

**Building High-Growth Firms Through Training the Owner vs. Through Linking the Firm
to Business Service Markets**

PRE-ANALYSIS PLAN

February 2019

Principal Investigators

David McKenzie, *World Bank*

Stephen J. Anderson, *Stanford University GSB*

Version 1: submitted 25th March 2018

(Prior to Follow-up Survey #1 conducted in April-June 2018)

Version 2: submitted 27th February 2019

(Prior to Follow-up Survey #2 conducted in February-April 2019)

The associated trial was registered on the AEA Social Science Registry on 14th March 2018, and the pre-analysis plan was registered on 25th March 2018. The assigned registry number is AEARCTR-0002769. See also: <http://www.socialscienceregistry.org/trials/2769>

Key Changes in Version 2 compared to Version 1

This version updates the original pre-analysis plan with the following key changes:

- *Updating information on the program's roll-out, in which a planned second year cohort of inductees did not take place, resulting in the final sample size being lower than initially planned.*
- *Updating the dates for a second round follow-up survey, and making clear that this is the last survey round that will be done under the program.*
- *Adding in additional outcomes that will be measured in the second round follow-up survey – these are intended to capture additional channels of potential impact.*

These changes are indicated with an explanation in italics, in the following plan.

1. INTRODUCTION

1.1. Overview

Many small firms lack the finance and marketing skills needed for firm growth. The standard approach in many development programs has been to attempt to train the owner to develop these skills, through business training sessions or personalized consulting services. However, an alternative is to link firms to these skills in the market through insourcing workers with these skills, or outsourcing these tasks to professionals specializing in these services. We test which approach works best to grow small firms through a randomized experiment in the context of the Growth and Employment (GEM) project in Nigeria.

The Nigeria GEM project is a World Bank project with the development objective of increasing firm growth and employment. It aims to do this indirectly through improving the investment climate in five economic sectors – light manufacturing, construction, hospitality, ICT and entertainment – and directly, by offering programs to improve the performance of firms in these sectors. These direct support activities are channeled through the Business Innovation and Growth (BIG) Platform, which was launched in 2016 and through which Micro, Small and Medium Enterprises (MSMEs) apply for these programs. By December 2016, when the registration was closed for the first year, 48,167 firms had registered on the BIG Platform throughout the country.

This impact evaluation intends to measure the impact of the GEM program on the performance of Nigerian firms. To do so, the evaluation will use a randomized control trial (RCT) methodology. The project aimed to recruit 2,000 firms (*but ended up with an actual experimental sample of 753 firms*) (with 2-15 workers each) and randomly assign them into one of five equal-sized experimental groups: (i) a control group; (ii) a group given business training for the owner; (iii) a group given consulting services; (iv) a group linked to HR specialists who will find a worker to insource these skills; and (v) a group linked to companies with professionals specializing in business services to outsource these skills. Impacts on firm practices and performance (e.g., sales, profits, employment) will be measured through multiple post-

intervention survey rounds. To the best of our knowledge, this is the first test of such a market-based approach to skill-development in firms.

This pre-analysis plan outlines the hypotheses to be tested and specifications to be used in the analysis of the GEM program's impact. Since the authors completed the plan before the follow-up data was collected and analyzed, the plan can serve as a commitment and marker of the key outcomes to be examined, and help guard against the risks of data mining and multiple hypothesis testing.

1.2. Motivation

The inability of firms to operate with advanced business practices, in particular to effectively manage finances and market products, is seen as key constraint that inhibits the growth of small firms. An additional (and related) challenge is the lack of specialization across business functions. Being able to focus the resources of each functional unit (e.g., finance, HR, production, marketing, sales, etc.) not only allows firms to efficiently scale up, but also gives owners the ability to delegate the more 'standardized' business practices (operational level) while freeing up time for them to focus on 'growth' related activities (strategic level). McKenzie and Woodruff (2017) show that small firms with better business practices earn more, are more likely to survive, and grow faster. What is less clear is the best way for firms to overcome these constraints. Experience with typical business training programs has been mixed (McKenzie and Woodruff, 2014), and it can be difficult to get business owners to take time away from their firms to attend lengthy trainings. Further, even if an owner attends training, it is unlikely she will have the time to personally implement all of her newly acquired skills across multiple business functions. Indeed, being a "jack-of-all-practices" will be especially difficult as the firm starts to grow and scale its business.

As an alternative to training, consulting services also have the potential to improve business practices and, in turn, increase employment (Bruhn, Karlan and Schoar, 2017) and productivity (Bloom et al, 2013). In addition, there are other ways firms can augment skills such as through insourcing or outsourcing services. Either approach could be effective at both improving business practices and increasing functional specialization so owners can delegate and focus on growth activities. Nonetheless, despite the potential returns, firm owners may be reluctant to go to the marketplace and pay for these types of business services (e.g., consulting, insourcing, outsourcing) because they are too expensive, there is a lack of information about providers, or the quality is uncertain. The interventions designed for this study therefore aim to test the training approach against different business service marketplaces.

2. RESEARCH DESIGN

2.1. Study Context

Nigeria enjoyed a period of sustained growth during the 2000s as a combination of high oil prices and domestic reforms enabled growth rates exceeding 6 percent per year. However, the recent drop in oil prices caused Nigeria to enter into recession in 2016, its worst economic

performance in thirty years. Diversifying into non-oil sources of growth, including developing the private sector, is crucial for the continued growth of the economy. Nigeria has a strong reputation for entrepreneurship, with Nollywood and the ICT sector being particularly prominent. Yet more than 99 percent of all firms in Nigeria still consist of only the owner, and there is a need to better identify and grow promising firms. Thus, this under-developed private sector represents an ideal research context for testing whether (and how) high-growth firms can best be developed by supporting the owner (e.g. via training) or by linking the firm to business service markets (e.g., via consulting, insourcing, outsourcing).

2.2. Interventions to be Evaluated

Implementation of the program is occurring through a project implementation unit set up under the Nigerian Federal Ministry of Industry, Trade and Investment. Firms that meet the eligibility criteria have been randomly assigned to one of the following experimental groups (with an intervention offered in four of them):

- (i) Control: Firms in this group are not offered an intervention. However, they will be tracked over time and interviewed in follow-up surveys.
- (ii) Training: Firms in this group are provided with a mix of online and in-class training focused on strengthening their capacity. The curriculum is based on the IFC Business Edge and adapted to the local context. The core modules are in financial management, marketing, general operations, and human resource management. Firms can also choose from among modules on personal productivity, enterprise governance, and tourism and hospitality. Firms have to complete a minimum of 12 days in-class along with online courses. The business owners can choose the dates, locations, and topics for the modules. This training intervention is offered for free to the firms.
- (iii) Consulting: Firms in this group are offered 11 days of business consulting services (~88 hours) by consultants over 6 to 9 months, meeting at least once per month. The initial visit (about 8 hours at the business site) helps the firm define a Needs Assessment and a Growth Strategy. The consulting services provider then proposes a list of business areas (Scope of Support) where it can advise the owner for the remaining 80 hours. These activities are customized to each firm but typically focus on Management, Finance, Sales and Marketing, Operations, and Human Resources. This consulting intervention is offered for free to the firms.
- (iv) Insourcing: Firms in this group are given access to an online marketplace where they can choose a Human Resource (HR) specialist from a list of vetted service providers. This HR specialist then helps recruit one accounting worker or one marketing worker to join the firm and perform tasks in the respective functional area. This ‘internally hired’ worker spends at least 5 days per week (~40 hours/week) at the business site implementing activities for the firm. The firm receives a subsidy that fully pays the cost of such a worker in the first few months, but then gradually declines over 9 months (after which time the firm covers the worker’s entire wage). Firms will also

leave “yelp-like” ratings for HR specialists, enabling a market with reputations to develop over time. *This ratings marketplace was intended to be used on new cohorts, but now will be only used for a second experiment on how ratings shape provider choice.*

- (v) Outsourcing: Firms in this group are given access to an online marketplace where they choose an Accounting specialist or a Marketing specialist from a list of vetted service providers. The firm then outsources its accounting or marketing functions to a professional from the chosen service provider. This ‘externally contracted’ professional spends at least 1 day per week (~8 hours/week) at the business site implementing activities for the firm. The firm receives a subsidy that fully pays the cost of such a professional in the first few months, but then gradually declines over 9 months (after which time the firm covers the professional’s entire cost of service). Firms will also leave “yelp-like” ratings for service providers, enabling a market with reputations to develop over time. *This ratings marketplace was intended to be used on new cohorts, but now will be only used for a second experiment on how ratings shape provider choice.*

Three of the four business development interventions are value-equivalent. That is, they have been designed so that the cost of offering each intervention is approximately the same, at around \$2,000 per firm. (The only exception is consulting, which costs twice the price of the others.)

2.3. Sample Recruitment and Screening

To be included in the study, firms are recruited through the following steps:

- Step One: Firms must apply online through the BIG platform in response to an advertising campaign.¹ McKenzie (2017) notes that requiring online application already screens on firms that are more sophisticated and likely to be plausible candidates for high-growth. Moreover, by applying for the program firms indicate their interest in developing their skills.
- Step Two: Firms must then pass an initial screening based on the government’s eligibility criteria:
 - Have complete data in their online application form.
 - Operate in one of the five GEM priority industries (light manufacturing, construction, ICT, hospitality, entertainment).
 - Be 18 years of age or older.
 - Have more than one and fewer than 100 workers.
- Step Three: Next, eligible firms are invited and must attend an induction workshop in the region where they operate. At this workshop, firm owners receive additional information about the program (generic and not intervention specific), as well as complete a baseline survey (conducted by an independent enumerator). This step further screens on motivation and effort.

¹ <https://bigportal.org.ng/>

- Step Four: Finally, using data from the baseline survey, the researchers apply an additional set of screening criteria to select firms into the study sample:
 - Have a score of between 5 and 8 (out of 10) on their current level of business practices. Firms that receive a score of below 5 are excluded from the sample (with some offered basic training), while those with scores above 8 that may have less room to improve are also excluded (with some offered consulting services or grants).
 - Have between 2 and 15 workers.
 - Must not already be outsourcing or insourcing *both* marketing and finance functions.
 - Be headquartered (with a physical business location) in Lagos or Abuja.

These recruiting and screening steps generate a sample of high-growth potential firms who: have scope to improve their businesses and increase performance; have already overcome the barrier of hiring at least one paid worker (apart from the owner); and operate in a region where linkages to professional service markets are possible.

The total planned sample size was 2,000 firms. However, obtaining this study sample is dependent on program constraints such as the number of participant firms attending induction workshops, government funding, project timelines, etc. *In practice, the program was part of a much larger World Bank loan to the Government of Nigeria. There were concerns about the performance of other components of this loan package, and considerable uncertainty about whether the program would get extended one year or not. As a result, the program implementors did not induct new batches, because they were unsure of whether there would be time. After over one year of delays, there was a last minute extension, but after more delays, no more firms were able to be added to the experimental sample.*

2.4. Random Assignment

Identification of effects will be by randomized controlled trial. The sample of 753 firms was randomly assigned at the individual level into five experimental groups (each with approximately 150 firms), stratified by induction workshop (denoted induction batch).

After the recruitment and screening steps are completed, firms are randomized by induction batch (i.e., all firms enrolled on a given date in the same city). For example, the set of firms that attended the Lagos induction workshop on a particular date would form an induction batch. Firms, within the same batch, are then randomly assigned in equal proportions (subject to rounding) to the four treatment groups and a control group.² Randomization is done in an office using a computer program (e.g., Stata).

2.5. Balance Checks

² A second sample outside of Lagos and Abuja are randomly assigned to consulting, training, and control only (for a second evaluation that is being used to examine business consulting compared to training).

The key variables used to check baseline balance (across the five experimental groups) are listed below.

- Firm Performance: sales in the last month, sales in the best month, sales in the worst month, sales in an average month, number of full-time employees.
- Business Practices: composite measure of finance practices, composite measure of marketing practices, composite measure of HR practices, overall business practices score out of 10.
- Current use of Business Services: currently insources or outsources accounting services, currently insources or outsources marketing services, has previously used a human resource specialist.
- Business Characteristics: dummies for each of the five GEM sectors (construction, ICT, entertainment, hospitality and tourism, light manufacturing), location (Lagos dummy), age of firm (years), registration with the CAC, percent of sales made in state,
- Owner Demographics: gender, age, marital status education (has completed undergraduate, has completed a masters degree), has previously held a salaried job for 6 or more months.

We will also test whether take-up (of a given intervention) varies according to these variables, and whether there is differential survey non-response according to these variables.

2.6. Key Data Sources

Regarding the outcomes of interest and the data sources, this evaluation will use both Survey Data (collected by independent enumerators using tablets) and Admin Data (information captured by the BIG platform). Indeed, the GEM program is heavily relying on the BIG platform to capture all data on participants, from firm signup and recruitment to intervention monitoring and compliance. For instance, instructors (for the training intervention) need to track firm attendance and completion of course modules. Likewise, business service providers (for the consulting, insourcing and outsourcing interventions) are required to track and enter the progress made by the firms (e.g., attendance, payments, business practices implemented, etc.).

The Survey Data includes the following:

- Baseline Survey: These data are collected as part of the online application and at the induction workshop. This baseline survey will collect information on the background of the firm owner, the current business practices of the firm, sales and profits, main products, and other such information. These data will be used in the screening stage to determine whether firms are eligible for the program, and whether they meet the additional criteria for inclusion in our experimental sample. In addition, baseline variables will be used to check balance (i.e., confirm randomization holds), to examine correlates of take-up, and to control for baseline values of outcomes.

- Follow-up Surveys: We planned to conduct 3 rounds of follow-up surveys, depending on funding availability (which itself depends on the lifecycle of the project).
 - The first follow-up survey took place in April-June 2018 and measures impacts 9-12 months into the program (approximately 1.5 to 2 years after baseline). This survey will mainly focus on intermediate outcomes and changes occurring in the business shortly after the intervention ends.
 - The second follow-up survey is planned to take place a year later (around Spring of 2019) with the aim of measuring impacts about 36 months after the baseline (or at least one year after the intervention ends). *This survey is being conducted in February-April 2019, and will be done as the government program closes. It is therefore the last survey that will be paid for by the government. We then plan on writing up the results after this survey.*
 - A third follow-up survey was tentatively planned for Spring 2020, *but the end of the World Bank loan for this program means that this is no longer funded.*

The Admin Data includes the following (collected through the BIG platform):

- Online Monitoring System: Administrative data on take-up, compliance, and drop-outs will be recorded for each of the four interventions using an online dashboard.
 - The monitoring of firms is done at different stages in the process: (i) by service providers and information reported on the BIG platform; (ii) by government staff and information changed or edited in the BIG platform; (iii) by independent, non-government monitors in charge of supervising all activities (training, consulting, insourcing, outsourcing) and information reported through a private Monitor Dashboard.
- Administrative “Implementation Tracker”: Firms receiving the consulting, insourcing and outsourcing interventions must report on a Firm Owner Dashboard about the Workplan for their staff, hours worked, and main activities undertaken.
 - For example, for the insourcing intervention this dashboard records: whether the firm was emailed about acceptance into the program, whether it then logged into the dashboard, whether it signed an agreement with the HR firm, whether it set up a bank account to receive the subsidy payments, whether it signed the grant agreement, the date of the first disbursement of funds, whether it conducted a needs assessment with the HR firm to decide what type of worker is needed, whether a worker has been hired and a Workplan entered for this worker, the results of phone audits to check the worker is present and what tasks they are doing, dates and amounts of payment, reasons for non-payment, reasons for dropping out in any step of the process, the work schedule of the worker, the main activities carried out by the worker, etc.
- Auditor Visits: Each firm also receives at least two ‘surprise’ visits by auditors who will check whether firms are carrying out the activities recorded on the platform and accurately reporting such activities.

2.7. Wave 1: Group Sizes and Intervention Details

In the 2016 launch, more than 2,500 firms attended the induction workshops in Lagos and Abuja, of whom 943 firms passed these screening criteria and became the first wave of the experimental sample. These are divided into two cohorts: a first cohort of 753 firms that received their baseline surveys between March 2016 and December 2016 and who will be surveyed starting at the end of March 2018, and a second cohort of 190 firms who were inducted in March 2017 and did not start their interventions until late-2017, and so will be surveyed later in 2018. *Unfortunately this second cohort of 190 firms were affected by the uncertainty of the project timeline, and so experienced severe delays and did not receive the intervention as intended. They have therefore been dropped from the study.*

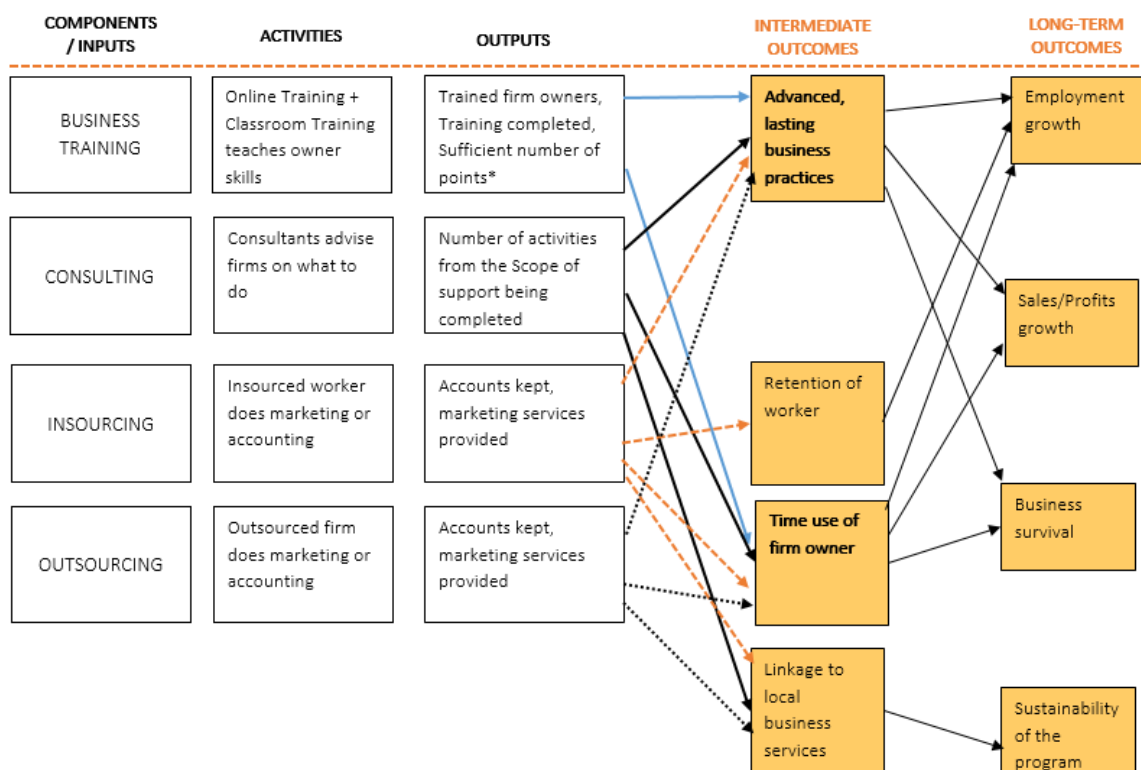
The intervention dates, take-up rates (i.e., percentage completing the first session/visit), and compliance rates (i.e., percentage completing at least half of the intervention) for the first cohort of Wave 1 are also included below.

- (i) Control: 149 firms.
- (ii) Training: 152 firms.
 - Intervention dates: training started April and May 2017, and ended June and July 2017 for most firms in this batch.
 - Take-Up rate: 92.8% started the online training, 73.9% completed at least 5 online classes, and 58.2% completed 12 days of in-class training.
- (iii) Consulting: 149 firms.
 - Intervention dates: *first meetings occurred in January through March 2017 for most firms, and consulting ended January and February 2018 for most firms (with several months of breaks).*
 - Take-Up rate: *We are still trying to get a final version of the administrative data, but our data as of June 2018 show 87.3% completed phase 1 of this intervention, which involved meeting with consultants and setting out a workplan, and 42.6% completed the second phase of all activities.*
- (iv) Insourcing: 152 firms.
 - Intervention dates: began April 2017
 - Take-Up rate: 138 firms hired a worker (90.7%)
- (v) Outsourcing 150 firms.
 - Intervention dates: began April 2017
 - Take-Up rate: 144 firms hired a company (96%)

3. THEORY OF CHANGE

The figure below outlines our proposed theory of change for the four interventions evaluated in this project. In the short-run we expect the interventions to **improve the business practices** in firms and **change the time allocation of the firm owner**. Once the interventions end, we are

then interested in testing whether insourced workers are retained in the firm, and whether firms that are exposed to the marketplace for professional business services continue to use that market. In the long-term, we then will measure whether better business practices and a change in the owner's time management result in improved firm outcomes in terms of business survival, employment, sales, and profits.



There are of course likely to be other impacts on channels between the intermediate outcomes and ultimate outcomes, such as innovation, investment, and the use of finance. We do not expect these to change in the short-run, *and so did not measure those in the one-year follow-up. However, we do measure these effects in the second year follow-up. In addition, we now seek to measure whether there were impacts through another channel, of market expansion.*

3.1. Main Effects (Long-Term Outcomes)

The different interventions propose alternative ways of building skills in the firm to generate high growth and achieve long-term outcomes (e.g., improvements in firm employment, sales, profits and survival).

3.2. Mechanism (Intermediate Outcomes)

Organizational Focus: In terms of intermediate outcomes, each of the interventions is hypothesized to improve business practices in the firm. Inducing firms to focus on improving and adding new business activities can, in turn, lead to greater employment growth, higher

profitability, and greater sales. The pathway from business practices to firm growth will depend on the type of practices employed. For instance, marketing practices are expected to grow the firm through generating new clients and increasing the demand for the firm's products; whereas financial practices are expected to help the firm grow by lowering production costs, identifying which products are most profitable to focus on, and allocating capital more efficiently.

Owner Focus: In addition to operating through better business practices, the interventions may work through other intermediate channels. One such mechanism is through changing the time use of the owner. Although training and consulting may lead the owner to run her business more efficiently, these interventions may also cause the owner to focus more time to practices at an 'operational' level (be it in finance, HR, production, marketing or across all functions), which takes time away from 'strategic' level tasks. In contrast, insourcing or outsourcing some of their business processes could both improve specialization across functional units and free the owner to focus more time on meeting customer demand, product innovations, new market development, strategic planning, and other investments important for high growth. In other words, we expect owners to delegate more and be a "jack-of-all-practices" less as they shift their time from managing many different standardized tasks (operational level) to focusing on a key set of growth related activities (strategic level).

Outward Focus: Third, having an initial experience with a business service marketplace may induce the owner to overcome her inertia (i.e., reluctance due to high costs, information frictions or quality uncertainty) towards hiring an outside company for business skills development. And so, the linkage to the market for local business services – which occurs through the consulting, insourcing and outsourcing interventions – may increase a firm's outward focus and encourage it to return to the same market (or a new one) to procure other business services necessary for growth.

3.3 Intermediate Channels

Changing the business practices, time use, and focus of the owner is hypothesized to enable the business to expand and grow. While we ultimately expect this to show up in the business outcomes of sales, profits, employment and survival, there are several intermediate channels in which this pathway to growth may first show up:

Innovation: *Firms may develop and introduce new products and processes, enabling them to produce their existing products better and diversify their product offerings.*

Investment: *Firm owners may need to increase the capacity for producing more through making large investments in capital, which could be financed by either new loans or equity.*

Market expansion: *Firm owners may diversify their market in order to expand sales, by targeting new geographic regions, new sales channels, or new customer segments.*

4. HYPOTHESES TO BE TESTED

We will test this theory of change through a set of hypotheses, which are presented in groups, summarized in the table below.

A.	Impact on Firm Performance (Long-Term Outcomes): Each of the four interventions may have positive average impacts on business survivorship, employment, profitability, and sales for treated firms. [Main Effects]
B.	Impact on Business Practices (Intermediate Outcomes): Each of the four interventions may lead to greater implementation of finance/accounting practices, marketing/sales practices, HR/operations practices, and overall business practices for treated firms. [Mechanism 1: Organizational Focus]
C.	Impact on Owner Time Use (Intermediate Outcomes): The Insourcing and Outsourcing interventions may lead the owners of treated firms to concentrate their time on fewer functional areas, to focus more on growth related activities, and to delegate more to others. [Mechanism 2: Owner Focus]
D.	Impact on Use of the Market for Professional Business Services (Intermediate Outcomes): The Consulting, Insourcing, and Outsourcing interventions may lead to greater use of the business service marketplaces for treated firms. [Mechanism 3: Outside Focus]
E.	<i>Impact on Innovation, Investment and Market Expansion (Intermediate Channels):</i> Each of the four interventions may lead to firms introducing new products or processes, investing more to build capital stocks, and expanding the range of markets they sell in.
F.	<i>Firm Boundaries and Persistence:</i> Transaction cost and agency theory models of the theory of the firm suggest that asset specificity, trust, the ease of verification, and the ease of enforcing contracts should affect the extent to which firms receiving insourcing and outsourcing continue to use these services after the subsidy, and the relative effectiveness of the two treatments.

We set out specific hypotheses regarding each of these outcome sets, and explain how these outcomes will be measured, noting the specific variables in our follow-up survey that will be used to measure these outcomes.

4.1. Group A: Impact on Firm Performance (Long-Term Outcomes)

Survival

Hypothesis A1: Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to higher survivorship.

This will be measured by the following outcome:

1. Business is still operating at the time of the survey. This will be measured directly for those answering the follow-up survey, and assessed via interviewer observation and third-party reporting for those unable to be re-interviewed. Since all firms were in existence at baseline, the baseline value of the dependent variable will not be included here. [Followup1: A1=1 or Q3=1; *Follow-up 2: Coversheet currently operate a business=1 or Section 1, Q1 = 1, or else attrition section 0 S3=1*]

Employment

Hypothesis A2: Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to an increase in employment.

Hypothesis A3: The Insourcing and Outsourcing interventions will lead to a greater increase in employment (compared to the Control group, and also the Training and Consulting interventions).

This will be measured by the following outcomes:

1. Total number of employees observed working in the firm (by enumerator), winsorized at the 99th percentile. [Followup1: F1; *Follow-up 2: Section 6Q4_1*]
2. Current number of wage or salaried workers, winsorized at the 99th percentile. [Followup1: F4_1; *Follow-up 2: Section 6Q4_1*]
3. Current number of casual or daily workers, winsorized at the 99th percentile. [Followup1: F4_2; *Follow-up 2: Section 6Q4_2*]
4. Current number of partners, winsorized at the 99th percentile. [Followup1: F4_3; *Follow-up 2: Section 6Q4_3*]
5. Current number of apprentices or interns, winsorized at the 99th percentile. [Followup1: F4_4; *Follow-up 2: Section 6Q4_4*]
6. Current number of unpaid workers, winsorized at the 99th percentile. [Followup1: F4_5; *Follow-up 2: Section 6Q4_5*]
7. Total number of workers (sum of all five categories), winsorized at the 99th percentile. [Followup1: F4_6, emp_num (employment reported on cover sheet if F4_6==.; *Follow-up 2: Section 6Q4_6, or, if missing, employment reported on cover sheet*)]
8. Number of hours per week worked in the business by the spouse, winsorized at the 99th percentile [Followup1: E12, coded as zero for those with no spouse; *Follow-up 2: Section 6Q10, coded as zero for those with no spouse*].
9. Firm has 10 or more workers [Followup1: F4_6>=10, or emp_num>=10 for firms missing F4_6; *Follow-up 2: Section 6Q4_6>=10, or, if missing, employment reported on cover sheet>=10*]
10. Number of new paid, full-time employees hired in past six months, winsorized at the 99th percentile. [Followup1: F7a; *Follow-up 2: Section 6 unnumbered question after Q4*]

11. Employment Index, this will be a standardized z-score average of 1-8.
12. *Given the long right tails of employment, we will also look at the inverse hyperbolic sine transform of 1) and 7).*

Hypothesis A4: Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to an increase in wages paid to employees.

This will be measured by the following outcomes:

1. Total amount the business paid in wages, salaries, piece-rates, commissions, and other payments to workers in the last month, winsorized at the 99th percentile. [Followup1: F8; *Follow-up 2: Section 6 Q5*]
2. Average monthly wage/salary for one worker in the last month, winsorized at the 99th percentile. [Followup1: F8 (total salary/wage costs) / F4_6 (total number of employees); *Follow-up 2: Section 6 Q5 (total salary/wage costs)/Section 6Q4 (total number of employees)*]
3. Average monthly wage/salary for the five highest paid workers in the business last month, winsorized at the 99th percentile. [Followup1: average of (E11e_1, E11e_2, E11e_3, E11e_4, E11e_5); *Follow-up 2: average of Section 6Q8e1-e5*]
4. Wage Index, this will be a standardized z-score average of 1-3.

Sales and Profits

Hypothesis A5: Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to an increase in sales and profits.

Hypothesis A6: The Insourcing and Outsourcing interventions will lead to a greater increase in sales and profits (compared to the Control group, and also the Training and Consulting interventions).

The sample is likely to be heterogeneous in terms of sales and profits outcomes. We expect the treatments to be more likely to have a proportional treatment effect than a level treatment effect. For example, we believe it is more likely that a firm earning sales of 200,000 and one earning sales of 2,000,000 will both experience a 10% increase in sales from receiving one of the treatments, than that the treatment will raise sales for both by a constant value. For these two reasons, we use the inverse hyperbolic sine transformation $\log(y+(y^2+1)^{1/2})$ – which is similar to the log transformation, but can deal with zeros – for our primary measures of sales and profits. Even these transformations may not be enough to reduce the influence of massive outliers, and so we apply these transforms after winsorizing the upper tail of the data at the 99th percentile.

This then leads to the following set of outcomes:

1. Inverse hyperbolic Sine of Total sales in the last month, winsorized at the 99th percentile. [Followup1: G16. For firms not answering G16, midpoint of the sales range G16_1, for those answering sales are in range above 10,000,000, sales will be coded as the median of firms with sales in this range, or 12,000,000 if no other firms are in this range; *Follow-up 2: Section 7Q12, same rules as for follow-up 1 for those not answering*]

Note that in addition, follow-up 2, Section 7 Q13a-f asks monthly sales for each of the six months from July 2018 through December 2018. The goal of these measures is to increase power for this measure by providing data on more months, thereby averaging out seasonality. However, we are unsure of whether response rates will be as high for these retrospective measures. We will take the months for which the response rates are no lower than 10 percentage points less than the response rate for Section 7 Q12, and then use these as additional data points (months) for this measure round 2 effect (see methods section). That is, we will stack these responses as additional panel months for this outcome.

2. Inverse hyperbolic Sine of Total sales in the last year, winsorized at the 99th percentile. [Followup1: G17. For firms not answering G17, midpoint of the sales range G17_1, for those answering sales are in range above 10,000,000, sales will be coded as the median of sales for other firms with sales in this range, or 12,000,000 if no other firms are in this range; *Follow-up 2, section 7 Q11, using same rules for range responses as for round 1.*]
3. Inverse hyperbolic Sine of Total profits in the last month, winsorized at the 99th percentile. [Followup1: G23. For those not answering G23, midpoint of the profits range G23_1, for those answering profits are in range above 10,000,000, profits will be coded as the median of profits for other firms with profits in this range, or 12,000,000 if no other firms are in this range; *Follow-up 2: Section 7, Q17, using same rules as round 1 for range responses*]
We also asked in round 2 for monthly profits for October, November, and December 2018. The goal of these measures is to increase power for this measure by providing data on more months, thereby averaging out seasonality. However, we are unsure of whether response rates will be as high for these retrospective measures. We will take the months for which the response rates are no lower than 10 percentage points less than the response rate for Section 7 Q17, and then use these as additional data points (months) for this measure round 2 effect (see methods section). That is, we will stack these responses as additional panel months for this outcome.
4. Inverse hyperbolic Sine of Total profits in the last year, winsorized at the 99th percentile. [Followup1: G24. For those not answering G24, midpoint of the profits range G24_1, for those answering profits are in range above 10,000,000, profits will be coded as the median of profits for other firms with profits in this range, or 12,000,000 if no other firms are in this range; *Follow-up 2: Section 7, Q18; using same rules for range responses as follow-up 1.*]
5. A standardized profits and sales impact will be obtained by aggregating these different effects as described below in our methods section as a standardized z-score.

Measurement of sales and profits in small firms is difficult, and there are several issues that we will need to deal with. Our planned approach to deal with several key issues is as follows:

1. Large heterogeneity across firms: as discussed, our primary approach will be to use a log-like transform. Nevertheless, as a robustness check, we will also examine the impact of treatment on the winsorized levels of total profits in the last month, and total sales in the last month: We also will examine quantile treatment effects on sales and profits, to examine how these change across the distribution.

2. Item non-response: High item non-response rates for profits and/or sales will be addressed in several ways. First, improved training of enumerators will stress techniques for better eliciting such information. Second, the use of netbooks and survey software, rather than paper surveys, in the follow-up round can help prompt better for this information. Third, for those individuals refusing to give point estimates, range questions have been included. For businesses not answering the exact answer, but answer the range question, the midpoint of the range will be used. For firms in the top range, a value equal to the median of firms with sales in this top range will be used. Missing baseline values will be then dummied out in ANCOVA specifications.
3. Dealing with firms which close down: Since treatment may affect both the proportion of firms which shut down, as well as the selectivity of which firms close down, simply comparing profits or sales for surviving firms may be biased. Our primary approach will therefore be to record profits and sales as zero for firms which are closed. If we find no significant impacts of treatment on either the rate or selectivity of survivorship, we will also conduct exploratory analysis which looks at impacts of treatment conditional on surviving.

Hypothesis A7: Each of the four interventions (training, consulting, insourcing, outsourcing) will not affect reporting errors.

There are two concerns here. The first is that with any program involving business skills or improvements in record-keeping (e.g., training, consulting, insourcing, outsourcing), it may lead to changes in the accuracy of the information being reported, even if the firm's underlying financial position does not change. If firms systematically under- or over-state sales and profits, this will lead to a bias in the measured treatment effect. A second concern is that participants may wish to show gratitude for being given training (i.e., exhibit desirability bias) and give biased responses they think that interviewers wish to hear.

We will address this issue in a couple of ways. The first is to test whether the interventions have affected the reporting of firms through estimating the treatment impact on the number of reporting errors made. The following will be deemed a reporting error, and our measure will be the total number of such errors:

- a) Total profits last month exceed total sales last month. [Followup1: (G23) > (G16);
Follow-up 2: Section 7, Q17 > Section 7 Q12]
- b) Total profits last year exceed total sales last year. [Followup1: (G24) > (G17);
Follow-up 2: Section 7 Q18 > Section 7 Q11]
- c) Total sales last month exceed total sales last year. [Followup1: (G16) > (G17);
Follow-up 2: Section 7 Q12 > Section 7 Q11]
- d) Total profits last month exceed total profits last year. [Followup1: (G23) > (G24);
Follow-up 2: Section 7 Q17 > Section 7 Q18]
- e) Total amount spent on wages/salaries last month exceed total sales last month.
[Followup1: (F8) > (G16); *Follow-up 2: Section 6 Q5 > Section 7 Q12*]

Second, as a robustness check, we will examine how sensitive the impact of treatment on the inverse hyperbolic sine of winsorized total profits in the last month, and the inverse hyperbolic sine of winsorized total sales in the last month, is to controlling for the accounting score from our business practices measured in baseline.

4.2. Group B: Impact on Business Practices (Intermediate Outcomes)

Business Practices

Hypothesis B1: Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to an increase in the implementation of business practices.

Hypothesis B2: The Insourcing and Outsourcing interventions will lead to a greater increase in the implementation of business practices (compared to the Control group, and also the Training and Consulting interventions).

This will be measured by the following outcomes:

1. An index of ten **Finance and Accounting practices**, averaging the practices that a firm carries out (coded: '1' if the practice was implemented in the last three months; and '0' if not). This is the average of the following practices:
 - a. Records all money in to the business on a daily or weekly basis (D1A==1|D1A==2; *Follow-up 2: Section 4Q1A==1 or 2*)
 - b. Records all money out of the business on a daily or weekly basis (D2A==1|D2A==2; *Follow-up 2: Section 4Q2A==1 or 2*)
 - c. Prepared an income statement (profit and loss statement) (D5A==1|D5A==2; *Follow-up 2: Section 4Q3A==1 or 2*)
 - d. Prepared a balance sheet (D4A==1|D4A==2; *Follow-up 2: Section 4Q6A==1 or 2*)
 - e. Analyzed a statement of cash flow (D5A==1; *Follow-up 2: Section 4Q7A==1*)
 - f. Analyzed which products/services are the most profitable (D6A==1; *Follow-up 2: Section 4Q8A==1*)
 - g. Used cost-control methods (D7A==1; *Follow-up 2: Section 4Q9A==1*)
 - h. Prepared and used a budget (D8A==1|D8A==2; *Follow-up 2: Section 4Q10A==1 or 2*)
 - i. Set financial goals or financial performance targets (D9A==1; *Follow-up 2: Section 4Q11A==1*)
 - j. Conducted feasibility studies before starting a new venture or investment (D10A==1; *Follow-up 2: Section 4Q12A==1*)
2. An index of nine **Marketing and Sales practices**, averaging the practices that a firm carries out. Each practice is measured using a binary variable (coded: '1' if the

- practice was implemented in the last three months; and '0' if not). This is the average of:
- a. Carried out structured research on customers (D11A==1; *Follow-up 2: Section 4Q13A==1*)
 - b. Carried out structured research on competitors (D12A==1; *Follow-up 2: Section 4Q14A==1*)
 - c. Carried out structured research on market potential (D13A==1; *Follow-up 2: Section 4Q15A==1*)
 - d. Promoted products/services (D14A==1|D14A==2; *Follow-up 2: Section 4Q16A==1 or 2*)
 - e. Changed prices to increase sales or profits (D15A==1; *Follow-up 2: Section 4Q17A==1*)
 - f. Used a branding strategy (D16A==1; *Follow-up 2: Section 4Q18A==1*)
 - g. Use a Customer Relationship Management (CRM) system (D17A==1|D17A==2; *Follow-up 2: Section 4Q19A==1 or 2*)
 - h. Used post-purchase loyalty programs (D18A==1|D18A==2; *Follow-up 2: Section 4Q20A==1 or 2*)
 - i. Has a professional sales force (D19A==1; *Follow-up 2: Section 4Q21A==1*)
3. An index of eleven **Digital Marketing practices**, averaging the practices carried out. Each practice is measured as a binary variable (coded: '1' if the practice was implemented in the last three months, and '0' otherwise). This is the average of the following practices:
- a. Has a functioning website (D20A==1; *Follow-up 2: Section 4Q28A==1*)
 - b. Used search engine marketing tools to improve customer traffic (D21A==1; *Follow-up 2: Section 4Q22A==1*)
 - c. Used e-commerce platform to sell products (D22A==1; *Follow-up 2: Section 4Q23A==1*)
 - d. Has a business facebook page/account (D24A==1; *Follow-up 2: Section 4Q24A==1*)
 - e. Has a business twitter profile (D24C==1; *Follow-up 2: Section 4Q24C==1*)
 - f. Has a business instagram profile (D24E==1; *Follow-up 2: Section 4Q24E==1*)
 - g. Has a business snapchat account (D24G==1; *Follow-up 2: Section 4Q24G==1*)
 - h. Has a business Tumblr account (D24I==1; *Follow-up 2: Section 4Q24I==1*)
 - i. Has a business Whatsapp account (D24K==1; *Follow-up 2: Section 4Q24K==1*)
 - j. Has a business Youtube account (D24M==1; *Follow-up 2: Section 4Q24M==1*)
 - k. Has a business Pinterest account (D24S==1; *Follow-up 2: Section 4Q24S==1*)
4. An index of eleven **Operations and HR practices**, averaging the practices that a firm carries out. Each practice is measured using a binary variable (coded: '1' if the practice was implemented in the last three months; and '0' if not). This is the average of the following practices:

- a. Uses a formal system to carry out maintenance checks (D25A==1; *Follow-up 2: Section 4Q25A==1*)
 - b. Uses a formal system to organize and clean the workplace regularly (D26A==1; *Follow-up 2: Section 4Q26A==1*)
 - c. Uses a formal system to carry out quality inspections (D27A==1; *Follow-up 2: Section 4Q27A==1*)
 - d. Uses electronic payments system in business (D28A==1|D28A==2; *Follow-up 2: Section 4Q30A==1 or 2*)
 - e. Compares the prices and quality of different suppliers (D29A==1; *Follow-up 2: Section 4Q31A==1*)
 - f. Uses a formal system to manage inventory (D30A==1; *Follow-up 2: Section 4Q32A==1*)
 - g. Uses an IT system for enterprise resource planning (ERP) (d5f_b==1; *Follow-up 2: Section 4 d5f_b==1*)
 - h. Uses a formal system to measure individual performance of employees (D31A==1|D31A==2; *Follow-up 2: Section 4Q34A==1 or 2*)
 - i. Provides formal training to employees (D32A==1; *Follow-up 2: Section 4Q35A==1*)
 - j. Uses a formal payroll system (D33A==1; *Follow-up 2: Section 4Q36A==1*)
 - k. Uses financial rewards for top performers (d6c__1==1| d6c__2==1| d6c__3==1; *Follow-up 2: Section 4Q d6c__1==1/ d6c__2==1/ d6c__3==1*)
5. An **overall business practices index of all forty-one business practices**, averaging the binary outcomes, so that this aggregate index ranges between 0 and 1.
 6. An index of the subset of **ten ‘verifiable’ business practices**, averaging the number of such practices that firms carry out. Each practice is measured using a binary variable (coded: ‘1’ if the practice was implemented in the last three months and could be verified by the enumerator; and ‘0’ if not). [Followup1: average of:
 - a. Records all money in to the business on a daily or weekly basis (D1A==1; *Follow-up 2: Section 4Q1A==1*)
 - b. Records all money out of the business on a daily or weekly basis (D2A==1; *Follow-up 2: Section 4Q2A==1*)
 - c. Prepared an income statement (profit and loss statement) (D3A==1; *Follow-up 2: Section 4Q3A==1*)
 - d. Prepared a balance sheet (D4A==1; *Follow-up 2: Section 4Q6A==1*)
 - e. Prepared and used a budget (D8A==1; *Follow-up 2: Section 4Q10A==1*)
 - f. Promoted products/services (D14A==1; *Follow-up 2: Section 4Q16A==1*)
 - g. Use a Customer Relationship Management (CRM) system (D17A==1; *Follow-up 2: Section 4Q19A==1*)
 - h. Used post-purchase loyalty programs (D18A==1; *Follow-up 2: Section 4Q20A==1*)
 - i. Uses electronic payments system in business (D28A==1; *Follow-up 2: Section 4Q30A==1*)
 - j. Uses a formal system to measure individual performance of employees (D31A==1; *Follow-up 2: Section 4Q34A==1*)

(Note that the digital marketing practices are all verifiable as well, apart from using search engine marketing tools to improve customer traffic and using an e-commerce platform. However, we treat these measures as a separate index)

Note that the baseline survey conducted during the induction workshop had a different set of questions on business practices than the more detailed follow-up survey. These questions were used to construct a marketing index, accounting index, and HR index, as well as an overall score out of 10. Ancova estimation of the treatment effects on business practices will control for this baseline score out of 10, except for the marketing and sales practices index above, which will control for the baseline marketing measure; the accounting and finance index above, which will control for the baseline accounting index; and the HR and operations index above, which will control for the baseline HR index. An advantage of Ancova estimation is that it only controls for these proxy measures to the extent that they are correlated with outcomes in the follow-up survey.

4.3. Group C: Impact on Owner Time Use (Intermediate Outcomes)

Amount of Time Owner Spends Working in the Business

Hypothesis C1: The Insourcing and Outsourcing interventions do not cause the firm owners to cut back on the amount of time they devote to their business (compared to the Control group, and to firms receiving Training and Consulting).

This will be measured by the following outcome:

1. Hours worked in the last week, winsorized at the 99th percentile. [Followup1: C1a; *Follow-up 2: Section 4Q1a*]

Time Concentration

Hypothesis C2: The Insourcing and Outsourcing interventions will lead firm owners to concentrate their time on fewer functional areas of the business (compared to the Control group, and also the Training and Consulting interventions).

This will be measured by the following outcomes:

1. A count of the number of different functional areas out of thirteen categories (e.g., production, accounting, marketing, investments, partnerships) in which the owner allocates more than 5 percent of his or her time every two weeks. Each “functional area time” category is measured in total hours spent out of 100 during the previous two weeks. First, each category will be converted into a binary variable (coded: ‘1’ if the number of hours entered was greater than 5; and ‘0’ if not). Second, the thirteen binary variables will be summed up to create a composite variable (ranging from 0 to 13).
 - Lower scores on this composite will represent greater concentration in time use by the owner.

- Higher scores will represent less concentration, or time use that is spread out across more functional areas of the business.
- [Followup1: C3A_1 to C3A_13; *Follow-up 2: Section 3 Question 3A_1-13*]

Our round 1 pre-analysis plan also specified constructed a homogeneity-heterogeneity score based on these responses, using the Gibbs-Martin Formula to measure concentration in time use. However, the round 1 responses showed that owners typically gave responses that were in multiples of 10 percent, suggesting noise in this measure and a likely difficulty in detecting changes. We therefore have dropped this concentration measure, and in round 2, ask owners to directly consider changes:

2. *A count of the number of different functional areas out of ten (10) categories (e.g., production, accounting, marketing, investments, partnerships) in which the owner has decreased the amount of time spent. First, each category is evaluated by the owner in terms of whether he is currently spending more or less time on this function compared to two years ago (measured using a five-point scale: 1 = time spent decreased a lot; 2 = time spent decreased; 3 = no change; 4 = time spent increased; 5 = time spent increased a lot). Second, each category will be converted into a binary variable coded as follows: '1' if the time spent has decreased (scale response = 2) or time spent has decreased a lot (scale response = 1); and '0' if not. Third, the ten binary variables will be summed up to create a composite variable (ranging from 0 to 10).*
 - *Higher scores on this composite will represent greater concentration in time use by the owner.*
 - *[Followup2: Section 3, Q5_1-10]*

Growth-Focused Activities

Hypothesis C3: The Insourcing and Outsourcing interventions will lead firm owners to focus their time on more growth related activities for the business (compared to the Control group, and also the Training and Consulting interventions).

This will be measured by the following outcomes:

1. A count of the number of 'growth-focused' functional areas out of thirteen categories in which the owner spends more than 5 percent of their time every two weeks. Each "functional area time" category is measured in total hours spent out of 100 during the previous two weeks. First, each category will be converted into a binary variable (coded: '1' if the number of hours entered was greater than 5; and '0' if not). Second, the six binary variables representing 'growth-focused' activities will be summed up to create a composite variable (ranging from 0 to 6).
 - Of the thirteen categories, the following *five* can be considered more focused on serving customer demand, expanding sales, and otherwise growing the business:³

³ Our first version of the pre-analysis plan also included Category #1: Production (the buying/sourcing of inputs, manufacturing of goods, making/assembling of products, or delivery of services offered by the business). Upon reflection, we do not consider this as growth-focused necessarily, so have dropped from this measure.

- Category #6: Marketing (the market research, advertising (e.g. tv, radio, flyers), promoting, pricing, branding, or digital marketing (search, e-commerce, social media) of existing products/services).
 - Category #7: Sales (the management of customer relationships, generating leads and closing sales, need/problem evaluation, cold calling, loyalty programs, or professional salesforce plans/pay).
 - Category #8: Innovations (the research and development of new products/services to improve benefits, new processes to increase efficiency, or new market opportunities to grow the business).
 - Category #9: Investments (the pursuit of external funding sources (e.g. loans, grants, equity) to invest in projects that can increase business performance and expansion).
 - Category #10: Partnerships (the development of strategic partnerships (e.g. trade associations, global distributors or suppliers, standard boards, endorsements) to enhance reputation and reach).
 - Higher scores on this composite mean the owner spends more time on growth-focused activities for the business.
 - [Followup1: C3A_1 to C3A_13, which includes 13 categories of time use; *Follow-up 2: Section 3 Question 3A_1-13*]
2. In addition, two other questions will be used to measure the proportion of time that the owner spent on ‘growth-focused’ activities during the previous two weeks. We will take an average of the following two questions to construct this composite measure:
- The percentage of time spent on externally focused activities in the last two weeks. [Followup1: C3B; *Follow-up 2: Section 3Q3B*]
 - The percentage of time spent on future focused activities in the last two weeks. [Followup1: C3C; *Follow-up 2: Section 3Q3C*]
 - Composite = average of (C3B and C3C; *Follow-up 2, average of: Section 3Q3B and : Section 3Q3C*)
3. *A count of the number of growth-focused functional areas in which the owner has increased the amount of time spent. First, each category is evaluated by the owner in terms of whether he is currently spending more or less time on this function compared to two years ago (measured using a five-point scale: 1 = time spent decreased a lot; 2 = time spent decreased; 3 = no change; 4 = time spent increased; 5 = time spent increased a lot). Second, each category will be converted into a binary variable coded as follows: ‘1’ if the time spent has increased (scale response = 4) or time spent has increased a lot (scale response = 5); and ‘0’ if not. Third, the six binary variables will be summed up to create a composite variable (ranging from 0 to 6).*
- *The following five categories can be considered more focused on serving customer demand, expanding sales, and otherwise growing the business:*
 - *Category #6: Marketing (the market research, advertising (e.g. tv, radio, flyers), promoting, pricing, branding, or digital marketing (search, e-commerce, social media) of existing products/services).*
 - *Category #7: Sales (the management of customer relationships, generating leads and closing sales, need/problem evaluation, cold calling, loyalty programs, or professional salesforce plans/pay).*

- *Category #8: Innovations (the research and development of new products/services to improve benefits, new processes to increase efficiency, or new market opportunities to grow the business).*
- *Category #9: Investments (the pursuit of external funding sources (e.g. loans, grants, equity) to invest in projects that can increase business performance and expansion).*
- *Category #10: Partnerships (the development of strategic partnerships (e.g. trade associations, global distributors or suppliers, standard boards, endorsements) to enhance reputation and reach).*
- *Higher scores on this composite mean the owner spends more time on growth-focused activities for the business.*
- *[Followup2: Section 3 Q5_1-10, which includes the five 'growth' focused categories.]*

Delegation of Tasks

Hypothesis C4: The Insourcing and Outsourcing interventions will lead firm owners to delegate more work to others (compared to the Control group, and also the Training and Consulting interventions).

This will be measured by the following outcome:

1. Five questions will be used to measure how often the firm owner has delegated work over the last three months. Each of these questions is measured using an identical five-point scale (where 1 = Never, 2 = Rarely, 3 = Occasionally, 4 = Often, 5 = Always). We will take an average of the following five questions (all focused on delegation of business activities) to construct this composite measure:
 - Let someone else take responsibility. [Followup1: C11A; *Follow-up 2: Section 3, Q11A*]
 - Gave up control over a certain duty. [Followup1: C11B; *Follow-up 2: Section 3, Q11B*]
 - Empowered someone to do more. [Followup1: C11C; *Follow-up 2: Section 3, Q11C*]
 - Delegate a job to a different person or unit. [Followup1: C11D; *Follow-up 2: Section 3, Q11D*]
 - Get someone to help you reach a business goal. [Followup1: C11E; *Follow-up 2: Section 3, Q11E*]
 - Composite = average of (C11A, C11B, C11C, C11D, C11E; *Follow-up 2: average of Section 3, Q11A-Q11E*)

4.4. Group D: Impact on Access to Business Services (Intermediate Outcomes)

Hypothesis D1: The Consulting, Insourcing and Outsourcing interventions will lead firm owners to be more likely to use the market for professional business services than firms in the Training or Control groups

Hypothesis D2: Specific experience matters, so that the insourcing treatment group will be more likely to use Human Resource Services, the outsourcing group more likely to use the services of a professional accounting or marketing firm, and the consulting group more like to use business consultants.

This will be measured by the following outcomes:

1. Use Human Resource Specialists: This will be a binary variable coded as 1 if the owner used an HR specialist to find new employees in the last 12 months, or used an HR consultant to find employees in the last 6 months.

[Followup1: E1_6==1 or d6e2==6; *Follow-up 2: Section 5 Q1_6=1 or Section 4 Qd6e==6*]

2. Uses Outside Accounting Agency at least monthly: This is the combination of a question that asks firms who was the main person or entity that managed the record-keeping and prepared the formal accounts for the business in the last six months. The owner can choose one answer from a list of four options, and then a second question asking how frequently the accountant visits the business or meets with them. We will construct a binary variable that is coded '1' if the owner acknowledges using an outside Accounting agency and meets at least monthly with the firm; and '0' if not. [Followup1: E6==4 & (E8B==4|E8B==5|E8B==6|E8B==7 and period specified less than 1 month); *Follow-up 2: Section 5 Q6==4 & (Section 5 Q8B==4,5,6, or 7 and period specified less than 1 month)*]

3. Uses Outside Marketing Services at least monthly: This is the combination of a question that asks firms who was the main person or entity that managed the marketing and communication activities of the business in the last six months, and a question on frequency of use. The owner can choose one answer from a list of four options. We will construct a binary variable that is coded '1' if the owner acknowledges using an outside Marketing agency at least monthly; and '0' if not. [Followup1: E10==4 & (E12B==4|E12B==5|E12B==6|E12B==7 and period specified less than 1 month); *Follow-up 2: Section 5 Q10=4 & (Section 5 Q12B==4, 5, 6, or 7 and period specified less than 1 month)*]

4. Used a Business Consulting Service in the past year: This is a binary variable that comes from a combination of asking whether the firm has used a business consulting service in the past year, and from a question on how many hours of business consulting they used. We count use as at least 8 or more hours of general business consultants, in order to ensure at least one full day of services were provided. [Followup1: E14==1 & E16>=8; *Follow-up 2: Section 5 Q14==1 & Section 5 Q15>=8*].

5. Professional Business Services Index: this is an aggregate index that is formed by averaging measures 1-4 in order to get an average use of professional services across the four service categories. [Follow-up1 and *Follow-up 2*: average of used an HR firm, used a professional accounting firm, used a professional marketing firm, and used a business consulting service, as defined above].

Group E: Impact on Innovation, Investment and Market Expansion (Intermediate Channels)

These intermediate channels were not measured in follow-up 1, nor at baseline. Hence the pre-intervention baseline values will not be included in the analyses.

Hypothesis E1: *Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to greater **innovation**.*

*This will be measured by an **innovation index** consisting of the simple average of the following 17 innovation components.*

- *Firm introduced a new product or service (IN1=1)*
- *Firm introduced a product or service that is new for its city (IN1=1 and IN3>=2)*
- *Firm significantly improved an existing product or service (IN5=1)*
- *Firm introduced a new or improved process (IN6=1)*
- *Firm implemented new design or packaging [IN8a==1]*
- *Firm Introduced a new channel for selling [IN8b==1]*
- *Firm Introduced a new method of pricing [IN8c==1]*
- *Firm Introduced a new way of promoting [IN8d==1]*
- *Firm Changed or introduced new business processes [IN8e==1]*
- *Firm Changed the way work is organized in your firm [IN8f==1]*
- *Firm Introduced new quality control standards for suppliers [IN8g==1]*
- *Firm Opened a new shop or production location [IN9a==1]*
- *Firm Outsourced a production activity [IN9b==1]*
- *Firm Licensed a new technology [IN9c==1]*
- *Firm Brought in-house a production activity [IN9d==1]*
- *Firm Introduced a new in-house training program [IN9e==1]*
- *Firm Obtained a new quality accreditation [IN9f==1]*

Hypothesis E2: *Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to greater investment.*

*This will be measured by an **investment index** consisting of the simple average of the following three channels for making new capital investments:*

1. *New Loan Financing. This will be measured by asking if the firm obtained any loans in the previous year from a formal source/institution [Followup2: Section 7Q1==1 (yes)].*
2. *New Equity Financing. This will be measured by asking if the firm obtained any equity investments in the previous year [Followup2: Section7Q2==1 (yes)].*
3. *Substantial Firm Investment. This will be measured by asking if the firm invested more than 100,000 Naira into new equipment, technology, buildings, etc. in the previous year [Followup2: Section7Q3==1 (yes)].*

Hypothesis E3: *Each of the four interventions (training, consulting, insourcing, outsourcing) will lead to greater market reach.*

*This will be measured by a **market expansion index** consisting of the average of standardized z-scores of the following three ways of increasing market reach:*

1. *Expanding Geographic Markets.*
 - a. *Number of new geographic markets the firm has expanded to in the last two years, winsorized at the 99th percentile [Section C, QC4]*
 - b. *Percent of sales that come from a different region [Section C, QC6]*
2. *Expanding Sales Channels*
 - a. *Number of new channels the firm used to get its offering to market (Section C, QC7)*
 - b. *Percent of sales that come from these new channels (Section C, QC9)*
3. *Expanding Customer Segments.*
 - a. *Number of new customer segments the firm targets [Section C, QC10]*
 - b. *Percentage of total sales that now come from these new segments [Section C, QC12].*

Group F: Firm Boundaries and Persistence

Based on the work of Williamson, Hart, Holmstrom and Milgrom, a literature on the boundary of the firm suggests the willingness of a firm owner to insource or outsource accounting and marketing activities should depend on how specific these activities are to the firm, how much they trust others, whether they can verify the work done by others, and how easy they find it to enforce contracts. This hypothesis is therefore intended to help directly test those theories, as well as explain any difference in the relative effectiveness of the two treatments.

Hypothesis F1: Persistence of Insourcing: Firms assigned the insourcing treatment are more likely to have kept the worker on when the subsidy ends if:

- *The owner believes marketing and accounting workers can be trusted (Follow-up 2: section 9, Q4=4 or 5)*
- *The owner believes he or she can verify the work of an internal accountant (Follow-up 2: section 9, Q6=4 or 5)*
- *The owner believes he or she can verify the work of an internal marketing worker(Follow-up 2: section 9, Q8=4 or 5)*
- *The owner believes he or she can enforce contracts with internal workers (Follow-up 2: section 9, Q10=4 or 5)*

This hypothesis will be tested by defining a dependent variable equal to 1 if a firm assigned to the insourcing treatment still has this worker working for them at the time of the second follow-up survey (Follow-up 2: section 8, Question 3 = 1), and then estimating a probit of this outcome on the above variables. We will do this with and without the following set of baseline controls for owner and firm characteristics (owner age, owner is female, owner has masters or more education; firm size (number of employees, log of monthly sales), firm sector (dummies for each of the five GEM sectors), firm location (dummy for Abuja), baseline marketing practices score, baseline finance practices score).

Hypothesis F2: Persistence of Outsourcing: Firms assigned the outsourcing treatment are more likely to have kept using the provider when the subsidy ends if:

- *The way the firm generates and spends cash is not very firm-specific (Follow-up 2: section 9, Q1 = 1)*
- *The type of customers and customer acquisition strategies are not very firm-specific (Follow-up 2: section 9, Q2=1)*
- *The owner believes marketing and accounting business service providers can be trusted (Follow-up 2: section 9, Q3=4 or 5)*
- *The owner believes he or she can verify the work of an outside accountant (Follow-up 2: section 9, Q5=4 or 5)*
- *The owner believes he or she can verify the work of an outside marketing firm (Follow-up 2: section 9, Q7=4 or 5)*
- *The owner believes he or she can enforce contracts with outside business service providers (Follow-up 2: section 9, Q9=4 or 5)*

This hypothesis will be tested by defining a dependent variable equal to 1 if a firm assigned to the outsourcing treatment still has this provider working for them at the time of the second follow-up survey (Follow-up 2: section 8, Question 12 = 1), and then estimating a probit of this outcome on the above variables. We will do this with and without the following set of baseline controls for owner and firm characteristics (owner age, owner is female, owner has masters or more education; firm size (number of employees, log of monthly sales), firm sector (dummies for each of the five GEM sectors), firm location (dummy for Abuja), baseline marketing practices score, baseline finance practices score).

Hypothesis F3: Insourcing will lead to more of a persistent improvement in business practices than outsourcing the higher the transaction costs and agency costs of using outsiders

*We will form an **inside vs outside transaction gap index** as an average of standardized z-scores of the following variables:*

- **Accounting specificity:** *coded as 1 if the way the firm generates and spends cash is very firm-specific (Follow-up 2: section 9, Q1 = 2, and 0 otherwise)*
- **Marketing specificity:** *coded as 1 if the type of customers of the firm and the way it acquires them is very firm-specific (Follow-up 2: section 9, Q2=2, and 0 otherwise)*
- **Inside-Outside Trust Gap:** *the difference between how much they think they can trust inside workers and outside workers (Follow-up 2: section 9, Q4 minus Q3)*
- **Inside-Outside Accounting Verifiability Gap:** *the difference between how much they can verify the work of inside accountants vs outside accountants (Follow-up 2: section 9: Q6 minus Q5)*
- **Inside-Outside Marketing Verifiability Gap:** *the difference between how much they can verify the work of inside marketers vs outside marketing firms (Follow-up 2: section 9, Q8 minus Q7)*
- **Inside-Outside Enforceability Gap:** *the difference between how much they can enforce contracts with inside accountants and marketers compared to outside service providers (Follow-up 2: section 9, Q10 minus Q9)*

We will then use just the control, insourcing, and outsourcing treatment groups, and take the business practices outcomes defined in hypothesis B2. We will then use the round 2 follow-up data to estimate:

$$Y_{i,t} = \beta_0 + \beta_1 \text{Insource}_i + \beta_2 \text{Outsource}_i + \delta_1 \text{Insource}_i * \text{TransGap}_i + \delta_2 \text{Outsource}_i * \text{TransGap}_i + \delta_3 * \text{TransGap}_i + \pi Y_{i,t=0} + \gamma M_{i,t=0} + X'_{k,i} \theta + \varepsilon_{i,t}$$

Where *TransGap* is the inside-outside transaction gap index, and the treatment dummies, baseline outcome, and strata controls are defined as in equation (M1) below. We are then interested in testing whether there is treatment heterogeneity by this gap index ($\delta_1 = 0$ and $\delta_2 = 0$) and whether the higher this index, the more effective is insourcing relative to outsourcing ($\delta_1 > \delta_2$).

A key caveat to note in this analysis is that the measures of transaction specificity, verifiability, trust, and enforcement are all measured after the treatment. A potential concern is then that firm owners learn about these aspects through their experiences with treatments. To address this concern, we will test for treatment effects in these variables by comparing the responses of the insourcing and outsourcing groups to those of the control. We will then examine robustness to only including those variables that do not change with treatment.

5. ESTIMATION METHODOLOGY

5.1 Estimation of Treatment Effects

We will estimate the intention-to-treat effect of being offered each of the four treatments through the following ANCOVA specification for firm i in follow-up survey round t

$$Y_{i,t} = \beta_0 + \beta_1 \text{Insource}_i + \beta_2 \text{Outsource}_i + \beta_3 \text{Train}_i + \beta_4 \text{Consult}_i + \pi Y_{i,t=0} + \gamma M_{i,t=0} + X'_{k,i} \theta + \varepsilon_{i,t} \quad (\text{M1})$$

Where $Y_{i,t}$ is the given outcome variable measured post-treatment, $Y_{i,t=0}$ is its baseline value (where available) and $M_{i,t=0}$ a dummy variable indicating whether or not this baseline value is missing for a particular individual, given that it was measured for some individuals; *Insource*, *Outsource*, *Train*, and *Consult* are dummy variables for being assigned to treatments of insourcing, outsourcing, training and consulting respectively; X_k is a vector of randomization strata dummy variables (induction batches) and $\varepsilon_{i,j,t}$ is the error term. β_1 , β_2 , β_3 , and β_4 will then provide the intent-to-treat effects, which are the effects of being assigned to a specific treatment relative to being a firm in the control. Note that the randomization strata dummy variables will also control for different cohorts being surveyed at different points in time, so long as induction batches are all surveyed at once. Since random assignment is at the individual level, robust (Eicker-White) standard errors will be used.

Since the batch-based random assignment precluded us from stratifying on covariates, equation (M1) does not include any controls for baseline variables other than the baseline outcome. However, including controls can boost our power, which is particularly important given the new

smaller sample. We will therefore use the double-lasso method of Belloni et al. (2014) to select the control variables to include, using the list of pre-specified baseline controls in section 2.5. These controls will also help correct for any non-random attrition related to these baseline variables.

In addition, the LATE/TOT will be estimated by instrumenting receipt of these different treatments with their random assignment. Take-up for this case will be defined as follows: for insourcing and outsourcing, using at least one month of the services; for training, completing at least 1 in-person class; for consulting, receiving at least one visit from the BDS consultant.

Key hypotheses to then be tested are i) no treatment has any impact ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$); ii) the treatments all perform equally well ($\beta_1 = \beta_2 = \beta_3 = \beta_4$); and iii) the new interventions outperform the traditional training and consulting approaches: ($\beta_1 > \beta_3, \beta_2 > \beta_3$; and $\beta_1 > \beta_4, \beta_2 > \beta_4$). Since randomization is at the individual level, standard errors will only be clustered when multiple follow-up rounds of data are used (if this is done for sales and profits), in which case clustering will be done at the firm level.

5.2 Procedure for Utilizing Multiple Measurement

The specifications above are intended for use with the first follow-up survey. When multiple measures of follow-up outcomes are available after additional surveys are taken, we will test to see whether the treatment effects vary with time. If we cannot reject equality of treatment effects over time, the multiple measurements will be combined to increase power following the procedures set out in McKenzie (2012). *For profits and sales, this will also mean stacking the monthly data collected in follow-up 2 so that the panel includes multiple months for this outcome, and then pooling together the months around follow-up 2 to get an average effect.*

5.3 Missing data from item non-response

The procedure for handling missing data will depend on the reason the data are missing. If data on firm outcomes are missing because the firm has closed down, then employment, sales, and profits will be coded as zero. Similarly, the firm owner will be coded as allocating no time to any business tasks, and using no business practices, and not using the market for professional services. If the firm shuts down and a new one is started, outcomes for the new firm will be used where available. If data are missing because of item non-response, then if item non-response rates are less than 10 percent, we will assume data are missing-at-random and not carry out any imputation. If item non-response is higher than 10 percent, we will check to see if the rate of refusal is balanced by treatment status, and whether baseline characteristics of those who respond are balanced by treatment. If the differences are small by treatment status (less than 5 percentage points), we will operate under a missing-at-random assumption, and then examine robustness to other assumptions. If data are missing in a way that is strongly correlated with observables, we will use baseline data to impute the missing data, and again examine robustness to other alternatives.

5.4 Procedures for Dealing with Testing for Multiple Outcomes

Three methods will be used to address multiple hypothesis testing.

1) Using index measures to form a restricted set of primary outcomes:

a. In the short-term (first follow-up) survey, the most important outcomes are the intermediate channels. The primary outcomes are then: i) **overall business practices index of all forty-one business practices; ii) time concentration (composite score out of 13 of number of areas in which at least 5 percent of time is spent); iii) task delegation index of the average of the five delegation questions; and iv) professional business services index.**

b. In the longer-term (second and later follow-up surveys), in addition to these intermediate channels, our primary outcomes are the key firm growth measures of i) **firm survival; ii) standardized z-score index of firm employment; iii) binary variable of whether firm has 10 or more workers; and iv) standardized z-score index of business profits and sales.**

2) Since we have four treatments, we will use F-tests to test for equality of the treatments, and that all treatments are jointly zero.

3) When it comes to testing impacts on individual outcome measures, we will calculate both single-estimate p-values that can be used when comparing results for this outcome to those for the same treatment and outcome in other studies, as well as sharpened q-values that hold constant the false discovery rates when testing multiple treatments against multiple outcomes.

5.5 Survey attrition

We will test for non-random attrition and non-random survey item non-response based on the baseline variables specified above for balance testing, and by treatment status. If we observe significant differences by groups, we will employ several approaches to examine the robustness of our results to this, following the methodology used in McKenzie (2017). These will include: i) Lee bounds to account for differential attrition; ii) imputation of the missing observations; and iii) Behagel et al. (2015) bounds which use the amount of effort required to reach firms.

5.6 Inflation

Once multiple rounds of survey data are available, nominal values will be converted to real Naira using the monthly urban consumer price index from the National Bureau of Statistics. If these price indices are available for Lagos and Abuja separately, then firms will have their monetary values converted using the price index corresponding to their location at the time of the baseline survey.

5.7 Questions with Limited Variation

In order to limit noise caused by variables where there is no room for a treatment effect to lead to any improvement, questions for which 95 percent or more of observations in the control group have the maximum value of the variable will be omitted from the construction of any aggregate index. For example, if there is a business practice that 99 percent of the control group is already doing, it will be omitted from construction of the business practices aggregate index.

5.8 Spillovers

The number of firms in the study (753) is small relative to the universe of firms in Lagos and Abuja, and as a result, we expect spillovers to be limited given the size of these cities. However, as a robustness check, we will geo-locate each firm, and then examine whether outcomes vary with the number of other treated firms in the same sector located within 5 kilometers of the firm

(constructing four variables, one for each treatment), conditional on the total number of sample firms in the same sector within this radius (a fifth variable). These five variables will be added to equation (M1) and tested for the primary outcomes.

6. POWER CALCULATIONS

These power calculations have been updated to reflect the sample size of 753 firms, with 150 firms in each treatment arm. We stick with the assumption of 90% take-up, and conservatively assume that take-up is uncorrelated with any treatment effect heterogeneity, so that the $ITT=0.9 \cdot TOT$. We report the minimum detectable TOT effect assuming this take-up rate. The calculations now use the control group mean and standard deviations from follow-up 1. These calculations do not account for any gains in power from stratification on randomization batch.

Table 1: Minimum Detectable Effect (80% power, 5% significance level, 90% take-up)

	Business Practices Score	Employment	Log Profits
Unit Level	Firm	Firm	Firm
Variable Mean	0.47	8.95	11.26
Variable SD	0.17	8.69	4.90
MDE in units	0.08	2.53	1.74
MDE (in SD)	0.48	0.29	0.35
MDE (as % of reference mean)	17%	28.3%	15%

Notes:

Business practices: this calculation is made using only a single round of follow-up data, and assumes autocorrelation is 0.25 with the baseline measure, as is the case for the control group. MDE for the ITT is then 0.074, so for the TOT is $1.1 \cdot 0.074 = 0.084$. This is 0.48 s.d., or a 17% increase in the mean.⁴ With two rounds of pooled follow-up, assuming the autocorrelation between round 1 and round 2 is 0.5, then the MDE falls to 0.046 for the ITT, or 0.051 for the TOT.

Employment: power calculations assume Ancova estimation, controlling for baseline value of employment. Autocorrelation is assumed to be 0.58. MDE is then 2.3 workers⁵, so a TOT of $1.1 \cdot 2.3 = 2.53$ workers. This is 0.18 S.d., or a 4.3% increase on the control mean. For employment we believe the increase in the number of workers is the most useful metric. Using

⁴ sampsi 0.47 0.524, sd1(0.17) n1(150) n2(150) pre(1) post(1) r01(0.24)

⁵ sampsi 8.95 11.25, sd1(8.69) n1(150) n2(150) pre(1) post(1) r01(0.58)

two rounds of pooled follow-up, assuming the autocorrelation between round 1 and round 2 is 0.5, then the MDE falls to 1.8 workers (ITT), or 1.98 workers for the TOT.

Inverse hyperbolic sine of profits: power calculations are based on post-estimation only, since profits were not collected on the baseline application form. MDE is then 1.58 for the ITT, or 1.74 for the TOT. This is a 0.35 S.D. effect, which is 15% of mean IHS of profits, but corresponds to a large (>100%) increase in the level of profits. Adding a second follow-up and pooled estimation, assuming an autocorrelation of 0.5, reduces this MDE for the ITT to 1.37, while using 5 follow-up months pooled (by using the three additional monthly measures) reduces it further to 1.24.

REFERENCES

- Anderson, Stephen J., Rajesh Chandy and Bilal Zia (2016) “Pathways to Profits: Identifying Separate Channels of Small Firm Growth through Business Training”, *World Bank Policy Research Working Paper* no. 7774.
- Anderson, Stephen J. and Bilal Zia (2017) "Measuring the Unmeasured: Combining Technology with Survey Design to Filter Noise in Self-Reported Business Outcomes" Manuscript.
- Behaghel, Luc, Bruno Crépon, Marc Gurgand, Thomas Le Barbanchon (2015) “Please Call Again: Correcting Non-Response Bias in Treatment Effect Models”, *Review of Economics and Statistics*, 97(5): 1070-1080
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014) “High-Dimensional Methods and Inference on Structural and Treatment Effects.” *Journal of Economic Perspectives* 28 (2): 29-50.
- Bloom, Nicolas, Benn Eifert, Aprajit Mahajan, David McKenzie and John Roberts (2013) “Does Management Matter? Evidence from India”, *Quarterly Journal of Economics*, 128(1): 1-51.
- Bruhn, Miriam, Dean Karlan and Antoinette Schoar (2017) “The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico”, *Journal of Political Economy*, forthcoming.
- McKenzie, David (2017) “Identifying and Spurring High Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition”, *American Economic Review*, 107(8): 2278-2307.
- McKenzie, David (2012) “Beyond Baseline and Follow-up: The Case for more T in Experiments”, *Journal of Development Economics*, 99(2): 210-21
- McKenzie, David and Christopher Woodruff (2014) “What are we learning from business training evaluations around the developing world? “, *World Bank Research Observer*, 29(1): 48-82.
- McKenzie, David and Christopher Woodruff (2017) “Business Practices in Small Firms in Developing Countries”, *Management Science*, 63(9): 2967-81.

