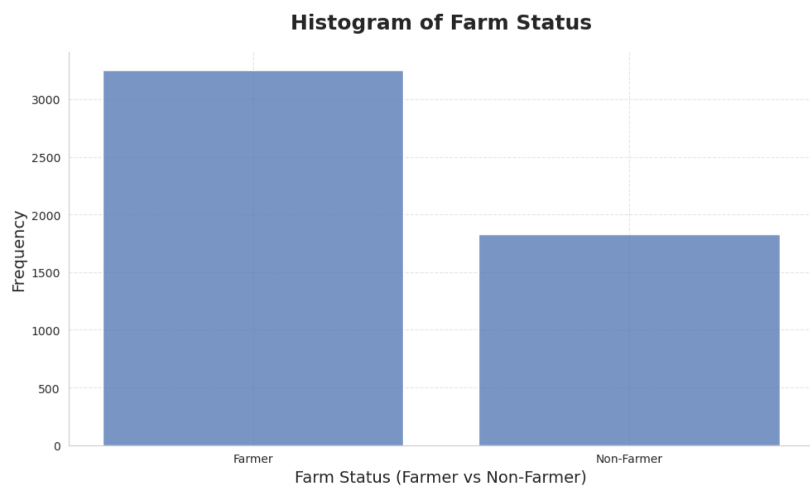

FIGURE 12. HISTOGRAM OF FARM STATUS (FARMER VS NON-FARMER)

The histogram in Figure 12 illustrates the distribution of household heads by occupation. The majority of household heads are farmers, with a smaller proportion classified as non-farmers. This occupation split will allow us to examine the differential impact of village fund credit on short-term credit uptake across these two groups.

IV. Data Limitations

Despite the comprehensive nature of the Townsend Thai Project dataset, several data limitations must be acknowledged. First, the study experienced attrition over the eleven-year period, with the panel size decreasing from 960 to 655 households due to factors such as migration. This reduction may introduce missing data issues and potential biases if the households that left differ systematically from those that remained. Second, measurement errors are a concern, given that key variables—including income, expenditures, borrowing, and saving behaviours—are based on self-reported information. Participants may inaccurately recall financial details or intentionally misreport sensitive data, leading to inaccuracies. Additionally, the dataset exhibits significant variability, particularly in financial transactions where a few large investments or loans can disproportionately influence the results. The presence of such outliers may skew statistical analyses and affect the robustness of the findings. Lastly, the regional focus on specific rural areas within four provinces of Thailand may limit the generalizability of the results to other regions or urban settings, thereby constraining the broader applicability of the study's conclusions.

V. Regression Specification

In this study, we estimate a series of instrumental variable (IV) regressions using a two-stage least squares (2SLS) approach with fixed effects to assess the impact of village fund credit on new short-term credit levels across three household characteristics: age, education, and occupation. We restrict our analysis to households in villages with between 50 and 250 households to reduce the influence of extreme outliers. Our empirical strategy consists of a first-stage regression predicting village fund credit using the instrumental variable inverse village size, followed by three second-stage regressions—one for each characteristic under study.

A. First-Stage Regression

The first stage predicts the household-level village fund credit using the inverse village size as an instrument:

$$\begin{aligned} \text{Village Fund Credit}_{it} = \quad & \alpha + \beta_1 \cdot \text{Inverse Village Size}_{1,it} \\ & + \beta_2 \cdot \text{Inverse Village Size}_{2,it} \\ & + X_{it} \cdot \gamma + \mu_i + \lambda_t + \epsilon_{it} \end{aligned} \tag{1}$$

where Inverse Village Size$_1$ and Inverse Village Size$_2$ represent the instrumental variables derived from village size, X_{it} is a vector of household-level control variables, including demographic and economic characteristics such as household composition, gender of the household head, occupation, age, and education level. μ_i represents village fixed effects, λ_t represents time fixed effects, and ϵ_{it} is the error term.

B. Second-Stage Regressions

After obtaining the predicted values of village fund credit from the first stage, we estimate three separate second-stage regressions, each interacting the predicted credit variable with one of the household characteristics (age, education, and occupation) to explore heterogeneity in responses.

Age Regression: We interact the predicted village fund credit with an indicator for Young household heads, using Old as the base case, to capture differences in response by age:

$$\begin{aligned}\text{New Short-Term Credit}_{it} = {} & \alpha + \beta_1 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{Young}_{it} \\ & + X_{it} \cdot \gamma + \mu_i + \lambda_t + \epsilon_{it}\end{aligned} \tag{2}$$

where Village $\widehat{\text{Fund}}$ Credit$_{it}$ $\cdot$ Young$_{it}$ represents the interaction term.

Education Regression: We interact the predicted village fund credit with indicators for Medium and High education groups, using Low as the base case, to examine the effects of education levels:

$$\begin{aligned}\text{New Short-Term Credit}_{it} = {} & \alpha + \beta_1 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{Medium}_{it} \\ & + \beta_2 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{High}_{it} \\ & + X_{it} \cdot \gamma + \mu_i + \lambda_t + \epsilon_{it}\end{aligned} \tag{3}$$

where Village $\widehat{\text{Fund}}$ Credit$_{it}$ $\cdot$ Medium$_{it}$ and Village $\widehat{\text{Fund}}$ Credit$_{it}$ $\cdot$ High$_{it}$ are the interaction terms.

Occupation Regression: We interact the predicted village fund credit with an indicator for Non-Farming households, using Farmer as the base case, to analyze occupation-based differences:

$$\begin{aligned}\text{New Short-Term Credit}_{it} = {} & \alpha + \beta_1 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{Non-Farmer}_{it} \\ & + X_{it} \cdot \gamma + \mu_i + \lambda_t + \epsilon_{it}\end{aligned} \tag{4}$$

where Village $\widehat{\text{Fund}}$ Credit$_{it}$ $\cdot$ Non-Farmer$_{it}$ is the interaction term.

In each of these second-stage regressions, the dependent variable is new short-term credit, which measures the amount of new loans taken by households for short-term borrowing. By estimating these models, we capture how different household characteristics—age, education, and occupation—modify the effect of village fund credit on short-term borrowing, while controlling for village and time fixed effects.

C. Directed Acyclic Graphs (DAGs) of the Regression Specifications

To complement our regression specifications, we present a series of Directed Acyclic Graphs (DAGs) to visually depict the causal pathways of our instrumental variable (IV) approach. These DAGs illustrate how inverse village size, used as an instrument for village fund credit, impacts new short-term credit with heterogeneous effects across age, education, and occupation. Each DAG mirrors the structure of our two-stage least squares (2SLS) model, showing how household characteristics mediate the effect of village fund credit. In this approach, inverse village size serves as the instrument in the first stage, and the predicted values are then interacted with specific household characteristics in the second stage to capture differential responses. Control variables and fixed effects are included to address potential confounders, ensuring that the observed relationships are isolated from unobserved village-specific or time-varying factors.

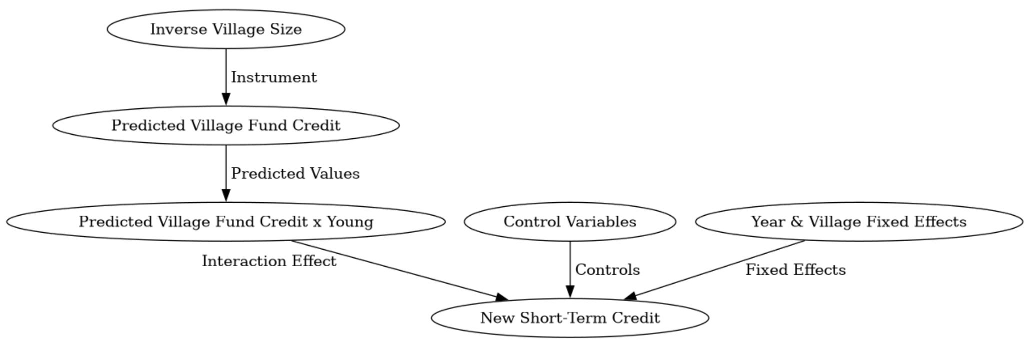

Figure 13. DAG for Age Specification

The first DAG, shown in Figure 13, illustrates the age specification. Here, the predicted village fund credit is interacted with an indicator for young household heads, with older households as the base case. This interaction enables us to assess whether younger households, possibly more inclined toward credit use due to investment opportunities or longer planning horizons, respond differently to village fund credit than older households.

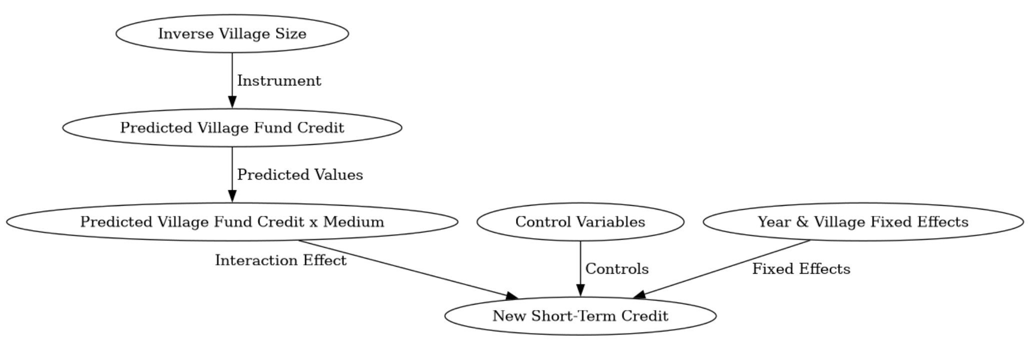

Figure 14. DAG for Education Specification: Medium Education Level

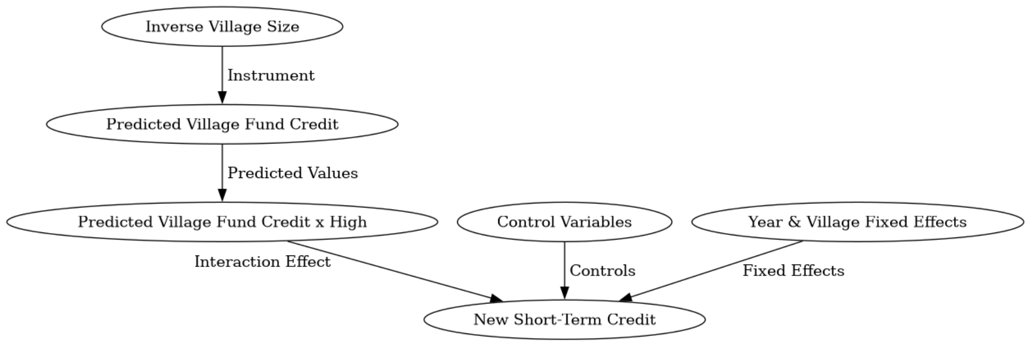

FIGURE 15. DAG FOR EDUCATION SPECIFICATION: HIGH EDUCATION LEVEL

Figures 14 and 15 depict the education specification. Here, we interact predicted village fund credit with medium and high education indicators, using low education as the base case. This allows us to examine if educational attainment modifies the response to credit access, as more educated households might have greater financial literacy or access to investment opportunities, enabling them to make more effective use of village fund credit.

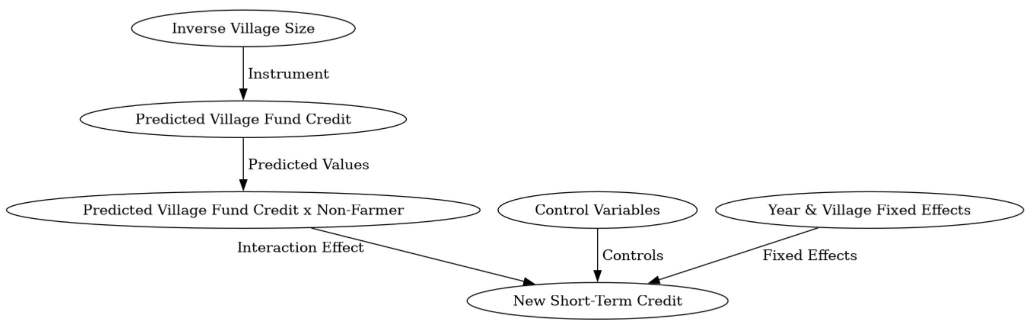

FIGURE 16. DAG FOR OCCUPATION SPECIFICATION

The final DAG, shown in Figure 16, represents the occupation specification, where predicted village fund credit is interacted with a non-farmer indicator. This model allows us to capture occupational differences in credit response, hypothesizing that non-farming households, with potentially greater or different economic opportunities compared to farming households, may exhibit distinct borrowing behavior when village fund credit becomes available.

Together, these DAGs reinforce our regression specifications by providing a structured view of how household characteristics—age, education, and occupation—mediate the effect of village fund credit on short-term borrowing. The DAGs clarify the flow from instrument to outcome, capturing the nuances of credit responses within demographic and socioeconomic subgroups while controlling for village and time fixed effects.

VI. Identification Strategy

To estimate the causal impact of the Million Baht Village Fund program on household borrowing behavior, we use a combination of a difference-in-difference (DiD) framework and an instrumental variable (IV) approach. The DiD framework leverages our panel data, spanning five years before and six years after the program's introduction in 2001. Since households could not borrow from the village fund before its implementation (the village fund credit variable only exists post-program), we compare the changes in new short-term credit levels within the same households before and after the program. This method controls for unobserved, time-invariant household characteristics that could otherwise confound the relationship between access to credit and borrowing behavior. However, the DiD method alone may not fully address endogeneity concerns, as unobserved time-varying factors could influence both village fund credit and borrowing behavior, thus we use the inverse of village size as an instrumental variable.

The IV approach exploits the exogenous variation in per-household credit availability created by the program's design, which allocated a fixed one million baht to each village, regardless of its population size. As a result, smaller villages received more funds per household than larger ones. This variation, independent of household characteristics and borrowing decisions, makes the inverse of village size a valid instrument. By using it to instrument for village fund credit, we can isolate the variation in credit availability attributable to village size alone, addressing endogeneity concerns and ensuring that our estimates of the effect on borrowing behavior are not biased by unobserved variables. Therefore, by combining the DiD and IV approaches, we leverage both the temporal variation from the program's implementation and the cross-sectional variation from village size differences. This strengthens the credibility of our causal estimates by controlling for unobserved heterogeneity and mitigating potential endogeneity arising from unobserved, time-varying factors.

VII. Results

In this section, we present the results from our second-stage regressions in a combined results table. Each second-stage regression uses the predicted values of village fund credit from the first-stage regression, where we regressed the instrumental variable, inverse village size, on village fund credit. First-stage regression results are available in the appendix for reference (see Table 8).

TABLE 4—2SLS IV REGRESSION: HETEROGENEOUS EFFECTS ON NEW SHORT-TERM CREDIT

Village Fund Credit Interactions	**Age Specification**	**Education Specification**	**Occupation Specification**
Young	5238.2***	-	-
	(2381.9)		
Medium Education	-	4846.9**	-
		(2208.4)	
High Education	-	10110*	-
		(5585.5)	
Non-Farmer	-	-	3118.1
			(2003.3)
Control Variables	✓	✓	✓
Year Fixed Effects	✓	✓	✓
Village Fixed Effects	✓	✓	✓
Observations	4967	4967	4967
R-squared (Within)	0.0195	0.0199	0.0183

Notes: The regressions control for household characteristics, including the number of adults, female adults, children, male household head, farming status, age, age squared, and years of education. These controls are included except for the specific subgroup being estimated (e.g., age is excluded when estimating by age groups). The model also includes village-level fixed effects and year fixed effects. Statistical significance is indicated by *** (1%), ** (5%), and * (10%). Standard errors are in parentheses.

The results from the age specification reveal significant differences in the effect of village fund credit on new short-term credit based on the age of the household head. For young household heads (with older household heads as the base case), the coefficient for village fund credit is 5,238.2 baht, statistically significant at the 1% level. This finding suggests that younger households are more responsive to the availability of village fund credit, potentially due to greater investment opportunities or a longer time horizon that encourages borrowing. Older households, by contrast, may be less inclined to take on new debt, possibly due to life-stage considerations.

The education specification demonstrates a clear gradient in responsiveness to village fund credit across different education levels. Using low education (0-4 years) as the base case, households with medium education (5-8 years) show a statistically significant coefficient of 4,846.9 baht, while those with high education (9-16 years) exhibit an even greater response, with a coefficient of 10,110 baht, significant at the 10% level. This pattern suggests that households with higher levels of education are better equipped to utilize village fund credit effectively. Higher educational attainment may correspond to greater financial literacy or access to investment opportunities, enabling these households to make productive use of borrowed funds.

The occupation specification highlights differences in credit responses based on the household head's occupation. Non-farming households (with farming households as the base case) exhibit a coefficient of 3,118.1 baht, though it is not statistically significant. This result indicates that non-farming households may have varied opportunities to employ credit in non-agricultural ventures, yet the effect is less robust. The lack of significance for non-farming households suggests that while they may have the potential for higher returns in certain cases, this effect is not consistent across the group.

Across all three specifications, the results suggest that household characteristics—age, education, and occupation—play a critical role in determining how households respond to the availability of village fund credit. Younger, more educated, and non-farming households are more likely to increase their short-term borrowing in response to increased credit availability. These findings indicate that these groups may benefit more from the program's implementation, likely due to better access to investment opportunities or a greater capacity to manage debt effectively.

Robustness Check. — To assess the robustness of our models, we compare the results from the instrumental variable (IV) 2SLS regression with those from a simple ordinary least squares (OLS) regression that does not use the instrument. Across all models stratified by age, education, and occupation, the OLS results consistently yield positive and statistically significant coefficients for the key interaction terms, as shown in Table 5. In contrast, some of the IV estimates lack statistical significance. For example, while the coefficients for young household heads and medium and high education levels remain significant in the IV model, the non-farmer interaction in the occupation specification is not statistically significant. This difference suggests that the IV method, which addresses endogeneity, produces more conservative estimates that highlight underlying variability across groups.

Notably, the IV estimates are consistently smaller than their OLS counterparts, underscoring potential upward bias in OLS estimates due to endogeneity. This divergence between IV and OLS coefficients implies that OLS may overstate the true effects, with IV offering a more reliable estimate of the impact of village fund credit on new short-term credit. Despite the lack of statistical significance for some IV interaction terms, the consistency of positive coefficients across both estimation methods strengthens the validity of our findings.

TABLE 5—OLS REGRESSION: HETEROGENEOUS EFFECTS ON NEW SHORT-TERM CREDIT

Village Fund Credit Interactions	**Age Specification**	**Education Specification**	**Occupation Specification**
Young	12150*** (1699.5)	-	-
Medium Education	-	11370*** (1312.6)	-
High Education	-	13170*** (4020.2)	-
Non-Farmer	-	-	8328.6*** (2700.8)
Control Variables	✓	✓	✓
Year Fixed Effects	✓	✓	✓
Village Fixed Effects	✓	✓	✓
Observations	4967	5037	4967
R-squared (Within)	0.0363	0.0378	0.0228

Notes: The regressions control for household characteristics, including the number of adults, female adults, children, male household head, farming status, age, age squared, and years of education. These controls are included except for the specific subgroup being estimated (e.g., age is excluded when estimating by age groups). The model also includes village-level fixed effects and year fixed effects. Statistical significance is indicated by *** (1%), ** (5%), and * (10%). Standard errors are in parentheses.

To further ensure robustness, we assess the first-stage regression of the IV model (detailed in the appendix), where the instrument—inverse village size—exhibits strong explanatory power, with an F-statistic of 66.343, significantly above the standard threshold of 10, underscoring its relevance. The coefficients for *invHHtvf1* and *invHHtvf2* (46.392 and 85.367, respectively) are statistically significant, providing further support for instrument validity. The within R-squared value of 0.4662 also confirms that a substantial portion of the variation in village fund credit within entities is explained by the model.

We further apply the Dubin-Wu-Hausman (DWH) test to statistically assess endogeneity across each subgroup specification. As shown in Table 6, all subgroup regressions (age, education, and occupation) reject the null hypothesis of no endogeneity, with p-values of 0.000. These findings indicate that the IV approach is preferable to OLS, supporting unbiased estimation.

TABLE 6—DUBIN-WU-HAUSMAN TEST FOR ENDOGENEITY

Regression	**DWH Statistic**	**P-value**
Age Regression	1,231,095.59	0.000
Education Regression	1,410,911.14	0.000
Occupation Regression	2,845,455.58	0.000

Note: The table reports Dubin-Wu-Hausman test statistics and p-values for each regression. A p-value of 0.000 indicates rejection of the null hypothesis, providing evidence of endogeneity.

Collectively, these robustness checks confirm the reliability of the IV estimates, validating the heterogeneous effects observed across age, education, and occupation subgroups in the primary analysis.

VIII. Future Works

In future work, we plan to build upon our two-stage IV regression analysis by incorporating additional methods and robustness checks. One key extension will involve the use of quantile regression to explore how the effects of village fund credit vary across households with different borrowing levels (quantiles), offering a more detailed understanding of how socio-economic groups respond to credit access. We also aim to apply double machine learning (DML) techniques, using machine learning in the first-stage regression to predict village fund credit from the instrumental variable (inverse village size), which could further reduce bias and improve estimate precision. Additionally, the variable selection process highlighted the potential importance of other variables, including the number of children in the household, which may warrant further investigation to understand its influence on borrowing behavior. Moreover, the statistical insignificance of the occupation variable (farmer vs. non-farmer) suggests that further investigation is warranted to assess potential improvements or alternative methods that may reveal statistically significant results. We also plan to employ causal machine learning models where applicable, enhancing the accuracy and robustness of our regression analysis. Furthermore, we will implement more robust checks to address endogeneity concerns, including exploring non-linear models to capture complex relationships between variables. To better visualize our findings and illustrate causal relationships, we plan to incorporate enhanced summary statistics and graphical tools, such as scatterplots. These steps, along with peer feedback and further refinements, will strengthen the rigour and comprehensiveness of the final version of this paper.

IX. Conclusion

Our findings reveal significant differences in how various household groups utilize village fund credit, with younger, more educated, and, to a lesser extent, non-farming households showing distinct borrowing responses. Specifically, younger household heads exhibit a stronger borrowing response, with a 5,238.2 baht increase in new short-term credit, indicating a greater propensity to seize new credit opportunities. This may be attributed to longer investment horizons and a higher willingness to undertake financial risks, as younger individuals often have more time to realize returns on investments and recover from potential losses. Households with medium and high education levels demonstrate notable increases in borrowing, with coefficients of 4,846.9 and 10,110 baht, respectively. This suggests that education enhances the ability to leverage credit effectively, likely improving financial literacy and decision-making skills that enable these households to assess investment opportunities and risks more proficiently. Although non-farming households exhibit a positive but not statistically significant response to the program, this finding is nonetheless insightful. Non-farming households may have access to more immediate or diverse investment opportunities in non-agricultural sectors, which could offer quicker or higher returns. This trend may reflect a broader transition in rural economies toward diversification and non-agricultural enterprises.

These results directly address our research question by highlighting how socio-economic characteristics influence household borrowing behavior in microfinance programs. The significant variations across age, education, and occupation emphasize the importance of demographic diversity in designing and implementing such programs. By targeting younger, more educated, and non-farming households, policymakers and financial institutions can enhance microfinance program effectiveness and expand financial inclusion.

Our study extends the work of Kaboski and Townsend (Kaboski and Townsend, 2012*a*) by offering a micro-level analysis that reveals how the impact of village fund credit varies across socio-economic groups. While previous research, such as Mahjabeen's (Mahjabeen, 2008) study on microfinance in Bangladesh and Awojobi's (Awojobi, 2011) work on Nigeria, highlights the broad benefits of microfinance in improving household welfare and reducing poverty, our findings show that these benefits are unevenly distributed. Specifically, younger, more educated, and non-farming households demonstrate a stronger borrowing response, suggesting they are better positioned to leverage credit productively. This insight advances the understanding of how socio-economic characteristics shape the effectiveness of microfinance programs. Furthermore, our results contribute to the discourse on credit constraints and financial inclusion, with significant implications for microfinance programs and financial institutions. By focusing on younger, educated, and non-farming households, financial institutions can expand inclusion and promote sustainable economic growth, aligning with Vanroose and D'Espallier's (Vanroose and D'Espallier, 2009) view that microfinance can complement traditional banking in reaching underserved populations.

In conclusion, our study enhances understanding of how microfinance programs can be optimized by addressing the diverse needs of different demographic groups. By recognizing which groups are more likely to respond to credit availability, these findings offer valuable insights for improving financial inclusion, reducing credit constraints, and refining microfinance programs to better serve underserved populations.

REFERENCES

Awojobi, Oluyemi. 2011. "Microfinancing for Poverty Reduction and Economic Development: A Case for Nigeria." *International Research Journal of Finance and Economics*, 159–168. SSRN: `https://ssrn.com/abstract=1926422`.

Kaboski, Joseph P., and Robert M. Townsend. 2012*a*. "The Impact of Credit on Village Economies." *American Economic Journal: Applied Economics*, 4(2): 98–133.

Kaboski, Joseph P., and Robert M. Townsend. 2012*b*. "The Impact of Credit on Village Economies: Dataset." Dataset used from Kaboski and Townsend, "The Impact of Credit on Village Economies".

Mahjabeen, Rubaiyat. 2008. "Microfinancing in Bangladesh: Impact on Households, Consumption and Welfare." *Journal of Policy Modeling*, 30(6): 1083–1092.

Vanroose, Anja, and Bert D'Espallier. 2009. "Microfinance and Financial Sector Development." Centre Emile Bernheim. Research Institute in Management Sciences CEB Working Paper no. 9. 040.

APPENDIX

A. Variable Selection: Performance Evaluation

To ensure rigor in our variable selection process, we evaluated several models using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as performance metrics. Table 7 presents the MSE and RMSE values for each model, with lower values indicating better performance. Among the models tested, Boosting achieved the lowest MSE and RMSE, suggesting it is the best-performing model in this context.

Table 7—Model Performance Comparison

Model	**MSE**	**RMSE**
Decision Tree	2,751,382,977.05	52,453.63
Random Forest	2,386,860,740.61	48,855.51
Boosting	1,495,585,006.52	38,672.79
Bagging	2,386,863,179.56	48,855.53

Notes: Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are reported for each model. Lower values indicate better performance. The Boosting model demonstrated the best performance, with the lowest MSE and RMSE values.

B. First-Stage Results for 2SLS IV Regression

To validate our instrumental variable approach, we present the first-stage regression results in Table 8. This table reports the coefficients and standard errors for the two instrumental variables, 'invHHtvf1' and 'invHHtvf2', used to predict village fund credit. These results confirm the relevance of the instruments in explaining variation in the endogenous variable.

Table 8—First-Stage Regression

Instrumental Variables	**Inverse Village Size**
invHHtvf1	46.392**
	(19.075)
invHHtvf2	85.367***
	(9.7400)
Control Variables	✓
Year Fixed Effects	✓
Entity Fixed Effects	✓
Observations	4,967
R-squared (Within)	0.4662
F-statistic (robust)	66.343
P-value (F-statistic)	0.0000

Notes: The regressions control for household characteristics, including the number of adults, female adults, children, male household head, farming status, age, age squared, and years of education. Statistical significance is indicated by *** (1%), ** (5%), and * (10%). Standard errors are in parentheses.

C. *Cross-validated MSE for LASSO*

Figure 17 shows the cross-validated MSE for LASSO, plotted against -log(λ). This plot helps identify the optimal value of the regularization parameter λ in the LASSO model, allowing for effective feature selection by shrinking some coefficients to zero.

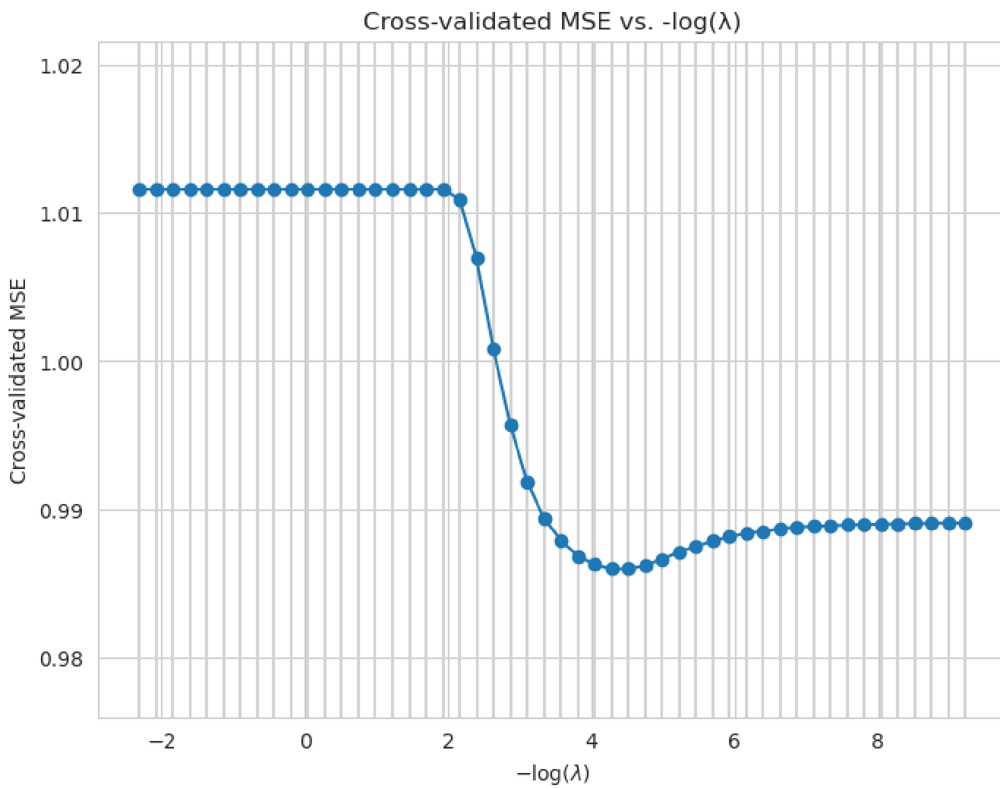

FIGURE 17. CROSS-VALIDATED MSE VS. -LOG(λ) FOR LASSO

D. LASSO Coefficient Paths

Figure 18 depicts the LASSO coefficient paths, illustrating how features are selected as λ increases. The paths provide insights into which variables remain significant as regularization enforces sparsity.

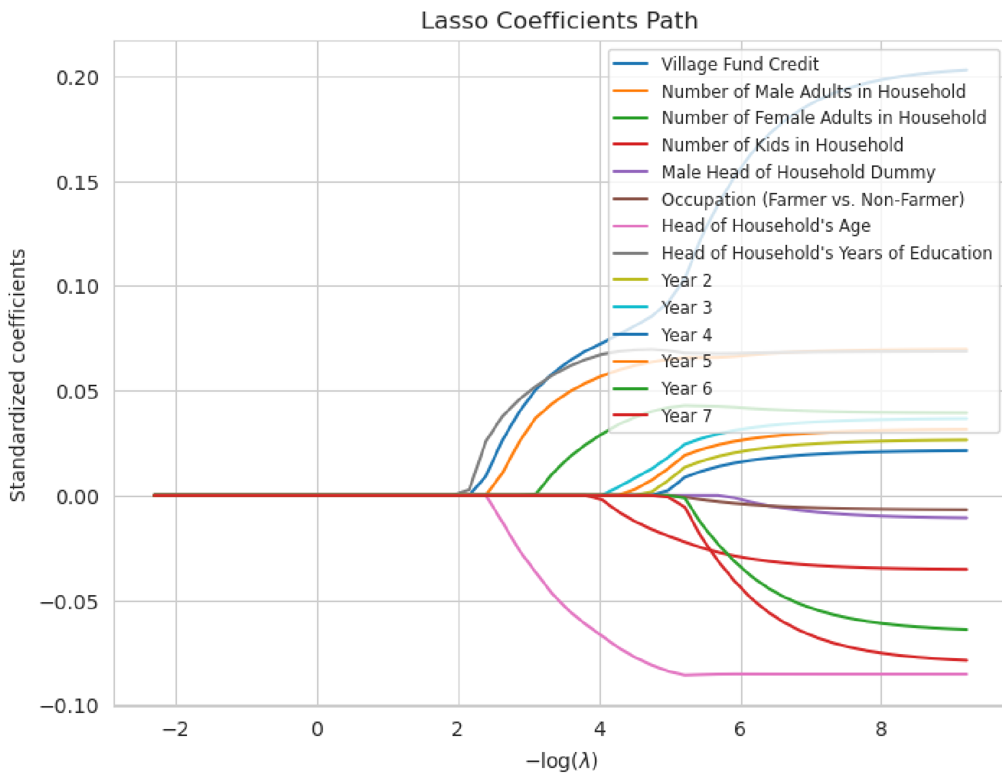

Figure 18. LASSO Coefficients Path for Feature Selection

E. Cross-validated MSE for Ridge

Figure 19 illustrates the relationship between the cross-validated MSE and the regularization parameter λ. The results highlight the optimal level of regularization for feature selection within the Ridge framework.

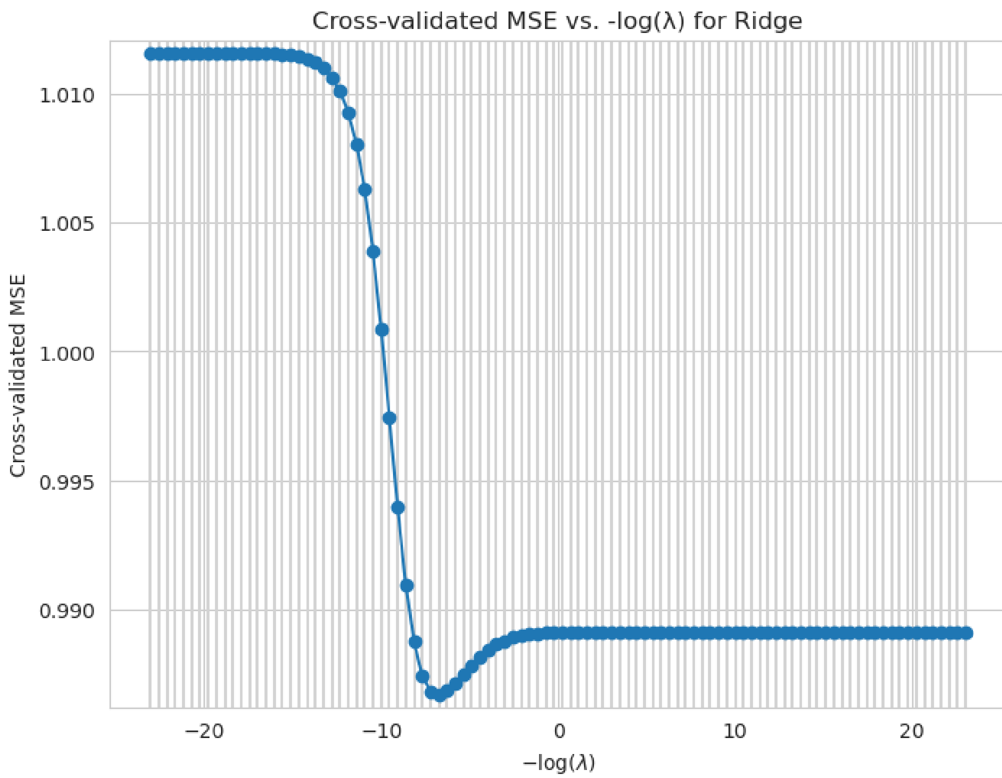

FIGURE 19. CROSS-VALIDATED MSE VS. -LOG(λ) FOR RIDGE

F. Ridge Coefficient Paths

The coefficient paths for Ridge regression, shown in Figure 20, demonstrate how each feature's coefficient changes as the regularization parameter λ varies. This visualization aids in understanding which features remain stable or diminish in importance with regularization.

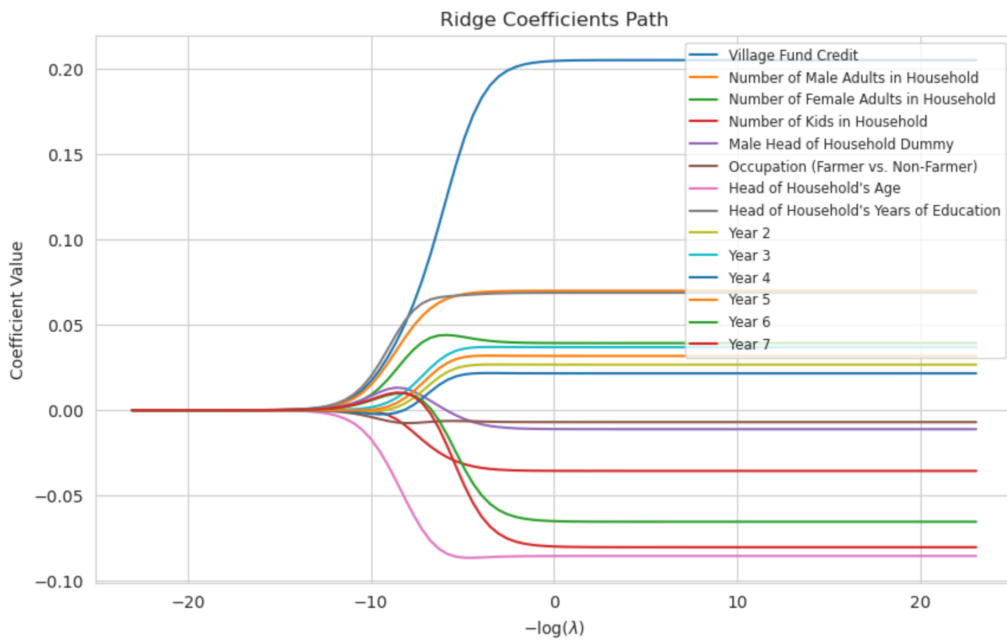

FIGURE 20. RIDGE COEFFICIENTS PATH FOR FEATURE SELECTION