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, evaluating its impact on household borrowing across age, education, and occupation groups. To address endogeneity in the relationship between credit access and borrowing behavior, we use a two-stage least squares (2SLS) approach, leveraging inverse village size as an instrument for village fund credit. This instrument exploits the exogenous variation in per-household credit allocation, arising from the program's uniform fund distribution across villages of varying sizes. Our findings reveal that younger, more educated, and non-farming households exhibit stronger borrowing responses, emphasizing the role of socio-economic characteristics in shaping credit utilization. Advanced analyses, including Meta-Learners, Propensity Score Matching, Inverse Probability Weighting, Deep IV, and Causal Forests, reinforce these results, uncovering heterogeneous treatment effects. These insights extend the work of Kaboski and Townsend, offering micro-level evidence to inform targeted microfinance strategies. Policy implications highlight the potential to enhance financial inclusion and economic growth by prioritizing demographic groups with the highest borrowing potential.
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, can also affect credit needs and repayment capabilities.

My 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. By doing so, we contribute to the literature by offering a more nuanced understanding of the micro-level dynamics

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

that underlie the aggregate outcomes observed in previous studies.

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, our approach provides a more granular understanding of how microcredit can be optimized to foster economic growth.

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. Notably, those who benefit most from microfinance are often the very same individuals most likely to transition into formal banking relationships by opening bank accounts. This highlights a nuanced link between microfinance participation and formal financial inclusion, underscoring how targeted microcredit can act as a gateway to broader financial engagement.

Our findings highlight notable differences in how various household groups respond to village fund credit availability. Given that our data captures the immediate period following the program's launch, and that new short-term credit refers to loans with a term of one year or less, we concentrate specifically on short-term borrowing patterns. 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.

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.

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 and semi-urban 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. We define rural villages as those primarily engaged in agriculture with limited access to urban infrastructure, while semi-urban areas show a higher concentration of non-farm activities and closer integration with nearby towns. 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. We note that while short-term impacts are the primary focus of this study, the dataset's extended time horizon allows for analysis of longer-term effects. For the purpose of this paper, we emphasize short-term borrowing behavior, as this horizon provides cleaner insights into the immediate credit response following the program's introduction.

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.

A. 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. To address this, we focus on the short-term period following the program's implementation and restrict the analysis to households with complete observations, removing rows with missing values. While this approach improves data consistency, it may still introduce selection bias if the excluded households differ systematically from those in the retained sample. 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. To continue, while all villages eventually received treatment, it is worth noting that the selection of villages for the survey may not have been entirely random at the outset. The sampled villages are geopolitical administrative units, and some level of administrative or infrastructural setup may have influenced their inclusion in the panel data. However, we verify that the variation in village size, which drives our exogeneity, is uncorrelated with geographic features like roads, rivers, or forest areas, as shown in previous studies (Kaboski and Townsend, 2012*a*,*b*). Furthermore, there is an imbalance in the data, particularly for village fund credit, as the inclusion of pre-program implementation data results in numerous observations where no village fund credit was available. This imbalance may affect estimates and requires careful interpretation of the program's impact. Lastly, the regional focus on specific areas within four provinces of Thailand may limit the generalizability of the results to other regions or settings, thereby constraining the broader applicability of the study's conclusions.

II. Summary Statistics

A. General Summary

Table 1—Summary Statistics for Key Variables

Variable	**Count**	**Mean (SD)**	**Min**	**Max**
New Short-Term Credit (All periods)	4,967	22,075.64		
		(53,437.76)	0	1,023,000
Village Fund Credit (Post-program)	714	9,381.44		
		(8,577.40)	0	50,000
Age of Household Head	4,967	53.99		
		(13.50)	21	93
Years of Education of Household Head	4,967	6.04		
		(3.20)	0	16
Age				
Young	2,741	(55.18% of sample)		
Old	2,226	(44.82% of sample)		
Education				
Low (0-4 years)	1,255	(25.27% of sample)		
Medium (5-8 years)	3,124	(62.90% of sample)		
High (9-16 years)	588	(11.84% of sample)		
Occupation				
Farmer	3,238	(65.19% of sample)		
Non-Farmer	1,729	(34.81% of sample)		

The summary statistics in Table 1 provide an overview of key variables in the dataset. New short-term credit, which includes loans with a term of one year or less, averages 22,076 baht (SD = 53,438) across all periods. Village Fund credit exists across all periods; however, since no household received it before the program's implementation, the summary statistics are influenced by a large number of zeros. To provide a clearer understanding of the funds distributed after the program's implementation, we report Village Fund credit only for the post-program period, where it averages 9,381 baht (SD = 8,577) and ranges from 0 to 50,000 baht. Household heads have an average age of 54 years, with 55.2% classified as younger ($\leq$ 55 years) and 44.8% as older. Educational attainment is modest, with most household heads in the medium education group (5–8 years, 62.9%). Lower education (0–4 years) accounts for 25.3%, while higher education (9–16 years) is the least common (11.8%). Occupation data highlights a predominantly agricultural setting, as 65.2% of household heads are farmers, while 34.8% are non-farmers. In the analysis that follows, we will explore the distribution of these key characteristics and provide justification for the categorizations applied.

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.

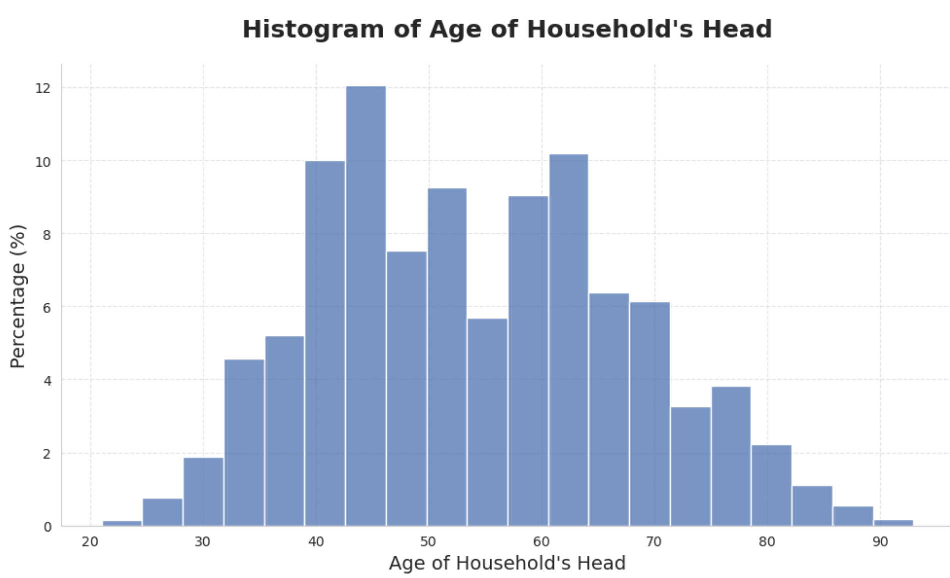

Figure 1. Histogram of Age of Household's Head

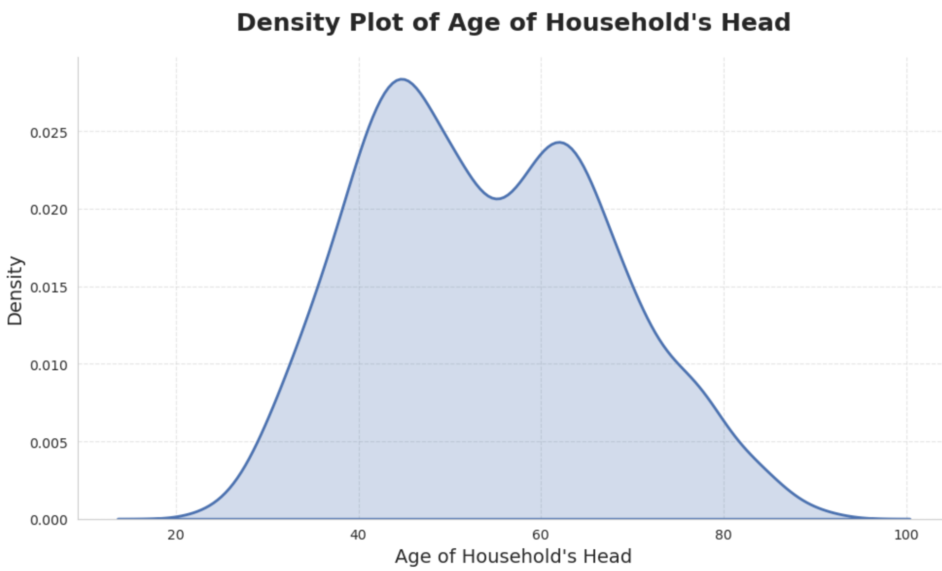

Figure 2. Density Plot of Age of Household's Head

The density and histogram plots of household head ages (Figures 1 and 2) 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.

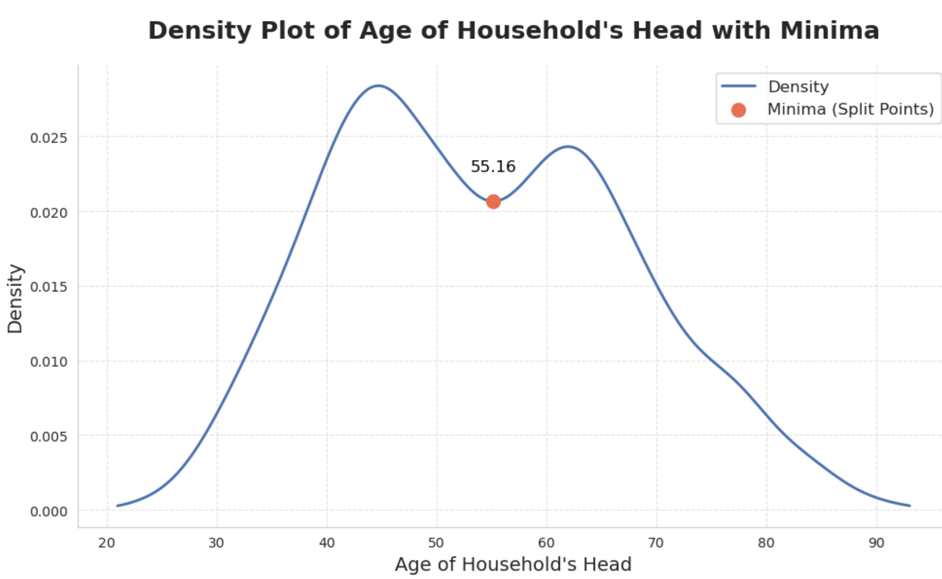

Figure 3. Density Plot of Age of Household's Head with Minima

Based on the density plot in Figure 3, 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

Age Group	**Count**	**Mean**	**SD**	**Min**	**Max**
Young	2,741	43.73	6.86	21	55
Old	2,226	66.63	7.70	56	93

As shown in Table 2, young household heads have an average age of around 44, ranging from 21 to 55, while old household heads average around 67 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.

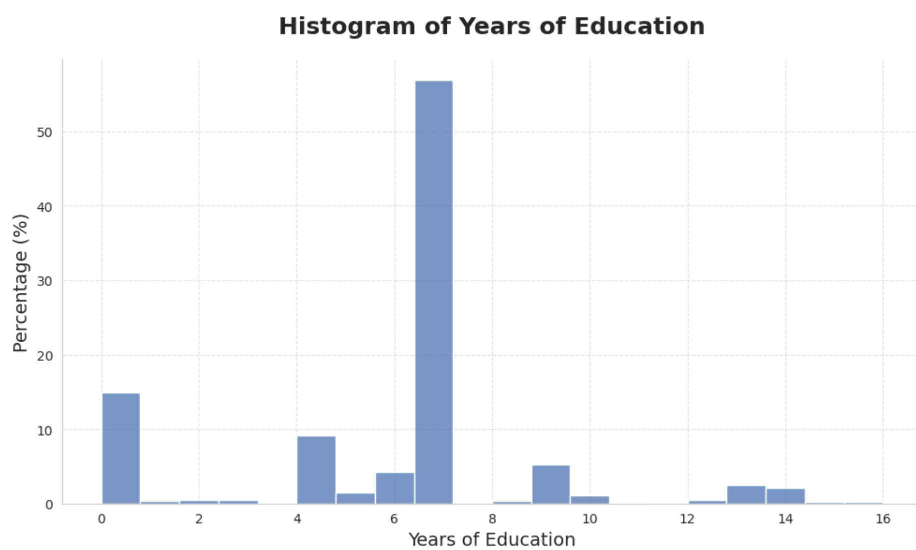

Figure 4. Histogram of Years of Education

Based on the distribution in the histogram (Figure 4), 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 and those with some or full tertiary education. This categorization provides a structured way to analyze education levels in the dataset.

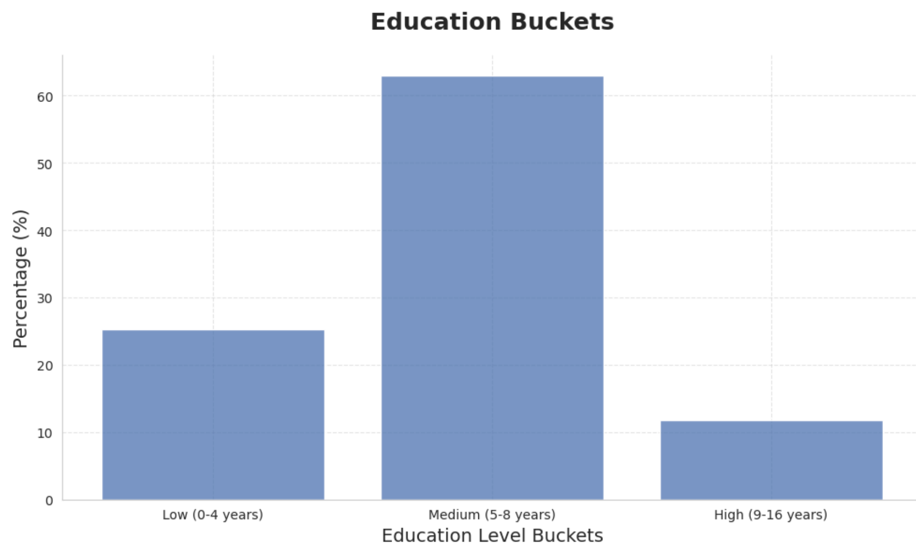

Figure 5. Histogram of Education Level Buckets

Table 3—Summary Statistics for Education Groups

Education Group	Count	Mean	SD	Min	Max
Low (0-4 years)	1,255	1.54	1.90	0	4
Medium (5-8 years)	3,124	6.89	0.40	5	8
High (9-16 years)	588	11.16	2.22	9	16

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 use this classification in our analysis. As shown in the summary table, approximately 64% of household heads are engaged in farming, a distribution similarly reflected in the histogram in Figure 6.

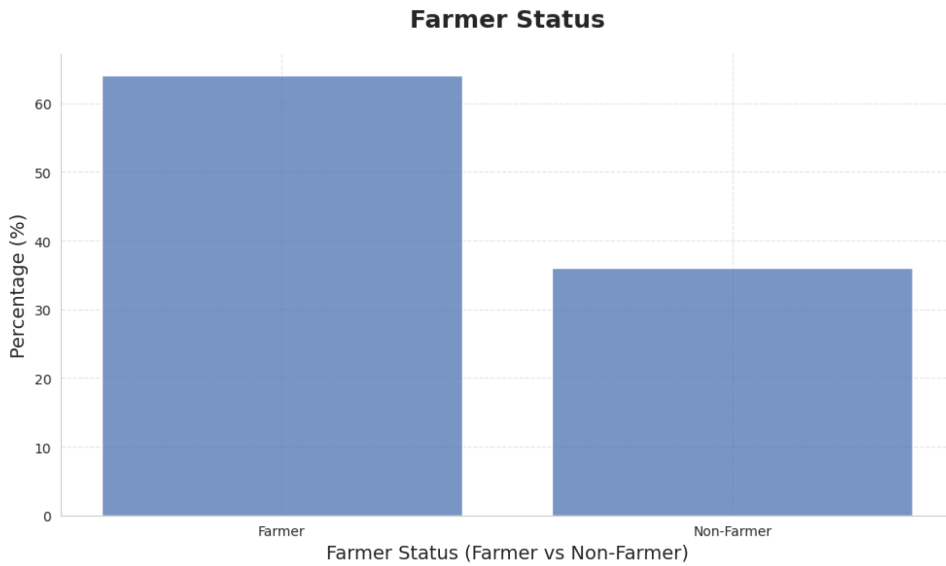

Figure 6. Histogram of Farmer Status (Farmer vs Non-Farmer)

III. Variable Selection

To build on insights from the summary statistics, we determined which heterogeneous groups to investigate by systematically evaluating the importance of household characteristics using a range of machine learning models. Specifically, we applied decision trees, random forest, boosting, bagging, LASSO, and Ridge regression to identify variables that consistently influenced new short-term credit under the Million Baht Village Fund program. This approach provided a robust justification for our variable selection, ensuring we focused on the most relevant predictors. Our analysis highlights Village Fund Credit as the central factor, with education and age of the household head emerging as critical determinants across all models. These variables align with theoretical expectations: education supports better financial decisions, while age shapes credit utilization. Household composition—including the number of male and female adults—also proved significant, reflecting its role in shaping borrowing capacity. Although occupation (farmer vs. non-farmer) ranked lower in predictive importance, it remains conceptually relevant for understanding credit behavior differences across economic activities. By focusing on Village Fund Credit, education, age, and occupation, we are able to explore how these key variables contribute to heterogeneous responses to the program.

In the sections that follow, we discuss each model's role in the selection process and how their outcomes informed our choice of variables.

A. Decision Tree

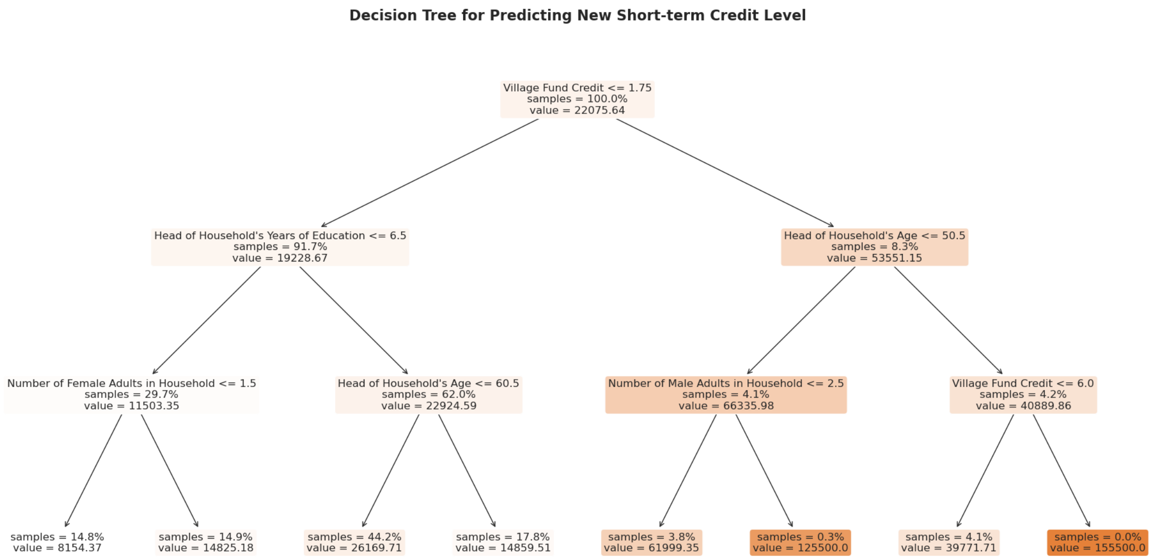

FIGURE 7. DECISION TREE FOR PREDICTING NEW SHORT-TERM CREDIT LEVEL

To identify the key predictors of new short-term credit, we constructed a decision tree (Figure 7) with a depth of three. The first split occurs at Village Fund Credit ($\leq$ 17,500 baht), indicating that households receiving less than this amount borrow less. Among these households, lower education ($\leq$ 6.5 years) further reduces borrowing, with additional splits based on the number of female adults ($\leq$ 1.5) or the household head's age ($\leq$ 60.5). Households with lower education and fewer female adults borrow the least (8,154 baht). For households with Village Fund Credit exceeding 17,500 baht, borrowing increases significantly, with further splits based on the household head's age and the number of male adults. Older household heads ($>$ 50.5 years) receiving Village Fund Credit above 60,000 baht exhibit the highest borrowing (155,500 baht).

B. Random Forest

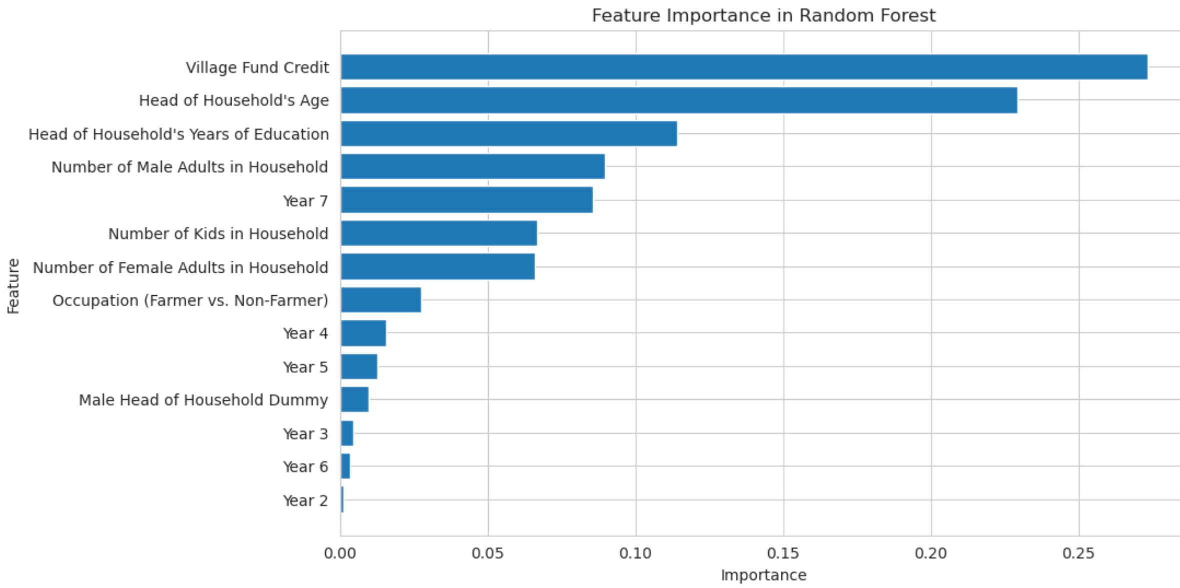

Figure 8. Feature Importance in Random Forest Model

Using a random forest model (Figure 8), we identify Village Fund Credit as the most influential factor, highlighting its central role in borrowing behavior. The household head's age and education follow in importance, while household composition, including the number of male and female adults and children, also plays a role. In contrast, occupation (farmer vs. non-farmer) is less significant, and year indicators capture only minor temporal effects.

C. Boosting

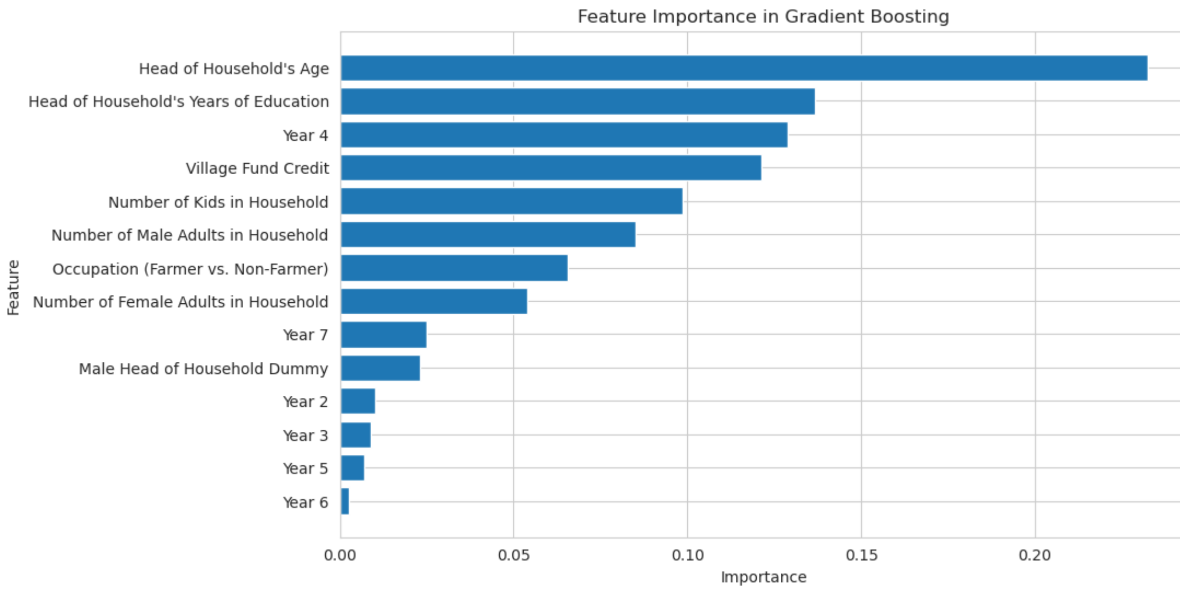

Figure 9. Feature Importance in Gradient Boosting Model

Using a gradient boosting model (Figure 9), we observe that the household head's age is the most influential factor, followed by years of education and Village Fund Credit. Household composition variables and occupation (farmer vs. non-farmer) have moderate contributions, while year indicators capture minor temporal effects, with year 7 (a post-program period) being the most notable.

D. Bagging

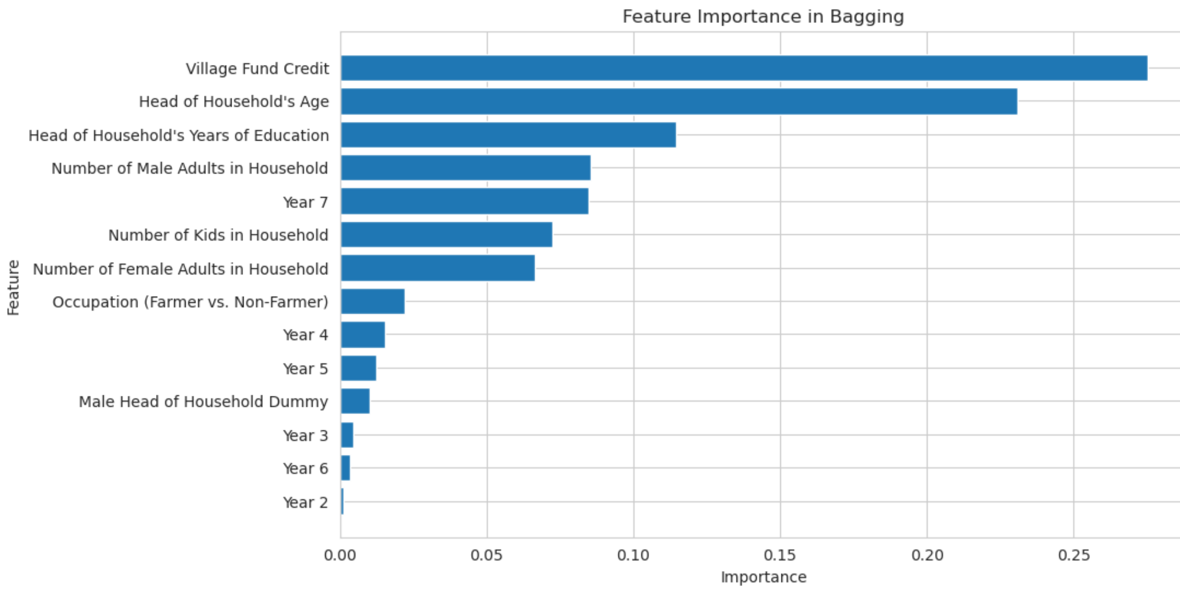

Figure 10. Feature Importance in Bagging Model

The bagging model in Figure 10 highlights Village Fund Credit as the most influential factor, followed closely by the household head's age and years of education. Household composition variables contribute moderately, while occupation (farmer vs. non-farmer) plays a minor role. Among the year indicators, year 7 exhibits the highest temporal effect, with the others showing minimal contributions.

E. LASSO

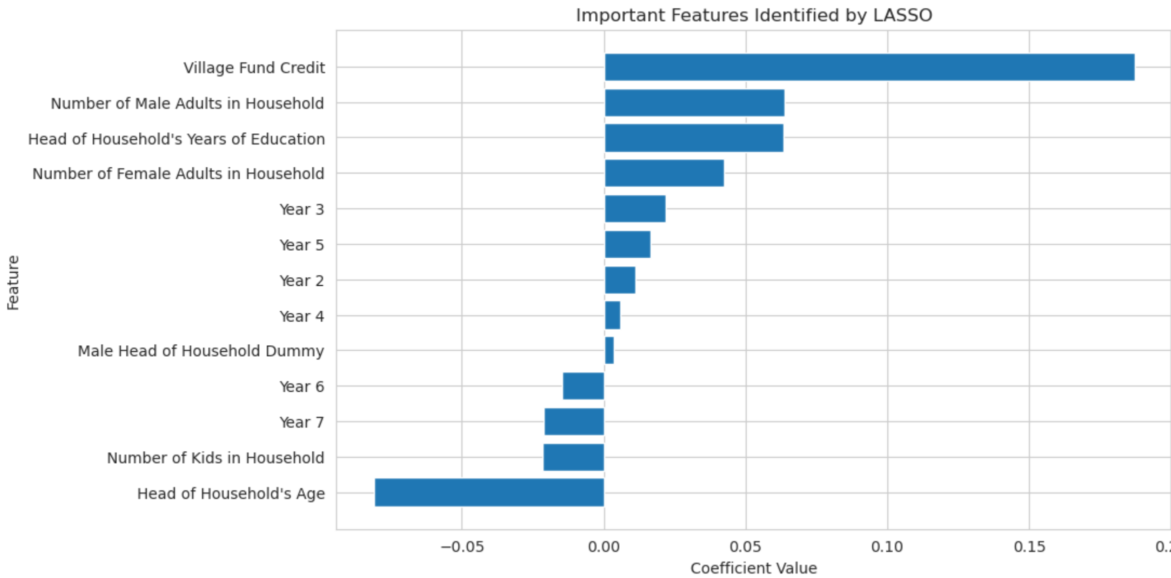

Figure 11. Important Features Identified by LASSO

Applying LASSO regression, which shrinks less influential variables to zero, we find that Village Fund Credit is the strongest predictor (Figure 11), followed by the number of male adults and years of education. The number of female adults also contributes positively, while the household head's age has a negative coefficient, suggesting that younger households are more responsive to credit availability. Temporal effects are minimal, and occupation (farmer vs. non-farmer) is dropped by LASSO. The cross-validated mean squared error (MSE) and coefficient paths are presented in the Appendix (Figures 14 and 15).

F. Ridge

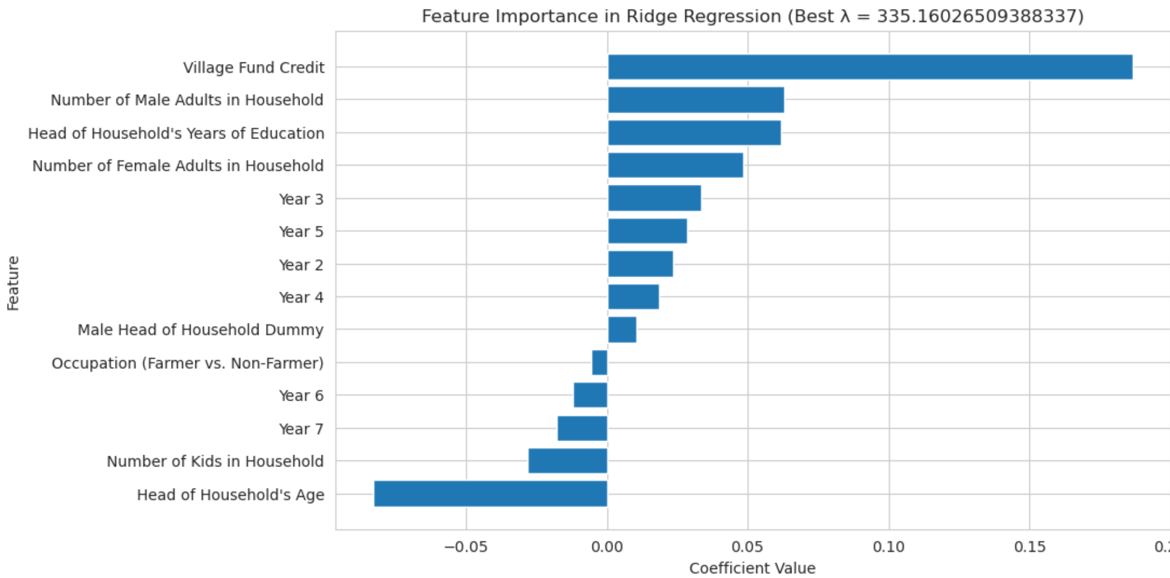

Figure 12. Feature Importance in Ridge Regression (Best λ)

Applying Ridge regression, which shrinks coefficients without reducing any to zero, we find Village Fund Credit to be the dominant predictor (Figure 12). Household composition, particularly the number of male and female adults, along with education, also play significant roles. Age, however, exhibits a notable negative relationship. Year dummies, especially Years 3 and 7, capture minor temporal effects, while occupation has minimal influence. The cross-validated mean squared error (MSE) and coefficient paths are provided in the Appendix (Figures 16 and 17).

IV. Identification Strategy

To estimate the causal impact of the Million Baht Village Fund program on household borrowing behavior, we adopt a combination of a difference-in-differences (DiD) framework and an instrumental variable (IV) approach. The DiD framework leverages panel data spanning five pre-program years (1997–2001) and six post-program years (2002–2007). Since households could not borrow from the village fund prior to 2001, changes in short-term borrowing behavior after program implementation are attributed to the availability of credit. This strategy relies on two key assumptions: parallel trends, which posits that borrowing levels would have followed similar trajectories across households in the absence of the program, and no anticipation effects, as the program's rapid and unexpected rollout following the 2000 Thai election precluded households from adjusting behavior in advance. Both assumptions are supported by pre-program trends and the program's design as described in prior work (Kaboski and Townsend, 2012*a*). However, the DiD approach alone may not fully address endogeneity concerns, such as time-varying unobservables or the selection of higher-ability households into more productive villages. To address this, we exploit exogenous variation in per-household credit availability arising from the program's design, which allocated a uniform one million baht to all villages regardless of population size. This variation makes inverse village size (1/number of households) a valid instrument for village fund credit, as smaller villages received more funds per household than larger ones. The IV approach satisfies three critical assumptions. First, the instrument is relevant, as confirmed by a highly significant first-stage regression with a robust F-statistic exceeding 10 ($F = 66.343^{***}$); see Table 11. Second, exogeneity holds because village boundaries are geopolitical units often redrawn arbitrarily for administrative purposes, and village size is uncorrelated with geographic features like roads, rivers, or pre-program borrowing trends. Third, the exclusion restriction is plausible: inverse village size influences borrowing only through credit availability, as we control for household and time fixed effects to rule out alternative pathways. Finally, the monotonicity assumption holds, as households in smaller villages consistently face higher per-household credit intensity due to the program's uniform allocation. To further mitigate concerns about the selection of more productive households into certain villages, we focus on household-level heterogeneity by interacting village fund credit with key household characteristics such as age, education, and occupation. This allows us to examine whether the response to increased credit availability varies systematically across demographic groups, addressing micro-level endogeneity concerns. By combining DiD and IV approaches, we leverage both temporal variation from program implementation and cross-sectional variation in village size, ensuring robust identification of the program's causal effects on household borrowing behavior.

V. Regression Specification

From the identification strategy, we employ an instrumental variable two-stage least squares (IV-2SLS) approach, using two variations of the inverse of village size as instruments for Village Fund Credit: invHHtvf1_{it}, which represents the inverse village size in year 6 (immediately post-implementation), and invHHtvf2_{it}, corresponding to year 7.

The first-stage regression models Village Fund Credit as a function of these instruments:

$$\text{Village Fund Credit}_{it} = \alpha + \beta_1 \cdot \text{invHHtvf1}_{it} + \beta_2 \cdot \text{invHHtvf2}_{it} + X_{it} \cdot \gamma + \mu_i + \lambda_t + \epsilon_{it}.$$

In this equation, invHHtvf1_{it} and invHHtvf2_{it} are the instruments for Village Fund Credit. X_{it} represents control variables, including the number of male and female adults and children in each household, the gender of the household head, and the household head's age, years of education, and occupation. Household fixed effects are captured by μ_i, year fixed effects by λ_t, and ϵ_{it} denotes the error term. Standard errors are clustered at the village-year level to account for potential within-cluster correlation.

The second stage estimates the impact of predicted Village Fund Credit on new short-term credit. To capture heterogeneous effects across age, education, and occupation subgroups, we include interaction terms between predicted Village Fund Credit and group indicators. The regression is specified as:

$$\begin{aligned}\text{New Short-Term Credit}_{it} = \alpha &+ \beta_1 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{Young}_{it} \\ &+ \beta_2 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{Medium}_{it} \\ &+ \beta_3 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{High}_{it} \\ &+ \beta_4 \cdot \text{Village } \widehat{\text{Fund}} \text{ Credit}_{it} \cdot \text{Non-Farmer}_{it} \\ &+ \beta_5 \cdot \text{Young}_{it} + \beta_6 \cdot \text{Medium}_{it} + \beta_7 \cdot \text{High}_{it} + \beta_8 \cdot \text{Non-Farmer}_{it} \\ &+ X_{it} \cdot \gamma + \mu_i + \lambda_t + \epsilon_{it}.\end{aligned}$$

In this equation, $\text{Village } \widehat{\text{Fund}} \text{ Credit}_{it}$ represents the predicted values from the first stage. Interaction terms, such as Village $\widehat{\text{Fund}}$ Credit$_{it}\cdot$Young$_{it}$, capture the differential effects across demographic subgroups. The subgroup indicators, Young$_{it}$, Medium$_{it}$, High$_{it}$, and Non-Farmer$_{it}$, enable direct estimation of group-level effects. Control variables X_{it}, such as the number of male and female adults and children in each household and the gender of the household head, are included alongside household fixed effects (μ_i) and year fixed effects (λ_t). Standard errors are clustered at the village-year level to ensure robust inference.

A. Directed Acyclic Graph (DAG)

To complement our regression specification, Figure 13 presents a Directed Acyclic Graph (DAG) that visually depicts the causal pathway of our instrumental variable two-stage least squares (IV-2SLS) approach. In the first stage, inverse village size serves as an instrument for Village Fund Credit, generating predicted values. These predicted values are interacted with subgroup indicators (e.g., age, education, and occupation) to capture heterogeneous effects on new short-term credit. Control variables account for confounding factors to mitigate bias from observed household characteristics. Fixed effects, though not explicitly shown in the DAG due to their association with multiple variables, are incorporated in the regression model to control for household- and year-specific effects.

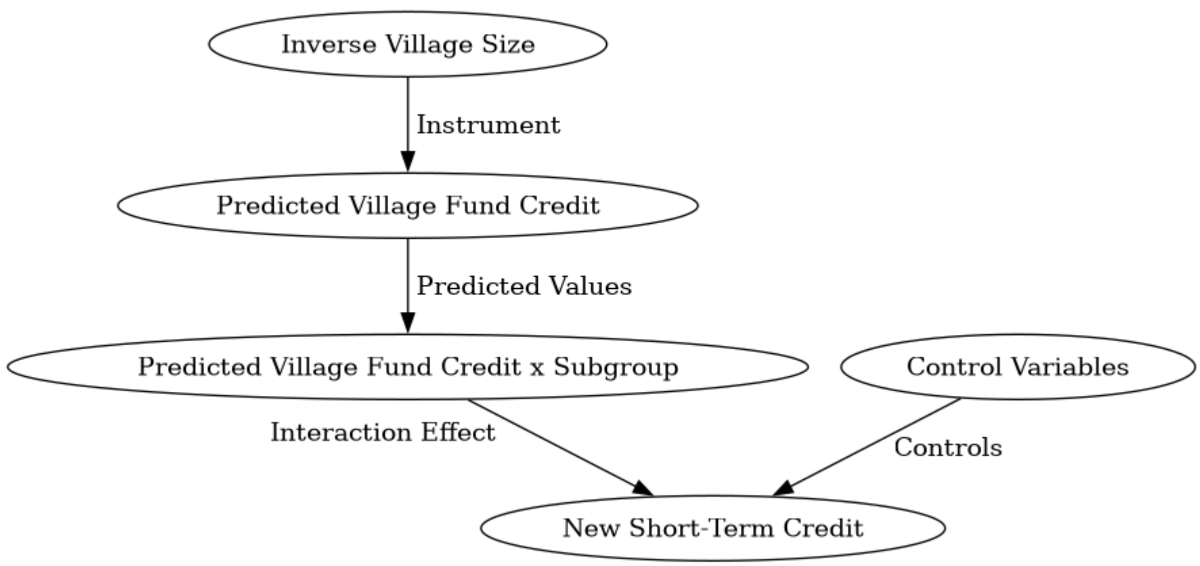

FIGURE 13. DAG FOR IV-2SLS SPECIFICATION

VI. Results

In this section, we present the results of the second-stage IV-2SLS regression, which estimates the heterogeneous effects of Village Fund Credit using predicted values from the first stage. In the first stage, Village Fund Credit is instrumented with the inverse village size (invHHtvf1 and invHHtvf2), both of which are highly significant, with coefficients of 46.392 and 85.367 at the 5% and 1% levels, respectively (Appendix Table 11). The robust F-statistic of 66.343 confirms strong instrument relevance, well above the critical threshold of 10, alleviating concerns about weak instruments. These first-stage results ensure the validity of the predicted values used in the second stage, where we analyze the heterogeneous borrowing responses across age, education, and occupation groups.

Table 4—2SLS IV Regression: Heterogeneous Effects on New Short-Term Credit

Interactions	**Second Stage Results**
Young	9,065.6***
	(3,401.7)
Medium Education	4,097.1
	(3.282.9)
High Education	7,910
	(7,521.5)
Non-Farmer	-533.3
	(3,463.6)
Control Variables	✓
Year Fixed Effects	✓
Village Fixed Effects	✓
Observations	4,967
R-squared (Within)	0.0232

Notes: The regressions control for household characteristics, including the number of male and female adults, children, and male household head. Year and village fixed effects are included. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors, clustered by village-year, are reported in parentheses.

The results in Table 4 reveal significant heterogeneous effects of Village Fund Credit on new short-term credit across age, education, and occupation groups, with all effects measured per 10,000 baht increase in Village Fund Credit.

For younger households (relative to older households), a 10,000 baht increase leads to an additional 9,065.6 baht borrowed, statistically significant at the 1% level. This effect translates to approximately 262.6 USD (at the 2007 exchange rate of 34.518 baht/USD (University of British Columbia, 2007)), which is 123.2% of Thailand's average monthly earnings (7,357.4 baht (International Labour Organization, 2007)). Relative to the standard deviation of new short-term credit (53,437.8 baht), this represents 16.96%, highlighting a substantial borrowing response among younger households. This result suggests that younger households, with their longer investment horizons and higher risk tolerance, are more likely to leverage credit when it becomes available.

For education, the results show a positive but statistically insignificant relationship. Households with medium education (5-8 years) borrow an additional 4,097.1 baht per 10,000 baht increase in Village Fund Credit, equivalent to 118.7 USD and 55.7% of average monthly earnings (7,357.4 baht (International Labour Organization, 2007)). This effect represents 7.67% of the standard deviation of new short-term credit. For households with high education (9-16 years), borrowing increases by 7,910 baht, or approximately 229.1 USD, equivalent to 107.5% of average monthly earnings. Relative to the standard deviation, this effect accounts for 14.8%. These results suggest that higher education levels may enhance the ability to utilize credit effectively, likely through improved financial literacy or access to productive investments, though the lack of statistical significance indicates these relationships are not robust.

For occupation, the results reveal that farming households borrow slightly more than non-farming households. The coefficient for non-farming households is -533.3 baht, implying that farmers borrow 533.3 baht more per 10,000 baht increase in Village Fund Credit. This effect translates to 15.4 USD, or 7.2% of average monthly earnings, and represents 1.0% of the standard deviation of new short-term credit. Although this result is not statistically significant, it indicates that farming households may depend more on credit to finance agricultural inputs, seasonal production, or working capital needs.

In summary, the results demonstrate that younger, more educated, and farming households exhibit a more pronounced borrowing response to Village Fund Credit. Among these, only the effect for younger households is statistically significant, suggesting that age is the most robust driver of credit utilization. The economically meaningful but statistically insignificant effects for education and farming highlight potential benefits for these groups that may warrant further investigation.

Robustness Check. — To assess the robustness of our results, we compare the IV-2SLS estimates (Table 4) with those from a simple ordinary least squares (OLS) regression that does not use the instrument (Table 5). The IV-2SLS results show that younger households borrow 9,065.6 baht per 10,000 baht increase in Village Fund Credit, significant at the 1% level. By contrast, the OLS estimate of 7,134.3 baht is lower, suggesting that unobserved constraints or credit limitations for younger households may bias the OLS estimates downward. For education, OLS estimates significant effects of 7,679.5 baht for medium education and 7,203.6 baht for high education. However, the IV-2SLS results yield smaller coefficients of 4,097.1 baht and 7,910 baht, respectively, which lose statistical significance. This suggests that OLS may overstate education effects by capturing unobserved traits, such as ability or financial acumen, that correlate with both education and borrowing. For occupation, both methods produce small and insignificant estimates. The OLS result of -917.6 baht and the IV-2SLS result of -533.3 baht indicate minimal differences in borrowing behavior between farming and non-farming households. Overall, IV-2SLS yields larger and more reliable estimates for younger households while providing more conservative effects for education groups, reinforcing the importance of addressing endogeneity. Despite the lack of statistical significance for some estimates, the consistency of coefficients across both estimation methods strengthens the validity of our findings.

Table 5—OLS Regression: Heterogeneous Effects on New Short-Term Credit

Interactions	**OLS Results**
Young	7,134.3***
	(2,257.2)
Medium Education	7,679.5***
	(1,720.3)
High Education	7,203.6*
	(4,153.4)
Non-Farmer	-917.6
	(2,675.6)
Control Variables	✓
Year Fixed Effects	✓
Village Fixed Effects	✓
Observations	4,967
R-squared (Within)	0.0420

Notes: The regressions control for household characteristics, including the number of male and female adults, children, and male household head. Year and village fixed effects are included. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors, clustered by village-year, are reported in parentheses.

To further evaluate the validity of our findings, we apply the Hausman test to assess endogeneity. As shown in Table 6, we fail to reject the null hypothesis of no endogeneity, with a p-value of 0.3170. This indicates that the OLS estimates are consistent and unbiased, as there is no strong evidence of correlation between Village Fund Credit and the error term. Consequently, the need for IV estimation becomes less critical, suggesting that OLS sufficiently captures the variation in borrowing behavior. Nonetheless, our instrument remains strong, with a first-stage F-statistic of 66.343, well above the conventional threshold of 10. While the IV approach addresses potential endogeneity concerns, the Hausman test confirms the reliability of OLS estimates. Importantly, both OLS and IV-2SLS models yield similar coefficient estimates, reinforcing the robustness of our results. We conclude that younger and more educated households borrow more in response to Village Fund Credit, while farming households exhibit a slightly higher borrowing response than non-farming households. However, the statistical insignificance of education in the IV-2SLS results and occupation effects in both models warrants further investigation. In the next section, we introduce additional models to enhance the precision and depth of our conclusions.

TABLE 6—HAUSMAN TEST FOR ENDOGENEITY

Hausman Statistic	**P-value**
1.0013	0.3170

Note: The table reports the Hausman test statistic and p-value for endogeneity. The null hypothesis of no endogeneity cannot be rejected ($p = 0.3170$), indicating that OLS is consistent.

VII. Additional Results

In this section, we introduce new models to provide further insights into the heterogeneous effects of Village Fund Credit on new short-term credit. The results are organized into three subsections. First, we explore IV-based models, such as Double Machine Learning (DML) and Deep IV, to address potential endogeneity and validate our IV-2SLS estimates. Second, motivated by the Hausman test results, which suggest that OLS estimates are consistent, we run models that do not rely on the IV framework. These include Propensity Score Matching (PSM), Inverse Probability Weighting (IPW), and meta-learners (S-Learner, T-Learner, and X-Learner), as well as Doubly Robust (DR) learners. To incorporate the DID framework, Village Fund Credit is treated as a binary treatment: households with credit equal to zero are assigned 0, and households with credit greater than zero are assigned 1. Finally, in the third subsection, we employ a Causal Forest model to analyze Village Fund Credit under three scenarios: using the predicted values, the actual (unpredicted) values, and the binary treatment. This approach allows us to compare results across multiple modeling frameworks and deepen our understanding of the program's impact.

A. IV-Based Models

In this subsection, we apply Double Machine Learning (DML) and Deep IV models to further evaluate the heterogeneous effects of Village Fund Credit. Table 7 presents the results.

Table 7—IV-Based Models: Heterogeneous Effects on New Short-Term Credit

Interactions	**DML Results**	**Deep IV Results**
Young	2,936.9	1,805.2**
	[-2,922.4, 8,796.3]	[1,318.6, 2,221.2]
Medium Education	2,389.2	2,225.9**
	[-5,735.3, 10,513.8]	[1,627.5, 2,747.7]
High Education	-1,245.8	338.8**
	[-12,440.0, 9,952.9]	[248.9, 414.5]
Non-Farmer	1,298.5	1,485.1**
	[-4,796.5, 7,393.5]	[1,083.5, 1,832.4]
Control Variables	✓	✓
Year Fixed Effects	✓	✓

Notes: Results are presented for Double Machine Learning (DML) and Deep IV models, with 95% confidence intervals in brackets. Control variables include the number of male adults, female adults, children, male household head, farming status, age of the household head, and years of education. Year fixed effects are included. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The DML model, which controls for confounders using Random Forests, yields smaller and statistically insignificant effects across all subgroups. For younger households, a 10,000 baht increase in Village Fund Credit is associated with an additional borrowing of 2,936.9 baht, much lower and insignificant compared to the IV-2SLS estimate of 9,065.6 baht. Medium education households borrow 2,389.2 baht more than low education households, while high education households borrow 1,245.8 baht less, both effects remaining statistically insignificant. Non-farming households borrow 1,298.5 baht more than farming households, but this effect is also insignificant. These results suggest that the DML model fails to produce conclusive evidence, likely due to the high dimensionality of machine learning methods, which may struggle to identify causal effects in smaller samples.

The Deep IV model, on the other hand, generates more precise and statistically significant estimates, offering clearer insights. For younger households, a 10,000 baht increase in Village Fund Credit leads to an additional borrowing of 1,805.2 baht, significant at the 5% level, though notably smaller than the IV-2SLS estimate. Medium education households borrow 2,225.9 baht more than low education households, and high education households borrow 338.8 baht more, both significant at the 5% level. Non-farming households borrow 1,485.1 baht more than farming households, also significant at the 5% level. These results contrast with both IV-2SLS and OLS models, where education groups and non-farming households exhibited insignificant effects. By addressing endogeneity and incorporating non-linearities, Deep IV appears to uncover more refined treatment effects.

Overall, Deep IV identifies significant borrowing responses for younger, medium education, high education, and non-farming households, albeit with smaller magnitudes compared to IV-2SLS. In contrast, the DML model fails to detect significant effects across any subgroup. These findings suggest that Deep IV provides a more robust and nuanced understanding of the treatment effects, particularly for younger, more educated, and non-farming households, where other methods, including IV-2SLS and OLS, yielded inconclusive results.

B. Binary Treatment Models

In this subsection, we examine the heterogeneous effects of Village Fund Credit by treating it as a binary variable. Specifically, households that receive any credit (*Village Fund Credit ¿ 0*) are classified as treated (1), while those receiving no credit (*Village Fund Credit = 0*) are classified as control (0). Given the Hausman test results, which confirmed the consistency of OLS, we apply this framework to a range of non-IV models to provide alternative estimates. We begin with Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW), followed by meta-learner models. This approach allows us to assess the robustness of our findings using methods that do not rely on instrumental variables.

PSM and IPW Models. — In this subsection, we evaluate the heterogeneous effects of Village Fund Credit using Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW). Table 8 presents the results.

Table 8—PSM and IPW Models: Heterogeneous Effects on New Short-Term Credit

Interactions	**PSM Results**	**IPW Results**
Young	25,054.8**	37,969.5**
	[16,624.9, 32,858.2]	[15,626.0, 61,343.7]
Medium Education	21,441.4**	25,840.8**
	[15,816.1, 27,191.6]	[9,023.5, 43,504.4]
High Education	27,621.3**	25,777.3**
	[3,930.0, 49,977.7]	[9,514.3, 43,480.0]
Non-Farmer	21,719.2**	23,333.1**
	[12,501.1, 31,241.3]	[1,922.5, 50,808.1]
Control Variables	✓	✓
Year Fixed Effects	✓	✓

Notes: Results are presented for Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW) models, with 95% confidence intervals in brackets. Control variables include the number of male adults, female adults, children, male household head, farming status, age of the household head, and years of education. Year fixed effects are included. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The PSM model estimates significant borrowing responses across all subgroups. Younger households borrow an additional 25,054.8 baht upon receiving Village Fund Credit, significant at the 5% level. Medium education and high education households borrow 21,441.4 baht and 27,621.3 baht, respectively, while non-farming households show a borrowing response of 21,719.2 baht, all statistically significant. Balance diagnostics (Appendix Table 13) confirm that PSM effectively reduces covariate imbalances, with most standardized mean differences nearing zero post-matching. Minor imbalances persist for a few covariates, but Appendix Figure 18 demonstrates substantial improvement in the overlap between treated and control groups, enhancing comparability.

The IPW model similarly produces significant effects. Younger households borrow an additional 37,969.5 baht, significant at the 5% level. Medium and high education households borrow 25,840.8 baht and 25,777.3 baht, respectively, while non-farming households show a borrowing response of 23,333.1 baht, all significant at the 5% level. Covariate balancing diagnostics (Appendix Table 14) show that IPW improves balance across most covariates, though small residual imbalances persist. Figure 19 confirms the reweighting effectiveness, as treated and control groups align more closely, albeit imperfectly.

Compared to IV-2SLS, both PSM and IPW yield substantially larger treatment effects. For example, while IV-2SLS estimated a borrowing effect of 9,065.6 baht for younger households, PSM and IPW report 25,054.8 baht and 37,969.5 baht, respectively. Similarly, medium and high education groups, which showed insignificant effects under IV-2SLS, exhibit significant positive responses in both PSM and IPW. Non-farming households, previously insignificant under IV-2SLS and OLS, also show significant effects under PSM and IPW. These differences may stem from the binary treatment specification, which simplifies the treatment variable and can amplify effect estimates relative to the continuous specification in IV-2SLS. However, the alignment of PSM and IPW results with the Deep IV estimates—where significant effects were also observed for younger, more educated, and non-farming households—reinforces the robustness of these findings. Overall, the PSM and IPW Modelsmodels provide strong evidence that younger, more educated, and non-farming households exhibit significant borrowing responses to Village Fund Credit.

Meta-Learner Models. — In this subsection, we analyze the heterogeneous effects of Village Fund Credit using meta-learner models: S-Learner, T-Learner, X-Learner, and Doubly Robust (DR) Learner. Table 9 presents the results.

Table 9—Meta-Learner Models: Heterogeneous Effects on New Short-Term Credit

Interactions	**S-Learner**	**T-Learner**	**X-Learner**	**DR-Learner**
Young	24,450.5**	26,625.4**	30,037.2**	24,070.3**
	[17,686.2, 31,742.6]	[18,363.8, 35,727.5]	[18,437.0, 39,125.6]	[12,855.8, 34,898.0]
Medium Education	18,875.8**	20,723.9**	23,179.2**	24,535.6**
	[13,328.1, 26,019.0]	[13,489.4, 29,848.4]	[15,092.7, 31,819.1]	[13,484.6, 35,913.2]
High Education	26,034.8**	28,760.5**	32,293.8**	24,385.9**
	[12,418.0, 41,919.8]	[14,460.3, 46,459.1]	[13,286.7, 52,304.7]	[13,559.5, 34,988.8]
Non-Farmer	16,397.4**	17,938.6**	21,120.7**	24,042.7**
	[9,456.2, 25,298.6]	[8,336.8, 27,941.6]	[8,252.6, 32,673.9]	[12,875.0, 35,706.4]
Control Variables	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓

Notes: Results are presented for S-Learner, T-Learner, X-Learner, and Doubly Robust (DR) Learner models, with 95% confidence intervals in brackets. Control variables include the number of male adults, female adults, children, male household head, farming status, age of the household head, and years of education. Year fixed effects are included. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The meta-learner models produce consistent and statistically significant effects across all subgroups. For younger households, receiving Village Fund Credit leads to additional borrowing ranging from 24,070.3 baht (DR-Learner) to 30,037.2 baht (X-Learner), all significant at the 5% level. These estimates closely align with the large effects identified by PSM and IPW while exceeding the smaller effects from IV-2SLS and Deep IV. For medium education households, receiving Village Fund Credit results in borrowing increases between 18,875.8 baht (S-Learner) and 24,535.6 baht (DR-Learner), confirming significant borrowing responses consistent with PSM, IPW, and Deep IV models. High education households exhibit borrowing effects ranging from 24,385.9 baht (DR-Learner) to 32,293.8 baht (X-Learner), providing strong evidence of significant responses in contrast to the insignificant results from IV-2SLS. For non-farming households, receiving Village Fund Credit increases borrowing by 16,397.4 baht (S-Learner) to 24,042.7 baht (DR-Learner), all significant at the 5% level. These findings align closely with the PSM and IPW models, highlighting significant borrowing effects for non-farming households. This contrasts with the insignificant and negative results from the IV-2SLS and OLS models.

Therefore, compared to IV-2SLS, which estimated smaller or insignificant effects for most subgroups, the meta-learner models reveal consistently larger and significant borrowing responses. These results align closely with the findings from Deep IV, PSM, and IPW, providing robust evidence that receiving Village Fund Credit significantly increases borrowing among younger, more educated, and non-farming households.

C. Causal Forest Models

In this section, we consolidate the Causal Forest results to assess the heterogeneity of Village Fund Credit effects across key subgroups—age, education, and occupation—using binary, unpredicted, and predicted treatments. This analysis provides a comprehensive view of treatment effect variation, offering complementary insights to earlier models such as IV-2SLS, OLS, and binary treatment approaches. Table 10 summarizes the subgroup effects, while Appendix Figures 20, 21, and 22 illustrate the treatment effect distributions.

Table 10—Causal Forest Models: Heterogeneous Treatment Effects by Subgroup

Subgroup	**Binary Treatment**	**Unpredicted Treatment**	**Predicted Treatment**
Young	24,291	16,698	28,261
	(4,424)	(2,235)	(11,323)
Medium Education	22,941	16,354	24,278
	(3,838)	(1,775)	(8,192)
High Education	24,629	15,322	30,968
	(6,282)	(3,392)	(18,850)
Non-Farming	22,834	16,366	32,152
	(4,467)	(2,093)	(11,721)
Control Variables	✓	✓	✓
Year Fixed Effects	✓	✓	✓

Notes: Results are presented for Causal Forest models using three treatments: binary treatment (0/1), unpredicted treatment, and predicted treatment. Unpredicted treatment refers to the continuous Village Fund Credit given in the data, while predicted treatment refers to the values predicted in the first stage of the IV-2SLS. Means are reported with standard deviations in parentheses. Control variables include the number of male adults, female adults, children, and the gender of the household head. Year fixed effects are included.

Under the binary treatment, the largest effect is observed for high-education households at 24,629 baht, followed closely by younger households (24,291 baht), medium-education households (22,941 baht), and non-farming households (22,834 baht). As shown in Appendix Figure 20, treatment effects for younger and high-education households are more dispersed, with slight left skewness, reflecting greater borrowing variability and stronger responses in these groups.

For the unpredicted treatment, the largest borrowing response comes from younger households at 16,698 baht, followed by non-farming households (16,366 baht), medium-education households (16,354 baht), and high-education households (15,322 baht). Appendix Figure 21 reveals more concentrated treatment effects, particularly for younger and medium-education households, indicating a relatively uniform borrowing response under this treatment.

Under the predicted treatment, the largest effect is seen for non-farming households, which borrow an additional 32,152 baht. This is followed by high-education households (30,968 baht), younger households (28,261 baht), and medium-education households (24,278 baht). As depicted in Appendix Figure 22, the treatment effects are significantly right-skewed, especially for non-farming and high-education households, reflecting their higher borrowing potential when treatment is predicted.

By analyzing all treatment types together, the Causal Forest models reveal a consistent pattern: non-farming households, more educated households, and younger households exhibit the largest borrowing responses, with predicted treatment effects showing the largest magnitudes.

These results align with earlier findings from meta-learners, PSM, IPW, and Deep IV models, but contradict some findings from IV-2SLS and OLS models, such as the lower borrowing observed for non-farming households compared to farming households.

VIII. Conclusion

This study investigates the heterogeneous effects of Thailand's Village Fund Credit program across key socio-economic groups—age, education, and occupation. While initial IV-2SLS estimates suggested a stronger borrowing response among younger households but insignificant effects for education and occupation, further analysis using advanced methods (Meta-Learners, Propensity Score Matching, Inverse Probability Weighting, Deep IV, and Causal Forests) consistently highlights three significant findings: younger, more educated, and non-farming households borrow significantly more in response to the program. These results suggest that prior models may have understated the roles of education and occupation, particularly the borrowing potential of non-farming households.

Younger households demonstrate a pronounced borrowing response, likely due to longer investment horizons and greater risk tolerance. High- and medium-education households also respond strongly, reflecting the role of education in enhancing financial literacy and credit decision-making. Non-farming households, initially underestimated, emerge as substantial borrowers across newer models, underscoring their access to diverse, high-return investment opportunities outside of agriculture. These findings expand upon Kaboski and Townsend's (2012) aggregate effects, offering a more granular understanding of the program's impact.

Our results align with broader microfinance research, including Awojobi (2011) on Nigeria and Mahjabeen (2008) on Bangladesh, which emphasize poverty alleviation through microfinance. However, our analysis reveals that benefits are unevenly distributed, favoring younger, more educated, and non-farming households who are better positioned to leverage credit productively. This underscores the need for targeted program designs to maximize economic impact.

The robustness of our findings across diverse methods provides actionable insights for policymakers and financial institutions. Targeting younger, educated, and non-farming households can enhance financial inclusion and economic growth by allocating credit to those most capable of utilizing it effectively. This aligns with Vanroose and D'Espallier (2009), who emphasize the complementary role of microfinance in bridging gaps left by traditional banking systems.

Despite these contributions, our study has limitations. The IV-2SLS approach depends on the exogeneity of the instrument, while binary and predicted treatments in advanced models may misestimate responses relative to continuous treatments. Regional focus on specific Thai provinces may constrain generalizability, and complex models like Deep IV or Causal Forests may pose challenges for policy communication.

Future research should address these limitations by employing quantile regression to analyze responses across borrowing levels, testing robustness in diverse contexts, and advancing causal machine learning techniques. These efforts will further refine microfinance evaluations and deepen the understanding of heterogeneous treatment effects.

In conclusion, this study provides robust evidence that younger, more educated, and non-farming households exhibit stronger borrowing responses to Village Fund Credit. By integrating multiple methods, we offer a nuanced perspective on microfinance impacts, emphasizing the importance of targeted interventions to promote inclusive economic development and reduce financial disparities in underserved communities.

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`.

International Labour Organization. 2007. "Average monthly earnings by major economic activity." Retrieved from ILO Laborsta database.

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.

University of British Columbia. 2007. "Exchange Rate Conversion Tables." Retrieved from the UBC FX Sauder website.

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. First-Stage Results for 2SLS IV Regression

To validate our instrumental variable approach, we present the first-stage regression results in Table 11. 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 11—First-Stage Regression

Instrumental Variables	**Coefficient Estimates**
invHHtvf1	46.392**
	(19.075)
invHHtvf2	85.367***
	(9.740)
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 first-stage regression includes household characteristics as control variables, such as the number of male and female adults (*madult, fadult*), children, male household head, farming status, age, and years of education. Year and entity fixed effects are included. Standard errors, clustered at the village-year level, are reported in parentheses. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B. 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 12 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. These metrics also guided the development of the models presented in the additional results section.

Table 12—Model Performance Comparison

Model	**MSE**	**RMSE**
Decision Tree	2,690,333,000	51,868.42
Random Forest	2,435,029,000	49,346.01
Boosting	1,817,375,000	42,630.68
Bagging	2,434,179,000	49,337.40

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.

C. Cross-validated MSE for LASSO

Figure 14 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.

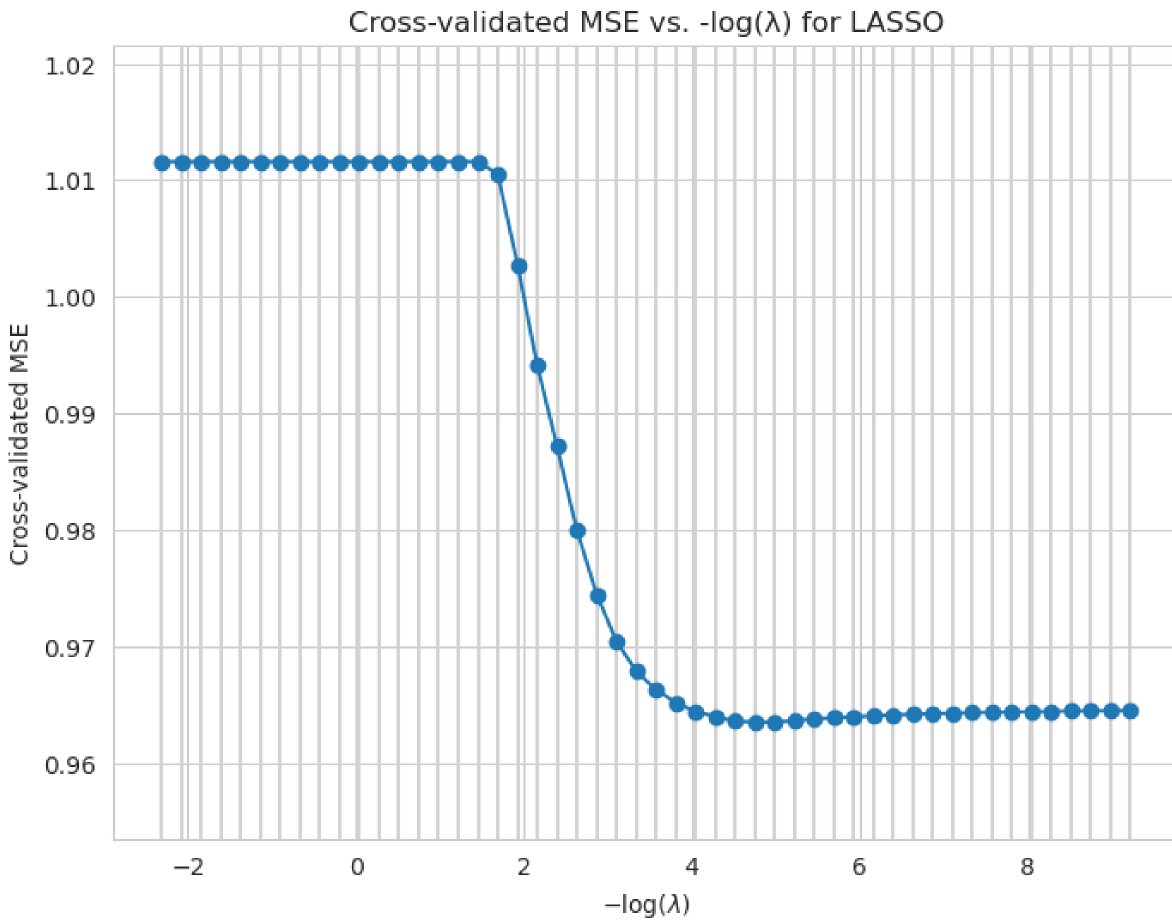

Figure 14. Cross-validated MSE vs. -log(λ) for LASSO

D. LASSO Coefficient Paths

Figure 15 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.

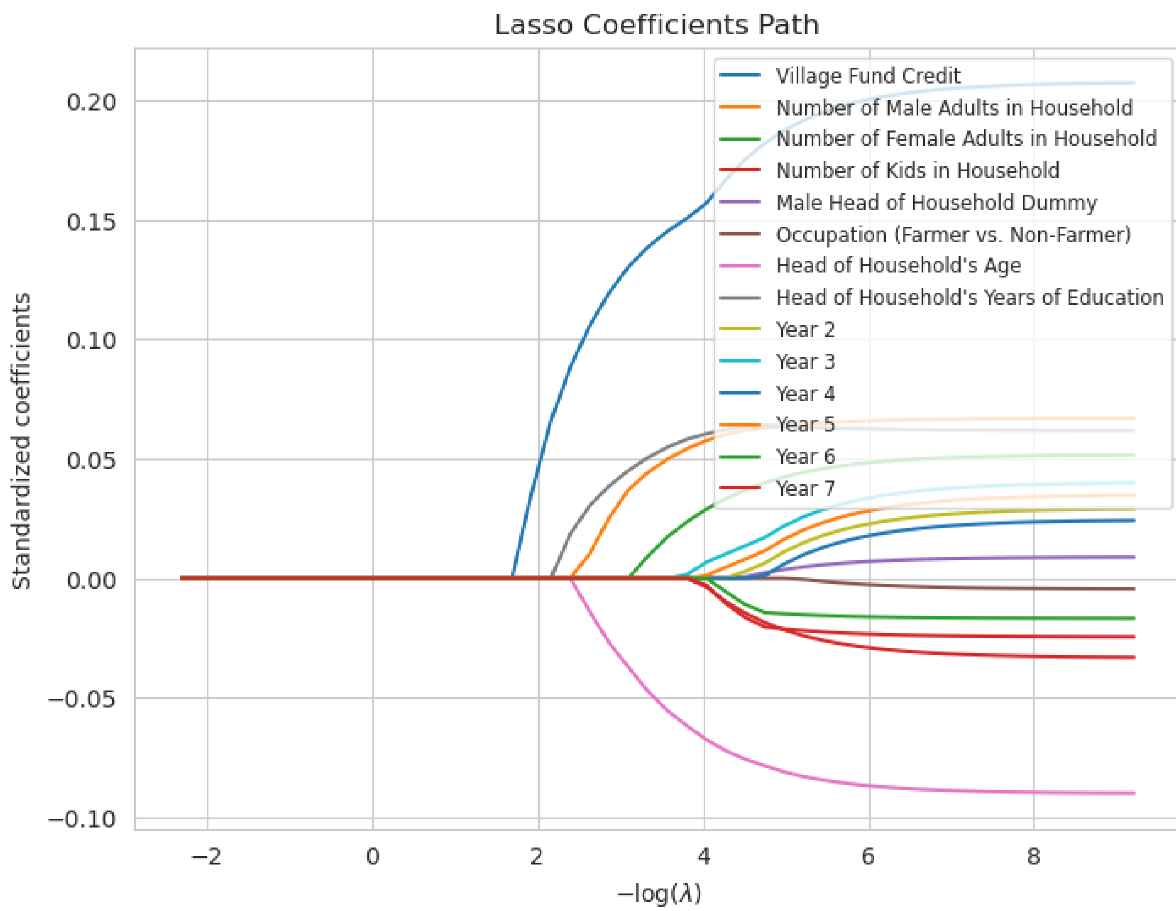

Figure 15. LASSO Coefficients Path for Feature Selection

E. Cross-validated MSE for Ridge

Figure 16 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.

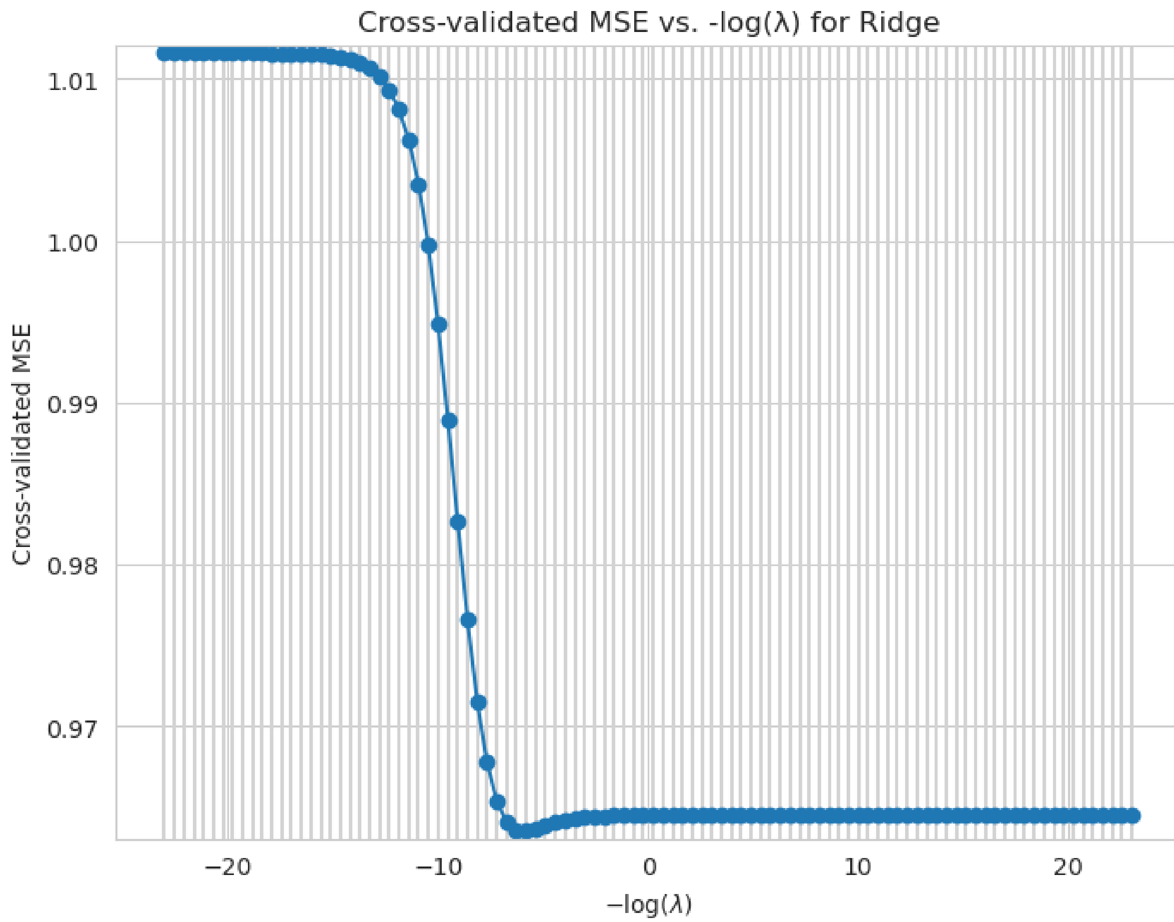

FIGURE 16. CROSS-VALIDATED MSE VS. -LOG(λ) FOR RIDGE

F. Ridge Coefficient Paths

The coefficient paths for Ridge regression, shown in Figure 17, 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.

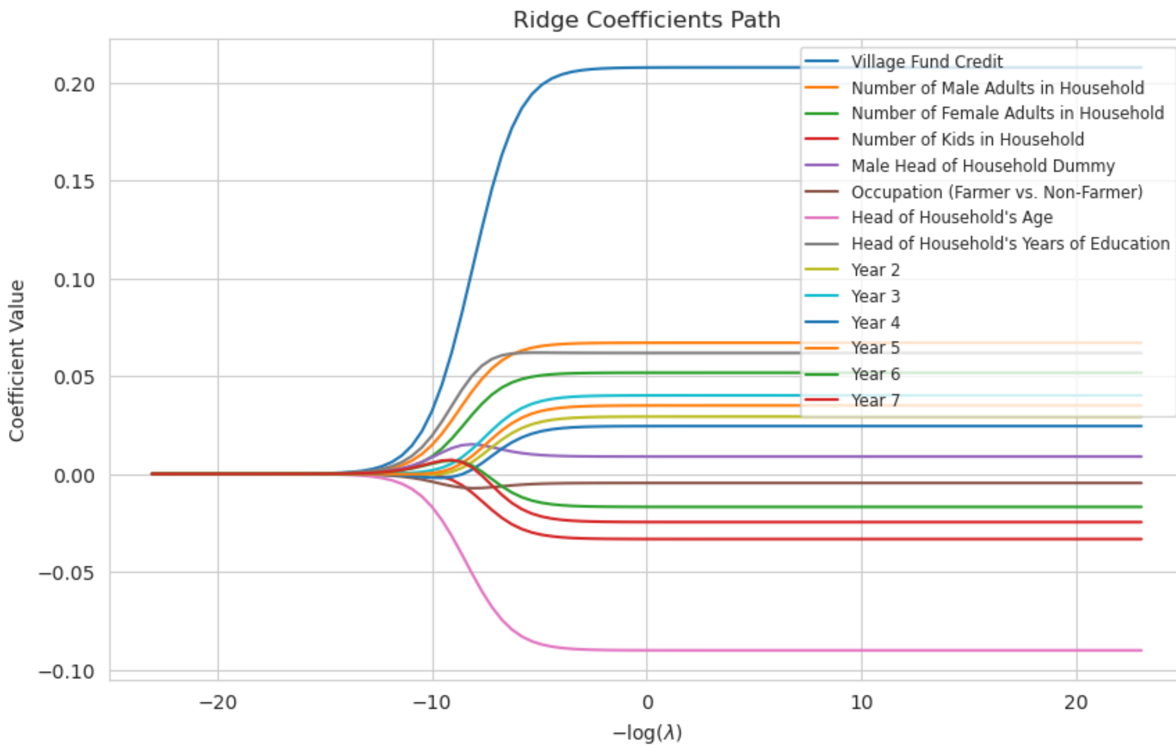

Figure 17. Ridge Coefficients Path for Feature Selection

G. Propensity Score Matching (PSM) Analysis

Table 13 compares standardized mean differences for covariates before and after Propensity Score Matching (PSM). The absolute changes indicate the improvement in balance across treated and control groups.

Table 13—Balance Table: Standardized Mean Differences Before and After Matching

Variable	**Pre-Matching**	**Post-Matching**	**Absolute Change**
Male Adults (madult)	0.0600	-0.0102	0.0702
Female Adults (fadult)	0.0852	0.0841	0.0011
Children (kids)	0.0540	0.0351	0.0189
Male Household Head (maleh)	-0.0177	-0.0272	0.0095
Age of Head (ageh)	-0.0132	0.0167	0.0299
Education (educh)	0.2483	-0.0692	0.3175
Farming Status (farm)	-0.0593	-0.0354	0.0239
Year 2	-0.6093	0.0187	0.6280
Year 3	-0.6063	-0.0149	0.5914
Year 4	-0.6280	0.0000	0.6280
Year 5	-0.5903	0.0447	0.6350
Year 6	0.9789	0.0610	0.9179
Year 7	1.0844	-0.0608	1.1452

Notes: The table reports standardized mean differences for covariates before and after Propensity Score Matching (PSM). Absolute change reflects the reduction in imbalance after matching.

Figure 18 shows the propensity score distribution for the treated and control groups, demonstrating the effectiveness of matching in aligning the two groups.

Figure 18. PSM Model: Propensity Score Distribution

H. Inverse Probability Weighting (IPW) Analysis

Table 14 compares standardized mean differences for covariates before and after Inverse Probability Weighting (IPW). The absolute changes indicate improvements in balance across treated and control groups after weighting.

Table 14—Balance Table: Standardized Mean Differences Before and After IPW

Variable	**Pre-IPW**	**Post-IPW**	**Absolute Change**
Male Adults (madult)	0.0600	0.2098	-0.1498
Female Adults (fadult)	0.0852	-0.4279	0.5131
Children (kids)	0.0540	-0.3547	0.4087
Male Household Head (maleh)	-0.0177	-0.1916	0.1739
Age of Head (ageh)	-0.0132	0.1429	-0.1561
Education (educh)	0.2483	0.1360	0.1123
Farming Status (farm)	-0.0593	0.1541	-0.2134
Year 2	-0.6093	-0.0841	0.5252
Year 3	-0.6063	0.0907	0.6970
Year 4	-0.6280	-0.1120	0.5160
Year 5	-0.5903	0.1618	0.7521
Year 6	0.9789	0.0532	0.9257
Year 7	1.0844	0.0574	1.0270

Notes: The table reports standardized mean differences for covariates before and after Inverse Probability Weighting (IPW). Absolute change reflects the reduction or increase in imbalance after weighting.

Figure 19 illustrates the propensity score distribution before and after weighting, showing the alignment achieved through Inverse Probability Weighting (IPW).

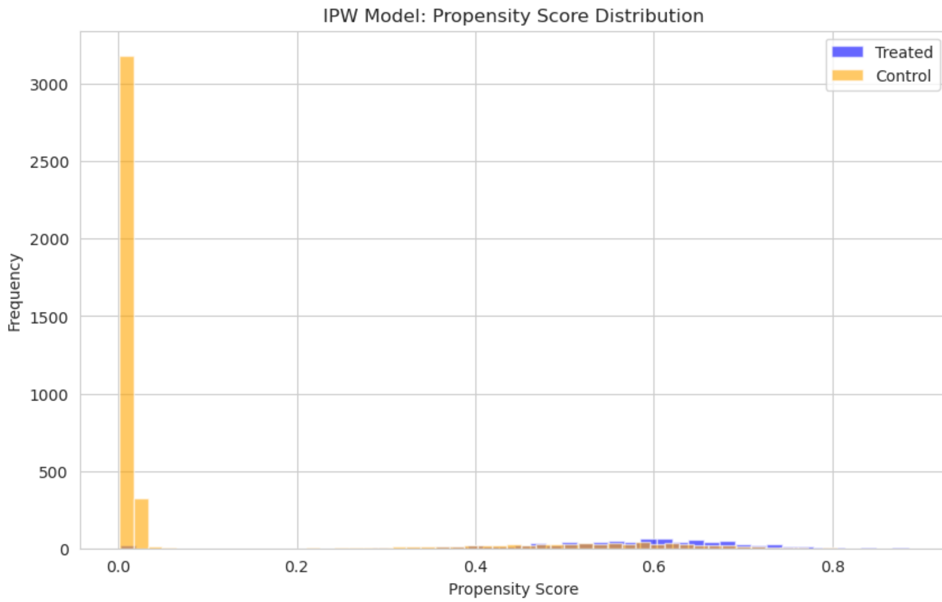

FIGURE 19. IPW MODEL: PROPENSITY SCORE DISTRIBUTION

I. Causal Forest Analysis

Figure 20 shows the distribution of treatment effects across subgroups when using binary treatment in the Causal Forest model.

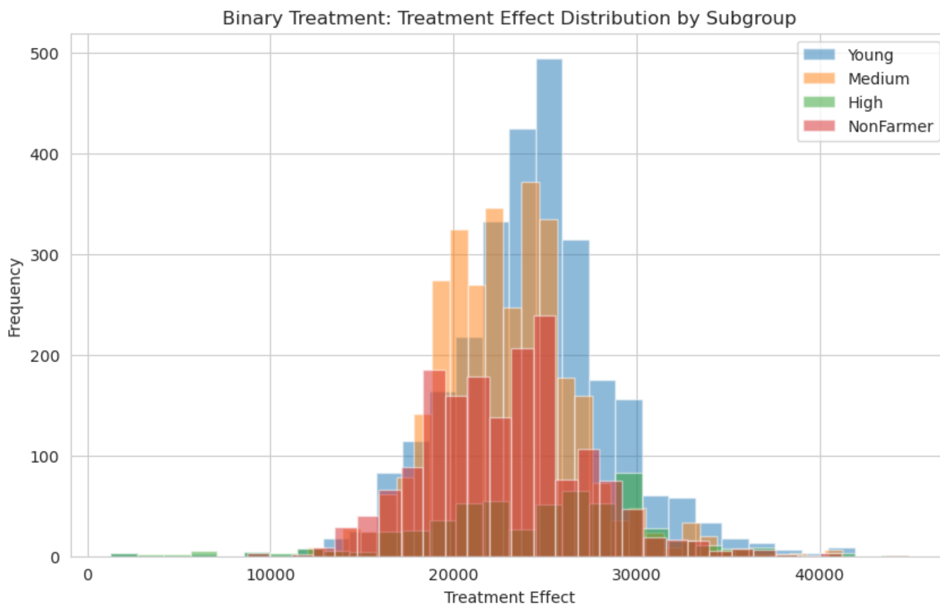

Figure 20. Binary Treatment: Treatment Effect Distribution by Subgroup

Figure 21 presents the distribution of treatment effects across subgroups using unpredicted treatment in the Causal Forest model.

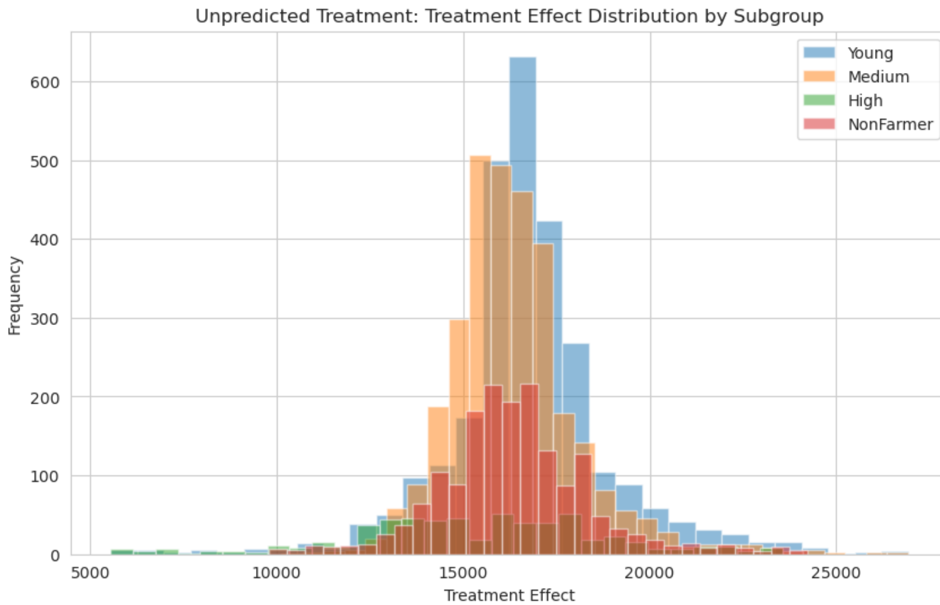

FIGURE 21. UNPREDICTED TREATMENT: TREATMENT EFFECT DISTRIBUTION BY SUBGROUP

Figure 22 displays the distribution of treatment effects across subgroups using predicted treatment in the Causal Forest model.

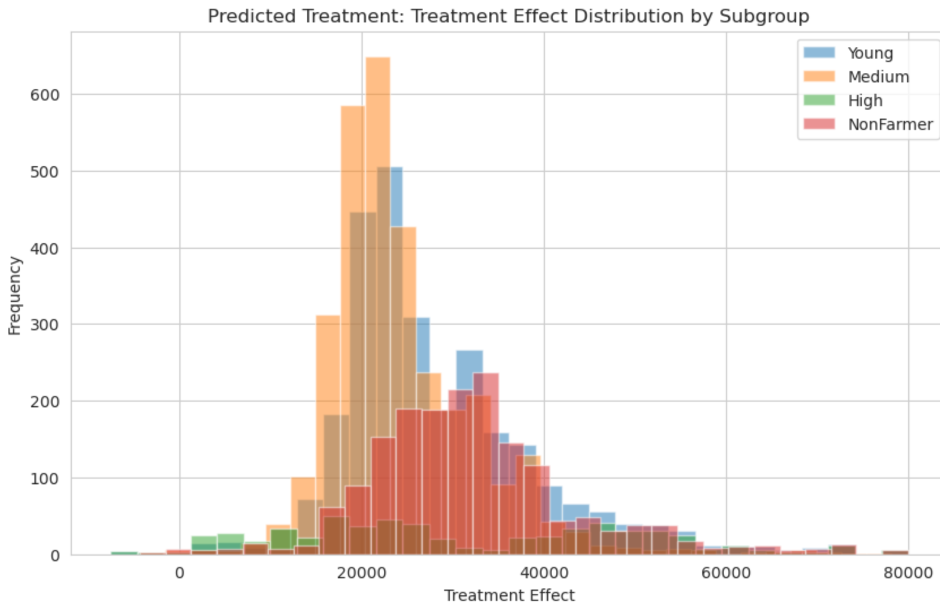

FIGURE 22. PREDICTED TREATMENT: TREATMENT EFFECT DISTRIBUTION BY SUBGROUP

These figures rank the importance of variables influencing treatment effects across binary, unpredicted, and predicted treatments in the Causal Forest model.

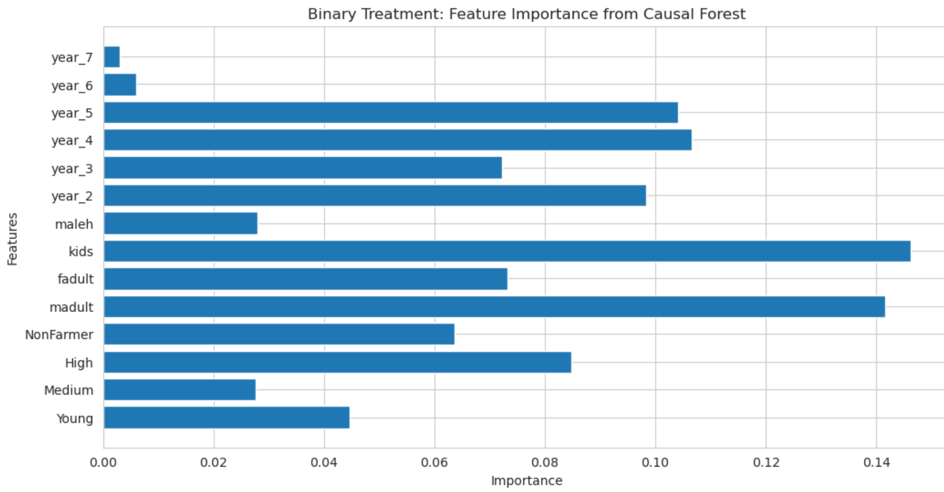

Figure 23. Binary Treatment: Feature Importance from Causal Forest

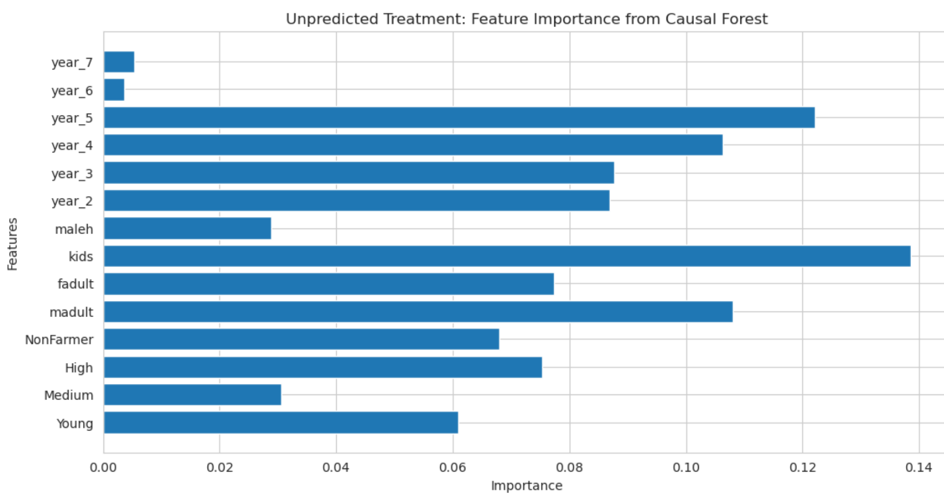

Figure 24. Unpredicted Treatment: Feature Importance from Causal Forest

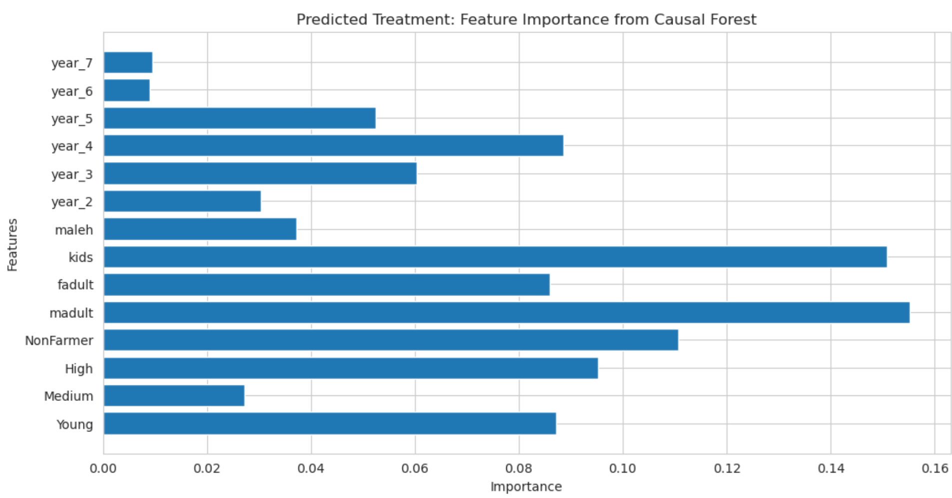

Figure 25. Predicted Treatment: Feature Importance from Causal Forest