Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi

Ezekiel Tano, Abraham Dabié, Mohamed Nana

April 18, 2025

1 Introduction

This project replicates and extends the study by Brune et al. (2016), which examines the effects of savings accounts on agricultural and financial outcomes in Malawi. The authors implemented a randomized controlled trial (RCT) offering farmers access to either a regular savings account or a commitment savings account with withdrawal restrictions. Their goal was to test whether disciplined savings mechanisms improve agricultural investment and household welfare. This study is policy-relevant, as many rural households struggle to save due to competing needs and a lack of financial tools.We focus on heterogeneity in one key outcome: agricultural input spending, a direct measure of farm investment. After replicating the original results using OLS, we apply the Generic Machine Learning (GenericML) framework by Chernozhukov et al. (2023) to flexibly explore how treatment effects vary with baseline characteristics, helping identify subgroups that benefit most.

Ok

2 Methodology

To analyze treatment heterogeneity in agricultural input spending, we employ the Generic Machine Learning (GenericML) framework developed by Chernozhukov et al. (2023). This approach provides robust estimates and valid inference on key features of treatment heterogeneity in randomized experiments, particularly when there is limited theoretical guidance but rich covariate data. GenericML is suitable for our analysis because it makes minimal assumptions about the form of treatment effect heterogeneity, enabling the data itself to identify heterogeneity patterns. It handles high-dimensional covariates without overfitting and facilitates valid inference on clear interpretable measures: the Best Linear Predictor (BLP) of the Conditional Average Treatment Effect (CATE), Group Average Treatment Effects by Sorted Groups (GATES), and Classification Analysis (CLAN), which contrasts characteristics of most and least affected units.

The method operates in two stages. First, machine learning (ML) algorithms are applied to an auxiliary sample from random sample splitting to estimate a proxy predictor for the CATE. In our application, we specifically use LASSO, Random Forest, Support Vector Machine, Elastic Net, Neural Networks, and Boosted Trees due to their strong performance with structured, high-dimensional data. Second, the estimated ML proxy is post-processed using an independent main sample to generate interpretable summaries and avoid overfitting. The procedure involves repeated random sample splitting and quantile aggregation, typically by taking medians, to ensure robust inference and replicability. Our main goal is to determine if the treatment effects on agricultural input spending differ according to baseline characteristics. By focusing on features of the CATE, we can effectively identify subgroups benefiting differently from the intervention and explore the associated characteristics. Thus, GenericML provides a rigorous yet flexible approach aligned with our empirical objectives.

3 Empirical Analysis and Results

We begin by replicating the two primary estimations from the original study, applying them to agricultural input spending using the following specifications:

$$Y_{ij} = \delta + \alpha \text{Savings}_j + \beta' X_{ij} + \varepsilon_{ij} \tag{1}$$

where Y_{ij} represents agricultural input spending for farmer i in club j, and Savings_j is an indicator for club-level assignment to either savings treatment (ordinary or ordinary plus commitment savings accounts). The coefficient α measures the effect of offering direct deposits into individual savings accounts. The vector X_{ij} includes stratification cell dummies and 17 baseline household characteristics. Standard errors are clustered at the club level, reflecting the randomization unit. The second specification differentiates between the two savings treatments:

$$Y_{ij} = \delta + \gamma_1 \text{Ordinary}_j + \gamma_2 \text{Commitment}_j + \beta' X_{ij} + \varepsilon_{ij} \tag{2}$$

why show this?

where Ordinary_j indicates club-level assignment to ordinary savings, and Commitment_j denotes assignment to ordinary plus commitment savings. γ_1 captures the effect of ordinary savings relative to control, and γ_2 captures the effect of commitment savings relative to control. The difference $\gamma_2 - \gamma_1$ tests the marginal impact of commitment savings relative to ordinary savings, with p-values reported for the hypothesis $\gamma_1 = \gamma_2$.

Results are detailed in Table 1. Panel A demonstrates a positive and statistically significant (at 10%) treatment effect on agricultural input spending during the 2009–10 planting season. Panel B separately identifies effects for ordinary and commitment treatments, both showing positive signs, with the commitment treatment statistically significant at the 5% level and the ordinary treatment not significant. Although the commitment treatment has a larger coefficient magnitude, we cannot reject at conventional levels the hypothesis that both treatment effects are equal. We next explore heterogeneity in the commitment treatment effect based on household size and gender. For household size, we test competing hypotheses: larger households may invest less in agricultural inputs due to competing expenditures, or invest more by leveraging

Nice, but why pre-specify?

increased savings and available labor. For gender, we hypothesize that men may invest more in agriculture due to social roles emphasizing farm output, while women might prioritize other uses for savings, potentially viewing agriculture as a male domain. To adjudicate between these mechanisms, we apply the Generic Machine Learning (GenericML) framework. Using 250 random splits for robustness, we estimate Group Average Treatment Effects (GATES) across predicted heterogeneity quintiles to understand which factors drive differential impacts. We apply a suite of machine learning algorithms—Lasso, Random Forest, Support Vector Machine, Elastic Net, Neural Network, and Boosting—to predict Conditional Average Treatment Effects (CATE). Their predictive performance is summarized in Table 2. Random Forest emerges as the best linear predictor (BLP) for CATE, while Support Vector Machine (SVM) provides the best GATES predictor. Accordingly, we rely on these two methods for the remainder of the analysis.

Table 3 presents the BLP estimates of the CATE regressed on the machine learning proxies. We report the coefficients β_1 (the average treatment effect, ATE) and β_2 (the heterogeneity loading, HET). The ATE estimate indicate that the commitment treatment increases the total value of inputs by 10470 Malawian Local currency. Reassuringly, the BLP ATE closely aligns with the overall ATE of the commitment treatment (10297). The estimated HET coefficient suggests a moderate degree of treatment effect heterogeneity, as indicated by the statistically significant estimate. Moreover, the estimate (0.65) is more close to 1 than 0, suggesting that the ML proxies are good predictors of the CATE.

We further explore this by estimating GATES by quintiles of the ML-predicted CATE. Figure 1 plots the GATES estimates, including the difference in ATE between the most and least affected groups, along with joint confidence bands. Specifically, the effect is slightly negative in the first quintile, approximately zero in the second and third quintiles, and becomes well positive in the fourth and fifth quintiles. Consequently, the overall difference between the first and last quintile is positive, though none of these differences are statistically significant. While these results do not reach statistical significance—likely due in part to reduced statistical power caused by sample splitting—they nonetheless suggest a subtle underlying heterogeneity in the treatment's impact on agricultural input spending, further confirming the slightly strong heterogeneity in the treatment's impact on agricultural input spending.

Given these findings, we next investigate whether this moderate observed heterogeneity correlates with observable pre-treatment characteristics. Using the CLAN method, we explore which covariates are associated with variation in treatment effects as detected through the BLP and GATES. Figure 2 presents the CLAN analysis for gender. We observe that the least affected group (G1) has the highest mean value of the gender variable (coded as 1 for female and 0 for male). This indicates that female farmers are the least affected, meaning they spend relatively less on agricultural inputs following the treatment, while male farmers, conversely, appear most affected and thus spend relatively more. However, the differences between these groups are not statistically significant, likely due to reduced statistical power resulting from sample splitting. Additionally, Figure 3 examines heterogeneity by household size. The analysis shows that the most affected group (G5) corresponds to the largest households, suggesting that households with more members allocate more resources to agricultural inputs than smaller households. Interestingly, the difference between the least affected (G1) and most affected groups (G5) is statistically significant. These findings help identify the relevant mechanisms driving heterogeneity in treatment effects.

4 Conclusion

This study replicates and extends findings from Brune et al. (2016) on the positive effects of commitment savings accounts in Malawi. Applying GenericML, we detect moderate heterogeneity in treatment impacts, with household size emerging as a significant driver. While gender-based differences are suggestive, they are not statistically significant. These results underscore the value of commitment savings in promoting farm investment and highlight the importance of tailoring financial inclusion programs based on household characteristics. GenericML proves to be a powerful tool for uncovering nuanced patterns in treatment response, enriching our understanding of program effectiveness.

5 Tables and Graphs

Table (1) Effects on Agricultural Input Expenditure

	Total Value of Inputs	
	(1)	(2)
Panel A:		
Treatment (Any)	8022.649*	
	(4130.825)	
Panel B:		
Treatment (Commitment)		10297.406**
		(4563.369)
Treatment (Ordinary)		5945.866
		(4503.707)
p-value for test of equality		0.245
Control mean	60371.795	60371.795
Observations	2835	2835
R-squared	0.458	0.459

TABLE 6

IMPACT OF TREATMENTS ON AGRICULTURAL OUTCOMES IN 2009–10 SEASON AND HOUSEH

	Land under Cultivation (Acres) (1)	Total Value of Inputs (MK) (2)	Proceeds from Crop Sales (MK) (3)
Panel A:			
Any treatment	.30**	8,023*	19,595**
	(.15)	(4,131)	(8,996)
Panel B:			
Commitment treatment	.33**	10,297**	26,427***
	(.16)	(4,563)	(9,979)
Ordinary treatment	.27*	5,946	13,358
	(.16)	(4,504)	(9,518)
P-value of *F*-test: coefficients on commitment and ordinary treatments are equal[a]	.614	.246	.086
Mean dependent variable in control group	4.28	60,372	91,747

Notes: Standard errors in parentheses are clustered at the club level. All specifications include stratification dummies and baseline controls: gender, marital status, age, education, household size, assets, livestock, land area, crop sales, input expenditure, formal account access, cash savings, hypertension, hypotension, network transfers, and missing indicators.

* p<0.10, ** p<0.05, *** p<0.01

Table (2) Comparison of Causal ML Methods: Total Value of Inputs

	Lasso	Random Forest	SVM	Elastic Net	Neural Network	Boosting
Best BLP (Λ)	386.80	492.69	342.25	174.12	53.49	388.12
Best GATES ($\bar{\Lambda}$)	362.15	390.73	427.40	286.84	292.20	378.22

Notes: Medians over 250 splits for all the machine learning methods.

2. What does this show?

Table (3) BLP of Total Value of Inputs

Parameter	Estimate	CI lower	CI upper	p value
ATE (β_1)	10470.43	-4378.86	25099.39	0.166
HET (β_2)	0.65	0.20	1.14	0.005

Notes: Medians over 250 splits using Elastic Net.

Figure (1) GATES of Total Value Inputs

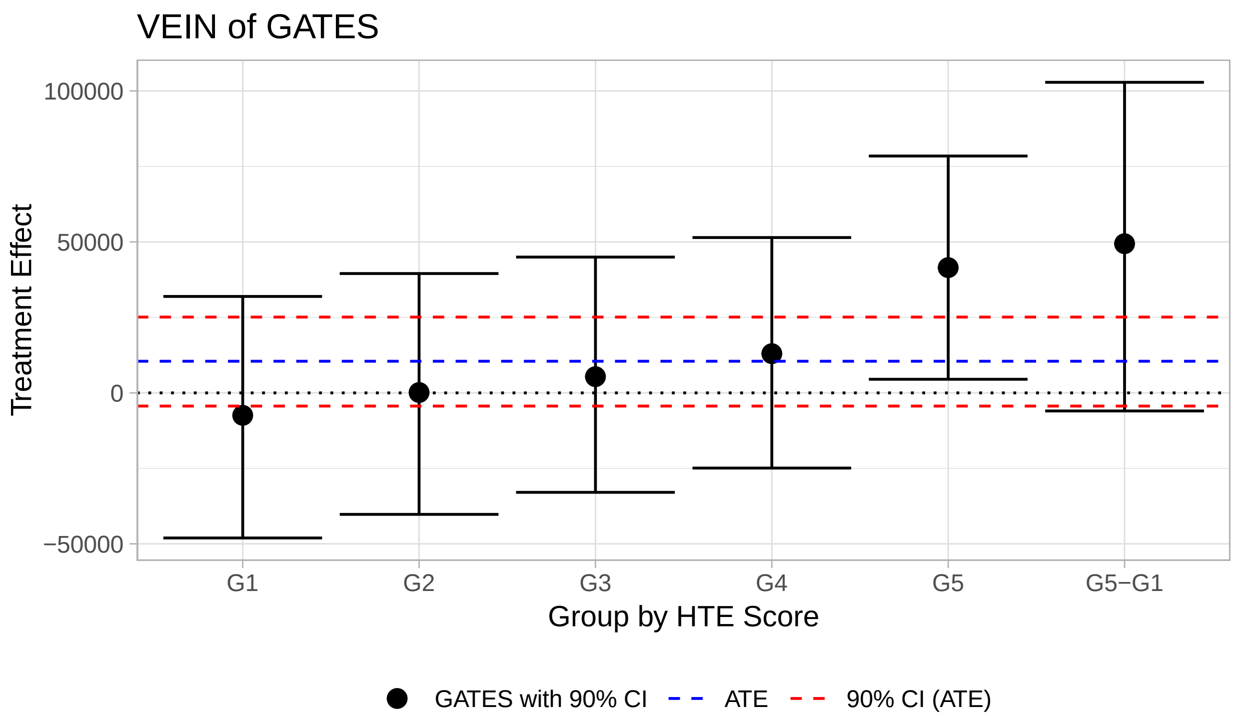


Figure (2) GATES based on gender

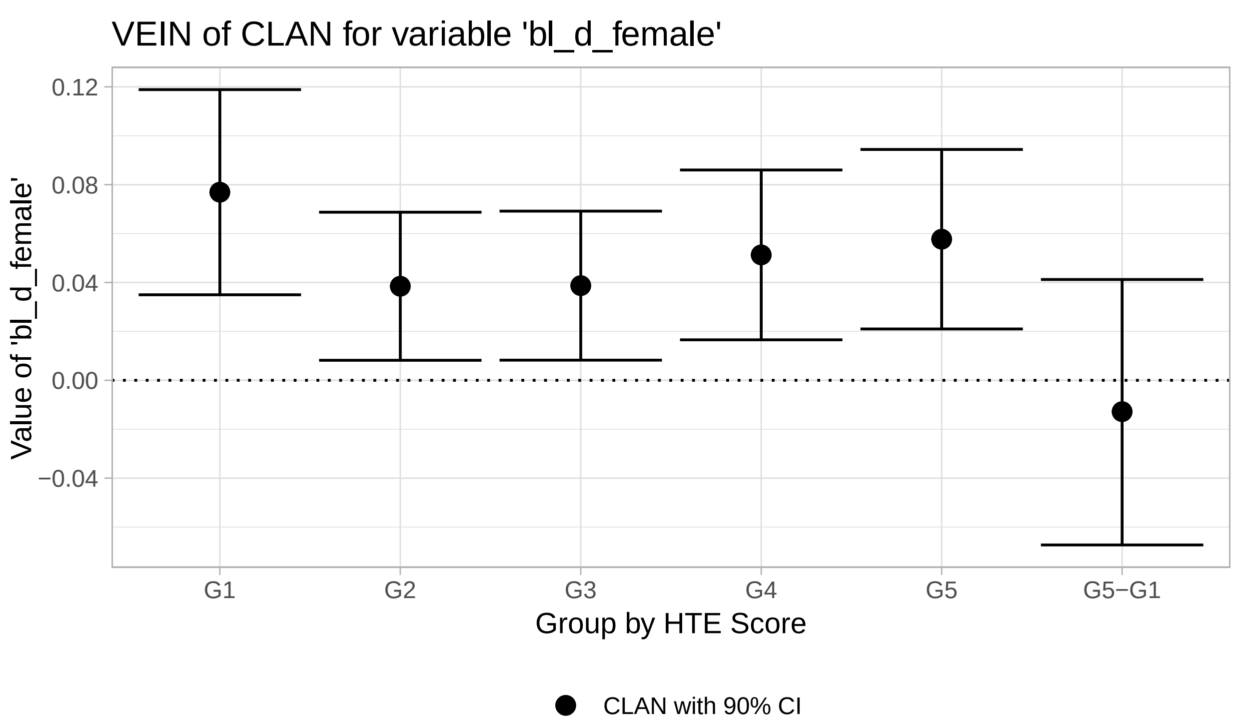
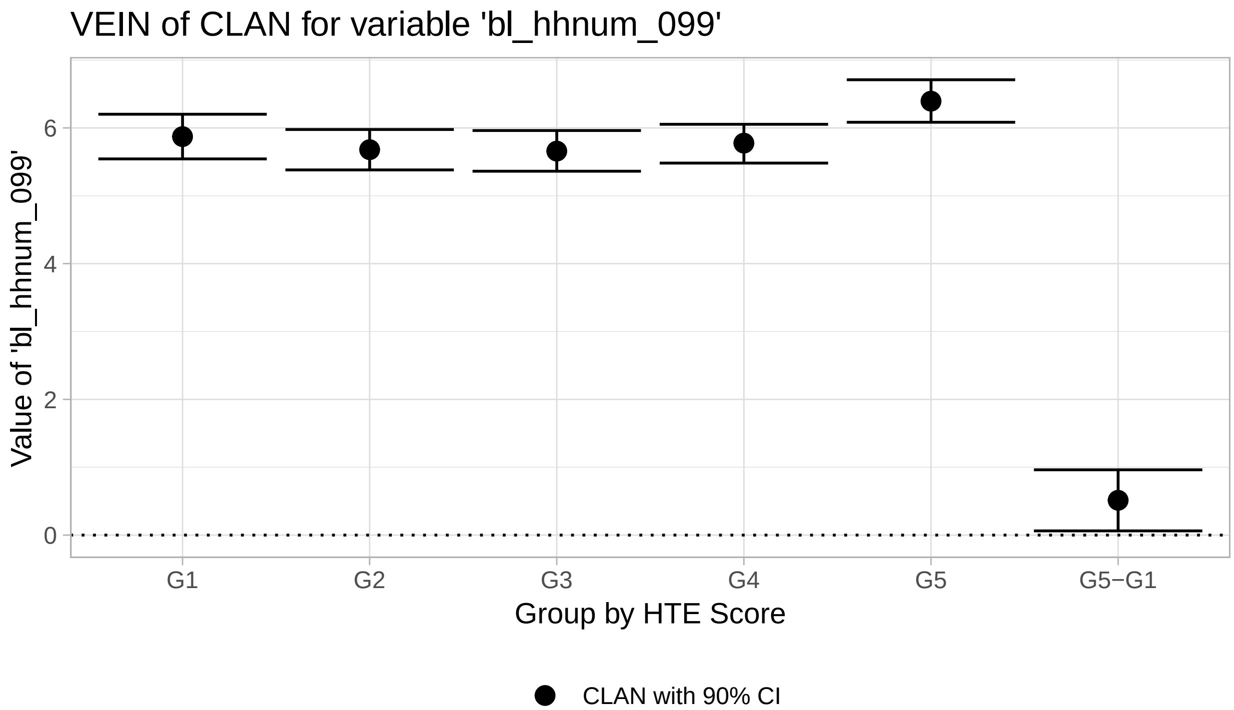


Figure (3) GATES based on household size



References

Brune, L., Giné, X., Goldberg, J., and Yang, D. (2016). Facilitating savings for agriculture: Field experimental evidence from malawi. *Economic Development and Cultural Change*, 64(2):187–220.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2023). Fisher-schultz lecture: Generic machine learning inference on heterogenous treatment effects in randomized experiments, with an application to immunization in india.