

DOING IMPACT EVALUATION

No.

8

Impact Evaluation for Land Property Rights Reform



THE WORLD BANK

Poverty Reduction and
Economic Management

PREM

Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation

Impact Evaluation for Land Property Rights Reforms

November 2007

Acknowledgement

This paper was written by Jonathan Conning and Partha Deb¹. The authors would like to thank Markus Goldstein, Malcolm Childress, Karen Macours, and participants at a November 8th, 2006 World Bank workshop hosted by the Poverty Analysis, Monitoring and Evaluation Thematic Group for excellent suggestions, and Papa Seck for excellent research assistance. Comments welcome.

This work program was task managed by Markus Goldstein and financed by the Norwegian Trust Fund for Impact Evaluation in Infrastructure, the Trust Fund for Environmentally and Socially Sustainable Development supported by Finland and Norway, and by the Bank-Netherlands Partnership Programs.

¹ Department of Economics, Hunter College and The Graduate Center, City University of New York New York, NY 10021. Contact information: jonathan.conning@hunter.cuny.edu and partha.deb@hunter.cuny.edu

TABLE OF CONTENTS

INTRODUCTION	1
I. WHY AN IMPACT EVALUATION?	5
II. EXPECTED IMPACTS OF LAND PROPERTY RIGHTS REFORMS	7
III. IMPACT EVALUATION CHALLENGES	10
A. NON-RANDOM PROGRAM PLACEMENT	11
B. PARTICIPANTS SELF-SELECT	12
IV. STUDY DESIGNS AND EVALUATION METHODS	12
A. MOTIVATING EXAMPLE	12
B. EXPERIMENTAL RANDOMIZATION	15
<i>Compliance Issues</i>	17
C. QUASI-RANDOMIZED INTERVENTIONS	17
D. OBSERVATIONAL DESIGNS	20
<i>Selection into Treatment Based on Observed Characteristics</i>	20
<i>Propensity Score Matching</i>	21
<i>Selection into Treatment Based on Unobserved Characteristics</i>	23
E. CONCLUDING THOUGHTS ON DESIGNS AND METHODS	30
V. MEASUREMENT AND DATA COLLECTION ISSUES	31
A. ADAPTING STANDARD SURVEY QUESTIONNAIRES	32
B. MEASURING PROPERTY RIGHTS AND IMPACTS	32
<i>Gender and Intra-household Allocation Issues</i>	34
<i>Subjective Measures of Property Rights Security</i>	34
C. SPILLOVERS AND GENERAL EQUILIBRIUM EFFECTS	35
<i>Packages of Reforms</i>	36
D. DIFFERENTIATED IMPACTS	36
CONCLUSIONS	38

Introduction

A large number of land property rights reforms (LPRRs) are taking place around the world today. While the news headlines may continue to concentrate on the often more conflictive redistributive land reform programs in places such as Zimbabwe, South Africa, Brazil, or Venezuela, a larger quiet revolution to transform property rights systems has also been taking place through a myriad of other perhaps less glamorous but nonetheless far-reaching land regularization and titling programs as well as other efforts to upgrade and transform land administration systems (Box 1 describes a typology of such interventions).

Governments and multilateral aid organizations have steadily increased their funding commitments to such programs over the past decades. The World Bank alone committed nearly 1 billion dollars in fiscal year 2004 to land administration, land titling, and other land reform projects, a significant real increase over its activities a decade prior (World Bank 2005; Holstein 1996; USAID 2005). This rising activity has been driven by a variety of factors. At a most basic level, it represents a struggling and delayed response to the inexorably rising demand for services by citizens, particularly where rising population density and/or new market opportunities have led to increased competition for land and upward pressure on land values. In such contexts, citizens demand clearer property rights to secure a place to live, to protect and recoup the value of their existing holdings and investments, or as a mechanism to help reduce the probability and cost of disputes and externalities. Better land administration systems are also demanded by local governments and neighborhood associations as they struggle to keep up with the pace of new settlement or the transformation of existing areas and the need to deliver roads, water, sanitation, schools, green spaces, and other local public goods. On the supply side, the availability of new hardware and software technologies have also helped to spur the rising tide of activities as many mapping and land registration solutions can now be implemented in new more decentralized and cost-effective ways.

The push to focus on land property rights has also been promoted by the influence of big ideas. Popular press books such as Hernando de Soto's *The Mystery of Capital* (2000) as well as the analyses of other economists and historians such as Douglass North (1990) have highlighted the central role of property rights systems and other institutions in shaping the long-run economic development outcomes of nations.² Some have interpreted the lessons of these studies as pointing to an urgent need to modernize or re-design property rights systems to strengthen the protection of property rights in general and, quite often also, to extend and better recognize the property claims of the poor (HCLEP 2006). In the view of many proponents, increased security and clarity of property rights should lead to a number of impacts including the reduction of costly conflicts, better incentives for agents to invest, increases in the supply of credit, new transactions on land and labor markets, and generally higher levels of economic activity and revenues for local government investments in public goods. More clearly defined and

² See also Sokoloff and Engerman (2000) and Acemoglu, Johnson and Robinson (2001).

securely enforced property rights over land are also prescribed as a solution to a range of environmental problems such as overgrazing, soil erosion, and the over-exploitation of forests.

As important and urgent as these reforms may seem to some, they have been neither always-successful nor uncontroversial. Although property rights reforms are often packaged in the language of providing access to land and empowerment of the poor, critics have worried that, like so many other planned projects of the past, this latest bandwagon may end up doing more harm than good if not carried out right (Easterly 2006). The last great period of world-wide property rights reforms in the late 19th century were often sold using similar explanations about the potential benefits of more modern property systems on access to land by new small property holders and its role in expanding markets and growth (Scott 1999; Lauria-Santiago 1999; Swinnen 2000). Many of these reforms did work to expand property ownership and citizen rights in parts of North America and Europe by recognizing and strengthening property rights of tenants or frontier settlers. In other parts of the world, however, the reforms of this era are instead often remembered for having usurped or undermined rather than strengthened the property claims of the poor and the indigenous (Binswanger, Deininger, and Feder 1995). More recent reforms to transform property rights in parts of East Africa in the post-colonial period, again under the ostensible purpose of modernizing the economy and helping small farmers, are also viewed by many as having failed (Berry 1993; van den Brink et al. 2006; Scott 1999; Firmin-Sellers and Sellers 1999).

In short, while some property rights reforms have brought important successes, in other cases the anticipated beneficial impacts of reforms were eventually overshadowed by sometimes disastrous unintended consequences. The latter often occurred where governments attempted to impose new property rules that failed to recognize and incorporate existing community-sanctioned arrangements and/or where the implementing agencies were too weak to avoid capture by elites and/or too unaccountable to respond to citizen demands when programs were not working.

Today's land property rights interventions claim to be different. Programs will supposedly now be designed to involve and elicit community participation. Monitoring, accountability, and feedback mechanisms are to be incorporated into program designs to make sure that programs meet their targets and are not just imposed by bureaucrats from above or captured from below by local elites and rent-seekers. Pilot programs and regular impact evaluation studies will be carried out to measure intended and unintended consequences and to experiment, learn, and make adjustments before scaling up costly program interventions.

This seemingly rational approach would represent significant progress if only one were able to demonstrate that this was actually happening. Unfortunately, that cannot yet be done. Although tens of billions of dollars have been invested into land property reform projects, to date there is not, to the best of our knowledge, one successfully completed impact evaluation study of a land reform intervention using a rigorous study design built into the program design where comprehensive measurement and appropriate, modern statistical methods were used. The few studies that have tried to measure the impact of

past property rights reforms have usually been carried out after the fact using retrospective data or, in the best of cases, using surveys tacked on half-way through project implementation. Table 1 summarizes the empirical approaches employed in a selection of existing studies, many of which are described in further detail below.³ This dearth of rigorous socioeconomic impact evaluation based on randomization or on a pre-post design may be attributed to a variety of reasons including late recognition of the value and the practicability of integrating impact evaluation studies into normal project planning activities but perhaps partly also to some unique challenges that arise in evaluating the impacts of property reforms.

The purpose of this paper is to describe some of these challenges as well as survey designs and methods for purposeful impact evaluation of land property rights reform projects. Typically, impact evaluations seek to obtain reliable estimates of the average effect of the program intervention, i.e., on average, the change in the outcome of interest between a household that receives the intervention compared to an otherwise identical household that does not receive the treatment. For example, does titling increase the value of household investments in improving the land or buildings compared to similar households on land with similar characteristics without titles? In the language of evaluation, measures of such effects are known as the Average Treatment Effect (ATE) of an intervention and will be the focus of our analytical approach in this paper.

This paper presents a practical approach to evaluation of programs and should be accessible to non-specialists interested in impact evaluation. It is not intended to be at the frontiers of the research literature in the field, which is quite active but also technically quite challenging. We take a fairly optimistic view of the possibility of rigorous evaluation without randomized design provided deliberate alternative data collection strategies can be pursued. For a more comprehensive and critical examination of non-randomized program design, we refer the reader to Duflo and Kremer (2003). The statistical methods we recommend are relatively straightforward to implement but are not necessarily the most general methods available, which can be much more complicated to implement. For a sophisticated, critical examination of the available statistical methods, we refer the reader to Ravallion (forthcoming). Finally, we focus exclusively on the ATE, which is not without technical issues. Interested readers should refer to work by Heckman (2001) and Angrist (2004) for discussion of these issues.⁴

Impact evaluation presents many challenges in general, and several challenges specific to land property rights reforms.⁵ Almost all existing studies to date have been

³ Several good reviews exist of the literature on property rights security and investment incentives. Feder and Nishio (1999) highlight findings from an early literature and provide a nice summary of some of the expected benefits and costs of titling reforms, while Besley (1995) identifies issues that have been at the center of more recent research. Pande and Udry (2005) provide an excellent recent survey of retrospective studies including summary tables of the methods employed in approximately thirty different studies.

⁴ The frontier of the literature on program evaluation discusses the pros and cons of the ATE as compared to other measures; e.g., the Average Treatment Effect on the Treated (ATET) and the Marginal Treatment Effect (MTE).

⁵ Baker (2000) provides a very useful general guide to impact evaluation in developing countries. This short book includes sample terms of reference and evaluation findings for a number of World Bank financed projects.

based on observational data where reliable comparison groups for those receiving the program treatment are difficult to identify because of non-random program placement and self-selected beneficiary households. For this reason the results of such studies have often been open to challenge or have remained controversial because the doubt persists that they may have confounded estimates of program impacts with existing differences between households in treatment and comparison groups. But if data collection is built into program design – a relatively easy task that may produce several other program benefits – then more reliable estimates of treatment effects can be obtained post implementation of the program.

Impact evaluation studies based on a baseline and follow-up surveys allow for far more reliable estimates of program impacts than studies based on simple cross-section of retrospective data, even in the case of randomized designs. For this reason impact evaluation studies should be designed as much as possible along with the project interventions themselves. Researchers, project leaders, and beneficiaries should ideally interact as much as possible to collect data which should be collected, if possible, before, during, and after program interventions. Such a design can create an interactive environment where researchers and project leaders can gain deeper knowledge of beneficiary behavior as well as of the institutional and social characteristics of the communities where programs are being implemented. This will prove useful to allow better identification of expected and unexpected impacts, as well as in generating practical lessons to inform program adjustments for better implementation.

Box 1: Types of Land Property Rights Reforms

We use the term Land Property Rights Reforms to encompass a variety of different types of interventions that have the ultimate aim of altering and/or enforcing property rights relationships over land. Adapting terminology laid out by FAO (2000), these interventions might be classified as follows:

Land Administration, Land Titling and Other Tenure Reforms including land registration and titling projects (regularization of holdings, adjudication, title issuance, registration). These interventions are often accompanied by cadastral surveying and mapping, and land settlement components and infrastructure improvements.

Imposed Redistributive Reforms refer to traditional redistributive land reforms with or without compensation to owners (e.g. Korea, Japan, Mexico, Peru, Chile) as well as tenancy reforms (e.g. India, many Western European countries). We also include here third-wave reforms that often transform earlier land reform properties into more transferable titles (e.g. Mexico).

Negotiated or Market-led Land Reforms are similar to redistributive reform but seek to match a willing buyer to a willing seller (e.g. South Africa, some programs in Brazil, Colombia). Leads to different selection of beneficiaries and types of properties.

Land Reform via Restitution refers to the return or re-assignment of properties to private owners that may have been seized by a state (e.g. reforms in Eastern Europe).

We motivate our analysis with several variants of a simple constructed example of a land titling program in an urban setting. We illustrate in simple visual examples how the distribution of observable and unobservable characteristics of treatment and comparison group *samples* might change according to the nature of the program

intervention and treatment selection rules (e.g. how the project targets geographic areas or population groups, whether and how households are allowed to self-select, etc.). This visual approach focuses attention on the key importance of survey design and data collection strategies to avoid confounding effects, and eschews a good deal of the math usually required to present these issues. Most methods for impact evaluation analysis can be explained as strategies to anticipate and adjust to these sample selection issues and as efforts to maintain a balance between observable and unobservable characteristics in treatment and comparison groups.

The rest of the paper is organized as follows. The next section asks the essential question: why carry out an impact evaluation at all? Given the conspicuous absence of impact assessments this seems an important question to answer. We then delve into a brief discussion of some of the possible expected benefits and costs of property rights reforms, factors that affect the timing and placement of property rights reforms, and the main challenges associated with impact evaluation of property rights reforms. Next is the main section of the paper on study designs and sampling methods, illustrated by instructive examples of an urban titling project. Having laid out these main empirical challenges and proposed solutions, the penultimate section of the paper returns to a more in-depth discussion of the kind of hypotheses that might be usefully posed and tested with an impact evaluation study, as well as the types of data that can and should be collected. A final section concludes.

I. Why an Impact Evaluation?

There are different types of evaluation. *Direct program impact assessments* aim to measure how well a program is meeting its operational targets, for instance the number of titles distributed, the number of poor families reached, the cost per title distributed, etc. *Qualitative* and *participatory assessments* are another essential element of program assessment (Rao and Woolcock 2003). These may include such tools as public meetings with beneficiaries and stakeholders, and qualitative interim evaluations conducted by knowledgeable national or international observers based on field visits, focus groups and interviews. A third type of evaluation, and the main focus here, are purposeful *socio-economic impact evaluations* based on the collection of survey data and quantitative analysis. Each of these types of evaluation can serve as an essential complement to the others. In what follows we shall reserve the term “impact evaluation” to refer to evaluation studies that construct a carefully chosen comparison group or counterfactual in order to estimate a quantitative measure of program impacts.

Why should impact evaluation studies be an integral element of program design? A number of very practical reasons can be offered, including the following:

Generating Useful Questions: The *process* of carrying out an impact evaluation can be as valuable as the results themselves because talking in-depth about how a program will be evaluated focuses debate on expected outcomes and the pathways to achieving those outcomes. An impact evaluation requires careful thinking about the nature and timing of program interventions and about the causal chain from inputs

(money and resources) to likely and unexpected impacts as well as about other affected groups. These are the types of questions that program planners and implementing agencies should already be asking themselves, but do not always get raised systematically. Planning for and carrying out a baseline socioeconomic survey, particularly when it is done with the involvement of local independent researchers and participatory information gathering to involve direct stakeholders, can generate knowledge about the working of household and community economies that might have otherwise not been understood by planners and program officers. This can lead to critical examination of project design for improved interventions that better match community demands.

Immediate Monitoring Indicators and Design Improvements: The collection of data for a baseline survey can lead to immediately useful monitoring indicators and practical lessons for improving the program interventions *ahead* of program implementation and that will serve in the construction of useful and realistic socioeconomic objectives. For instance a rural survey questionnaire complemented by participatory discussions with rural village assemblies (which can be incorporated explicitly into the design of a baseline survey to gather community-level data and to build greater community understanding and support) would generate valuable information on how to lower program costs, as well as about the demand for different levels of tenure security, and other services. This helps to contribute to a culture of professionalism, open debate, accountability and project ownership.

Pilot Study Assessments before Scaling Up: Governments and development programs have been criticized for top-down design and bureaucratic execution of their interventions. Interventions to register land in rural areas may seem overly expensive given a country's many other pressing needs. Are these concerns legitimate, and if so can they be addressed by better program design? Do the benefits of the program intervention exceed the costs? Are there unintended benefits or costs? Do the interventions undermine tenure security for the weak and vulnerable as critics might contend, or strengthen security as the program intends? Are the benefits of interventions more or less uniform across program areas and individuals or can differential impacts be predicted to construct indicators that might help to better target or sequence the program interventions to reach areas where demand or need is highest?

These are all important and legitimate questions that should be asked about any intervention which is why impact evaluations are increasingly being built into projects (Baker 2000; Ravallion forthcoming; The Economist 2005). The day may soon arrive when project proposals that fail to include rigorous monitoring and impact evaluation will be routinely rejected.

Experiments for Program Design Innovation: Simple variations on interventions can be randomly assigned to program areas to generate information to help judge whether adding these program features will be cost effective or not on a larger scale (Duflo 2004). More specifically, a test of the hypothesis of whether a suggested program feature such as adding a land use management component to a village lands certification project is an

essential complementary intervention or not can be run at relatively small incremental cost before a project has had time to be rolled out to new areas.

Credible Evidence to Advance Public Debates and Overcome Political Resistance: Property rights reforms have remained stalled in some countries due to the lack of a properly informed public debate aimed at overcoming often legitimate apprehensions and concerns over the reasons and nature of the reforms. Debates over property rights are often highly politicized. In such charged atmospheres many actors and would-be advocates or beneficiaries often adopt a ‘wait and see’ attitude. Such concerns are not likely to be addressed by “internal reviews” within a ministry or by retrospective qualitative studies, and these are hardly the basis for good policymaking in any case. As one South African Land Affairs official recently despaired at a land conference, the evidence available to guide policymaking on the critical and very sensitive issue of property rights reforms in his country – after years of land initiatives – still added up to little more than “assertions, impressions, ad-hoc consultancies, newspaper clippings, and anecdotes (Mbongwa 2005).”

More systematic impact evaluations carried out by qualified independent local researchers promise to create the hard data and evidence around which a more informed public debate can be framed and which could move policymaking forward either by helping to demonstrate program effectiveness or by helping policymakers back away from, or fundamentally redesign, programs that are clearly failing.⁶ Impact evaluation can also be an important complement to other monitoring frameworks, for instance to assure that program interventions are being targeted to reach their intended beneficiaries.⁷

For all of the above reasons, evaluations are becoming increasingly important for the economic and political sustainability of programs, and are more and more required to make the case for continued financial support for programs in the face of tight budgets and competing needs.

II. Expected Impacts of Land Property Rights Reforms

In this section we present a brief initial summary of some of the claimed expected impacts of land property rights reforms, focusing in particular on the impacts of land titling programs. This will provide an entry point for discussing the nature of property

⁶ On this van den Brienk et al. (2006) write: “controversy should not become an excuse for inaction...the optimal way forward is to agree on a policy framework which allows a menu of options to be pursued, the results of which can then be evaluated as the program proceeds, and corrections made when ex post evaluation shows some negative aspects. Rather than endlessly debating the pro’s and con’s of each particular approach, we propose to create a policy arena in which the particular models can show their relative performance in competition with each other (p. 47).”

⁷ The failure to build transparent monitoring and evaluation procedures into programs increases the chances of political capture. A study of Cameroon’s 1974 Lands Ordinance found that over 82 percent of all titles intended to create a new ‘middle class’ of small and medium farmers ended up being assigned to state elites (politicians and bureaucrats) and connected businessmen and only slightly more than 3% of titles in one large study area where assigned to women (Firmin-Sellers and Sellers 1999).

rights and how property rights reforms tend to come about and become targeted to specific communities. The manner in which these processes occur is what defines many of the unique challenges involved in identifying the impact of such interventions. Later sub-sections will then discuss research designs and methods that may be used to address some of these evaluation problems, using land registration and titling reforms as a motivating example. Later in the paper, after we have discussed these designs and methods, we shall return to a more complete discussion of other types of land-related interventions and expected outcomes.

All societies have institutions to define and regulate the use of resources including property rights rules and norms and conventions of behavior (North, 1990). Together these define and delimit the set of privileges and obligations individuals may have over the use of specific resources and assets such as a parcel of land (Libecap, 1989). Property rights institutions may range from formal arrangements codified in statutes and laws to more informal conventions, customs and obligations upheld by local community norms and traditions. Property rights systems are however also constantly evolving and being adapted to changing constraints and opportunities brought about by such forces as rising population densities, new technological opportunities and markets, as well as changes in politics and relative bargaining power (Boserup 1965; Binswanger, Deininger, and Feder 1995). There is no presumption however that property rights will always evolve in the most efficient or fair manner, and property rights systems may at times become dysfunctional or too weakly enforced to serve their purpose. In those situations property rules may fail to provide the right incentives leading agents to forego or to fail to discover valuable opportunities for investment, trade and natural resource conservation and possibly also may lead to excessive and wasteful forms of contestation or conflict over resources.

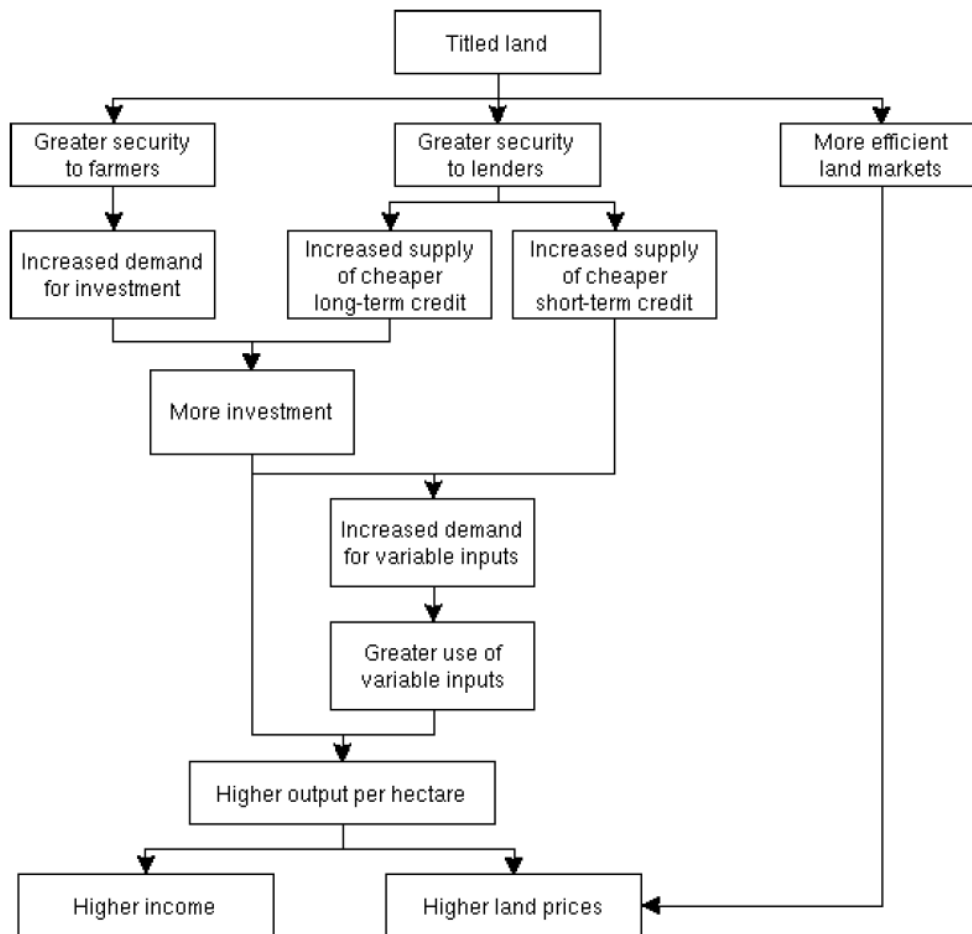
The aim of land property rights reforms is to attempt to strengthen or transform existing property relations to avoid such situations and to achieve more efficient and/or equitable outcomes. Simplifying vastly, the case for government-sponsored land registration and titling reforms has typically been made on the claim that establishing more clearly delineated and secure property rights can lead to the following expected effects (Besley 1995; World Bank 2003; Feder 1999):

- ***Investment Demand Effect:*** the claim that with more secure titles households and communities will have greater incentive to invest without fear of challenge or expropriation. This should increase investment, raising productivity and cashflows, which in turn should stimulate incomes, land values and general levels of economic activity. Reforms may also improve community abilities and incentives to invest in local public goods including environmental protection;
- ***Credit Supply Response:*** the claim that financial services will be expanded both in response to expected increased cashflows as well as because reforms that increase transferability may expand the use of land as collateral;
- ***Gains-to-Trade Effect:*** the claim that more secure and clearly defined property rights will lead households to expand the volume of land lease and sale transactions, leading to a more efficient overall allocation of resources and output. In some circumstances the activation of land lease markets may expand land access to the landless poor;

Note that, depending on circumstances, improvements in the *security* of possession might be conferred under either community or privately managed property systems. Note also that increased *transferability* is neither a synonym nor a necessary corollary of increased security – there are reforms that may increase security of possession without improving transferability, and vice-versa.

Pagiola (1999) summarizes several of these expected relationships in a rural setting with the following diagram, which is also similar to a diagram by Feder and Nishio (1999) , which we display in Figure 1a.

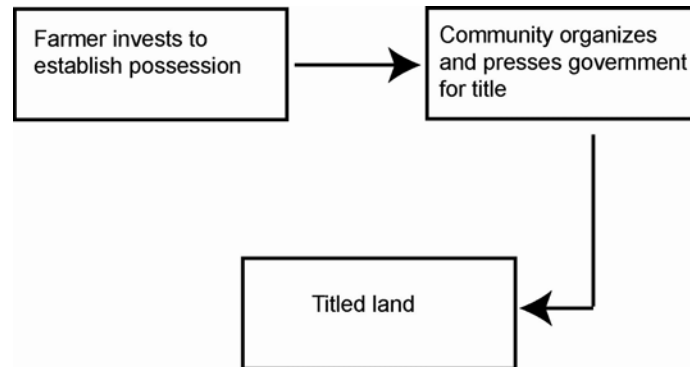
Figure 1a: Standard Presentation of Expected Impacts from Titling



These are plausible causal relationships. But the real challenge of identifying the implied impacts has as much to do with what the diagram does *not* indicate as with what it does. Notice that the diagram begins at a node labeled ‘Titled Land’ suggesting that the pathway to impacts begins with an exogenous intervention whereby a previously untitled farmer is issued a ‘title.’ In fact the issuance of title is very rarely the beginning of the relationship between a farmer, his community, the state, and his or her land rights. Most land registration and titling programs provide title primarily to households that have already made costly investments that position them to be more likely to become

beneficiaries. We could have elaborated upon the diagram above by adding the following possible pathway leading up to the ‘titled land’ intervention.

Figure 1b: The Likely Endogeneity of Title Placement



Note how this simple extension suggests a possibly quite different relationship between ‘title’ and plot-level investments. Whereas the original diagram in figure 1 suggested that more secure title should lead to an increase in investment, the extended diagram opens the possibility that households might have invested to establish and defend possession (e.g. the building of structures and fences, the establishment of residency to build up legitimacy and relationships within the community, etc) in order to increase the probability of eventually receiving title. If this is the case then the establishment of more secure title, enforced now via the courts rather than mostly through the private efforts of the farmer, could very well lead the household to *reduce* plot-level investments (Besley 1995).

This example raises another important, related complication. Farmers will typically differ in terms of the length of their existing possession of land, the level of security they enjoy under informal community sanctioned property rights regimes, and how well organized their communities might be at solving collective action problems, including lobbying the government to have their communities titled ahead of others. These are important observed and unobserved, household and community characteristics that may confound the estimation of intervention impacts. For instance, higher measured outcomes for households that received title might be explainable by the titling program having had a large impact and/or simply by the fact that the most enterprising farmers in the best organized communities who would have had higher outcomes even in the absence of any titling effect, are more likely to have applied for title.

III. Impact Evaluation Challenges

The aim of the impact evaluation study is to measure the average impact of the program intervention on a broad array of community, household, and/or individual outcome indicators compared to outcomes under the counterfactual of no program

intervention. Since households obviously cannot be both treated and not treated by the program at the same time, actual counterfactuals can never be observed. Instead we must attempt to estimate the value of this unobserved counterfactual by obtaining a measure of the value of the variable(s) of interest in a comparison group of untreated households. A carefully designed and deliberate sampling strategy will be required to ensure that the chosen comparison groups provide as good of a representation of the unobserved counterfactual as possible.

Finding an appropriate comparison group is not an easy task. If we are not careful about how we collect data we might use a comparison group that has a quite different distribution of observed or unobserved characteristics amongst its members compared to the members of the treatment group. We then run the risk of incorrectly attributing differences in measured outcomes between the treated and untreated to program impacts when in fact they are mainly due to initial differences between the two populations.

The ideal method for constructing a comparison group to avoid this problem is to assign the program intervention or ‘treatment’ randomly across households and then compare outcomes in treated and untreated households. If treatment is assigned randomly then the distribution of observed and unobserved characteristics can be expected to be the same across treated and untreated households. A random sample of untreated households then forms an ideal comparison group for the set of treated households. A simple comparison of mean outcomes between treated and control observations would then provide an unbiased measure of program impacts.

A. Non-random Program Placement

Property rights reforms are typically not assigned on a random basis, both because program placement tends to follow a non-random criterion and because beneficiaries often have a choice whether to participate or not. For example urban property formalization or titling programs, are very often first targeted at longer established or better organized neighborhoods, perhaps because those neighborhoods have been most vocal or proactive, or because those are the areas where program officers believe they will be most clearly able to demonstrate immediate results to justify future program expansion. In some contexts there may be pressures to target areas where incomes or other outcome variables are expected to be lower, i.e., to prioritize poorer neighborhoods.

Purposeful program placement or beneficiary targeting of this sort may serve important social or political objectives, but it greatly complicates evaluation because the characteristics of the treated and untreated comparison groups will typically differ considerably. A simple comparison of mean outcomes will then be very likely to provide a biased and misleading measure of program impacts.

For this reason it will be important for those who carry out the impact evaluation to try to carefully understand the manner in which the programs are being targeted and to adapt evaluation methods, survey design and the timing of data collection to control for the effects of targeting decisions as carefully as possible. It will also be extremely important to collect as much data as possible about outcome-relevant variables to control

for observable differences wherever possible. For example, in rural projects it might be very important to collect data on plot-level soil conditions, on distance to markets and weather shocks, and such variables as household and individual membership and participation in kin-based or social networks.

B. *Participants Self-select*

Even if program placement could be randomized across geographic areas, many types of programs are demand-driven within a locality in that they will provide treatment only to those households or individuals that apply or those who pay a user fee. For instance, a rural Village lands certification program in Tanzania formalizes village boundaries and supports the establishment of village land registries but individual certificates of customary residential occupancy are granted only to households that take the steps to apply for and obtain approval from the Village assembly (Government of Tanzania 2005). This design may be very good economic policy but it also means that the individuals who obtain title within a village are likely to be quite different from who do not apply. Those that are more likely to apply might also be those who might be expected to have higher outcomes even in absence of the program. A simple comparison between those who receive and do not receive certificates would then clearly falsely attribute differences in outcomes to the program when in fact these measured differences would be partly or even mostly due to a failure to control for observable and unobservable differences between treatment and comparison groups. Households that take up the program might, for instance, have more political clout within the village, be more enterprising or entrepreneurial, have better access to credit, etc.

IV. Study Designs and Evaluation Methods

In this section we describe a variety of possible study designs, starting with the ideal and ending with the most challenging or troublesome designs, from an evaluation point of view. To fix ideas, we begin by constructing a simple hypothetical example of a land titling program to illustrate a set of challenges that an analyst might face in a range of other real program evaluation contexts. We describe study designs and appropriate statistical methods for estimation of the effect of treatment in each of these circumstances.

A. *Motivating Example*

In the sections that follow we employ variants of the following simple constructed example of estimating the impacts of a hypothetical urban titling program. As depicted in figure 2 the greater urban area is divided into four neighborhoods (or ‘zones’ for short) labeled I through IV. Within each neighborhood, households differ according to both observable and unobservable characteristics which will affect the measured outcome of interest. We assume that outcome variable y (e.g. household income) in household i in the post-intervention period is determined by the following simple relationship:

$$y_i = \beta_1 t_i + d_i + e_i$$

where d_i is an *observed* characteristic of the household (e.g. the education or degree of the household head, etc), e_i is an *unobserved* (to the researcher) characteristic such as entrepreneurial drive, ability or effort, and t_i indicates an indicator those who participate in the titling program ($t_i = 1$) or not ($t_i = 0$). In all the examples below we will assume that the true impact of treatment is $\beta_1 = 1$ but that this value is unknown to the investigator. The challenge of impact evaluation is to form an unbiased estimate of β_1 from sample data that reveals y_i , t_i , and d_i (but not e_i) for each household in the sample.

To simplify further, suppose that d_i and e_i are binary variables which can only take on values 0 (high) or 1 (low). Figure 2 summarizes the four types of household and corresponding eight values of the outcome y_i that will be found in the data depending on whether the household received treatment or not. Observed degree is denoted by the mortarboard (hat) and high unobserved effort or enterprise by the holding of a rake. Households that receive treatment are shaded, while comparison households are not.

Figure 2: Household Characteristics and Outcomes









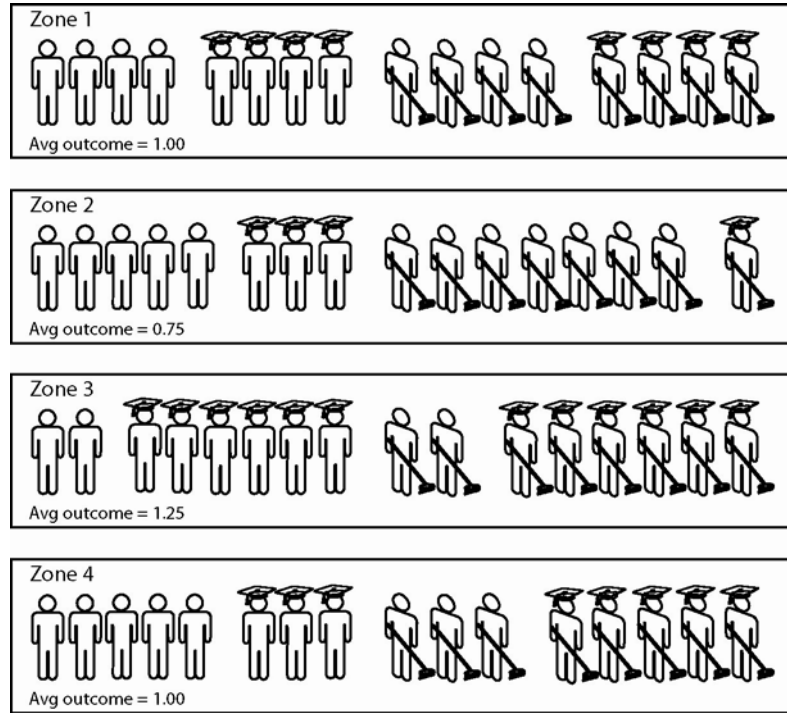
		Outcome $Y_i = t_i + d_i + e_i$			
		Comparison ($t_i = 0$)		Treated ($t_i = 1$)	
Observed Degree	Unobserved Effort	Symbol	Outcome	Symbol	Outcome
Low ($d_i = 0$)	Low ($e_i = 0$)		0		1
High ($d_i = 1$)	Low ($e_i = 0$)		1		2
Low ($d_i = 0$)	High ($e_i = 1$)		1		2
High ($d_i = 1$)	High ($e_i = 1$)		2		3

Figure 3 presents the assumed distribution of characteristics across a city population of 64 households residing in 4 neighborhood zones of 16 households each. The number reported at the bottom-left corner of each box, representing a zone, is the average value of the outcome in that zone. We will keep to this as an illustrative example in the sections below.

Figure 3: Pre-treatment Distribution of Household Characteristics and Outcomes



There is an equal number of households of each type in the city, i.e., the distributions of observed and unobserved characteristics are balanced across the city. This assumption is made without loss of generality. On the other hand, the distribution of households in each zone is *not* balanced on the observable characteristic d . Zone 2, for example, is assumed to have a three times higher concentration of low d households compared to Zone 3, i.e., Zone 2 is a neighborhood characterized by lower educational attainment and socioeconomic status. To keep the examples below simple we assume that the distribution of unobserved ability is balanced across zones (i.e. the ratio of rake-wielding to non-rake wielding households is the same within each zone). This assumption may not always hold in practice; households in some neighborhoods may have higher average unobserved effort than households in other neighborhoods. Unbalanced unobserved heterogeneity of this sort would complicate the statistical analysis. In some cases, the required adjustments to the proposed methods below will be relatively simple, in others less so. Finally, we assume that the distributions of observed and unobserved characteristics are independent of each other, which is not important for our examples but can be an important statistical assumption in the context of regression models.

By assumption, the true impact of a titling program would be to raise household income by $\beta_1 = 1$, in all types of households. The question to be explored in each of the sections below is how to correctly estimate this average treatment effect (ATE) by comparing mean sample outcomes of treated and untreated households. We must take care to not attribute differences in outcomes to the effect of program treatment where

they can instead be explained by differences in the distribution of observed and unobserved characteristics in the two respective samples.

As mentioned earlier, the ideal design is a controlled randomized experiment, and this is the design we describe first. Sometimes, when pure randomization is not feasible, it may be possible to implement quasi-random designs. Generally, however, when randomization is not feasible, assignment into treatment is self-selected. Such designs are labeled ‘observational’; we describe variants of such designs next.

Although we shall illustrate each of the cases below with simple variants on this constructed example, the analysis can be extended very easily to a more general model. In most studies the outcome variable is assumed to be given by a more general linear relationship of the form

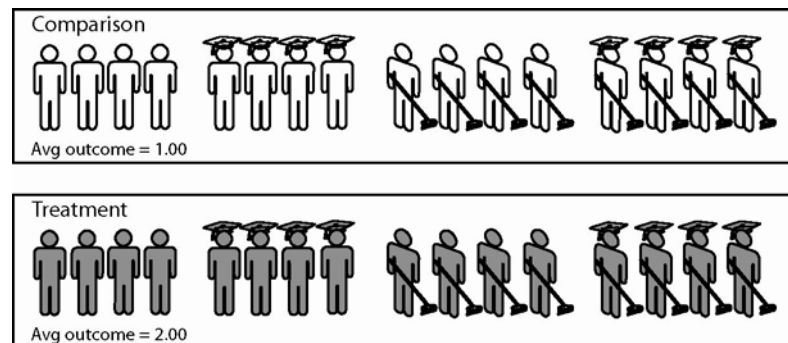
$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \varepsilon_i$$

where X_i is a vector of observed household characteristics and ε_i is a classical regression error term which includes unobserved effort e_i . We use this specification, which implies a constant program effect, as the benchmark for each method discussed below, but later also extend this specification to allow for program effects to vary by household characteristics.

B. Experimental Randomization

Consider experimental randomization into a tiling program in this context. In the simplest implementation of such a design, households would be selected by random lottery from anywhere in the town. For each household assigned to receive the program intervention, another household is randomly selected to be in the control group. Then, a simple comparison of mean outcomes between the treated and untreated samples some time after the program intervention, provides an unbiased estimate of the average treatment effect (ATE). The figure below indicates two representative random samples drawn from treated and untreated groups. A simple difference of sample means would provide an unbiased estimate of program impact.

Figure 4: Random Assignment



In more sophisticated randomization schemes, designed to increase the precision of the estimates of program effects for a given sample size, the pool of agents could be

stratified by key observable characteristics, and then treatment assigned randomly within each strata. In the context of our example, one more sophisticated scheme simply entails selecting households randomly within each neighborhood and within each educational attainment group, i.e., for every household selected to receive the program intervention, select a household from the same neighborhood and educational group to be in the control group.

Although not particularly valuable in this particular randomized assignment context where the simple comparison of means is sufficient, one could also estimate the ATE by estimating the coefficient β_1 in the regression specification below via ordinary least squares.

$$y_i = \beta_0 + \beta_1 t_i + \varepsilon_i$$

One could then trivially incorporate covariates, X_i into the regression specifying it as

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \varepsilon_i.$$

To the extent that the covariates X_i affect outcomes, incorporating them into the model provides a more precise estimate of ATE, i.e., there is an efficiency gain. In our example, X_i consists of measures of household and neighborhood characteristics. Unobserved ability is in ε_i , but that is uncorrelated with assignment to the program treatment or control groups.

If the treatment effect is expected to vary by socioeconomic status and/or neighborhood characteristics, the regression framework also provides a natural way of incorporating differential effects. Consider the regression:

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \beta_3 X_i t_i + \varepsilon_i$$

which allows for the treatment effect to vary by socioeconomic status. The ATE for a household with characteristics X_i is given by $\hat{\beta}_1 + \hat{\beta}_3 X_i$. If all the covariates are indicators, ATE estimates for each socioeconomic group are identical to those which would be obtained if stratified sample means were used to calculate differences between treated and untreated groups. When one or more of the covariates is continuous, the regression approach provides a way to estimate varying ATEs when the stratification approach is no longer useful.

An important feature of our baseline example is that the distribution of unobserved characteristics within each zone is the same, i.e., there are as many high effort and low effort households in each zone. Therefore, random selections of households within each zone produce, on average, sets of households with the same average effort level. Random selections of households across zones would also produce sets of households with the same average unobserved effort. In fact, a much simpler randomization device is possible in such a scenario. One could simply choose two of the zones to be assigned to treatment with the two remaining zones being the control group.

Any imbalances in observed covariates across zones could be controlled for either by the regression methods described above, or by more sophisticated matching methods, including propensity score matching methods which we describe below. But if the distribution of unobserved ability is unbalanced across zones, then a scheme that selects entire zones for assignment into treatment may not produce groups balanced on unobserved characteristics. The consequences of such imbalances are much more difficult to overcome via statistical methods. Therefore, if randomization is indeed feasible, randomization of households within neighborhoods is preferred to a random selection of neighborhoods where balance on unobservable characteristics may be harder to achieve.

Compliance Issues

While experimental designs are often considered the ideal method for program evaluation, pre-program randomization does not guarantee that randomization will be retained post-program for a variety of reasons. First, some households assigned to the treatment group may choose not to take up the program, perhaps because in typical programs, the receipt of treatment may not be costless, or because they fail to complete the required paperwork. In this case, it may be reasonable to consider an *intent-to-treat* analysis, i.e., to treat the take-up issue as part of the state of nature and thus estimate a treatment effect by comparing those who were *targeted* to be in the treatment group with those who were in the control group. Indeed, the intent-to-treat estimate may be the measure of impact with external validity, i.e., one that would be observed if the program were scaled up or replicated elsewhere. Second, some households assigned to the control group may receive the program treatment through alternative sources, an issue sometimes known as contamination. In this case too, an intent-to-treat analysis may be of interest to provide a lower bound for the true program impact. Thus a significant intent-to-treat effect would be informative. An insignificant intent-to-treat effect would, however, not provide any insight into the significance of the true program effect.⁸ Third, households slated to be in the experiment may move out of the project area. This case poses a more complicated problem with no simple solutions. Such issues would have to be tackled on a case-by-case basis. These concerns can be quite important in practice. A study of a blanket-titling program in urban centers of Peru reported finding a number of households with COFOPRI (Committee for the Formalization of Private Property) titles in neighborhoods that had supposedly not yet been reached by the titling program, as well as a number of households without title in neighborhoods that had been supposedly reached (Field 2003).

C. Quasi-randomized Interventions

In many situations, experimental randomization is very often simply not feasible, often for political or ethical reasons. It may, however, still be possible to assign treatment in a way that might be considered *quasi-random*. For example, waiting lists may create quasi-random assignment to treatment when all households must be administered the

⁸ The true program effect may be estimated in these situations, but it requires more sophisticated statistical techniques which we describe below.

program intervention. Unless the program intervention is to be administered to all eligible households in a short period of time, some households will inevitably receive the treatment substantially later than others. In such a situation, a quasi-randomization device might be used to select households who receive the intervention early (the treatment group) and late (the comparison group). The design can then, under some circumstances, be effectively treated as if it were a randomized trial.⁹

Sometimes accidents of nature or geography or arbitrary program implementation decisions may create a quasi-randomization. In a classic early study of the impacts of property rights security Feder (1987) and Feder et al. (1988) compared household investments on plots of titled land to those on plots held by households in adjacent state forests where similar title could not be legally issued. According to the authors the assignment of lands to forest reserves was, from the standpoint of the cultivators, considerably arbitrary and unexpected, and therefore quasi-random. Reforms introduced in the 1980s meant that many farmers who had been cultivating land for years suddenly found all or part of their plots assigned to forest reserves where land could not be transferred or mortgaged, while other nearby plots ended up assigned to the other side of a boundary where they were not subject to the same restrictions.

Given the resulting quasi-random allocation of plots, farmers who cultivated titled plots on lands in areas adjacent to forest reserves served as a good group against which to compare farmers who found their properties ‘treated’ by the suddenly more restricted environment of the forest reserves. Feder and his co-authors found significant impacts. After the intervention farmers who had legal title outside of the reserves invested more and had better access to credit. Their plots yielded approximately 15-21 percent higher revenue per acre and sold at land prices between 25 to 152 percent higher (Feder 1999).

In an interesting recent study Galiani and Schargrotsky (2007) took advantage of a ‘natural experiment’ in title allocation to study the impacts of strengthened property rights security in an urban squatter settlement in Argentina. More than two decades ago, after a large group of squatters occupied suburban lands on the outskirts of Buenos Aires, Argentina’s Congress intervened to entitle the new occupants by passing a law to expropriate and compensate the former owners. Some of the original owners rejected the government’s compensation package however and instead chose to fight the order in the court, where the case remains to this date (The Economist 2006). As a result some of the original squatters received secure title but another group remains under more insecure tenure so long as the matter remains undecided in the courts. Although the decision by former owners to either accept or challenge the expropriation with compensation was not necessarily random (it might perhaps be explained by those owners’ characteristics) the authors argue that from the standpoint of the squatters the granting of title was an exogenous event. What matters are that the factors that affected the prior owners’ decision to accept or challenge can be assumed to not affect squatters’ subsequent behavior. Working with this assumption, the authors found that households with more

⁹ Issues of compliance will in general however usually be more pronounced under a quasi-randomization compared to true randomization. For instance, people will often find ways to jump ahead in a queue.

secure title had significantly higher levels of housing investment, reduced household sizes, and their children had better educational outcomes. They found only very modest effects on access to credit and no real effect on labor incomes.

The claim of having an exogenous or quasi-random allocation rules can sometimes come under challenge. In the above referenced study of Peruvian urban titling programs, Erica Field and Maximo Torrero carried out a ‘pipeline comparison’ taking advantage of the fact that large scale titling programs are sometimes interrupted or take several years to reach all intended beneficiaries (Field 2003; Field 2003, 2003). By surveying households while the titling program was still being rolled out as part of a staggered phase-in they were able to compare households in neighborhoods that had already received the program to households drawn from neighborhoods that had not yet received the program. This strategy will on average produce comparable groups so long as the researchers can argue that the decision to follow a particular sequenced choice of neighborhoods was not influenced by variables that might also affect measured program outcomes. The authors argued that this was the case by explaining that the sequencing of neighborhoods chosen could be explained as accidents of history or matters of simple expediency. As a check they compared variable means for several observable pre-intervention variables (malnutrition, literacy, school attendance, residential crowding, adequacy of roofing, access to water, sewage, and electricity, and demographic variables such as the age and education of household head) in treated versus untreated neighborhoods. Finding few significant differences in the means of these variables lead the authors to be more confident that the two groups seemed to be balanced on observables and therefore more likely to be comparable. They found a strong impact of title on market labor supply as well as household fertility.

A recent paper by Mitchell (2005) challenges the claim of this study that the treatment and comparison groups were in fact balanced on all outcome-relevant characteristics. He claims that the first neighborhoods in Lima to receive treatment were in fact more central, longer established, and in general less impoverished, and that a similar pattern can be found in the other cities to which the program spread. He argues that the first neighborhoods reached were part of a demonstration program and that the implementing group would have chosen neighborhoods where they expected the biggest impacts. Finally he argues that many of the households in the comparison group in the cities of Huancayo and Lima would have been refugees fleeing the rural conflicts of the late 80’s and early 90s. For all these reasons Mitchell suggests that the program impacts are likely overestimated. In particular the high measured female labor supply response may have been primarily due to the fact that labor supply was higher to begin within longer established neighborhoods.

This discussion points to the potential limitations of basing impact estimates on data obtained from a single cross-section. Below we describe how a pre-intervention baseline survey would have allowed the researchers to avoid several of these challenges and to be more confident of impact estimates.

D. Observational Designs

Although experimental or quasi-experimental designs are desirable for most program evaluations, it is still the case that political and operational realities often make such schemes infeasible. For example, local authorities might insist that households receive treatment on a first-come, first-served basis; or that the poorest households receive treatment first. In such cases, the data collected are of a purely observational nature. We anticipate such situations to be common in the areas of land titling and reform.

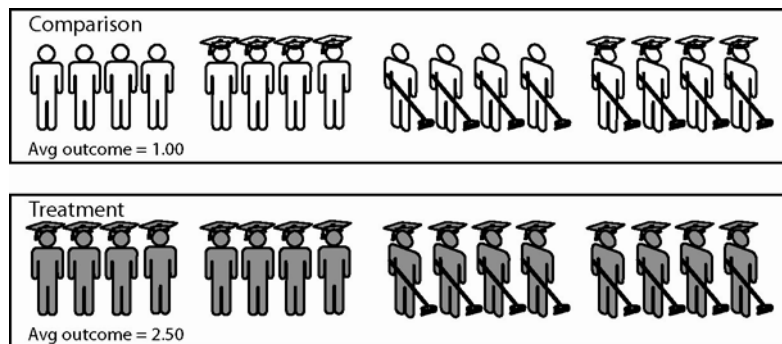
When this is the case, evaluation must rely on observational designs and there will now be concerns that estimates of program impacts might possibly be confounded with pre-existing differences in the treatment and comparison groups due to non-random program targeting and/or the self-selection of beneficiaries. It may still be possible to eliminate, or at least reduce, estimation biases that arise due to selection into treatment by a careful design of the data collection efforts. In what follows we distinguish between two types of selection mechanisms (possibly related) and describe data collection strategies and other strategies to address each.

Selection into Treatment Based on Observed Characteristics

As the last section suggests, when program placement can be assumed to be random or quasi-random, program impacts can be reliably estimated by a simple comparison of means of treated and untreated groups.

In many, if not most land reform contexts, beneficiaries will be self-selected. In some situations, selection into treatment may be influenced by observed characteristics of potential beneficiaries, but not by their unobserved characteristics. For example, in many programs beneficiaries must pay taxes or purchase fees to obtain title, or to convert newly acquired lands into functioning properties. Suppose $d=1$ is a prerequisite for a household's access to credit and only households with easy access to loan financing are able to access a titling program. In terms of figure 3 we assume now that only individuals with high X apply to the program. Figure 5 below indicates a typical set of samples drawn in such circumstances.

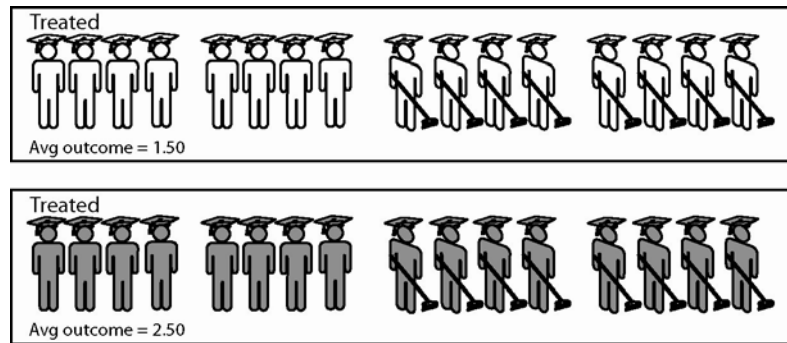
Figure 5: Selection on Observables (only $d_i > 0$ select into treatment)



The simple comparison of means in this example provides a biased estimate of 1.5 ($=2.5 - 1.0$), whereas the true program impact is only 1.0. The problem is that this estimator incorrectly attributes higher outcomes to the treatment effect which are instead due to the fact that the treated group contains a heavier selection of high d households relative to the comparison group.

How can we avoid this bias? The problem is that the treatment and comparison group are unbalanced on observables, a problem that could have been easily spotted by doing a simple comparison of means on the observable variable(s). In this case we would find out that the average value of d is 1.0 in the treatment group but only 0.5 in the comparison group. A better comparison can be obtained by building a matched comparison group. What we would like to measure are the difference in outcomes for households with treatment against households with the same (or very similar) observable values of d that did not receive treatment. In the simple case of just one observable variable d a matched comparison group is very easy to build. If we limit the selection of households into our comparison group to include only high d households then we are left with a simple comparison of means that yields an unbiased estimate of the project impact:

Figure 6: Matching on Observables



Note that we also could have calculated the difference in outcomes between each household in the treated group with that of a ‘matched’ household in the comparison group, and then averaged over those differences to arrive at our estimate of impacts.

The most common statistical approach to implement matching is on the basis of a regression model, as has been described above. In most applications, a linear functional form in X is specified, but more general specifications are possible. In the context of our example, a linear regression would achieve matching perfectly. But in general, a limitation of such parametric regression methods is that one must specify a functional form for the relationship between outcome and observed covariates. Other matching methods require fewer assumptions. We briefly describe the method of propensity score matching, which has gained in popularity recently, as it is easily implemented.

Propensity Score Matching

The intuition of matching methods is that, by comparing treated and untreated households with otherwise similar observable characteristics, one can more confidently

attribute differences in outcomes purely to the treatment effect. In practice, there are potentially many characteristics one could match on. The propensity score matching method selects “similar” households on the basis of their predicted probabilities of participation or propensity scores. Under certain conditions, matching on the propensity score eliminates selection bias. A standard logit or probit regression can be used to estimate the propensity score for each observation. The control variables in the logit or probit regression are often the same as those in X_i but this is not a requirement. For example, pre-program values of the outcome is often a valuable control variable in the propensity score regression. Although it would not pass exogeneity tests if one thought of the system of equations from a structural point of view, balancing on pre-program outcomes often helps reduce selection bias on unobserved characteristics in practice. An important corollary of the approach is that if the propensity score equation does not have good predictive power, one should expect greater residual selection bias on unobserved characteristics.

Once the propensity score is estimated, a comparison observation for each treated observation is created by choosing the “nearest neighbor”, defined as the untreated household with the closest propensity score. It may sometimes be useful to create a comparison “group” by choosing more than one nearest neighbors.

As in a randomized experiment, the ATE can be estimated as the difference in means for the treated and the matched comparison group. Specifically, now the estimate of ATE is given by

$$ATE = \frac{1}{N_{t=1}} \sum_{j=1}^{N_{t=1}} \left(y_j^{t=1} - \frac{1}{\sum_{i=1}^{N_{t=0}} w_{ij}} \sum_{i=1}^{N_{t=0}} w_{ij} y_j^{t=0} \right)$$

where $N_{t=1}$ denotes the number of households receiving the treatment, $N_{t=0}$ denotes the number of households not receiving treatment and the non-participant households and the w_{ij} 's are weights constructed to describe the quality of the match. Two popular weighting schemes are the nearest-neighbor weights in which non-matched neighbors implicitly have zero weights, and kernel-based weights in which all households in the control group have a non-zero weight, but one that diminishes in the absolute difference in the propensity score between the treated observation and the candidate control household.

LaIglesia (2004) employed propensity score matching methods to arrive at estimates of the impacts of land property registration on investment and credit access, working with a dataset of Nicaraguan farm properties. Property registration was demand-driven rather than compulsory in the regions studied. The decision to register a plot was predicted using observed characteristics of the plot and household. This probability of registering or propensity score was then used to match observations from households with registered plots to observations on unregistered plots in households with similar characteristics.

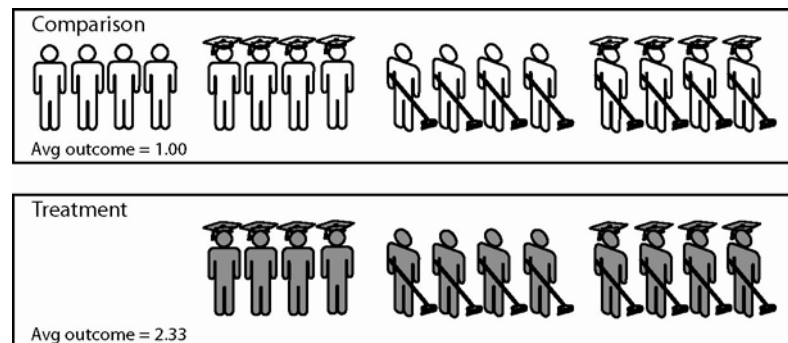
As in the methods for randomized designs, regression methods may be used to construct ATE after propensity score matching, but such applications are not common.

Selection into Treatment Based on Unobserved Characteristics

In many situations, if potential beneficiaries are able to self-select into treatment, it is reasonable to believe that they do so, at least in part, on the basis of their unobserved characteristics. Reconsider the example in which beneficiaries must pay taxes or purchase fees to obtain title. We argued earlier that a household's observed access to credit may influence selection into treatment. In addition, it is plausible to believe that high ability households are more likely to select into treatment. Because household ability is unobserved by the researcher, a solution to this feature of selection is more difficult to take into account.

Consider the situation where, even though households have been assigned to treatment randomly, only individuals with high a or high X actually apply for or take-up the program. Because only three of four types of households will have chosen to take up the program intervention we show only 12 households receiving treatment in the figure below. Such situations are quite common, i.e., typically one expects lack of balance on observable and unobservable characteristics to go hand in hand. In this example, the simple comparison of means will provide a biased estimate. $2.33 - 1.0 = 1.33$ in the example, whereas the true impact is only 1.0. The problem with using this estimator is that it incorrectly attributes higher outcomes to the treatment which are instead due to the fact that the treated group contains a heavier selection of high a_i households relative to the comparison group.

Figure 7: Selection on Unobservables



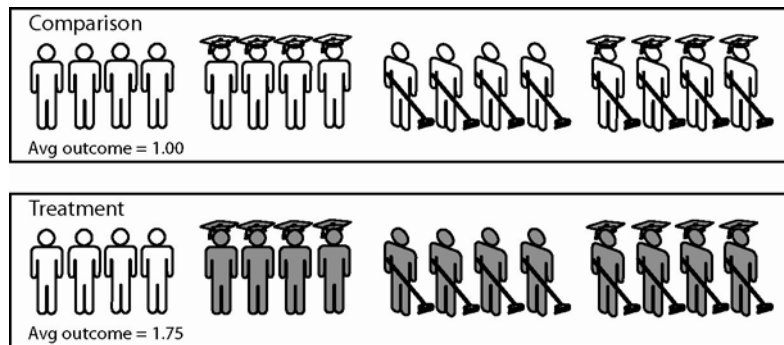
How to avoid this bias? Unlike the situation where there is selection on observed characteristics, here it is not possible to redress the imbalance via matching methods. It is possible, however, to adjust for this type of bias using instrumental variables regression methods.

Instrumental Variables Estimation

Loosely speaking the idea behind an instrumental variables estimator (IV) is to rebalance treatment and comparison groups on observable and unobservable characteristics by using a treatment group predicted by exogenous 'instrumental

variables' rather than actual participation. In our simple example we have a very good instrumental variable in the form of the random assignment to treatment. Suppose that rather than estimate outcomes for the actual treatment group as in the figure above we instead measured the average outcome for households that were initially assigned to the treatment group whether or not they actually participated. In our example this means that households with low a and low X are now considered to be part of the treatment group, as indicated in the figure below. Unobservable ability a is now balanced between our new 'treated' and the comparison groups. The comparison of means between those who 'received treatment' and those who did not yields an effect of 0.75 ($=1.75-1.0$), which is the *intent-to-treat effect*. The intent-to-treat effect underestimates the true effect in this example because one fourth of the households in the 'treatment group' did not in fact receive treatment. The instrumental variables method works by scaling up this initial estimate by the ratio of intended to actual size of the treatment group. In our example, this ratio is $4/3$, thus the ATE obtained via instrumental variables methods is $4/3 \times 0.75 = 1.0$ and so we have measured the impact without bias.

Figure 8: Intent to Treat as an Instrumental Variable



The instrumental variables estimator in the regression context applies more generally than the example we have used to illustrate the procedure. In the general setup, consider again the basic regression model:

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \varepsilon_i$$

but now p_i is assumed to be correlated with ε_i . In the context of our example, selection into treatment is, in part, determined by household ability, an unobserved characteristic that also determine the outcome, since ε_i includes unobserved effort e_i . In this case, the least squares estimate of β_1 is no longer an unbiased estimate of the ATE. Note that t_i may also be correlated with X_i (degree or education in our example) but this does not pose any particular statistical issue. An unbiased estimate of β_1 may be obtained if one is able to identify in the data a variable z_i which is correlated with treatment t_i but uncorrelated with ε_i , or in other words, uncorrelated with the outcome conditional on observed characteristics. An alternative way to think about z_i is that it should, in part, determine selection into treatment but should not determine the outcome, except through

its effect on treatment. The variable z_i is often called the instrument, and the statistical approach to estimating the parameters of the regression is known as the instrumental variables regression. The technical details of instrumental variables regressions can get complicated, but most standard statistical software provides canned tools to obtain such estimates.

In the case of experimental or quasi-experimental designs with compliance issues that create some selection into treatment, the original experimental assignment into treatment and control groups serves as an excellent instrument as in the example above. This variable should be highly correlated with receipt of treatment, and being a consequence of randomization, should be uncorrelated with the outcome, except via the indicator for actual receipt of treatment. Even when randomization or quasi-randomization is not feasible, it may be possible to provide random encouragement, i.e., randomizing the intensity of the information provided to households about a program. Although it would not be appropriate to use methods for randomized designs in such contexts, the intensity of information provided could be used as a powerful instrument in an instrumental variables analysis of an otherwise observational design.¹⁰ Unfortunately it is often rather difficult to think of other plausible variables that might predict participation but are not at the same time correlated with the outcomes of property rights reforms, so instrumental variables estimation will often prove difficult or unconvincing where there was no experimental design.¹¹

As discussed above, interactions of the treatment indicator with exogenous covariates is a simple, yet useful way to allow for treatment effects to vary across households. Therefore, we describe a simple strategy for incorporating such heterogeneous effects into an instrumental variables framework. As before, the outcome regression is specified as

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \beta_3 X_i t_i + \varepsilon_i$$

with t_i assumed to be correlated with ε_i . One way to approach the fact that now, in addition to t_i being endogenous, $X_i t_i$ is also endogenous, is to use z_i and $X_i z_i$ as instruments for t_i and $X_i t_i$ and proceed with standard instrumental variables. Note that although straightforward in principle, such models may in fact be quite fragile in practice because each interaction is treated as a separate endogenous regressor. This may explain why applications of instrumental variables regressions in which the endogenous regressor is interacted with other exogenous regressors remain sparse in the literature.

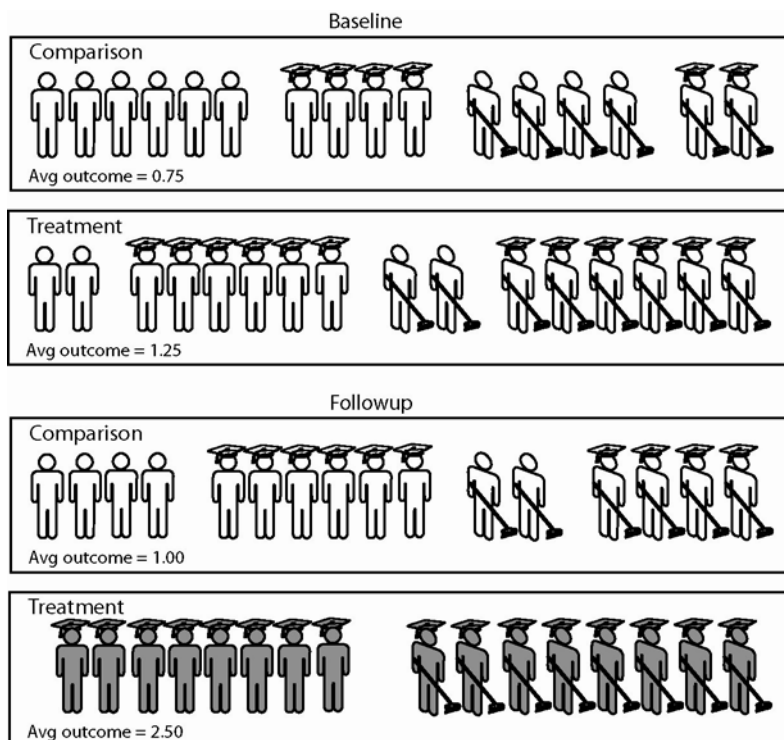
¹⁰ Deininger, Ayalew and Yamano(2005) argue that differences in the intensity of ‘sensitization’ to which different Ugandan beneficiary communities were exposed prior to a pilot program for systematic adjudication – which they argue can be treated as exogenous – had a large impact on outcomes such as tree investment and land productivity.

¹¹ The IV approach has nonetheless been employed in many studies to measure the impact of property rights security on investment and other outcomes using observational data. Besley (1995) used possession of an existing title deed as an instrument for his measure of ‘transfer rights.’ Similarly, LaGlesia (2004) used a measure of existing documents held as an instrument for registration in a new titling program. Pande and Udry (2005) provide a critical review of several of these studies.

Pre-post Data (Difference-in-difference Design)

Consider the situation in which the program is implemented in entire zones or communities. Untreated communities exist, but they have not been chosen in a random fashion. Continuing with our stylized example in Figure 9 below, suppose that all households in zone III are targeted for treatment while all households in zone II are targeted to be in the comparison group. As before, households in zone III have a pre-program outcome of 1.25 while households in zone II have a pre-program outcome of 0.75. By the time of the follow-up survey, households in zone II have an average outcome of 1.00 (in general, we might think of this as being due to an upward ‘drift’ in the value of d ; in the figure, we depict this by 4 individuals in zone II receiving degrees in the interim) while those in zone III, who have received the treatment, have an average outcome of 2.50 (which also includes an upward drift). Note that without the intervention, zone III households would have a follow-up outcome of 1.50. Again, if we only had a cross-section survey taken after the program intervention the ex-post difference between treated and untreated, 1.50, would overestimate the impact of titling. Data from a baseline survey would allow us to calculate a difference-in-difference estimator, which would provide the correct program impact.

Figure 9: Difference in Difference (DiD) Estimator



There are two equivalent ways to envision the DiD estimator. First, consider the difference in baseline and follow-up outcomes for the comparison group, $1.00 - 0.75 =$

0.25, and the treated group, $2.50 - 1.25 = 1.25$. Next, use the estimate of the trend effect obtained from the comparison group to remove the trend effect in the observed change in outcomes for the treated group. Thus the program effect for the treated group is $1.25 - 0.25 = 1.00$. Second, consider the pre-program difference between households in zone III and those in zone II, $1.25 - 0.75 = 0.50$, and the follow-up difference between outcomes for those two sets of households, $2.50 - 1.00 = 1.50$. Then use the pre-program difference to remove the effect of neighborhood heterogeneity to get the true program effect, $1.50 - 0.50 = 1.00$.

More generally, let the outcome y_i for household i be measured twice ($p=1$ for the post-program or follow-up survey and $p=0$ for the pre-program or baseline survey). The DD treatment effect is calculated as

$$ATE = (E(y_{i1} | t_i = 1) - E(y_{i0} | t_i = 1)) - (E(y_{i1} | t_i = 0) - E(y_{i0} | t_i = 0))$$

In typical implementations using regression methods, the model is given by

$$y_i = \beta_0 + \beta_1 p_i + \beta_2 X_i + \beta_3 t_i + \beta_4 p_i t_i + \varepsilon_i$$

and the DD estimate of ATE is given by $\hat{\beta}_4$. As with previously described designs, it is possible that the ATE is not a constant, and indeed varies according to observed characteristics of the households. In such a case, the regression can easily be augmented to allow for such effects:

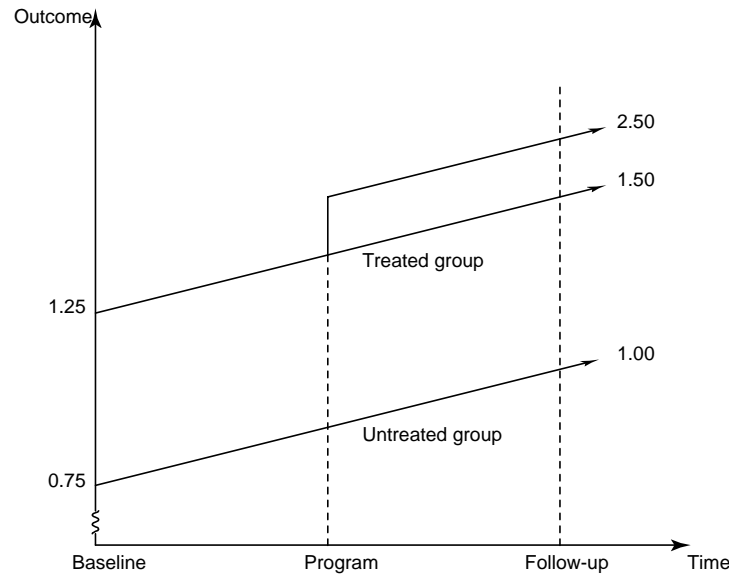
$$y_i = \beta_0 + \beta_1 p_i + \beta_2 X_i + \beta_3 t_i + \beta_4 p_i t_i + \beta_5 p_i X_i + \beta_6 t_i X_i + \beta_7 p_i t_i X_i + \varepsilon_i.$$

The ATE for a household with characteristics X_i is given by $\hat{\beta}_4 + \hat{\beta}_7 X_i$.

Note that both illustrations and the regression analysis proposed for the DiD design assume a constant rate of change between baseline and follow-up points in time, and the same rate of change for treated and untreated groups. If either of these assumptions is violated, the DiD estimator as described above will not provide correct estimates of program impact. For example, suppose that the rate of change is greater at baseline for households that select into treatment. This may not be an uncommon situation since the demand for property may be highest precisely in those dynamic growth regions where households are on a faster upward trajectory. Then the DiD estimator would overestimate the program effect, as it would attribute the greater baseline rate of growth to the program. The situation would be made even worse if, in addition to the baseline difference in growth rates, the program caused growth rates in program households to increase further. In such situations, program impacts can be identified if additional surveys are implemented. For example, in the case of a pre-program difference in the rate of growth, in principle one requires two pre-program surveys to pin down the differential rates of growth in treated and untreated households. If the growth rate was influenced by the program intervention, then (possibly in addition to two pre-program surveys) two post-program surveys would be needed to identify the program impacts. In

practice, it is typically believed that the standard DD estimator removes most, if not all, of the bias, and is widely applied.¹²

Figure 9: Difference in Difference (DiD) Estimator



Cross-sectional Data with Multiple Observations per Unit (Household Panel)

In some situations, households may possess more than one plot of land. In the simplest instance, suppose that each household has two plots of land. Further, assume that for at least a subset of households, one plot of land receives the program intervention while the other plot does not. Then it is possible to eliminate the bias introduced by the lack of balance on unobserved characteristics without fielding surveys at multiple points in time but by instead implementing a regression with household fixed-effects.

Consider the following regression specification:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_i + \beta_3 Z_{ij} + u_i + \varepsilon_{ij}$$

where the subscript j indicates the plot within the household. The specification distinguishes between observed household characteristics X_i , observed plot characteristics Z_{ij} and also household unobserved characteristics u_i which includes unobserved ability and ε_{ij} , an idiosyncratic random error term which includes unobserved characteristics of

¹² There are a number of more technical issues associated with the DiD approach that we have not discussed but one is worth mentioning briefly. The issue is that in measuring left hand side outcome variables such as ‘investment’ one may find that one has quite a few zero observations. The standard solution of running a Tobit regression will not work in the panel data case that is required for the DiD approach because the so-called “incidental parameters problem” (Greene 2003) may lead to biased coefficients. There is no clear solution to this problem but good practice would include reporting Tobit-FE results (obtained by including a dummy for each observation unit) as well as random effects estimates. Ayalew et al. (2005) provide an example of this approach.

the plot of land. The fixed effect method eliminates the effects of X_i and u_i as they do not vary within households, and produces unbiased estimates of the ATE is given by $\hat{\beta}_1$. Note that although we have presented the method assuming that each household has two plots of land, this is not necessary. It is only necessary that some households possess more than one plot of land and, among households that possess more than one plot of land, that some households have plot-level variation in whether the plot is subject to the program intervention or not. If households either have all plots subject to the intervention, or all plots not subject to the intervention, then the fixed effects method cannot identify the ATE.

In ideal situations, this within-household variation could be built into the program design. Indeed, this might be one of the few instances where randomization within households may survive political, ethical and administrative scrutiny. In purely observational studies, such variation would be fortuitous. But there are examples of such situations in the literature.

In a recent paper Gine (2004) explored the Thai context studied earlier by Feder (1988). He uses the fact that in his sample many households had plots within the forest reserve (where property rights are more insecure) and outside of its boundaries. He is able to use a household level fixed-effect estimator in which the household itself acts as its own counterfactual since he can compare outcomes on plots that were ‘treated’ (made more insecure) against outcomes on plots within the same household that were outside the forest reserve. Estimates of the program impact can then be obtained by averaging the measured difference in outcomes between treated and untreated plots across all households. Gine uses this approach to argue that tenure insecurity significantly dampened activity on the land lease market. Jacoby and Minten (2006) used a similar approach to study the impact of formal titling in Madagascar. They find that the private economic benefit from extending land titling is relatively minor. Although, as the authors readily admit, the assignment of plots to titled or untitled status is much less obviously exogenous in the Malagasy case, the study provides a good example of the important uses of plot-level data.

Cross-sectional Data with a Program-eligibility Rule

When randomization or quasi-randomization is not possible and only data at a single point in time (post program) is feasible, it might still be possible to create sharp program eligibility thresholds such that households on either side of the threshold may be reasonably considered similar. Such eligibility thresholds can often be justified on the basis of equity, financial or administrative considerations, and can be used to either to create assignment into program treatment or simply to create a waiting list for program treatment. In our example, suppose the program is designed so that only households with land holdings below a certain arbitrary threshold are eligible for the intervention. Clearly, we expect land size to potentially affect the outcome, so such a mechanism is not a randomization device. It is, however, reasonable to expect that land size will affect the outcome only in a smooth, continuous way. In other words, one does not expect a discontinuous change or even a very nonlinear but smooth change in the outcome due to land holdings, especially in the neighborhood of the proposed eligibility threshold. In

addition, although land size may generally be correlated with unobserved household ability, it is reasonable to believe that households on either side of, but close to the arbitrary cutoff have the same (or at least very similar) ability. Finally, one expects that if the program treatment has an effect on outcomes, it would create a sharp, discontinuous effect at the eligibility threshold.

If all the conditions stated above are met, we have what is known as a regression discontinuity design. In such designs, the program eligibility rule can be used as an instrument in an instrumental variables regression. The arbitrary eligibility threshold satisfies both conditions for validity of the instrument. First, the indicator for eligibility should be highly correlated with actual receipt of treatment. Second, assuming that the smooth, continuous effect of the characteristic underlying the threshold is taken into account (land holdings in our description above), eligibility should influence outcome only through the program intervention. See Pitt & Khandker (1995) for a leading example of the regression discontinuity design in the development context. But this only works when the rule is well enforced and is not subject to manipulation by the beneficiaries (one can imagine selling a wee bit of land if it gets you big benefits for instance).

Cross-sectional Data, No Eligibility Rules

In terms of study design, the worst possible scenario arises when only cross-sectional data is available and when assignment into treatment is not random. Such designs are often necessitated by post-intervention calls for evaluation, when it is too late to influence study design or data collection efforts in substantive ways. Program evaluation is still feasible but results are not always reliable. In addition, often more complicated statistical methods are needed to overcome the basic shortcomings of the design. Nevertheless, we discuss appropriate statistical methods and their limitations briefly.

In the case of observational studies, where selection into treatment based on unobservables cannot typically be ruled out, it is often difficult to identify valid instruments in the data. Typically, it is not difficult to find variables that affect selection into treatment, but it is much more difficult to argue or provide evidence that such candidate variables are exogenous to the outcomes under consideration. Nevertheless, if a valid instrument can be identified in the data, instrumental variables methods can be used to estimate ATE.

It is also becoming increasingly popular to use propensity score matching to obtain estimates of ATE in this context. In principle, propensity score matching assumes away selection on unobserved characteristics of households, which makes it inappropriate, in general, for such situations. In practice, however, one hopes that selection on observed and unobserved characteristics affect outcomes in the same direction, so that eliminating selection on observables also eliminates part of the selection due to unobserved characteristics. Such a belief seems plausible in many applications, but it cannot be verified nor assured a priori.

E. Concluding Thoughts on Designs and Methods

For ease of exposition, we have provided a considerably stylized description of alternative study designs and associated statistical methods. In practice, the implementation of programs can involve a mix-and-match of designs, requiring judicious mixing-and-matching of methods. Even when experimental randomization is implemented, evaluators would do well to be aware of post-implementation complexities. Consequently, in terms of data collection efforts, we strongly recommend a pre-intervention baseline survey in addition to a post-intervention survey, as well as careful monitoring of implementation for potential mid-course corrections

We have described study designs in two distinct ways. In the first, randomization across households or households selected into treatment, i.e., the study is designed with the household as the focal point. In the second, neighborhoods or villages are selected randomly, or otherwise chosen to participate in the intervention and we have implicitly assumed that participation within such units is complete and without contamination. In practice, however, even when communities are selected at random, often households within such communities are allowed to self-select into the program. Thus selection needs to be considered at multiple levels. If analysis is conducted at the neighborhood or village level, participation rates can be used as measures of treatment intensity. If analysis is conducted at the household level, which is the approach we prefer, the discussion of intent-to-treat designs and instrumental variables methods are relevant even if randomization at the community level is perfect.

We have described parametric instrumental variables and difference-in-difference methods but also recommended less parametric propensity score matching methods for better adjustments for selection on observed characteristics. One could, in principle, implement instrumental variables or difference-in-difference estimators after using matching methods to reduce or eliminate selection on observed characteristics. Alternatively, one could use semi- or non-parametric regression approaches in the difference-in-difference context. Such models are more general than the parametric methods described here, but are also substantially more complex and often of unknown practical reliability. In our view, they should be used with caution.

V. Measurement and Data Collection Issues

Throughout our description of statistical methods, we have noted the importance of repeated measurements of potential impacts. Regardless of the nature of the program design, reliable evaluations will likely require information from a pre-intervention or baseline survey coupled with a post-intervention or follow-up survey. A case can also be made for having more than two surveys, especially in the case of programs where impacts are likely to be relatively slow to materialize. If budget constraints and timeliness limit this possibility, we strongly recommend at least two surveys.

In order for a baseline survey to be fielded in a timely manner, it is imperative that its planning begin at the early stages of program design and prior to implementation. Lack of early planning often results in studies with only one, post-intervention survey. As emphasized in our earlier discussion, the cost of not being able to balance treatment and

control groups in observational designs can be quite large in terms of reliable estimates of the treatment effects.

A. *Adapting Standard Survey Questionnaires*

Fortunately, questionnaire design need not be tackled from scratch. There are two large and widely used general purpose surveys, the Living Standards Measurement Surveys (LSMS) and the Demographic and Health Surveys (DHS) that provide high quality, off-the-shelf questionnaires (Deaton 1997; Grosh and Glewwe 1995). Unless the specific situation warrants an alternative approach or design, several measures of household and neighborhood characteristics and of several potential impacts (e.g. on measures of production, income, consumption, health, education, human capital, etc.) exist in these multipurpose surveys.

While there are several obvious measures of impact matched to the main expected claims from property rights reforms that have been briefly described above – household and community level investment, consumption and income, demand for and access to credit, male and female labor supply, land sale and lease activity, etc – it may be useful and important to consider outcomes that are typically more difficult to measure such as home production and self-employment. Children’s health and nutrition and female fertility are also outcomes often overlooked in studies of property rights reforms.

Some important variables are not measured in much detail in the standard LSMS and DHS surveys, for instance questions pertaining to the physical characteristics of household plots or detailed questions concerning formal and informal property rights claims and their history. However a LSMS land module is available which does specifically address several of these issues and a recent very good World Bank study of gender issues in land administration projects discusses data collection issues for studying gendered rights and intra-household bargaining issues (World Bank 2005). This last study contains an appendix with a sample land module questionnaire.

While these resources should be consulted and are very useful, we recommend caution in implementing an off-the-shelf approach. The best approach is to adapt available land modules to suit the type of land intervention being studied and to design questions and data collection strategies after having spent considerable effort to understand the nature of existing property arrangements, program design, and potential impacts via preliminary field visits, interviews and studies of the existing literature.

B. *Measuring Property Rights and Impacts*

There is an immense diversity of property rights arrangements around the world. Property rights may be private, cooperative or communal but even these three different labels fail to adequately even begin to describe the universe of actual arrangements. Communities typically will also have more than one possible mechanism for defining and enforcing property rights, and different property rights systems may overlap or come into conflict at times. Informal community-sanctioned rights or rights of possession may or may not match rights established through custom or through formal law (Shipton and

Goheen 1992; Ellickson 1991; Deininger 2003; Platteau 1995). Property rights are also multi-dimensional and layered. For instance, a family may enjoy the right to cultivate a plot of land or even to rent or sell it to others within their community, but they may not enjoy the right to make the same transfers to outsiders, or the right to exclude sanctioned other users from deriving certain other uses from the same plot, for instance grazing their animals on crop stubble during certain time periods. Men and women and children within the same household may live on and farm the same private or communal land but may enjoy different rights to use or inherit it.

One view of this is to simply take such complexity as given and apply standard impact evaluation methods to measure how observable outcomes of interest changed in ‘treated’ communities compared to other similar communities without the same interventions. A more serious approach would however attempt to make an effort to understand and measure different dimensions of property rights in a given community for two reasons. First, such efforts would incorporate relevant information into the observable covariates that may be expected to affect and possibly interact with the program intervention. Second, such efforts would better identify and understand the principal stakeholders and outcomes of interest. This will often provide a clearer context for understanding unintended or unexpected impacts, for instance how existing stakeholders rights may have been enhanced or diminished.

Because property rights are multi-dimensional it will be important to ask more than just whether an individual ‘owns’ a plot. An individual may have tenure security without the ability to transfer (e.g. a villager enjoys secure access to communal land but cannot sell or lease land to outsiders) and vice-versa (e.g. someone who possesses a plot may have the ability to buy or sell land in an informal urban settlement but may feel the insecurity of being possibly evicted or having others squat on their own land). What rights and privileges accompany possession? Can the plot be leased, transferred, sold or mortgaged? When formal titles do not exist, informal claims to the property may nonetheless still be quite strong and secure and it will be important to understand the mechanisms that establish these *de jure* rights, and the distribution of these rights within a community.

Property rights security will often depend on the nature of one’s trading partners. Where enforcement relies on informal community sanction or custom, individuals may only trust transactions with other members of one’s immediate kinship group or trading network, and many communities impose explicit bans on certain types of land transactions with outsiders. If one of the anticipated outcomes of an intervention is a more active land rental market, then one may want to collect detailed information not just on the household’s level of land market transactions, but also data on the characteristics of their trading partners to understand the changing pattern of contracting. Macours (2004) offers an excellent example of such a data collection effort and the purposes to which it can be put.

It will often also be useful to collect data on land accumulation trajectories, for instance data on inheritance, the history of land market lease and sale transactions by individuals and their parents. This will be useful for tracing evolution over time. In some

cases such historical variables may also serve as potential instruments. It is also often useful to ask households to recall data on potential outcome measures. For instance in the above mentioned cross-sectional study, Field (2003) used recall data on family demographic data to identify a significant impact of land title on household fertility using a difference in difference approach.

Gender and Intra-household Allocation Issues

Attention to the detail of local property rights systems and production processes can also end up affecting how one interprets outcome variables. Goldstein and Udry (2005) have argued that ‘fallowing’ to improve soil productivity in village lands in Ghana is more likely to be carried out by individuals who enjoy more secure property rights. They find that village women who enjoy less secure rights than their husbands tend to fallow land less often as a way to establish continued possession of a plot.¹³ Their understanding of property relationships and gender politics within the study region led them to separately collect detailed data for husbands and wives within each household. This allowed them to empirically discern relationships and patterns of resource use that would have been obscured had they followed the more conventional practice of using the household as the unit of observation.

Several other studies have found that property rights reforms have quite different impacts on male-headed and female-headed households. Lanjouw and Levy (2002) and Field (2003) have both found evidence to support the claim that the granting of more secure title in urban squatter settlements in Ecuador and Peru increased outside the household labor supply considerably more in female-compared to male-headed households. The reason they suggested this may happen is similar to that offered by Goldstein and Udry: women may have to spend more time at home protecting insecure property rights through continued possession, and therefore will have a larger labor supply response when their tenure security is improved by new legal external enforcement.

A recent World Bank synthesis report entitled “Gender Issues and Best Practices in Land Administration Projects” offers an in depth discussion of gendered rights under different forms of property ownership around the world, how land administration and property rights reforms have attempted to re-shape those rights in several contexts, as well as a list of indicators and outcome measures as well as sample questionnaires for gathering gender-disaggregated information for impact assessments (World Bank 2005). The report underscores the importance of collecting baseline data ahead of program implementation not only as a tool for impact assessment but also as a way to know the gender-related issues that might be addressed through the intervention.

Subjective Measures of Property Rights Security

Several researchers have made use of questions that have asked individuals about their own perceptions of property rights security and how the impacts that they *expect* to

¹³ On the topic of gendered property rights see also (World Bank 2005; Deere and Leon 2001)

come from property rights reforms. For instance Lanjouw and Levy (2002) asked households how much they thought their properties had been worth before property rights reforms were announced, as well as by how much they expected land values to increase after a reform. They were then able to use this data to apply a difference approach to measure impacts.

Asking individuals about how they subjectively perceive property rights to be changing is important for several other reasons as well. Several studies have noted that households may be quite sensitive to perceptions of how likely future policies will affect them. Ayalew, Dercon and Gautam (2005) used panel data and self-reported perceived threat of expropriation along with measures of investments in Coffee, Eucalyptus and Q'at. They asked households about whether they expected to be able to transfer land (including via inheritance) and whether they expect to lose land through land reform and found that their perception of insecurity clearly affected investment.

C. *Spillovers and General Equilibrium Effects*

If households in a comparison group are well aware of a titling program in other neighborhoods and have become quite certain that they will receive title themselves shortly, then they may begin to adjust their behavior in anticipation. Unless the researcher has some measure of this perception and takes it into account, the estimated ATE may not reflect just the true effect of the intervention. Fort et al. (2005) argue that perceptions had powerful effects in land titling programs in Peru and point to the possibility of interesting and complex spillover and feedback mechanisms. They report findings which suggest that being in a geographic area with a high concentration of other households had as important an impact as receiving title oneself, a finding that is possibly consistent with such anticipatory behavior.

The designs and methods we described above assumed away the possibility of general equilibrium, spillover, anticipation or peer effects. In any of these situations, it becomes considerably more difficult to isolate pure program effects because individuals in the comparison group might benefit from the existence of treated households, or because treated households might interact with each other in ways that might magnify treatment effects. Such effects can increase or decrease the observed treatment effects depending on the nature of the spillovers, and the problem of properly identifying the impact of endogenous social effects (i.e. how these behavior-modifying perceptions are formed in equilibrium) can become quite a challenging empirical issue in practice (Manski 1993). As usual the best way to be prepared for these challenges is to collect richer data, including data on perceptions.

Finally it is worth mentioning that one way to address the issue is to take a step back and attempt to measure impacts at a higher level of aggregation, focusing for instance on county, or region wide impacts using panel data. Besley and Burgess (2000) followed this approach to estimate positive impacts of land and tenancy reforms on state-wide measures of poverty and economic activity in India. More recently, Do & Iyer (2005) used province-level data from Vietnam to conclude that property rights reforms

had thus far mainly affected the types of crops adopted by farmers but had not yet led to several other anticipated impacts.

Packages of Reforms

A related concern is that land property rights reform interventions are quite often introduced in the context of much larger package of interventions. The existence of concurrent broader programs have implications for general equilibrium effects and spillovers, thus should also be seriously considered at design and implementation stages. Careful planning and thoughtful design may provide a way to measure the additionality of a particular project component, as well as interaction effects. For instance one might imagine a large project that combines mapping, property registration and land use management planning components. Although it may be difficult to convince policymakers to randomize the order in which the first two components are introduced, it may be somewhat more feasible to randomize the order in which the land use management component is assigned. Under certain assumptions it may then be possible to identify the incremental additive value of the land use module even if the impact of the base intervention was less convincingly estimated.

One of the main surprises, and some would say disappointment, of many existing impact assessments of land property rights reforms to date has been the failure to find much significant credit supply response (The Economist 2006) as advertised by some of the key proponents of property rights reforms including Hernando de Soto and others (de Soto 2000).¹⁴ Several explanations have been offered to try to explain this puzzle. One is that in certain contexts property rights reforms have only formalized informal property rights arrangements that already provided tenure security, so that the incremental value of the reforms should have been expected to have been relatively small at least in the short run (Brasselle, Gaspart, and Platteau 2002). Another explanation is that that many reforms may have strengthened property rights security to existing occupants without increasing transferability which is what creditors care about. In this view property registration and titling are just a few of the preliminary steps necessary to begin to stimulate private sector activity and the growth of finance in particular. Other necessary complementary steps such as improving the efficiency of judicial systems, rewriting bankruptcy codes, etc. may need to be in place before the credit supply response becomes appreciable (Woodruff 2001).

D. Differentiated Impacts

Since the stated purpose of many interventions is to improve access to land and security of tenure to the poor, we will be often quite interested in knowing how benefits and costs are distributed across households. An Average Treatment Effect (ATE) estimate of program impacts can be very useful for demonstrating program effectiveness but it tells us relatively little about the differentiated impact of the intervention across different groups of households within communities. Standard specifications of models

¹⁴ The early studies by Chalamwong and Feder are an important early exception (1988).

used for estimating ATE assume a constant ATE. A relatively simple way around this limitation, one that we have described above, is to include interactions of the treatment indicator with exogenous covariates. We could, for instance, interact a dummy variable representing the ‘receipt of a title’ with the size of the possession or the gender of the household head, to arrive at a size- or gender-differentiated measure of impacts. The idea of course in all cases will be to insure that the interacted variable is itself exogenous, and there may be reasons to believe that this assumption sometimes fails in the case of these and other variables. Note too, that this has implications for the desired sample size.

The length of time that reforms have had to take effect may also lead to differentiated impacts. Since it may take a considerable amount of time to roll out a nationwide property rights program – years or even decades in some cases – communities that are sampled in a post-intervention survey may differ significantly in how long ago they received treatment. A simple interaction between treatment and time elapsed since the intervention will create a simple ‘intensity of treatment’ variable that may allow one to discern how impacts are graduated over time. The time horizon under consideration will also affect the outcomes of interest, e.g., crop choice (annuals versus tree crops) or access to credit rather than yields and income.

Several recent studies have used non-parametric regression techniques as a way to extend this general method in more flexible ways. Whereas a simple linear regression equation interaction would imply a linear relationship between land size and program impacts, a non-parametric regression approach allows this relationship to be more flexible. Boucher, Barham and Carter (2005) applied such an approach to visualize changes over time in the shape of the relationship between probability of having title (or a formal loan) as a function of owned area, attributing the changes to in the relationships to the effects of market and titling reforms in Nicaragua and Honduras over the period. Their findings suggest a generally higher rate of take-up of title amongst medium and larger sized farms, a pickup in rental market activity across the spectrum, but credit market impacts that maintain a strong skew against small farmers. A similar approach was been pursued by Finan et al (2005) to measure the poverty reduction effects of recent *ejido* reforms in Mexico. Olinto and Carter (1999) adopt a more sophisticated semi-structural model approach to measuring land and credit market interactions in studying land titling in frontier regions of Paraguay.

The study of differentiated impacts remains an important item on the research agenda particularly as a significant anthropological and economic literature has emphasized the role of existing inequality within a community in shaping the evolution of property rights definitions and their distribution over time (Boserup 1965; Baland and Platteau 1997; Baland and Platteau 1999). A recent study by Munoz-Pina et al. (2003) found a significant relationship between initial inequality and the decision to privatize communal land *ejidos* in Mexico.

Conclusions

The idea that property rights matter, and matter a lot for investment incentives, efficient resource use, and economic growth has been one of the central tenets of political economy for centuries. A large wave of new policy interventions have been carried out to identify, secure or otherwise transform land property rights systems around the world and to build the infrastructure and support systems to create more modern and efficient property rights systems. Given globalization and new technology, the transformation of rural spaces, and the rapid pace of urbanization and growth in many parts of the world, many more large scale programs to modernize and transform property rights systems can be expected. Unfortunately, there are very few rigorous impact evaluation studies carried out to measure the distribution of benefits or the cost effectiveness of such interventions.

In this paper we have argued that situation could be fairly easily remedied. Rigorous impact evaluation studies can be carried out at relatively modest expense and doing so may strongly complement other program objectives. The end result should be better program designs and project experimentation, better targeting, greater transparency and credible data to build support and legitimacy for projects and contribute to a more informed public debate.

While we have highlighted some factors that can complicate the statistical identification of program impacts we have tried to eschew technical issues by focusing on a sequence of practical examples that underline the importance of data collection strategies including, most importantly of all, planning for baseline and follow-up surveys, and (where possible) randomized designs. We have also deliberately focused on straightforward statistical techniques. Although arguments can be made against such methods, we feel that the methods we have described have the advantages of being easy to implement, are well understood, and work reliably in a wide variety of data contexts. The newer, more sophisticated methods are much less well understood and are much more difficult to implement. Therefore, the potential costs of using such methods may, in fact, outweigh their gains, at least for the near future.

Table 1: Summary of Selected Study Designs and Methods

Country, Authors	Outcomes	Program or intervention	Data collection, # observations	Statistical methods
Observational designs				
Ecuador Levy & Lanjouw (2002)	Land Prices Self-reported Welfare Ability to contract	Effect of hypothetical titling	Cross-section 400 households	Inverse-mills ratio to control for selection
Ethiopia Ayalew, Dercon & Gautam (2005)	Tree Investment Land allocated to commercial crops	Effect of stronger transfer rights	4 wave panel 1470 households	Household FE, IV
Ghana Besley (1995)	Land Rights Investment Productivity	Households decision to register	Cross-section 384 households, 1568 plots	Household FE, IV
Ghana Goldstein and Udry (2005)	Investment Productivity	Effect of stronger property rights	15 round panel 575 plots	Household-year and spatial FE, IV
Madagascar Jacoby and Minten (2005)	Plot-level investment land productivity Land values	Households decision to register	Plot-level cross-section 1604 households, 3232 plots	Household FE
Nicaragua LaGlesia (2004)	Investment Credit access	Households decision to register	Cross-section 2193 households, 3214 plots	Propensity scores, IV, Household FE
Nicaragua Deininger and Chamorro (2004)	Investment Land Values	Households decision to register	Cross-section 2475 households, 3659 plots	Household FE
Nicaragua and Honduras Boucher, Barham and Carter (2005)	Title Credit	Households decision to register	Repeated cross-sections	Semi-parametric regression
Paraguay Carter and Olinto (2003)	Investment Credit Access	Households decision to register	2 wave panel 300 households	Switching regression, Household FE
Uganda Deininger, Ayalew & Yamono (2005)	Tree investment Soil conservation Productivity Land values	Information campaign within systematic adjudication program	Cross-section 970 households, 2185 plots	IV
Quasi-randomized designs				
Argentina Galiani & Schargrodsky (2007)	Housing investment Household Size Education Credit Access	Urban property formalization	Cross section 530 parcels	QR 'Discontinuity'
Peru Field (2003) Field and Torrero (2004)	Labor Supply Credit access	Urban property formalization	Cross section 2750 households	QR 'Pipeline'
Thailand Feder et. al (1988)	Credit access Land values Capital formation	Reduced of security/transferability of land in forest reserves	Cross section 230 households	Switching regression
Thailand Gine (2004)	Rental rates	Reduced of security/transferability of land in forest reserves	Cross section 2874 households	Household FE
Vietnam Do & Iyer (2005)	Cultivated area Irrigated area Labor inputs Investment Land transactions	Changes in Land Law that increased transferability	Province-level panel 61 provinces, 6000 households	DiD

References

- Acemoglu, Daron, Simon Johnson, and James Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation," *The American Economic Review* 91 (5):1369-1401.
- Angrist, J. D. (2004). "Treatment Effect Heterogeneity in Theory and Practice," *Economic Journal* 114 (494):C52-C83.
- Ayalew, Daniel, Stefan Dercon, and M. Gautam (2005). "Property Rights in a Very Poor Country: Tenure Insecurity and Investment in Ethiopia," Department of Economics working paper, Oxford University.
- Baker, Judy L (2000). *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. The World Bank.
- Baland, J. M., and J. P. Platteau (1999). "The ambiguous impact of inequality on local resource management," *World Development* 27 (5):773-788.
- Baland, Jean Marie, and Jean Philippe Platteau (1997). "Wealth Inequality and Efficiency in the Commons: Part I: The Unregulated Case," *Oxford Economic Papers* 49 (4):451-482.
- Berry, Sara (1993). *No condition is permanent : the social dynamics of agrarian change in sub-Saharan Africa*. Madison, Wis.: University of Wisconsin Press.
- Besley, Timothy (1995). "Property Rights and Investment Incentives: Theory and Evidence from Ghana," *Journal of Political Economy* 103 (5):903-937.
- (1995). "Property Rights and Investment Incentives: Theory and Evidence from Ghana," in: University of Chicago.
- Besley, Timothy J., and Robin Burgess (2000). "Land Reform, Poverty Reduction, and Growth: Evidence From India," *Quarterly Journal of Economics* 115 (2):389-430.
- Binswanger, Hans P., Klaus Deininger, and Gershon Feder (1995). "Power, Distortions, Revolt and Reform in Agricultural Land Relations", in *Handbook of Development Economics*. Volume 3B, Amsterdam; New York and Oxford: Elsevier Science, North Holland, 2659-2772.
- Boserup, Ester (1965). *The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure*.
- Boucher, Stephen R., Bradford L. Barham, and Michael R. Carter (2005). "The Impact of "Market-Friendly" Reforms on Credit and Land Markets in Honduras and Nicaragua," *World Development* 33 (1):107-128.

- Brasselle, Anne-Sophie, Frederic Gaspart, and Jean-Philippe Platteau (2002). "Land Tenure Security and Investment Incentives: Puzzling Evidence from Burkina Faso," *Journal of Development Economics* 67 (2):373-418.
- Carter, Michael R., and Pedro Olinto (2003). "Getting Institutions "Right" for Whom? Credit Constraints and the Impact of Property Rights on the Quantity and Composition of Investment," *American Journal of Agricultural Economics* 85 (1):173-186.
- Chalamwong, Yongyuth, and Gershon Feder (1988). "The Impact of Landownership Security: Theory and Evidence from Thailand," *World Bank Economic Review* 2 (2):187-204.
- de Soto, Hernando (2000). *The Mystery of Capital: Why Capitalism Succeeds in the West and fails almost everywhere else*. New York: Basic Books.
- Deaton, Angus (1997). *The analysis of household surveys: a microeconometric approach to development policy*. Baltimore, MD: Published for the World Bank [by] Johns Hopkins University Press.
- Deere, Carmen Diana, and Magdalena Leon (2001). *Empowering women: Land and property rights in Latin America*. Pittsburgh: University of Pittsburgh Press.
- Deininger, Klaus (2003). *Land policies for growth and poverty reduction*. Washington, D.C. Oxford and New York: World Bank Oxford University Press.
- Deininger, K., D. Ayalew and T. Yamano (2005). "Legal knowledge and economic development: The case of land rights in Uganda," World Bank Development Economics Research Group.
- Deininger, Klaus, and Juan Sebastian Chamorro (2004). "Investment and Equity Effects of Land Regularisation: The Case of Nicaragua," *Agricultural Economics* 30 (2):101-116.
- Do, Quy-Toan, and Lakshmi Iyer (2005). "Land Titling and Rural Transition in Vietnam" [cited. Available from http://www.people.hbs.edu/liyer/Do_Iyer_June2005.pdf
- Duflo, E., and M. Kremer (2003). "Use of Randomization in the Evaluation of Development Effectiveness," Fifth Biennial World Bank Conference on Evaluation and Development. "Evaluating Development Effectiveness: Challenges and the Way Forward." Washington, DC July:15–16.
- Duflo, Esther (2004). "Scaling Up and Evaluation," Annual World Bank Conference on Development Economics 2004.
- Easterly, William (2006). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good..* The Penguin Press HC.
- Ellickson, R. C. (1991). *Order Without Law: How Neighbors Settle Disputes*. Harvard University Press.

- FAO (2006). *Contemporary thinking on land reform*. Food and Agriculture Organization 2000 [cited 2006]. Available from <http://www.fao.org/sd/ltdirect/ltan0037.htm>.
- Feder, Gershon and Akihiko Nishio (1999). "The Benefits of land registration and Titling: Economic and Social Perspectives," *Land Use Policy* 15 (1):25-43.
- Feder, Gershon, and Tongroj Onchan (1987). "Land Ownership Security and Farm Investment in Thailand," *American Journal of Agricultural Economics* 69 (2):311-320.
- Feder, Gershon, Tongroj Onchan, Yongyuth Chalamwong, and C Hongladarom (1988). *Land policies and farm productivity in Thailand*. Baltimore, MD.: Johns Hopkins University Press for the World Bank.
- Field, Erica (2003). "Fertility Responses to Urban land Titling programs: The Roles of Ownership Security and the Distribution of Household Assets".
- (2003). "Property Rights and Household Time Allocation in Urban Squatter Communities: Evidence from Peru".
- Field, Erica Marie (2003). "Entitled to Work: Urban Property Rights and Labor Supply in Peru," Harvard University working paper.
- Finan, Frederico, Elisabeth Sadoulet, and Alain de Janvry (2005). "Measuring the Poverty Reduction Potential of Land in Rural Mexico," *Journal of Development Economics* 77 (1):27-51.
- Firmin-Sellers, Kathryn, and Patrick Sellers (1999). "Expected Failures and Unexpected Successes of Land Titling in Africa," *World Development* 27 (7):1115-1128.
- Fort, Ricardo, Ruerd Ruben, and Javier Escobal (2005). "Spillovers and Externality Effects of Titling on Investments," Wageningen University, The Netherlands & GRADE, Lima, Peru.
- Galiani, Sebastian, and Ernesto Schargrotsky (2007). "Property Rights for the Poor: Effects of Land Titling," Universidad Torcuato Di Tella working paper.
- Giné, Xavier (2004). "Cultivate or Rent Out? Land Security in Rural Thailand," mimeo, The World Bank.
- Goldstein, Markus, and Christopher Udry (2005). "The Profits of Power: Land Rights and Agricultural Investment in Ghana."
- Government of Tanzania (2005). "Strategic Plan for the Implementation of the Land Laws," in: Ministry of Lands and Human Settlement Development (consultants Dr. Lugoe, Prof. Mtatifikolo, Ostberg).
- Greene, William (2003). "Fixed Effects and Bias due to the Incidental Parameters Problem in the Tobit Model," Department of Economics, Stern School of Business, New York University, mimeographed.

- Grosh, M., and Paul Glewwe (2006), *A Guide to Living Standards Surveys and Their Data Sets*. LSMS Working Paper #120, The World Bank, 1995 1995 [cited 2006]. Available from <http://www.worldbank.org/LSMS/guide/describe.html>.
- Heckman, J., J. L. Tobias, and E. Vytlačil (2001). "Four Parameters of Interest in the Evaluation of Social Programs," *Southern Economic Journal* 68 (2):210-223.
- High Level Commission on Legal Empowerment of the Poor (2006). *Agreed Principles and Conceptual Framework: Outcome Document from the First Commission Meeting*. URL:http://legalempowerment.undp.org/pdf/Agreed_principles_conceptual_framework.pdf.
- Holstein, Lynn (1996). "Towards best practice from World Bank Experience in Land Titling and Registration," in, World Bank Consultant Report.
- Jacoby, Hanan, and Bart Minten (2006). "Is Land Titling in Sub-Saharan Africa Cost-Effective? Evidence from Madagascar" [cited. Available from http://siteresources.worldbank.org/INTISPM/Resources/Training-Events-and-Materials/Land_Titles_MG.pdf
- Lanjouw, Jean O, and Philip I. Levy (2002). "Untitled: A Study of Formal and Informal Property Rights in Urban Ecuador," *Economic Journal* 112 (482):986-1019.
- Lauria-Santiago, Aldo (1999). "An agrarian republic: commercial agriculture and the politics of peasant communities in El Salvador, 1823-1914," Pitt Latin American series. Pittsburgh: University of Pittsburgh Press.
- Macours, Karen, Alain de Janvry, and Elisabeth Sadoulet (2004). "Insecurity of Property Rights and Matching in the Tenancy Market," Paul Nitze School of Advanced International Studies, Johns Hopkins University.
- Manski, C. F. (1993). *Identification of Endogenous Social Effects: The Reflection Problem*. University of Wisconsin--Madison, Institute for Research on Poverty.
- Mbongwa, M. (2006). *Land Reform for South Africa* by Masiphula Mbongwa (Director-General Agriculture) (2005). [cited May 16 2006]. Available from <http://www.info.gov.za/speeches/2005/05060312451001.htm>.
- Mitchell, Timothy (2005). "The work of economics: how a discipline makes its world," *European Journal of Sociology* 6 (02):297-320.
- Munoz-Pina, Carlos, Alain de Janvry, and Elisabeth Sadoulet (2003). "Recrafting Rights over Common Property Resources in Mexico," *Economic Development and Cultural Change* 52 (1):129-158.
- North, Douglass Cecil (1990). *Institutions, institutional change, and economic performance*. Cambridge; New York: Cambridge University Press.
- Olinto, Pedro, Maria Correia, K Deinenger, B Barham, M Carter, A de Janvry, E Katz, and E Sadoulet (1999). "Land Market Liberalization and the Land Access of the

- Rural Poor: Lessons from Recent Reforms in Central America," World Bank Research Proposal.
- Pagiola, Stefano (1999). "Economic Evaluation of Rural Land Administration Projects," The World Bank Land Economics Policy and Administration Thematic Team.
- Pande, R. and C. Udry (forthcoming). *Institutions and Development: A View from Below. Proceedings of the 9th World Congress of the Econometric Society.* edited by R. Blundell, W. Newey and T. Persson. Cambridge, Cambridge University Press.
- Pitt, Mark Martin, and Brown University. Population Studies and Training Center. (1995). "Credit programs for the poor and reproductive behavior in low income countries: are the reported causal relationships the result of heterogeneity bias?" PSTC working paper series ; no. 95-04. Providence: Brown University.
- Platteau, Jean-Philippe (1995). "Reforming Land Rights in Sub-Saharan Africa: Issues of Efficiency and Equity," United Nations Research Institute for Social Development Discussion Paper # 60.
- Rao, Vijayendra, and Michael Woolcock (2003). "Integrating Qualitative and Quantitative Approaches in Program Evaluation," in *The impact of economic policies on poverty and income distribution: Evaluation techniques and tools.* Washington, D.C. Oxford and New York: World Bank and Oxford University Press, 165-190.
- Ravallion, Martin (forthcoming). "Evaluating Anti-Poverty Programs", in Schultz, T.P. and John Strauss (eds.), *Handbook of Agricultural Economics*, Vol. IV, Amsterdam: North Holland.
- Scott, J. C. (1999). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed.* Yale University Press.
- Shipton, P., and M. Goheen (1992). "Introduction. Understanding African Land-Holding: Power, Wealth, and Meaning," *Africa: Journal of the International African Institute* 62 (3):307-325.
- Sokoloff, Kenneth, and Stanley Engerman (2000). "Institutions, Factor Endowments, and Paths of Development in the New World," *Journal of Economic Perspectives* 14 (3).
- Swinnen, Johan (2000). "The Political Economy of Institutional Change: A historical perspective on land tenure in Western Europe," Policy Research Group, Department of Agricultural and Environmental Economics, Katholieke Universiteit Leuven.
- The Economist (2005). "Aid to Africa: The \$25 billion question," *The Economist*,.
- (2006). "The mystery of capital deepens: giving land titles to the poor is no silver bullet", September 9-15th.

- USAID (2005). Tierra Americas website 2005 [cited November 1 2005]. Available from <http://www.landnetamericas.org/>.
- van den Brink, Rogier, Thomas Glen, H Binswanger, J.W. Bruce, and Frank F. K. Byamugisha (2006). *Consensus, Confusion, and Controversy: Selected Land Reform Issues in Sub-Saharan Africa*.
- Woodruff, C. (2001), "Review of de Soto's 'The Mystery of Capital,'" *Journal of Economic Literature* 39 (4):1215-1223.
- World Bank (2003). *Land policies for growth and poverty reduction*. Washington, D.C.: World Bank - Oxford University Press.
- (2005). "Gender Issues and Best Practices in Land Administration Projects: A Synthesis Report."
- Land Policy and Administration Web Page, November 1, 2005. [cited. Available from <http://www.worldbank.org/landpolicy/>.