# Introduction

In this paper, we extend "Liquidity Constraints, Informal Institutions, and the Adoption of Weather Insurance" by Belissa et al. (2019) to account for heterogeneous treatment effects via a Multi-Arm-Causal-Forest model.

The original study finds that drought insurance uptake among smallholder farmers in Ethiopia could be higher, as smallholders suffer from seasonal liquidity constraints associated with harvest periods. The paper builds upon previous work in the literature, which finds that downside production risks impede farm modernisation (Emerick et al., 2016). By introducing a novel insurance product which delays repayment (IOUs) until after the harvest, the authors find that insurance uptake can be considerably increased.

The authors conducted a Randomised Control Trial (RCT) in 2015 using a pre-existing index-based drought-insurance product offered by Oromia Insurance Company (OIC) in the Ethiopia Rift Valley area. The product is typically sold during the rainy season before the harvest, and payouts are determined based on local rainfall levels.

To improve the effectiveness and uptake of the insurance product, the authors made three modifications to the original product. Firstly, they allowed deferred IOU payment post-harvest to alleviate smallholder liquidity constraints. Secondly, they varied the market channel and advertised the product through local (Iddir) leaders to increase trust and information. Finally, they imposed harsher contract terms to reduce the likelihood of default. In total, six different index-based insurance contracts were offered. The availability of these modified insurance products was randomised at the Iddir level to test their effectiveness.

Our extension examines whether increases in insurance uptake are consistent with the author’s theory that financial constraints limit uptake. If true, one would expect increased drought insurance uptake among poorer smallholders. The study's findings have implications for policymakers and stakeholders working to improve the financial resilience of smallholder farmers in developing countries and increase smallholder investment in new technologies.

The design of the original papers’ novel insurance product is motivated by the findings of two previous studies, which it builds upon.

<Contribution was not clear> The first study by Casaburi and Willis (2018)

The second study, by Dercon et al. (2014), investigates the effectiveness of marketing through informal local groups to boost insurance uptake. Additionally, they consider the impact of delaying insurance premiums on insurance uptake. Belissa et al. (2019) test both measures in several combinations to examine their effects on insurance updates and defaults.

The findings of Belissa et al. (2019) share similarities with Casaburi and Willis (2018) outside a contract farming context. The study finds that the demand-increasing effect of the IOUs may be more significant for players with low savings or income, in line with Dercon et al. (2014). The result suggests that liquidity constraints impede the uptake of drought-based insurance. By building on and extending previous research, the study provides a more nuanced understanding of the factors that affect insurance uptake among smallholder farmers in developing countries.

Our heterogeneity analysis focuses on two household characteristics, household income and savings. The authors undertake limited heterogeneity analysis in the paper, comparing below and above median income and participants with and without savings. We take a more nuanced approach using machine learning to calculate heterogeneity across the income and savings spectrum. Specifically, we implement a Multi Arm Causal Forest model developed by Wager and Athey (2017) to estimate these effects across several treatments. This extension aims to determine whether the author’s theory that liquidity issues hamper drought-based insurance uptake among agricultural smallholders in Ethiopia is consistent with the results.

(More details of results and conclusions needed.)

The paper is structured into six sections. The first section is the introduction, which overviews the study and its objectives. The second section is the literature review, which discusses the research on the topic and identifies the research gaps the study aims to address. The third section is the methodology, which describes the data sources, study design, and statistical techniques used to analyse the data. The fourth section is the results, which present the study findings. The fifth section, the extension, discusses the multi-arm causal forest approach used to estimate heterogeneous treatment effects. The sixth and final section is the conclusion, summarising the study's key findings and their implications for policy and future research.

# Methodology

This subsection discusses the methodology used to replicate Belissa et al. (2019). We provide a detailed explanation of the data and methods used in the original study, including any necessary assumptions or model specifications. We also explain the steps to ensure our replication accurately reproduces the original study's results.

## Data

As established in the introduction, the authors randomise the availability of several alternative index-based insurance contracts at the Iddir-level to test the effect of their design on insurance uptake and premium defaults.

The data gathered by the authors is cross-sectional in structure. All observations date to 2015? and contains 8,579 household observations. Six variations of index-based drought insurance were offered to smallholders. These are (1) Standard Index Insurance (control), (2) Standard Index Insurance via \textit{Iddir} promotions, (3) IOU insurance, (4) IOU insurance with Contract, (5) IOU insurance via \textit{Iddir} promotions; and (6) IOU insurance via \textit{Iddir} promotions with Contract. The availability of these treatments/products was randomised over 144 \textit{Iddirs}. In addition to randomising treatment the researchers collected data on socio-economic status[[1]](#footnote-1), farm production[[2]](#footnote-2) and savings to test whether randomisation was successful and to facilitate the estimation of causal effects.

## Replication Method

The authors employed a multi-arm Randomized Control Trial (RCT) methodology to estimate the causal effect of their novel insurance products on index-based insurance uptake among smallholders. In contrast to traditional two-arm RCTs, a multi-arm RCT compares more than two intervention groups to a control group, in this case, five. Each intervention group receives a different treatment, and the primary objective is to identify the most effective treatment(s), with the control group serving as the baseline comparison. Compared to two-arm RCTs, multi-arm RCTs offer several benefits, including increased study efficiency, potential identification of multiple effective treatments, and decreased risk of false-negative results. However, multi-arm RCTs can be more challenging due to increased trial design, analysis, and interpretation complexity. As a result, careful planning and execution are necessary to ensure the comparability of intervention arms and the study's ability to detect meaningful differences between the groups.

Estimating causal effects for cross-sectional data via ordinary least squares involves regressing the outcome variable on the treatment variable and any other relevant predictor variables. The coefficient for the treatment variable represents the average difference in the outcome variable between the treatment and control groups, controlling for any other relevant variables.

OLS (Ordinary Least Squares) is a widely used statistical technique often applied in an RCT (Randomized Control Trial) to estimate the effect of an intervention on an outcome variable. OLS offers several advantages, including its ability to control for other factors that might influence the outcome, making it easier to communicate study results to a broader audience, and providing estimates of treatment effects that can be used to calculate the cost-effectiveness of the intervention.

To verify whether randomisation was successful and that treatment is not confounded by other variables balancing tests are required to ensure no significant differences between groups before intervention. If groups differ significantly, it may bias the results making it difficult to mark causal inference.

However, OLS has some disadvantages that should be considered in an RCT. One major drawback is its assumption of a linear relationship between the outcome variable and explanatory variables. The results may be biased or inaccurate if this assumption is not met. OLS also does not account for potential confounding variables that were not included in the regression model, leading to biased estimates of treatment effects. Lastly, OLS requires a large sample size to provide reliable estimates of treatment effects, and if the sample size is too small, the results may be unreliable.

## Extension Method

We apply a multi-arm causal forest model methodology to extend the original report’s findings to estimate the heterogeneous treatment effects. A multi-arm causal forest is a machine-learning algorithm that can estimate the causal effect of multiple treatments on an outcome variable in a randomised control trial (RCT) setting. It extends the Random Forest algorithm commonly used for prediction tasks.

The multi-arm causal forest model builds on the traditional Random Forest model by incorporating information about the treatment assignments in the RCT. The key idea of a causal forest model is to build causal trees, which are decision trees that partition the data based on the causal effect of a specific feature. These trees can then be aggregated to create a causal forest, which provides an estimate of the treatment effect. A multi-arm causal forest model broadens the base causal forest algorithm to model for multiple treatment groups compared to the control group.

We implement the multi-arm casual forest model Nie and Wager (2021) suggested. This approach extends the causal forest algorithm, a machine learning algorithm that estimates treatment effects by partitioning the data into subgroups based on observed covariates. The quasi-oracle approach extends the causal forest algorithm by using kernel weights to estimate the treatment effect for multiple treatment arms simultaneously.

The equation is given by:

$\tilde{\tau}(x) = \operatorname{argmin}{\tau} \left{ \sum{i=1}^n \alpha\_i(x) \left(Y\_i - \bar{m}^{(-i)}(X\_i) - c(x) - \langle W\_i - \hat{e}^{(-i)}(X\_i), \tau(X\_i) \rangle \right)^2 \right}$

In this equation, $\tilde{\tau}(x)$ represents the estimated conditional average treatment effect for the target sample $x$. The subscript $i$ refers to the $i$th observation in the sample, and $n$ represents the total number of observations. The observed outcomes are denoted by $Y\_i$, and the observed covariates by $X\_i$. The estimated propensity score is denoted by $\hat{e}(X\_i)$, and is a vector-valued function that assigns a probability to each treatment arm, conditional on the observed covariates.

The weights $\alpha\_i(x)$ reflect the contribution of each observation to the estimation of the treatment effect for the target sample $x$. The intercept term $c(x)$ is a nuisance parameter that captures any systematic bias in the treatment effect estimates that is not captured by the observed covariates or the estimated propensity score.

The term $\bar{m}^{(-i)}(X\_i)$ represents the average outcome for the control group, which is estimated by excluding the $i$th observation from the sample. The notation $(-i)$ indicates that the $i$th observation is excluded from the calculation. Similarly, the term $\hat{e}^{(-i)}(X\_i)$ represents the estimated propensity score for the control group, which is also estimated by excluding the $i$th observation from the sample.

The term $\langle W\_i - \hat{e}^{(-i)}(X\_i), \tau(X\_i) \rangle$ represents the estimated treatment effect for each treatment arm, which is a linear combination of the estimated propensity score and the observed covariates. The notation $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors.

The optimisation problem seeks to find the value of $\tilde{\tau}(x)$ that minimises the sum of squared residuals, given the observed covariates, the estimated propensity score, and the weights $\alpha\_i(x)$. The resulting value of $\tilde{\tau}(x)$ represents the estimated treatment effect for each treatment arm, conditional on the observed covariates for the target sample $x$.

However, the quasi-oracle approach has some weaknesses. The approach requires a large sample size for sufficient statistical power, especially when estimating treatment effects for rare subgroups or interactions. Secondly, the approach is computationally intensive and can be time-consuming to implement, especially when simultaneously estimating treatment effects for multiple outcomes. Additionally, the approach may suffer from overfitting and may be sensitive to the choice of hyperparameters. Lastly, the quasi-oracle approach assumes that the treatment assignment mechanism is ignorable, meaning that no unobserved confounding variables affect both the treatment assignment and the outcome. If this assumption is violated, the treatment effect estimates may be biased.

<point out that we have a small number of observations. Overfitting is a potential issues as we have a large number of covariates. We assume no confounders, so balancing must be correct otherwise we have bias.>

# RESULTS

This chapter aims to replicate the findings of Belissa et al. (2019) in their study on drought insurance uptake among smallholder farmers in Ethiopia. The original study uses a randomised controlled trial (RCT) to test the effectiveness of modified index-based insurance products, which aim to alleviate liquidity constraints and increase trust among smallholders. Our replication examines the key findings of the original study, including the impact of deferred IOU payment, marketing through local leaders, and harsher contract terms on insurance uptake and defaults. By replicating the study, we aim to verify the robustness of the original findings and contribute to the reproducibility and transparency of research in development economics.

## Randomisation

To begin our replication, we first verified the consistency of the randomisation process in our dataset with that reported by Belissa et al. (2019). The original paper states that randomisation occurred at the Iddir level, with 144 iddirs across the treatment and control arms. However, upon examination of the data, we found that the total number of iddirs across treatment arms was 226, with several Iddirs appearing in multiple arms. Only the Standard Insurance via Iddir Promotions seems unaffected. We confirmed that our replication code did not alter the data and obtained a second dataset from another source, which yielded the same result. Thus, we conclude that the replication data was systematically distorted before uploading. We have contacted the original author concerning the paper.

<dummy table/counts by iddir and obs>

Note that using a distorted dataset will affect our ability to replicate further material from the original paper. In particular, this will affect our calculation of heterogeneous treatment effects in the extension due to the distortion of the randomisation arm, which relies on randomisation for unconfoundedness.

## Balancing

Next, we assess the balancing of the randomisation process. The original paper used this to verify the success of the randomisation process for causal inference. We match the author’s approach and regress a battery of socio-economic, production and savings variables on the treatment groups. We do not include the control group, the “Standard Index Insurance” dummy, which is therefore reflected in the constant.

Comparing the constants of our balancing test to Table 1 in the original paper, we attempt to trace the impact of the data distortion. Constants for Age, Sex, Education, Family Size, Monthly Income, Drought Severity and Previous Insurance are unaffected. The marriage dummy is significantly higher at 0.95 compared to 0.90 in the original paper. Point estimates for each independent variable adhere closely to those presented in the original paper. However, they do not match exactly in almost all cases. This would suggest that distortions in the dataset are relatively small.

Table 1

<Drought variable is too high. Coding error on our part. Need to fix. Must be if any Yes=1>

We test whether the treatments’ coefficients are equal via the Wald test. In general, the results are highly significant except for the education variable. High significant levels would normally lead us to conclude that coefficients are unequal. However, in the context of data

distortions noted in the dataset between randomisation and clusters, particularly as we are required to clusters robust errors by iddir, we cannot discount the possibility that our results are distorted.

Distortions of the constants in Table 2 remain small but are more frequent. Only the constant for Wheat (5.09) matches the original paper. All other constants align closely with the original paper but differ somewhat. The increased frequency of departures from the original paper may not represent any distortion pattern, but this data contains more variance, and therefore distortions are more visible.

Table 2

<savings variable is too high. May be a coding error on our part. Need to fix. Must be if any Yes=1>

Wald tests for joint significance between treatment arms in production variables are again extremely significant. These results are substantially more significant than in the original paper. The high significance level likely represents the earlier distortions noted in the dataset between randomisation and clusters, particularly as Iddir clusters robust errors.

As noted previously regarding Table 1. The results of our Wald test are highly significant. Usually, this would lead us to conclude that the coefficients are significantly different. However, we are reluctant to place faith in the results in light of our data issues.

## Uptake Rates

Figure 2 illustrates the insurance uptake rates across various treatment arms. The IOU product's delayed payment option shows a substantial increase in uptake compared to standard insurance, jumping from 8% to 24%. The combination of IOU and promotion through Iddir outperforms all other treatments, with uptake rates reaching approximately 43%. These results are consistent with those reported in the original paper, except for Group 6, which received the most comprehensive intervention package, including IOU, promotion by Iddir leaders, and a binding contract. In our study, the uptake rate of this group was slightly lower than in the original experiment, specifically 27% compared to 32%. However, upon further examination of the data, we discovered that around 5% of households in this group took up an IOU via Iddir without signing a contract, which explains the discrepancy between our findings and those of the authors. Consequently, we have adopted the authors' calculation method in our subsequent research.

<Fig. 2. Uptake rates across IOU treatments, 95% CI clustered at Iddir level.>

## Main results (Have not revised)

The regression analysis results (Table 3) are similar to those of the authors. Individual IOU contracts don't have a significant uptake advantage over standard insurance with a binding contract. IOU insurance with and without Iddir promotion both have a significant positive impact on uptake rates. Introducing a binding contract to IOU dampens uptake rates, suggesting that delayed payment adoption is driven by strategic default or farmer uncertainty about future premiums. Signing a binding contract requires trust, which Iddir could mitigate, increasing uptake rates even with a contract. However, index insurance via Iddir significantly impacts uptake in the parsimonious model, potentially due to data issues.

Table 3

Insurance uptake rates increase under IOU.

<moved from methodology to results. Would like to pepper into resulst>

To verify whether randomisation resulted in balanced groups they regress household observables and farming observables on treatment group dummies and a constant. The constant reflects the comparison group and the coefficients indicate whether other groups are significantly different from the comparison group. They test for differences between other groups by Wald tests. The results suggest the randomization worked well.

Then, they present insurance uptake across treatment arms and conduct regression analysis with and without additional controls including \textit{Kebele} fixed effects and all baseline socio-economic characteristics, all showing that uptake change induced by \textit{Iddir} promotions in isolation is statistically insignificant, as is IOU with binding contract. They also exclude the subsample from a certain \textit{Kebele}, Dalota Mati, which all the defaults in the dataset come from, to increase the statistical power of the analysis, and the result shows robustness.

They perform a heterogeneity analysis, to figure out if they can attribute the increase in uptake under IOU insurance to the relaxation of the liquidity constraint. To proxy for liquidity, they distinguish between households with above and below-median income, and between households with and without savings (self-reported). And for both proxies, the coefficients of the simple IOU product are higher for the liquidity-constrained. However, while the IOU coefficient of the (more) constrained subsample is consistently different from zero, and the coefficient for the complementary sample is not, the relevant coefficients are not statistically different from each other (according to a Wald test). This can also be seen by the insignificance of the coefficients using the interaction term instead of subsamples.

To accurately replicate the results reproduced in the original study, we follow the steps shown in the paper and try to explain the dicrepancies if there are any of them. The steps of our replication are described as follows.

We generate five dummies indicating whether each individual is in one of the five treatment groups and a dummy indicating the status of uptake. Then we generate the controls used in the article, including demographic variables: Age (in years), Sex (male=1; female=0), Marital status (married=1; not-married=0), Education (years of schooling), Family size, Total income in the last month (in Birr), Drought (a dummy taking value of 1 if the household experienced a drought in the last three years), and Insurance (a dummy taking the value of 1 if the household had purchased index insurance during the past three years); and farming variables: capturing quantities of crops produced in the last cropping season (maize, haricot, teff, sorghum, wheat, and barely), a measure of total land under cultivation, and a dummy taking the value 1 if the household had any formal savings.

To conduct balancing tests, we regress observable controls, including demographic variables and farming variables, on treatment group dummies and a constant, to see if the coefficients of the group dummies are statistically significant. The randomisation works well if we see the coefficients of the treatment dummies are not significant, which shows that covariates do not affect the treatment assignment and therefore there is no severe selection bias.

For regression analysis, we regress the uptake status on five group dummies, and then add controls and \textit{Kebele} fixed effects to the parsimonious model. Finally, we exclude the subsample from a the \textit{Kebele} Dalota Mati and run the same regression. We also conduct Wald tests to test if the coefficients of each two treatment dummies are statistically the same, verifying whether the treatment effects of different types of insurance designs differ.

For extension of the replication, we apply a non-parametric causal forest, which is developed by Wager and Athey (2017) and can achieve better matching with many covariates, for estimating heterogeneous treatment effects that extends Breiman’s widely used random forest algorithm. In the potential outcomes framework with unconfoundedness, causal forests are pointwise consistent for the true treatment effect, and have an asymptotically Gaussian and centred sampling distribution. The causal forests give us a better understanding of treatment effect heterogeneity, so we apply this method to the original paper, which reveals the treatment effects of a drought insurance, to analyse the heterogeneity of the effects on individuals with different characteristics.

# Conclusion

Replication component: Our study replicated Belissa et al.'s (2019) findings on the effects of delayed payment and Iddir promotion on insurance uptake. We found that delayed payment (IOU) and promotion via Iddir leaders increased uptake rates, while binding contracts had a negative effect. However, we were unable to fully reproduce the authors' results due to a data issue that led to discrepancies in the numbers of Iddirs in certain treatment groups.

Extension component: Additionally, we used causal forests to analyze heterogeneity in the treatment effects across households with different socio-economic and production characteristics. Our results showed that the effects of the insurance design varied depending on these characteristics. Specifically, we found that IOU had a greater effect on households with lower asset levels and lower livestock numbers, while Iddir promotion had a greater effect on households with higher asset levels and higher levels of education. Our findings highlight the importance of considering heterogeneity in treatment effects when designing weather insurance programs in Ethiopia.

1. Age, Sex, Marital Status, Education, Family Size, Monthly Income, Drought, Insurance History etc. [↑](#footnote-ref-1)
2. Maize, Haricot, Teff, Sorghum, Wheat, Barley, Farm Size etc. [↑](#footnote-ref-2)