ESTIMATING TREATMENT EFFECTS FROM SPATIAL POLICY EXPERIMENTS:
An Application to Ugandan Microfinance*

Craig McIntosh

September 9, 2004

Abstract

This paper demonstrates a method for estimating treatment effects in spatial tests, utilizing a second control group to measure unexplained spatial phenomena. The technique is implemented on two innovations in Ugandan microfinance, and we measure the ways in which concurrent shocks such as an ebola outbreak and a contentious presidential election altered outcomes differentially across regions. By correcting for this spatial heterogeneity we measure the impact of the policies; a program which increased borrowers' control over the terms of their loans improved outcomes, while the results of a program which bundled health insurance into the lending contract were more mixed.

*Thanks to Fulbright IIE for research funding, and to Alain de Janvry, Gordon Hanson, Guido Imbens, Ethan Ligon, Elisabeth Sadoulet, Michael Ward, and seminar participants at U.C. Berkeley, U.C. San Diego, University of Maryland, RAND, Dalhousie University, WEAI, and NEUDC for helpful comments.

I Introduction

Spatial tests, while subject to biases which random tests are not (Kremer 2003), are likely to remain a permanent feature of the policy landscape. States, school districts, administrative branches and the rule of law all describe physical spaces, and so designating physical regions as treatment and controls will continue to be the least intrusive way for many kinds of institutions to experiment (Card & Krueger 1994). The problem with this methodology is that it takes unexplained changes in the control region as a counterfactual for what would have transpired in the absence of the test, and so will be biased by unexplained spatial shifts. Many such tests also contain an eligibility requirement, whether as a result of a voluntary program, or some screening criteria. We show here that if the ineligible agents in the control can be clearly identified, we have a population that allows us to estimate and remove exactly the kinds of unexplained spatial heterogeneity that would otherwise bias impact estimates. The result is a fine-grained form of triple-differencing which accounts for localized shocks or mis-specification in the estimating equation. This not only makes spatial impact analysis robust to unexplained shocks and endogenous placement, but allows us to measure a kind of impact invisible to standard techniques: the ability of a treatment to insulate agents against shocks.

In recent years, an increasing number of researchers have been attempting to overcome the limitations of difference-in-difference techniques by some form of triple differencing. Examples include the comparison of pre-treatment *growth* to post-treatment *growth* (McKenzie & Mookherjee 2003), or comparison of current and past members of a training program to a matched sample of non-participants (Ravallion, Galasso, Lazo & Philipp 2002). Closer to the application presented here are efforts to compare eligible and ineligible agents in regions which offer and do not offer a program (Rosenbaum 1982), (Gruber 1994), (Morduch 1998), (Pitt & Khandker 1998). Much of this work, however, has used the second control group simply to calculate lump-sum spatial effects, which will not remove the bias if spatial distributions differ. This paper implements techniques designed for propensity-score matching across the selection criterion (Rosenbaum & Rubin 1983), (Angrist 1995), (Dehejia & Wahba 2002) in order to match across the simple and observable criterion used to select agents into the treatment in a spatial test; namely physical location. If physical space defines the treatment and control region, we need only remove unexplained heterogeneity within this dimension in order to recover an unbiased impact estimate.

We apply the technique to two new policies introduced by FINCA Uganda, the country's largest micro-finance institution. The first innovation allowed clients to change the repayment frequency of their loans, and the second bundles a health-insurance package into the lending contract. The insurance program was sufficiently innovative to be discussed in the Economist of January 11, 2001, and we present here the first quantitative analysis of its impact. A straightforward spatial testing strategy was implemented, wherein whole administrative branches of the institution were designated as treatment and control regions. A standard analysis of a voluntary, spatial test would either compare treated and untreated regions to estimate the intention-to-treat effect, or identify those likely to choose the program in the control in order to estimate the treatment effect on the treated. Here, we dealt with the selection problem by conducting mock elections in the control to establish the groups which would have chosen the programs had they been offered them. With this extra degree of identification we can use differences between choosers and non-choosers as well as the differences between those offered and not offered the program to identify the impact of the two treatments.

The results of the analysis have several implications for best practices in microfinance. Weekly repayment is widely perceived to be central to the low default rates observed in microfinance (Morduch 1999), yet this system entails enormous transaction costs for both lender and borrower. We find that moving from repaying loans every week to repaying every two weeks causes none of the predicted negative effects, instead causing dropout to fall by 10 percentage points (a 40% reduction) and triggering a slight improvement in repayment performance. The implication of this study is that *when clients are allowed to decide* whether their fellow members are capable of repaying reliably biweekly, the joint-liability mechanism allows transaction costs to be reduced without pushing up default. The entire impact of the health insurance program arose due to insulation against shocks, rather than from directly observable changes in outcomes. Both savings and loans are weakly sensitive to the removal of health expenditure risk, and thus may be precautionary in nature. New client enrollment increased sharply, but client composition shifts for the worse in insured groups indicating that the asymmetric information problems in credit and insurance markets need to be solved separately rather than using a single joint-liability contract (Stiglitz 1990), (Ghatak 1999) to solve them both.

II Bias in the Difference-in-Differences

The joint-liability group is the unit of observation. The change in outcomes is denoted by Y_i, which is determined by some function $Y_i = f(X_i, s_i, Z_i) + \epsilon_i$. We use simple changes in outcomes rather than a panel because most of our control data is cross-sectional, as well as because of the concerns raised in Bertrand, Duflo & Mullainathan (2004) over the inconsistency of estimates of standard errors in the DID when outcomes are serially correlated. X_i is a vector of observable control variables, s_i is group i's location in physical space, and Z_i is a vector of unobservables. Treatment status is a binary variable indicating whether a group was offered the treatment, and is denoted by T_i. Our application involves a spatial treatment-control strategy, which implies a mapping from location s_i to the treatment status; we denote this mapping by τ, so $\tau(s_i) = T_i \in \{0, 1\}$. In addition, the treatment is voluntary, and so groups may either choose or not choose to accept the treatment, a decision denoted by ω_i (in a different context, ω_i might represent eligibility). Receipt of the treatment, then, indicates that $T_i = \omega_i = 1$.

In potential outcomes notation, Y_{1i} represents changes in outcomes for a treated group, and Y_{0i} represents the counterfactual untreated outcome for the same group. Without loss of generality, we can think of the treatment effect as additive, so that $Y_{1i} = Y_{0i} + t(X_i, s_i, Z_i)$. Since both choices and location determine treatment status in this kind of test, we have a two-tiered selection problem. The rule by which agents select into $\omega_i = 0$ and $\omega_i = 1$ is called the selection criterion, and $T_i = 0$ and $T_i = 1$ the treatment criterion; both rules are, in general, endogenous. Estimating impact in such two-tiered tests requires us to account for variation in the determinants of Y_i across both criteria.

Under perfect randomization, the incidence of the treatment is orthogonal to the determinants of Y_i, and so the control population forms a perfect counterfactual for the treated population. Thus,

$$E(Y_{0i} \mid T_i = 1) = E(Y_i \mid T_i = 0),$$

and so we can easily recover the 'intention to treat' effect (ITE) because

$$E(Y \mid T=1) - E(Y \mid T=0) = E(t) + E(Y_0 \mid T=1) - E(Y \mid T=0) = E(t).$$

If all of the agents offered the treatment accept it, then this intention-to-treat effect is the same as the average treatment effect on the treated (ATET). If self-selection of agents into the treatment is an inherent feature of the program (as is the case in both of these innovations) then the average treatment effect (ATE) is not of general policy interest because the average treated individual will not correspond to the average individual in the population.

Ignorability:

In situations where participation is not total, the usual approach is to assume ignorability, or 'selection on observables' (see Heckman, Ichimura, Smith & Todd (1998) for a thorough discussion of selection issues). If we use as controls groups which were offered the program but did not choose it, then the assumption of conditional mean ignorability can be written as:

$$E(Y_{0i} \mid X_i, \omega_i) = E(Y_{0i} \mid X_i), \ \textit{and}\ E(Y_{1i} \mid X_i, \omega_i) = E(Y_{1i} \mid X_i)$$

If we use as controls a second population which was denied access to the treatment, ignorability requires us to make assumptions about the role of unobservables across both the selection and the treatment criteria:

$$\textit{Selection Assumption}: \ E(\omega \mid X, T=1) = E(\omega \mid X, T=0)$$

(Ignorability of treatment on selection decisions)

$$Outcome\ Assumption:\ E(Y_0 \mid X, \omega, T=1) = E(Y \mid X, \omega, T=0).$$

(Ignorability of treatment and selection on outcomes)

These two assumptions allow us to identify impact because they imply that $E(Y \mid X, T=0)$ is an appropriate counterfactual for $E(Y_0 \mid X)$.

In a linear difference-in-differences (DID) regression, the ITE is identified by:

$$Y_i = \beta X_i + \delta_1 T_i + \mu_i,$$

and the ATET by:

$$Y_i = \beta X_i + \delta_2 T_i + \mu_i,\ \forall\ \omega_i = 1.$$

Failure of Ignorability

We investigate the bias due to spatial heterogeneity by explicitly denoting the portion of outcomes which is systematic but not explained by the linear fit on observables as $\phi(X_i, s_i, Z_i)$. Therefore the equations determining outcomes are:

$$Y_{0i} = \beta X_i + \phi(X_i, s_i, Z_i) + \epsilon_i,$$

and

$$Y_{1i} = \beta X_i + \phi(X_i, s_i, Z_i) + t(X_i, s_i, Z_i) + \epsilon_i,$$

Proposition 1: The Difference in Differences estimator will be biased in a spatial test unless unexplained spatial effects are identical in the treatment and control regions.

Proof of Proposition 1:

When we estimate the DID $Y_i = \beta X_i + \delta_1 T_i + \mu_i$ on outcomes generated by the process above, $\phi(X_i, s_i, Z_i)$ is an omitted variable which is orthogonal to X by definition. Consequently, it will project into T but not into X, and we estimate impact to be $\hat{\delta_1} = E(t) + P_T(\phi)$. Since $T_i = \tau(s_i)$ is a binary variable defined over space, the projection of ϕ into T will equal $E(\phi \mid \tau(s) = 1) - E(\phi \mid \tau(s) = 0)$. This is the difference in unexplained spatial effects between treatment and control regions. Thus $\hat{\delta_1} = E(t) + E(\phi \mid \tau(s) = 1) - E(\phi \mid \tau(s) = 0)$, and the DID will only unbiased if this difference is zero, so

$$\hat{\delta_1} = E(t) \Leftrightarrow E(\phi \mid \tau(s) = 1) = E(\phi \mid \tau(s) = 0). \diamond$$

This simply rewrites the outcomes assumption for spatial tests. If estimating the ITE (using the whole population) then $E(t) = E(t \mid T_i = 1)$, and for the ATET (using only choosers) then $E(t) = E(t \mid T_i = 1, w_i = 1)$.

III Overcoming Spatial Bias

This paper shows how failure of the Outcome assumption can be overcome through a modification of the Selection assumption. Consequently, this technique is most appropriate for circumstances in which the selection process is either straightforward or directly observed in the control, and where we suspect the presence of spatial shocks during the test.

The intuition of the idea is that even if the treatment and control regions are not directly comparable, the differences in outcomes between choosers and non-choosers of the program should provide a measure of the treatment effects. This involves altering the assumption on the *selection* criterion in order to be able to investigate spatial effects across the *treatment* criterion. In this way, matching techniques designed to control for selectivity bias can be implemented instead to control for bias arising from a non-randomly assigned treatment.

The key to identification in this approach lies in our ability to clearly specify the rule which selects agents into the treatment and control region. If this effect is highly complex (such as in a failed randomization) then we cannot proceed. In a case, however, where simple geographic regions are used as treatment-control regimes, the dimension across which we must be concerned with bias is very clear.

If we consider the non-choosers of a program to represent a counterfactual for the spatial variation in outcomes that would have been present in the absence of a treatment, then we are provided with a spatial surface which forms a natural baseline from which to estimate the true effects of the treatment. A simple way to proceed would be to subtract off locally-averaged outcomes among non-choosers from the outcomes of each chooser. Upon further reflection, however, we see that it is only unexplained spatial effects which will bias estimates, and so we should not include in this local outcome estimate anything which is directly explained by observables. This suggests that it is the residuals among non-choosers in a local area that possess the most information about unexplained spatial effects. The reason for the concern with shocks *which vary across space* is that it is only unexplained effects which have some spatial component that will project into the treatment and cause bias. The direct implication is that if we are able to estimate residual spatial effects, we can recover an unbiased impact term.

We now introduce two assumptions which allow us to estimate and utilize these spatial effects. The first is essentially an extension of the Selection assumption:

$$\textit{Spatial Assumption}: \; E(\phi \mid s_i, \omega_i = 1) = E(\phi \mid s_i, \omega_i = 0).$$

Because we include a constant term in the X-vector, $E(\phi \mid \omega_i = 1) = E(\phi \mid \omega_i = 0) \equiv 0$. So the Spatial Assumption says that a chooser and a non-chooser located at the same place should experience unexplained spatial effects, relative to their own group mean, that are the same.

The second required assumption is that there be no spillover effects of the treatment from choosers to non-choosers within this same region.

$$\textit{No Spillovers Assumption}: \; E(t_i \mid \omega_i = 0) = 0$$

Homogeneous Spatial Distributions:

If there is a perfect one-to-one spatial mapping of choosers and non-choosers, then these assumptions imply that the spatial effects we estimate in the non-choosers are precisely those we expect to see among the choosers. This suggests two straightforward ways of using our assumptions. The first is the **False Difference-in-Differences (FDID) regression** among non-choosers;

$$Y_i = \beta_1 X_i + \gamma_1 T_i + \mu_i \;\forall\; \omega_i = 0.$$

$\hat{\gamma}_1$ will equal $E(P_T(\phi) \mid \omega = 0)$, which is a spurious treatment effect arising from unexplained spatial phenomena. Since the spatial dispersion of choosers and non-choosers is identical, the spatial assumption implies that

$$E(P_T(\phi) \mid \omega = 0) = E(P_T(\phi) \mid \omega = 1),$$

and so $\hat{\gamma}_1$ is a precise measure of the spatial bias among choosers. In this case, we can recover an unbiased estimate of the ATET by subtracting the coefficient from the FDID off of the DID, which is analogous to conducting a **Triple Difference regression**:

$$Y_i = \omega_i \beta_2 X_i + (1 - \omega_i)\beta_1 X_i + \alpha_1 T_i + \alpha_2 \omega_i + \delta_3 (T_i * \omega_i) + \mu_i.$$

In other words, $\hat{\delta_3} = \hat{\delta_2} - \hat{\gamma}_1$.

In the absence of spatial effects, the estimate from the triple-difference will be the same as from a difference in differences, but less efficient. Since the FDID will also be insignificant in such cases, this suggests that where this second control group exists we should use the FDID as a preliminary test for the presence of spatial effects which will bias standard methods. Only in the presence of a significant false impact in the FDID should we proceed to utilize methods designed to remove spatial bias.

Heterogeneous Spatial Distributions:

In general the spatial distribution of choosers and non-choosers is not identical, which means that even under the spatial assumption, the FDID coefficient is not the proper measure of unobserved spatial effects. If, however, we have precise location information on agents, we have what we need to do recover an unbiased treatment measure by estimating spatial effects at every location rather than lump-sum effects in the treatment and control regions as a whole. We do this by estimating the residual surface among non-choosers:

$$\hat{\mu}_i = Y_i - \hat{\beta} X_i \ \forall \ \omega_i = 0.$$

Proposition 2: Spatial expectations of the residual estimated among the non-choosers can be used as a counterfactual for unexplained spatial effects among the choosers, or $E(\hat{\mu}_0 \mid s_i) = E(\hat{\mu} \mid s_i, \omega_i = 0)$.

Proof of Proposition 2: In the absence of the treatment, $\hat{\mu}_i = Y_i - \hat{\beta} X_i = \phi_i + \epsilon_i$. ϵ is an i.i.d. error term, and the assumption that there are no spillovers means that the untreated receive no treatment effect, so the spatial expectations of residuals contain only spatial effects: $E(\mu \mid s_i, \omega_i = 0) = E(\phi \mid s_i, \omega_i = 0)$. For an agent who chooses the program, the counterfactual expectation of unexplained spatial effects in the absence of a treatment effect is $E(\mu_0 \mid s_i, \omega_i = 1) = E(\phi \mid s_i, \omega_i = 1)$. The spatial assumption says that $E(\phi \mid s_i, \omega_i = 1) = E(\phi \mid s_i, \omega_i = 0)$, so

$$E(\hat{\mu}_0 \mid s_i) = E(\hat{\mu} \mid s_i, \omega_i = 0).\diamond$$

Proposition 3: $Y_i - E(\hat{\mu} \mid s_i, \omega_i = 0)$ is an outcome which allows for unbiased estimation of the ATET even when $P_T(\phi) \neq 0$.

Proof of Proposition 3:

From Proposition 2, subtracting this spatial residual off of outcomes will give us

$$Y_{0i} - E(\hat{\mu} \mid s_i, \omega_i = 0) = \beta X_i + \phi_i - E(\phi \mid s_i) + \epsilon_i.$$

The term $\phi_i - E(\phi \mid s_i)$, represents the unobserved effect minus the unexplained effect conditional upon being in that location. Crucially, the resulting term is itself orthogonal to space, and so can be written as $M_s(\phi_i)$: this is the residual vector that remains when ϕ has been projected into s, or the component of the outcome vector that projects off of both X and s. This modified dependent variable can be written as:

$$Y_{0i} - E(\phi \mid s_i) = \beta X_i + M_s(\phi_i) + \epsilon_i.$$

When this term is explained with a linear vector of controls and a spatial treatment dummy, we have no residual unexplained spatial effects which could bias the treatment term. This is true because our unexplained residual is orthogonal to the Xs by definition, and is orthogonal to the treatment by construction, since:

$$M_s(\phi) \perp [\tau(s), \beta X]$$

This means that $Y_i - E(\phi \mid s_i)$ satisfies the Outcomes assumption by construction; that is:

$$E((Y_{0i} - E(\phi \mid s_i)) \mid X_i, \tau(s_i) = 1) - E((Y_i - E(\phi \mid s_i)) \mid X_i, \tau(s_i) = 0) \equiv 0.$$

Thus, from the proof of Proposition 1, when use our modified dependant variable

$$Y_i - E(\hat{\mu} \mid s_i, \omega_i = 0) = \beta X_i + M_s(\phi_i) + t(X_i, s_i, Z_i) + \epsilon_i$$

to run a DID regression $Y_i - E(\hat{\mu} \mid s_i, \omega_i = 0) = \beta X_i + \delta_4 T_i + \mu_i \ \forall \ \omega_i = 1$, $E(\hat{\delta_4}) = E(t) + P_T(M_s(\phi_i)) = E(t)$ because $P_T(M_s(\phi_i)) \equiv 0$, and so we have an unbiased estimate of the ATET. $\diamond$

One feature of this approach which is very attractive is that it allows for consistent estimation of treatment effects even under endogenous placement. If a program is (intentionally or unintentionally) placed in a region which has inexplicably different rates of growth from the control, this is usually fatal to our ability to estimate impact. The spatial assumption, however, will apply as long as it is the case that the non-choosers of the program share the outcome differences seen among nearby choosers. In other words, if a certain ethnic group, region, or administrative unit is non-randomly assigned a treatment, they only bias spatial impact assessments because they introduce spatial heterogeneity. To the extent, then, that this spatial effect is *not* related to selection into the program in a way that varies across space, the effects of the endogenous placement will be removed by the use of spatial matching to the second control.

IV The Spatial Matching Estimator

To operationalize these propositions, we match each agent who chose the treatment to the residual of the nearest agent who did not choose the treatment (denoted by i'), and subtract this off of the dependent variable.

So, the **Spatial Matching estimator** estimates the ATET as follows:

$$Y_i - \hat{\mu_{i'}} = \beta X_i + \delta_4 T_i + \mu_i \; \forall \; \omega_i = 1.$$

Although there is no way to directly estimate the ITE when spatial effects are present, we can back out an estimate from the ATET. The reason is that we have already assumed no spillover effects, which means that $E(t_i \mid \omega_i = 0) = 0$, and so the intention-to-treat effect can be estimated by:

$$ITE = \hat{\delta}_4 * Pr(\omega = 1),$$

where $Pr(\omega = 1)$ is the fraction of choosers.

A technical problem arises in the use of OLS residuals due to the potential clustering of the Xs in space, which will cause observables to proxy for underlying spatial effects. If we do nothing to address this issue, then the spatial assumption in reality requires a similar spatial distribution of the Xs between choosers and non-choosers, which is an unpalatable assumption. To address the issue, we use an orthogonalizing procedure known as Gauss-Seidel Regression (Telser 1964) to 'backfit' the residual vector among non-choosers. This procedure uses a spatial smoother to extract spatial information present in the residuals, subtract it off of the dependant variable, and iterates until there is no spatial component remaining. The resultant βs are used to predict a vector of residuals which contain the full degree of spatial variation. This procedure is deemed to make the spatial assumption more realistic, and so it is implemented throughout the paper.

Having identified spatial matching as the appropriate way to deal with spatial effects when two control groups are available, the matching itself is mechanically standard. Thus, we draw on the extensive literature developed to solve matching problems across the selection criterion. We utilize a **bias-correction** technique suggested by Abadie & Imbens (2004) to estimate

$$\hat{\mu}^N = \lambda s^N + \epsilon$$

(in this case, s is longitude and latitude). The estimated shock at each location is modified using λ as follows; letting i index the group and i' be the nearest neighbor to i,

$$\hat{\phi} = \hat{\mu}^N + \hat{\lambda}(s_i - s_{i'}).$$

This corrects the estimated shock estimated for i by the distance and direction to the nearest neighbor, given any overall tilt that may exist in the spatial distribution of shocks.

The second technique taken from that paper is to establish a **double counterfactual**; not only do we estimate what all of the choosers would have looked like in the absence of the treatment, but we also estimate what all of the non-choosers would have looked like had they received the treatment. Because we backfit residuals for both choosers and non-choosers, we can estimate a residual vector among choosers which contains the treatment term, and so the regression used to estimate impact can be written in block matrix form as follows:

$$\left| \begin{array}{c} Y^C - \hat{\mu}_{i'}^{NC} \\ \\ Y^{NC} - \hat{\mu}_{i'}^{C} \end{array} \right| = \left| \begin{array}{ccc} X^C & 0 & \text{T} \\ \\ 0 & X^{NC} & \text{-T} \end{array} \right| \left| \begin{array}{c} \beta^C \\ \beta^{NC} \\ \delta \end{array} \right| + \quad \mu$$

Finally, we can vary the number of agents used to form the estimate of the local spatial residual; in McIntosh (2004) the mean-variance properties of the number of matches is investigated, and it is shown that we can nest the estimator outlined in this paper and the DID estimator by varying the number of matches used. In our empirics we use a single match, which is the minimum-bias estimate, as well as 16 matches, to investigate how altering the size of the local area affects the impact estimates.

V Two Policy Innovations

FINCA Uganda is among the oldest formal microfinance institutions in the country and is one of the largest and best-established in Africa. Their standard lending product utilizes a group-lending methodology, wherein the thirty members of 'village banks' are jointly liable for each others' loans. Loans are made almost exclusively to women. There is no formal screening of new clients, so membership in groups is constrained only by the selection imposed by current clients on members of their community for whom they will accept liability. Loans begin at fifty dollars, and subsequent increases are based on fixed multiples of clients' savings determined by the client's grade, which in turn is based on repayment and attendance of weekly meetings. The standard loan has a 16-week cycle, and clients pay 4 percent per month flat interest (87% effective). Each client is covered by a life-insurance policy, whose premiums are included in the interest payments. FINCA Uganda now has more than 25,000 clients in 1,000 village banks, spread over most of the conflict-free parts of Uganda.

At the beginning of 2000 FINCA Uganda began offering two new policies, as shown in Figure 1. The first, the so-called 'flexibility program' allowed groups to elect (by a unanimous vote) to change from weekly to biweekly repayment. While it is clear that making fewer payments is preferable to clients, the widespread perception that frequent repayment is central to preventing default has made institutions slow to offer this service, and indeed makes members of joint-liability groups reluctant to accept it when it is offered. Thus, the primary concern for those groups which switch to biweekly repayment is whether the reliability of repayment drops. Since clients in groups that have switched to biweekly payment have lower transaction costs, dropout should fall. The same effect would cause new client enrollment to rise, however biweekly groups may also be more *selective* in admitting new clients; thus the effect on the fraction of new clients is ambiguous. Weekly repayment also places a very tight cash-flow constraint on client businesses; from anecdotal evidence, the amount which can be repaid in the worst typical week often determines what clients are willing to borrow. Thus, we hypothesize that biweekly repayment will cause loan volume to increase.

The second new policy offered a health insurance package to village banking clients and their families. The package costs roughly $13 per four-month cycle and covers the client, spouse, and four dependants against routine medical expenses. As an attempt to control adverse selection, more than sixty percent of the clients in any village banking group were required to enroll in the program (in effect, causing the adverse selection of unhealthy *groups* rather than unhealthy *individuals*. Uganda is an environment characterized by high mortality and morbidity; consequently medical costs can constitute a major burden for poor families. Focusing on the risk that health expenditures entail for households, we can identify several interesting behavioral effects from the removal of this source of risk. First, if health shocks are causing delinquency and business failure, then default will fall when this source of risk is removed. Secondly, if households are engaged in precautionary savings or borrowing to buffer against this source of risk, we will observe the elimination of this buffer among newly insured households. Additionally, it is almost certainly the case that FINCA will more attractive to new clients as it adds the health insurance option.

Another feature of the insurance program in practice is that it suffered from major cost overruns; premia calculated on the basis of pre-insurance health expenditures turned out to be far too low to cover costs among the insured. Given the relative destitution of this client base, there is good reason to think that this is a result of more than the standard moral hazard story: if agents were truly constrained in their pre-insurance health expenditures, then the problem they face is fundamentally not one of insurance (which relates to smoothing) but one of limited income. Hence, even in the presence of a well-designed program, we might expect to encounter cost overruns. These were likely exacerbated by the fact that the 60% voting rule does very little to control adverse selection, and does nothing to utilize information possessed by clients about their fellow group members.

Given the somewhat chaotic time period over which this data was collected, we have prior reason to believe that spatial shocks may be present in outcomes. An ebola outbreak occurred in October of 2000; the disease was contained to the northern part of the country but the town of Masindi, a part of the biweekly treatment, was effected. An unexpectedly close presidential election in March of 2001 led to insecurity and some minor rioting in the capital (the health treatment region) but had little impact on business in other parts of the country.

The data are taken from the accounts of FINCA Uganda, and from surveys conducted at the individual and the group level during the test. The number of usable groups, restricted to those which were in existence prior to the beginning of the test and are in the relevant areas, is 450. Data was missing from roughly 10 percent of the surveys, and groups which participated in either treatment were eliminated from the analysis of the other, reducing the overall number of observations to roughly 400 for each program. Table I lists the outcome and explanatory variables used in the analysis. The chooser/non-chooser status of groups in control areas was established through a mock election as a part of the group survey; voting rules were applied in the same way that they had been in the treatment, e.g. 60 percent required for health insurance, and unanimity for biweekly repayment.

In Table II, we see the number of groups that were offered and that chose the treatments. In each case there is a slightly higher fraction of choosers in the control than the treatment, possibly indicating that hypothetical decisions are more easily taken than real decisions; to the extent that this bias exists it creates problems for any assumption across the selection criterion. In Table III, we compare means of the exogenous and (pre-treatment) endogenous variables for both programs. We see evidence of program placement bias in both cases through differences across the treatment criterion. Selection effects are present for both programs; choosers of the biweekly program were more urban and optimistic, had higher grades and lower dropout than non-choosers, and choosers of the insurance program came from older groups and had larger families.

VI Investigating Spatial Effects

We begin our empirical analysis by investigating spatial heterogeneity, using the variation in loan volume among non-choosers of the biweekly treatment as a test case. In order to visualize these spatial effects, we create a moving average composed of the twelve nearest neighbors for each outcome. In Fig. 2, we plot the spatial surface of deviations from the average loan size. The view is from the south-west, over Lake Victoria, and the defining feature of this image is the large spike over the capital Kampala, located on the north-west side of the lake. Obviously, conditions in this city induce clients to take vastly larger loans on average than anywhere else in the country.

Figure 3 is the same except that we now we use *changes* in loan volume. At first glance, this picture appears to be the inverse of the previous, and again we see the enormous difference between the capital and the rest of the country. It would appear that changes in loans are inversely related to size, however the correlation between loan size and loan growth in this subsample is positive (.014). Thus it is more likely that the strong negative spike under Kampala is related to shocks surrounding the 2001 presidential elections, which were disproportionately felt in the capital.

It is unsurprising to find large spatial differences in raw rates of change. The use of a DID estimator in spatial tests, however, relies crucially on the assumption that the control variables will remove all spatial heterogeneity other than that imposed by the test itself. Consequently, if we examine the spatial surface of the *residuals* among non-choosers (who have no treatment effect), under the assumptions of the DID we should see a flat surface, meaning that no significant spatial information remains in the residuals. The FDID, indeed, is measuring the difference in the average height of this surface between the treatment and control regions. In Figure 4 we carry out the exercise of plotting residuals of loan changes, and we see that our control variables have achieved almost nothing in terms of removing the kinds of spatial heterogeneity that will bias the DID; in fact, figures 3 and 4 appear virtually identical.

Under the two assumptions presented in section 3, the surface in Figure 4 is the counterfactual surface for choosers, showing what the spatial heterogeneity in residuals would have been in the absence of the test. The presence of such extreme spatial heterogeneity in this picture, coupled with the fact that the capital is a part of the control for the Flexibility treatment, implies that the DID estimate of impact will be sharply biased upward. The lack of difference between the contours in Figures 3 and 4 indicates that the DID estimate would be almost as biased as an impact estimate arrived at by simply subtracting the raw changes in the control of the raw changes in the treatment.

We can verify the problems inherent to the use of the DID regression in a context with this kind of spatial shocks in several other ways. First, we compose a simple table of the means and t-tests of differences for loan size, loan changes, and residuals of loans taken by choosers and non-choosers of the program across the treatment and the control:

Comparing Differences Across Treatment and Control:

	Non-Choosers			Choosers		
	T	C	t-test	T	C	t-test
Loans	113.9	119.8	1.10	117.4.9	110.4	-.96
Loan Changes	39.0	14.4	-3.57**	36.0	15.2	-2.28**
Residuals	8.8	-14.5	-3.97**	7.06	-8.18	-1.99**

While these residuals are not backfitted, this table tells us several things. First, the use of differencing can exacerbate bias in a spatial DID if the source of bias is spatial shocks. We see this from the fact that the levels of loans are not significantly different between the treatment and control in the non-choosers, but loan changes and residuals are significantly different in both groups. Secondly, we see the inefficacy of this vector of controls at removing spatial heterogeneity, despite the fact that they are fairly standard and were chosen from a large groups of potential controls for their explanatory power. Thirdly, under the assumptions in Section 3, we see that the DID cannot be unbiased when the counterfactual surface (residuals among non-choosers) shows such strong differences between treatment and control. In this example we see that far from being equal to zero, the difference between treatment and control among non-choosers is stronger than among choosers, leading us to believe that the DID coefficient may actually have the wrong sign.

In order to investigate the surprising failure of our vector of controls at removing spatial heterogeneity, we regressed both changes in loan size and the residuals on a set of district-level dummies. In the former regression the R^2 was .073 and in the latter .065, meaning that according to this rough measure our control variables only removed ten percent of the spatial heterogeneity. While we can make no claims as to the controls used in other studies, this result raises serious concerns about the efficacy of the typical battery of cross-sectional controls to achieve ignorability in a spatial treatment/control setup.

Finally, we bootstrap the residual surface in Figure 4. Not only would we like to know whether the spatial effects observed are significant, but we would like to know whether there are significant effects at the locations of the choosers. This achieved by matching each chooser to the residual that is estimated at the location of the closest non-chooser. We then sample with replacement from the empirical marginal distribution of the residuals, randomizing over location, and so we bootstrap datasets that are spatially i.i.d by construction. We then run a spatial smoother (a Gaussian smoother with a bandwidth of .2 standard deviations) over these bootstrapped datasets, and select the envelope that contains 95% of the smoothed surfaces. Finally, we smooth our observed nearest-neighbor residuals with the same smoother, and then compare these pointwise smoothed outcomes with the pointwise smoothed i.i.d. outcomes. Any observations that lie below the confidence region have experienced a significant negative residual effect, and any observations above have experienced a positive effect.

Figure 5 shows the locations of all of the choosers. Those observations which experienced insignificant shocks are denoted with an X, and those with significant shocks are denoted by an arrow pointing in the direction of the effect. It is clear by inspection that the area around and to the west of Kampala, which is the control region, displays more significantly negative residuals, while residuals in the northern and eastern sections of the country, the treatment, are disproportionately positive. While the treatment and the control contain six groups each that experienced significant negative effects, the treatment region contains twelve groups that experienced positive effects, while the control region contains only two. This is consistent with our prediction of a significant positive treatment effect in the False DID. We thus proceed confident that we do indeed experience significant spatial heterogeneity in our data, and thus the standard DID estimates will be biased (in this case upwards).

VII Regression Results

The first column in Table IV shows the DID impact estimates for five outcomes under each of the treatments. The programs appear to have few significant effects, with the exception of an increase in loan volume under the biweekly treatment, significant at the 95% level. These will be unbiased estimates of impact only in the case that no other unexplained spatial effects coincide with the treatment and control regions. To test this, we run the False DID for the same five outcomes in both programs, and these results are reported in column 2. We reject the absence of such effects among the group of non-choosers in three out of ten cases, implying that significant unexplained spatial phenomena are present. Therefore, the DID estimator will be biased in this spatial test, and only because the non-choosers provide us with a counterfactual for these spatial effects are we able to back out an unbiased estimate of the impact of the programs.

In the cases where the false treatment effect observed is similar to the treatment effect measured in the DID, we suspect that the DID is picking up only spatial bias, and there is unlikely to be any real treatment effect. The increase in loan volume in the biweekly program measured by the False DID, for example, is similar to that observed in column 1, implying that all of the 'treatment effect' picked up by the DID is bias. Where we have significant false effects in the FDID which differ from those found in the DID, we suspect that the bias in the DID may be masking real underlying treatment effects.

If the spatial mapping of choosers and non-choosers were identical, then we could get an unbiased treatment estimate by subtracting the coefficient from the second column off of the first. The spatial matching estimator differs from this triple-difference estimate because of differences in the spatial distributions of the two populations. Column 3 gives the estimate of the Spatial Matching estimator using only the nearest neighbor as a match, and column 4 uses the 16 nearest neighbors. Column 4 thus uses a larger area to define the 'local' shock.

These estimates demonstrate no impact of the biweekly program on loan volume, despite the significant positive DID estimate. Where we do see significant effects for the biweekly program are in strongly decreased dropout, and grades that are somewhat higher. So, while we have not seen the jump in loan volumes that was predicted by theory, we do observe that these groups which have tailored products to the clients have managed to become significantly more attractive to current members, and so have improved retention. The costs of lending are, obviously, dramatically lower when the credit officers halve the number of meetings they are required to oversee, and so if this transition leaves loan volumes unchanged and actually improves client quality, it suggests that extending Biweekly flexibility to the rest of the country will both increase the sustainability of the institution and help to tailor products towards the needs of groups.

Intriguingly, every outcome for which the FDID is significant and the DID is not is found under the insurance program. The obvious intuitive interpretation is that the insurance program has insulated participants against a shock observed in an uninsured population, which is precisely what we would expect from such a treatment. New client enrollment is weakly lower, dropout weakly higher, and loans significantly lower for clients within areas offered the insurance treatment who did not choose it. For those who chose and received the treatment, however, we see no evidence of these spatial differences, leading us to infer that the treatment has played a causal role in insulating them.

The insurance program has been strongly effective in attracting new clients, or at least in avoiding a decrease in enrollment which was otherwise occurring. A study of individual clients who have joined groups after they received the insurance product shows that these clients are low-quality borrowers (low grades, small loans, small growth in loans) and hence are probably participating in FINCA only to get access to the Insurance product. The fact that average default does not increase indicates that the screening process continues to be effective in preventing the admission of deadbeat clients. It is also important, however, that insuring clients against health expenditure risk creates no worsening in repayment, implying that illness is not driving default even in this high-mortality environment. It may be that the zero average effect on repayment calculated here masks separate negative effects caused by the newly enrolled clients and positive impacts upon existing clients. Surprisingly, the estimate using only a single nearest neighbor shows a significant *increase* in dropout as a result of the insurance program despite the fact that the DID estimate is lower than the FDID, but with more neighbors forming the match the sign has flipped and the relationship is insignificant.

The fact that the treatment weakly increases borrowing and decreases savings is also interesting, and open to several interpretations. Because the number of new clients has gone up, we expect savings (which start at zero) to drop; however estimates of this effect only explain about \$4, or one third, of the effect seen. This also fails to explain why loans have gone up. A second explanation would be that clients are paying for the health insurance using savings and simply increasing borrowing to compensate. However, the most common pattern is that the premiums were paid out of loan funds, leaving approximately \$6-8 of the fall in savings (out of an average client savings of \$50) unexplained. This is seen as suggestive evidence for the existence of precautionary savings in microfinance institutions being used to buffer households against health risk.

We perform two robustness checks. First, we modify a new method of calculating the variance of matching estimators presented in Abadie & Imbens (2004). Because we perform the spatial matching without replacement, agents may be used as a match multiple times, requiring an upwards adjustment in the estimator of the variance. We re-calculate the standard errors of the spatial matching estimator using the square roots of the corresponding diagonal elements of:

$$(X'X)^{-1}\frac{1}{N-K}\sum_i\Big(\frac{1+K_M(i)}{M}\Big)^2\Big(\frac{M}{M+1}(\mu_i-\hat{\mu}_{i'})^2\Big)$$

where N is the number of agents, M is the number of nearest neighbors used as matches, and $K_M(i)$ is the number of times that each control agent is used as a match. The results are presented in Table V; the sample standard errors from this method are roughly twice as large as those calculated by the normal method. Only the decrease in dropout in the Biweekly program remains significant at the 95% level, and the increase in new clients for the Insurance program falls to 90%

significance. Different FINCA groups often meet at the same location, meaning that multiple groups in the parallel population may qualify as a 'nearest neighbor'. We dealt with this problem in the data by averaging residuals for all units of a given status at a given location, but this multiple matching led to a substantial increase in the estimated variance. A study population with a more dispersed, uniform spatial distribution would likely not see this large difference between standard errors calculated by the two different methods. The additional uncertainty introduced by multiple matches to the same units substantially decreases our confidence in the conclusions, but the main results are still present.

Our second robustness check takes advantage of the fact that for all of our outcomes except savings, we have two pre-treatment observations. This allows us to test the spatial assumption because we can verify whether the spatial shocks incurred by choosers and non-choosers were similar. The results of this exercise are reported in Table VI. Because there are no treatment effects experienced anywhere in this population, both of our DID regressions are now false. The first column reports the False DID among choosers, and the second the False DID among non-choosers. We see once again that significant spatial phenomena exist that were not properly picked up by our control variables. While the number of observations among choosers is limited, we see that in every case where there are significant spatial effects among non-choosers, the magnitude of the effects among choosers is similar. When we run the Spatial Matching estimator (using the normal SE calculation and M=1) we see that there are no significant differences between the spatial shocks experienced by these two populations in the 8 months prior to treatment. We thus fail to reject the spatial assumption.

VIII Conclusion

In the presence of a second control group, we have a way of testing the assumptions underlying the use of a DID estimator of spatial treatment effects. Not only does the application in this paper display heterogeneity of a type that will result in bias in the DID, but our battery of control variables have been largely ineffective in removing this heterogeneity. While this result needs to be tested in other data sets, the implication is that studies which appear to be estimating spatial DID impacts having controlled for all observable differences between regions may be very sensitive to any differences in raw growth rates across the space of the treatment criterion. Particularly when legislative or administrative units are used to define treatment and control regions, we have myriad reasons to expect that unexplained differences will exist. Data collection, managerial talent, differing incentive structures, shocks, and endogenous policy placement can all cause units to differ in unexplained ways across the treatment rule, and hence the DID will be inconsistent.

Through the use of the spatial matching estimator we can directly estimate other forms of unexplained spatial heterogeneity and recover unbiased impact estimates. Using the non-choosers of the program to eliminate this spatial heterogeneity, we find the biweekly repayment program improving retention of existing clients while improving repayment. Since it also lowers transaction costs on both sides of the contract, we can unambiguously recommend extending the choice of this program to all of the groups in Uganda.

Insuring clients against health expenditure risk has no effect on their repayment performance. Since few environments on earth have a higher incidence of endemic diseases than Uganda, this is relatively strong evidence for the fact that health shocks are not a major determinant of microfinance delinquency worldwide. The insurance program retained existing clients more successfully and attracted new ones at a much greater rate than control groups. In addition, it caused a weak drop in savings and increase in loans. This indicates that agents are engaged in precautionary savings to buffer against health shocks, but not in precautionary borrowing; indeed once freed from this source of risk clients apparently feel comfortable accepting *higher* levels of debt. It is very likely that the insurance caused increases in the household welfare of participants, however the program as currently constituted represents a mixed blessing from the perspective of the lender.

The impact of these programs is closely related to the manner in which groups choose whether or not to accept them, and hence to the difference between the average treatment effect and the treatment effect on the treated. We have strong reason to suspect that the ATE of the biweekly repayment program would be negative, particularly in terms of repayment performance. What this experiment shows is that as long as a mechanism exists to expose group members to the negative repercussions of their choices (in this case, joint liability), then the groups themselves will make the correct decisions regarding insurance and risk. The more mixed success of the insurance program illustrates these same concepts. Insured groups were able to accept a large number of new clients who appeared less attractive without sustaining a drop in repayment, indicating an ability to overcome the asymmetric information present in the lending contract. The insurance program itself, however, suffered from adverse selection in client health, huge cost overruns, use of the system which was dramatically higher than predicted, and many reports of misuse of the system. These forms of adverse selection and moral hazard are not covered by FINCA's joint liability, and hence local information was of no use in overcoming these problems. If we wished to learn the lessons of microfinance, a financially sustainable insurance program might include features like a joint co-payment to be made collectively by all members of a group; this will mobilize local information and induce agents to screen.

It has become increasingly clear that program evaluation needs to be built into the design of programs in order for us to have any hope of measuring their effects accurately. In many contexts, the best way of doing this is to randomize some aspect of the design or implementation of the program. What is suggested here is a different path; we show that the use of simple, clear rules for the treatment and selection criteria present us with a measurable dimension along which we need to worry about bias. Because this dimension can be used for matching purposes, we are likely to have much higher-quality controls in situations where assignment into the treatment is based on easily observable criteria. The implication is that, in environments where randomization is not feasible, a spatial testing design with transparent eligibility requirements is an alternate way of producing high-quality impact estimates.

GRADUATE SCHOOL OF INTERNATIONAL RELATIONS AND PACIFIC STUDIES, 9500 GILMAN DR, LA JOLLA CA, 92093-0519. Phone: (858) 822 1125. Fax: (858) 534 3939. Email: ctmcintosh@ucsd.edu.

References

Abadie, A. & Imbens, G. (2004). Large sample properties of matching estimators for average treatment effects. Unpublished working paper.

Angrist, J. (1995). Conditioning on the probability of selection to control selection bias. *NBER Technical Working Paper 181.*

Bertrand, M., Duflo, E. & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. *Quarterly Journal of Economics, forthcoming.*

Card, D. & Krueger, A. (1994). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *The American Economic Review* **8**(4), 772–793.

Dehejia, R. & Wahba, S. (2002). Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics* **84**(1), 151.

Ghatak, M. (1999). Group lending, local information, and peer selection. *Journal of Development Economics* **60**, 27–50.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American Economic Review* **84**(3), 622–641.

Heckman, J., Ichimura, H., Smith, J. & Todd, P. (1998). *Econometrica* **66**(5), 1017–1098.

Kremer, M. (2003). Randomized evaluations of educational programs in developing countries: some lessons. *AEA Papers and Proceedings* **93**(2), 102–106.

McIntosh, C. (2004). The use of two control groups in quasi-experimental program evaluation. Unpublished working paper.

McKenzie, D. & Mookherjee, D. (2003). Distributive impact of privatization in latin america: an overview of evidence from four countries. *Economia* pp. 161–218.

Morduch, J. (1998). Does microfinance really help the poor? new evidence from flagship programs in bangladesh. Unpublished working paper.

Morduch, J. (1999). The microfinance promise. *The Journal of Economic Literature* **37**(4), 1569–1614.

Pitt, M. & Khandker, S. (1998). The impact of group-based credit programs on poor households in bangladesh. does the gender of participants matter?. *Journal of Political Economy* **106**(5), 958–995.

Ravallion, M., Galasso, E., Lazo, T. & Philipp, E. (2002). Do workfare participants recover quickly from retrenchment?. World Bank Working Paper.

Rosenbaum, P. (1982). The role of a second control group in an observational study. *Statistical Science* **2**(3), 292–316.

Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Stiglitz, J. (1990). Peer monitoring and credit markets. *World Bank Economic Review* **4**, 351–366.

Telser, L. (1964). Iterative estimation of a set of linear regression equations. *Journal of the American Statistical Association* **59**(307), 845–862.

IX Tables

I: Dependent and Independent Variables:

The following outcome variables are used in the analysis:
1. Dropout (percent that took a loan last cycle and do not return)
2. New clients (percent of clients starting this cycle that are new)
3. Grades (repayment performance and meeting attendance)
4. Average loans in a village bank
5. Average savings in a village bank

The control variables used are:
1. Loan cycle number (e.g. loans taken by this VB)
2. Ethnic homogeneity of the group
3. Borrower's perception of their local business climate
4. A dummy equal to one if the VB is in a rural area
5. The average number of children in clients' households
6. The average number of non-working adults in clients' households
7. The share of clients in a VB that own their own homes
8. Does the VB conduct other informal internal savings and lending
9. Did the group pre-exist in some form prior to formation of the VB

II: Numbers of Groups in Treatment and Selection Categories:

Biweek:

		Treatment Criterion		
		$T = 0$	$T = 1$	Total
Selection	$\omega = 0$	102	168	270
Criterion	$\omega = 1$	51	59	110
	Total	153	227	380

Insurance:

		Treatment Criterion		
		$T = 0$	$T = 1$	Total
Selection	$\omega = 0$	140	91	231
Criterion	$\omega = 1$	99	18	117
	Total	239	109	348

III: T-tests of Differences in Means Across the Treatment and Selection Criteria:

	Treatment Criterion E(treated)-E(untreated)		**Selection Criterion** E(comp.)-E(non-comp.)	
	Biweek	Insurance	Biweek	Insurance
Exogenous:				
Located in Village	0.0334	-0.1285**	-0.1418**	-0.0073
Cycle No. of Group	0.6935**	-0.0859	0.2689	0.8595**
Group conducts ROSCA	-0.0106	-0.0133	-0.0106	0.1042**
Ethnic Heterogeneity	0.0881	0.0867	-0.0279	-0.1762**
Perceived Econ. Climate	-0.0502	0.2185**	0.2885**	-0.1048
No. of Children in HH.	0.1088	0.1643	-.1019	0.1996**
No. of Adults in HH.	0.1739**	-0.0829	-0.0345	0.2842**
Own Home?	0.0096	0.0237	0.0081	-0.0231
Pre-existing Group	0.0111	-0.0199	-0.0606	-0.0157
Endogenous: (Pre-treatment)				
% Dropout	0.1581	-0.3488	-0.8554**	-0.4072
% New Clnts	-0.5844	-0.1820	-0.1757	-0.7421
Av. Ind. Loans, USD	-11.712**	12.0221**	-1.874	-6.2119
Av. Ind. Savings, USD	-12.0557**	16.4137**	-0.4804	-0.4287
Grades, A=4, D=1	-0.0418	0.0884	0.3845**	0.0018

(**=Difference in means significant at 95% level.)

IV: Regression Results:

IV: Regression Results:	Difference-in-Differences	False Difference-in-Differences	Spatial Matching M=1	Spatial Matching M=16
Biweekly: # obs:	110	270	380	380
% New Clients:	-5.3331 (-1.0223)	.2736 (.0890)	-2.4280 (-.7521)	-.4867 (-.1818)
% Dropout:	-5.0194 (-1.2097)	3.0012 (1.2474)	-12.9544** (-4.7618)	-9.5690** (-4.5962)
Savings (USD):	-4.9995 (-1.1161)	-9.3477 (-1.0105)	13.3042 (1.4288)	8.0221 (1.1692)
Loans (USD):	17.5045** (2.0508)	24.6750** (4.0369)	2.8716 (.4455)	-3.8805 (-.7796)
Grades (A=4, D=1):	-0.0674 (-.4842)	-.0403 (-.4418)	.2468 ** (2.4766)	.0743 (.9671)
Insurance: # obs:	117	231	348	348
% New Clients:	.2263 (.0549)	-3.6515 (-1.6399)	5.2698** (2.1727)	3.8479** (1.9612)
% Dropout:	2.5454 (.508)	4.6773* (1.6934)	6.5050** (2.0535)	-1.5620 (-.6322)
Savings (USD):	0.4842 (.0787)	12.9900 (1.2227)	-10.4638 (-1.2494)	-12.9625 (-1.6056)
Loans (USD):	-4.6143 (-.4485)	-14.7495** (2.1586)	4.2655 (.5716)	9.0664* (1.6566)
Grades (A=4, D=1):	.2357 (1.3368)	.0195 (.2104)	-.0598 (-.5621)	-.0243 (-.2832)

(*=90% significance, **=95% significance)

V: Regression Results using Robust Variance Estimator:

V: Results, robust variance estimator	Spatial Matching M=1	Spatial Matching M=16
Biweekly: # obs:	380	380
% New Clients:	-2.4280	-.4867
	(-.3955)	(-.0867)
% Dropout:	-12.9544*	-9.5690**
	(-1.7730)	(-2.0624)
Savings (USD):	13.3042	8.0221
	(.8912)	(.6094)
Loans (USD):	2.8716	-3.8805
	(.1842)	(-.3860)
Grades (A=4, D=1):	.2468	.0743
	(.9105)	(.4311)
Insurance: # obs:	348	348
% New Clients:	5.2698	6.8398*
	(.9924)	(1.6609)
% Dropout:	6.5050	-1.5620
	(.7695)	(-.2858)
Savings (USD):	-10.4638	-12.9625
	(-1.0575)	(-.8947)
Loans (USD):	4.2655	9.0664
	(.3129)	(.8164)
Grades (A=4, D=1):	-.0598	-.0243
	(-.1838)	(-.1304)

VI: Counterfactual Test of Spatial Assumption:

False Impact Estimates:	Choosers False DID	Non-Choosers False DID	False Spatial Matching
Biweekly: # obs:	72	162	234
% New Clients:	-4.6611 (-1.4165)	-4.0120* (-1.7716)	-3.0326 (-1.1252)
% Dropout:	-.2362 (-.0415)	4.9902 (1.4189)	-2.2653 (-.5778)
Loans (USD):	8.0202 (1.1398)	12.2790** (3.5661)	-2.3783 (-.5098)
Grades (A=4, D=1):	-.0387 (-0.1554)	-.0039 (-.0314)	-.0066 (-.0423)
Insurance: # obs:	61	173	234
% New Clients:	1.7701 (.3329)	-1.0621 (-.5213)	-.0663 (-.0313)
% Dropout:	4.3708 (.6838)	-.3149 (-.0907)	2.5592 (.6696)
Loans (USD):	-9.8486 (-1.0493)	-11.1089** (-3.2083)	-1.6184 (-.3545)
Grades (A=4, D=1):	.0908 (.3328)	.2259* (1.7591)	-.1672 (-1.3548)

X Figures:

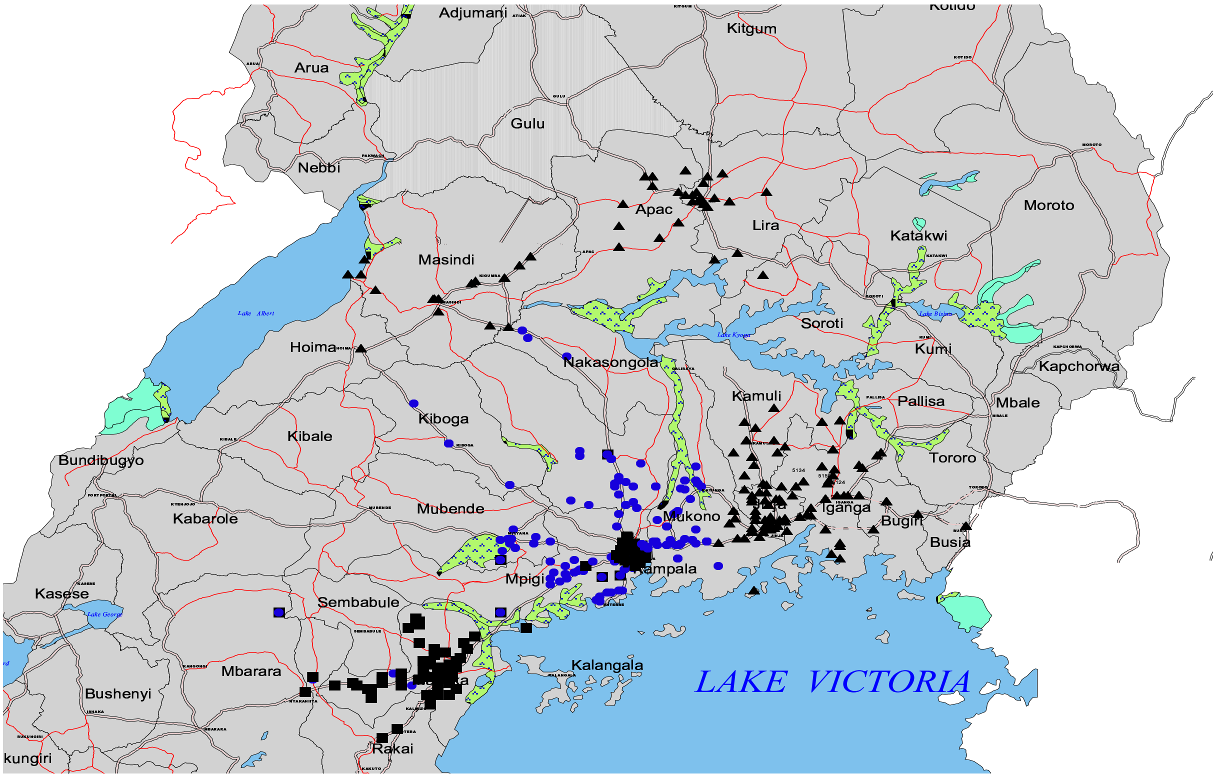

Figure 1: Treatment-control regimes: Circle=control, Square=Health Treatment offered, Triangle=Flexibility Treatment offered

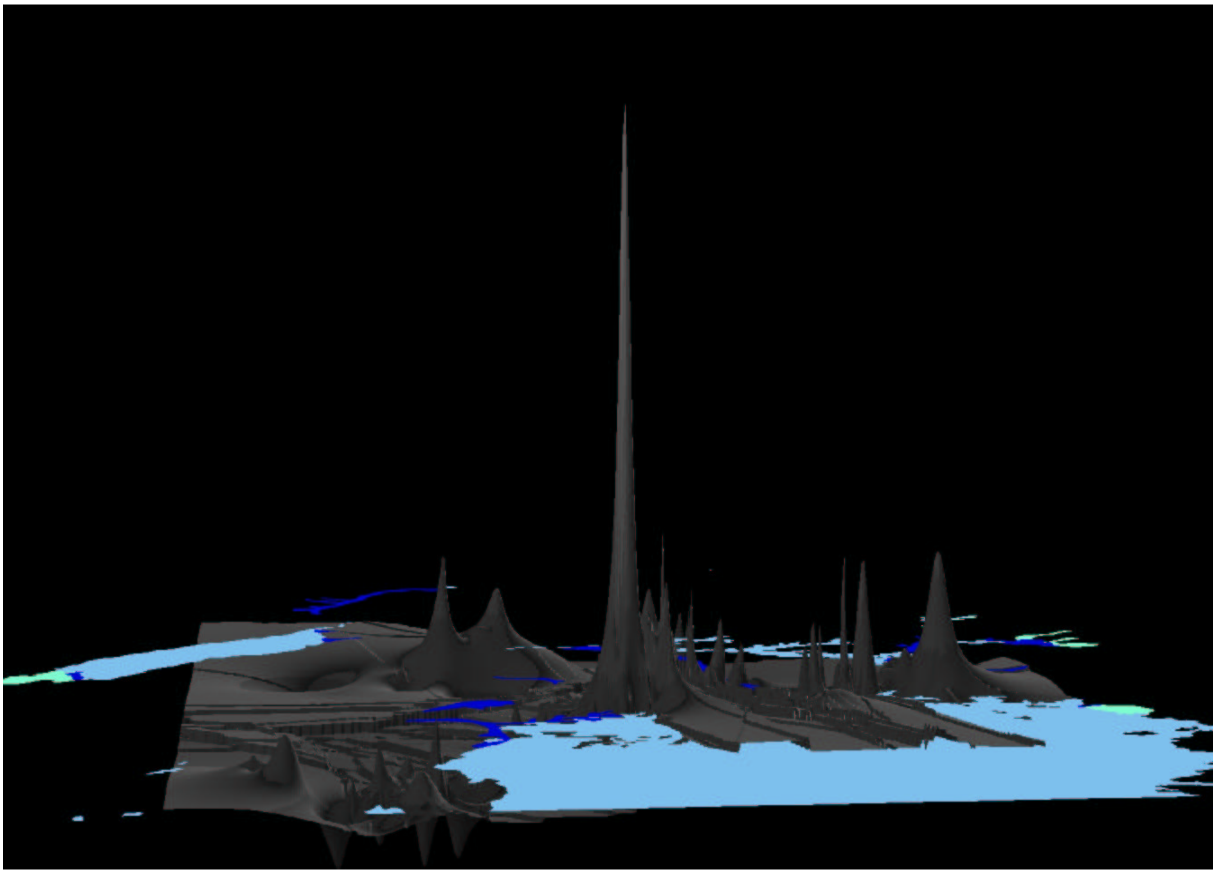

Figure 2: Spatial distribution of Loan Sizes

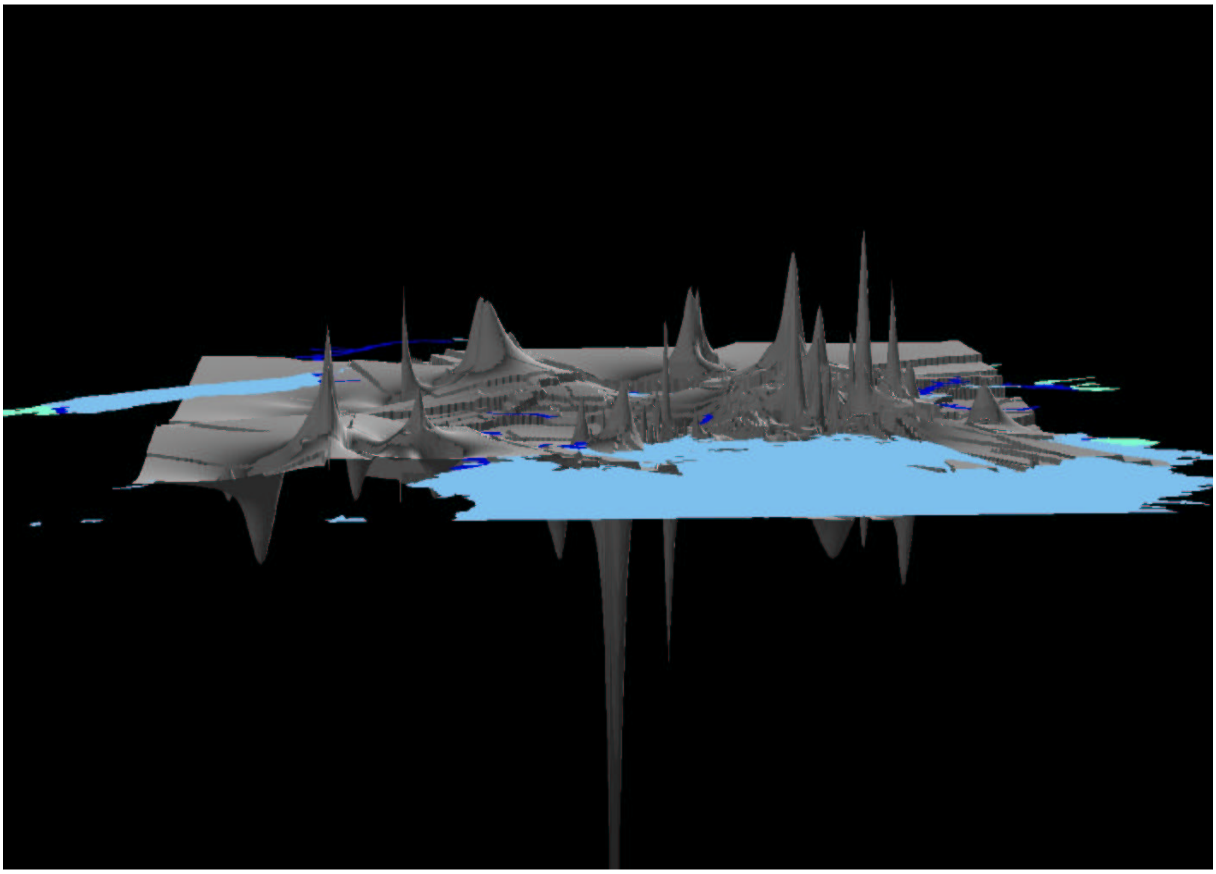

Figure 3: Spatial distribution of Loan Changes

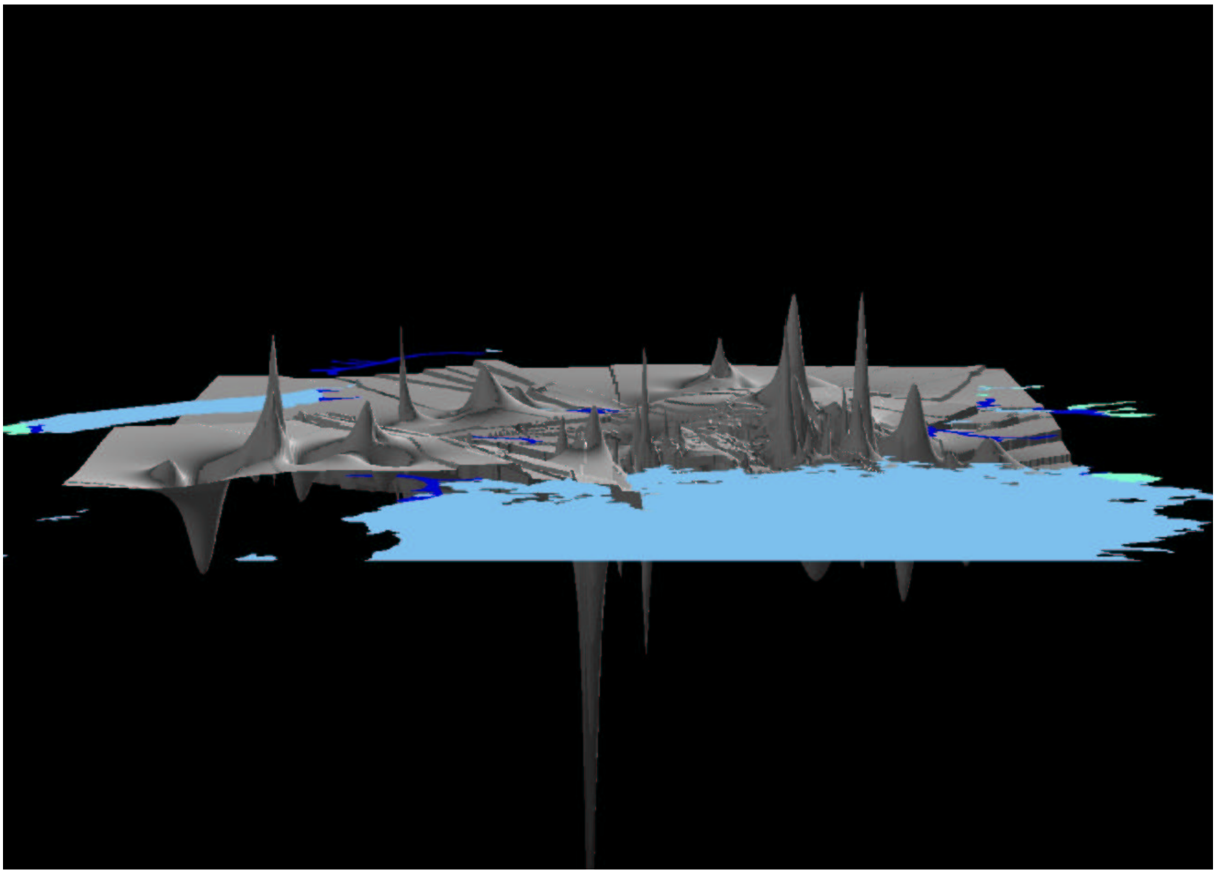

Figure 4: Spatial distribution of Loan Change Residuals

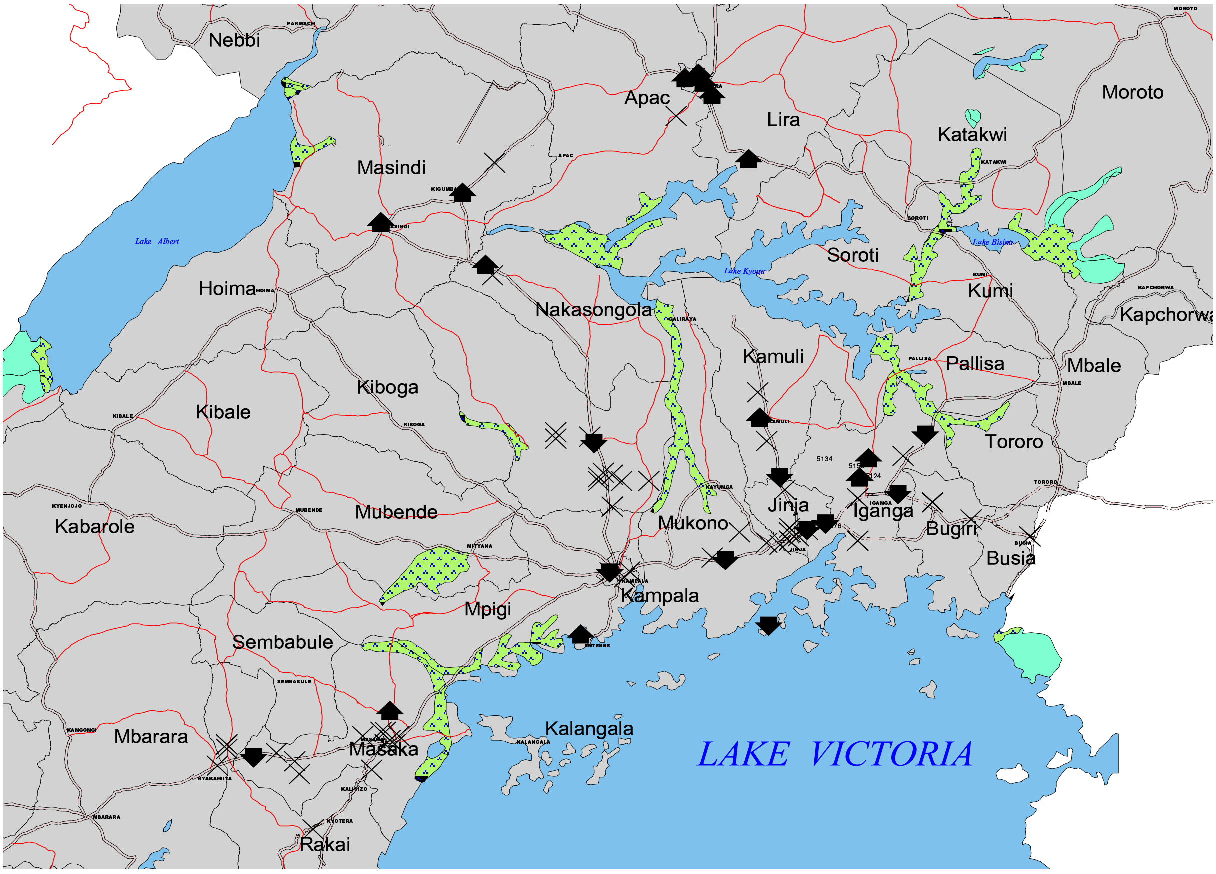

Figure 5: Location of Significant Shocks for Choosers