How can we learn whether firm policies are working in Africa?

Challenges (and solutions?) for experiments and structural models[*]

David McKenzie, *World Bank*

Abstract

Firm productivity is low in African countries, prompting governments to try a number of active policies to improve it. Yet despite the millions of dollars spent on these policies, we are far from a situation where we know whether many of them are yielding the desired payoffs. This article establishes some basic facts about the number and heterogeneity of firms in different sub-Saharan African countries and discusses their implications for experimental and structural approaches towards trying to estimate firm policy impacts. It shows that the typical firm program such as a matching grant scheme or business training program involves only 100 to 300 firms, which are often very heterogeneous in terms of employment and sales levels. As a result, standard experimental designs will lack any power to detect reasonable sized treatment impacts, while structural models which assume common production technologies and few missing markets will be ill-suited to capture the key constraints firms face. Nevertheless, I suggest a way forward which involves focusing on a more homogeneous sub-sample of firms and collecting a lot more data on them than is typically collected.

Keywords: Firm productivity; Randomized Experiments; Structural Models; Impact Evaluation

JEL codes: O12, N67, D22, C93

1. Introduction

The last decade was a good one for growth in many African countries, as a reduction in conflict, improved political and macroeconomic stability, booming commodity prices, and a number of microeconomic reforms led to a growth surge (Radelet, 2010). The McKinsey Global Institute (2010) calculates that labor productivity grew annually at 2.7 percent in Africa during the 2000s, compared to negative labor productivity growth in the 1980s and 1990s. Yet despite this reversal, productivity remains low in many firms. Van Dijk (2003) finds that labor productivity in South Africa, the most advanced economy in sub-Saharan Africa, was only 20 percent of U.S. levels in 1999. Likewise, Harrison et al. (2010) find labor productivity in sub-Saharan African countries during the mid-2000s to be 36 percent lower than performance in the top half of the set of non-African countries with income levels below $3000 in PPP terms. Increasing this productivity is vital for long-run growth prospects and generating the jobs needed to employ Africa's young and rapidly growing labor force.

Private sector development policy agendas of most governments in the region pursue two interrelated approaches to spurring private sector growth and increasing firm productivity. The first is to provide a better enabling-environment for businesses through maintaining macroeconomic stability, reducing administrative and regulatory barriers, and investing in infrastructure. The second, and more activist, policy approach is to enact programs which work directly with micro, small, medium and large firms to enhance their growth and create employment. Examples of these programs include matching grants, business trainings, partial credit guarantees, and wage subsidies. Despite, the millions of dollars invested in such programs, to date there is little rigorous evidence as to their effectiveness. The purpose of this paper is to discuss the challenges and potential for experimental and structural methods to help policymakers in understanding whether or not these policies are working and why.

Recently there has been considerable debate in the broader development literature about whether the profession is overemphasizing randomization (Rodrik 2009; Deaton 2010; Ravallion 2009; Imbens 2010). I agree with Imbens that, given the question which one is interested in answering is possible to answer with randomization, there is little to gain and much to lose by not randomizing. I also agree that there are many important policy questions we wish to answer that it may not be possible to randomize for, and that economics still has something to offer in answering those questions. Nevertheless, it is worth pointing out that when it comes to rigorous evaluation of private sector policies in Africa, there has not been much in the way of either experimental or structural work – so there is plenty of scope for both types going forward – and hopefully this paper will help researchers considering either approach.

I begin section 2 by establishing a number of facts about the private sector and the firms participating in private sector programs in different African countries. Most importantly, the number of medium and large firms is relatively small in most countries, as are the number of firms participating in any given firm-level intervention. Moreover, these firms are quite

heterogeneous in terms of employment size, industry, and sales levels. Section 3 examines the implications of these facts for the power of experiments to detect impacts of firm policies, and for the ability of structural models to credibly estimate firm behaviors. The number of firms and heterogeneity of these firms poses challenges for both approaches. In section 4, I then discuss a way forward, which involves focusing attention on a smaller, more homogeneous sample of firms and collecting a lot more data on these firms. Section 5 then concludes.

2. Background on Small, Medium and Large Enterprises in Africa

Before discussing which methods of evaluation and estimation are likely to be most appropriate for assessing policies directed at firms in Africa, it is useful to establish some facts about the likely populations of interest for such studies.

2.1 Fact 1: The number of SMEs and large firms in many African countries is relatively small.

I begin by first asking how many firms there are of different sizes, as collected through industrial censuses in selected countries. These censuses differ somewhat in their coverage: some only cover manufacturing and some exhibit an incomplete coverage of home-based enterprises. Nevertheless, they give a sense of the scale of small, medium, and large scale enterprise activity.

Table 1 shows that there are many firms with fewer than 10 workers in most African countries, and there are more of such firms in wholesale and retail trade than in manufacturing. Within the category of firms with fewer than 10 workers, the vast majority have at most one or two workers. For example, in Mauritius, out of the 91,980 units with 9 or fewer workers, 70.6 percent have only one or two workers, and only 6.2 percent have five to nine workers.[1] Likewise, among manufacturing firms in Tanzania, only 8.9 percent of firms with less than 10 workers have 5 to 9 workers, whereas 60.8 percent have 1 or 2 workers. Based on Table 1, these numbers imply populations of around 770 to 2150 manufacturing firms with 5 to 9 workers. This is about the same size as the number of manufacturing firms with 10 or more workers, which is in the 1000 to 2000 firm range for most countries in the table. Once one starts getting towards large firms, with 100 or more workers, the population sizes in most African countries get quite small – there are 80 of such firms in Tanzania, 114 in Uganda, and 251 in Ghana according to this data. Some policies are targeted just at exporting firms. Mauritius is the only one of the countries in Table 1 for which data on the number of export-oriented firms is available, reporting 386 such firms in 2010.[2]

Second, note that there is considerable heterogeneity in what firms produce within the broad category of manufacturing. For example, of the 2,203 Ethiopian manufacturing firms with 10 or

---

[1] http://www.gov.mu/portal/goc/cso/ei717/toc.htm
[2] http://www.gov.mu/portal/goc/cso/ei869/toc.htm. Export-oriented firms are those with an export processing zone certificate and those manufacturing goods for export.

more employees, 562 were involved in the manufacture of food products (itself a heterogeneous category), 127 in paper and paper products, 363 in furniture, 47 in textiles, 87 in rubber and rubber products, 75 in chemical and chemical products, and the remainder in a variety of other products. As a consequence, the number of medium and large firms in most African countries in any given manufacturing sector is likely to be relatively small.

2.2 Fact 2: There is considerable heterogeneity in firm performance among these firms

There is also tremendous heterogeneity in firm outcomes. Table 2 uses the World Bank Enterprise Surveys to illustrate the variation among firms in three key outcomes that policy interventions often try to affect: employment, firm sales, and whether or not firms export. I break the data down into three size categories: fewer than 10 workers, 10-100 workers, and more than 100 workers.[3] Few of the Enterprise Surveys contain data on firms with less than 5 workers, so the first size category should be considered as encompassing small firms with 5 to 9 workers. First, consider employment. We see that for firms with 5 to 9 workers, the standard deviation of employment is about one-third of the mean. This coefficient of variation (the standard deviation divided by the mean), increases to 0.72 on average for firms with 10 to 100 workers and 1.25 on average for firms with more than 100 workers. The heterogeneity is much larger with regard to revenues – on average the cross-sectional standard deviation is 3.1 times the mean for firms with 5-9 workers, and 4.2 times the mean for firms with 10 to 100 workers. The potential for a large outlier to increase this dramatically is seen in a couple of cases where the coefficient of variation exceeds 10. Finally, only a minority of firms export, with quite a lot of variation across countries in this percentage.

2.3 Firms participating in SME projects are typically also relatively small in number and very heterogeneous.

Last, it is useful to get a sense of the number of firms actually involved in major policy efforts to actively benefit firms in Africa, and what these policy efforts are. A typical World Bank loan to enhance private sector competitiveness in Africa has two main elements: a component desired to improve the business climate and reduce regulatory burden and a component which actively tries to enhance the productivity of firms through policies which work at the firm level. One of the most common forms of actively helping firms is through matching grant programs, whereby the government reimburses firms for 50 percent of the cost of business services like hiring a consultant, launching a marketing campaign, training workers, or attending a trade fair. The typical justification for such subsidies is a belief that firms underinvest in these services because of externalities to other firms: for example, firms might hesitate to train workers if there is a chance that the workers will then leave and start their own firms or go to work for competitors. A few of these programs focus solely on the export sector, with the aim of getting firms to

---

[3] All numbers reported are for the unweighted data, to illustrate the variation among the sample of firms. This illustrates the variation in the actual data one might have in using the data that is currently regularly collected on African firms; using the survey weights also gives large heterogeneity.

diversify into other markets or other products, but most of these matching grant programs are open to a wide variety of sectors.

Table 3 uses data from World Bank project completion reports to highlight the number of firms participating in these projects. A typical project seems to involve giving out matching grants totaling $1-5 million to between 100 and 500 firms, and has a duration of around 5 years. Projects focusing on exports have fewer firms: with a range of 13 to 149 firms seen in the data. Within some of these programs there can be considerable heterogeneity in both the size of the intervention, and in the types of firms receiving them. For example, the Zambia matching grant program gave out 63 grants, of which 23 were for less than $5000, 33 between $5,000 and $50,000, and 7 above $50,000; 39 percent of the firms were in manufacturing, 30 percent in services, 14 percent in agriculture, 6 percent in tourism, and the rest in other sectors.[4] In other programs the range of the grants offered can be smaller, but there is usually still large heterogeneity in the sectors. For example, the matching grants under the South Africa Black Business Supplier development project were capped at US$17,000, and restricted to firms with fewer than 20 employees. Even given these restrictions which make the firms more homogenous, a survey of 50 participating enterprises showed full-time employment with a mean of 9.9, standard deviation of 8.0, and 2004 revenues with a mean of 983,000 rand and standard deviation of 2,726,000 rand: that is, like the enterprise survey data, the standard deviation of sales is several times its mean. The Kenya project listed in Table 3 is notable as an exception, and an illustration that projects focused on microenterprises may serve many more firms than those focusing on SMEs.

Table 4 summarizes a half-dozen projects on SMEs being undertaken by the International Finance Corporation (IFC) in Africa. These projects are typically directed at firms in a particular sector, such as firms in, or linked to, the mining industry. As a result, the firms in these projects appear to be typically smaller in number, but more homogeneous, than those in the World Bank projects. In addition, the IFC offers business training through its *Business Edge* program[5] in a number of African countries. Table 4 shows that numbers of firms participating in these programs in the last year are small in most countries (e.g. 16 in South Africa, 19 in Senegal), but in a couple of countries get up to a size of around 300-400 (Nigeria and Kenya).

Numbers are also often small for other organizations working to generate firm growth in African countries. Technoserve is one of the better known NGOs, which runs business plan competitions in different countries. In 2006 and 2007 they trained 120 people in Ghana, with 11 businesses launched; in Kenya in 2007 and 2008 they had 38 businesses launched, and in Swaziland between 2006 and 2009 they had 87 businesses launched.

---

[4] Source: World Bank Implementation Completion Report.
[5] http://www.ifc.org/ifcext/africa.nsf/Content/SMED_BusinessEdge. In many of these training sessions there are several individuals per firm being trained, so the total number of people trained exceeds the number of firms trained. But for the purpose of assessing firm outcomes, the number of firms trained is the more relevant metric.

Finally, it is worth keeping in mind that *none* of these projects has been subject to a rigorous experimental evaluation *or* to a rigorous structural evaluation to date. Although a couple of experimental evaluations are in the early stages for matching grant projects, this lack of evaluation to date is consistent with the general lack of rigorous evaluation of SME projects in most areas of the World (McKenzie, 2010), and is indicative of the practical and political challenges one will face in doing evaluations. This is especially the case in experimental studies where the thought of randomly deciding which firms get a program is often anathema to self-professed experts who believe that they know how to decide which firms would most benefit from a program (despite any evidence to support this I would argue), and to politicians who sometimes view the benefits of such programs as rents to be dished out to favored firms. Thus whilst the remainder of the paper concentrates on what is technically feasible to do, it also requires considerable effort, skill, and luck on the part of researchers to make it also politically feasible.

3. Implications of these Facts for Evaluation of SME policies

These facts pose several large challenges for both experimental and structural modeling approaches to determining whether development interventions involving firms meet their goals. Typical goals of these projects are to increase firm sales and/or employment. For example, the 2009 Mauritius Manufacturing and Services Development and Competitiveness Project has a goal of a 20 percent increase in sales over three years from its matching grant program; and the 2005 Tanzania Private Sector Competitiveness Project has a goal of a 10 percent increase in sales after two years, and 15 percent increase in sales after five years for firms supported by the project.

3.1 Implications and Challenges for Experimental Design

The first challenge any experimental evaluation also faces is convincing policymakers of the need for an experiment. Given the relatively large amounts of money being spent on these SME programs (Table 2), the lack of existing rigorous evidence for the impact of such programs (McKenzie, 2010), and the potential for such programs to be seen as merely subsidizing wealthy firms with little benefits for the remainder of the population, there is certainly a strong case to be made for experimentation and testing whether the programs work. The toolkit of Duflo et al. (2008) discusses a number of strategies for implementing an experiment, and along with their advantages and disadvantages. Four such strategies are over-subscription, randomized phase-in, randomization among marginal applicants (see also Karlan and Zinman, 2010), and encouragement designs.

The best case scenario is to be able to implement an over-subscription design. In such a case, there is excess demand for the program beyond what project resources can support. Then a fair and equitable way of allocating the program's scarce resources can be through randomly choosing which firms will participate in the program. Based on Table 1, an over-subscription

design will be a lot easier to implement for projects focusing on microenterprises or small enterprises than SMEs because there are simply many more micro and small enterprises. In fact, some SME projects have struggled to disburse all available funding and achieve their target number of participating firms. In the Zambia project listed in Table 3, only 45 percent of available funding for the matching grant program was disbursed. The project review concluded that this was in part due to a lack of corporate depth in Zambia, with an insufficient number of companies to meet targeted project numbers.

A randomized phase-in design can help in this situation. Under such a design, some firms are randomly selected to receive the project first, while others remain eligible to receive it in later periods. Oftentimes, projects lack the capacity to serve everyone at the same time. From an operational standpoint, this approach then allows the project to learn as it goes and does not require turning down interested firms. The downside of the phase-in design is that its potential to reveal a program's impact is limited to the short-term: the time until other firms are brought into the program. Furthermore, if firms know for sure that they will get the project in the future, then firms might change their current behavior. In order to avoid creating expectations and changing behavior, it is preferable to tell firms they can re-apply for future rounds, rather than guarantee that they will get the program in the future.

Randomization among marginal applicants involves program implementers selecting firms that they identify as top priorities (e.g. perhaps the most likely to grow firms, firms in specific target groups or the best credit risks), rejecting firms deemed well below the eligibility bar, and then randomly select which of the remaining firms can participate. Given the tremendous uncertainty which surrounds efforts to ex ante determine credit risk or firms' growth prospects, it is realistic to think that a large fraction of firms applying for loans, credit guarantees, or matching grant schemes may fall into this "marginal" category. In this case, the treatment parameter will be the effect of the program on marginal candidates, which, although it doesn't tell policymakers the overall impact of their program, is useful for deciding whether to expand access to more or less firms.

Under an encouragement design, the program is open to all firms, but some firms are randomly chosen to be "encouraged" to participate, perhaps through targeted visits and marketing or hand-holding efforts to help firms apply for the program. Such designs typically require a large number of eligible firms and very effective encouragement strategies for success. This is likely to be difficult to achieve given the firm populations in many African countries and the lack of proven ways to dramatically increase program participation.

An additional implication of Table 3 is that with an average duration of 5 years, evaluation of these types of firm programs by means of randomized experiments will take years to show results. Note however, that this does not mean that research will only be able to show impacts 5 years later – ideally results from experimentally implementing the project in the first year or two of the project can guide how the project is implemented in the remaining years.

3.2 Implications and Challenges for Experimental Power

The power of a statistical test is the probability that it will reject a null hypothesis given that the null hypothesis is false. A starting point in most experiments is to test the null hypothesis that the intervention had no effect, so the power of the experiment is a measure of the ability to detect an effect of a given policy intervention if such an effect does exist. Suppose an oversubscription design has been implemented. Then, the small number of firms in many programs and the tremendous heterogeneity amongst them will still pose a severe power challenge for randomized experiments. To illustrate this, Table 5 presents power calculations for a hypothetical experiment on firms exhibiting the same degree of heterogeneity as witnessed in the data in Table 2. To take a best case scenario, we assume 100 percent compliance with treatment allocation, so that all firms randomly assigned to participate in a program actually do participate, and none of the control firms participate. In practice for many programs, especially those which involve a cost to the firm for participating, compliance may be lower. For example, in an experiment to evaluate the effect of financial literacy training on female entrepreneurs in Uganda, only half of the treatment group invited to participate in the training actually attended (McKenzie and Weber, 2009). Other recent business training experiments have found higher attendance rates when given to members of microfinance groups (e.g. Berge et al. (2010) report 83% of the women in their business training experiment in Tanzania attended often enough to get the completion certificate), but 100 percent compliance is unlikely.

Table 5 reports the power of an experiment with 300 treatment and 300 control firms to detect the types of treatment effects often targeted as results of projects like those in Tables 3 and 4.[6] It also reports the size of the treatment group one would need in order to achieve 90 percent power, assuming an equal-sized control group. The results in the first column show that if a one-off follow-up survey is conducted of treatment and control firms, the sample sizes needed to detect program effects of interest are larger than the number of firms participating in typical SME programs, and indeed larger than the total number of manufacturing firms with 10 or more employees that exist in some African countries. For example, with treatment and control groups of 300 firms each, the power is only 12.9 percent for being able to detect a 20 percent increase in sales, 41.7 percent for detecting a 10 percent increase in employment, and 36.8 percent to detect a 5 percentage point increase in the proportion of firms exporting (from an assumed baseline rate of 12% of firms exporting).

The next three columns consider the improvement in power from also using baseline data, under a variety of assumptions about the autocorrelation in firm outcomes (ρ). Firm data exhibit considerable variability, much of it seemingly genuine (Fafchamps et al. 2010a), and available data suggest that the autocorrelation of profits or sales over periods of 6 months to 1 year is likely to be 0.5 or less (McKenzie, 2011). As a result, adding baseline data improves power

---

[6] Specifically, I assume mean sales of 1000 with a standard deviation of 3000, mean employment of 29 with a standard deviation of 21, and that 12% of firms in the control group export.

slightly, but not dramatically – with ρ = 0.5, we would still need treatment and control groups each of 3547 to detect a 20 percent increase in sales with 90 percent power.

Carrying out multiple rounds of post-treatment follow-up surveys and pooling these multiple measures can be used to further increase power (McKenzie, 2011). For example, Fafchamps et al. (2010b) use two pre-treatment and four post-treatment rounds of data collection in an experiment designed to alleviate capital constraints in Ghanaian microenterprises. This approach offers some hope for achieving adequate power levels to detect increases in employment or exporting with treatment and control groups of each 300 to 400 firms, but will still not yield enough power to detect a 20 percent increase in sales.

The bottom line therefore from this analysis is that many SME projects, even if experimentally implemented with 100 percent compliance, are unlikely to be able to tell whether they achieved their desired outcomes given the number and heterogeneity of the firms participating. Next, I discuss my thoughts on the ability of structural methods to resolve this problem, before returning to ask what we can identify with firm experiments.

3.3 Implications and Challenges for Structural Methods

I should acknowledge here that I am at most a consumer, rather than producer, of research on firms using structural models. With this in mind, let me discuss what I see of three important challenges that structural modeling is likely to face given the nature of firms and projects aiming to assist firms in many African countries.

The first challenge is that structural modeling of firm production and productivity is more convincing when homogeneous firms can be reasonably assumed to be using the same common production technology, leading to a focus of many industrial organization papers on very specific industries such as ready-mix concrete, ready-to-eat cereals, minivans, etc. Nevertheless, when it comes to productivity estimation, such narrow focus is the exception, and it is more typical for studies to group together firms at the 2-digit ISIC industry level (e.g. textiles and apparel, or food products). This is done for sample size reasons rather than any theoretical or empirical justification that the same production function should apply for all such firms. The reality of many interventions directed at firms in Africa is instead one of a wide mix of heterogeneous firms operating in many different sectors, for which assuming a common production technology is highly inappropriate. Assuming that this problem can be solved by the inclusion of sector dummies also appears wishful thinking. Alternatively, one could focus on specific industries, but the small number of firms in many industries in most African countries makes this more difficult to do than in larger developed country markets.

The second key challenge to structural modeling approaches to assessing whether firm policies are working in African countries is the pervasive market failures and potential externalities that these same policies are intended to overcome. Consider for example, attempting to use a structural approach to estimate the increase in production and employment from a matching

grant program.[7] Such a grant could induce additional investment by overcoming liquidity constraints, subsidizing firms for taking on uninsured risk, act as a spur to force entrepreneurs to avoid procrastination, change the decision-making process of firm owners by getting them to consider inputs (like consultants) that they had never considered before or had information about before, etc. The parameters in any structural model which has been estimated without taking into account these potential missing markets are unlikely to be informative about the effects of overcoming them. But standard approaches to production function estimation (e.g. Levinsohn and Petrin, 2003) do not account for such constraints. While structural models have been developed to incorporate liquidity constraints (Schündeln, 2010), such models rely heavily on functional form and distributional assumptions and are complex to solve – adding the full range of potential market failures facing these firms to such models would likely be extremely difficult to solve and even more reliant on additional questionable assumptions.

Note, however, that since we never directly observe productivity, some approach to productivity estimation is needed if productivity is the outcome of interest, even in an experimental impact evaluation. I have only seen experiments look at labor productivity (defined in terms of the observed quantities sales/employment) as an outcome, but if one wished to get the experimental impact on total factor productivity, then these structural approaches to productivity estimation would typically be needed (unless one conducts an experiment that independently shocks each input to the production function).

I am very sympathetic to the idea that, given the large cost and long time frame of existing SME interventions, it is desirable to learn what we can about the impact of these existing projects via ex post evaluation. But the third challenge facing both structural estimation and ex post non-experimental impact evaluations is the general lack of panel data on firms in most African countries. Both structural estimation and non-experimental impact evaluations are more convincing if there are many periods of pre-intervention data on which to either estimate structural parameters or from which to create non-experimental control groups, coupled with detailed data on firm participation in different programs, and several rounds of post-intervention data.[8] I do not know of a single African country for which such data are readily available.

4. A Way Forward?

So, where does this leave us? Is there scope for actually learning what works in firm programs in Africa given these issues? I think so, provided we are a little more modest in our aims and a lot more data intensive. While my belief is that there is a lot we can learn from well targeted

---

[7] Note, I am not aware of any paper which has actually tried to do this, but I use this to illustrate the difficulty a purely structural approach would have in answering this policy question of interest.

[8] See for example the rich data used by Kaboski and Townsend (forthcoming) in a structural evaluation of a microfinance program in Thailand. They also use quasi-experimental variation in helping to identify their structural model, illustrating the potential of these methods to work together.

experiments, many of the recommendations here will also aid in making structural modeling and non-experimental estimation more believable as well.

4.1 Focus on a smaller number of more homogenous firms

Given that I have said that one of the problems facing firm studies in Africa is the small number of firms, it may seem a little counterintuitive to say we should reduce the number of firms studied. But lumping together a firm with 5000 workers with one with 101 workers, or one with $50,000 annual sales with one with $5 million annual sales loses more through the increase in heterogeneity than it gains through the benefit in sample size. The key problem facing the power calculations in Table 5 is the large coefficient of variation. Restricting our set of attention to a set of more homogenous firms will shrink this coefficient, thereby increasing our power to detect program effects. Restricting ourselves to studying more homogeneous firms will also aid structural analysis by making it more reasonable to assume a common production function.

A first approach to doing this is just to screen firms based on baseline size, and focus any evaluation on identifying the average treatment effect of the intervention for a more homogenous subgroup of the overall pool of firms participating in the program. While we might not be able to say whether an export facilitation program works for all firms, separating out the few largest firms that dwarf the rest of the sample before randomizing the rest into treatment and control may allow us to cleanly identify treatment effects for the majority of interested firms.[9] One can go further and restrict attention to a smaller number of firms from a specific industry, as Bloom et al. (2011) do for textile firms in India. Homogenous firms operating in the same industry are likely to face many of the same seasonal effects and industry-level shocks, which can then be differenced out of the data, making it easier to distinguish the impact of an intervention from all these other factors that change firm outcomes from one period to the next. Such an approach is also helpful for structural modeling, since the production function and the nature of potential market failures facing the industry can likely be better understood and modeled than is the case with a mix of firms from different industries.

Table 6 illustrates using the World Bank Enterprise Surveys the reduction in the coefficient of variation (C.V.) of sales that one obtains by focusing on different sets of more homogeneous firms. I take the set of surveys where there are at least 200 firms with 10 to 100 workers, and then examine the C.V. for all firms in this group, then for the firms in specific industries, for the firms with sales below a certain threshold, and for firms with 10 to 30 workers. We see that focusing on a specific industry often, but not always, reduces the C.V. For example, in Ghana, the C.V. reduces from 3.0 for all firms to 1.2 for garment firms and 2.5 for firms in the food sector. However, in South Africa, the C.V. actually increases from 1.9 for all firms to 2.0 for garment firms and 2.1 for food firms, with the industry-level samples only about 12 percent of the size of the full sample. Focusing on more homogeneous firms in terms of employee size

---

[9] One can also increase power further within this subgroup by stratified or matched pair randomization designs (see Bruhn and McKenzie, 2009).

likewise has a general tendency to reduce the C.V., but some cases with small increases. The strategy that is most successful in reducing the C.V. of sales is to concentrate on firms with sales below a certain range (here $100,000 annual sales). This reduces the C.V. below one in all country cases. However, note that if sales are not strongly autocorrelated, then the C.V. of future sales may be larger than that of baseline sales in this group.

Reducing the C.V. can make an enormous difference to the sample size required to achieve adequate power. For example, to detect the 20 percent increase in sales with $\rho = 0.5$ and a single baseline and follow-up, Table 3 shows a treatment group size of 3547 firms is required. Reducing the coefficient of variation from 3 to 2 reduces the required treatment group size to 1577 firms, and reducing it to 1 (mean = standard deviation) reduces it to 395 firms. Based on this and experiences in other countries, my recommendation is therefore to aim for an initial pool of firms whose standard deviation of sales is no greater than their mean. However, as we have seen, the number of firms in any particular industry in most African countries can be quite small, so we are unlikely to get treatment and control groups of 300 firms each using such an approach. This brings me to the second step needed:

4.2 Collect a lot more data on these firms

The typical SME project collects very little data on the firms in the project, and often none on comparable firms not participating in the project. Even in cases where experiments are being planned, the default option is often to conduct a baseline survey, a midline survey (at say 1 year) in some cases, and then an endline survey. While such an approach is often appropriate for educational and health interventions in which outcomes like anthropometrics and test scores are highly persistent, but it is less well-suited to learning about dynamic firms, whose sales and profits can differ dramatically from one month to another due to seasonality, idiosyncratic demand shocks, shocks to the supply of labor, and a host of other reasons.

In McKenzie (2011), I provide theory and evidence to show the increases in power possible from collecting more waves of data than the usual baseline and follow-up. There are two important uses for more frequently collected data. The first, and better known, is for understanding the trajectory of program impacts, and for helping to unpack causal chains. For example, if we want to know whether an effort to formalize firms increases firm growth by allowing firms better access to formal credit, having multiple data points allows measurement of whether changes in credit precede firm growth or instead follow it. The second, and less practiced, use of more frequent data is to collect multiple measurements on noisy and weakly autocorrelated outcomes. For example, in a microenterprise survey, measuring several months of profits allows noise in a given month to be averaged out, increasing the power to detect genuine effects of a program.

With larger firms it may be possible to collect even more frequent data on key performance indicators like daily production, electricity usage, quality defects, and sales. Such data can then be used for both purposes above – averaging out daily fluctuations, and also providing detail on

the trajectory of an effect. Such data are also incredibly valuable for structural modeling, and for constructing comparable control groups in non-experimental impact evaluations. Moreover, with a large number of time periods on a smaller number of firms, one can move from standard estimators that rely on large-N asymptotics to estimators that rely on large-T. Bloom et al. (2011) provide an example of this, using over 100 observations per firm and the Ibramigov and Muller (2010) method to implement a t-statistic based estimator that is robust to substantial heterogeneity across firms as well as to considerable autocorrelation across observations within a firm.

4.3 Recognize what experiments can and cannot do

Employing the two recommendations offers the potential to estimate the average treatment effect of a firm policy for the types of firms that one has restricted the analysis to. Thus if we reduce the C.V. to a manageable level by focusing on firms with 10 to 100 workers and baseline sales below $100,000, this is the group for which we are able to estimate the policy impact. If there are only a couple of firms in the whole country with much larger sales and employment, an experiment is not going to be able to reveal the impact of the policy on these two firms, and pooling these firms in with the remainder of the firm population will make it hard to detect the impact for *any* type of firm. Note that this is not a limitation of experiments alone – if there are really only two firms like this in the entire country with no firms comparable to them, then non-experimental methods of impact evaluation are not going to be informative about these largest firms either.

Then what can we say about the effects of a program on the largest firms? This is where I believe time series analysis and structural modeling can play their most important role provided that *lots* of data is collected on such firms. If we have daily production data on the same firm for a decent period of time pre-intervention, it is more credible to believe that we can approximate well its production function than is the typical case of only a couple of periods of data on many firms. Detailed data on a lot of different dimensions of firm behavior also allows for construction of a spatial and temporal causal chain, and testing of falsifiable predictions of competing theories.

5. Conclusions

African SMEs are small in number and heterogeneous in performance, which poses a challenge for both experimental and structural methods of estimating the impact of firm policies designed to facilitate employment creation and firm growth. In reviewing African data on firms and on the types of policies governments are using with these firms, it seems there are three sets of firm types, for which different evaluation methods may be suited.

1. The vast majority of firms are microenterprises. Policies which focus on providing training, credit, or other assistance for these firms can often rely on standard

randomization of a large number of firms to treatment and control groups, with a small number of surveys taken on each firm.

2. For SME projects and policies targeting smaller number of firms, one should attempt to reduce heterogeneity by focusing on a more homogeneous sub-sample within this target group – such as firms in particular industries or size categories – and then collect data in many survey rounds or time periods on these firms. Randomized experiments can be used to obtain policy impacts for this sub-sample, and the rich data gathered can provide a good basis for structural modeling.

3. For the largest firms there may be no other firms in the country that are comparable, making the use of experimental or non-experimental methods that require a comparison group for estimating the counterfactual not possible. In such cases, rich time series data on multiple dimensions of firm behavior and firm outcomes can enable time series analysis and structural modeling to get some sense of how the policy has performed.

Finally, for years now there has been talk of how we can use experimental methods and structural modeling together to get the best of both worlds – using the experiment to identify structural parameters and impacts of a particular policy, and then the structural model to undertake simulations of how firms would respond under alternative policies (e.g. Todd and Wolpin, forthcoming). At present this is more promise and rhetoric than reality when it comes to firm experiments. The collection of much more data and the use of randomization in the implementation of more firm policies is a necessary first step in permitting the future development of such research.

References

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn and Bertil Tungodden (2010) "On the role of human and financial capital for microenterprise development: Evidence from a field experiment in Tanzania", Mimeo.

Bloom, Nick, Benn Eifert, Aprajit Mahajan, David McKenzie and John Roberts (2011) "Does management matter? Evidence from India", World Bank Policy Research Working Paper no. 5573.

Bruhn, Miriam and David McKenzie (2009) "In pursuit of balance: Randomization in practice in development field experiments", *American Economic Journal: Applied Economics* 1(4): 200-32.

Deaton, Angus (2010). "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48(2): 424-55.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008) "Returns to capital in microenterprises: Evidence from a field experiment", *Quarterly Journal of Economics* 123(4): 1329-72.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. (2008). "Using randomization in development economics research: A toolkit." In Handbook of Development Economics, Vol. 4, ed. T. Paul Schultz and John Strauss, 3895–3962. Amsterdam, NH: North Holland

Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff (2010a) "Using PDA consistency checks to increase the precision of profits and sales measurement in panels", *Journal of Development Economics*, forthcoming.

Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff (2010b) "When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana", Mimeo. World Bank.

Harrison, Ann, Justin Lin and Lixin Colin Xu (2010) "Explaining Africa's (dis)advantage", Mimeo. World Bank.

Ibragimov, Rustam and Ulrich Muller (2010), "t-statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28: 453 – 468

Imbens, Guido (2009). "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." *Journal of Economic Literature* 48(2): 399-423.

Kaboski, Joseph and Robert Townsend (forthcoming) "A structural evaluation of a large-scale quasi-experimental microfinance initiative", *Econometrica*.

Karlan, Dean and Jonathan Zinman (2010) "Expanding microenterprise credit access: Using randomized supply decisions to estimate the impacts in Manila", Mimeo. Yale University.

Levinsohn, James and Amil Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables","*Review of Economic Studies* 317-342

McKenzie, David (2010) "Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn?", *World Bank Research Observer*, 25(2): 209-33.

McKenzie, David (2011) "Beyond Baseline and Follow-up: The Case for more T in Experiments", Mimeo. World Bank

McKenzie, David and Michaela Weber (2009) "The results of a financial literacy and business planning training program for women in Uganda", *Finance and PSD Impact note no. 8.* World Bank.

McKinsey Global Institute (2010) *Lions on the Move: The progress and potential of African economies.* McKinsey Global Institute, Washington D.C.

Radelet, Steven (2010) *Emerging Africa: How 17 countries are leading the way.* Center for Global Development, Washington D.C.

Ravallion, Martin. (2009). "Should the Randomistas Rule?" The Economists' Voice, February 2009. (www.bepress.com/ev).

Rodrik, Dani. (2009). "The New Development Economics: We Shall Experiment, but How Shall We Learn?" pp.24-47 in Jessica Cohen and William Easterly (eds.) *What works in development: Thinking Big and Thinking Small*. Brookings Institute Press.

Schündeln, Matthias (2010) "Modeling firm dynamics to identify the cost of financing constraints in Ghanaian manufacturing", Mimeo. Goethe Universit, Frankfurt.

Todd, Petra and Kenneth Wolpin (forthcoming) "Structural estimation and policy evaluation in developing countries", *Annual Review of Economics*, forthcoming.

Van Dijk, Michiel (2003). "South African manufacturing performance in international perspective 1979-1999", *The South African Journal of Economics*, 71(1):119-142.

Table 1: Total Number of Firms in Selected African Countries

	Botswana 2005	Ethiopia 2008/09	Ghana 2003	Mauritius 2007	Madagascar 2005	Tanzania 2008	Uganda 2006/07
Total number of manufacturing firms	1302	n.a.	26088	13639	19334	24979	n.a.
With <10 employees	872[b]	n.a.	22181	12798[a]	18030	24204	n.a.
With 10 or more employees	430	2203	3907	841	1304	775	1024
With 100+ employees	78	605 with 50+	251	n.a.	190 with 200+	80	114
Total number in trade	5817	n.a.	n.a.	n.a.	159594	122622	n.a.
With <10 employees	5064[b]	n.a.	n.a.	35132	157080	121589	n.a.
With 10 or more employees	753	n.a.	n.a.	n.a.	2514	1033	n.a.
With 100+ employees	26	n.a.	n.a.	n.a.	89 with 200+	0	n.a.

Notes:

n.a. denotes not available

Firms are defined as establishments in many cases

a) Mauritius data is for firms with 10 or fewer employees, and for firms with 10+ employees

b) Botswana data was grouped as 1-4, then 5-29 workers, so 5-29 split according to Ghana and Tanzania proportions within this category

Sources

Botswana data from Central Statistics Office, Republic of Botswana, http://www.cso.gov.bw/index.php?option=com_content&task=view&id=88&Itemid=88

Ethiopia data from Central Statistical Agency of Ethiopia, Large and Medium Manufacturing Survey http://www.csa.gov.et/index.php?option=com_content&view=article&id=62&Itemid=489

Ghana data from 2003 industrial census, http://www.statsghana.gov.gh/Industrial_Census.html

Mauritius data from Central Statistics Office, Government of Mauritius, http://www.gov.mu/portal/goc/cso

Madgascar data from http://www.instat.mg/pdf/enq_entreprises_2005.pdf

Tanzania data is from National Bureau of Statistics, Tanzania from 2007 Business Survey and excludes Zanzibar http://www.nbs.go.tz

Uganda data from Uganda Bureau of Statistics http://www.ubos.org/onlinefiles/uploads/ubos/pdf%20documents/more%20notes%20on%20Business%20Register07.pdf

Table 2: Heterogeneity in Firm Outcomes among African Firms

Country	Survey Year	Sample Size	Employment <10 employees Mean	S.D.	10-100 employees Mean	S.D.	>100 employees Mean	S.D.	Sales <10 employees C.V.	10-100 employees C.V.	>100 employees C.V.	% of Firms Exporting 10+ workers
Ghana	2007	494	6.3	1.3	22.3	16.0	330	522	6.1	3.0	1.7	9.1
Kenya	2007	657	6.5	1.3	36.7	26.5	386	502	1.5	6.6	4.1	3.2
Mali	2007	490	6.2	1.1	23.0	16.5	215	80	2.7	2.3	1.1	5.9
Mozambique	2007	479	6.0	1.3	29.7	20.9	214	145	3.3	15.1	5.0	5.2
Nigeria	2007	1,891	6.5	1.3	24.1	18.4	188	125	9.6	2.6	1.8	9.2
Senegal	2007	506	6.3	1.3	22.9	19.9	429	623	1.7	1.9	1.8	10.3
SouthAfrica	2007	937	6.4	1.3	33.8	23.3	446	1005	1.5	1.9	2.6	22.6
Zambia	2007	484	6.3	1.4	33.9	24.7	275	324	2.2	1.4	1.4	11.6
Benin	2009	148	4.6	2.0	24.6	11.9	252	156	1.8	2.6	0.9	8.8
BurkinaFaso	2009	391	6.3	1.9	27.6	20.3	211	410	3.6	5.5	1.3	12.8
Cameroon	2009	363	6.1	1.6	29.2	19.6	424	804	1.8	5.3	1.8	2.2
CapeVerde	2009	152	6.1	1.8	33.5	22.3	225	150	2.9	6.3	4.1	11.2
Chad	2009	150	6.1	1.5	28.4	22.0	183	112	3.6	2.7	1.0	7.3
Congo	2009	139	5.1	2.6	31.4	21.6	327	209	3.5	2.0	1.3	7.9
Eritrea	2009	177	7.0	1.2	23.7	16.9	125	9	1.3	2.2	0.5	23.2
Gabon	2009	170	6.6	1.5	23.2	18.0	336	367	2.5	1.9	1.1	10.0
Ivory Coast	2009	526	4.6	2.2	25.0	18.8	425	1008	4.4	4.2	1.5	4.8
Lesotho	2009	146	5.6	1.8	29.5	20.7	3207	12688	2.9	2.1	1.9	21.2
Liberia	2009	150	6.0	1.5	24.1	17.4	193	72	6.1	6.2	1.2	28.7
Madagascar	2009	444	7.0	1.4	35.8	25.2	344	332	1.5	3.5	1.5	16.4
Malawi	2009	149	5.8	1.5	40.2	26.9	522	1038	1.0	1.8	1.7	12.8
Mauritius	2009	398	5.5	2.2	33.6	23.8	397	523	4.1	12.6	1.8	20.4
Niger	2009	150	6.2	1.6	28.9	21.1	211	210	2.4	3.7	1.4	14.0
Sierra Leone	2009	150	6.6	1.2	25.2	21.6	278	250	3.0	4.8	1.1	2.0
Togo	2009	154	5.7	1.6	28.5	21.1	254	157	2.5	2.0	1.0	19.5
Average		396	6.1	1.6	28.8	20.6	416	873	3.1	4.2	1.8	12.0

Source: World Bank Enterprise Surveys Harmonized Database

Table 3: Examples of World Bank Private Sector Projects involving Firms in Africa

Country	Project Name	Year	Type of Intervention	US$ amount (000s)	Number of Beneficiary Firms
Cote d'Ivoire	Private Sector Capacity Building Project	1999-2004	Matching Grant program for Exporters	2200	149 firms, 27 associations
			Capacity building for SMEs	n.a.	not implemented
South Africa	Industrial Competitiveness and Job Creation Project	1998-2004	Matching Grant I: Competitiveness fund	17850	984 firms
			Matching Grant II: Black-Business Suppliers	1870	672 firms
			Matching grant fund for sub-sector partnerships	7620	96 partnerships
Zimbabwe	Enterprise Development Project	1996-2002	Matching grants for associations	n.a.	4
			Matching grants for export firms	1000	13
			SME finance facility	n.a.	1079
			Export finance facility	1500	28
Nigeria	Second National Fadama Development Project	2004-2009	Matching grants for productive assets	15320	7,766
Mali	Private Sector Assistance	1992-2002	Matching grants	3300	468 firms, 303 associations
Niger	Niger Agro-Pastoral Export Promotion Project	2001-2005	Matching grants for Exports	3950	64 associations, 20 enterprises
Mozambique	Enterprise Development Project	2000-2006	Matching grants for firms	1860	328 firms
			SME credit line	4335	52
Zambia	Enterprise Development Project	1992-2003	Credit facility to SMEs	59830	192 firms
			Matching grants	1100	63 firms
Mauritius	Technical Assistance to Enhance Competitiveness Project	1994-99	Matching grants	5500	199 firms
Benin	Private Sector Development Project	2000-07	Matching grants		
Nigeria	Private Small and Medium Enterprise Development Project	1989-1994	SME loans	132770	211 firms
Mozambique	Small and Medium Scale Enterprise Development Project	1990-1997	SME credit line	28800	134 firms
Malawi	Financial Sector and Enterprise Development Project	1992-98	SME credit line	25300	126 firms
Zimbabwe	Small Scale Enterprise Project	1986-94	SME credit line	8500	246 firms
Ghana	Private Small and Medium Enterprise Development Project	1989-1997	SME credit line	28200	109 firms
			Equipment leasing	1400	25 firms
			business training for entrepreneurs		587 individuals
Senegal	Private Sector Capacity Building Project	1996-2001	Matching grants	2750	301 firms
Kenya	Micro and Small Enterprise Training and Technology Project	1998-2002	Training for microenterprises	7500	34778 individuals

Sources: World Bank Implementation Completion Reports

19

Table 4: Selected IFC SME Projects in Africa

Country	Project	Year	Type of Intervention	Number of Firms
Ghana	Ahafo Linkages	2007-	Managerial Mentoring	101
Guinea	Guinea SME Linkages Project	2008-	Capacity Building through Training	18
Mozambique	Mozambique SME Initiative	2004-2010	Royalty Loans	18
Mozambique	Mozlink	2006-	Managerial and Technical Training	80 (planned)
South Africa	Thandi Land Reform Project	2007-	Package of finance, skills and equity	30 (planned)
Zambia	Copperbelt Suppliers SME Development Program		Capacity Building through Training	300
			In-depth Advisory Services	36
Cameron	Business Edge SME training courses	2010-11	Business training	125
Chad	Business Edge SME training courses	2010-11	Business training	48
Ghana	Business Edge SME training courses	2010-11	Business training	48
Kenya	Business Edge SME training courses	2010-11	Business training	382
Madagascar	Business Edge SME training courses	2010-11	Business training	115
Mozambique	Business Edge SME training courses	2010-11	Business training	132
Nigeria	Business Edge SME training courses	2010-11	Business training	320
Rwanda	Business Edge SME training courses	2010-11	Business training	57
Senegal	Business Edge SME training courses	2010-11	Business training	19
South Africa	Business Edge SME training courses	2010-11	Business training	16

Source: http://www.ifc.org/ifcext/africa.nsf/Content/SMED_Programs and IFC Business Edge team.

Table 5: Power calculations for detecting impact of firm policies

	Single	Baseline + Single Follow-up			Baseline + Multiple Follow-ups (ρ=0.5)		
	Follow-up	ρ=0.3	ρ=0.5	ρ=0.7	2 follow-ups	3-follow-ups	4-follow-ups
10% increase in sales							
Power with 300 treated	0.069	0.071	0.076	0.088	0.089	0.097	0.102
Treatment Sample size needed for 90% power	18914	17212	14186	9646	9457	7881	7093
20% increase in sales							
Power with 300 treated	0.129	0.137	0.156	0.208	0.211	0.244	0.266
Treatment Sample size needed for 90% power	4729	4303	3547	2412	2365	1971	1774
50% increase in sales							
Power with 300 treated	0.532	0.571	0.654	0.815	0.823	0.885	0.915
Treatment Sample size needed for 90% power	757	689	568	386	379	316	284
10% increase in employment for firm with 10-100 workers							
Power with 300 treated	0.417	0.450	0.524	0.688	0.697	0.774	0.815
Treatment Sample size needed for 90% power	1030	938	773	526	515	430	387
5 percentage point increase in exporting							
Power with 300 treated	0.368	0.518	0.599	0.764	0.772	0.842	0.878
Treatment Sample size needed for 90% power	1080	784	646	439	431	359	323

Notes: equal size treatment and control groups assumed

Table 6: Reducing the Coefficient of Variance for Sales by Focusing on more homogeneous firms

Country	Survey Year	All Firms 10-100 employees		Firms in Garment Sector 10-100 employees		Firms in Food Sector 10-100 employees		All sectors 10-30 employees		Firms with Sales less than $100k 10-100 employees	
		Number of Firms	C.V. Sales	Number of Firms	C.V. Sales	Number of Firms	C.V. Sales	Number of Firms	C.V. Sales	Number of Firms	C.V. Sales
Angola	2006	249	9.5			59	0.6	234	9.8	82	0.4
Tanzania	2006	207	3.7			47	4.1	137	2.7	86	0.7
Uganda	2006	323	4.5			59	3.3	255	3.2	172	0.7
Ghana	2007	211	3.0	54	1.2	44	2.5	177	3.4	135	0.6
Kenya	2007	356	6.6	39	2.2	68	2.6	201	9.3	51	0.5
Mozambique	2007	250	15.1			59	7.3	168	2.4	89	0.5
Nigeria	2007	925	2.6	75	1.2	226	1.8	744	2.7	397	0.5
SouthAfrica	2007	558	1.9	66	2.0	70	2.1	347	1.6		
Zambia	2007	255	1.4			69	1.6	155	1.2	44	0.4
Madagascar	2009	239	3.5					141	5.5	64	0.6
Mauritius	2009	207	12.7			37	2.0	126	10.7	38	0.5
Average		343.6	5.9	58.5	1.7	73.8	2.8	244.1	4.8	115.8	0.5

Source: World Bank Enterprise Surveys Harmonized Database