An Internship Programme for Young Ethiopian Entrepreneurs: Pre-Analysis Plan

Girum Abebe, Marcel Fafchamps, Michael Koelle and Simon Quinn

Fieldwork location:	Addis Ababa
Fieldwork dates:	November 2015 to January 2018
Date of Pre-Analysis Plan:	March 11, 2018

1 Introduction

We are currently running a field experiment in Addis Ababa, in which we place treated respondents in a four-week 'management internship', to work alongside middle and senior managers of established Ethiopian firms. This Pre-Analysis Plan outlines our primary hypotheses and accompanying identification strategy. We intend to register it with the AEA Hypothesis Registry. At the time of writing, we are finalising 12-month follow-up interviews, and cleaning the data.

2 Data

In this section, we outline the majority of the variables that we plan to use in our analysis (collected both from interns and from firms). There are two additional kinds of outcome variables that we will use: namely, firms' and interns' ranking of the relative importance of different kinds of management practices, and firms' rankings of hypothetical interns. We will use a different identification strategy for these additional variables; for clarity of presentation, we will describe these variables immediately before we outline that alternative identification strategy.

2.1 Data on interns

For simplicity, in this Pre-Analysis Plan, we use the term 'interns' to refer collectively both to the treatment and control groups. We plan to use the following variables about interns.

Note that, for any continuous outcomes (including, for example, profits, earnings, hours worked, *etc*), we will winsorise at the 95th percentile.

The following table summarises our intended outcome variables from our face-to-face surveys.

Table 1: **Variables collected through face-to-face survey**

VARIABLE	DEFINITION	SOURCE
OUTCOME FAMILY 1.1: MAIN OUTCOMES ABOUT EMPLOYMENT		
`self_employed`	Dummy: Respondent is self-employed	Dummy: e1 = 1
`self_employed_hours`	Hours worked (last weekday) in self-employment	u2_a + u2_b/60
`profit_earnings`	Profit for the last month (zero if not self-employed)	e15_01
`wage_employed`	Dummy: Respondent is wage-employed	Dummy: w1 = 1
`wage_employed_formal`	Dummy: Respondent has a permanent wage job	Dummy: w14 = 1
`wage_employed_manager`	Dummy: Respondent has a wage job with managerial responsibilities	Dummy: w5_4 = 1
`self_employed_hours`	Hours worked (last weekday) in wage employment	u1_a + u1_b/60
`wage_earnings`	Wage earnings for the last month (zero if not wage employed)	w11

OUTCOME FAMILY 1.2: PERCEPTIONS OF MANAGEMENT ABILITY

(*Note:* Variables here are dummies for respondents having answered ‘agree’ or ‘strongly agree’.)

perception_idea	Dummy: Has a good idea	Dummy: p1
perception_skills	Dummy: Has necessary technical skills	Dummy: p2
perception_costs	Dummy: Could accurately estimate costs	Dummy: p3
perception_demand	Dummy: Could accurately estimate demand	Dummy: p4
perception_sales	Dummy: Could sell to a new customer	Dummy: p5
perception_findemp	Dummy: Could identify good employees	Dummy: p6
perception_inspire	Dummy: Could inspire/encourage/motivate employees	Dummy: p7
perception_suppliers	Dummy: Could find suppliers to offer a good price	Dummy: p8
perception_seed	Dummy: Has seed money to start	Dummy: p9
perception_banklend	Dummy: Could persuade a bank to lend to finance a business	Dummy: p10
perception_friendlend	Dummy: Could persuade friend/family to lend to finance a business	Dummy: p11
perception_networks	Dummy: Has necessary business networks	Dummy: p12
perception_complicated	Dummy: Too complicated to handle business tasks	Dummy: p13
perception_luck	Dummy: Business success is mostly determined by luck, not skill	Dummy: p14

OUTCOME FAMILY 1.3: MANAGEMENT PRACTICES

(*Note:* These outcomes will be missing for any respondents not running a business.)

practices_all	Score for management practices (weighted using covariance matrix from the control group, as in Anderson (2008))	Weighted sum of all variables listed in the following three rows
practices_marketing	Score for marketing practices (weighted using covariance matrix from the control group)	Weighted sum of e27_1, e27_2, e27_3, e27_4, e27_5, e28, e29, e30

`practices_records`	Score for costing and record-keeping practices (weighted using covariance matrix from the control group)	Weighted sum of e34, e35, e36, e37, e38, e39
`practices_financial`	Score for financial planning practices (weighted using covariance matrix from the control group)	Weighted sum of e40, e42, e43, e44

OUTCOME FAMILY 1.4: PREPARATION FOR SELF-EMPLOYMENT

`business_plan_start`	Dummy: Respondent has plans to start a business	Dummy: n2 = 1 or n2 = 4
`business_plan_expand`	Dummy: Respondent has plans to expand a business	Dummy: n2 = 2 or n2 = 3 or n2 = 4
`business_plan_steps`	Score for preparatory steps taken (weighted using covariance matrix from the control group)	Weighted sum of n11_1 to n11_17
`business_knowledge`	Score for business knowledge (weighted using covariance matrix from the control group)	Weighted sum of k1-k9
`reservation_profit`	Minimum monthly profit to open a business	s4

OUTCOME FAMILY 1.5: SEARCH FOR WAGE EMPLOYMENT

`wage_search_any`	Dummy: Any steps taken to search for a wage job in the past four weeks	Dummy: any of s1-s7 = 1
`wage_search_manual`	Dummy: Search for manual work (set to 0 if `wage_search_any` = 0)	Dummy: s2 = 1
`wage_search_clerical`	Dummy: Search for clerical or administrative work (set to 0 if `wage_search_any` = 0)	Dummy: s2 = 2
`wage_search_prof`	Dummy: Search for professional work (set to 0 if `wage_search_any` = 0)	Dummy: s2 = 3
`wage_search_management`	Dummy: Search for management work (set to 0 if `wage_search_any` = 0)	Dummy: s2 = 4
`reservation_wage`	Minimum monthly wage to accept a job	s3

OUTCOME FAMILY 1.6: BUSINESS NETWORKS

`networks_years`	Total years of contacts' experience	Sum of b3
`networks_count`	Number of contacts listed (maximum of 5)	Count of b2
`networks_senior`	Number of senior contacts	Sum of (b2 $\in \{1, \ldots, 4\}$)
`networks_middle`	Number of mid-level contacts	Sum of (b2 $\in \{5, \ldots, 12\}$)

The following table summarises our intended outcome variables from our monthly phone surveys.

Table 2: **Variables collected through phone survey**

VARIABLE	DEFINITION	SOURCE
OUTCOME FAMILY 2.1: EMPLOYMENT		
`phone_wage`	Whether respondent worked last week in a wage job	(q1 > 0)
`phone_self_employed`	Whether respondent worked last week in own business	(q2 > 0)
`phone_wage_hours`	Hours worked last week in a wage job	q1
`phone_self_hours`	Hours worked last week in own business	q2
`phone_search_wage`	Whether respondent searched for a wage job	Dummy: (q4 == 1)
`phone_search_self`	Whether respondent planned/researched starting own business	Dummy: (q4 == 2)
OUTCOME FAMILY 2.2: BELIEFS ABOUT EMPLOYMENT		
`phone_satisfied`	Whether respondent is satisfied with current employment situation	q5
`phone_wage_belief`	Whether respondent believes that, 12 months from now, (s)he will have a wage job	Dummy: (p8 = 4) or (p8 = 5)
`phone_self_belief`	Whether respondent believes that, 12 months from now, (s)he will be self-employed	Dummy: (p9 = 4) or (p9 = 5)

OUTCOME FAMILY 2.3: PERCEPTIONS OF MANAGEMENT ABILITY		
`phone_supervise`	Respondent knows how to supervise production workers	p11_1
`phone_customers`	Respondent knows how to deal with customers	p11_2
`phone_suppliers`	Respondent knows how to deal with suppliers	p11_3
`phone_advertise`	Respondent knows how to market products or services	p11_4
`phone_materials`	Respondent knows how to source raw materials	p11_5
`phone_accounts`	Respondent knows how to deal with accounts	p11_6
`phone_newwork`	Respondent knows how to hire new workers	p11_7
`phone_debtors`	Respondent knows how to deal with people who do not pay	p11_8
`phone_banks`	Respondent knows how to deal with banks and other financial institutions	p11_9
`phone_prioritise`	Respondent knows how to prioritise his/her time	p11_10

2.2 Data on firms

We have the following data on respondent firms.

Table 3: **Firm outcomes**

VARIABLE	DEFINITION	SOURCE (QUESTION NUMBER)
OUTCOME FAMILY 3.1: ADVERTISING FOR NEW EMPLOYEES		
`ad`	Did any advertising for new hires	l4_1
`ad_board`	Did advertising on the job boards	l4_2_1
`ad_newspapers`	Did advertising in the gazette or other newspapers	l4_2_2
`ad_post`	Did advertising outside premises	l4_2_3
`ad_online`	Did advertising online	l4_2_4

`ad_agency`	Did advertising by agency/broker	l4_2_5
`ad_university`	Did advertising on university/college campuses	l4_2_6
`ad_fairs`	Did advertising through job fairs	l4_2_7

OUTCOME FAMILY 3.2: LABOUR FLOWS

`hires_total`	Total hires (last two months)	Sum of the next four variables
`hires_professional`	Professional hires (last two months)	l3_2_p (= 0 if none)
`hires_services`	Client services hires (last two months)	l3_2_c (= 0 if none)
`hires_production`	Production worker hires (last two months)	l3_2_w (= 0 if none)
`hires_support`	Support services hires (last two months)	l3_2_s (= 0 if none)
`sep_total`	Total separations (last 12 months)	Sum of the next four variables
`sep_professional`	Professional separations (last 12 months)	l1_2_p (= 0 if none)
`sep_services`	Client services separations (last 12 months)	l1_2_c (= 0 if none)
`sep_production`	Production worker separations (last 12 months)	l1_2_w (= 0 if none)
`sep_support`	Support services separations (last 12 months)	l1_2_s (= 0 if none)

OUTCOME FAMILY 3.3: MANAGEMENT PRACTICES

`mgmt_total`	Overall management practices z-score	z-score of average of `mgmt_operations`, `mgmt_monitoring`, `mgmt_target` and `mgmt_incentives`
`mgmt_operations`	Operations practices z-score	m25 z-score
`mgmt_monitoring`	Monitoring practices z-score	Average of `mgmt_monitoring1` to `mgmt_monitoring7`
`mgmt_monitoring1`	How many production performance indicators (PPI)	m26 z-score (recode -9 = 1)
`mgmt_monitoring2`	How frequently PPI collected	m27 z-score (recode -9 = 1)

mgmt_monitoring3	How frequently PPI shown to managers	(m28 + 1) z-score (recode 8 = 1, recode 'other')
mgmt_monitoring4	How frequently PPI shown to workers	(m29 + 1) z-score (recode 8 = 1, recode 'other')
mgmt_monitoring5	Where PPI displayed	m30 z-score (recode -9 = 1)
mgmt_monitoring6	How often PPI reviewed	recoded m31 z-score (recode 1 = 3, 3 = 1)
mgmt_monitoring7	Are PPI compared	recoded m32 z-score (recode 1 = 2, 2 = 1)
mgmt_target	Target practices z-score	recoded m33 z-score (recode 1 = 2, 2 = 3, 3 = 4, 4 = 1, -9 = 1)
mgmt_incentives	Incentive practices z-score	Average of mgmt_incentives1 to mgmt_incentives3
mgmt_incentives1	Rewarding target achievements	m34 z-score
mgmt_incentives2	Promoting employees	recoded m35 z-score (recode 1 = 3, 3 = 1)
mgmt_incentives3	Moving employees	m9 z-score
mgmt_records	Record-keeping practices z-score	Average of mgmt_records1 to mgmt_records5
mgmt_records1	Issue invoices	recoded m20 z-score (recode 1 = 4, 2 = 3, 3 = 2, 4 = 1)
mgmt_records2	Pay on invoice	recoded m21 z-score (recode 1 = 4, 2 = 3, 3 = 2, 4 = 1)
mgmt_records3	Minute of meetings	recoded m22 z-score (recode 1 = 2, 2 = 1, -9 = 1)

mgmt_records4	Keeping of archives	recoded m23 z-score (recode 1 = 2, 2 = 1, -9 = 1)
mgmt_records5	Written reports	recoded m24 z-score (recode 1 = 2, 2 = 1, -9 = 1)
mgmt_marketing	Marketing practices z-score	Average of mgmt_marketing1 and mgmt_marketing2
mgmt_marketing1	Advertising	recoded m19 z-score (recode 1 = 2, 2 = 1)
mgmt_marketing2	Warranties	recoded m16 z-score (recode 1 = 5, 2 = 4, 4 = 2, 5 = 1)

For management practices, we only will use the six area scores and the overall management practices scores as outcome variables in the analysis, but list the component variables for completeness and to document their coding. All the components of `mgmt_total` are based on Bloom, Schweiger, and Van Reenen (2012); we keep the questions and the coding identical to ensure comparability of management in Ethiopian fims to other countries. We elicit additional scores for record-keeping and marketing practices; we will not include these two categories in the overall management practices score.

We will calculate z-scores of variable x_i for observation i as follows:

$$z_i = \frac{x_i - \bar{x}}{\sigma_x}. \tag{1}$$

We will calculate the mean $\bar{x}$ and the standard deviation σ_x from the baseline data; and apply these moments for the z-score calculation both at baseline and at endline.

2.3 Testing balance

We will begin our analysis by testing balance. We will test balance both for the interns and for the firms.

To test balance for the interns, we will take each of the variables described earlier in Table 1, and will run the following regression:

$$y_{ip0} = \beta_0 + \beta_1 \cdot T_i + \delta_p + \varepsilon_i, \tag{2}$$

where i indexes interns, p indexes the intern pairs used for randomisation, where y_{ip0} refers to the baseline value of the variable and T_i is a dummy for being treated. We will allow for robust standard errors.

That is, using Stata code, we will estimate:

```
ivreg2 y_pre treat pair*, partial(pair*) robust
```

To test balance for the firms, we will take each of the variables described earlier in Table 3, and will run the following regression:

$$y_{fg0} = \beta_0 + \beta_1 \cdot T_f + \delta_g + \varepsilon_f, \tag{3}$$

where f indexes firms, g indexes the 'gathered fields' used for randomisation, where y_{fg0} refers to the baseline value of the variable and T_f is a dummy for being treated. We will allow for robust standard errors.

That is, using Stata code, we will estimate:

```
ivreg2 y_pre treat group*, partial(group*) robust
```

As a formal test that our randomisation worked, we will then conduct an omnibus F-test for the joint hypothesis that all β_1 coefficients are equal to zero (that is, a single test across all variables for interns and all for firms).

3 Treatment effects on interns

3.1 Basic estimating specification

We will test the effects of the internship on a variety of outcomes for the individual (we will shortly outline our intended structure of outcome variables for doing this). We begin by outlining our preferred estimation specification for interns.

For some individual respondent i, denote T_i as a dummy for whether i was assigned to treatment. Treatment status was assigned using matched pairwise dummies; we index these dummies by p. We observe each individual at baseline (which we denote as $t = 0$), at a six-month follow-up (which we denote $t = 1$) and at a 12-month follow-up ($t = 2$). Our preferred estimating equation is ANCOVA with pairwise dummies; that is, for individual i in pair p at time $t > 0$, we intend to estimate:

$$y_{ipt} = \beta_1 \cdot T_i + \beta_2 \cdot y_{ip0} + \delta_p + \varepsilon_{ipt}. \tag{4}$$

Neither our sampling process nor our assignment mechanism was clustered; therefore, following the recent guidance of Abadie, Athey, Imbens, and Wooldridge (2017), we will use robust standard errors rather than by clustering at any higher level of aggregation. (When we pool across waves, we will — of course

— allow for clustering at the level of the respondent.)

Using Stata code, denote `y` as some outcome variable of interest, `y_pre` as the baseline value, `treat` as a dummy for treatment, `pair*` as the set of pairwise dummy variables, and `IndividualID` as the identifier for the individual participants. Then we will estimate the following by OLS (pooling across $t = 1$ and $t = 2$):

```
ivreg2 y treat y_pre pair*, partial(pair*) cluster(IndividualID)
```

We will interpret $\hat{\beta}_1$ as the estimated 'Intent to Treat'.

In our primary specification, we will pool across follow-up waves; we also plan to report (either in the main paper or in an appendix) separate estimations for each follow-up wave.

For each hypothesis test, we will report two values:

(i). The usual p-value from a Wald test; and

(ii). We will report False Discovery Rate q-values, within the relevant family of outcomes (Benjamini, Krieger, and Yekutieli, 2006).

We anticipate that, prompted by our results on these outcomes, we will run further analysis on other outcomes, in order to further explore any mechanisms at work. We will acknowledge in the paper where analysis goes beyond the regressions pre-specified here.

3.2 Analysis by outcome families

We now outline a series of regression families. We structure these outcomes to represent primary outcomes of interest, followed by potential mechanisms (see Olken (2015)). Our experimental design has three primary hypotheses: namely, that the internship changes employment outcomes (both in the sense of wage employment and self-employment), that the internship changes attitudes about management practices, and that — among self-employed respondents — the internship changes management practices.[1] We will then test a set of secondary outcomes (*i.e.* in this context, mechanisms) — namely, we will test effects on steps taken to search for wage employment, effects on steps taken to search/prepare for self-employment, and effects on business networks.

3.2.1 Primary outcome: Occupation

To estimate the effect of treatment has any effect on occupation, we will estimate equation 4 using the outcomes in Outcome Family 1.1.

3.2.2 Primary outcome: Attitudes about management practices

To estimate the effect of treatment has any effect on attitudes about management practices, we will estimate equation 4 using the outcomes in Outcome Family 1.2. We intend to estimate with each of these outcomes in turn.

We view all of these outcomes as reflective of a latent variable — namely, the respondent's perceptions about his/her skills/ability to run a business. Therefore, in addition to estimating with each of these outcomes in turn, we will also estimate (i) using the sum across dummy variables, and (ii) using an index of outcomes (constructed following the recommendation in Anderson (2008)). Note that, for both of these additional estimations, we will reverse the dummy variable coding for the last two variables

[1] In our original funding proposal, for example, we emphasise the value of this experimental design for testing how exposure to established managers facilitates changes in views on management practices among aspiring entrepreneurs.

(perception_complicated and perception_luck), so that an increase in each variable indicates a perception of being more skilled/able.

3.2.3 Primary outcome: Management practices among self-employed respondents

To estimate the effect of treatment on management practices among the self-employed we will estimate for each of the outcomes in Table 1.3 (with outcomes set to missing for those respondents not running a business).

3.2.4 Mechanism: Effects on preparing for self-employment

To estimate whether the treatment encouraged respondents to prepare for self-employment, we will estimate using the outcome variables in Table 1.4.

3.2.5 Mechanism: Effects on search for wage work

To estimate whether the treatment encouraged respondents to search for wage employment, we will estimate using the outcome variables in Table 1.5.

3.2.6 Mechanism: Effects on networks

To estimate whether the treatment encouraged respondents to search for wage employment, we will estimate using the outcome variables in Table 1.6.

3.3 Treatment effects by month

We conducted monthly phone surveys with both treated and control interns. We can estimate the trajectory of treatment effects by pooling all phone observations and estimating quadratic trends over time of the treatment effect. To do this, we estimate the following, subject to quadratic constraints on the treatment effects (where $m > 0$ indexes the months after treatment, c indexes calendar months, and p again indexes the intern pairs used for randomisation):

$$y_{ipmc} = \beta_m \cdot T_i + \delta_p + \eta_m + \omega_c + \varepsilon_{ipmc},$$

subject to:

$$\beta_m = \phi_0 + \phi_1 \cdot m + \phi_2 \cdot m^2.$$

That is, instead of estimating parameters β_m, we will estimate ϕ_0, ϕ_1, and ϕ_2. (We also also anticipate estimating equation 5 separately for each month, in which case η_m drops from the regression.) We anticipate producing graphs of the form generated in Abebe, Caria, Fafchamps, Falco, Franklin, and Quinn (2016); that is, showing both point estimates on a monthly basis (estimated as just described), with the quadratic fit superimposed.

We anticipate running this estimation for each of the variables in Outcome Family 2.1, Outcome Family 2.2 and Outcome Family 2.3.

4 Treatment effects on firms

4.1 Effects on firm practices

We have four hypotheses for the possible effect on firms of hosting interns: namely, that treatment caused firms to perceive prospective interns differently, that treatment encouraged firms to increase advertising for future hires, that treatment changed firms' labour flows (whether by increasing hiring or separations, or both), and that treatment changed firms' labour management practices more generally.

We view the first of these hypotheses as being a primary/direct potential outcome. We view the remaining three hypotheses as secondary, and we structure our analysis accordingly.

4.1.1 Primary outcome: Perceptions of prospective interns

Both treated firms and control firms were asked to provide a hypothetical ranking of interns, after each batch of interns finished. On the firm side, our primary hypothesis is that treatment causes changes in perceptions about prospective interns. We propose to test for such an effect as follows. For firm f ranking hypothetical intern i, we will use the following latent-utility difference specification:

$$y_{fi}^{*} = \beta_0 \cdot X_i + \beta_1 \cdot X_i \cdot T_f + \varepsilon_{fi}, \tag{5}$$

where T_f is a dummy for firm f having been treated, and X_j is a characteristic of hypothetical intern i (which we specify shortly). We will use this specification to estimate a rank-ordered logit, using data from the hypothetical rankings of interns. The key parameter of interest here is β_1; the null hypothesis of no treatment effect is tested as $H_0 : \beta_1 = 0$.

We plan to estimate for the following definitions of X_i:

(i). X_i as a dummy for whether the hypothetical intern is female;

(ii). X_i as a dummy for whether the hypothetical intern has age at or above the sample median;

(iii). X_i as a dummy for whether the hypothetical intern has education at or above the sample median; and

(iv). X_i as a dummy for whether intern i is currently running his or her own business.

We interpret this as a test of whether, across the sample as a whole, the internship changed firms' attitudes towards interns. For example, if hosting an intern made firms more favourable towards potential female interns, we would estimate $\beta_1 > 0$. We plan to treat this as a family of estimations, and report sharpened q-values across the family.

We then plan to extend the specification as follows:

$$y^*_{fi} = \beta_0 \cdot X_i + \beta_1 \cdot X_i \cdot T_f + \beta_2 \cdot X_i \cdot T_f \cdot Z_f + \varepsilon_{fi}, \tag{6}$$

where Z_f is a dummy variable for whether firm f hosted an intern having the same characteristic that defines X_i. (For example, if X_i is a dummy variable for whether the hypothetical intern is female, then Z_f will be a dummy for whether firm f hosted a female intern. We plan to estimate using the same set of specifications for X_i.) Our parameter of interest is β_2. We interpret this as a test of whether the internship changed firms' attitudes towards interns *who are similar to the intern hosted*. For example, if hosting a *female* intern made firms more favourable towards potential female interns, we would estimate $\beta_2 > 0$. We plan to treat this as a family of estimations, and report sharpened q-values across the family.

4.1.2 Secondary outcome: Advertising for new employees

For some respondent firm f, denote T_f as a dummy for whether f was assigned to treatment. Treatment status was assigned using 'gathered fields' of firms; we index these gathered fields ('groups') of firms by g. We observe each firm at baseline (which we denote as $t = 0$), and immediately after the time of the internship (which we denote $t = 1$).

For secondary outcomes, our preferred estimating equation is ANCOVA with group dummies; that is, for firm f in group g at time $t = 1$, we intend to estimate:

$$y_{fg1} = \beta_0 + \beta_1 \cdot T_f + \beta_2 \cdot y_{fg0} + \delta_g + \varepsilon_{fg}, \tag{7}$$

That is, using Stata code, we will estimate:

```
ivreg2 y treat y_pre group*, partial(group*) robust
```

For each hypothesis test, we will report two values:

(i). The usual p-value from a Wald test; and

(ii). We will report False Discovery Rate q-values, within the relevant family of outcomes (Benjamini, Krieger, and Yekutieli, 2006).

We anticipate that, prompted by our results on these outcomes, we will run further analysis on other outcomes, in order to further explore any mechanisms at work. We will acknowledge in the paper where analysis goes beyond the regressions pre-specified here.

To test for effects on advertising for new employees, we will estimate using the outcome variables in Outcome Family 3.1. In addition to each of these separate measures, we will add two additional outcome variables: the sum of different types of advertising, and the weighted sum of different types of advertising (with weights calculated as recommended in Anderson (2008)). We view these final two outcomes as providing summary measures of the extent of advertising activities (in the sense of measuring an increase in scope, through a variety of different advertising methods).

4.1.3 Secondary outcome: Labour flows

To test for effects on labour flows, we will estimate using the outcome variables in Outcome Family 3.2.

4.1.4 Secondary outcome: Management practices

Finally, to test for effects on host firm management practices, we will estimate using the outcome variables in Outcome Family 3.3.

5 Testing diffusion from host firms to hosted interns

5.1 Testing diffusion of ideas

This relies on our rankings of the relative importance of different management practices. We have this from five separate sources:

(i). The treated and control interns at baseline;

(ii). The treated firms (at the initial ranking interview);

(iii). The treated and control firms (at the immediate follow-up);

(iv). The treated interns (at immediate follow-up);

(v). The treated and control interns at six months;

(vi). The treated and control interns at 12 months.

We need to construct a distance measure between the ranking given by intern i and the ranking given by firm j. For this, we will use "Kendall's τ". That is, we use the ranking to construct a matrix of pairwise comparisons — showing, for each pair, which of the two is preferred. Then, between two matrices, we have τ as 'proportion of pairwise comparisons where intern and firm agreed, minus proportion of pairwise comparisons where intern and firm disagreed'. Note that $\tau = 1$ for complete concordance, $\tau = -1$ for perfect discordance, and $\tau = 0$ (in expectation) for one or both rankings being random.

We continue to index interns by i and firms by f. This estimation will use both treated and control interns, and treated firms. (We do not use control firms, because we did not obtain rankings from them prior to the internship.) We will construct a large dyadic dataset within each session — in each case forming τ_{if}. Then we will run the following ANCOVA regression, clustering by session:

$$\tau_{ift} = \beta_1 \cdot A_{if} + \beta_2 \cdot T_i + \beta_3 \cdot \tau_{if0} + \delta_{ip} + \varepsilon_{ift}, \tag{8}$$

where A_{ij} is a dummy for intern i being assigned to firm f, T_i is a dummy for i being treated, and δ_{ip} are pairwise dummies to record the way that we randomised interns. We use t to denote the time of follow-up (discussed shortly), and we denote τ_{if0} as the baseline measure of concordance.

Therefore, we can test the following:

(i). $H_0 : \beta_1 = 0$ is a test for *specific* diffusion of management ideas from firm f to intern i;

(ii). $H_0 : \beta_2 = 0$ is a test for whether, in *general*, the treatment moved the treated group 'closer' to firm managers.

Note that, to construct τ_{ij}, we will use the firm ranking at the ranking interview, then the interns' rankings respectively at immediate follow-up (*i.e.* just for the treated), then six month follow-up and 12 month follow-up. That is, we will report three different estimations: one for immediate follow-up (using just the treated interns, so T_i will be dropped from the estimation), one for the six month follow-up, and one for the 12 month follow-up.

5.2 Testing diffusion of management practices

For those interns running a business, we will test directly for diffusion of implemented management practices. We will do this in two ways: by testing for diffusion in the level/quality of management practices, and by testing for diffusion in the relative importance (that is, the ranking) of management practices.

5.2.1 Testing diffusion on individual measures of management practice

Our estimating specification is as follows:

$$y_{ift} = \beta_1 \cdot M_{f0} + \beta_2 \cdot T_i + \beta_3 \cdot y_{if0} + \beta_4 \cdot s_{i0} + \varepsilon_{ift}, \tag{9}$$

where i indexes interns, f now indexes the host firm for intern i and t refers to the survey round ($t = 0$ for baseline, $t = 1$ for the six-month follow-up, and $t = 2$ for the 12-month follow-up). M_{f0} refers to a measure of host firm management practices at baseline, and is set to zero for interns in the control group. y_{if0} refers to the baseline value of the explanatory variable; if the intern was not running a firm at baseline, this is set to zero. s_{i0} is a dummy for whether the intern was running a firm at baseline. T_i refers to whether intern i was treated. Note that this estimation will only be run for those interns who are running a firm; we will therefore cluster by treatment assignment pair, rather than including pairwise dummies.[2]

We will apply this specification in two ways. First, we will define y_{ift} as `practices_all` (*i.e.* the overall management score for intern i). We will define M_{f0} as `mgmt_total` (*i.e.* the overall management score for host firm f). This is a general test of whether management quality (in the aggregate) diffuses from firms to interns. Second, we will run three separate regressions, respectively defining y_{ift} as `practices_marketing`, `practices_records` and `practices_financial`. For each regression, we will expand M_{f0} to be a vector, comprising `mgmt_operations`, `mgmt_monitoring`, `mgmt_target` and `mgmt_incentives`. This is a more specific test for whether particular aspects of management quality are differentially affected by the different components of host firm management quality. For this estimation, we will report individual tests on each of the regressors, and will also report an omnibus test for the null hypothesis that the coefficients on `mgmt_operations`, `mgmt_monitoring`, `mgmt_target` and `mgmt_incentives` are all zero.

5.2.2 Testing diffusion on rankings of management practice

By direct analogy to section 5.1, we will test for diffusion in the relative importance of management practices. (For example, one might imagine that the internship experience does not affect the overall quality of an intern's management, but may affect the relative importance that the intern attaches to different aspects of management.)

[2] If we were to include pairwise dummies, we would be estimating only for the narrow subset of observations where both members of a dummy are running a business at follow-up.

To do this, we will form five common categories between large and small firms, according to the following table.

Firm management practice area	Intern management practices ranking codes
Marketing	01, 02, 03, 04
Record-keeping	05, 06
Targets	07
Incentives	08, 09
Uncategorised	10

Both for host firms and for interns, we will rank the relative importance of these five categories. We will do this using an ordering of z-scores across the areas; for each firm and for each intern enterprise, we will therefore have a ranking of the relative quality of management in different categories. Our analysis will then proceed in the same way as in section 5.1; namely, we will construct τ_{ift} as before, and estimate again using equation 8. For the reasons discussed in the previous subsection, we will no longer use pairwise dummies; instead, we will cluster by stratification pairs. As in equation 8, we will include τ_{if0}. If intern i was not running a firm at baseline, we will set $\tau_{if0} = 0$; we will also include a dummy variable to measure whether intern i was running a firm at baseline.

6 Heterogeneity

We plan to study treatment effects for a number of relevant sub-groups. Sub-groups are identified by categorical variables capturing characteristics at baseline. When characteristics are continuous, we create subgroups by separating individuals (i) below the median of the characteristic and (ii) at or above the median level of the characteristic.

For each intervention, we will run the following specification:

$$y_{ipt} = \beta_1 \cdot T_i + \beta_2 \cdot I(x_{i0} = \nu) + \beta_3 \cdot I(x_{i0} = \nu) \cdot T_i + \alpha \cdot y_{ip0} + \delta_p + \varepsilon_{ipt}, \quad (10)$$

where $I(x_{i0} = \nu)$ is a dummy whether the categorical variable x_{i0} measured using data at baseline belongs in category ν. All other variables are defined as before.

We plan to study heterogeneity in impacts for the subgroups defined in Table 4.

For each hypothesis test, we will report two values:

(i). The usual p-value from a Wald test; and

(ii). We will report False Discovery Rate q-values, taken across the relevant family of mediators (Benjamini, Krieger, and Yekutieli, 2006).

We anticipate that, prompted by our results on these outcomes, we will run further analysis on other outcomes, in order to further explore any mechanisms at work. We will acknowledge in the paper where analysis goes beyond the regressions pre-specified here.

Table 4: **Subgroups for heterogeneous treatment effects**

VARIABLE	DEFINITION	SOURCE (QUESTION NUMBER)
	MEDIATOR FAMILY 4.1: HUMAN CAPITAL	
`degree`	Respondent has a degree	dummy (f18 = 20 to 22)
`cognitiveskills`	Skills tests: mathematical ability (T1), English language (T2), digit span (T3)	Average of z-scores from t_1, t_2, and t_3
	MEDIATOR FAMILY 4.2: FINANCIAL CAPITAL	
`assets`	Above median baseline assets	Assets are sum of all rows of a5_ and a8 and a9

pred_loan	Predicted access to loan	Predicted endline value of (a12 + a17 + a23 + a30 + a35)

MEDIATOR FAMILY 4.3: PROSPECTS IN WAGE-EMPLOYMENT

base_perm_wagejob	Dummy: respondent had a permanent wage job at baseline	Dummy: w14 = 1
pred_perm_wagejob	Dummy: respondent predicted to have a permanent wage job at endline	Predicted endline w14 = 1
wage_search_any	As defined previously	

MEDIATOR FAMILY 4.4: GENDER

female	Dummy: respondent is female	Dummy: f7 = 2

When our mediator is a predicted variable, we obtain predictions as follows: we train a machine learning algorithm using features from the baseline survey. We regard all features (variables) from the baseline survey as potentially relevant. (For computational reasons, we might drop some variables before we start training the algorithm; especially if a variable contains many missing values.) We will use empirical tuning to define the optimal degree of complexity of the algorithms, which governs the set of variables included in the final prediction.

When the variable to be predicted is measured at endline, we use only the control group as our training sample.

7 Dealing with potential problems

7.1 Attrition

Our attrition is very low, and we do not anticipate this to be an empirical problem in this setting. Our field team placed a lot of emphasis on tracking respondents; including following respondents to other parts of Ethiopia and conducting some interviews on the phone. Our tracking sheets indicate that we successfully surveyed 96% of our initial sample at the 6-month follow-up, and 95% at the 12-month follow up survey.

7.2 Intent-to-treat: firms and interns

We have imperfect compliance of both firms and interns. Our experimental sample only consists of firms who had, in principle, agreed to host interns, and interns who turned up to an indication session we invited them to based on an initial expression of interest. Nevertheless, some firms subsequently refused to act as a host firm based on operational or capacity reasons. Similarly, some interns randomised to treatment did not complete their internship and dropped out at various stages after the randomisation. Some of these dropouts occurred because we had to ask interns to defer their placement due to capacity constraints; and anecdotally, such interns did not always come back to take up their placements. In such cases, we randomly chose which interns we asked to defer. This does not pose a conceptual challenge for our identification strategy; we are using an Intent-to-Treat interpretation, so we use the assignment to treatment as the relevant explanatory variable.

7.3 Re-assignment of interns

In a small number of cases, host firms refused to offer placements to interns after they were assigned to their specific firm. In such a case, we would re-invite the affected treated intern to complete another ranking exercise and be placed in a new firm. In other words, we treated them as if they had simply deferred their placement, without any further differences in the assignment mechanism. For the purpose of our diffusion analysis, we will assign as the matched firm the firm that actually hosted the intern (i.e. the firm they were re-assigned to).

References

ABADIE, A., S. ATHEY, G. W. IMBENS, AND J. WOOLDRIDGE (2017): "When Should You Adjust Standard Errors for Clustering?," Discussion paper.

ABEBE, G., S. CARIA, M. FAFCHAMPS, P. FALCO, S. FRANKLIN, AND S. QUINN (2016): "Curse of Anonymity or Tyranny of Distance? The Impacts of Job-Search Support in Urban Ethiopia," *NBER Working Paper No. 22409*.

ANDERSON, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 103(484).

BENJAMINI, Y., A. M. KRIEGER, AND D. YEKUTIELI (2006): "Adaptive Linear Step-up Procedures that Control the False Discovery Rate," *Biometrika*, 93(3), 491–507.

BLOOM, N., H. SCHWEIGER, AND J. VAN REENEN (2012): "The Land that Lean Manufacturing Forgot?," *Economics of Transition*, 20(4), 593–635.

OLKEN, B. A. (2015): "Promises and Perils of Pre-Analysis Plans," *The Journal of Economic Perspectives*, 29(3), 61–80.