Replication Study

Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda

Blattman, Fiala, Martinez, 2014

POLS 4768

Jacqueline Woo

5/15/2019

Introduction

Blattman et al's (2014) paper evaluates the Youth Opportunities Program (YOP), a government program in Uganda aimed at helping young adults in the conflict-afflicted north become self-employed artisans. Screened and eligible groups of youth from ages 16 to 35 were randomly given one-time unsupervised grants worth $385 per member on average—about the average annual income—to invest in skills training, tools and materials (Blattman et al., 2014). On average, the program increased business assets by 57%, work hours by 17%, and earnings by 38% relative to the control group (Blattman et al., 2014). The researchers find evidence of heterogenous treatment effects in line with their conceptual model of credit constraints—in particular, that a cash windfall should have a larger effect on those that are most constrained, such as women, those not engaged in a skilled trade, and those with low working capital, human capital and patience. However, the relationships are not statistically significant using the standard linear regression model hypothesis test (Blattman et al., 2014).

Methodology

The researchers collected data on 535 groups containing nearly 12,000 members (Blattman et al., 2014). Five people per group were surveyed three times over four years: first at baseline in 2008, then at two endlines, one in 2010 (24 – 30 months after grant disbursement) and another in 2012 (44 – 47 months after grant disbursement) (Blattman et al., 2014). Both endline surveys resulted in around a 40% attrition rate, and as a result researchers used a two-phase double-sampling strategy that ultimately resulted in a combined response rate of 80 – 85% for both control and treatment groups (Blattman et al., 2014). Attrition could be related to potential outcomes, as it varied with treatment status: in 2012, the treatment group was 7 percentage points more likely to be found than control, and it is conceivable that unsuccessful youth could have moved away in search of better opportunities (Blattman et al., 2014). Thus, the two-phase double-sampling strategy seems to be the best method of addressing attrition in this case, and the researchers weight observations by the inverse of the probability of selection into the second phase of surveying (Blattman et al., 2014).

Replication, Extension, and Findings

I replicate and extend three aspects of the Blattman et al (2014) study. First, I will replicate the full sample regression models from Table IX: Treatment Heterogeneity by Initial Working Capital, Human Capital, Patience, and Engagement in a Skilled Trade that examine treatment heterogeneity on business assets and monthly cash earnings in 2010 and 2012 (pooled). As noted previously, while the study finds evidence of heterogenous treatment that is in line with the researchers' model of credit constraints, this evidence is not statistically significant using the standard linear regression hypothesis test (Blattman et al., 2014). I additionally confirm this lack of statistical significance by conducting hypothesis tests that the interaction coefficients are zero using randomization inference. In particular, 10,000 replications of the F-statistic calculation under the sharp null hypothesis yields p-values of 0.17 for business assets and 0.78 for monthly cash earnings.

Second, I replicate the ITT with controls, TOT with controls, difference-in-differences ITT with controls and ITT without controls models in Table VII: Sensitivity Analysis of Intent-To-Treat Estimates to Alternate Models and Missing Data Scenarios. Additionally, I estimate the same models without the double-sampling and IPW procedure and compare ATEs and standard errors to determine the impact of the double-sampling method in this case. I find that excluding double-sampling results in both over- and under-estimates depending on the outcome variable, further indication that attrition may not be systematic.

Third, I examine the weighting methodology under double sampling and am unable to replicate their weighting. I attempt to identify if probability of selection is influenced by any particular combination of covariates and cannot identify a model that fits, an indication that the methodology was indeed random. I also determine that the summary statistics reported for phase 2 of endline 1 are incorrect. Additionally, I attempt to replicate weighting using stratification by district and proportion unfound, but my replication does not match the data. Lastly, I model another method of selecting second phase sampling candidates, taking into account initial distribution of observations at baseline and stratifying by district and proportion unfound.

Heterogenous Treatment Effects

I conducted a replication of Table IX: Treatment Heterogeneity by Initial Working Capital, Human Capital, Patience, and Engagement in a Skilled Trade that examine treatment heterogeneity on business assets and monthly cash earnings in 2010 and 2012 (pooled) both in R and STATA, since the original data and replication code is available in STATA but the ability to conducted randomization inference is more easily available in R. I was able to exactly replicate the regression results in STATA, but the standard errors from my replication in R were slightly different from the original despite the use of IPW and clustered standard errors by participant ID. Table I (end of paper) illustrates the difference. Estimates and standard errors differ by 0 – 4% for the variables of interest.

To conduct a hypothesis test of the F-statistic for both the business assets and monthly cash earnings models, I use the non-robust F-statistic calculation under the assumption of homoskedasticity:

$$\hat{F} = \frac{\dfrac{SSR_{Null} - SSR_{Alternative}}{Parameters_{Alternative} - Parameters_{Null}}}{\dfrac{SSR_{Alternative}}{N - Parameters_{Alternative}}}$$

I also use non-robust, non-clustered linear regressions of the models. For business assets, the F-statistic is 1.81 with a p-value of 0.17 under 10,000 hypothetical replications. For cash earnings, the F-statistic is 1.34 with a p-value of 0.78 under 10,000 hypothetical replications. We are unable to reject the null of no treatment-by-covariate interactions.

Sensitivity Analysis without Double-Sampling

The authors conduct sensitivity analysis to test the robustness of their initial results, particularly in light of baseline imbalance and systematic attrition, and calculate the ITT with controls, TOT with controls, ITT without controls and difference-in-differences ITT with controls for level of business assets, hours engaged in skilled trade work, monthly cash earnings and durable assets in 2010 and 2012. To additionally explore the impact of systematic attrition, I replicate these models with and without the second round of sampling (Table II—end of paper). I was able to successfully replicate and extend these calculations for all models except that of the difference-in-differences ITT with controls model for level of business assets as the baseline measurements of business assets were unavailable in the replication dataset.

I find that while estimates for the first phase of sampling are in the same direction as that of double-sampling estimates and have roughly the same level of statistical significance, the magnitude of these estimates varies in different directions depending on the outcome variable, making it difficult to infer any underlying pattern of systematic attrition. A first-phase sampling strategy tends to overestimate the treatment effect on business assets in 2010 and 2012, underestimate the treatment effect on monthly cash earnings and durable assets in 2010 and 2012, and overestimate the effect on skilled trade work hours in 2010 and underestimate the effect on the same variable in 2012. If some systematic attrition did occur, one would expect these estimates to vary in the same direction. For example, the authors suggest that since the treatment group is more likely to be found, if unfound controls are particularly successful, treatment effects could be overstated. However, we only find that to be the case for business assets. Given the different directions of estimation, I am unable to conclude if systematic attrition did occur.

Inverse Probability Weights

The authors weight all their calculations using the inverse of the probability of selection into the second round of surveying and state that their double-sampling methodology was random but stratified by district and proportion unfound in the group (Blattman et al., 2014).

I first attempted to identify if the probability of selection into the second phase is influenced by any particular combination of covariates and cannot identify a model that fits, an indication that sampling was indeed random. I use five models:

TABLE III

Relationship between Weights and Covariates

	Model	Adjusted R^2
(1)	Regression of the inverse of weights on being found at baseline and covariates that significantly predicted attrition	0.02
(2)	Robust version of (1)	0.02
(3)	Regression of the inverse of weights on being found at baseline and all baseline covariates, including district	0.03
(4)	(1) with only double-sampled data	0.02
(5)	(3) with only double-sampled data	0.13

Any model with a high adjusted R^2 would provide an indication that selection is systematic, and I do not find this to be the case (Table III). This is puzzling, as one would expect a relationship between probability of selection and district since selection was stratified by district.

I additionally find that the summary statistics for phase 2 of endline 1 (in Table 1: Survey Response Rates) are incorrect. Table I states that 63.4% of the total sought (2,677) were found in Phase 1 for the first endline survey, implying around 1,697 found. The dataset has two variables denoting the first endline results—ind_found_e1 and ind_unfound_p1_e1—and two variables denoting the second endline results—ind_found_e2 and ind_unfound_p1_e2. Ind_found_e1 denotes whether an observation was found at the end of both phases of the first endline, as it yields 2,005 found observations, which aligns with the final number of observations found for the first endline in Table I. Similarly, ind_found_e2 denotes whether an observation was found at the end of both phases of the second endline, as it yields 1,868 observations, which aligns with the final number of observations for the second endline found in Table I of the paper. That leaves one to conclude that to identify the observations found during phase 1 of endline 1, ind_unfound_p1_e1 should equal 0, and similarly for ind_unfound_p1_e2 for observations found during phase 1 of endline 2. However, filtering on ind_found_p1_e1==0 and e1==1 yielded 1,902 observations, or 71.0%, not 63.4%. In fact, the authors note that they found 75% of the subset of the unfound that they attempted to find, but this number coincides more closely with the final proportion of observations found, 74.9%. Using the selection and found probabilities in Table I for phase 2 of endline 1 would result in 840 observations being found for phase 2, but the data seems to imply that 913 observations were found (Table IV). On the other hand, filtering on ind_found_p1_e2==0 and e2==1 yielded 1,632 observations, which does align with the 61% found after the end of phase 1 for endline 2 in the paper. The data for endline 2 phase 2 also aligns with the summary statistics in Table I. It seems clear then that the ind_unfound_p1_e1 variable is meant to indicate whether the observation was unfound after phase 1 of the survey, but the data reported in the paper is incorrect.

TABLE IV

Replication of Survey Response Rates

Survey	Total Sought	Found, Phase 1	Unfound, Phase 1	Selected, Phase 2	Found, Phase 2	Found, Phase 2 (From Paper)
Endline 1	2677	1092	1585	840	913	627
Endline 2	2677	1632	1045	403	236	236

Focusing on endline 1, I attempted to use the “unfound after Phase 1” variable to calculate the proportion unfound by district in an attempt to replicate the sampling weighting as stated in the paper (Table V). Proportion unfound can have two definitions: it can be the percentage of unfound in a district as a proportion of total unfound (column 3) and it can also be the percentage unfound within a district as a proportion of the total initial observations in that district (column 5). I calculate both of these proportions by district and their inverse an in attempt to see if they match the weights in the dataset—they do not, as the weights in the dataset range from 1 to 4 and the inverse weights calculated span beyond that (columns 4 and 6).

TABLE V

Replication of Sampling Weighting

District	(1) Total Obs	(2) # Unfound Phase 1 Endline 1	(3) Proportion Unfound	(4) Inverse	(5) Proportion Unfound in District	(6) Inverse
ADJUMANI	95	22	3%	35.23	23%	4.32
APAC	585	222	29%	3.49	38%	2.64
ARUA	237	93	12%	8.33	39%	2.55
KABERAMAIDO	75	19	2%	40.79	25%	3.95
KOTIDO	160	31	4%	25.00	19%	5.16
KUMI	220	63	8%	12.30	29%	3.49
LIRA	370	112	14%	6.92	30%	3.30
MOROTO	130	12	2%	64.58	9%	10.83
MOYO	110	20	3%	38.75	18%	5.50
NAKAPIRIPIRIT	95	12	2%	64.58	13%	7.92
NEBBI	90	26	3%	29.81	29%	3.46
PALLISA	240	86	11%	9.01	36%	2.79
SOROTI	135	34	4%	22.79	25%	3.97
YUMBE	135	23	3%	33.70	17%	5.87
Grand Total	2677	775	100%	N/A	25%	N/A

Lastly, I calculated another method of stratifying second-round sampling by district and proportion unfound for endline 1 that takes into account initial distribution of observations by district (Table VI). According to the paper, 53% of the unfound after phase 1 were selected into phase 2. This means that 411 observations were selected to be found in phase 2 for a final goal of 2,313 observations. Since phase 2 sampling was stratified by district and proportion unfound, researchers would presumably wish to obtain the same proportion of observations as there had been at baseline, assuming symmetrical attrition rates by district for phase 2 of surveying. I calculate this final number assuming 100% of the selected unfound were found (column 3), then calculate the total to be selected for Phase 2 by district (column 4). Finally, I calculate the selection probability as the total to be selected for Phase 2 as a proportion of total unfound at the end of Phase 1 (column 5) and the corresponding inverse weight (column 6). These weights are closer to the weights found in the dataset. However, the goal of obtaining 2,313 observations in proportion to the initial distribution of observations translated to a need to reduce the number of found observations in two districts, Moroto and Nakapiripirit, as these districts had higher than proportional response rates in the first round. That translated to negative selection probabilities and weightings; if this were indeed the methodology used to select second phase observations, I would recommend increasing the target observations such that negative probabilities do not occur.

TABLE VI

Selection Method Accounting for Initial Distribution of Observations

District	(1) Total Obs	(2) # Unfound Phase 1 Endline 1	(3) Total After Phase 2	(4) Total to be found during Phase 2	(5) Selection Probability	(6) Weight
ADJUMANI	95	22	82	9	0.41	2.44
APAC	585	222	505	142	0.64	1.56
ARUA	237	93	205	61	0.66	1.52
KABERAMAIDO	75	19	65	9	0.47	2.11
KOTIDO	160	31	138	9	0.29	3.44
KUMI	220	63	190	33	0.52	1.91
LIRA	370	112	320	62	0.55	1.81
MOROTO	130	12	112	-6	-0.50	-2.00
MOYO	110	20	95	5	0.25	4.00
NAKAPIRIPIRIT	95	12	82	-1	-0.08	-12.00
NEBBI	90	26	78	14	0.54	1.86
PALLISA	240	86	207	53	0.62	1.62
SOROTI	135	34	117	16	0.47	2.13
YUMBE	135	23	117	5	0.22	4.60
Grand Total	2677	775	2313	411	0.53	1.89

Conclusion

This attempt to replicate and extend the Blattman et al. (2014) confirms their main findings with regards to attrition and heterogenous treatment effects. However, the authors could clarify their weighting and double-sampling methodology further, particularly since their findings could potentially be impacted by their weighting methodology. Overall, this exercise has been an informative way of integrating concepts into practice.

TABLE I
Heterogenous Treatment Effects: Comparison of STATA and R Replications

	Dependent variable (2010 and 2012 endline data pooled)					
	Business assets (000s 2008 UGX), Full sample			Monthly Cash Earnings (000s 2008 UGX), Full sample		
	STATA Replication (Exact Match)	R Replication	Difference	STATA Replication (Exact Match)	R Replication	Difference
Assigned to treatment	419.8	419.8	0%	15.9	15.9	0%
	[68.9]***	[69]***	0%	[4.2]***	[4.3]***	1%
2012 endline	144.3	144.3	0%	19.2	19.2	0%
	[70.5]**	[70.7]**	0%	[4.7]***	[4.7]***	1%
Assigned x 2012 endline	-238.3	-238.3	0%	1.5	1.5	0%
	[90.8]***	[90.7]***	0%	[5.8]	[5.8]	0%
Female	-298.1	-298.1	0%	-11.4	-11.4	0%
	[65.6]***	[65.8]***	0%	[4.3]***	[4.3]***	1%
Female x 2012 endline	-84.0	-84.0	0%	-13.3	-13.3	0%
	[80.4]	[80.3]	0%	[5.6]**	[5.6]**	0%
Engaged in skilled trade	201.8	201.8	0%	15.5	15.5	0%
	[229.1]	[235.1]	3%	[15.3]	[15.8]	3%
Assigned x skilled trade	-90.3	-90.3	0%	-10.0	-10.0	0%
	[220.7]	[224.6]	2%	[14.9]	[15.3]	2%
Working capital index (z-score)	127.8	127.8	0%	15.7	15.7	0%
	[50.6]**	[52.2]**	3%	[5.2]***	[5.4]***	4%
Assigned x working capital index	-83.7	-83.7	0%	-4.3	-4.3	0%
	[66.8]	[68.7]	3%	[6.3]	[6.5]	4%
Human capital index (z-score)	45.9	45.9	0%	9.0	9.0	0%
	[32.3]	[32.6]	1%	[2.4]	[2.4]	1%
Assigned x human capital index	25.1	25.1	0%	3.0	3.0	0%
	[48.6]	[48.9]	1%	[3.6]	[3.6]	1%
Patience index (z-score)	23.8	23.8	0%	5.7	5.7	0%
	[31.5]	[31.8]	1%	[2.1]	[2.2]	2%
Assigned x patience index	-27.6	-27.6	0%	2.0	2.0	0%
	[50.2]	[50.7]	1%	[3.3]	[3.4]	1%
Observations	3,873	3,873		3,873	3,873	

TABLE II

Sensitivity Analysis: Comparison of Double-Sampling and First Phase Estimates

	ITT with controls			TOT with controls			ITT without controls			Diff-in-diff ITT with controls		
Dependent variable	Double-Sampling	First Phase	% change	Double-Sampling	First Phase	% change	Double-Sampling	First Phase	% change	Double-Sampling	First Phase	% change
Business assets												
2010	377.023	479.713	-21%	442.138	567.175	-22%	407.250	501.251	-19%	407.250	N/A	N/A
Std. err	[78.217]***	[71.412]***		[89.274]***	[80.563]***		[82.511]***	[70.985]***		[82.511]***	N/A	N/A
2012	224.986	255.544	-12%	275.556	305.308	-10%	250.532	299.964	-16%	250.532	N/A	N/A
Std. err	[62.601]***	[67.130]***		[72.083]***	[75.661]***		[68.404]***	[66.754]***		[68.404]***	N/A	N/A
Skilled trade work hrs												
2010	4.703	4.897	-4%	5.394	5.694	-5%	4.551	4.676	-3%	4.763	5.451	-13%
Std. err	[0.612]***	[0.523]***		[0.675]***	[0.588]***		[0.621]***	[0.521]***		[0.664]***	[0.606]***	
2012	3.776	3.479	9%	4.380	4.022	9%	3.666	3.455	6%	4.092	3.753	9%
Std. err	[0.548]***	[0.490]***		[0.618]***	[0.550]***		[0.569]***	[0.489]***		[0.604]***	[0.569]***	
Monthly cash earnings												
2010	14.605	12.645	16%	17.087	15.086	13%	15.044	11.529	30%	9.112	13.100	-30%
Std. err	[4.073]***	[3.938]***		[4.636]***	[4.462]***		[4.324]***	[3.948]***		[5.472]*	[4.797]***	
2012	18.186	12.054	51%	22.045	14.867	48%	19.049	12.695	50%	16.453	10.347	59%
Std. err	[4.898]***	[4.725]**		[5.560]***	[5.324]***		[5.475]***	[4.827]***		[5.911]***	[5.628]***	
Durable assets (z-score)												
2010	0.101	0.100	1%	0.119	0.1160	3%	0.106	0.098	8%	0.053	0.050	6%
Std. err	[0.047]**	[0.0414]**		[0.053]**	[0.047]**		[0.054]**	[0.043]**		[0.043]	[0.053]	
2012	0.181	0.141	28%	0.203	0.158	28%	0.190	0.155	23%	0.104	0.110	-5%
Std. err	[0.055]***	[0.046]***		[0.062]***	[0.052]***		[0.061]***	[0.049]***		[0.049]**	[0.057]*	

References

Blattman, C., Fiala, N., & Martinez, S. (2014). Generating Skilled Self-Employment in Developing Countries : Experimental Evidence From Uganda. *Quarterly Journal of Economics*, 697–752. https://doi.org/10.1093/qje/qjt057

Appendix 1: Replication Data

Replication code in STATA: STATA Replication.do

Replication code in R: YOP_Jac

Dataset: yop_analysis.dta

Excel file: yop_doublesampling.xlsx