Do climate resilient seeds hurt farmer profitability?

Analysis of the results of a randomized controlled trial (synthetic data)

Fanni Varhelyi

I. Introduction

Climate change is an immense danger to humanity with profound implications for all areas of our lives, including agriculture. As the planet warms, weather patterns will become more unpredictable, and extreme heat and drought will become more common phenomenons. In order to prepare for that, scientists developed a more resistant wheat seed that can withstand these effects better than traditional seeds. While the seed is better equipped to withstand climate change effects, this variant might also have an impact on farmer's earnings who use them. For these new seeds to be a sustainable, long-term solution, it is crucial to ensure farmers can still make a living wage from selling the crop from this new seed variant.

To test for this, an experimental project was conducted involving 283 towns in a province. Farmers were randomly assigned to treatment or control group (given seeds or not), and treatment intensity varied within the treatment groups as well. Since the farmers were randomly assigned to these groups, we can assume that this is a random controlled trial - and we will be conducting further tests to ascertain this by checking if control and treatment groups are balanced and comparable.

Using earnings data collected after this experiment was conducted, the goal of this analysis is to understand the effect the new seeds had on the income of the farmers. First, I will investigate the basic attributes of the study to ascertain if any elements of it seem out of order or not random, and adjust the data if needed based on my findings. After that, I will analyze the impact of the new seeds on farmer earnings, also looking at potential spillover effects.

II. Study set-up

As the farmers live in towns, we can understand these towns as clusters, and for precise analysis, I will use clustering on a town level when conducting all my analyses. I will also assume that earnings are in thousands of U.S. dollars. In this section, I will check three things. First, I will establish if the randomly assigned treatment and control groups are similar enough to be used as the basis of comparison. Then, I will check for attrition, and whether the attrition is random or not. Finally, I will conduct a permutation test to see if the results from the study are statistically unlikely to happen by random chance. If necessary, I will adjust the data using weights before moving on to the next section of analyzing the results.

To begin, I checked if treatment and seed usage had a relationship with earnings, and found that both of these had a statistically significant relationship with earnings. Based on the results in Table 1, it seems that on average, farmers in the control group earn around $10.12-10.22 thousand dollars, and being part of the treatment group increased this earning by $0.15 ths, and using the new seeds increased it by $0.25 thousand. These results indicate that some spillover happened, and some farmers not assigned to treatment also used the seeds.

```
treat_earn <- felm(lnearnings ~treatment | 0 | 0 | town_id, data = data)
seed_earn <-  felm(lnearnings ~uses_new_seeds | 0 | 0 | town_id, data = data)
```

Before investigating these spillovers further, I checked if the control and treatment groups were balanced. To do this, I analyzed whether there's a statistically significant relationship between any known attribute of farmers and the treatment - namely, I checked this for household size, age, and marital status. Table 2

Table 1: Preliminary regression results

	Dependent variable:	
	lnearnings	
	(1)	(2)
treatment	0.153***	
	(0.006)	
uses_new_seeds		0.251***
		(0.006)
Constant	10.224***	10.124***
	(0.004)	(0.004)
Observations	64,689	64,689
R^2	0.010	0.028
Adjusted R^2	0.010	0.028
Residual Std. Error (df = 64687)	0.720	0.713
Note:	*p<0.1; **p<0.05; ***p<0.01	

displays the results of this balance check, and shows that treatment and control groups are similar enough to call treatment assignment random, and continue with the analysis without any readjustments.

```
hhldcheck <- felm(hhld_size ~treatment | 0 | 0 | town_id, data = data)
agecheck <-  felm(age_head_hhld ~treatment | 0 | 0 | town_id, data = data)
maritalcheck <-  felm(married ~treatment | 0 | 0 | town_id, data = data)
multicheck <-  felm(treatment ~hhld_size + age_head_hhld + married | 0 | 0 |
                      town_id, data = data)
```

Next, I investigated attrition. About ~9.8% of farmers dropped out of this experiment, and this attrition appears to be correlated with treatment: control group participants were more likely to drop out, and this difference between the groups (around 2.2%) is statistically significant.

```
attrition_original <- felm(attrition ~treatment | 0 | 0 | town_id, data = data)
```

To address this imbalance in attrition, I trimmed the data to readjust the control and treatment group to ensure optimal results. To do this, I estimated the probability of dropping out, and kept only observations in treatment and control group that had similar probabilities of dropping out by trimming observations that had either high or low probabilities of attriting. While this ensured balance in the data, it also led to a large loss in the number observations (~18 ths observations remained out of ~71 ths) and thus a significant decrease in precision. As the difference between the attrition of the control and treatment groups was only ~2%, I decided to continue working with the original dataset and not trim it. The slight difference in attrition can be kept in mind that we might slightly overestimate the effect of treatment on income, but as the magnitude is small, this shouldn't affect the results by much. Table 3 displays the results of regressing attrition on treatment before and after trimming.

Finally, I conducted a permutation test on the original dataset to check how likely it is that these results are only due to random chance. Based on the results on my permutation test, which can be seen in Figure 1, it is highly unlikely that these results are due to random chance. Thus, I will continue with my analysis as I checked if treatment assignment was random, ensured that control and treatment groups are comparable and balanced, and checked if the results are plausibly non random. While I did not correct for a slight difference in attrition, the magnitude of the difference is small enough that it can be disregarded.

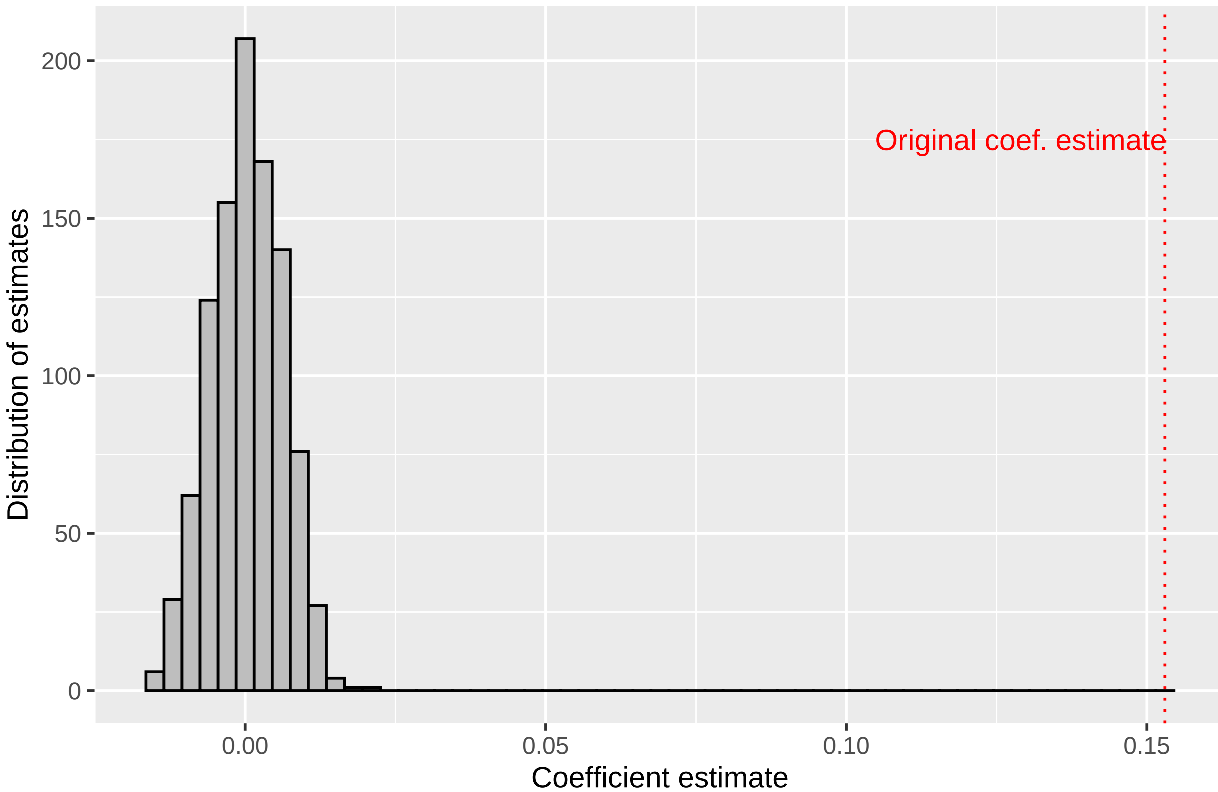
Table 2: Balance check

	Dependent variable:			
	hhld_size	age_head_hhld	married	treatment
	(1)	(2)	(3)	(4)
treatment	0.001	0.081	−0.001	
	(0.010)	(0.057)	(0.004)	
hhld_size				0.0002
				(0.001)
age_head_hhld				0.0004
				(0.0003)
married				−0.001
				(0.004)
Constant	3.994***	36.626***	0.570***	0.359***
	(0.006)	(0.034)	(0.002)	(0.020)
Observations	71,714	71,714	71,714	71,714
R^2	0.00000	0.00003	0.00000	0.00003
Adjusted R^2	−0.00001	0.00002	−0.00001	−0.00001
Residual Std. Error	1.303 (df = 71712)	7.055 (df = 71712)	0.495 (df = 71712)	0.484 (df = 71710)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3: Attrition before and after trimming

	Dependent variable:	
	attrition	
	(1)	(2)
treatment	−0.022***	−0.013
	(0.007)	(0.008)
Constant	0.106***	0.103***
	(0.005)	(0.006)
Observations	71,714	18,158
R^2	0.001	0.0005
Adjusted R^2	0.001	0.0004
Residual Std. Error	0.297 (df = 71712)	0.297 (df = 18156)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 1: Permutation test distribution



III. Analysis

In my analysis, I checked for both the effect of being assigned to treatment and actually receiving treatment (i.e., using the seeds), and what were the implications of either. First of all, I checked for the effect treatment assignment had on earnings, or the intent to treat effect (ITT). Treatment resulted in a \$0.153 ths or ~1.5% increase in earnings. The result is statistically significant, and while it's relatively small in magnitude, earnings are expressed as ln function of earnings. Table 4 displays the regression results. These results might slightly overestimate treatment impact as attrition was slightly higher in the control group.

```
itt_treat_earn <- felm(lnearnings ~treatment | 0 | 0 | town_id, data = data)
```

Table 4: ITT regression results

	Dependent variable:
	lnearnings
treatment	0.153***
	(0.006)
Constant	10.224***
	(0.004)
Observations	64,689
R^2	0.010
Adjusted R^2	0.010
Residual Std. Error	0.720 (df = 64687)
Note:	*p<0.1; **p<0.05; ***p<0.01

Next, I checked the treatment effect looking at everyone who used the treatment, i.e., the seeds, or the TOT effect. In theory, this could be quite different from treatment assignment due to noncompliance in both the treatment group and the control group. Table 5 displays the regression results: using the seeds resulted in a $0.256 thousand increase in earnings. The difference between ITT and TOT indicates that some spillover happened, and farmers in either the control group or in the treatment groups where they were not assigned treatment used the seeds as well.

Table 5: TOT regression results

	Dependent variable:
	lnearnings
uses_new_seeds	0.256^{***}
	(0.010)
Constant	10.122^{***}
	(0.007)
Observations	64,689
R^2	0.028
Adjusted R^2	0.028
Residual Std. Error	0.713 (df = 64687)
Note:	*p<0.1; **p<0.05; ***p<0.01

The farmers in the experiment were assigned to four groups: Group 0 was onlt control, Group 1 was 25% treatment, group 2 was 50% treatment, and group 3 was 75% treatment. To assess spillover effects on different participants who were not intended to receive the treatment, I will take these groups into consideration.

Table 6 displays the split between control and treatment group and the usage of the seeds. It seems like a significant number of farmers in the control group used the seeds. Similarly, some farmers in treatment groups where they were not assigned treatment also used the seeds. Noncompliance was not present in the treatment groups for farmers who were assigned treatment: all of them used the seeds.

Table 6: Spillover in groups

	town_group	treatment	uses_new_seeds	n
1	0	0	0	11,738
2	0	0	1	2,927
3	1	0	0	8,196
4	1	0	1	5,278
5	1	1	1	4,154
6	2	0	0	3,361
7	2	0	1	4,839
8	2	1	1	8,910
9	3	0	0	756
10	3	0	1	3,105
11	3	1	1	11,425

To further investigate this, I created indicator variables for being assigned treatment (T_) or control (SO_), and for being part of a specific treatment group (1,2, or 3). Table 7 displays the results of regressing these indicators on both earnings and seed usage. Seed usage shows that even in the pure control group, about 20% of farmers used the seeds. In groups 1-3, as treatment intensity increased, spillover increased as well: in group 1, ~49% of control assigned farmers used the seeds, in group 2, 59% of them did, and in group 3, ~80% of them did. All farmers assigned to treatment used the seeds. Similarly, the effect in income increased as the % of treated increased in the groups. All results are significant.

```
spillover <- felm(lnearnings ~T_1 + T_2 + T_3 + SO_1 + SO_2 + SO_3 | 0 | 0 |
                    town_id, data = data)
spill2 <- felm(uses_new_seeds ~T_1 + T_2 + T_3 + SO_1 + SO_2 + SO_3 | 0 | 0 |
                 town_id, data = data)
```

Table 7: Spillover regression results

	Dependent variable:	
	lnearnings	uses_new_seeds
	(1)	(2)
T_1	0.195***	0.800***
	(0.013)	(0.003)
T_2	0.210***	0.800***
	(0.009)	(0.003)
T_3	0.199***	0.800***
	(0.010)	(0.003)
SO_1	0.039***	0.192***
	(0.008)	(0.005)
SO_2	0.105***	0.391***
	(0.009)	(0.006)
SO_3	0.153***	0.605***
	(0.014)	(0.007)
Constant	10.175***	0.200***
	(0.005)	(0.003)
Observations	64,689	64,689
R^2	0.014	0.461
Adjusted R^2	0.013	0.461
Residual Std. Error (df = 64682)	0.719	0.355
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 2 and Figure 3 illustrates this effect on income across the different treatment groups. Overall, increasing treatment intensity directly did not affect earnings by much, the different group participants earned similarly. On the other hand, it had an impact indirectly: as treatment intensity increased, earnings on average also increased for those farmers who were in these groups but not assigned treatment. It is clear from Figure 4 why: as treatment intensity increased, not treated farmers were using the new types of seeds more intensely as well.

Figure 2: Direct treatment effects:
Increasing the percentage treated had minor or no effect on income

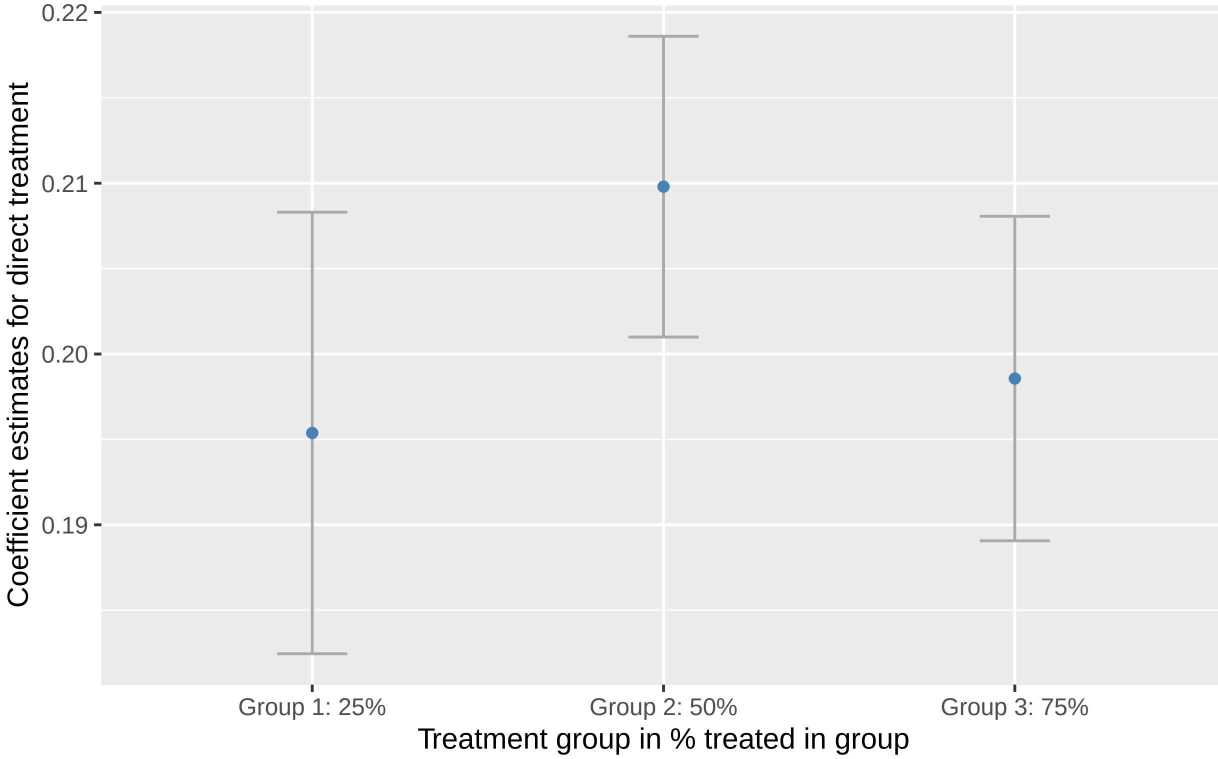
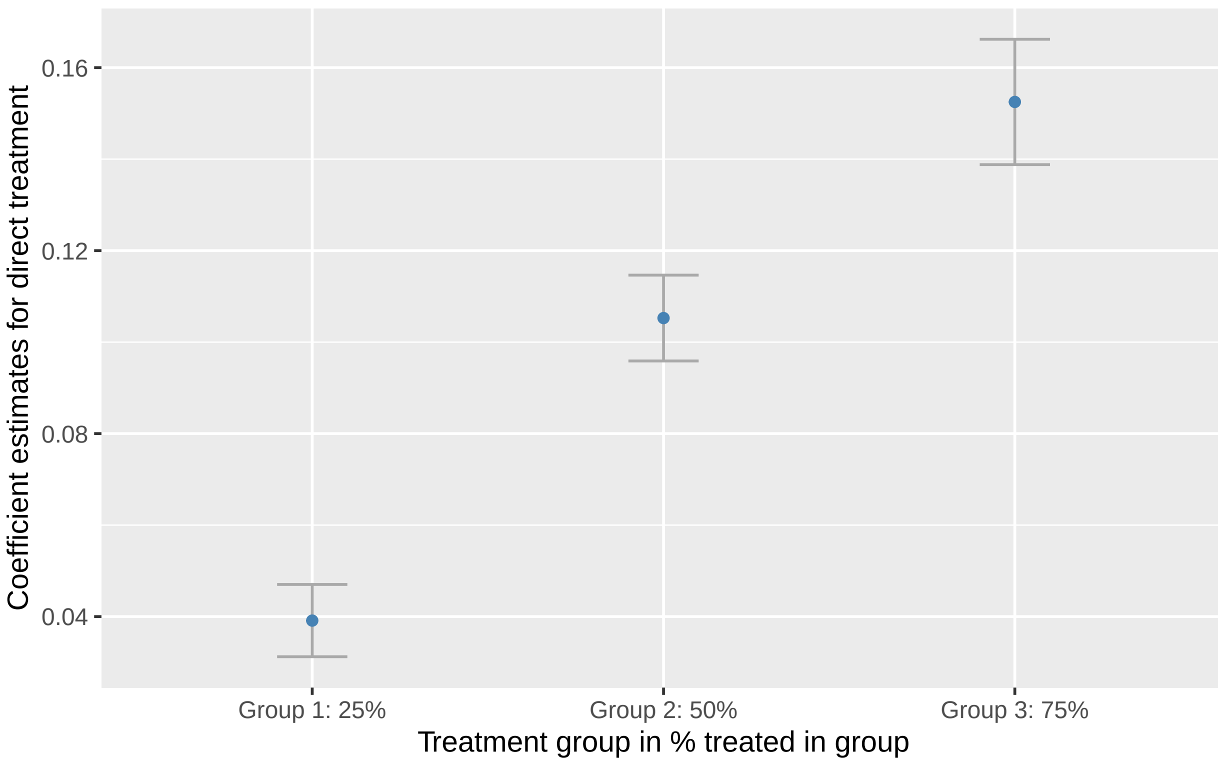
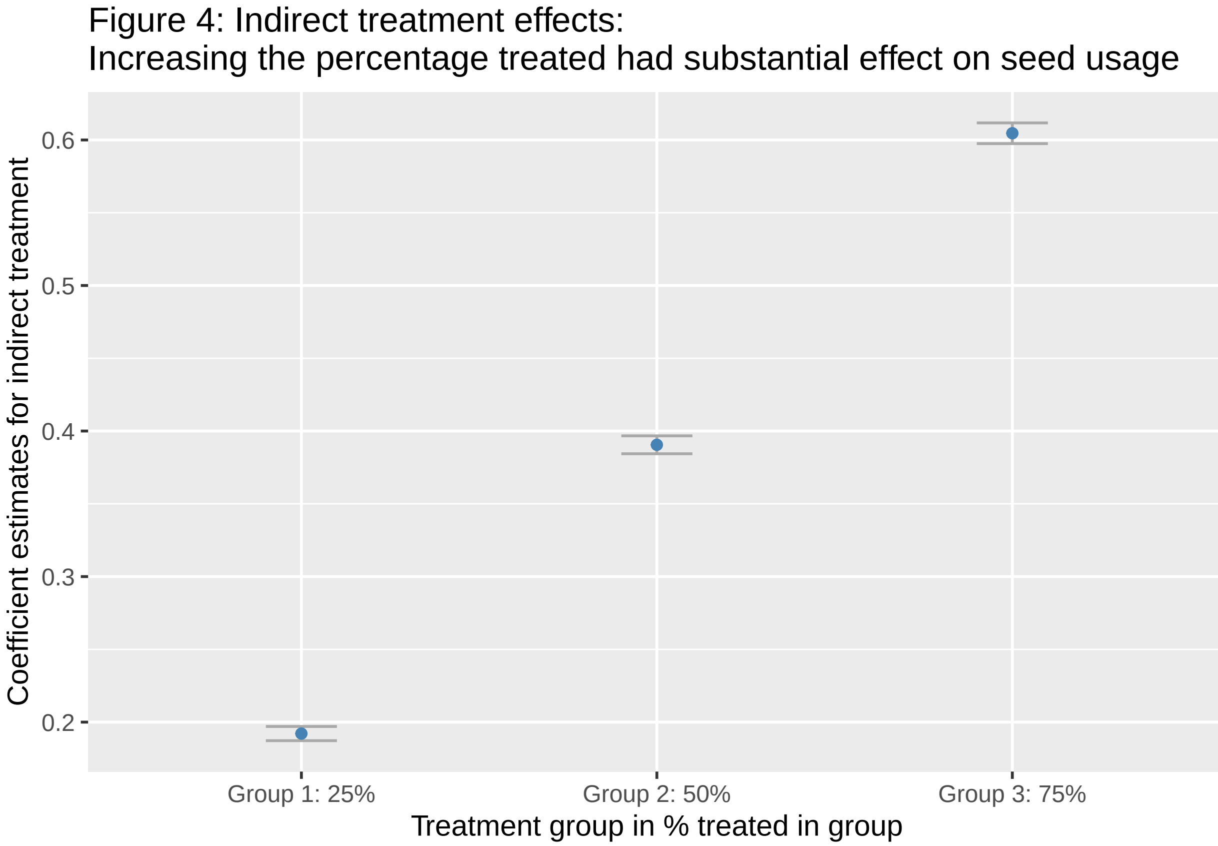


Figure 3: Indirect treatment effects:
Increasing the percentage treated had substantial effect on income





IV. Conclusion

To conclude, my analysis showed a relatively small, but positive impact on earnings when using the seeds: about $0.25 thousand increase on earnings. There was also significant spillover even for the pure control group, and this spillover increased with treatment intensity. About 20-80% of farmers assigned to control received treatment as treatment intensity increased from 0 to 75%, and there was no noncompliance in treatment. Thus, my finding is that the new climate resistant seed is a viable alternative to the current variant, as it increased earnings, and likeliness of usage was relatively high even in control groups.