☆

Poverty and Land Ownership in South Africa ☆,☆☆

M. Keswell[a,b,**,1], M. R. Carter [c,*,2], K. Deininger [d,1]

[a]*Southern Africa Labour & Development Research Unit, University of Cape Town, 10 University Avenue, Rondebosch, 7700, Cape Town, South Africa*
[b]*Department of Economics, Stellenbosch University, Stellenbosch, 7602, South Africa*
[c]*Department of Agricultural & Applied Economics, University of Wisconsin Madison, WI 53706 USA*
[d]*Development Research Group, Agriculture & Rural Development Department, The World Bank, MSN MC 3-300, 1818 H St NW, Washington DC 20433*

Abstract

The paper analyses the welfare impact of the current programme of land redistribution in South Africa using a quasi-experimental survey design. We show that the impact of redistribution on household per capita consumption is positive, and remains positive and significant once we have controlled for selection bias. While it is hard to quantify exactly what this means in terms of poverty reduction we find some evidence to suggest that even our lower bound estimates of impact are significant enough to bump households out of poverty in the short term.

Key words: Land Reform, Poverty, Impact Evaluation
JEL Codes: O10, O12, O13

1. Introduction

A vast literature dealing with the consequences of incomplete contracts has shown that wealth and asset inequality can prevent the poor from fully engaging in productive activities, by restricting the types of contracts and exchanges open to them, thereby perpetuating the cycle of poverty. Non-market transfers of assets from the wealthy to the less wealthy might therefore have positive efficiency and poverty reducing effects, in addition to the desired equity enhancements that such transfers bring.[3]

☆Draft, not for circulation. Please do not cite without permission of authors.
☆☆Poverty and Land Ownership in South Africa: Quasi-Experimental Evidence from South Africa.
*Corresponding author
**Principal corresponding author
Email addresses: keswell@sun.ac.za (M. Keswell), mrcarter@aae.wisc.edu (M. R. Carter), kdeininger@worldbank.org (K. Deininger)
[1]Thanks to Simon Halliday, Tim Brophy, Heather Warren and Ronelle Ogle for research assistance.
[2]Thanks to Victor Orozco for research assistance.
[3]See the review in Bardhan, Bowles and Gintis (2000). Also see Legros and Newman (1997), Moene (1992), Mookherjee (1997), Shetty (1987), Banerjee, Gertler and Ghatak (2002).

Historically however, the redistribution of land is one type of non-market transfer that has generally not led to the type of positive effects predicted by this recent work. With a few exceptions, history is littered with failed attempts at reforms that have been undertaken by fiat. In countries where some success has been achieved, little is known about whether the observed improvements in outcomes can be attributed to the policy innovations associated with the transfer of land, or to the other events occuring at the same time as when these innovations were introduced.

On the other hand, reforms undertaken purely through the market-mechanism have been equally flawed mainly because the pre-existing institutions that determine past land distributions might be hard to displace in some instances, thereby muting or reversing reforms aimed at changing the pattern of land distribution.

It is not surprising therefore, that recent interest has centered on countries adopting more "market-friendly" approaches to reform, where the aim is to strike a balance between private interests and state involvement. Land reforms undertaken under this rubric are often described as "negotiated" settlements, where redistribution happens through the market-mechanism but with extensive state and community involvement. The leading latter day examples of this approach are Brazil and South Africa.

The South African case is instructive because of the wide ranging reforms that have been undertaken over the past decade. Having started with a broad-based programme of reforms in the mid 1990's that were aimed primarily at maximising the amount of land transferred, it has since adopted a more selective stance toward redistribution. The current programme requires prospective participants to provide an own contribution towards the purchase price of the land that will be transferred to them. Moreover, like its Brazilian counterpart, the South African land reform programme is also community-based, with extensive participation from the intended beneficiaries at the local level. Yet there remains extensive state involvement in the process, in addition to a host of other role players - local government, NGOs, and lenders to name but a few.

From a purely theoretical perspective, it seems reasonable to conjecture that this type of community-based redistribution would lead to positive impacts on the outcomes of beneficiaries because of: (a) the own contribution required from a prospective beneficiary changes the incentives implicit in the contracts that arise out of these asset transfers by introducing a negative limited liability constraint where there previously would have been none; (b) the community-based nature of the process creates a platform for mobilising support to exercise de-facto rights once established.[4]

This paper focuses on estimating the welfare impact of land reforms in South Africa during the post-apartheid period. Specifically, we seek to estimate the short-term impact of recent land reforms on consumption, using the Quality of Life Survey described in the earlier chapters.

We begin in section 2 by outlining the key challenge – that of statistically identifying the impact of interest. Section 3 then outlines some salient features of the survey design as well as other qualitative work undertaken alongside the survey that speak to the issue of identification *ex ante*. In section 4, we describe how we measure consumption – our outcome variable of interest – as well as a range of additional covariates that we constructed out of the raw data for estimation purposes. We show that these variables are important predictors of treatment status and therefore need to be taken into account when estimating impact. The remainder of the paper then turns to these core estimation issues, beginning in section 5 with a discussion of how we

[4] On the salience of the latter, witness the experience of the village *panchayats* under *Operation Barga* in West Bengal in the late 1970s (Banerjee, Gertler and Ghatak, 2002).

matched using the propensity score. Section 6 then looks at the sensitivity of our estimates of impact to the assumptions underlying the propensity score approach. To test the robustness of our results, we construct an instrumental variable estimator that is aimed at examining the exclusion restrictions implicit in both in the propensity score regressions of section 5, as well as other more parsimonious IV approaches.

We show that the impact of the current program of redistribution on household per capita consumption is positive, and remains positive and significant even once we have controlled for selection bias. While it is hard to quantify exactly what this means in terms of poverty reduction because: (a) the magnitude of this impact tends to vary according to the methods we employ; and (b) there is some controversy over which is the correct poverty line to employ, there is mixed evidence that the lowest estimate of impact we find is significant enough to bump households out of poverty in the short term. These and other issues of interpretation are taken up more fully in section 7, the concluding section of the paper.

2. Evaluating Land Redistribution

Land policy in South Africa has been through several phases encompassing a wide range of reforms in the period since 1995. From inception however, all of these reforms have had the same underlying structure in that they have usually been undertaken through a once-off grant made to beneficiaries followed by voluntary market transactions. The sole purpose of the grant generally is to facilitate the purchase of land. The state's role is to lubricate the bargaining process between the prospective beneficiaries and the seller.

2.1. Programs of Focus

Since 1995 there have been two main grant-making mechanisms for redistributing land under this market assisted framework. The present mechanism is the *Land Redistribution for Agricultural Development* or LRAD program, which was introduced in 2001 and targeted at individual applicants. This program works on the basis of a grant that is awarded to beneficiaries on a sliding scale, depending on the amount of the applicants' own contributions. In practice, the grants are pooled into a fund that is administered on behalf of the beneficiaries by the Land Affairs Department or a Communal Property Association, elected by the members of a "project", where a project is defined as a group comprising individuals, family members, or going concern that will eventually own the land. The impact analysis that follows focuses mainly on this non rights-based program that mandates outright transfers of land.

2.2. Analytical Challenge

The key analytical challenge confronting us is *selection bias*: if there are some special pre-existing features (of the participants or the program) that determines beneficiary status, then any estimate of impact is biased. To see how this problem arises consider the following trivial version of our problem: let y_{1i} refer to average consumption across all households in a given "community" i if the community has been given land titles to their plots, and let y_{0i} refer to average consumption across all plots in this same community i if no land titling had taken place. We are interested in what difference the transfer of land has made to the average consumption of households in this community; i.e., the difference $y_{1i} - y_{0i}$. The problem is that we will never have a given community both with and without title deeds at the same time.

Now given that we have data on many communities, where some communities have title deeds and others not, we could approximate this difference with $\delta = E[y_{1i}|T = 1] - E[y_{0i}|T = 0]$. This estimate, known as the single-difference estimate, is only accurate as an estimate of the impact of the policy when allocation to the beneficiary group (i.e., those communities where land titling takes place) is randomly determined among all eligible individuals. To see why this is so, imagine that we could observe the counterfactual $E[y_{0i}|T = 1]$ - i.e., we can compute average consumption across all households in non-beneficiary communities in an alternative state of the world in which these communities were part of the beneficiary group. Now add and subtract this conditional mean from the one used previously to give:

$$\delta = \underbrace{E[y_{1i}|T = 1] - E[y_{0i}|T = 1]}_{\textit{treatment effect}} - \underbrace{E[y_{0i}|T = 0] + E[y_{0i}|T = 1]}_{\textit{selection bias}}$$

The first term in this expression is what we want to try to isolate: the effect of the intervention on those that received it. We call this the *treament effect*, or more precisely, the average treatment effect on the treated (ATT). The last two terms together constitute selection bias and picks up systematic unobservable differences between treatment and control households. The inability to separate out the treatment effect from selection bias is the our identification problem.

A key focus of this paperis to describe how we dealt with this issue in the study, as it is of first-order importance when trying to statistically estimate the impact of any program or policy. Given the central nature of this issue, and the amount of time we devoted to resolving it in the course of our work, the next section devotes considerable space to discussing the issues arising out of this problem and the multiple ways in which we have sought to address concerns about selection bias.

3. Ex-Ante Identification Strategies

Broadly speaking, the design of the study involved two key innovations that were aimed at minimising selection bias. First, we used a quasi-experimental survey design, with a control-treatment stratification that was limited only to individuals already participating in the program (i.e., they were already in the system at the time of being interviewed). Second, we used an iterative fieldwork design: the study began with a stratified random sample of households to be interviewed. Once the fieldwork had begun, we then embarked on a detailed qualitative study aimed at getting a detailed picture of the most important supply-side factors influencing selection into the treatment group. Once this process was complete, this new information was then used to fine-tune the sampling of the control group by pre-screening projects deemed unlikely to be approved, the main objective being to reduce the level of heterogeneity between beneficiary and non-beneficiary households. Following this, a further round of fieldwork (predominantly focused on the control group was condiucted. This section outlines each of these steps in some detail.

3.1. Phase I

The initial leg of the fieldwork was conducted between September 2005 – January 2006. A two-stage stratified random sample design was followed. In the first stage, collections of households (called projects) that had received land grants were sampled randomly from individual districts within provinces. A similar approach was followed in drawing a comparison sample comprising of households that had applied for but not yet received land grants. In the second stage, a random sample of households from within each project was drawn and interviewed.

Households that were members of projects where land transfer had taken place are thus our treated sample, whereas households still awaiting grant approval (and therefore where transfer has yet to occur) comprise our control group. In total, the sample comprises 1963 households in the treatment group and 1703 households in the control group.

Although the treatment group are a random sample of all treated households, this sampling strategy clearly does not randomly assign treatment status. For this to be true, the sampling frame for assignment to the treatment group should have been limited to the pool of 3666 households at baseline (i.e., before the 1963 treated households received land). The sampling strategy is therefore only quasi-experimental.

Despite this drawback, the selection bias induced by a non-random assignment to the treatment group is likely to be less pronounced in this quasi-experimental design than it would be in a non-experimental design based on some or other synthetic approach to constructing a control group from non-participants. This is because the survey makes use of a "pipeline" strategy where the control group is constituted of *applicants* who are in the pipeline to become beneficiaries. In principle, this approach should attenuate the effect that unobserved individual differences in characteristics will have on an individual's decision to participate. Therefore selection biases emanating from pre-existing differences between participants and non-participants can safely be assumed to be approximately zero, since the treatment effect is only defined for participants.

In this respect, our study design is similar to perhaps the most famous example of a pipeline strategy – the Program for Education, Health and Nutrition (PROGRESA) (now called *Opportunidadas*), which was introduced by the Government of Mexico in 1997. However, unlike PROGRESA, because our pipeline comparison is not randomised, our sampling strategy can only partially resolve the selection bias problem.

The best possible sample to estimate the treatment effect on would have been one where assignment to the treatment group was completely random: i.e., one where each control group project has an equally high probability of becoming a treatment group project as the next. In our case, this was explicitly not the case when the control group was formed as the data on applications was too coarse for a call to be made on the likelihood of this – the level of detail required meant that much of this information was qualitative in nature, scattered across numerous paper records kept at relatively remote locations at the district level, and therefore was not easily usable for sampling purposes.[5] Yet the import of the pipeline approach as an identification strategy rests on getting such such details right.

To address this issue we combined the pipeline design with a more careful matching exercise in order to reduce the heterogeneity among the potential control group projects in the pipeline. Our approach to matching has two distinct components. First we embarked on a largely qualitative exercise of refining our sampling by screening out control group projects unlikely to be approved. Second, we combined our pipeline comparison with a statistical matching exercise based on the propensity score (Rosenbaum and Rubin, 1983). The combination of these two approaches is sometimes referred to as "pipeline-matching" in the evaluation literature. Examples of its application in other contexts can be found in Chase (2002), Galasso and Ravallion (2004), and Ravallion (2007). In section 5, we outline the details behind this approach. However, before proceeding to this discussion, we first discuss the *ex-ante* screening we subjected the sample to.

[5] Indeed, much about the administrative, legal, and regulatory frameworks governing land reform in South Africa suggests that one should anticipate a great deal of heterogeneity in the applications entering the pipeline. Stated differently, the complex labyrinth of hurdles entailed in gaining approval for a grant exists precisely to screen out inviable project applications.

3.2. Phase II

Applications that are in the pipeline to become beneficiaries have to pass through several key milestones before final approval of the grant is obtained. At each milestone, projects are either approved to pass on to subsequent stages, referred back to the the government appointed planner for further development, or rejected altogether. Failure to reach a required milestone is therefore measurable, and such information could therefore be used in principle as an indicator of the likelihood of eventual selection into the treatment group. When the sample for the control group was initially drawn, this level of information was not available and so the control group contained applications across the full spectrum of approval likelihood. It therefore became imperative to collect this information where possible in an attempt to fine-tune the pipeline strategy.

During November 2005 – April 2006, and then again during January – May 2007, extensive qualitative work was carried out with a view to factoring in some of this detail for sampling purposes. In order to see clearly how we made use of this information, it is necessary to briefly outline the stages an application passes through before it reaches the final approval milestone. The key stages are as follows:

Stage 1 (Project Registration): Once an application is received, the state appointed official (hereafter referred to as the "planner") does a needs assessment by visiting the site on which the applicants live as well as the land they have applied to purchase (which need not coincide with the the current place of occupancy of the applicants). Once the application has been verified, the application is "registered" as as candidate land redistribution project.

Stage 2 (Approval of Planning Grant) The planner then asks the district line authority of the land affairs department to release a nominal sum of money to begin developing a proposal on behalf of the applicants. The funds are meant to be used to commission various specialised activities that will culminate in a portfolio of sorts that will ultimately be used by the planner both in negotiating a purchase price for the land, as well as in motivating the grant application to the state in the final analysis. Examples of such activities are valuations, soil assessments, quantity surveys, and business plans.

Stage 3 (Preparation of Project Identification Report): Once these commissioned studies start to materialize, the planner begins to collate a report that summarizes the merits of the application. This document, which is called the project identification report (PIR), is the first important milestone that can be used to measure the likelihood of approval. The existence of this document indicates that the application was serious enough to warrant the release and expenditure of state resources to begin making the case for the grant.

Stage 4 (Approval of District Screening Committee): The planner then submits this document to a district-level screening committee of the land affairs department. This group then screens out all applications deemed inviable, too expensive, or incongruent with infrastructure roll-out plans by local municipalities. An application that is not approved by the district screening committee (DSC) is generally referred back to the drawing board, so to speak, if not rejected altogether. The primary purpose of the DSC is to vet applications so as to improve their likelihood of approval when submitted for consideration to the provincial grants approval committee (PGAC).[6]

[6]The PGAC is the main grant-making authority. It usually has broad representation from all role players including officials from the agriculture department, surveyor general's office and local municipalities.

Stage 5 (Approval of Provincial Government): Once an application has been approved by the DSC, a formal request to designate the land for redistributive purposes is made. At this stage a quasi-legal document called the "designation memo" is prepared, which is what the provincial grants approval committee deliberates over when making their final decision. This document must ultimately be signed by the directors general and minister of land affairs and agriculture. A key hurdle of these meetings that applications usually have to overcome is that there must be complementarity around basic service provision (roads, irrigation, electrification), before the PGAC gives it's final approval.

The above process conveys the sense in which land reform in South Africa is both market-assisted and state-negotiated. While in practice this process tends to vary by province in terms of the details, the broad stages outlined above tend to be fairly standardised.

Our qualitative work centered around collecting project identification reports and designation memos for all control group projects. In the course of this activity, we travelled to many of the land affairs district offices to interview planners and delve into archived records of projects to locate this information. Our goal was to collect updated information on pipeline projects and thereby piece together a timeline. Generally, if a PIR or designation memo could not be found for a given project, a replacement project was found that did meet this criteria. This requirement effectively screened out any observations in the control group that had not passed at least stage 2.

For projects that were at stage 3, we then had to ascertain, through a process of interviewing land affairs officials, whether any further progress had been made that had not yet been reflected in the archived records. Ultimately, we needed to make sure that we only selected pipeline projects that had at least passed stage 4 so that no dormant, rejected, or disbanded projects were included.[7] Our study of the administrative process governing land redistribution suggests that passing the stage 4 milestone tends to be a key predictor of grant approval and so we screened out any applications that ultimately did not meet this criterion.

4. Data Description

4.1. Outcomes

While a number of possible outcomes could be considered, for the purposes of this study we use per capita consumption expenditure as our welfare metric since we are interested in the impact of land transfers on poverty alleviation.[8] We explicitly do not consider using a binary indicator of poverty status since this is arguably a more restrictive approach, as Ravallion (2007) has argued.[9]

[7] An application could be rejected or become dormant for several reasons. The two most commonly cited reasons were: (a) complications surrounding the pending sale agreement (e.g., renegotiation over the the offer to purchase), or; (b) some aspect about the proposed productive enterprise was deemed infeasible by the PGAC, such as the size and/or suitability of the land to be designated. An application could also be "de-registered" (i.e., rejected outright) because of a competing claim through the restitution programme.

[8] We have conducted our analysis using alternative measures of welfare – for example, measures of consumption based on the number of adult-equivalents in the household. These alternate measures of consumption do not change any of our substantive conclusions or interpretations, so we only report on the per-capita measure of consumption.

[9] The basic point here is that collapsing a continuous welfare metric such as expenditure or income into a binary indicator amounts to throwing away information. We also do not normalise expenditure by a poverty line, because there is some controversy in the case of South Africa as to which is the most appropriate line to use (Woolard and Leibbrandt, 2007).

Table 1: Mean Per Capita Consumption

Program	Total	Treatment	Control	N	p-val	Δ
All	459.17	453.26	465.99	3666	0.58	0
LRAD	497.52	547.76	472.61	1925	0.05	+
SLAG	375.51	373.55	386.93	456	0.84	0
Restitution	487.30	471.29	550.18	596	0.16	0
Tenure Reform	307.22	280.66	365.12	493	0.07	–
All	5.67	5.65	5.69	3665	0.22	0
LRAD	5.73	5.78	5.70	1940	0.05	+
SLAG	5.50	5.48	5.62	460	0.24	0
Restitution	5.41	5.37	5.49	498	0.09	0
Tenure Reform	5.77	5.74	5.91	606	0.05	–

The measure of consumption used in this table and throughout the paperis per capita consumption expenditure in 2005 Rands. The first five rows are in levels, wheres the last five are logged values.The second last column shows the p-value for a two-sided *t*-test for equality of means between the treatment and control groups, and the last column shows the sign of the difference, accounting for whether it is significant or not. The total treated sample consists of 1963 households, whereas the total control sample consists of 1703 households.

Table 1 shows the means and standard deviations of the dependent variable by programme. What is immediately noticeable is that consumption appears smaller in the treatment group when we aggregate all programs, although this difference is not statistically significant . One reason that might account for this is that the earlier programme of redistribution (SLAG) was less restrictive because of the absence of an own-contribution. Thus participation in the older program is likely to have exhibited a greater extent of unobserved heterogeneity, and larger fraction of poorer households, than participation in the the current program (LRAD). Moreover, the problem is potentially exacerbated by the presence of the rights-based programme, where one would expect an even greater degree of heterogeneity among beneficiaries. The two programs for which this naive estimate of the treatment effect is significant are LRAD (positive) and Tenure (negative). In what follows, we limit our focus to the LRAD program. Our main purpose is to investigate whether the apparent positive effect of the LRAD program is robust to corrections for selection bias.

4.2. Covariates

In this section we give a brief description of the variables that will be used in the the analysis to follow. Table 2 presents a summary of tests for differences in means for these variables between the treatment group and control group before matching on the propensity score.

One of the strategies we have followed is to construct variables that could mirror in a quantitative setting what we set out to do in the screening processes discussed above. We do this for two reasons. First, the qualitative information we used in our screening exercise is by nature imprecise. Second, a not insubstantial number of projects were not subjected to this screening process because they were interviewed during phase I.

We therefore use responses from the survey to construct two variables that will be put to use in our various econometric methods to follow. The first of these is the variable *Doserec* which measures the number of days elapsed between the date of grant approval and date of interview.[10]

[10]We do not actually observe the date of approval in the survey. However, it is possible to proxy it with the question

The second variable, called *DoseIV*, captures the length of time spent in the pipeline. This variable measures the speed with which an application is approved, and is given by the number of days elapsed between the date of application and date of grant approval. In section 5, we use *DoseIV* as a key regressor of interest.

Table 2: Test of Difference in Means for Covariates

Variables	Total	Treatment	Control	N	p-val
Number employed in agriculture	0.54	0.77	0.44	1725	0.00
Log days in pipeline	6.74	5.94	7.08	1725	0.00
Days in pipeline (DoseIV)	1423.26	844.27	1666.97	1725	0.00
Days since treatment (Doserec)	352.01	1188.30	0.00	1725	0.00
Household head is male	0.69	0.76	0.67	1725	0.00
Education of household head (yrs)	5.98	6.31	5.85	1725	0.06
Mean farming experience (yrs)	1.51	1.62	1.46	1725	0.40
Number plots accessed pre-95	1.15	0.65	1.34	1663	0.00
Distance to DLRO (100 km)	0.93	0.94	0.92	1718	0.54
Area plots accessed pre-95 (hectares)	51.55	31.60	59.18	1663	0.26
Land allocated by municipality (post-94)	0.13	0.03	0.21	916	0.00
Land allocated by other farmer (post-94)	0.09	0.00	0.15	916	0.00
Land allocated by tribal authority (post-94)	0.06	0.00	0.09	916	0.00

The last column shows the p-value for a two-sided *t*-test for equality of means between the treatment and control groups.

As discussed earlier, our qualitative work included extensive interviews with Land Affairs officials involved with actual implementation. During these discussions it was often reported by planners that 3-6 months is a good rule of thumb for the length of time it takes for a "good" application to be approved once an application has been officially registered.[11] If the approval process proceeds smoothly, then transfer of the land often happens more or less predictably. However, it is often the case that if even one of the milestones is held up, the approval timeline is rendered unpredictable. What is clear is that the longer an application takes to meet these milestones, the less likely it is that the sale agreement will be signed by the seller. Therefore the length of time in the "pipeline" will vary negatively with a household's probability of being in the treatment group.

Another variable of importance that will feature in subsequent analysis is the variable *Distance to DLRO*. This variable is constructed by using the georeferencing of interview sites to map the shortest distance that would need to be travelled by road from each visiting point to the nearest land affairs office in the same district as the visiting point.

We also constructed a range of additional variables to be used as controls in regressions to follow, because they describe some aspect of program structure or emphasis that we hypothesize to be important. The variable *Number employed in agriculture* refers to the number of individuals in the households that reported some history of working on a farm or other agricultural enterprise.

"When did [. . .] first receive a grant from the Department of Land Affairs?" We also assume that the household had not been a grant recipient before that point. While this possibility is not specifically precluded by the LRAD program rules, our qualitative work on the approval process suggests that such occurrences are not likely to be practically important as they are extremely rare.

[11] Once an application is registered, a needs-assesment meeting with the applicants is conducted, a land-valuation is conducted, a business plan is drawn up, an agricultural assessment report is prepared and a draft offer to purchase is prepared and presented to the prospective seller. Extensive workshops with key role-players are conducted throughout the process.

Since LRAD is a grant targeted towards agricultural activity in the first instance, this variable is likely to feature prominently in predicting selection into the treatment group, as is the variable *Mean farming experience*, which averages the farming experience over all household members. Likewise, the variable *Household head is male* is likely to be important as LRAD emphasizes the targeting of women.

Finally, we constructed a set of dummy variables meant to capture previous access to land. This category of variable is likely to matter because it would likely disqualify participation when observable (to the planner) but would introduce a confound on the treatment effect when not observed (again by the planner). Thus, it is plausible to expect these variables to be negatively related to treatment status.

Table 2 shows that most of these variables are partially correlated with treatment status in that the p-value of the test of equality of means in most cases turns out to be significant, illustrating the point that assignment to the treatment group is clearly not random.

5. Estimating Impact – Pipeline Matching

As with any quasi-experimental design, the lack of complete randomization necessitates the *ex-post* use of non-experimental statistical methods in order to construct the types of counterfactual cases that would approximate that of an idealised experiment. In the case of our study, this is further necessitated by the fact that some control group projects were not subjected to the the type of *ex-ante* screening discussed in section 3.2.[12]

The main analytical approach we have followed in this regard was to combine the pipeline design with propensity score matching – a technique sometimes referred to in the evaluation literature as pipeline matching. This section reports on how we went about applying this methodology as well as the results obtained.

5.1. Key Assumptions

The key idea behind matching methods is to match treatment households to control households on the basis of characteristics we can observe about the actual household, and thereby remove the selection bias induced by the role played by these observable characteristics in affecting selection into the treatment group. The treatment effect is then computed by taking the average of the difference in mean outcomes for that subset of the data for which the match is a good one. Exactly how this is done is taken up in section 5.4.

Ideally, we would want to match individuals/households directly on their characteristics. Angrist (1998) provides a good illustration. However, this technique of exact matching is often not practical. There are two reasons for this. First, when some of the more important variables we wish to condition on are continuous, we would need to find a useful way of transforming the relevant variables into a discrete form. Secondly, when the number of covariates we wish to match on is of large dimension, then we often run into degrees of freedom problems. As an example, consider what would be required if we tried to match exactly using 11 of the covariates described in table 2. The simplest possible match we could make is to divide each covariate into just two levels, say above or below its median. This would result in $2^{11} = 2048$ possible patterns for which we would need matches. As table 1 reports, we only have 1703 unique control

[12]These projects were interviewed at an early stage in the study before the need for such a screening exercise was identified.

households. Moreover, we would probably need to group the data more finely if our main goal is to minimize selection bias (Cochran, 1968 as cited in Rosenbaum, 2004). Assuming that instead we divided the covariates into 4 rather than 2 groups (say, by quartiles of the distribution of each covariate), then we would need $4^{11} = 4194304$ control group observations.

A now standard technique used to address such data issues is to match not on the multidimensional vector of covariates but rather on a scalar index such as the propensity score – i.e., the predicted probabilities that are computed from a regression where the outcome variable is a binary indicator of treatment. There are two variants of this approach. The approach used here is the standard model of the propensity score (Rosenbaum and Rubin, 1983; Heckman and Robb, 1985; Heckman, LaLonde and Smith, 1999) where we use a binary variable for our treatment measure (i.e., either a household is in the treatment group or it is not).

Formally, if we let $\mathbf{x}$ be a vector of pre-treatment variables, then we can define the propensity score as the conditional probability of receiving the treatment T, given $\mathbf{x}$

$$p(\mathbf{x}) = \Pr[T = 1|\mathbf{x}] = E[T|\mathbf{x}]$$

For the purposes of the analysis to follow, two key theoretical results proved by Rosenbaum and Rubin (1983) are noteworthy:

Lemma 1. *(Balance): If $p(\mathbf{x})$ is the propensity score, then $\mathbf{x} \perp T|p(\mathbf{x})$. Stated differently, the distribution of the covariates for treatment and control is the same once we condition on the propensity score: $F(\mathbf{x}|T = 1, P(\mathbf{x})) = F(\mathbf{x}|T = 0, P(\mathbf{x}))$*

Lemma 2. *(Ignorability): If there is no omitted variable bias once $\mathbf{x}$ is controlled for, then assignment to treatment is unconfounded given the propensity score.*

The first result says that once we condition on the propensity score, assignment to the treatment group is random. In the limit, for two identical propensity scores, there should be no statistically significant differences in the associated $\mathbf{x}$ vector, independent of how these scores are distributed between the treatment group and the control group. This property must be met if we are to move forward after computing the propensity score.

The second result says that selection into treatment depends only on what we can observe, i.e., $\mathbf{x}$. In other words, while the propensity score balances the data (i.e., removes the influence of the observables on assignment to the treatment group), it also assumes no confounding on the basis of unobservables. Whether or not this assumption is plausible rests on whether the specification of the propensity score regression accurately reflects the key factors that might influence the process of selection.

Our strategy in this regard is to represent features of the selection and screening process which we know to be important from our qualitative work, using the variables described in table 2. Where we can construct variables that relate to the targeting of LRAD in the sense that such variables are directly observable (like the fact that the program emphasizes the targeting of women), then the variable in question is not to be interpreted as a proxy. A variable like *DoseIV* on the other hand is to be interpreted as a proxy variable as it summarizes all unobservable factors that influence the speed of progression through the pipeline. Unobserved effects that are orthogonal to *DoseIV* or the other explanatory variables remain a black box.

This is a potential problem for the propensity score approach, since unobserved effects can only be picked up through observable proxies. To test the sensitivity of our estimates to this possibility, one needs to make different assumptions to Lemma 1 and Lemma 2. One possible

alternative is to use instrumental variables (IV) methods. If *DoseIV* were used to instrument treatment status in a regression predicting consumption, then any remaining unobserved effects must be orthogonal to *DoseIV* in order to avoid confounding the treatment effect. Therefore, if our treatment effect based on the propensity score is confounded because Lemma 2 does not hold, a good way of checking for this confound is to use an IV approach. We return to this issue in section 6.

5.2. Specification Issues

While the propensity score regression is of immediate interest to us as it serves as a diagnostic tool for describing how well we have captured the latent process of selection into the treatment group, we pay only passing attention to the magnitudes of the estimated coefficients because ultimately our main interest is in estimating the *average* treatment effect on beneficiaries. Since we are less interested in magnitude, this would seem to suggest running a linear probability model but we impose the restriction that $\hat{p}(\mathbf{x}) \in [0, 1]$, because Lemma 1 is predicated on this assumption. Since the logistic distribution imposes this restriction by construction, we use a logit regression to model the propensity score. Of course, there are many other reasons why one would want to do this, but one practical reason has to do with the fact the the linear probability model would require a re-scaling of the propensity score distribution before a test of the balancing property can be performed, whereas the logit (and probit) obviates the need for such an exercise.

Table 3 shows two specifications of the logit regression. The table reports the index coefficients and not the marginal effects for the reasons pointed out above. The first specification excludes the variables relating to past access to land as well as our distance measure, whereas the second specification includes these variables. Immediately noticeable is the fact that the sample size is much larger for the first specification than the second. In part, this has to do with the fact (as evidenced by their significance in table 3) that the variables relating to previous access to land are more likely to negatively predict selection into the treatment group, and this appears to have been anticipated by the survey respondents themselves as reflected in the poor response rate received on those questions in the survey by both applicants and beneficiaries.[13] To account for this possible bias, we compute average treatment effects on the treated for both sets of regressions.

The number of days spent in the pipeline has a negative estimated effect. This finding seems reasonable: applications that spend longer in the system are more likely to become dormant. In spite of the targeting of women, female-headed households seem to suffer a distinct disadvantage in getting into the treatment group. The fact that the coefficient on education is negative could be interpreted as evidence that LRAD is predominantly a program affecting rural households, and therefore average education on balance is likely to be lower compared to individuals in the control group who are more heterogenous in general. Of course it might also be true that the temporal dimension that separates the treatment group from the control group might account for this negative coefficient in the sense that the first applications to be approved under the LRAD program (i.e, say during the first year of operation when the program rules were not fully understood on the implementation side), were made by relatively poorly educated individuals. We remain agnostic on this point, but merely offer this reasoning as a possible interpretation.

[13] This result also resonates with a widely held perception picked up during the fieldwork that past access to land would serve to disqualify an applicant, even though this is not explicitly stated in the program rules. It is therefore not surprising that many respondents evidently chose not to answer questions relating to past access to land.

Table 3: Propensity Score Regressions

Variable	(1)	(2)
Number employed in agriculture	.375 (.055)***	.645 (.107)***
Log days in pipeline	-.761 (.063)***	-.844 (.104)***
Household head is male	.302 (.132)**	.616 (.200)***
Education of household head (yrs)	-.004 (.013)	-.069 (.020)***
Mean farming experience (yrs)	-.004 (.016)	-.016 (.023)
Number plots accessed pre-95		1.017 (.181)***
Distance to DLRO in 100 Km		.238 (.146)
Size of plots accessed pre-95 (Hectares)		.00004 (.0004)
Ever been allocated land by the municipality (post-94)?		-2.180 (.351)***
Ever been allocated land by other farmer (post-94)?		-4.915 (1.094)***
Ever been allocated land by the tribal authority (post-94)?		-4.649 (1.084)***
Const.	3.813 (.457)***	4.993 (.760)***
N	1725	913

The regressions are based on the logit model. The dependent variable equals one if the household is in the LRAD treatment group and zero if it is in the LRAD control group.

5.3. Testing the Balancing Property

A key challenge in getting the right specification for the propensity score is making sure that the balancing property is satisfied. Practically speaking, the balancing property of the propensity score implies that we need to make sure that the control group and beneficiary group are not statistically different from each other, once we've conditioned on $\mathbf{x}$. This requires that we check that $\mathrm{E}(p(\mathbf{x})|T = 1) = \mathrm{E}(p(\mathbf{x})|T = 0)$ as well as that $\mathbf{x} \perp T_i|p(\mathbf{x})$. On way to accomplish this test is to aggregate the estimated propensity score $\hat{p}(\mathbf{x})$, into mutually exclusive intervals (blocks) over its distribution and then check that the average propensity score within each block is uncorrelated with treatment assignment. Then using this same procedure, we can then check that each covariate is uncorrelated with treatment assignment within each block.

This obviously means that the balancing property can only be tested in proximate sense. We have used the algorithm proposed by Dehejia and Wahba (2002), as encoded in the implementation developed by Becker and Ichino (2002). The approach works by arbitrarily grouping the data by blocks (intervals) of the propensity score, where initially the scores within a block are quite similar. An equality of means test between treatment and control observations is performed for each of the regressors contained in $\mathbf{x}$. If there are no statistically significant differences between treatment and control for each of the covariates in the propensity score regression, then the regressors are balanced. If a particular regressor is unbalanced for a particular block, then that block is split into further groups and the test is conducted again. This iterative process continues until all the regressors are balanced or the test fails. Tables 4 – 5 shows a summary of our results

from testing the balancing property using the Dehejia and Wahba (2002) algorithm.[14]

There are 6 blocks in the final analysis, and as table 4 shows, in each case the computed t-statistic for the equality of means of the propensity score in each block is smaller than the associated critical value (which in turns depends on the sample size within each block). The null hypothesis that the means of the propensity score are the same within each block is therefore not rejected.

Table 5 essentially shows the results of a similar test, but in this case we test the null that $\mathbf{x}$ is balanced across the various blocks. The table reports that the computed t-statistics in each case is less than the critical values shown in table 4, thus confirming that $\mathbf{x}$ plays no role in predicting selection into the treatment group once we have conditioned on the propensity score.

Table 4: Propensity Score Balance

Block	$\min \hat{p}(\mathbf{x})$	N_0	N_1	$\bar{p}_0(\mathbf{x}) - \bar{p}_1(\mathbf{x})$	SE	t	t_{cv}
1.00	0.03	133.00	11.00	-0.01	0.16	-0.66	2.58
2.00	0.20	64.00	23.00	-0.02	0.01	-2.50	2.64
3.00	0.30	48.00	29.00	-0.01	0.01	-1.81	2.64
4.00	0.40	80.00	73.00	-0.01	0.01	-1.55	2.58
5.00	0.60	38.00	119.00	0.00	0.01	-0.32	2.58
6.00	0.80	19.00	139.00	-0.03	0.02	-2.19	2.58

"Block" refers to an interval placeholder from among 6 mutually exclusive intervals of the propensity score distribution. These intervals are defined by the cut-off points given by $\min \hat{p}(\mathbf{x})$. The fifth column in the table reports on the magnitude of the difference in means for the propensity score between treatment and control for each block. t refers to the t-statistic for testing that the reported difference in column 5 is significant.

Table 5: Covariate Balance

Variable	1	2	3	4	5	6
$onfarmemp$	-0.01	0.88	-0.30	-0.42	-0.05	-0.90
$ldoseIV$	1.81	2.55	-0.55	-0.40	-0.63	1.79
$sexhhead$	-0.46	0.68	1.41	-0.59	0.12	-0.40
$hheadeduc$	-1.01	0.92	0.98	0.11	-0.80	0.40
$farmexper$	-1.53	0.44	1.88	-0.89	0.67	-0.19
$pre95sum$	0.75	-0.70	0.53	-1.40	-1.02	-0.07
$dist100$	-0.90	-1.53	0.87	-0.19	0.74	0.53
$pre95size$	0.29	1.05	0.77	-1.83	-0.72	-0.60
$MUNpl$	0.21	-1.81	1.37	0.96	1.78	-0.37
$FARMERpl$	-0.68	–	–	–	–	–
$TRIBALpl$	0.58	–	–	-1.05	–	-0.37

The table shows that the covariates are balanced once we condition on the propensity score. The column headings refer to the 6 intervals of the propensity score distribution within which the estimated propensity score is balanced. The entries in each table report the t-statistic for an equality of means test of each regressor by treatment status.

[14]These test statistics are based on the second logit specification. We omit the diagnostic detail pertaining to the first specification, but the balancing property is also satisfied in that case.

5.4. Calculating the Average Treatment Effect

Our method of estimating the average treatment effect (ATT) rests ultimately on two approaches which can be viewed in some senses as being at opposite ends of the spectrum in terms of the trade-off between bias and efficiency. Non-parametric methods are an attractive option as they are very efficient (little or no loss of information), but when $\mathbf{x}$ contains more than three covariates, the problem of dimensionality arises. Parametric methods however work better when $\mathbf{x}$ is of large dimension but this class of approaches will typically be based on much smaller sample sizes than other alternatives. We use three different approaches, each reflecting this trade-off between bias and efficiency.

5.4.1. Blocking on the Propensity Score (Stratification)

Our first method is based directly on the blocking (or stratification) of the propensity score shown in tables 4–5. Our tests of the balancing property have already demonstrated that within each block, the treated and control households have, on average, the same propensity scores. A somewhat natural way to compute the treatment effect then is to take the difference between the mean consumption of the treated and control groups within each block, and weight each of these differences by the distribution of the treated households across the blocks in order to get the average treatment effect for the treated households. Formally, let i denote the ith treated household; let j denote the jth control household, and let b denote the bth block. Then a block-specific treatment effect is

$$\text{ATT}_b = (N_{b,1})^{-1} \sum_{i \in I(b)} y_{1i} - (N_{b,0})^{-1} \sum_{j \in I(b)} y_{0j}$$

where I_b is the set of households in the bth block, and where $N_{b,1}$ and $N_{b,0}$ are the subsets within I_b that fall either into the treatment group or control group. To get the average treatment effect by the method of stratification, we simply weight each of these block-specific treatment effects by the proportion of treated households falling into each block, and then sum the resulting weighted block-specific treatment effects over all 6 blocks Thus,

$$ATT^{Strat} = \sum_{b=1}^{6} \text{ATT}_b \times \frac{\sum_{i \in I_b} D_i}{\sum D_i}$$

5.4.2. Nearest-Neighbor Matching

The second approach we take is to match each treated household to the control household that most closely resembles it. There are various ways in which this can be done, one of which is to match directly on a chosen linear combination of $\mathbf{x}$, but given Lemma 1, a better way to proceed is to match on the propensity score. Since $p(\mathbf{x})$ is a scalar index, this method has the advantage of permitting a greater number of matches than matching directly on $\mathbf{x}$ would allow.

Formally, we can define the set of potential control group matches (based on the propensity score) for the ith household in the treatment group with characteristics $\mathbf{x}_i$ as

$$A_i(p(\mathbf{x})) = \{p_j | \min_j |p_i - p_j|\}$$

Again, there are a number of ways to implement this method. The most restrictive form of the nearest neighbor method would select a unique control group household for every treatment group household on the basis of computing the absolute value of the difference in propensity

scores for every pairwise match considered, and then selecting as a match the *j*th household with the smallest absolute difference in propensity scores.

Alternatively, all observations in the set $A_i(p(\mathbf{x}))$ could be matched against household i. In this case, a differential weight would be applied to each match falling into the matching set. The average treatment effect would then be computed as follows:

$$\text{ATT}^{NN} = (N_1)^{-1} \sum_{i \in \{T=1\}} (y_{1i} - \Sigma_j \omega(i,j) y_{0j})$$

where j is an element of $A_i(p(\mathbf{x}))$ and $\omega(i,j)$ is the weight given to j. For the restrictive one-to-one match mentioned above, we would then have $\omega(i,j) = 1$ when $j \in A_i(p(\mathbf{x}))$, and $\omega(i,j) = 0$ when $j \ni A_i(p(\mathbf{x}))$.

5.4.3. Kernel Matching

An alternative method, closely related to nearest-neighbour matching is to match non-parametrically using a kernel function. In this instance our formula for the ATT is as above, but the weight given to the *j*th control group household in matching it to the *i*th treated household is determined as follows

$$\omega(i,j) = \frac{K(p(\mathbf{x}) - p(\mathbf{x}))}{\sum_{j=1}^{N_{0j}} K(p(\mathbf{x}) - p(\mathbf{x}))}$$

$$K = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{p(\mathbf{x})^2}{2\sigma^2}}$$

where K is the Gaussian (normal) kernel. This method has the benefit of using the entire sample for each prediction with decreasing weights for more distant observations, where the rate of decline of these weights is determined by σ. In principle, ω could be determined in other ways (e.g., tri-cubic, caliper etc.) We present both nearest-neighbour and kernel estimates as a way of offering some way of controlling for the relative trade-offs between bias and efficiency of these two methods.

5.4.4. Specification of Support

A final challenge concerns defining a reliable range over which the ATT is valid. Because the range of values encompassed by the estimated propensity scores differ between the treatment and control samples, it only ever makes sense to compute the ATT for the intersection of the two supports. To accomplish this, let α and β denote the limits of the common support for the two propensity score distributions such that $0 < \alpha < p(\mathbf{x}) < \beta < 1$. We then trim the data such that all observations with a propensity score of less than α or greater than β are dropped before calculating ATT. Table below (table 6) shows our estimates of the average treatment effect, using the three approaches discussed.

Table 6: Summary of Treatment Effects

Program	Method	Definition	$T = 1$	$T = 0$	ATET	SE	t
LRAD	Single Difference	Per capita			75.18	37.97	1.98
LRAD	Stratification*	Per capita	511	2154	143.93	56.43	2.55
LRAD	Stratification	Per capita	394	1047	149.87	84.42	1.78
LRAD	Nearest Neighbour*	Per capita	511	303	65.32	73.16	0.893
LRAD	Nearest Neighbour	Per capita	394	143	148.28	91.54	1.62
LRAD	Kernel*	Per capita	511	1063	134.24	54.88	2.45
LRAD	Kernel	Per capita	394	382	169.18	77.55	2.18
LRAD	Stratification*	Log per capita	511	1063	0.19	0.07	2.57
LRAD	Stratification	Log per capita	394	1047	0.09	0.09	0.97
LRAD	Nearest Neighbour*	Log per capita	511	303	0.15	0.09	1.72
LRAD	Nearest Neighbour	Log per capita	394	143	0.03	0.11	0.31
LRAD	Kernel*	Log per capita	511	1063	0.17	0.05	3.23
LRAD	Kernel	Log per capita	394	382	0.10	0.079	1.29
LRAD	Stratification*	Household	511	1063	592.87	208.76	2.84
LRAD	Stratification	Household	394	1047	765.55	329.79	2.32
LRAD	Nearest Neighbour*	Household	511	305	603.92	291.25	2.07
LRAD	Nearest Neighbour	Household	394	143	911.38	337.94	2.70
LRAD	Kernel*	Household	511	1063	545.17	202.24	2.70
LRAD	Kernel	Household	394	382	867.15	276.94	3.13

The table summarizes estimates of the average treatment effect on the treated (ATT) using three different methods. Differences in sample sizes when using logged per capita consumption as opposed to level per capita consumption are the result of the combined effect of matching and trimming. The estimates based on nearest neighbour and kernel matching make use of the estimated propensity score and are not based directly on $\mathbf{x}$. Nearest neighbour matching is done with replacement. Stratification matching is based directly on the blocking identified in tables 4 and 5.

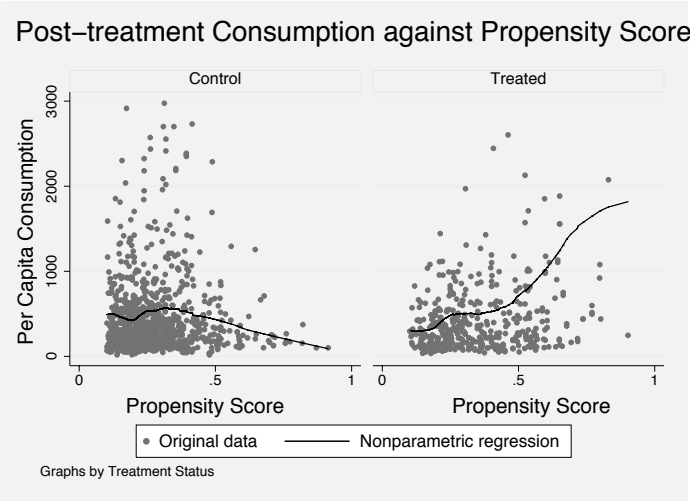

Figure 1: Semi-Parametric Treatment Effect on PCE

To get a better feel for the matched data on which these estimates are based, figures 1 and 2 shows a locally weighted regression of real per capita consumption expenditure (levels and logs) against the estimated propensity score, for treatment households and control households. The gradients of the fitted curves are clearly positive in both cases.

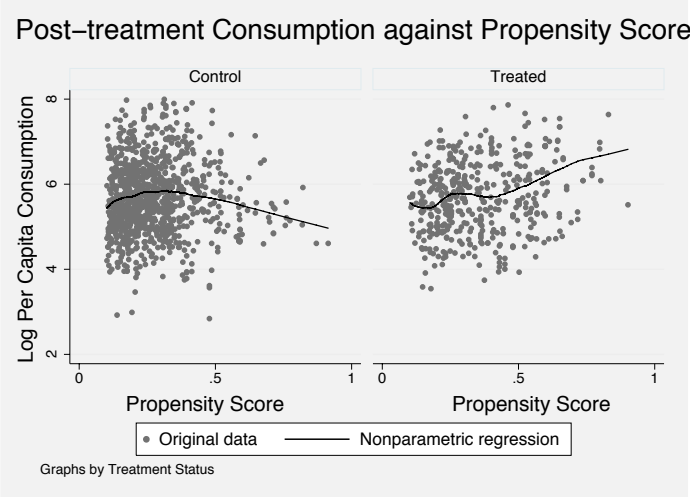

Figure 2: Semi-Parametric Treatment Effect on Log PCE

5.5. General Equilibrium Effects

As outlined earlier, a key factor in predicting whether an application is approved or not is whether the timing of the intended transfer is in line with the provision of basic services that would be required by the proposed enterprise. For example, a proposed commercial livestock enterprise (involving the transfer of land for grazing purposes) might require the establishment of dipping services, which would be difficult to operate without the availability of electricity. Therefore, it is not altogether uncommon for an otherwise promising enterprise proposal to get rejected at stage 5, because the relevant local municipality could not commit to rolling out electricity services to the district in question. But, in the vast majority of cases, planners will try to gain the support of the municipality ahead of the approval committee meeting, so that there is some guarantee that electricity would be rolled out to the area before the transfer date.

It is possible that for two communities that are spatially close to one another, and distinguished only by their treatment status (one community falls into the treatment group whereas the other is in the control group), that the benefits of complementary programs that were required for the first project to be approved, would now also extend the second group of households picked up by our survey as belonging to the control group. In such an instance, the treatment effect would then be confounded by the presence of a spillover effect from a complementary program.

While such an occurrence might might not be practically important, we attempt to address the issue nonetheless. The approach we adopt can be considered as a variant of the nearest-neighbourhood estimator. In short the approach amounts to minimizing the Euclidean distance between pairwise treatment and control propensity scores, while at the same time maximizing the spatial distance between the georeferenced visiting points that correspond to the neighborhood under consideration when doing the pairwise matches. Doing this calculation results in 394 matched pairs and a treatment effect on per capita monthly consumption expenditure of R29.18 and R612.85 on household monthly consumption expenditure.

6. Sensitivity to Alternative Statistical Assumptions

Two potential problems remain unexplored with the propensity score approach. The first, discussed already, concerns the possibility of remaining omitted variable biases. Since matching is based explicitly on the predicted probabilities of the treatment status regression, the presence of any remaining omitted variable bias will have a direct consequence for the resulting match, and therefore the estimated average treatment effect.[15] Fixing this problem requires an alternative solution to the omitted variable bias problem. The standard approach is instrumental variable estimation. In this section, we present an alternative estimator of the average treatment effect that exploits this feature of IV estimators (i.e., that unobservables don't bias the treatment effect as long as the instrument is uncorrelated with omitted factors).

This approach also allows us to relax some of the exclusion restrictions required by the propensity score regression. The linear projection implicit in the propensity score is based on the idea that if there are differences in the observable dimension, these are fully accounted once matching takes place such that no further conditioning is required when computing the average treatment effect.[16] This restriction would be quite natural in observational studies where outcomes are observed more or less contemporaneously to treatment. However, the underlying assumption behind the exclusion restriction is harder to motivate when dealing with non-experimental data generated from household surveys, where there is usually a time lag between when treatment status is observed and outcomes are measured. The IV approach which we now turn to offers a potential solution to this problem.

6.1. Binary Grouping (Wald) Estimator

In section 2.2, we introduced the most parsimonious approach to estimating treatment – the single difference estimator – where we simply take the mean outcome for the treatment group and subtract from it the mean outcome of the control group. A regression equivalent of this approach is

$$y_{ij} = \alpha + \delta T_{ij} + u_{ij} \tag{1}$$

where T is our treatment dummy; y is our outcome variable; and i, j indexes projects and households respectively.

A simple alternative to this naive approach is the grouping (Wald) estimator (Angrist, 1990). This estimator is a special case of the standard instrumental variable (IV) estimator where we make use of a binary variable (assumed to be exogenous to y) to predict T.

Let this variable be denoted as P_{ij}. Then as long as P_{ij} does not perfectly predict T_{ij}, it can be shown that β_1 is simply equal to the ratio of the diference in means for y (between households with $P = 1$ and $P = 0$) to the difference in means for T (between households with $P = 1$ and $P = 0$). For the most parsimonious case given above where we use a single IV, the IV estimate

[15] The propensity score regression uses proxies for the unobserved/omitted variables under the assumption that the omitted variables are redundant in explaining treatment assignment once their proxies are accounted for. This is obviously not true for omitted variables for which no proxies exist.

[16] In other words, this will be true if Lemma 1 is satisfied.

of the slope can be written as

$$\begin{aligned}\hat{\delta} &= \frac{(\sum_{i=1}^{N}(P_{ij}-\bar{P})(y_{ij}-\bar{y}))}{(\sum_{i=1}^{N}(P_{ij}-\bar{P})(T_{ij}-\bar{T}))} \\ &= \frac{(\sum_{i=1}^{N}P_{ij}(y_{ij}-\bar{y}))}{(\sum_{i=1}^{N}P_{ij}(T_{ij}-\bar{T}))} \\ &= \frac{\bar{y_1}-\bar{y_0}}{\bar{T_1}-\bar{T_0}}\end{aligned}$$

The complete proof is given in Appendix A.1. In the regressions that follow, $P_{ij} = 1$ if household j was part of project i that spent less than a year in the pipeline (quick approval), and 0 if household j was part of project i that spent a year or longer in the pipeline (stuck indefinitely); i.e., P_{ij} is a binary equivalent of $DoseIV$.

Tables 7–8 show the results from estimating the treatment effect using this basic Wald sstimator. The dependent variable is per-capita consumption expenditure. Four specifications are shown. The first specification does not attempt to control for selection bias in any way (i.e., it is a single-difference estimate). The second specification instruments the treatment dummy with the variable $DoseIV$. The third specification uses P_{ij} as the IV. The coefficient on the variable $Treated$ of 579.80 is therefore our estimate of $\hat{\delta}$. Table 7 presents results for the full sample (i.e., all four programs aggregated), whereas table 8 presents estimates only for the LRAD program.

The naive estimates reflect the same result as the simple difference in means test: the treatment effect is not significantly different from zero for the aggregated data, but positive and significant for LRAD.[17]

Specification 3 shows the treatment effect based on the Wald estimator. For both the aggregated and disaggregated data, the effect comes out significantly positive and much larger in magnitude than the estimates that are based on the propensity score.

One reason that might account for this difference is that the Wald estimators exclude other covariates from the regression specification. Specification 4 tests this possibility by looking at whether the IV used in (3) has any explanatory power in explaining consumption, where $Treated$ is not instrumented. We also include one other variable at this stage since it will feature in our subsequent regressions: $D_i = 1$ if a project is located less than 51 kilometers from the nearest DLRO and 0 if it is more remotely located. The results suggest a reconsideration of the Wald estimator since both D and P appear to predict consumption. We next turn to why this might be the case before motivating an alternative approach that extends the idea behind the Wald estimator to account for this possibility.

6.2. An Alternative Wald Estimator

The variable P_{ij} seems to be a plausible IV candidate. If an application spends a long time in the system, this is a signal that some complication has arisen. Therefore a plausible assumption is that the probability of approval declines with the passage of time, until the application reaches a stage of dormancy. We would expect this variable to be non-trivially correlated with the probability of selection into treatment.

[17](To see the comparison with the estimate presented in table 1, then in lines 1 and 2 of table 1, subtract column 4 (control) from column 3 (treatment) to get the coefficient on the variable $Treated$ in column 1 of tables 7 – 8)

Table 7: Average Treatment Effects (Single Difference and Basic Wald – All)

	Naive	DoseIV	P	Binary
	(1)	(2)	(3)	(4)
Treated	-12.700	426.897	579.802	-26.167
	(23.176)	(100.704)***	(156.786)***	(23.409)
D (*Distance to Nearest DLRO* ≤ 50 km)				84.407
				(26.630)***
P (*DoseIV* ≤ 365 days)				150.813
				(34.887)***
Const.	465.968	268.510	148.706	432.602
	(16.959)***	(50.069)***	(84.886)*	(18.563)***
Obs.	3666	2971	3666	3666
R^2	.00008	.	.	.008
F statistic	.3	17.97	13.675	9.247

Dependent variable is real per capita consumption expenditure in 2005 Rands. Standard errors in parentheses. Treated = 1 if a household falls into the treatment group and 0 if it falls into the control group. Column 1 is a single difference estimate of impact. Column 2 instruments $Treated$ with $DoseIV$ (the number of days a household spent in the pipeline). Column 3 reports the results from implementing the grouping estimator, where $Treated$ is instrumented with the variable P_{ij} which is equal to 1 if $DoseIV \leq 360$ and 0 otherwise. Column 4 does not instrument $Treated$.

Table 8: Average Treatment Effects (Single Difference and Basic Wald – LRAD)

	Naive	DoseIV	P	Naive2
	(1)	(2)	(3)	(4)
Treated	75.180	673.024	651.287	47.453
	(37.971)**	(141.112)***	(159.155)***	(39.025)
D (*Distance to Nearest DLRO* ≤ 50 km)				175.661
				(42.243)***
P (*DoseIV* ≤ 365 days)				221.350
				(52.155)***
Const.	472.584	301.871	281.646	408.903
	(21.860)***	(47.417)***	(56.037)***	(24.780)***
Obs.	1925	1766	1925	1925
R^2	.002	.	.	.019
F statistic	3.92	22.748	16.746	12.4

Dependent variable is real per capita consumption expenditure in 2005 Rands. Standard errors in parentheses. Treated = 1 if a household falls into the LRAD treatment group and 0 if it falls into the LRAD control group. Column 1 is a single difference estimate of impact. Column 2 instruments $Treated$ with $DoseIV$ (the number of days a household spent in the pipeline). Column 3 reports the results from implementing the grouping estimator, where $Treated$ is instrumented with the variable P_{ij} which is equal to 1 if $DoseIV \leq 360$ and 0 otherwise. Column 4 does not instrument $Treated$.

Yet P_{ij} might proxy for other unobserved factors that directly influence consumption so it is not immediately obvious that this variable should not also be included in the structural equation. For example, if an application contains a relatively small number of households (say an extended family working as labour-tenants) who happen to be relatively well endowed in terms of productive assets (which they include in the application as part of their own contribution), then such an application is likely to be approved quicker than some other application that is otherwise identical in terms of the fundamentals save for the fact that it involves many more households whose individual own contributions are small (or not monetary in nature).[18] Since their consumption will generally be positively related to asset endowments, and since their asset contributions to the project are not controlled for directly but proxied for by P_{ij}, this proxy might also belong in the structural equation. This effectively disqualifies the use of P_{ij} as a valid IV.

Another plausible candidate for an IV is the variable labelled D_i (used in tables 7–8). This variable might be thought to be a plausible IV candidate because households further away from land reform offices are more likely to be in remote rural locations, and since we know that LRAD targets rural households, D_{ij} is likely to be positively related to treatment status. Indeed, this is borne out to be true in the logit regression reported in table 3. However, here again it might be the case that this variable should be included in the structural model. Indeed if land reform is targeted to poorer households, and if poorer households are also more likely to be found in more remote locations, then D_i can't be assumed to be orthogonal to u.[19]

If D, P, and T, all belong in the structural model, then a more plausible data sampling process might be:

$$y_{ij} = \alpha + \beta D_i + \gamma P_{ij} + \delta T_{ij} + \underbrace{\{\eta(D_i \times P_{ij}) + v_i + \epsilon_{ij}\}}_{\text{composite error}} \tag{2}$$

where $i = 1, \ldots, N$ indexes projects, $j = 1, \ldots, M^i$ indexes the M^i sampled households in project i, and v_i and ϵ_{ij} are project and household-specific error terms respectively. As before, y_{ij} is per capita consumption expenditure, but our treatment variable, T_{ij}, is now generalised to the continuous case. We now interpret this variable in terms of length of exposure or the "dose" of treatment, to use the medical parlance. One such measure is the variable *Doserec* introduced earlier which measures the number of days elapsed from date of transfer to date of interview. To reduce notational clutter, in the results that follow we will denote this variable as T_{ij}.

We've already made the case that D and P are not necessarily good IV candidates because the implied exclusion restrictions are not that plausible under some conditions. However, could the same be said for the interaction between the two variables? Referring to the above data sampling process, if $(D_i \times P_{ij})$ is to be considered a valid IV, we must assume $\eta = 0$, otherwise it could be the case that $cov((D_i \times P_{ij}), u_{ij}) \neq 0$, where $u_{ij} = v_{ij} + \epsilon_{ij}$. On the other hand, if we assume $\eta = 0$, we can then construct a Wald type of estimator using $D_i \times P_{ij}$ as an IV for T_{ij}. We show in appendix A.2 that this IV turns out to resemble a Wald type of estimator that consistently

[18] For households that don't have any assets or cash to put to the prospective project, own contributions are stipulated in terms of (family) labour.

[19] The variable D_i proxies for relative remoteness since most district land reform offices are located in small towns or cities.

estimates the average treatment effect. Formally,

$$\begin{aligned}\tilde{\delta}_{IV} &= \frac{\Delta_{y|D,P}}{\Delta_{T|D,P}} \\ &\xrightarrow{p} \delta + \frac{\eta}{\Delta_{T|D,P}}\end{aligned}$$

where $\Delta_{y|D,P}$ and $\Delta_{T|D,P}$ are defined explicitly in appendix A.2. The exclusion restriction $\eta = 0$ is plausible if it can be assumed that non-remote and remote households are fairly homogenous groups. In other words, if this assumption holds, it won't be the case that time spent in the system differentially affects the consumption of these groups. Arguably this is a weaker exclusion restriction than $cov(y_{ij}, D_i) = 0$ or $cov(y_{ij}, P_{ij}) = 0$, one of which would have to be true if we excluded either variable from the outcome regression.

Table 9: Average Treatment Effects (Alternate Wald)

	All1	LRAD1	All2	LRAD2	All3	LRAD3
	(1)	(2)	(3)	(4)	(5)	(6)
D (*Distance to Nearest DLRO* $\leq$ 50 km)	94.083	185.580	.080	.235	.215	.728
	(29.619)***	(44.311)***	(.037)**	(.050)***	(.077)***	(.414)*
P (*DoseIV* $\leq$ 365 days)	149.722	212.559	.187	.314	-.514	-1.663
	(37.312)***	(53.953)***	(.046)***	(.061)***	(.290)*	(1.555)
T (*Doserec*)	-.022	.039	-.00003	.00005	.001	.004
	(.015)	(.031)	(.00002)	(.00004)	(.0005)**	(.003)
Const.	438.489	416.413	5.658	5.615	5.000	4.319
	(18.363)***	(24.959)***	(.023)***	(.028)***	(.267)***	(1.016)***
Obs.	3125	1823	3124	1823	3124	1823
R^2	.008	.019	.007	.027	.	.
F statistic	8.495	11.709	6.993	17.075	4.739	2.42

Dependent variable is real per capita consumption expenditure in 2005 Rands. Standard errors in parentheses. T is the variable *Doserec*, which measures the number of days a household has been in the treatment group. Columns 1–4 do not instrument T. "All" refers to all treatments aggregated into one. Columns 5–6 instrument T with $D_i \times P_{ij}$.

The results of estimating this model are shown in table 9. Columns 1–4 are alternate specifications of column (4) of tables 7 – 8 where we use our new dose measure of treatment T_{ij} instead of the binary variable $Treated$, but without any instrumenting. The dependent variable is either per-capita consumption (columns 1–2) or its logged equivalent (columns 3-4). The results show that D_i and P_{ij} remain significant in the structural equation and are consistent with our earlier discussion of redundancy in that they are all positively signed. The treatment effect is shown to be insignificant, but not much can be made of this result since we do not control for selection bias at this stage.

Columns 5–6 present the IV regressions where we do control for selection bias by instrumenting T_{ij} with $D_i \times P_{ij}$. Column 5 (all programs) shows that the treatment effect is positive and significant when instrumenting, but the same is not true of the estimates for LRAD (column 6). In order to check weather a less parsimonious specification might make a difference, we included all of the explanatory variables used in the propensity score regression. These results are reported in table 10.

Column 1 shows that the pattern evident in column 4 of table 8 does not change when we add additional covariates and use continuous measures of D and P, which again appear to have

predictive power in explaining consumption. What is different between the two sets of estimates however, is that the treatment effect now becomes significant. To some extent this is unsurprising as the single difference estimate of the impact of LRAD (column 1 of 8) was also positive and significant, and only marginally lower than this less parsimonious specification.

Columns 2, 3, and 4 from table 10 are the most directly comparable sets of estimates to the propensity score estimates reported earlier. In both cases, our treatment indicator is binary, and we instrument with $DoseIV$. The average treatment effect is of the same sign as the various estimates presented in table 6 and about double the magnitude of those estimates. However, notice that the standard errors of the IV estimates are large enough to suggest that the confidence intervals around the 2SLS estimates will overlap with the confidence intervals of the estimates based on the propensity score (table 6).

Table 10: Average Treatment Effects (2SLS)

	1	2	3	4	5	6
Dependent Variable	PCE	PCE	PCE	PCE	LnPCE	LnPCE
Treatment	Binary	Binary	Binary	Binary	Dose	Dose
Instrumental Variable	None	*DoseIV*	*DoseIV*	*DoseIV*	$(D_i \times P_{ij})$	$(D_i \times P_{ij})$
Distance to DLRO (km)	-.698	-.749				
	(.31)**	(.314)**				
DoseIV	-.03					
	(.018)*					
Treated	85.764	370.036	386.936			
	(47.458)*	(166.103)**	(166.455)**			
D (*Distance to Nearest DLRO* ≤ 50 km)			172.769	304.197	.005	.263
			(47.937)***	(145.506)**	(.05)	(.072)***
P (*DoseIV* ≤ 365 days)				-601.085	-.05	.098
				(621.066)	(.082)	(.097)
T (*Doserec*)				1.621	.0002	.0002
				(1.347)	(.0001)*	(.0001)*
Household Agricultural Employment	13.808	-11.329	-8.158	-79.176	-.054	-.076
	(20.065)	(23.7)	(23.502)	(85.447)	(.023)**	(.031)**
Male-Headed Household	66.07	46.679	45.684	-30.429	.047	.013
	(43.345)	(46.186)	(46.101)	(113.313)	(.046)	(.063)
Education of household head (yrs)	49.676	51.735	50.483	61.644	.085	.083
	(4.562)***	(4.57)***	(4.549)***	(13.329)***	(.006)***	(.007)***
Mean farming experience (yrs)	9.731	10.359	10.859	6.873	.024	.028
	(5.426)*	(5.48)*	(5.466)**	(8.586)	(.006)***	(.008)***
Number Plots Accessed Pre-1995	19.739	32.435	34.891	120.304	.047	.069
	(15.12)	(18.827)*	(18.886)*	(94.968)	(.041)	(.067)
Size of Plots Accessed Pre-1995 (Hectares)	.017	.016	.02	.06	.0002	.0002
	(.044)	(.045)	(.045)	(.076)	(.0001)*	(.0001)
Const.	198.581	81.128	-32.29	-465.542	4.979	5.019
	(64.619)***	(70.628)	(69.313)	(448.852)	(.099)***	(.086)***
N	1651	1651	1658	1709	1706	943

Dependent variable is real per capita consumption expenditure in 2005 Rands (columns 1–4), or its logged equivalent (columns 5–6) . Standard errors in parentheses. T is the variable *Doserec*, which measures the number of days a household has been in the treatment group. $Treated$ is the usual binary treatment for the LRAD program. Column 1 does not instrument $Treated$. "All" refers to all treatments aggregated into one. Columns 5–6 instrument T with $D_i \times P_{ij}$.

Finally, columns 5 and 6 implement an extension of the models used in columns 5 and 6 of table 9, the difference being the inclusion of additional covariates. Because of the potential for scale effects that could be introduced by the change in the treatment measure from a binary variable to a continuous variable, we now log the left hand side of this regression. The treatment effect, given by the estimated coefficient for T is now to be interpreted as a percentage change in log per capita consumption expenditure for a single day increase in exposure to the program. If we multiple this estimate by the mean exposure to the treatment for the LRAD program (roughly 350 days), we get an average treatment effect of about 7%, which is to say that after about a year of becoming an LRAD beneficiary (i.e., post-transfer), monthly per-capita consumption increases by 7%.

7. Conclusion

Several lessons emerge from this analysis. First, there is a clear need to control for selection bias, even though we have a quasi experimental design. Second, the impact of the current program of redistribution on household per capita consumption is positive, and remains positive and significant even once we have controlled for selection bias. Thus the direction of the effect for the LRAD program appears robust to changing the underlying statistical assumptions of our models.

In terms of the overall magnitude of the impact of LRAD, this is less clear cut. Our estimates of the average treatment effect on beneficiaries tend to vary according to the methods we employ. Having said that, it is clear that sampling error is quite large for the IV estimates. When compared against the results emanating out of our various propensity score matches, it is clear that the larger sampling error of the IV estimates would imply little difference in magnitudes between the two sets of estimates (i.e., the implied confidence intervals for the estimates shown in table 6 seem to overlap with those implied by columns 2 – 4 of table 10, judging from the standard errors associated with the treatment effects reported in those two tables). We therefore limit our concluding comments to the more conservative set of estimates based on the propensity score (table 6).

It is hard to quantify *exactly* what this means in terms of poverty reduction because there is some controversy over which is the correct poverty line to employ. Woolard and Leibbrandt (2007) provide an interesting sensitivity analysis of the national head count index of poverty based on a menu of 10 different poverty lines currently in use by various governmental agencies and departments responsible for tracking poverty in South Africa. Arguably the two most widely used lines are those used by Stats SA and the Household Effective Level (HEL) line. Woolard and Leibbrandt (2007) report that the lower bound of the Stats SA line stands at R416.99 (November 2005 prices, per capita) whereas the HEL line on the other hand is R555.55 (November 2005 prices, per capita). A conservative approach is to take the average of these two lines. This gives us a notional poverty threshold of R486.27 per capita. As table 1 indicates, average per capita consumption expenditure for households in the LRAD control group stood at R472.61. Thus, on average, households in that control group can be said to be poor. Now referring back to table 6 we see that the lowest estimate of treatment (based on the second logit specification) is R134.24. This means that mean per capita consumption expenditure in the treatment group is R606.85 once we've controlled for selection bias. These admittedly conservative calculations suggest that in the short term, the impact of land redistribution is significant enough to bump the average participating household out of poverty, even by the standards of the less generous HEL line.

Whether or not these calculations translate into a significant reduction in rural poverty remains an open question, in part because little is known about how well the the range of poverty lines currently in use translate to rural households – a point argued quite convincingly by Woolard and Leibbrant (2007). Moreover, even if these estimates are an accurate reflection of short term impact, nothing can be said about whether these effects would be sustained, muted, or reversed over time. Notwithstanding these caveats around interpretation however, what is quite clear is that the LRAD program does seem to benefit the consumption of its beneficiaries in the short term.

References

[1] Angrist, J. D. (1990) "Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records, *American Economic Review*, 80, 313-335.

[2] Angrist, J. D., and A. B. Krueger (1999) "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, O. Ashenfelter and D. Card (eds), Elsevier: Amsterdam, North Holland.

[3] Bardhan, P. Bowles, S., and H.Gintis (2001) "Wealth Inequalities, Wealth Constraints, and Economic Performance," in *Handbook of Income Distribution: Volume 1*, A. Atkinson and F. Bourguignon (eds), Elsevier: Amsterdam, North Holland.

[4] Becker and Ichino (2002) "Estimation of Average Treatment Effects Based on Propensity Scores," *The Stata Journal*, 2, pp.358–77.

[5] Berry, R.A., and W.R. Cline (1979) *Agrarian Structure and Productivity in Developing Countries*, Baltimore: Johns Hopkins University Press.

[9] Banerjee, A., P. Gertler and M. Ghatak (2002) "Empowerment and Efficiency: Tenancy Reform in West Bengal," *Journal of Political Economy*, 110: 2, pp 239-280.

[7] Besley, T., and R. Burgess (2000): "Land Reform, Poverty Reduction, and Growth: Evidence from India," *Quarterly Journal of Economics*, 115: 2, pp.389-430.

[8] Binswanger, H. P., K. Deininger, and G. Feder (1995): "Power, Distortions, Revolt and Reform in Agricultural and Land Relations," in *Handbook of Development Economics: Volume 3B*, Elsevier: Amsterdam, North-Holland

[9] Cochran, W.G. (1968) "The effectiveness of adjustment by subclassification in removing bias in observational studies, " *Biometrics*, 24, 205-13.

[10] Chase, R. (2002) "Supporting Communities in Transition: The Impact of the Armenian Social Investment Fund," *World Bank Economic Review,* 16: 2, 219-240.

[11] Dehejia, R.H. and S. Wahba (2002) "Propensity Score Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*, 84, 151-161.

[12] Dehejia, R.H. and S. Wahba (1999) "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94: 448, 1053-1062.

[13] Dutta, B., D. Ray, and K. Sengupta (1989): "Contracts with Eviction in Infinitely Repeated Principal-Agent Relationships," in *The Economic Theory of Agrarian Institutions*, P. Bardhan (ed.). Clarendon Press: Oxford.

[14] Galasso, E., and M. Ravallion (2004) "Social Protection in a Crisis: Argentinas Plan Jefes y Jefas," *World Bank Economic Review,*" 18: 3, 367-399.

[15] Hahn, J. (1998) "On the role of the propensity score in efficient semiparamentric estimation of average treatment effects," *Econometrica*, 66: 2, 315-331.

[16] Heckman, J. J., H. Ichimura, and P. Todd (1997) "Matching as an econometric evaluation estimator," *Review of Economic Studies*, 65, 261-294

[17] Heckman, J. J., H. Ichimura, and P. Todd (1998) "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66: 5, 261-294

[18] Heckman, J. J., Lalonde, R. J., and J. A., Smith (1999) "The Economics and Econometrics of Active Labor Market Programs in *Handbook of Labor Economics: Volumn 3A*, O. Ashenfelter and D. Card (eds), Elsevier: Amsterdam, North Holland.

[19] Heckman, J. J., and R. Robb (1985) "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, J. Heckman, and B. Singer (eds), Cambridge: Cambridge University Press.

[25] Rosenbaum, P. R. (2004) "Matching in Observational Studies," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Date Perspectives*, A. Gelman and Xiao-Li Meng (eds), John Wiley and Sons: Sussex.

[21] Legros, P. and A.F. Newman (1997) "Wealth Effects, Distribution and the Theory of Organisation", *Journal of Economic Theory*.

[22] Moene, K.O. (1992) "Poverty and Landownership," *American Economic Review*, 81, pp.52-64..

[23] Mookherjee. (1997) "Information Rents and Property Rights in Land," in John Roemer, ed., *Property Rights, Incentives and Welfare*, MacMillan, London.
[24] Ravallion, M (2007) "Evaluating Anti-Poverty Programs," in *Handbook of Development Economics: Volume 4*, R. E. Evenson and T. Paul Schultz (eds), Elsevier: Amsterdam, North-Holland.
[25] Rosenbaum, P.R. and D.B. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70: 1, 4155.
[26] Shaban, R. (1987) "Testing Between Competing Models of Sharecropping," *Journal of Political Economy*, 95, pp 893-920.
[27] Shetty, S. (1988) "Limited Liability, Wealth Differences, and the Tenancy Ladder in Agrarian Economies," *Journal of Development Economics*, 29, 1–22.
[28] Woolard, I. and M. Leibbrandt "The Measurement of Poverty in South Africa: Some Technical Issues," Paper presented at SANPAD conference, 26- 29 June 2007, Durban.

A. Derivation of the Wald Estimator

The proof is as follows: the numerator can be written as $\sum_{i=1}^{N} P_i(y_i - \bar{y}) = \sum_{i=1}^{N} P_i y_i - (\sum_{i=1}^{N} P_i)\bar{y} = N_1\bar{y_1} - N_1\bar{y} = N_1(\bar{y_1} - \bar{y})$ where $N_1 = \sum_{i=1}^{N} P_i$ is the number of observations in the sample with $P_i = 1$ and $\bar{y}_i$ is the average of the y_i over the observations with $P_i = 1$. Next write $\bar{y}$ as a weighted average: $\bar{y} = \frac{N_0}{N}\bar{y_0} + \frac{N_1}{N}\bar{y_1}$, where the zero/one subscripting refers to treatment and control. After some algebra it can be shown that $\bar{y_1} - \bar{y} = (\frac{N-N_1}{N})\bar{y_1} - (\frac{N_0}{N})\bar{y_0} = (\frac{N_0}{N})(\bar{y_1} - \bar{y_0})$. So the numerator of the IV estimate is $(\frac{N_0 N_1}{N})(\bar{y_1} - \bar{y_0})$. The same argument shows that the denominator is $(\frac{N_0 N_1}{N})(\bar{T_1} - \bar{T_0})$. Taking the ratio completes the proof.

B. Derivation of the Probability Limit of the Wald Estimator Using $D \times P$ as an IV

We begin by computing the following conditional expectations:

$$\begin{aligned}
E(y_{ij}|D_i = 1, P_{ij} = 1) &= \alpha + \beta + \gamma + \delta E(T_{ij}|D_i = 1, P_{ij} = 1) \\
&+ \eta + E(v_i|D_i = 1) \\
E(y_{ij}|D_i = 1, P_{ij} = 0) &= \alpha + \beta + \delta E(T_{ij}|D_i = 1, P_{ij} = 0) + E(v_i|D_i = 1) \\
E(y_{ij}|D_i = 0, P_{ij} = 1) &= \alpha + \gamma + \delta E(T_{ij}|D_i = 0, P_{ij} = 1) + E(v_i|D_i = 0) \\
E(y_{ij}|D_i = 0, P_{ij} = 0) &= \alpha + \delta E(T_{ij}|D_i = 0, P_{ij} = 0) + E(v_i|D_i = 0)
\end{aligned}$$

We will also need to compute:

$$\begin{aligned}
&E(T_{ij}|D_i = 1|P_{ij} = 1) \\
&E(T_{ij}|D_i = 1|P_{ij} = 0) \\
&E(T_{ij}|D_i = 0|P_{ij} = 1) \\
&E(T_{ij}|D_i = 0|P_{ij} = 0)
\end{aligned}$$

We can now construct difference-in-difference estimators for the effect of D and P on consumption, as well as on the dose variable:

$$\begin{aligned}
\hat{\Delta}_{y|D,P} &= [E(y_{ij}|D_i = 1, P_{ij} = 1) - E(y_{ij}|D_i = 1, P_{ij} = 0)] \\
&- [E(y_{ij}|D_i = 0, P_{ij} = 1) - E(y_{ij}|D_i = 0, P_{ij} = 0)]
\end{aligned}$$

$$\begin{aligned}
\hat{\Delta}_{T|D,P} &= [E(T_{ij}|D_i = 1, P_{ij} = 1) - E(T_{ij}|D_i = 1, P_{ij} = 0)] \\
&- [E(T_{ij}|D_i = 0, P_{ij} = 1) - E(T_{ij}|D_i = 0, P_{ij} = 0)]
\end{aligned}$$

Taking the ratio of these two estimators produces a Wald estimator with probability limit,

$$\begin{aligned}
\tilde{\delta}_{IV} &= \frac{\Delta_{y|D,P}}{\Delta_{T|D,P}} \\
&\xrightarrow{p} \delta + \frac{\eta}{\Delta_{T|D,P}}
\end{aligned}$$