PH 252D                                                                                    Spring 2018

# Entrepreneurship in Uganda

Shruti Bathia, David Contreras Loya, William Krinsman                          05/06/2018

## 1   Background/Motivation/Abstract/Scientific  Question/Causal  Question/etc.

Uganda, like most low-income countries, has a large share of youth who are either unemployed or underemployed. Living in economies where employment opportunities are scarce and self-employment often the only option, youth need the right combination of human, financial, and social capital to improve their welfare.

Many governments recognize that their economy would benefit from better-trained entrepreneurs. Uganda and 22 other African countries have mainstreamed entrepreneurship training in high school through support from the ILO. Other countries are developing short training programs in entrepreneurship, while others are expanding university level entrepreneurship training. However, the curricula in these programs are based primarily on hard skills and ignore the potential contribution of soft skills.

This proposed research seeks to address a gap in development literature by focussing on what specific business training techniques work. There have been a number of experimental business training evaluation studies including Karlan and Valdivia (2011) and Valdivia (2011) in Peru, Drexler et al. (2011) in the Dominican Republic, Berge et al. (2011) in Tanzania. These studies confirm that business training leads to improvements in knowledge of good business practices. However, these studies examine the impact of training on existing entrepreneurs. In Sri Lanka de Mel, McKenzie, and Woodruff (2012) examine the effects of an ILO business training program on business success of both existing female entrepreneurs and the general population of women. The proposed project wishes to expand on this research.

(Scientific question is the same as the causal question(s), so we should probably state somewhere that they are the same, rather than stating them twice) More specifically, we want to investigate if entrepreneurial training affects labor market outcomes by a) inducing individuals to start businesses sooner after graduation of secondary school and b) increasing revenues and profits for those businesses. We measure business creation and financial performance in a sample of 3,893 Ugandans between 22-30 years old who were eligible to receive a three-week, post-secondary intensive training camp on entrepreneurship skills. We will study economic outcomes of individuals under a non-parametric framework to estimate their treatment-specific counterfactual outcomes.

Does entrepreneurial training increase the likelihood of starting a business after graduation from high school?
Does entrepreneurial training increase business monthly revenue?
Does entrepreneurial training increase business monthly profit?

We focus on the differential effect of hard skills and soft skills on three economic outcomes: business creation, revenues and profitability. For the purpose of this class, we will pool both treatment arms into a single group.

Would people have gotten more financial performance had they been assigned to the business course?

## 2   Data and Experiment Description

We interviewed 4,400 individuals at baseline, and we reached 3,891 during the follow-up study 4 years after. Our baseline covariates $W_0$include basic sociodemographic characteristics (age, gender, region of residence, household socio-economic level) and several measures of cognitive development (e.g. Raven score), personality constructs (Big 5), and time and risk preferences. Distance from home village to training site was also recorded for all individuals. We also observe treatment status A in $\{0,1\}$. At follow-up, we obtained information about every economic activity undertaken in the period after graduation from high school and time of the follow-up interview (April 2016). Our outcomes Y are 1) a binary indicator for whether the individual started a business, 2) log of monthly revenue (USD) and 3) log of monthly profit (USD).

The target population was youth in Uganda who graduated high school and are in the job market. The sample consisted of students enrolled in the last year of high school in 4 regions of Uganda in 2013. Approximately, 40% of the sample attended schools in the West, 20% in Jinja, 20% in Mbale, and 20% in the North. The study was designed to be nationally representative with both students and teachers assigned to one of three groups (hard skills, soft skills, control) randomly. Students were recruited from 200 secondary schools, which represents a third of the total number of full time secondary schools in Uganda. Students interested in the program were asked to fill out an application form and a baseline survey. In total 8,080 students applied to the program and 7,431 complied with eligibility requirements (completeness of key baseline characteristics and no concurrent entrepreneurship or business training).

Power calculations showed that 1,200 students per arm were required, but sample size was incremented to account for attrition. We drew a random sample of 4,400 students out of the eligible pool of 7,421. Then, 1,600 students were randomly assigned to hard skills, 1,600 to soft skills and 1,200 for the control group. On each step of the sampling process we stratified by school and gender.

A two-arms intervention was implemented: a 3-week intensive entrepreneurship camp with a strong emphasis on (1) soft skills and (2)hard skills. All students had a basic overview entrepreneurship and worked on a business plan during the 3-week course. The intervention was implemented in May 2013.

Students in the hard skills program focused on financial decision making, while the soft skills arm focused on abilities such as negotiation and communication. The curricula for the training was designed by the International Labor Organization and the Haas Business School.

Teachers were recruited, hired and trained by Educate! a non-profit organization. Teachers were randomly assigned to training site, school and classroom. Each of the 20 host schools was staffed with 3 teachers: 2 regular curriculum instructors (hard or soft skills), who taught both the regular curriculum and 1 instructor who taught the business plan curriculum exclusively. Assignment was stratified by language and ability. The principal investigators of this study are Paul Gertler and Dana Carney at UC Berkeley.

Treatment was assigned randomly, i.e. using a random number generator. This was for identifiability of results.

Overall about one-third are female and they are 20 years old on average.

Sample balanced across 3 arms of the study. 9 of 144 p-values were less than 0.10. Teacher characteristics balanced as well. Of those assigned to treatment, 67.4% participated in the training. None of the controls participated in the training. Our sample consists of 1,021 controls, 1,448 individuals assigned to *hard skills*, and 1,422 individuals assigned to soft skills. Roughly 2/3 of the sample started a business during the recall period, and average monthly revenues and profits were 957*and*501 USD (adjusted for purchasing power parity, PPP).

# 3   Anticipated challenges

Even though assignment to treatment was randomized, compliance with treatment was not perfect (i.e. not every individual assigned to treatment attended the training). Moreover, we were able to reach 88% of the original sample in the follow-up interview. Therefore, estimation of causal effects in this setting entails dealing with a potential selection problem, because individuals who did not attend the training, or individuals who were lost to follow-up could differ in observable and unobservable characteristics correlated with the outcomes. Fortunately, we have baseline covariates of those were lost to follow-up for the original 4,400 individuals. By fully utilizing all the available baseline covariates, our aim is to estimate a double robust locally efficient substitution estimator that will be consistent and asymptotically linear if the selection mechanism is consistently estimated or if we can treat assignment to treatment and attrition as independent events (i.e. no differential attrition between treatment and control).[**?**] An empirical strategy to deal with this censoring issue is to model assignment to treatment A and attrition $\Delta$ as a single intervention node by estimating its joint distribution $f_{A,\Delta}(A, \Delta)$.

# 4   Causal Analysis of Experiment

For each of the three outcomes, the target causal parameter is the Average Treatment Effect, which is the difference in the expected counterfactual if all recruited students had taken the entrepreneurial training and the expected counterfactual if none of the students were assigned to the treatment.

$$\Psi^{\mathcal{F}}(\mathbb{P}_{U,X}) = \mathbb{E}_{U,X}(Y_1) - \mathbb{E}_{U,X}(Y_0)$$

Because this is an RCT (randomized control trial), we can conclude that A is a function of UA only, so that there must be an exclusion restriction between W and A. [Citation: p. 24 of the Mark van der Laan book]

UA is independent of UW and of UY

We may assume that UA is independent of UY and of UW, which corresponds with believing that there are no unmeasured factors that predict both A and the outcome Y: this is often called the no unmeasured confounders assumption.

We are randomizing A, this is often called the no unmeasured confounders assumption.

Any probability distribution of O can be obtained by selecting a particular data-generating distribution of (U, X).

causal assumptions we have made about how the data were generated in nature. Specifically, with the SCM represented in (2.1), we have assumed that the underlying data were generated by the following actions:

1. Drawing unobservable U from some probability distribution PU ensuring that UA is independent of UY, given W.

2. Generating W as a deterministic function of UW.

3. Generating A as a deterministic function of W and UA.

4. Generating Y as a deterministic function of W, A, and UY.

There are no assumptions on functional forms, but because of the randomization of UA, our model is semi-parametric.

We intervene on node A.

The time ordering of the variables is: W -> A -> Y. For the causal ordering, we have that W and A both precede Y, and neither W nor A precede each other.

UA independent UY and UA independent of UW, so there is no backdoor path from Y to A, and our causal estimand is identifiable from the statistical estimand. Thus, there should not be any backdoor paths, and therefore nor should there be any unmeasured confounders.

The intervention node A was deliberately randomized, so these independence assumptions should hold by design of the experiment. We can test this assumption for UA and UW by using a balance table. There is no way for us to test the independence assumption of UA and UY.

# 5   Variables of interest

| Variable type | Variable name | Description |
|---|---|---|
| Y | Business creation | =1 if respondent started a business after graduation from high school |
| Y | Log of revenue | Monthly revenue from all self-employment activities |
| Y | Log of profit | Monthly profit from all self-employment activities |
| A | Treatment | =1 if participated in entrepreneurial training |
| W | Sociodemographic characteristics | Gender, age, parent's income source and education level, boarding student, perceived socioeconomic level |
| W | Cognitive skills | Raven score, math score, GPA, O-level score, previous exposure to entrepreneurship |
| W | Risk and time preferences | Present-bias and time-inconsistency scores |
| W | Personality characteristics | Big 5 (extroversion, emotional stability, openness, conscientiousness, agreeableness), leadership, perceived control, anxiety, pro-social behavior, and more. |

# 6   Works cited

Van der Laan MJ, Rose S, (2011). Targeted learning: causal inference for observational and experimental data. *Springer Science & Business Media*.

Karlan and Valdivia (2011) in Peru

Valdivia (2011) in Peru.

Drexler et al. (2011) in the Dominican Republic.

Berge et al. (2011) in Tanzania.

In Sri Lanka de Mel, McKenzie, and Woodruff (2012)

Appendices

# 7   Appendix 1