Public Disclosure Authorized

Public Disclosure Authorized

WPS4944

Policy Research Working Paper 4944

Impact Assessments in Finance and Private Sector Development

What Have We Learned and What Should We Learn?

David McKenzie

The World Bank
Development Research Group
Finance and Private Sector Team
May 2009



Policy Research Working Paper 4944

Abstract

Until recently rigorous impact evaluations have been rare in the area of finance and private sector development. One reason for this is the perception that many policies and projects in this area lend themselves less to formal evaluations. However, a vanguard of new impact evaluations on areas as diverse as fostering microenterprise growth, microfinance, rainfall insurance, and regulatory reform demonstrates that in many circumstances serious evaluation is possible. The purpose of this paper is to synthesize and distil the policy and implementation lessons emerging from these studies, use them to demonstrate the feasibility of impact evaluations in a broader array of topics, and thereby help prompt new impact evaluations for projects going forward.

This paper—a product of the Finance and Private Sector Team, Development Research Group—is part of a larger effort in the department to conduct impact assessments of FPD policies. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at dmckenzie@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn?[#]

David McKenzie, *World Bank*

Keywords: Impact Evaluation; Finance; Private Sector Development; Randomized Experiment

[#] I thank Miriam Bruhn, Asli Demirgüç-Kunt, Xavier Gine, Bilal Zia, the editor, and three anonymous referees for useful comments and discussions. All opinions are of course my own and do not necessarily reflect the views of the World Bank.

Introduction

The recent external review of World Bank research noted that "perhaps the most important role of Bank research is to learn what works, and to widely disseminate the results" (Banerjee et al. 2006, p. 148). Rigorous impact evaluations, which compare the outcomes of a program or policy against an explicit counterfactual of what would have happened without the program or policy, are one of the most important tools that can be used along with appropriate economic theory for understanding "what works". Despite this, until recently impact evaluations have been rare, especially outside the areas of health and education.[1] This is now particularly apparent in the area of finance and private sector development, where the recent financial crisis has prompted renewed attention to knowing what works in terms of getting finance to consumers and firms, and in getting the private sector growing again.[2]

One reason for the lack of impact evaluations in this area is the perception that many finance and private sector development (hereafter FPD) policies and projects lend themselves less to formal evaluations.[3] Changes in laws or regulations may occur at an economy-wide level, or a large loan may only be given to one or two banks or firms. However, in many cases it is still possible to formally evaluate FPD policies or projects. Regulations may be implemented in some regions and not others, or apply only to firms of a certain industry or size. Generally available programs or policies may have low take-up that can be raised through targeted interventions. And in a non-trivial number of cases it will indeed be feasible to implement a randomized experiment. The purpose of this paper is to demonstrate the feasibility of such impact evaluations, distil the lessons of these new evaluations for policymakers and practitioners, and help prompt new impact evaluations for projects going forward.

---

[1] For example, the Development Impact Evaluation (DIME) Initiative has until recently focused on topics in health and education. See http://web.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDEVIMPEVAINI/0,,menuPK:3998281~pagePK:64168427~piPK:64168435~theSitePK:3998212,00.html [accessed February 4, 2009].

[2] See also the recent World Bank Policy Research Report on Access to Finance which calls for more impact evaluation (Demirgüç-Kunt et al. 2008).

[3] A second reason may be that research on FPD has historically worried less about the challenges of identification that are a prime concern of the labor and applied microeconomics literature. Financial economists are much less likely to be exposed to impact evaluation methods in their graduate classes than health, education, or labor economists. A further purpose of this paper is thus to better expose practitioners in the FPD field to the ideas and possibilities of impact evaluations.

We begin by highlighting policy and implementation lessons from four areas where impact evaluations are beginning to emerge: microenterprises, microfinance, rainfall insurance, and regulatory reform. We use impact evaluations in these areas to illustrate various methods which are possible when evaluating FPD reforms, as well as to note some of the key challenges to their effective use. We then discuss several reasons why these policy areas are at the forefront of FPD impact evaluations, which leads to a discussion of where the biggest opportunities appear to be going forward for new knowledge generation of what works.

Many of the examples discussed here will come from randomized experiments, which have increasingly become the preferred method of evaluation for many development economists (Duflo and Kremer, 2005). Randomized experiments offer many advantages for evaluation, chief among them being that they ensure that they only reason why some firms, consumers, or other units are subject to a policy or program and others are not is pure chance. This also makes the results easy to communicate to policymakers.

However, recently there has been a debate about whether the profession is over-emphasizing randomization (Rodrik, 2008; Deaton, 2009; Ravallion, 2009; Imbens, 2009). Many of the issues discussed, such as for whom the treatment effect is identified for, and whether the results are generalizable to other settings, are also important considerations in using non-experimental methods. There are three lessons from this debate that I consider important for the discussion in this paper. The first is that we must not let methodological purity determine which questions to try and address: just because a policy cannot be randomized does not mean we should give up on trying to understand whether it is working or not. Indeed this paper considers a range of approaches that can be used for ensuring more rigorous impact evaluation. Second, studies need to go beyond a simplistic black-box approach of "does this work or not" to try and understand why and how it works, and for whom? Finally, I agree with Imbens (2009) who argues 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. Randomization is not always feasible, but I do not know of a single study that has

credibly argued that they could have randomized, but choose not to do so because of a belief that they would get a more rigorous assessment of impact by not randomizing.

What Have We Learned?

Raising the incomes of the self-employed

Self-employment accounts for a large share of the labor force in most developing countries. For example, Gollin (2002) reports that in Ghana, Bangladesh, and Nigeria, 75-80 percent of manufacturing workers were self-employed. Self-employment is particularly important among the poor. Banerjee and Duflo (2007) find that between 47 and 69 percent of urban households who live on less than US$2 per day in Peru, Indonesia, Pakistan and Nicaragua own a non-agricultural business. A central question for policymakers is then how to raise the incomes of these poor businesses, and whether in fact the typical microenterprises owned by the poor have any ability to grow.

In the absence of market failures, a standard model of firm size determination (e.g. Lucas, 1978) would argue that the answer is no – the reason for firms being small in such models is that their owners have low entrepreneurial ability. Of course market failures are pervasive in many developing countries, with restrictions on access to credit being a notable example. However, an influential branch of theory suggests that in the presence of credit constraints, the prospect of microenterprise growth from small investments is low, due to production non-convexities (Banerjee and Newman, 1993). The argument is that the profitable investments facing a business are lumpy (e.g. a new machine), and that without sufficient access to external credit, individuals who start a business too small will be trapped in poverty, earning low returns. Conversely, if these non-convexities are not important, then if small firms are operating well below the optimal production point (given their entrepreneurial ability), we might expect the returns to additional capital investment to be particularly high.

However, assessing the extent to which a lack of capital hampers the income growth of microenterprises is complicated by the fact that firm owners with more capital stock or greater access to credit are likely to differ in a host of other ways from owners with less capital, such as in terms of entrepreneurial ability in the Lucas model. Two recent randomized experiments in Sri Lanka and Mexico (de Mel et al. 2008a and

McKenzie and Woodruff, 2008) illustrate one approach to impact evaluation which can resolve this problem and credibly identify the impact of additional capital on firms. Grants of between US$100 and US$200 were given to a randomly selected subset of poor microenterprises in each country. The authors can then compare the profits of firms which randomly received these grants to those which did not, to determine the extent to which grants raise business incomes.[4] Their results challenge the somewhat conventional wisdom that subsistence firms have no scope for group (see Table 1 for a summary of key results from studies in this paper). They find the grants do substantially raise incomes for the average firm receiving a grant, and estimate real returns to capital of 5.7 percent per month in Sri Lanka and 20 percent per month in Mexico, much higher than market interest rates in both countries. They explore heterogeneity in the treatment effects in an attempt to understand why the returns are so high. They find returns to be highest for high ability, credit-constrained firm owners, which is consistent with the view that credit market failures prevent talented owners from getting their firm to its optimal size.

These randomized experiments show grants work in raising incomes for the average microenterprise owner. In the particular research studies, the grants were not part of a government or NGO program, but rather given out by the researchers and funded through research grants. However, there are several cases where governments have employed grants as a way of raising the incomes of the self-employed. An example is the *Microemprendimientos Productivos* program in Argentina which provided financial support in the form of in-kind grants to finance inputs and equipment to beneficiaries with the aim of helping them obtain a sustainable source of income and reduce their dependence on welfare payments (Almeida and Galasso, 2007). The Mexican *Jovenes con Oportunidades* program provides grants to youth for completing the last few years of schooling, with these grants kept in bank accounts that can be accessed for paying for further study or for starting a business. Grants to microenterprises are also more common in disaster recovery situations, such as following the Indian Ocean tsunami of December 2004 (de Mel et al, 2008c).

[4] Comparing profits requires knowing how to measure the profits of microenterprises which are usually informal and keep few records. Impact evaluations have been useful for learning what works in this regard too (see de Mel et al., 2009).

A question which faces policymakers who wish to give grants to raise the incomes of microenterprises is whether these grants should be in the form of unrestricted cash, or made in-kind, as was the case with the Argentine program.[5] In the randomized experiments in Sri Lanka and Mexico, half of the grants were given as cash, and the other half as raw materials and equipment for the businesses (chosen by the owner). The authors find in both studies that there is no difference between the two forms of grant: they result in approximately the same change in capital stock and same increase in business profits. If business owners have profitable opportunities to expand they will invest additional cash in these opportunities. If they do not, then any inputs or equipment they provide will crowd out the investments they would have made on their own, and they can sell excess capital stock if it is not yielding a return. This suggests that policymakers can achieve the same results with the cheaper and easier to administer cash grants.[6]

Impact evaluations are not only useful for showing what works, but also what does not. This can guide new policy experiments. A first example of this from the microenterprise experiments is that while the grants succeeded in raising the incomes of male business owners, the average return to capital for women receiving the grant in Sri Lanka was zero (the Mexican study contained only men). Grants alone thus did not work in raising the incomes of self-employed women. In follow-up work, de Mel et al. (2008b) combine the experimental results with several theoretical models to try and understand why the grants did not work for raising business income for women. They find that women did not invest smaller grants in the business, while the larger grants invested in the business had low returns. They speculate that a possible explanation for this is inefficient household use of resources, with other household members capturing a share of the income and working capital held by women, leading women to use fixed business assets as a store of value rather than simply for production. They also find returns to be

[5] This parallels the debate in the conditional cash transfers literature as to whether the conditions attached to cash grants matter (see Fiszbein and Schady, 2009). Our finding of no differential effect of conditioning does not immediately carry over to other forms of conditioning, such as conditioning on school attendance or health clinic visits, since firm owners can undo the conditioning of being required to spend the money on their business more easily than they can undo other types of conditions – e.g. in theory they could devote less time to school work at home if children attend school more, but this seems less likely.

[6] Although conditional grants may be still prepared from a political economy perspective, since grants may be easier to sell to the public if they are conditioned on the recipients "using them properly".

particularly low in business sectors dominated by women. This has led to ongoing field experiments designed to determine the impact on business profits of getting women to shift into sectors which both men and women work in, as well as a replication of the study in Ghana to understand whether the same gender differences emerge in a country with much higher female participation rates in self-employment.

A second example of what does not work from this same body of work is that although the one-time grants succeed in raising the incomes of male poor business owners, they do not lead to significant employment creation. A comparison of the characteristics of microenterprise owners with those of wage workers and owners of firms with five or more employees suggests that only one-quarter to one-third of microenterprise owners have attributes such as ability, motivation, and ambition similar to that of larger firm owners (de Mel et al., 2008d). The key question for policymakers is then how to unleash the employment-creating potential of these select microenterprise owners. In addition to access to credit, business training and business development services have been the typical programs governments have tried to do this. However, to date there has been little rigorous evaluation of business training programs[7], something which ongoing evaluations hope to correct.

Rethinking the central precepts of the microfinance movement

The previous section demonstrated that one-off grants can raise the incomes of the average microenterprise owner. Grants to certain vulnerable groups, and perhaps even large sections of the poor, may be sustainable as part of a government social protection program (the *Oportunidades* program in Mexico covers 5 million households, almost one-quarter of Mexico's population).[8] However, in terms of finance and private sector development policies, most of the focus on households and microenterprises has been through microfinance. The most famous example of microfinance is that of the Grameen bank, and the model of microfinance most strongly associated with it is group lending to women at low interest rates. Recent impact evaluations (along with the success of

[7] An exception is Karlan and Valdivia (2008) who find that business training increases the sales and repayment rates of female microfinance clients in Peru.

[8] See http://www.oportunidades.gob.mx/Wn_Inf_General/Padron_Liq/Cober_Aten/index.html [accessed February 5, 2009].

microfinance institutions such as *Banco Compartamos* in Mexico which offers individual loans at quite high interest rates to both men and women) give strong reasons to question this archetypical model of microfinance as necessarily the best way to expand access to finance to the poor and to improve the small business sector going forward.[9]

Many microfinance organizations focus almost exclusively or largely on female borrowers. For example, 97 percent of Grameen Bank's seven million borrowers are women[10], as are 70 percent of FINCAs borrowers[11], and 65 percent of ACCIÓN's five million clients.[12] While part of this reflects a social mission, many of the justifications are economic in nature. Women are argued to be poorer than men on average (e.g Burjorjee et al., 2002; FINCA, 2007), have less collateral, and hence be more credit-constrained (e.g. Khandker, 1998, Armendáriz and Morduch, 2005). But if this is the case, when women do receive access to credit, it should generate higher returns than when men receive access. The experimental evidence from Sri Lanka (and supporting non-experimental evidence from Mexico and Brazil) in de Mel et al. (2008b) provides a reason to question this extensive focus on women, and a suggestion that more products need to be developed to fit the needs of urban male clients.

Group liability is often hailed as one of the central innovations of the microfinance movement, mitigating both the adverse selection and moral hazard problems which can give rise to credit market failures. The idea is that borrowers who know they will be liable for the debts of others in their group will have an incentive to screen others so that only reliable people will join their group, and then to monitor their group members to ensure they invest their funds wisely and exert enough effort. However, as Giné and Karlan (2008) note, group liability has several pitfalls which may cause it to be disliked by many borrowers. It may be particularly troublesome for small business owners, who might be discouraged from undertaking somewhat risky but high return projects by other group members, may need different size loans or different loan periods from other group members, and find frequent group meetings costly in terms of

[9] See Cull et al. (2009) for a description of the heterogeneity in the microfinance sector, and the debate generated by the successful stock offering of Banco Compartamos. Karlan and Morduch (2009) provide an excellent overview of recent research on access to finance.
[10] http://www.grameen-info.org/bank/index.html [Numbers as of May 2007], accessed August 15, 2007.
[11] http://www.villagebanking.org/site/c.erKPI2PCIoE/b.2604299/k.FFD9/What_is_Microfinance_What_is_Village_Banking.htm, accessed August 15, 2007.
[12] http://www.accion.org/about_key_stats.asp [all clients 1976-2006], accessed August 15, 2007.

time. Finally, there is also a concern that group liability loans are less useful for establishing credit records in credit bureaus than individualized loans, making graduation to larger loans more difficult (de Janvry et al., 2008).

Giné and Karlan (2008) carried out a randomized experiment with a microfinance bank in the Philippines to investigate the extent to which group lending really reduces the moral hazard problems. Half of the group-lending centers of the bank were randomly chosen to be converted to individual liability. They find no change in default rates after one and three years in the converted centers, and faster client growth in the converted branches. These results suggest that group liability is not that important for reducing moral hazard, but since the converted loans were all initially screened by groups, the paper can not say anything about the importance of groups for screening out bad risks. Ongoing work by the authors is examining this issue, comparing newly formed groups to new individual loan clients.

The third precept of microfinance that has been strongly challenged by recent impact evaluations is the belief that serving the poor requires low interest rates. Muhammad Yunus (2007) states "a true microcredit organization must keep its interest rate as close to the cost of funds as possible", criticizing the high interest rates being charged by Banco Compartamos. This lies at the heart of the debate on commercialization of microfinance (see Cull et al., 2009 and Harford, 2008). However, the high returns to capital for many microenterprises in Sri Lanka and Mexico suggest the ability to repay loans at rates significantly higher than market interest rates. The problem, especially for urban business owners seeking individual loans, is often one of access rather than interest rate. In follow-up work in Sri Lanka, de Mel et al. (2009b) find that few of the high return microenterprises qualify for a loan from microfinance banks, which lend on a basis of physical collateral and not on whether the owner's business shows high prospects for growth.

The most striking evidence that high interest rate loans can improve welfare comes from a study of consumer loans in South Africa. Karlan and Zinman (2008) conducted a randomized experiment with a microlender, in which applicants which were marginally rejected for consumer loans were randomly selected into two groups, one of which received a second look and higher probability of getting a loan. The loans were 4

month loans at a monthly interest rate of 11.75 percent (equivalent to an APR of 200% per year). Despite these high interest rates, the authors find that six to twelve months later the marginal loan recipients were more likely to have kept their job, had higher incomes, and experienced less hunger. This is not to argue that the customers would not have been even better off had loans been available at lower interest rates. But at the existing rates, not only did the customers benefit, but these marginal loans appear to have been profitable for the bank.

This study illustrates well some of the pros and cons of trying to build policy on the basis of a randomized experiment. The impacts estimated are credible and easily understood by policymakers. They are the impacts for marginally rejected consumers, a group of interest certainly to the bank. However, the fact that this group can benefit a lot from additional access to high interest rate credit is not informative about whether poorer individuals who are far from the creditworthy cutoff would stand to benefit from high interest rate loans – other studies are needed to look at this question.

To be sure, these existing impact evaluations consist of only a couple of rigorous studies from a couple of countries, and it will be important to see if the results are repeated in replication studies. Nevertheless, the results do suggest reasons to question the structure of the prototypical microfinance product. Moreover, despite the rampant expansion of microfinance worldwide and tremendous amount of attention this has received in the media, to date there has been little rigorous impact evaluation of the welfare effects of the basic microfinance product.[13] Several large-scale randomized trials of microfinance are currently nearing completion. The first preliminary results from a randomized trial involving 2,400 households in India were recently presented by Esther Duflo.[14] While the full results are not yet available, two points to note are first, take-up was only 17.5 percent. That is, most households offered a loan did not want one. Second, the preliminary results show very modest impacts, with no significant effects on health or education, and relatively little use for business purposes. As more results become

[13] See Armendáriz and Morduch (2005) for a summary of different non-experimental approaches that have been used to measure impact. The most well known of these is Pitt and Khandker (1998), who employ a regression discontinuity design. There is some debate as to the extent to which the regression discontinuity applied in practice, see the discussion in Armendáriz and Morduch.

[14] Presentation by Esther Duflo at the Innovations for Poverty Action 2008 Microfinance Conference at Yale University. Discussion of these results is covered at http://www.philanthropyaction.com/nc/the_real_impacts_of_micro_credit/ [accessed February 5, 2009].

available from this and other impact evaluations of microfinance going forward, it will lead new impetus to policy efforts in the microfinance domain.

Insuring poor farmers

Missing credit markets are one important reason why firms in developing countries are less productive than they could be. However, reluctance to take-up credit may be linked to the existence of another important market failure, the lack of an insurance market. This may be particularly important in occupations such as farming, which are subject to substantial income risk from rainfall variation during the growing season. One solution which has been proposed and introduced in a number of countries is Rainfall Index Insurance, which links payouts to rainfall at local rain gauges.

An important question of interest is then whether offering this rainfall insurance works in increasing the use of credit by risk-averse farmers. A randomized experiment conducted by Giné and Yang (2009) among farmers in Malawi finds evidence that it does not. The authors worked with the Malawian farmers' association, financial institutions in Malawi, and the Commodity Risk Management Group of the World Bank to offer smallholders credit to purchase high-yielding seed varieties. Farmers in some localities were randomly selected to be just offered credit, while those in other localities were offered a bundle of credit and insurance. Take-up of the credit was 33 percent for farmers offered the loan without insurance, and only 17.6 percent for farmers who were offered a loan bundled together with rainfall insurance.

Take-up rates of rainfall insurance have also been low elsewhere – Giné et al. (2008) report a take-up rate of only 4.6 percent for one product in India. In a cross-sectional non-experimental setting, they find that risk-averse households are actually less likely, not more likely, to purchase the insurance, especially when they are unfamiliar with other types of insurance and the insurance provider. They attribute this to uncertainty about the insurance product, which as a new technology requires some risk and trust to participate in it. In follow-up randomized experiments in India, Cole et al. (2008) investigate the sensitivity of the take-up decision to price, the presence of an endorsement from a third trusted party, means of presentation, and liquidity constraints.

Their results are consistent with the view that in addition to price and liquidity, trust and financial literacy influence take-up to a significant degree.

These studies have several implications for efforts to develop better insurance products for the poor. In addition to finding that price matters, the findings on trust and financial literacy suggest scope for modifying implementation and marketing in a way which will boost demand. To the extent that poor farmers are unable to understand complicated insurance products, as an introductory product to get people used to the idea of insurance, a simpler product design with fewer thresholds and payment schedules may be preferred to a more complicated product that offers more complete insurance.[15] For example, a product that pays out if rainfall is below 150mm during the specified period and does not if rainfall is above is simpler to understand than the more standard product that, in the example in Cole et al. (2008, p. 9), "pays zero when cumulative rainfall during a particular 45 day period exceeds 100mm. Payouts are then linear in the rainfall deficit relative to this 100mm threshold, jumping to Rs. 2000 when cumulative rainfall is below 40mm". It would be interesting in future impact evaluations to compare the take-up and efficacy of simpler designs to more complex designs.

Secondly, the authors find take-up to be much higher in villages where a positive past insurance payout has occurred. They conclude from this that it would be useful to modify the contracts to ensure they pay out a positive return with sufficient frequency as to engender trust in the population, whereas the standard contracts pay out very rarely. The trade-off here is that for the same insurance premium, more frequent payouts mean smaller amounts can be paid out each time, resulting in less complete coverage of catastrophic losses to compensate for greater coverage of more common losses. Third, since liquidity constraints mattered a lot for take-up, they suggest that it might be beneficial to bundle the insurance product together with a loan. The results in Malawi shows this results in less credit uptake than if pure loans were offered, but it might offer greater insurance uptake than if insurance alone was offered, and would not preclude offering a separate loan-only product.

[15] This is not to preclude also offering the more complicated products at the same time, and letting farmers choose between them. An alternative would be better financial education to teach the participants how to learn this product. Cole et al. (2008) implemented brief (5 to 10 minute) training sessions on this, which they found had no effect.

Learning from regulatory reform

The impact evaluations profiled above used randomized experiments to randomly offer the program to selected individuals, firms, banking branches, farming localities, and slum areas. However, this approach to evaluation may not be possible with some forms of FPD projects, such as reforms in the regulatory environment. Nevertheless, in many cases rigorous impact evaluation is still possible. We illustrate this through consideration of two recent studies which have conducted impact evaluations of regulatory reforms.

The view that burdensome regulations are an important barrier to private sector development was famously expressed by de Soto (1989), who calculated that it would take 289 days, 11 permits, and over $1,000 to legally register a small business in Peru. This emphasis on regulatory reform has been further spurred by the World Bank's *Doing Business* project, which ranks countries each year on both the overall ease of doing business, and on the extent of reforms undertaken in the previous year. The 2009 report notes that almost 1000 reforms have been recorded in the areas measured by Doing Business have occurred in the past six years, with the most common reform being one which makes it easier to start a business by reducing the costs and number of procedures needed. Yet despite the huge number of reforms, there is almost no rigorous impact evaluation of these reforms.

An exception is found in Bruhn (2008) and Kaplan et al. (2007), who study the impact of business registration reform in Mexico. The reform was organized by a federal agency, but implemented at the municipal level since many business registration procedures were set locally. Due to staffing constraints, the federal agency could not implement the reform in all priority municipalities at once, but instead staggered the reforms, introducing them first in some municipalities and then later in others. Among the municipalities identified as priorities for implementation, there was no specification of which should go first. This allows the author to use municipalities in which the reform was introduced later as a control group for the municipalities in which it was introduced earlier, using a difference-in-differences estimation methodology. This estimation essentially looks at the period where the first few municipalities had reformed and others had yet to. It then compares the change in the number of registered businesses (or in other outcomes of interest) for those municipalities where the reform was introduced early to

the change in these same outcomes for municipalities where the reform was introduced later. This is an estimation strategy that is likely to be applicable in understanding a number of other regulatory reforms, which might be phased in over time.[16]

The headline result from both Bruhn (2008) and Kaplan et al. (2007) is that the reform succeeded in increasing registrations. This is where the most simplistic measures of impact would stop. For example, World Bank (2008) reports that following a reduction in the minimum capital requirement, there was an increase in new company registrations of 55 percent in Georgia and 81 percent in Saudi Arabia.[17] But to know if a reform worked and why, we want to go beyond did it lead to more businesses, to understand how and why? In the specific example of business registration reform, an important question of interest is whether these new registrations are the results of existing firms registering, or of new firms starting up. Bruhn (2008) finds that the increase in registrations comes from new entry, not from the conversion of existing informal firms.

This result suggests there may be a group of potential self-employed for whom the burden of registering is a barrier to business formation; but once this pool of pent-up demand is exhausted, there may be much less long-term impact. The results here do not support de Soto's (1989) view that existing informal small business owners are individuals who wish to become formal, but are stymied by high barriers to registration. They are more consistent with the view that the majority of informal businesses are informal by choice, because becoming formal offers no benefit to them. Indeed, McKenzie and Sakho (2009) estimate that for Bolivian small firms, there are huge gains to becoming formal for the subset of informal firms who do not know how to become formal, but that becoming formal would be costly to the remainder of informal firms.

We also want to know what the consequences of these reforms are for the economic outcomes we ultimately care about, such as employment generation, consumer

[16] Note that the validity of this difference-in-difference estimation strategy relies on an assumption that the municipalities which reform later are a good comparison group for what would have happened to the earlier reform municipalities in the absence of early reform. Bruhn (2008) carries out a number of checks on pre-existing trends and municipality characteristics to argue this is the case. This strategy will be less applicable if countries decide to, for example, first introduce the reform in the capital city or business capital, and then roll the reform out to progressively smaller cities.

[17] Note that these numbers for Georgia and South Africa are not even the true impact on the number of registrations, since they are a simple before-after comparison and do not control for pre-existing trends or concurrent events in the economy.

welfare, and economic growth. Bruhn (2008) finds the Mexico reform increased employment by 2.8 percent after the reform, and benefited consumers by decreasing prices by 0.6 percent, likely as a result of additional competition. However, in doing so, it reduces the income of incumbent registered business owners. Since municipal level GDP is collected only every five years, it is not possible to look at the overall impact on economic growth.

Although in some cases reforms might be introduced in a staggered fashion into some regions of the country first, a more common experience is for the reform to be introduced for the entire country at once. But even in this situation, it is often the case that the reform only applies to, or should theoretically only have consequences for, a subset of the population. One special case of this is when the policy only applies to firms above (or below) some particular size threshold. A relatively common example of this occurring is in the area of labor regulation, where employment protection rules might apply only to firms above a certain number of workers.[18] For example, both Italy's employment protection legislation and Sri Lanka's termination of workmen act place much more onerous requirements on firms with 15 or more employees. In some circumstances this might allow evaluation of the effects of the reform by comparing firms just above the threshold to those just below, a regression discontinuity design. This is done for Italy by Leonardi and Pica (2006). However, in practice such regulations will often cause firms to sort themselves around the size threshold, making this approach to evaluation more challenging. Abidoye et al. (2008) find some evidence that this is the case in Sri Lanka, with firms slower to grow from 14 to 15 workers than from 13 to 14 workers or from 15 to 16 workers.

More typically reforms introduced at the country level may affect only some firms or industries, but not others.[19] This allows for a difference-in-differences estimation strategy in which unaffected firms or industries are used as a comparison group for those

[18] Priority lending also may have size thresholds. See Banerjee and Duflo (2008) who study a reform in India which increased the maximum size limit for firms to be eligible for priority-sector lending. They then use a triple-difference evaluation strategy, comparing the change in the rate of changes in outcomes before and after the reform for firms that were newly eligible for priority lending compared to firms that were already eligible.

[19] Another example is seen in Kugler et al. (2005), who study a reform of Spain's labor law, which applied only to some demographic groups such as young workers, older workers, women under-represented in their occupations, and disabled workers, but not other groups.

affected by the reform. An example of this is seen in Giné and Love (2006), who evaluate the impact of a bankruptcy reform in Colombia which reduced the costs of re-organizing a bankrupt firm. Their goal is to see whether the law change led to distressed, but viable firms, being more likely to reorganize when they would have previously liquidated. Since active, non-bankrupt firms are not affected by the law, they can use a difference-in-difference strategy to compute the difference in the characteristics of bankrupt firms selecting into re-organization rather than liquidation after the law was reformed relative to the characteristics of active firms, relative to this same difference pre-reform. They find that lowering the costs of re-organization led to an improvement in the efficiency of the bankrupt procedure, with more viable firms now more likely to be re-organized than liquidated relative to the pre-reform situation.

Lessons for Implementation of Impact Evaluations

The impact evaluations summarized above have begun to yield important policy lessons for work with microenterprises, microfinance, rainfall insurance, and regulatory reform. These are all important components of finance and private sector development policy, yet they only cover a fraction of the important policy tools and research areas in the FPD domain. The questions which then arise are why these few areas have been at the forefront of evaluation efforts to date, and what lessons do they hold for other evaluations going forward?

Why have these subject areas dominated evaluation efforts to date?

A substantive reason why these topics have been at the forefront of evaluation efforts is that they have close ties with important bodies of theoretical work in development economics, and that in many cases the theory suggests reasons both why the policy may have its intended effect, as well as reasons why it may not work in practice. For example, in the grants to microenterprises, one body of theory suggested returns to capital may be very low due to non-convexities, while another body suggested returns could be high due to credit constraints with convex production technologies. Likewise there are theoretical reasons why group lending may have benefits, as well as reasons why it may deter certain types of borrowers. These cases where the impact of the

program is theoretically uncertain motivate empirical studies to see what happens in practice.

A more practical reason is that these studies are all in areas where evaluation is most feasible for a variety of reasons. The first is one of sample size. The policies studied are ones where the units of analysis are consumers or firms, allowing the comparison of the impacts on many affected units to a control group of many other units. The second is one of data availability. The regulation studies relied on unusually good existing databases in Mexico (a quarterly labor force panel survey and administrative data from the Mexican social security system) and a comprehensive database on the universe of bankruptcy cases in Colombia. The other studies were designed as ex-ante evaluations, with data collection designed by the researchers. The randomized experiments done to date have generally been conducted by researchers working with NGOs or funding the programs through research grants. This has limited study to either programs which have been run by NGOs willing to work with researchers, or to projects which are cheap enough for research grants to fund.

Going forward this calls for a need for continued close interaction between theory and evaluation – we want to know not just whether or not something works, but why, and how? It also suggests that widespread rigorous evaluation of the many other types of FPD programs and policies implemented by governments and supported or advised by international financial institutions requires a much greater commitment to evaluation, and in particular, to planning ahead so the evaluation process (including data collection) can start before the program is implemented. It also suggests unexploited benefits exist from small modifications in currently collected sources of data which do not presently have policy evaluation in mind. For example, surveys of firms should include questions on participation in particular types of policies or projects (e.g. does your firm participate in a business cluster developed by the government under its regional clustering program), and include enough identifying information to link with administrative records on banks, firms, or consumers participating in such programs. And unfortunately even when such data is collected, access to the microdata is often limited in many countries, so greater data accessibility is also needed.

Evaluation of many FPD programs is possible

The studies highlighted above have demonstrated a variety of methods that can be used for evaluating FPD policies – randomized experiments, difference-in-differences, and regression discontinuity designs. There are a variety of other evaluation methods available which when used carefully can also be informative as to policy impacts. We highlight here three of these other methods which are also likely to be useful in evaluating a broad array of FPD policies.[20]

Propensity-score matching is a commonly used method for estimating a treatment impact. An example in the FPD literature is seen in Oh et al. (2008) who evaluate the impact of a credit guarantee policy used by the Korean government to support small and medium enterprises in the aftermath of the Asian financial crisis. The authors use plant-level panel data on manufacturing firms and match firms which received credit guarantees to similar firms which did not, finding that the guarantee program positively affected both survival rates and sales and employment growth of the firms receiving the guarantees. A concern with propensity score matching is that it assumes the process of which firm receives a guarantee and which does not can be adequately captured by a set of observable variables which the firms are matched on. How plausible this is will be a judgment call in any given setting, and benefit from detailed knowledge of how the program was actually implemented. In general the literature has found the results to be closer to those obtained in an experimental setting when a rich set of data can be used for the matching, including multiple periods of pre-program data to control for existing trends. The data used by Oh et al. (2008) do not meet these criteria, with only data from one year (2000) for a relatively limited set of firm characteristics being used. This suggests one should be cautious in accepting their results.

A second method is the control function approach introduced by Heckman, which involves explicitly modeling how unobservables which affect the outcome are related to the observables, including the choice of participation in a program or policy regime. This approach is used along with propensity-score matching by Fajnzylber et al. (2006) to look at the impact of access to credit, training, and membership in business associations on

[20] For a good recent general reference to different estimation strategies for impact evaluation, see Imbens and Wooldridge (2008). Instrumental variables is another common technique for evaluation which we won't explicitly discuss here – McKenzie and Sahko (2009) provide an example in FPD.

microenterprises in Mexico. Traditionally these methods have relied heavily on functional form assumptions and distributional assumptions such as joint normality, which can lead to significant bias when these assumptions do not hold and as a result such methods have fallen out of favor in much of development economics. However, recently semi-parametric approaches have been developed which rely less on these assumptions, but which still need an exclusion restriction to hold (see Heckman and Navarro-Lozano, 2004 for a review and comparison to matching). This need for an exclusion restriction takes us back to the need to answering the underlying question needed for evaluation: thinking of an exogenous reason why some firms, consumers, or other units participate in a program and others do not.

A third method which is likely to be applicable for a wide variety of FPD evaluations is an encouragement design (Diamond and Hainmueller, 2007). This can be useful when evaluating a program that is implemented at the country-level, such as a change in regulation or in policy. The basic idea behind this design is that firms (or other units of interest) are randomly divided into a treatment and a control group. While the program is available to all, the treatment group receives additional encouragement to participate in the program – for example they might receive marketing visits to make them more aware of the program. If the encouragement is successful it yields a difference in program take-up rates between the two groups which can then be used in evaluating the impact of the program. More precisely, what can be estimated is the impact of the program on units which would take up the program when offered encouragement but which would not otherwise.

An example of an encouragement design being successfully used is seen in de Janvry et al. (2008), who examine the impact of the introduction of a credit bureau in Guatemala. While the credit bureau is in place for everyone, knowledge of its implementation was found to be almost non-existent in surveys conducted soon after its implementation. The authors therefore randomly informed a subset of 5,000 microfinance borrowers about the existence of the bureau and how it works. They find this awareness of the bureau leads to a modest and temporary increase in repayment rates, and to microfinance groups ejecting their worse-performing members.

The IFC has recently attempted encouragement designs in two evaluations, an ongoing evaluation of business registration in Lima, Peru where firms receive encouragement to register; and an evaluation of an alternate dispute resolution (ADR) project in Macedonia. The preliminary results from the Macedonia project also illustrate the potential downside of this approach in estimating the impact – the encouragement might not encourage very many units to take-up the program.[21] The Macedonia project tried several methods of encouraging use of the ADR mechanism, but found that none of these encouragement methods succeeded in raising use. This prevents estimation of the effect of the ADR on firms. Nevertheless, a finding that no one wants to participate in a program, even when encouraged, is in of itself a useful result for understanding the likely program impact. More detailed analysis of why firms do not take-up the program can then be used to improve the program going forward.

The importance of take-up

A key difference between evaluation of most FPD programs and many impact evaluations in education and health lies in take-up. In programs such as vaccination campaigns or get children to school programs, the goal of the program is to have all eligible individuals participate. And in the case of cash transfers, participation can be close to universal. In contrast, universal take-up is not the goal of most FPD programs, and even when it is a goal, it is seldom the reality. Not all households or firms will want or need a loan, register formally, or wish to purchase insurance. This is evident in some of the studies profiled above: take-up rates of 17.5 percent for microfinance, 5 to 33 percent for rainfall insurance, and no increase in the number of firms in the informal sector registering to become formal when regulations changed.

Less than universal program take-up offers both challenges and opportunities for impact evaluation. Learning what the level of take-up is, and which characteristics predict take-up can be useful for refining and modifying the policy to enable it to better reach its goals in the future. For example, the low take-up of risk-averse individuals in the rainfall insurance papers, coupled with the fact that take-up was much higher when there had been a recent pay-out in the village or when there was an endorsement from a trusted

[21] Discussion of the Macedonia results is based on correspondence with Alexis Diamond in the IFC.

third party can help guide marketing efforts and product design in the future. It can lead to revealing other market or government failures where policy action is required.[22] Finally, take-up rates and characteristics can also be useful for gauging the potential market for taking pilot trials to scale.

However, low take-up also offers several challenges for attempts to rigorously evaluate FPD programs. The first is one of power to detect the program effect. For example, one of the ultimate goals of the work on rainfall insurance is to find out if rainfall insurance allows farmers to farm more efficiently, and protects their households against negative shocks. However, because few farmers purchased the insurance, and those who did purchase only purchased enough to cover a trivial fraction of their crops, the existing studies do not allow the researchers to determine the impacts on production and household welfare.

There are two solutions to this problem of power. The first is to employ a very large sample size, so that the resulting sample will still contain enough firms or households which take-up the program to enable the researchers to detect a program impact of a given size. However, the downside of this is that it can be very expensive. For example, consider a program such as a new loan product or business training that aims to raise the profits of microenterprises undertaking the program by 25 percent. A randomized experiment which offered the program to half the firms and used a single follow-up survey to estimate this impact would require a sample size of 670 firms if take-up was 100 percent, but need a sample size of 2,700 with 50 percent take-up, and of 67,000 with 10 percent take-up.[23] An example of a randomized experiment with sample sizes of this magnitude is seen in Karlan and Zinman (2009) who randomized 58,000 direct mail offers issued by a South African lender, with 8.7 percent of those contacted applying for a loan.

[22] For example, de Mel et al. (2009b) worked with a regional development bank to try and help microenterprises obtain loans. Despite 62 percent of firms showing up for information meetings, only 10 percent received loans. One reason was that in the absence of a credit bureau, applicants had to travel to other institutions and obtain letters from them attesting that they had no outstanding loans, thereby increasing the cost to applicants of applying for loans. This experience highlights the need for credit bureaus to cover microfinance.

[23] These calculations were made using the sampsi command in STATA, assuming a constant treatment effect, a coefficient of variation of 1, which is in line with what one typically sees in microenterprise data after trimming outliers, that the treatment has no effect on the variance of profits, and for power of 0.90 and test significance level alpha of 0.05.

The second solution to the problem of low power from low take-up is to restrict study to a group of units for whom take-up would be much higher. For example, a business training program could be advertised to all eligible firms or microfinance clients, and then the number of slots available in the program could be randomly allocated among the group of interested firms. A related example is seen in Karlan and Zinman (2008), in which consumers first apply for loans, and then the pool of marginally rejected candidates (all of whom wanted a loan) is then randomly assigned to receive a second look at getting a loan. The advantage of this second approach is that it requires much smaller samples to detect a treatment impact. The downside is that of external validity – the program impact estimated will apply only to the self-selected group of individuals or firms which expressed interest in the program, not to the general population. In some cases however this might be precisely the impact of interest – for instance, policymakers might want to know what the effect of their loan program is on firms interested in taking up credit.

The second challenge offered by low take-up is one of interpretation of program impact. Consider evaluating the impact of microfinance on microenterprise profitability in a situation where take-up of loans is only 10 percent. With a randomized experiment, comparison of the mean profits of firms *offered* the microfinance treatment to those which were not offered the microfinance treatment yields the average intention-to-treat effect. This is the impact on firms of being offered credit. This in itself is a parameter of interest, but in most cases we would also like to go further and know what the impact of the credit was if it was actually taken up. The standard approach is to instrument receipt of microfinance with the randomly determined offer of credit. However, if the impact of receiving credit varies by firm, what is recovered is known as a local average treatment effect (LATE) (Angrist and Imbens, 1994). This is the average effect of receiving credit for firms which would take-up the microfinance treatment when offered. If firms which stand to benefit more from credit are the ones who take it, this will overstate the gain in profit which the average firm would receive if it got microfinance.[24] What this means in practice is that there needs to be care taken in interpreting program effects with low take-

[24] See Heckman et al. (2006) and Deaton (2009) for more discussion on interpretation of treatment effects when the take-up decision is a choice which is related to the individual unit's program effect. Ravallion (2009) also discusses some related issues in the use and interpretation of experiments.

up, and in deciding whether the parameter estimated is in fact one of policy interest. Researchers can also go further in understanding the underlying observable sources of heterogeneity in the take-up decision and in treatment effects.

What Should We Learn?

The previous sections have shown that evaluation of FPD programs and policies is possible in a wide variety of contexts, and that the small number of evaluations to date are yielding useful lessons for both policy design and future evaluation efforts. The question is then where should we go from here? While we have argued that there is much greater scope for serious evaluations than is currently being realized, two general areas are particularly attractive for increased efforts.

The first is more evaluations in the areas that have been at the forefront of existing efforts: microfinance, microenterprises, insurance, and regulatory reform. We noted that there are a number of features of these policy domains that lend themselves to rigorous evaluation. Yet there are currently only a handful of rigorous studies. More are needed on a wider range of policies in a number of different institutional settings, to learn what works, where, and why?

The second general area where there appears to be unexploited gains to be made from impact evaluation is in looking at the effects of other programs and policies that are widely used to benefit large numbers of consumers and firms. Three such important policy areas where evaluation seems possible, yet is currently almost non-existent, are financial literacy and consumer protection, business training, and policies to enhance the SME sector.

The subject of financial literacy has received increased policy attention in recent years, with worldwide efforts underway to role out financial literacy training. For example, Citi Foundation is four years into a ten year, $200 million global program of financial education, operating in 65 countries and a number of governments have developed programs in this area.[25] The recent global financial crisis has also turned attention to issues of consumer protection, and the possible macroeconomic

[25] See http://www.citigroup.com/citi/financialeducation/ [accessed February 10, 2009].

consequences of consumers entering into credit transactions that they do not fully understand. Financial literacy programs are ripe for evaluation efforts, since despite the increasing amounts of money, there are always groups of consumers that receive the program and others that do not. The challenge for evaluation is making these two groups as similar as possible, and measuring the outcomes.

One preliminary study in Indonesia has found teaching financially illiterate individuals about the benefits of bank accounts did lead to an increase in bank account use among this group, with no increase for those who were already financially literate (Cole, Sampson, and Zia, 2009). Moreover, they find small incentive payments to have a much larger effect on getting individuals to open bank accounts and to be three times as cost-effective as financial education in this regard, suggesting a need for some skepticism in judging some of the lofty claims of proponents of financial education. This fledging effort provides a good base for future evaluations to build on, with the ultimate goal of finding out under what circumstances such programs work, when they do not work, and what the consequences on consumer welfare are.

A second area which is ripe for experimentation and impact evaluation lies in business training programs. Many microfinance organizations, NGOs, and governments worldwide offer short courses to budding or existing microenterprises to teach them the basics of running a small business. Public sector funding of such programs may be justified from a poverty alleviation standpoint, since even if the programs worked and had large benefits, credit constraints and risk aversion might prevent poor people participating. Again in these programs one can design impact evaluations by comparing firm owners which are offered the training to similar individuals that are not offered the training. Several randomized experiments currently in the field are attempting to do this.

The last area I wish to stress as being particularly full of unexploited possibilities for impact evaluations lies in policies directed at the SME sector. These include SME lending policies, trade credit policies, management training, and sector-specific technical assistance. These programs are typically carried out by governments and international financial institutions (IFIs) rather than NGOs, and are too expensive for researchers to

typically fund the program on offer themselves.[26] As a result, there is a real knowledge gap – and an opportunity to be grasped. If governments and operations staff at IFIs can work with researchers in evaluating the many projects being implemented, it should be possible to rigorously evaluate many of the policies being carried out for SMEs, and to learn where modifications of existing strategies are needed.

Conclusions

This paper has surveyed the existing literature on impact evaluations in finance and private sector development with two main aims. The first was to draw emerging policy lessons and implementation lessons from the slowly growing set of rigorous impact assessments that have been carried out in areas such as microfinance, microenterprise growth, rainfall insurance, and regulatory reform. The second aim was to use the lens of these existing evaluations to demonstrate some of the different strategies for evaluation that are possible, and to argue that much more impact evaluation is possible than has currently been attempted.[27] Hopefully policymakers and operational staff reading this paper will agree with this message, and join together with researchers in better understanding what works and why?

[26] A nascent effort to evaluate a few of the IFC's programs has been underway for a few years. IFC (undated) describes some of these efforts. However, to date these efforts have to my knowledge not resulted in any working papers or published articles.

[27] The Finance and Private Sector Development team of the Development Research Group has recently introduced a new impact note series to try and better disseminate the results of new impact evaluations which do occur. See http://econ.worldbank.org/programs/finance/impact to see the latest in FPD impact evaluations.

References

Abidoye, Babatunde, Peter Orazem and Milan Vodopivec (2008) "Firing Cost and Firm Size: A Study of Sri Lanka's Severance Pay System", Iowa State University Working Paper # 08014

Almeida, Rita and Emanuela Galasso (2007) "Jump-starting self-employment? Evidence among welfare participants in Argentina", *IZA Working Paper no. 2902.*

Angrist, Joshua D., and Guido Imbens (1994) "Identification and estimation of local average treatment effects," *Econometrica*, 62(2), 467–75.

Armendáriz, Beatriz and Jonathan Morduch (2005) *The Economics of Microfinance*. MIT Press: Cambridge, MA.

Banerjee, Abhijit, Angus Deaton, Nora Lustig, and Ken Rogoff (2006) "An evaluation of World Bank research, 1998-2005", http://siteresources.worldbank.org/DEC/Resources/84797-1109362238001/726454-1164121166494/RESEARCH-EVALUATION-2006-Main-Report.pdf [accessed February 4, 2009].

Banerjee, Abhijit and Esther Duflo (2007) "The Economic Lives of the Poor", *Journal of Economic Perspectives* 21(1): 141-67.

Banerjee, Abhijit and Esther Duflo (2008) "Do Firms want to borrow more? Testing credit constraints using a directed lending program", Mimeo. MIT.

Banerjee, Abhijit and Andrew Newman (1993) "Occupational Choice and the Process of Development", *Journal of Political Economy* 101: 274-298.

Bruhn, Miriam (2008) "License to sell: The effect of business registration reform on entrepreneurial activity in Mexico", *World Bank Policy Research Working Paper No. 4538.*

Burjorjee, Deena M., Deshpande, Rani, and Weidemann, C. Jean (2002), "Supporting Women's Livelihoods Microfinance that Works for the Majority. A Guide to Best Practices", United Nations Capital Development Fund, Special Unit for Microfinance. http://www.uncdf.org/english/microfinance/pubs/thematic_papers/gender/supporting/part_1.php

Cole, Shawn, Xavier Giné, Jeremy Tobacman, Petia Topalova, Robert Townsend and James Vickrey (2008) "Barriers to Household Risk Management: Evidence from India", Mimeo. World Bank.

Cole, Shawn, Thomas Sampson and Bilal Zia (2009) "Valuing Financial Literacy", Mimeo. World Bank.

Cull, Robert, Asli Demirgüç-Kunt and Jonathon Morduch (2009) "Microfinance meets the market", *Journal of Economic Perspectives*, forthcoming.

Deaton, Angus (2009) "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development", Mimeo. Princeton University.

De Janvry, Alain, Craig McIntosh and Elisabeth Sadoulet (2008) "The Supply- and Demand-Side Impacts of Credit Market Information", Mimeo. UCSD.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2009) "Measuring microenterprise profits: Must we ask how the sausage is made?",*Journal of Development Economics*, 88(1): 19-31.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2009b) "Getting Credit to High Return Microenterprises: The Results of an Information Intervention", Mimeo. World Bank.

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

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008b) "Are women more credit constrained ? Experimental evidence on gender and microenterprise returns", *American Economic Journal: Applied Economics*, forthcoming.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008c) "Rebound: How Enterprises Recover from a Natural Disaster", *Mimeo. World Bank.*

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008d) "Who are the Microenterprise Owners?: Evidence from Sri Lanka on Tokman v. de Soto", forthcoming in Joshua Lerner and Antoinette Schoar (eds.) *International Differences in Entrepreneurship*. NBER, Boston, MA.

Demirgüç-Kunt, Asli, Thorsten Beck and Patrick Honohan (2008) *Finance for All? Policies and Pitfalls in Expanding Access.* World Bank Policy Research Report. World Bank, Washington, D.C.

de Soto, Hernando (1989) *The Other Path: The Economic Answer to Terrorism*. Basic Books: New York, NY.

Diamond, Alexis and Jens Hainmueller (2007) "The Encouragement Design for Program Evaluation", http://www.ifc.org/ifcext/rmas.nsf/AttachmentsByTitle/Encouragement/$FILE/The+Encouragement+Design+for+Program+Evaluation.pdf [accessed February 10, 2009].

Duflo, Esther and Michael Kremer. (2005). "Use of Randomization in the Evaluation of Development Effectiveness" In *Evaluating Development Effectiveness*, ed. Osvaldo Feinstein, Gregory K. Ingram and George K. Pitman, 205-232. New Brunswick, NJ: Transaction Publishers.

Fajnzylber, Pablo, William Maloney, and Gabriel Montes Rojas. (2006). "Releasing Constraints to Growth or Pushing on a String? The Impact of Credit, Training, Business Associations and Taxes on the Performance of Mexican Micro-Firms." World Bank Policy Research Working Paper No. 3807. Washington, D.C.

FINCA (2007), "Frequently Asked Questions", http://www.villagebanking.org/site/c.erKPI2PCIoE/b.2394157/k.8161/Frequently_Asked_Questions.htm [accessed August 15, 2007].

Fiszbein, Ariel and Norbert Schady (2009) *Conditional Cash Transfers: Reducing Present and Future Poverty*, Policy Research Report, The World Bank: Washington, D.C.

Giné, Xavier and Dean Karlan (2008) "Peer Monitoring and Enforcement: Long Term Evidence from Microcredit Lending Groups with and without Group Liability", Mimeo. World Bank.

Giné, Xavier and Inessa Love (2006) "Do Reorganization Costs Matter for Efficiency? Evidence from a Bankruptcy Reform in Colombia", *World Bank Policy Research Working Paper No.* 3970.

Giné, Xavier and Dean Yang (2009) "Insurance, Credit, and Technology Adoption: Field Experimental Evidence from Malawi", *Journal of Development Economics*, forthcoming.

Giné, Xavier, Robert Townsend and James Vickrey (2008) "Patterns of Rainfall Insurance Participation in Rural India", *World Bank Economic Review* 22(3): 539-66.

Gollin, Douglas (2002) "Getting Income Shares Right, *Journal of Political Economy* 110(2): 458-474.

Harford, Tim (2008) "The battle for the soul of microfinance", *Financial Times* December 6.

Heckman, James and Salvador Navarro-Lozano (2004) "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models", *The Review of Economic and Statistics* 86(1): 30-57.

Heckman, James, Sergio Urzua and Edward Vytacil (2006) "Understanding Instrumental Variables in Models with Essential Heterogeneity" *Review of Economics and Statistics*, 2006, 88(3): 389-432.

IFC (undated) "Innovations in Impact Evaluation in IFC", *IFC Monitor: Results Measurement for Advisory Services*. http://www.ifc.org/ifcext/rmas.nsf/AttachmentsByTitle/Innovationsmonitor/$FILE/Innovations2.pdf [accessed February 10, 2009].

Imbens, Guido (2009) “Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)”, Mimeo. Harvard University.

Imbens, Guido and Jeffrey Wooldridge (2008) “Recent Developments in the Econometrics of Program Evaluation”, Mimeo. Harvard University.

Kaplan, David, Eduardo Piedra and Enrique Seira (2007) “Entry regulation and business start-ups : evidence from Mexico”, World Bank Policy Research Working Paper No. 4322.

Karlan, Dean and Jonathan Morduch (2009) “Access to Finance”, Chapter 2 in M. Rosenzweig and D. Rodrik (eds.) *Handbook of Development Economics, Volume 5*. forthcoming.

Karlan, Dean and Martin Valdivia (2008) “Teaching Entrepreneurship: Impact Of Business Training On Microfinance Clients and Institutions”, Mimeo. Yale University.

Karlan, Dean and Jonathan Zinman (2009) “Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment”, *Econometrica*, forthcoming.

Karlan, Dean and Jonathan Zinman (2008) “Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts”, *Review of Financial Studies*, forthcoming.

Khandker, Shahidur R. (1998), "Using microcredit to advance women", World Bank Premnote (November) No8. http://www1.worldbank.org/prem/PREMNotes/premnote8.pdf [accessed August 15, 2007].

Kugler, Adriana, Juan Jimeno, and Virginia Hernanz (2005) “Employment Consequences of Restrictive Permanent Contracts: Evidence from Spanish Labor Market Reforms” Mimeo. University of Houston.

Leonardi, Marco and Giovanni Pica (2006) “Effects of Employment Protection Legislation on wages: a Regression Discontinuity approach”, Mimeo. University of Milan.

Lucas, Robert E. (1978) "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9(2): 508-523

McKenzie, David and Christopher Woodruff (2008) “Experimental Evidence on Returns to Capital and Access to Finance in Mexico”, *World Bank Economic Review*, 22(3): 457-82.

McKenzie, David and Yaye Seynabou Sakho (2009) “Does it pay firms to register for taxes? The impact of formality on firm profitability”, *Journal of Development Economics*, forthcoming.

Oh, Inha, Jeong-Dong Lee, Almas Heshmati and Gyoung-Gyu Choi (2008) “Evaluation of credit guarantee policy using propensity score matching”, *Small Business Economics*, forthcoming.

Pitt, Mark and Shahidur Khandker (1998) “The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?” *Journal of Political Economy* 106(5): 958-996.

Ravallion, Martin (2009) “Should the Randomistas Rule?”, *The Economists’ Voice*, www.bepress.com/ev, February 2009.

Rodrik, Dani (2008) “The New Development Economics: We shall experiment, but how shall we learn?, Mimeo. Harvard University.

World Bank (2008) *Doing Business 2009*. World Bank, Washington D.C.

Yunus, Muhammed (2007) “Remarks by Muhammad Yunus, Managing Director, Grameen Bank.” Microcredit Summit E-News, Volume 5, No. 1, July 2007.

Table 1: Summary of Main Findings

Study	Policy or Program Studied	Main results
Panel A: Results that largely confirm or support conventional wisdom		
Bruhn (2008) and Kaplan et al. (2007)	Business registration reform in Mexico	- reform increased the number of registered firms and employment. Less in line with conventional wisdom, Bruhn shows this is from new entry, not formalization of existing firms.
Gine and Love (2006)	Bankruptcy reform in Colombia	- reducing reorganization costs improves efficiency of bankruptcy process, with more viable firms more likely to be re-organized and less viable firms to be liquidated.
Oh et al. (2008)	Credit guarantee policy in Korea to support SMEs during crisis	-guarantees improved survival rates, sales growth, and employment growth
de Janvry et al. (2008)	Introducing a credit bureau in Guatemala	- awareness of the bureau leads to a modest and temporary increase in repayment rates and to microfinance groups ejecting worst-performing members.
Panel B: Results that challenge or overturn conventional wisdom		
de Mel et al. (2008a,b)	Conditional and Unconditional grants to microenterprises in Sri Lanka	- returns to capital are high for male-owned firms, but zero for female-owned firms - no difference between conditional and unconditional transfers
Gine and Karlan (2008)	Removing group liability in microfinance groups in the Philippines	- no change in default rates when joint liability removed, and faster client growth in converted branches
Karlan and Zinman (2008)	High interest rate consumer loans in South Africa	- high interest loans let marginal recipients to be more likely to keep their jobs, have higher incomes, and experience less hunger
Gine and Yang (2009) and Cole et al. (2008)	Offering rainfall insurance to farmers in Malawi and India	- take-up is extremely low, so that insurance leads to little risk mitigation or changes in farmer behavior
Cole, Sampson and Zia (2009)	Financial literacy training in Indonesia	- program had zero impact on the general population, but increased bank account use for financially illiterate. However, small cash payments had much more effect than financial education.