

# How important is spatial correlation in randomized controlled trials?

Kathy Baylis

Department of Agricultural and Consumer Economics University of Illinois at  
Urbana-Champaign  
[baylis@illinois.edu](mailto:baylis@illinois.edu)

Andrés Ham

Department of Agricultural and Consumer Economics University of Illinois at  
Urbana-Champaign  
[hamgonz2@illinois.edu](mailto:hamgonz2@illinois.edu)

Selected Paper prepared for presentation at the 2015 Agricultural & Applied  
Economics Association and Western Agricultural Economics Association Annual  
Meeting, San Francisco, CA, July 26-28

Copyright 2015 by Kathy Baylis and Andrés Ham. All rights reserved. Readers may make  
verbatim copies of this document for non-commercial purposes by any means, provided that this  
copyright notice appears on all such copies

# How important is spatial correlation in randomized controlled trials?<sup>\*</sup>

Kathy Baylis  
Andrés Ham<sup>†</sup>

Department of Agricultural and Consumer Economics  
University of Illinois at Urbana-Champaign

This version: May 2015

## Abstract

Randomized controlled trials have become the gold standard for impact evaluation since they provide unbiased estimates of causal effects. This paper studies randomized settings where treatment is assigned over geographical units. We analyze how omitting spatial correlation in outcomes or unobservables affects treatment effect estimates. First, we study spatial dependence in Mexico's Progresa program. Second, we conduct Monte Carlo simulations to generalize our results. Findings reveal that spatial correlation is more relevant than the literature suggests, and may affect both the precision of the estimate and the estimate itself. Existing spatial econometric methods may provide solutions to mitigate the consequences of omitting spatial correlation.

**Keywords:** randomization, spatial correlation, treatment effects, estimation, inference

**JEL Classification:** C15, I38, R58

---

<sup>\*</sup>We would like to thank Giuseppe Arbia, Anil Bera, Leonardo Bonilla, Raymond Florax, Don Fullerton, Rafael Garduño-Rivera, Geoffrey Hewings, Roger Koenker, Esteban López, Daniel McMillen, Dusan Paredes, Paolo Postiglione, and Ignacio Sarmiento for fruitful and engaging discussions on this research; as well as seminar participants at the University of Illinois, the 2014 North American Regional Science Conference, and the 2015 Midwest International Development Conference. Marin Thompson provided outstanding research assistance. All remaining errors and omissions are entirely our own and do not necessarily reflect the views of the University of Illinois.

<sup>†</sup>Corresponding author: [hamgonz2@illinois.edu](mailto:hamgonz2@illinois.edu). Mailing address: Department of Agricultural and Consumer Economics, 405 Mumford Hall, 1301 W. Gregory, Urbana, IL 61801. MC-710.

# 1 Introduction

Randomized evaluation has many benefits. Perhaps the most important asset is that it presents social scientists with a simple way to estimate unbiased causal effects. This method has provided a vast amount of evidence on the effectiveness of educational, health, job training, and informational programs on household and individual outcomes. However, the approach also has its critiques. Some of the most debated issues include its lack of external validity and replicability ([Rodrik, 2008](#)), agnosticism with respect to behavioral changes ([Ravallion, 2009](#)), partial equilibrium approach ([Deaton, 2010](#)), and the largest concern: interference.

What is interference? The [Rubin \(1974\)](#) causal model employed by most randomized studies rests on the Stable Unit Treatment Value Assumption (SUTVA), which requires that outcome differences between treated and control units depend solely on own treatment status and not on what treatment others receive. If there is interference or spillover between units, this assumption does not hold, and randomization no longer provides unbiased causal effects because of contamination.

We identify three types of SUTVA violations from previous research which depend on the randomization unit. First, when treatment assignment occurs at the individual-level, interference may be driven by peer effects. Because contamination occurs inside shared networks such as schools or villages, we refer to these as ‘within’ spillovers. This potential violation of SUTVA has received the most attention in recent literature.<sup>1</sup> Second, if randomization occurs at the cluster-level, then the existence of externalities between clusters generates interference. Since contamination occurs among larger units where the internal population are all treated or untreated, we designate this type of interference as ‘between’ spillovers. This form of contamination has received limited

---

<sup>1</sup>For instance, some recent studies analyzing within interference in randomized settings include [Duflo and Saez \(2003\)](#), [Angelucci and DeGiorgi \(2009\)](#), [Bobonis and Finan \(2009\)](#), [Lalive and Cattaneo \(2009\)](#), [Babcock and Hartman \(2010\)](#), [Barrera-Osorio et al. \(2011\)](#), and [Baird et al. \(2014\)](#).

attention but we claim it is no less important.<sup>2</sup> Finally, both types of interference could be present at once, generating ‘mixed’ spillovers.

This study focuses on randomized controlled trials (RCTs) where treatment is assigned over geographical units. Hence, ‘between’ spillovers are the only potential source of contamination. Upon reviewing the RCT literature, we find that many studied outcomes are variables that at least in other contexts, show high degrees of spatial correlation (Barrios et al., 2012). We argue that spatial dependence may possibly lead to between interference. Unlike previous literature on spatial externalities in randomized settings, we explore the type of correlation. A spatially correlated outcome will have different implications than a spatially correlated unobservable. Using spatial econometric models, we show that the first violates SUTVA while the second does not.

To gauge the importance of spatial correlation in RCTs, we revisit Mexico’s Progresa program. Using existing spatial methods such as Moran’s spatial correlation index, robust LM tests, and parametric spatial regressions on village-level data, we find significant spatial dependence. Spatial diagnostic tests reveal evidence of some spatially correlated outcomes (spatial lag process), implying average outcomes in one village affect average outcomes in neighboring villages, raising a potential challenge to SUTVA. In other cases, we find evidence of spatially correlated unobservables (spatial error process), that affects the precision of the estimates. These findings suggest that some additional tests need to be conducted to ensure the internal validity of Progresa’s estimated effects.

We then generalize our results by conducting Monte Carlo experiments that simulate a scholarship program aimed at improving academic performance. We test the efficacy of difference-in-differences with clustered standard errors when the outcome follows a spatial lag and when the error term is spatially correlated. For the first scenario, assume that students from one school compete with students from nearby schools in academic and athletic events, and this interaction leads outcomes in one school to influence

---

<sup>2</sup>To our knowledge, only Miguel and Kremer (2004) and Bobba and Gignoux (2014) have studied and estimated between interference in cash transfer contexts.

outcomes in neighboring schools.<sup>3</sup> For the second scenario, suppose that agricultural productivity is spatially correlated, not easily observable, and affects the demand for child labor. School attendance and performance respond to household demand for child labor which in turn, is affected by a spatially correlated unobservable.

Results show that spatial correlation can affect both the precision of the estimate and the estimate itself. Using standard methods, we conclude that omitting spatial correlation generates noisier estimates, leading researchers to over-accept the null hypothesis that their treatment has no effect. This occurs because clustering standard errors only accounts for within-cluster correlation, assuming that between-cluster correlation is zero. If spatial correlation is present, this does not hold, and this correction requires further refinement. The policy implications are stark, since many interventions may be discarded because they will find no effects.

We propose solutions from existing methods to improve the evaluation of cluster-randomized experiments in the presence of spatial correlation. First, we highlight the importance of spatial statistics to test internal validity assumptions such as SUTVA. Second, we propose available parametric approaches to control for spatial correlation. Finally, we emphasize that results from spatial analysis might help pinpoint unexplored mechanisms that better characterize how programs work.

This article is organized as follows. Section 2 introduces spatial correlation into the potential outcomes framework and surveys the RCT literature, observing if and how it deals with interference and space. Section 3 explores spatial correlation in Progresa. Section 4 conducts the Monte Carlo experiments. Section 5 proposes potential solutions to account for spatial correlation in randomized evaluations. Finally, Section 6 concludes.

---

<sup>3</sup>Another scenario that could lead to a spatial lag in outcomes is to assume that student outcomes in one school affect the demand for community library resources also used by students in neighboring schools.

## 2 Randomization and Spatial Correlation

### 2.1 Randomization

Randomization has become the gold standard in the social sciences because it solves the counterfactual problem and facilitates the estimation of causal effects.<sup>4</sup> To illustrate, consider the potential outcomes framework first proposed by Rubin (1974), where outcomes are determined by a binary indicator,  $T$ , that is equal to 1 if an individual is exposed to a treatment and 0 otherwise. For individual  $i$  we observe:

$$y_i = Ty_i^1 + (1 - T)y_i^0$$

Suppose that  $y_i$  are test scores and  $T$  indicates whether the person receives a scholarship. Researchers observe the outcome for individual  $i$  if they have the scholarship ( $y_i^1$ ) or if they do not ( $y_i^0$ ). Hence, calculating individual-specific treatment effects is impossible since we cannot monitor the same person in both states. This situation is known as the counterfactual problem.

One solution to this problem is to randomly assign  $T$  across individuals. Such an assignment mechanism ensures that treatment status is uncorrelated with other characteristics that may affect academic performance. Randomization thus makes recipients and non-recipients comparable on average observable and unobservable characteristics, and different only in treatment exposure. This facilitates obtaining the unbiased causal effect of scholarships on test scores by estimating a regression of the group difference in means or Average Treatment Effect (ATE):

$$y_i = \alpha + \beta T_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \tag{1}$$

There is one unspoken assumption in this model, the Stable Unit Treatment Value

---

<sup>4</sup>For more in-depth treatment of the concepts we briefly describe here, see Morgan and Winship (2007) and Pearl (2009).

Assumption (SUTVA). SUTVA requires that outcomes for treated and control units depend solely on their own treatment status and not on what treatment others receive. If this assumption does not hold, we have interference or spillovers, and randomization no longer provides an unbiased ATE because of contamination.<sup>5</sup> Formally, SUTVA says that if we have two randomized units,  $i$  and  $j$ , with  $T_i = 1$  and  $T_j = 0$ , then  $y_i \perp T_j$  and  $y_j \perp T_i$ . In our example, say that student  $i$  is randomly selected to receive the scholarship and student  $j$  is randomized out. SUTVA requires that the academic performance of each individual will not be affected by their peer's scholarship status. This rules out the non-scholarship student forming a study group with the scholarship recipient. Is this truly how interactions within schools work? From the evidence in [Sacerdote \(2011\)](#), we believe not.

Despite potential issues with SUTVA, the ability to easily estimate an unbiased ATE has generated strong incentives for conducting randomized controlled trials. In practice, this procedure is carried out in many ways. While some studies randomize at the individual-level, others assign treatment at the cluster-level. In this study, we focus on the latter because chosen clusters are predominantly regions where individuals are all treated or untreated.<sup>6</sup> Researchers usually estimate Equation (1) and correct variance estimates for heteroskedasticity and within-cluster correlation. However, depending on whether spatial correlation exists and in what variables —outcomes or unobservables—, there may be further consequences on the properties of ATE estimates and their variance that have not been considered in previous research.

## 2.2 Spatial correlation in cluster-randomized experiments

Suppose that scholarships are randomly assigned at the village-level and our outcome of interest remains academic performance measured by test scores. Villages may be

---

<sup>5</sup>For more on interference, see [Manski \(1993\)](#), [Hudgens and Halloran \(2008\)](#), [Aronow \(2012\)](#), and [Aronow and Samii \(2013\)](#).

<sup>6</sup>This type of design eliminates within spillovers, so that the only potential form of interference would occur between clusters.

interdependent in a number of ways, but the First Law of Geography frames it succinctly: “everything is related to everything else, but near things are more related than distant things” ([Tobler, 1970](#)). Two types of spatial relationships have been the most widely studied: i) dependence in outcomes or *spatial lag*, and ii) dependence in unobservable characteristics or *spatial error*.<sup>7</sup>

The spatial lag model tells us that outcomes for village  $j$  are directly influenced by outcomes in neighboring villages  $k$ . Supply constraints may generate this situation. If schools are built to serve multiple villages in a 5 km. radius, then academic performance in villages without a school depends directly on how the village with the school performs. Formally, we can rewrite Equation (1) as:

$$y_j = \alpha + \beta T_j + \rho W y_k + \varepsilon_j \quad \text{for } j \neq k \quad (2)$$

where  $W$  is a proximity matrix, whose typical element  $w_{jk}$  is 1 if villages  $j$  and  $k$  are within a 5 km. radius and zero otherwise. The degree of direct spatial correlation is captured by  $\rho$ . If positive, it tells us that villages with higher test scores have neighbors whose average performance is also high. If negative, higher scoring villages will have lower performing neighbors. For most socioeconomic outcomes, we expect the spatial lag coefficient to be positive.

What does the existence of a spatial lag imply for a cluster-randomized treatment? If spatial correlation takes this form, then Equation (2) tells us that  $y_j = f(T_j, y_k(T_k))$ . SUTVA is violated because the treatment status of neighbors affects the outcomes of other villages. The consequences are identical to those from omitted variables bias. In particular, if  $\rho > 0$ , then Equation (1) will overestimate the ATE.

The spatial error model does not assume direct dependence between the outcomes of neighbors  $j$  and  $k$ , but rather that there is some unobservable attribute which is spa-

---

<sup>7</sup>For a textbook treatment of these models, see [Anselin \(1988\)](#), [Arbia \(2006\)](#), and [Ward and Gleditsch \(2008\)](#).

tially correlated and affects the outcome of interest. In our setting, a prime example is agricultural productivity. Productivity measures are generally absent in practice and tend to be correlated over space since some regions are better suited to produce certain goods. Previous evidence has found a link between agricultural productivity and educational outcomes that operates through the labor market ([Ferreira and Schady, 2009](#)). For instance, a positive productivity shock may increase the demand for child labor, attracting some students to the fields and out of the classroom. In this case, we have:

$$y_j = \alpha + \beta T_j + \varepsilon_j + \lambda W \varepsilon_k \quad \text{for } j \neq k \quad (3)$$

where  $W$  is the same proximity matrix as above, and  $\lambda$  measures spatial dependence in village-level agricultural productivity, which is unobservable. As before, we expect a positive association in productivity across villages, since more productive farming areas tend to be clustered together.

The spatial error model does not cause such dire consequences as the spatial lag. Since outcomes between neighbors are not directly affected, SUTVA is not violated and Equation (1) still provides an unbiased ATE. However, variance estimates will be inefficient because of between-cluster correlation in the error term. At higher values of  $\lambda$ , estimates will be noisier. Even while controlling for within-cluster correlation, between-cluster noise may not be eliminated ([Cameron and Miller, 2015](#)).

What is the main difference between both types of spatial correlation? The spatial lag model captures simultaneity because there is feedback between the randomization unit and its neighbors, i.e. the value of  $y_j$  influences  $y_k$  and viceversa. This leads to between interference, which violates SUTVA. In contrast, there is no interference in the spatial error model, but between-cluster correlation lowers precision. Either scenario reveals that spatial correlation may affect consistency or efficiency of ATE estimates.

There are methods to mitigate these issues. Increasingly, data used in RCT evalua-

tions is being geo-referenced. In upcoming sections, we argue that using spatial econometric tools that take advantage of locations will address some problems caused by unaccounted spatial correlation. However, before proceeding, we review existing RCTs to gauge the literature's treatment of interference and spatial dependence.

### 2.3 A survey of recent RCT studies

We compile data on published papers in six journals between 2000 and 2014.<sup>8</sup> Studies were included in this survey if they estimate the effect of a randomly allocated treatment. In total, we find 86 such papers.

Table 1 summarizes the randomization method, journal of publication, setting, and main outcomes of the selected studies. Thirty-one papers assess cluster-randomized programs (36%), while fifty-five evaluate an individual-level treatment (64%). Forty-two articles assess interventions in developing countries and forty-four evaluate programs in industrialized contexts. The most frequently represented nations are the United States (37), Kenya (8), Mexico (8), and India (7). Main outcomes are classified into six categories.<sup>9</sup> We count thirty-four articles analyzing educational outcomes, sixteen observing consumer behavior, seven discussing health, five dealing with micro-credit, four focusing on insurance, and four evaluating effects on investment.

Just over a third of the surveyed papers evaluate a cluster-randomized program, of which only twelve discuss interference. Most analyze forms of within spillovers, with eight studies investigating how ineligible individuals benefit from their treated peers ([Banerjee et al., 2007](#), [Kremer and Miguel, 2007](#), [Angelucci and DeGiorgi, 2009](#), [Bobonis and Finan, 2009](#), [Glewwe et al., 2010](#), [Duflo et al., 2011](#), [Muralidharan and Sundararaman, 2011](#), [Attanasio et al., 2012](#)). The remaining four papers discuss between interference,

<sup>8</sup>The journals are the American Economic Review, the American Economic Journal: Applied Economics, the American Economic Journal: Policy Economics, the Quarterly Journal of Economics, the Journal of Political Economy, and the Review of Economic Studies.

<sup>9</sup>Papers could fall into more than one category if, for example, they were studying the effect of education on investment. Unfortunately, some of the 86 papers were not directly classifiable and were thus omitted from these calculations.

specifically whether ineligible groups are indirectly affected (Olken, 2007, Bobonis, 2009, Kremer et al., 2011, Pradhan et al., 2014). This denotes a trend in the randomization literature. On one hand, within spillovers have so far received the most attention.<sup>10</sup> On the other, between interference has been largely ignored.

Nevertheless, our study is not the first to consider the existence of spillovers across clusters, their consequences, and how to estimate them. To our knowledge, two studies have addressed between interference in RCTs. Both consider interference generated by spatial externalities.

Miguel and Kremer (2004) calculate the effect of a de-worming treatment randomized across schools on education and health outcomes. Unlike most of the papers in our survey, they acknowledge that the intervention had potential externalities and estimate them. Using the density of treatment within villages as their source of regional treatment variation, they find that for every additional 1,000 dewormed children living within a 3 kilometer radius, enrollment further increases by 2 percentage points and rates of moderate to heavy worm infections fall by 26%. They conclude that the spatial externalities associated with de-worming are large enough to justify fully subsidizing de-worming treatment.

More recently, Bobba and Gignoux (2014) estimate the spatial externalities of the Progresa program on secondary school enrollment. They find that in addition to the 9.5% increase expected in a treated village, living within five kilometers of a treatment community further raises secondary attendance rates by 3.6%. This finding is taken as evidence of significant spatial spillovers.

While similarly grounded, our analysis differs from this last study in several ways. First, we are interested in the form of spatial correlation. As shown, the type of dependence has distinct implications on the ATE distribution. Second, we employ existing

---

<sup>10</sup>In fact, research in this area has led to theoretical developments, most notably Baird et al. (2014), who propose a method for designing randomized experiments to consistently estimate within spillovers. Their randomized saturation strategy demonstrates the increased importance that interference has begun to receive and how design-based methods can help.

tools to quantify the existence and degree of spatial correlation in Progresa. These methods are frequently used in spatial econometrics and have a history of applications in other fields. Third, Bobba and Gignoux's definition of neighbors varies over time since they exploit changes in Progresa's eligibility criteria. This approach allows them to estimate spillover effects without using spatial econometrics. Our concern is that most RCTs do not have frequent changes in cluster eligibility. While creative, their identification strategy would only be applicable to a handful of programs. Fourth, our analysis of Progresa extends to other outcomes besides secondary enrollment. Last, we attempt to generalize our findings by conducting Monte Carlo simulations.

### 3 Spatial correlation in Progresa

#### 3.1 Program description

Mexico's *Programa de Educación, Alimentación y Salud* (Progresa) began in 1997 and remains in place today. It is now the largest social program in Mexico and widely considered the landmark conditional cash transfer program.<sup>11</sup> The intervention's main objective was to alleviate poverty and foster the accumulation of human capital by providing transfers to eligible households conditional on regular school attendance and periodic health check-ups for young children and pregnant women.

We revisit the first phase of Progresa, that took place between 1997-1999. The advantage of using this period is that 506 rural villages were chosen for randomized evaluation.<sup>12</sup> Eligibility was determined in two-stages. First, villages were randomly assigned to either receive transfers or not. The resulting allocation selected 320 villages into treatment and 186 communities to control. Second, low income households within

---

<sup>11</sup>The program is now called *Prospera*, and has been renamed multiple times over the years, usually coinciding with its expansions.

<sup>12</sup>As [Gertler \(2004\)](#) notes, from 1997-1999, Progresa extended benefits to approximately 2.6 million families in 50,000 rural villages. This represented approximately 40% of rural households and 10% of all Mexican households.

treatment villages were classified as eligible or ineligible, with the former becoming transfer recipients.<sup>13</sup>

We exploit the cluster-level randomization of transfers to study spatial correlation in Progresa villages. Using publicly available surveys, we first aggregate individual data at the village-level. Then, we employ spatial methods to quantify the existence and type of spatial dependence.

### 3.2 Data and methods

The data for Progresa's rural phase is publicly available on the website of the Mexican Ministry of Social Development ([SEDESOL](#)). The microdata is a census of beneficiaries that spans four waves: before program implementation (September 1997 - March 1998) and for three follow-ups in November 1998, March 1999, and November 1999. The data contains individual information on outcomes similar to the World Bank's LSMS surveys.

Since our interest lies in spatial correlation in outcomes or unobservables across villages, we aggregate the individual data at the village-level. In addition, to define links between villages we require measuring distances between communities. The data is geo-referenced using official shapefiles from the National Statistics Institute (INEGI) that include the latitude and longitude of all villages in Mexico.<sup>14</sup> Figure 1 plots the location of Progresa villages. Targeted communities are located in the central-south region of the country, roughly coinciding with the poorest areas in Mexico.

To gauge the importance of spatial correlation before the transfers, we first analyze baseline data. We restrict the analysis to outcomes that have been previously evaluated in the literature, such as primary and secondary enrollment, labor supply, and per capita income and consumption. To determine if outcomes are spatially correlated we employ Moran's  $I$  ([Moran, 1948](#)):

---

<sup>13</sup>On average, 78% of households in treatment villages were eligible to receive the transfers.

<sup>14</sup>The shapefiles we employ for the geo-referencing are publicly available and may be downloaded [here](#).

$$I = \frac{N}{\sum_j \sum_k w_{jk}} \frac{\sum_j \sum_k w_{jk} (y_j - \bar{y})(y_k - \bar{y})}{\sum_j (y_j - \bar{y})^2}$$

where  $N$  denotes the total number of villages. For two villages  $j$  and  $k$ ,  $I$  calculates the correlation between the outcome in village  $j$ ,  $y_j$ , and its neighbors,  $y_k$ . Neighbors are identified by a weights matrix,  $W$ , with typical element  $w_{jk}$ . Any  $w_{jk} > 0$  tells us that villages are linked. This measure is a correlation coefficient, and therefore takes values between  $[-1, 1]$ . We also test the statistical significance of the index.<sup>15</sup>

Having tested if spatial correlation is present, the next step is to determine the type of dependence: lag or error. [Anselin et al. \(1996\)](#) developed a series of LM multiplier tests based on the residuals from an OLS regression. These tests are robust to local mis-specification and are suitable to determine the source of spatial dependence. Based on these diagnostics, we proceed to evaluate program effects controlling for the form of correlation by estimating spatial difference-in-difference models at the village-level.

### 3.3 How important is spatial correlation in Progresa villages?

Before proceeding, we need to define our spatial network, i.e. to determine how ‘neighbors’ are defined. We must be sure that villages are near each other. Figure 2 provides some insight. The map shows a close-up view of Progresa villages by treatment status. Visually, treatment and control villages seem fairly close together.

For simplicity, we choose to model interactions between villages by defining a  $k$ -nearest neighbor spatial network. Three possible links are considered: a village is related to its closest neighbor,  $k = 1$ ; its two nearest neighbors,  $k = 2$ ; and five nearest neighbors,  $k = 5$ .<sup>16</sup> A visual depiction of these interactions is presented in Figures 3a, 3b, and 3c, respectively. Note that some of the networks clearly delineate small clusters of villages

---

<sup>15</sup>The index’s expected value under the null of no spatial correlation is  $\mathbb{E}(I) = -1/(N - 1)$ . Hence, it is not actually zero, but tends asymptotically to this limit.

<sup>16</sup>We do not assume that links between villages are symmetric. Hence, village  $j$  may be linked to village  $k$ , but the converse need not apply.

such as  $k = 2$ . Our next step is to employ spatial tools to obtain evidence of whether spatial correlation exists before transfers are paid out and if so, what form it takes.

Table 2 provides estimates of Moran's  $I$  for the selected baseline outcomes and spatial networks. First, we notice that treatment assignment shows no significant spatial correlation. This provides a test of balance over space, since we should expect no spatial correlation if randomization was successful. Nevertheless, all remaining outcomes show positive and significant spatial dependence, regardless of the network we use. Since correlations are positive, villages with better outcomes tend to be clustered together. For example, communities with higher enrollment rates, income, and consumption levels are closer together.

We highlight that even while treatment assignment shows no spatial correlation, outcomes are geographically related. Therefore, an insignificant Moran's  $I$  in the treatment variable is insufficient evidence that SUTVA holds. For that, we need to identify the type of spatial dependence. On one hand, it may be driven by a spatially-correlated unobservable. Since treatment is uncorrelated over space, the existence of this spatially correlated unobservable that affects outcomes should not lead to bias, but will affect standard errors. On the other hand, spatial correlation in the outcomes could imply that outcomes in one village affect the outcomes of others. This scenario does violate SUTVA, in that if treatment affects outcomes in a treated village, those outcomes affect outcomes in untreated villages nearby. To address this, Table 3 presents the robust LM test statistics for spatial dependence.

[Anselin et al. \(1996\)](#) derived the distribution of these test statistics as  $\chi^2_2$ , so that the critical value for significance at the 5 percent level is 5.99. The results in Table 3 show that conditional on the network, outcomes have different types of spatial dependence. For instance, secondary enrollment, female labor supply, and per capita income seem to be spatially lagged and thus SUTVA may not hold. Conversely, primary enrollment, labor supply, male labor supply and per capita consumption have a spatially correlated

unobservable and SUTVA is valid. We note that the chosen network matters. Educational outcomes will have different definition of neighbors than labor market outcomes. Ultimately, researchers should assess regional dependence for each outcome and determine which network is best. In our case, we conduct the remainder of our analysis using a two nearest neighbor network for simplicity and because it seems to delineate more pronounced clusters in Progresa (see Figure 3b).

To determine the consequences of spatial correlation on treatment effect estimates in Progresa, we estimate difference-in-difference regressions on village-level data. We use a before/after approach, where all follow-up waves are considered the after period. The estimating equation is:

$$y_{jt} = \alpha_{post} + \beta(post \times T_j) + \gamma T_j + \theta X_{jt} + \mu_j + \varepsilon_{jt}$$

We estimate this specification in two ways. First, using linear methods that do not control for spatial correlation. Second, we use the spatial econometric model consistent with the LM test result for the outcome variable. Both specifications cluster standard errors at the village-level.

Table 4 presents the results. For all outcomes, the baseline DD results are consistent with effects found in the literature. Accounting for spatial correlation does not substantially change the estimates of Progresa's effect. However, the spatial coefficients are all significant and denote positive dependence, ranging between 0.15-0.37. Some program effects become insignificant, such as those on secondary enrollment and per capita consumption.

Assessing spatial correlation in Progresa tells us that village-level dependence exists, and in some cases may pose a threat to internal validity in theory. However, treatment effect estimates are not substantially affected when controlling for spatially correlated outcomes or unobservables. This suggests that spatial correlation might need to be

much higher than found in those villages. To determine if this is the case, we proceed to conduct a series of controlled simulations.

## 4 The Broader Implications of Spatial Correlation

### 4.1 Setting

Our story begins with a government wanting to implement a pilot cash transfer to improve secondary school performance in the poorest region of the country. If the pilot is successful then the intervention will be scaled up, otherwise it will be discarded. Thus, there are substantial policy implications riding on this evaluation.

Mindful about the benefits of randomization, a group of economists designs the intervention so transfers are randomly allocated at the village-level. Moreover, to avoid within spillovers, all households in treated villages receive the transfer and families in control villages do not. Furthermore, aware of inference issues with cluster-randomization, the researchers follow the literature and cluster their standard errors at the village-level.

Using simulated data, we test the performance of treatment effect estimates with clustered standard errors under two types of spatial dependence: a spatially lagged outcome and a spatially-correlated unobservable. For the first, assume that schools complete regionally, and through these athletic and academic contests, academic outcomes in one school affect the academic outcomes of the other. For the second, suppose the existence of a time-variant omitted variable measuring agricultural productivity that is correlated among villages.

### 4.2 Simulation setup

The simulations are carried out in six steps.<sup>17</sup> First, our procedure involves creating villages in the poorest region of the country. For simplicity, we simulate  $V = 100$  villages

---

<sup>17</sup>The replication code for the Monte Carlo simulations is available upon request.

and locate them using random  $(x, y)$  coordinates.<sup>18</sup> Second, we define a spatial network using a  $k$ -nearest neighbor approach. Results are presented for the simplest case,  $k = 1$ , but the simulations are also carried out with  $k = 2$ , and  $k = 5$  for robustness. Each network is summarized in a  $V \times V$  proximity matrix,  $W$ .

Third, villages are randomly selected into treatment. We draw a random number from a uniform distribution on  $[0, 1]$ , and set an exogenous eligibility threshold at  $\bar{T}^* = 0.50$ . All villages with a score above this threshold form the treatment group and those below the cut-off are classified as controls. In the fourth step, we populate villages. For simplicity, we assume that the selected region has  $I = 10,000$  beneficiaries and randomly place each individual in one of the 100 villages. We do not impose that every village has the same number of inhabitants, allowing a better approximation of actual population patterns where some areas are more densely populated than others.

The fifth step simulates a Data Generating Process (DGP), first at village-level and then by individuals. The specifics for each type of spatial dependence are described below. Following the literature, we assume that treatment is estimated using a difference-in-differences approach to control for time-invariant differences in treatment and control villages that persist after randomization. Finally, we cluster standard errors at the village-level. Our main results include the distribution of the estimate, bias calculations, and efficiency measures.

Not all the steps are repeated in the simulations. For instance, the spatial network remains fixed. Figure 4 plots a random draw of village locations, treatment status, and their links. Steps 3-6 are repeated  $R = 5,000$  times for different degrees of spatial correlation in outcomes or unobservables.

---

<sup>18</sup>The smaller number of villages considered in the simulations intends to reflect that most randomized evaluations tend to be quite small in the pilot phase, unlike Progresa.

### 4.3 Spatially correlated outcome

As mentioned previously, the existence of a spatial lag tells us that outcomes in a village depend on its neighbors' outcomes. In our hypothetical country, let us assume that neighboring schools compete, and thus high performance from one school spurs higher performance from neighboring schools.

Since academic performance at the village-level is spatially lagged we may write:

$$y_{jt} = \alpha post + \beta(post \times T_j) + \gamma T_j + \mu_j + \rho \sum_{k \neq j} w_{jk} y_{kt} + \varepsilon_{jt}$$

and after rearranging and gathering terms, we obtain the simpler form:

$$y_{jt} = (I - \rho W)^{-1} [\alpha post + \beta(post \times T_j) + \gamma T_j + \mu_j + \varepsilon_{jt}] \quad (4)$$

where  $y_{jt}$  is village  $j$ 's average academic performance at time  $t$ . The weights matrix,  $W$ , summarizes the neighbor network and  $\rho$  measures the strength of the spatial lag. The parameter  $\alpha$  captures aggregate time trends after program exposure,  $\gamma$  controls for any pre-existing differences between treatment and control villages, and the coefficient on the interaction term,  $\beta$ , is the program's effect on academic performance. Last,  $\mu_j$  is a randomly drawn fixed effect that captures time-invariant village-specific factors and  $\varepsilon_{jt}$  is a normally distributed mean zero disturbance term.

Intuitively, Equation (4) tells us that village-level academic performance is a spatially weighted average of time trends, treatment status, pre-existing differences between groups, unobserved heterogeneity, and an error term. Recall that interference arises because a village's average academic performance is a function of its own treatment status  $T_j$ , but also depends on the treatment status of neighboring villages,  $T_k$ .

Academic performance for student  $i$  in village  $j$  at time  $t$  is determined by average

village performance and individual-level covariates:

$$S_{ijt} = y_{jt} + \boldsymbol{\theta} \mathbf{X}_{ijt} + u_{ijt} \quad (5)$$

where  $y_{jt}$  is defined as in Equation (4),  $\mathbf{X}_{ijt}$  is a matrix of individual covariates that includes a constant, an indicator variable for female students, and father and mother's years of schooling. The error term,  $u_{ijt}$ , follows a standard normal distribution.

Researchers evaluating this simulated program will have access to panel data on individuals. To estimate the effect of the program, they are likely to specify the following empirical equation:

$$S_{ijt} = \alpha post + \beta(post \times T_j) + \gamma T_j + \boldsymbol{\theta} \mathbf{X}_{ijt} + \mu_j + u_{ijt} \quad (6)$$

and will generally estimate by difference-in-differences with clustered standard errors at the village-level.

We set the parameter values at:  $\alpha = 0.04$ ,  $\beta = 0.10$ ,  $\gamma = 0$ ,  $\boldsymbol{\theta} = (1, 0.025, 0.07, 0.12)$ , and vary  $\rho = \{0, 0.10, 0.25, 0.50, 0.75, 0.90\}$ . Our interest lies in the distribution of the treatment effect estimate,  $\hat{\beta}$ .

What results, if any, should we expect? SUTVA is violated since own outcomes are affected by neighbors' treatment status. To our knowledge, no evidence has been provided on spatially lagged outcomes in randomized evaluations, particularly when randomization occurs at the same level as the spatial correlation itself. Therefore, while we expect some bias, we are unsure of its magnitude.

The simulation results are presented in Table 5. In general, we do not find large bias from the estimation strategy unless the spatial spillover is high ( $\rho \geq 0.75$ ). In these extreme cases, the program effect tends to be overestimated anywhere between 60-270%. However, perhaps the most important finding in this scenario is that clustering standard errors at the village-level performs poorly. By construction, the simulations assume that

the effect of cash transfers on academic performance will lead to a statistically significant 10% increase. Hence, accurate inference should reject the null hypothesis of no effect 95% of the time. Using the standard difference-in-differences specification in the literature finds an effect only 6-7% of the time.

Figure 5 shows this result graphically. As the spatial lag becomes stronger, the distribution of the treatment effect estimate widens. Therefore, even while the point estimate may be correct, its variance is not. Ultimately, researchers will tend to over-accept the null hypothesis that the program has no effect. The policy implications are stark, since evaluators may wrongfully conclude that the program is ineffective and discard it because spatial dependence in outcomes contaminated their estimates.

#### 4.4 Spatially correlated unobservables

We now explore how treatment effect estimates respond to the existence of a spatially correlated omitted variable, focusing on the case of unobserved agricultural productivity.

In many randomized evaluations, the target population is located in rural areas where agriculture is the main occupation. Productivity is hard to measure empirically since it is directly affected by land quality and other unobservables in a particular geographic location ([Benjamin, 1995](#)). Moreover, it usually changes over time, responding to natural, economic, and other shocks. Previous evidence has found a link between agricultural productivity and education that operates indirectly through the labor market ([Ferreira and Schady, 2009](#)). For instance, a positive productivity shock may increase labor demand, attracting some students to the fields and out of the classroom. Negative shocks have the opposite effect.

Suppose agricultural productivity is determined at the village-level by:

$$\nu_{jt} = (I - \lambda W)^{-1} \varepsilon_{jt} \quad (7)$$

where as above,  $W$  is the weights matrix that summarizes the spatial network. To differentiate from the previous simulation, we measure the strength of this spatial correlation with the parameter  $\lambda$ . Random noise  $\varepsilon_{jt}$  follows a standard normal distribution. In words, Equation (7) tells us that village  $j$ 's agricultural productivity is related to its neighbors' productivity level, unless  $\lambda = 0$ .

Individual-level outcomes for student  $i$  in village  $j$  at time  $t$  are determined by:

$$S_{ijt} = \alpha post + \beta(post \times T_j) + \gamma T_j + \boldsymbol{\theta} \mathbf{X}_{ijt} + \mu_j + \nu_{jt} + u_{ijt} \quad (8)$$

where the variables are defined as before. The main difference is that agricultural productivity,  $\nu_{jt}$ , now affects student outcomes. Note that we assume that productivity is uncorrelated with all the independent variables, which means that it becomes part of a composite error term  $\psi_{ijt} = \nu_{jt} + u_{ijt}$ .

Once more, the estimated equation will take the form:

$$S_{ijt} = \alpha post + \beta(post \times T_j) + \gamma T_j + \boldsymbol{\theta} \mathbf{X}_{ijt} + \mu_j + \psi_{ijt} \quad (9)$$

where as before, we set the parameter values at:  $\alpha = 0.04$ ,  $\beta = 0.10$ ,  $\gamma = 0$ ,  $\boldsymbol{\theta} = (1, 0.025, 0.07, 0.12)$ , and vary  $\lambda = \{0, 0.10, 0.25, 0.50, 0.75, 0.90\}$ . Equation (9) is once again estimated by difference-in-differences with clustered standard errors at the village-level.

We know that a spatial error leads to inefficient estimates for any non-zero  $\lambda$ . So, program effect estimates are expected to become less precise as agricultural productivity is more spatially correlated. However, the amount of inefficiency is unknown *ex ante*.

Table 6 provides the simulation results from estimating Equation (9). Because treatment is uncorrelated with agricultural productivity, we find no bias from omitting this variable from the empirical specification. However, clustering the standard errors again leads to an over-acceptance of the null hypothesis that the program has no effect. While

a  $\lambda = 0$  provides accurately sized testing procedures, these become less precise as the degree of spatial correlation in the omitted variable grows. This ultimately leads to noisier treatment effect distributions, as depicted in Figure 6. Therefore, while controlling for within-cluster correlation remains important, not accounting for between-cluster correlation results in efficiency loss. Moreover, this increases the likelihood that a successful program is discarded even though it is effective.

## 5 How do we deal with spatial correlation?

We briefly discuss some lessons and potential solutions to address spatial correlation in randomized controlled trials. First, we highlight the importance of spatial statistics to test assumptions such as SUTVA. Second, we propose parametric approaches to control for spatial correlation. Finally, we emphasize that results from spatial analysis might help pinpoint unexplored mechanisms that explain how programs work.

Spatial analysis may complement existing tests for the internal validity of randomization. For instance, calculating Moran's  $I$  on treatment status will tell us if assignment is spatially correlated. Ideally, it will not be; but if it is, then [Barrios et al. \(2012\)](#) discuss the implications. However, this test is insufficient to rule out interference. To test if SUTVA is unaffected by spatial correlation, we propose using the diagnostic tests developed by [Anselin et al. \(1996\)](#). They are easily implementable since they only require estimating OLS regressions. Depending on the results from these computations, researchers may identify if their outcomes may suffer from bias or efficiency problems due to spatial correlation. While our results indicate that bias is not the greatest concern, these measures may be useful to further scrutinize the internal validity of the RCT.

Since estimating spatial models requires defining a spatial network and bias only occurs at high levels of spatial dependence, estimating such models may serve as a robustness check for evaluation results. Moreover, calculating the marginal effects may

quantify any spatial spillovers without the need for alternative identification strategies as in [Bobba and Gignoux \(2014\)](#). Since inference seems to be the biggest concern, perhaps multi-way clustering, that controls for both within and between-cluster variation may provide more efficient estimates. Spatial HAC corrections have been proposed as a viable solution ([Cameron and Miller, 2015](#)).

Additionally, a recurring theme in the interference literature is that outcomes can be affected by both individuals within the group and by neighboring groups. Since most RCTs are randomized by villages and evaluated using microdata, aggregating the data may be costly in terms of informational loss and power. A possible parametric solution to more appropriately control for spatial correlation across clusters and estimate program effects on microdata would be to consider hierarchical or multi-level spatial models ([Banerjee et al., 2003](#)). In future work, we will explore the efficacy of these methods to control for spatial correlation in randomized settings.

Conducting spatial analysis in randomized settings is useful for more than testing assumptions and identifying likely estimation issues. For instance, maps provide a visual depiction of intervention areas that allows thinking about geographical mechanisms. In fact, four studies in our survey suggest that spatial mechanisms may drive their results but conduct no quantitative analysis to confirm their hypothesis. Geo-referenced data is becoming more common, which facilitates conducting spatial analysis. Therefore, using available tools from regional science may potentially improve existing and future randomized evaluations.

## 6 Conclusion

In this paper, we studied the consequences of spatial correlation on cluster-randomized controlled trials where randomization is carried out over geographical units. We analyze how unaccounted spatial correlation in outcomes or unobservables affects treatment ef-

fect distributions in the Progresa program and also conduct Monte Carlo simulations to generalize our results.

The main findings reveal that spatial spillovers can affect both the precision of the estimate and the estimate itself. Using standard difference-in-difference methods, we conclude that omitting spatial correlation generates noisier estimates, leading researchers to over-accept the null hypothesis that their treatment has no effect. The policy implications are stark, since many interventions may be discarded because of unaccounted spatial correlation.

We thus propose some existing tools to improve the evaluation of cluster-randomized experiments in the presence of spatial correlation. First, we highlight the importance of spatial statistics to test internal validity assumptions such as SUTVA. Second, we propose parametric approaches to control for spatial correlation. Finally, we emphasize that results from spatial analysis might help pinpoint unexplored mechanisms that explain how programs work.

Overall, our results suggest that spatial methods may enhance the evaluation toolkit when spatial correlation is important, but will not be applicable to all RCTs. For instance, the benefits of spatial econometric solutions are clear-cut if correlation over space is unaccounted for in the design phase. Furthermore, we are unsure of the additional complications that could arise when within interference are also present. Finally, context matters. Spatial correlation may be relevant in some countries and for some outcomes more than others, so the examples presented here need not generalize everywhere. Nevertheless, we believe that analyzing spatial correlation in RCTs would improve existing and future evaluations by providing further robustness tests to obtain more credible and conclusive results.

## References

- Angelucci, M. and DeGiorgi, G. (2009). Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption? *American Economic Review*, 99(1):486–508.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. NATO Asi Series. Series E, Applied Sciences. Springer.
- Anselin, L., Bera, A. K., Florax, R., and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104.
- Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Advances in Spatial Science. Springer Berlin Heidelberg.
- Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16.
- Aronow, P. M. and Samii, C. (2013). Estimating Average Causal Effects Under Interference Between Units. *ArXiv e-prints*.
- Attanasio, O. P., Meghir, C., and Santiago, A. (2012). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *Review of Economic Studies*, 79(1):37–66.
- Babcock, P. S. and Hartman, J. L. (2010). Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment. NBER Working Papers 16581, National Bureau of Economic Research, Inc.
- Baird, S., Bohren, A., McIntosh, C., and Ozler, B. (2014). Designing experiments to measure spillover effects. Policy Research Working Paper Series 6824, The World Bank.
- Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedyng Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 122(3):1235–1264.
- Banerjee, S., Carlin, B., and Gelfand, A. (2003). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Barrera-Osorio, F., Bertrand, M., Linden, L. L., and Perez-Calle, F. (2011). Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia. *American Economic Journal: Applied Economics*, 3(2):167–95.
- Barrios, T., Diamond, R., Imbens, G. W., and Kolesár, M. (2012). Clustering, Spatial Correlations, and Randomization Inference. *Journal of the American Statistical Association*, 107(498):578–591.

- Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics*, 46(1):51 – 84.
- Bobba, M. and Gignoux, J. (2014). Policy Evaluation in the Presence of Spatial Externalities: Reassessing the Progresa Program. PSE Working Papers halshs-00646590, HAL.
- Bobonis, G. J. (2009). Is the Allocation of Resources within the Household Efficient? New Evidence from a Randomized Experiment. *Journal of Political Economy*, 117(3):453–503.
- Bobonis, G. J. and Finan, F. (2009). Neighborhood Peer Effects in Secondary School Enrollment Decisions. *The Review of Economics and Statistics*, 91(4):695–716.
- Cameron, A. C. and Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.
- Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48(2):424–55.
- Duflo, E., Dupas, P., and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5):1739–74.
- Duflo, E. and Saez, E. (2003). The Role Of Information And Social Interactions In Retirement Plan Decisions: Evidence From A Randomized Experiment. *The Quarterly Journal of Economics*, 118(3):815–842.
- Ferreira, F. H. G. and Schady, N. (2009). Aggregate economic shocks, child schooling, and child health. *The World Bank Research Observer*, 24(2):147–181.
- Gertler, P. (2004). Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA’s Control Randomized Experiment. *American Economic Review*, 94(2):336–341.
- Glewwe, P., Ilias, N., and Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–27.
- Hudgens, M. G. and Halloran, M. E. (2008). Towards Causal Inference with Interference. *Journal of the American Statistical Association*, 103(482):832–842.
- Kremer, M., Leino, J., Miguel, E., and Zwane, A. P. (2011). Spring cleaning: Rural water impacts, valuation, and property rights institutions. *The Quarterly Journal of Economics*, 126(1):145–205.
- Kremer, M. and Miguel, E. (2007). The Illusion of Sustainability. *The Quarterly Journal of Economics*, 122(3):1007–1065.

- Lalive, R. and Cattaneo, M. A. (2009). Social Interactions and Schooling Decisions. *The Review of Economics and Statistics*, 91(3):457–477.
- Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–42.
- Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1):159–217.
- Moran, P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243 – 251.
- Morgan, S. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Analytical Methods for Social Research. Cambridge University Press.
- Muralidharan, K. and Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1):39 – 77.
- Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, 115:200–249.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition.
- Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., and Artha, R. P. (2014). Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia. *American Economic Journal: Applied Economics*, 6(2):105–26.
- Ravallion, M. (2009). Should the Randomistas Rule? *The Economists' Voice*, 6(2):1–5.
- Rodrik, D. (2008). The New Development Economics: We Shall Experiment, but How Shall We Learn? Working Paper Series rwp08-055, Harvard University, John F. Kennedy School of Government.
- Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomised and Non-Randomised Studies. *Journal of Educational Psychology*, 66:688–701.
- Sacerdote, B. (2011). *Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?*, volume 3 of *Handbook of the Economics of Education*, chapter 4, pages 249–277. Elsevier.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:pp. 234–240.
- Ward, M. and Gleditsch, K. (2008). *Spatial Regression Models*. Number no. 155 in Quantitative Applications in the Social Sciences. SAGE Publications.

Table 1  
Survey of RCT Studies

Total Number of RCTs	86
<i>Randomization method</i>	
Individual-level	55
Cluster-level (regions, schools, villages)	31
<i>Published in:</i>	
AER	30
AEJ: Policy Economics	7
AEJ: Applied Economics	17
QJE	21
JPE	6
REStud	5
<i>Setting</i>	
Developing countries	42
Mexico	8
Kenya	8
India	7
Developed countries	44
USA	37
<i>Main outcome</i>	
Education	34
Consumer Behavior	16
Health	7
Micro-credit	5
Insurance	4
Investment	4

Notes: Data come from a survey of articles in six journals between 2000 and 2014: the American Economic Review, the American Economic Journal: Applied Economics, the American Economic Journal: Policy Economics, the Quarterly Journal of Economics, the Journal of Political Economy, and the Review of Economic Studies. We define an article as a randomized controlled trial if it estimates the impact of a randomly allocated program or intervention on one or more socioeconomic outcomes.

Table 2  
Spatial Correlation in Progresa at Baseline: Moran's  $I$

Mean	1 nearest neighbor		2 nearest neighbors		5 nearest neighbors	
	Moran's $I$	p-value	Moran's $I$	p-value	Moran's $I$	p-value
Treatment Assignment	0.632	-0.024	0.693	-0.020	0.653	-0.011
Primary Enrollment	0.965	0.107	0.053	0.141	0.000	0.111
Secondary Enrollment	0.485	0.281	0.000	0.290	0.000	0.236
Labor supply (All)	0.546	0.350	0.000	0.345	0.000	0.295
Labor supply (Men)	0.905	0.143	0.010	0.168	0.000	0.161
Labor supply (Women)	0.184	0.377	0.000	0.397	0.000	0.367
Per capita income	300.68	0.266	0.000	0.269	0.000	0.265
Per capita consumption	241.37	0.244	0.000	0.275	0.000	0.238

Source: Authors' calculations from Progresa baseline data aggregated at village-level.

Notes: Moran's  $I$  is estimated as described in the text. Reported p-values are obtained from conducting a two-tailed test of the null hypothesis,  $H_0 : \mathbb{E}(I) = -1/(N - 1)$ .

Table 3  
Robust LM tests for spatial dependence in Progresa outcomes

	1 nearest neighbor		2 nearest neighbors		5 nearest neighbors	
	Lag	Error	Lag	Error	Lag	Error
Primary Enrollment	0.10	0.31	0.01	3.14	15.46	10.89
Secondary Enrollment	0.45	2.54	6.06	0.54	15.92	4.42
Labor supply (All)	3.96	9.33	1.33	2.43	4.69	5.49
Labor supply (Men)	0.81	1.62	0.13	2.02	0.56	9.32
Labor supply (Women)	0.19	0.08	5.90	0.03	17.58	1.17
Household income	11.05	4.26	8.75	2.67	9.09	0.26
Household consumption	0.19	3.67	0.04	8.15	0.73	6.51

Source: Authors' calculations from Progresa baseline data aggregated at village-level.

Notes: These tests are two-directional, distributed as  $\chi^2_2$  with critical levels of 5.99 ( $p = 0.05$ ) and 9.21 ( $p = 0.01$ ). See [Anselin et al. \(1996\)](#) for details. The reported tests are obtained from the following regression:  $y_j = \alpha + \beta T_j + \theta X_j + u_j$ . Covariates in  $X_j$  include: average village household size, fraction of male headed households, average years of schooling of the household head, average number of children (0-17 years), average number of adults (+18 years), number of adult workers, fraction of dwellings owned, fraction of precarious dwellings, and the fraction households that have: drinking water, a toilet, sewer access, and electricity.

Table 4  
Village-level program effect estimates for Progresa

		DD	Spatial Lag	Spatial Error
Primary enrollment ( $V = 505$ )	$\hat{\beta}$	0.009 (0.004)**		0.010 (0.005)*
	$\hat{\lambda}$			0.158 (0.033)***
Secondary enrollment ( $V = 505$ )	$\hat{\beta}$	0.029 (0.014)**	0.036 (0.023)	
	$\hat{\rho}$		0.280 (0.029)***	
Labor supply (All) ( $V = 506$ )	$\hat{\beta}$	-0.009 (0.007)		-0.009 (0.008)
	$\hat{\lambda}$			0.351 (0.029)***
Labor supply (Men) ( $V = 506$ )	$\hat{\beta}$	-0.003 (0.008)		-0.007 (0.008)
	$\hat{\lambda}$			0.200 (0.032)***
Labor supply (Women) ( $V = 506$ )	$\hat{\beta}$	-0.012 (0.012)	-0.008 (0.013)	
	$\hat{\rho}$		0.355 (0.027)***	
Per capita income ( $V = 506$ )	$\hat{\beta}$	0.001 (0.039)	0.006 (0.045)	
	$\hat{\rho}$		0.300 (0.028)***	
Per capita consumption ( $V = 505$ )	$\hat{\beta}$	0.050 (0.025)**		0.050 (0.032)
	$\hat{\lambda}$			0.372 (0.027)***

Source: Authors' calculations from Progresa data aggregated at village-level.

Significance levels: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Notes: Reported coefficients are obtained from difference-in-difference regressions with clustered standard errors at the village-level. Covariates include: average village household size, fraction of male headed households, average years of schooling of the household head, average number of children (0-17 years), and average number of adults (+18 years).

Table 5  
Simulation results: Spatially lagged outcome

	$\rho$					
	0.00	0.10	0.25	0.50	0.75	0.90
<i>1 nearest neighbor</i>						
$\hat{\beta}$	0.096 (0.286)	0.095 (0.289)	0.104 (0.307)	0.117 (0.396)	0.175 (0.749)	0.389 (1.908)
% Bias $\hat{\beta}$	-4.28	-5.42	3.73	17.41	74.92	289.23
Rejects $H_0 : \hat{\beta} = 0$	6.58	6.46	6.50	5.76	5.72	5.52
<i>2 nearest neighbors</i>						
$\hat{\beta}$	0.096 (0.286)	0.094 (0.287)	0.102 (0.297)	0.110 (0.346)	0.142 (0.544)	0.234 (1.231)
% Bias $\hat{\beta}$	-4.28	-5.79	2.46	9.75	42.16	133.60
Rejects $H_0 : \hat{\beta} = 0$	6.58	6.54	6.34	6.26	5.72	5.94
<i>5 nearest neighbors</i>						
$\hat{\beta}$	0.096 (0.286)	0.094 (0.286)	0.101 (0.290)	0.105 (0.313)	0.113 (0.402)	0.153 (0.662)
% Bias $\hat{\beta}$	-4.28	-5.61	1.35	4.98	12.92	53.36
Rejects $H_0 : \hat{\beta} = 0$	6.58	6.42	6.88	6.22	6.08	5.36

Source: Authors' calculations from 5,000 Monte Carlo Simulations.

Reported values are average estimates from the total number of replications.  $\hat{\beta}$  is estimated as in Equation (6). Clustered standard errors at village-level in parentheses. The percentage of bias is calculated as the difference between the average estimate and the true value  $([\hat{\beta} - \beta]/\beta) \times 100$ . The rejection rate measures the percentage that the estimation strategy finds a significant effect when it should. A higher value represents good accuracy and lower percentages denote a large number of false negatives.

Table 6  
Simulation results: Spatially correlated unobservable

	$\lambda$					
	0.00	0.10	0.25	0.50	0.75	0.90
<i>1 nearest neighbor</i>						
$\hat{\beta}$	0.100 (0.028)	0.100 (0.04)	0.100 (0.076)	0.100 (0.143)	0.102 (0.213)	0.103 (0.255)
% Bias $\hat{\beta}$	0.02	0.24	-0.37	-0.02	2.22	3.09
Rejects $H_0 : \hat{\beta} = 0$	94.56	71.18	26.68	10.82	8.08	7.04
<i>2 nearest neighbors</i>						
$\hat{\beta}$	0.100 (0.028)	0.100 (0.035)	0.100 (0.057)	0.099 (0.103)	0.101 (0.151)	0.100 (0.181)
% Bias $\hat{\beta}$	0.02	0.08	-0.03	-0.54	1.43	0.12
Rejects $H_0 : \hat{\beta} = 0$	94.56	82.72	42.30	16.06	9.98	9.00
<i>5 nearest neighbors</i>						
$\hat{\beta}$	0.100 (0.028)	0.100 (0.031)	0.100 (0.042)	0.101 (0.068)	0.099 (0.096)	0.101 (0.114)
% Bias $\hat{\beta}$	0.02	0.37	0.13	0.72	-1.19	0.62
Rejects $H_0 : \hat{\beta} = 0$	94.56	90.18	66.24	31.74	17.92	15.04

Source: Authors' calculations from 5,000 Monte Carlo Simulations.

Reported values are average estimates from the total number of replications.  $\hat{\beta}$  is estimated as in Equation (9). Clustered standard errors at village-level in parentheses. The percentage of bias is calculated as the difference between the average estimate and the true value  $([\hat{\beta} - \beta]/\beta) \times 100$ . The rejection rate measures the percentage that the estimation strategy finds a significant effect when it should. A higher value represents good accuracy and lower percentages denote a large number of false negatives.

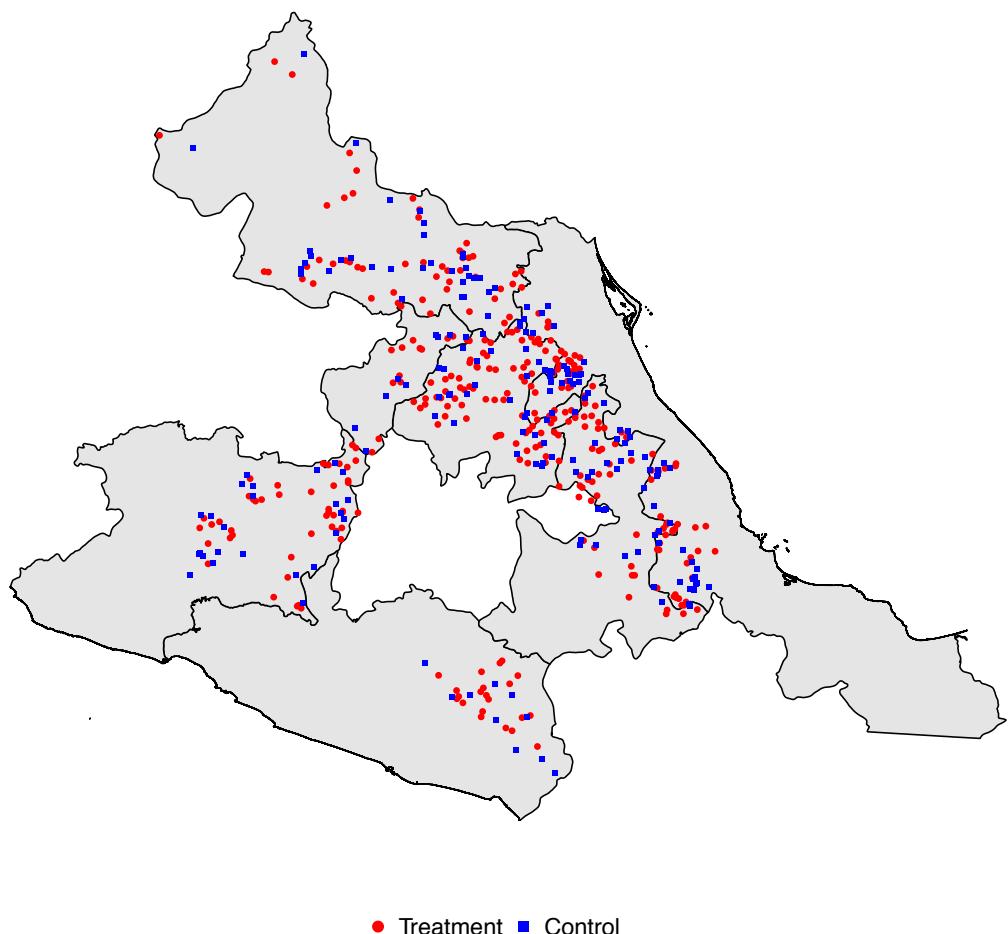
Figure 1  
Location of Progresa villages in Mexico

33



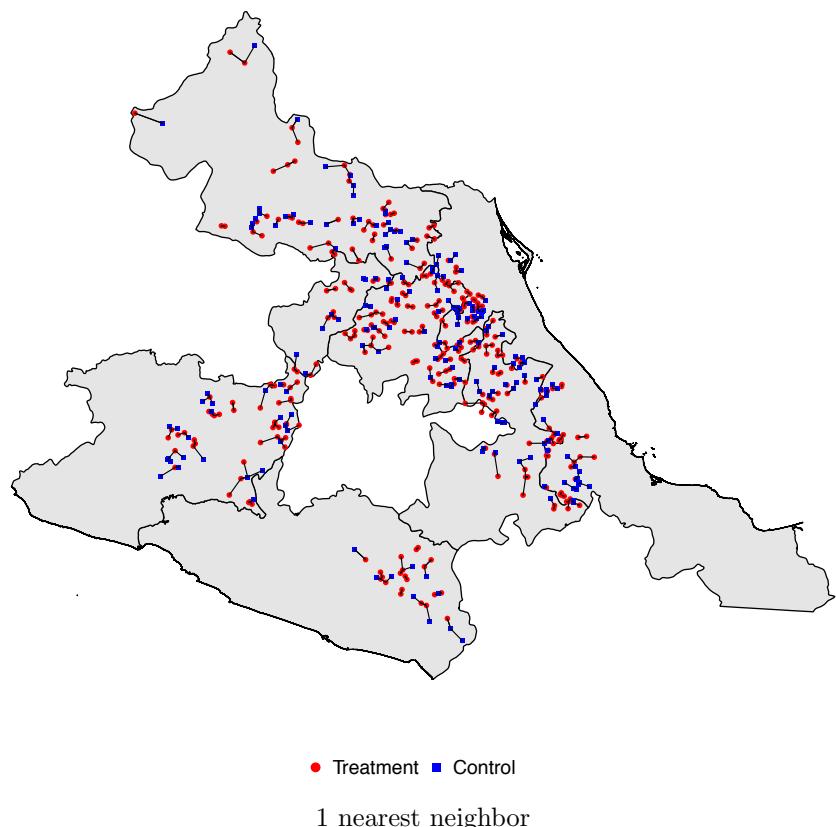
Source: Own elaboration from official shapefiles from the Instituto Nacional de Estadística, Geografía e Informática (INEGI). Progresa villages were identified by using the state, municipality, and village identifiers in the official program data downloaded from [SEDESOL](#).

Figure 2  
Close-up of Progresa villages by treatment assignment



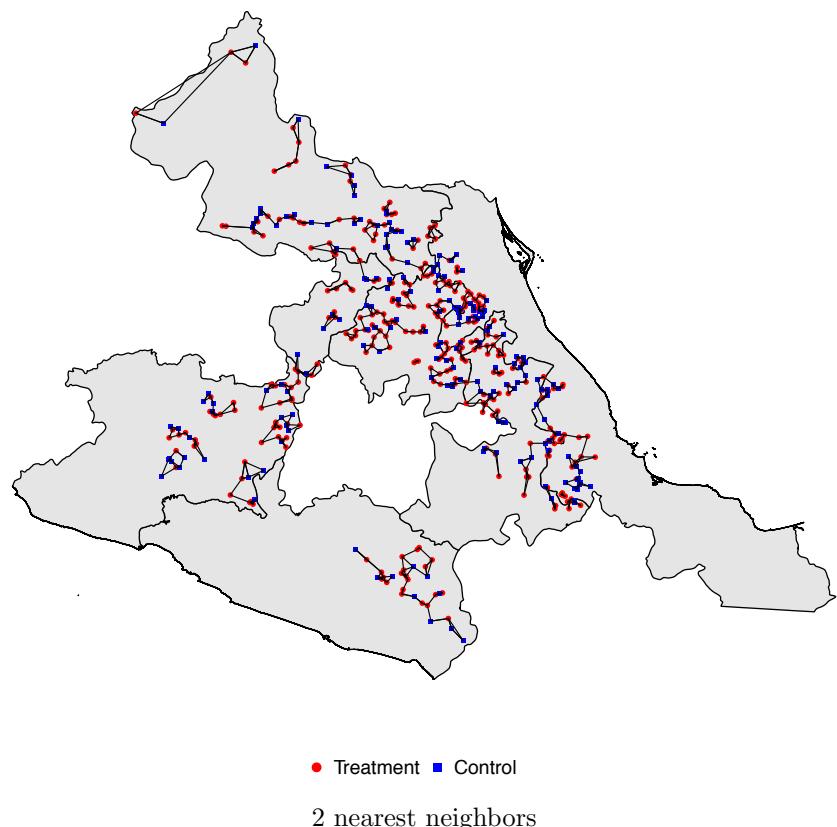
Source: Own elaboration from official shapefiles from the Instituto Nacional de Estadística, Geografía e Informática (INEGI). Progresa villages were identified by using the state, municipality, and village identifiers in the official program data downloaded from [SEDESOL](#).

Figure 3  
Spatial Networks in Progresa



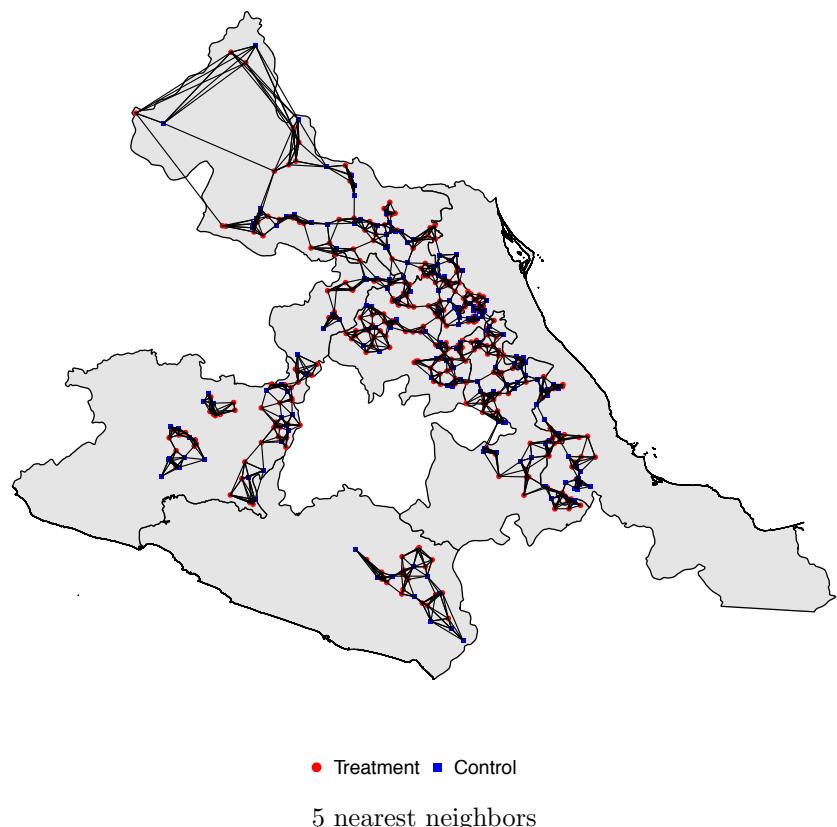
Source: Own elaboration from official shapefiles from the Instituto Nacional de Estadística, Geografía e Informática (INEGI). Progresa villages were identified by using the state, municipality, and village identifiers in the official program data downloaded from [SEDESOL](#). The black lines connect a village to its nearest neighbor.

Figure 3  
Spatial Networks in Progresa (continued)



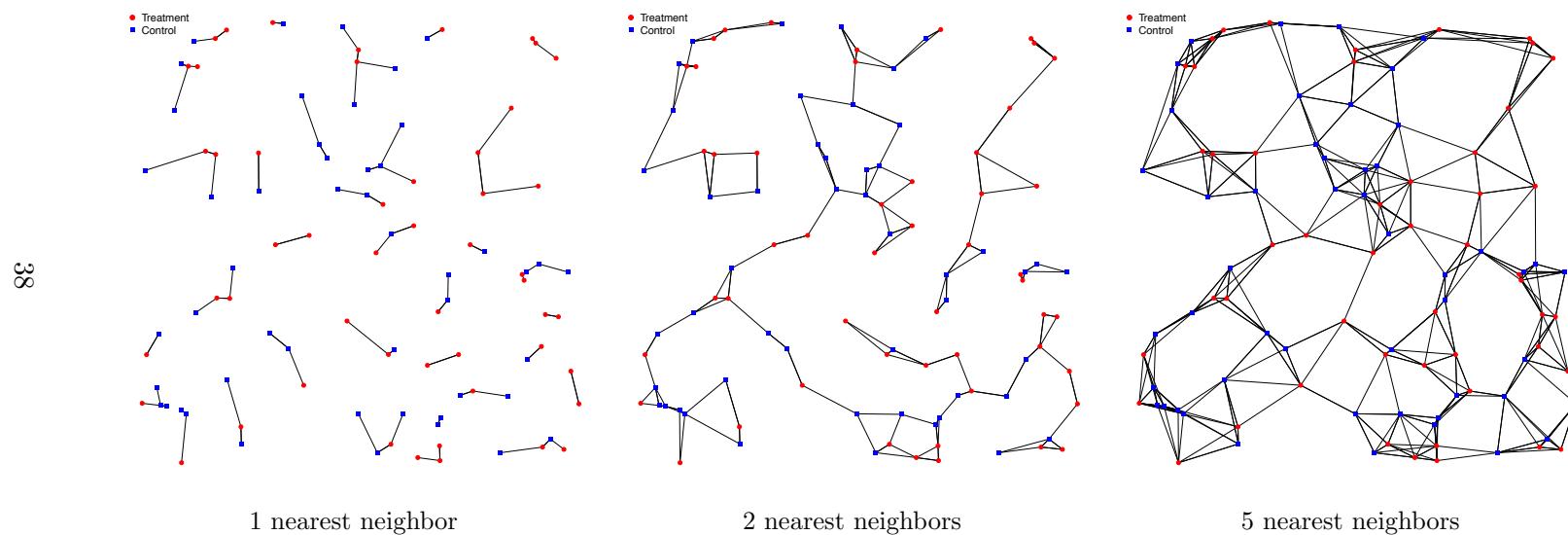
Source: Own elaboration from official shapefiles from the Instituto Nacional de Estadística, Geografía e Informática (INEGI). Progresa villages were identified by using the state, municipality, and village identifiers in the official program data downloaded from [SEDESOL](#). The black lines connect a village to its two nearest neighbors.

Figure 3  
Spatial Networks in Progresa (continued)



Source: Own elaboration from official shapefiles from the Instituto Nacional de Estadística, Geografía e Informática (INEGI). Progresa villages were identified by using the state, municipality, and village identifiers in the official program data downloaded from [SEDESOL](#). The black lines connect a village to its five nearest neighbors.

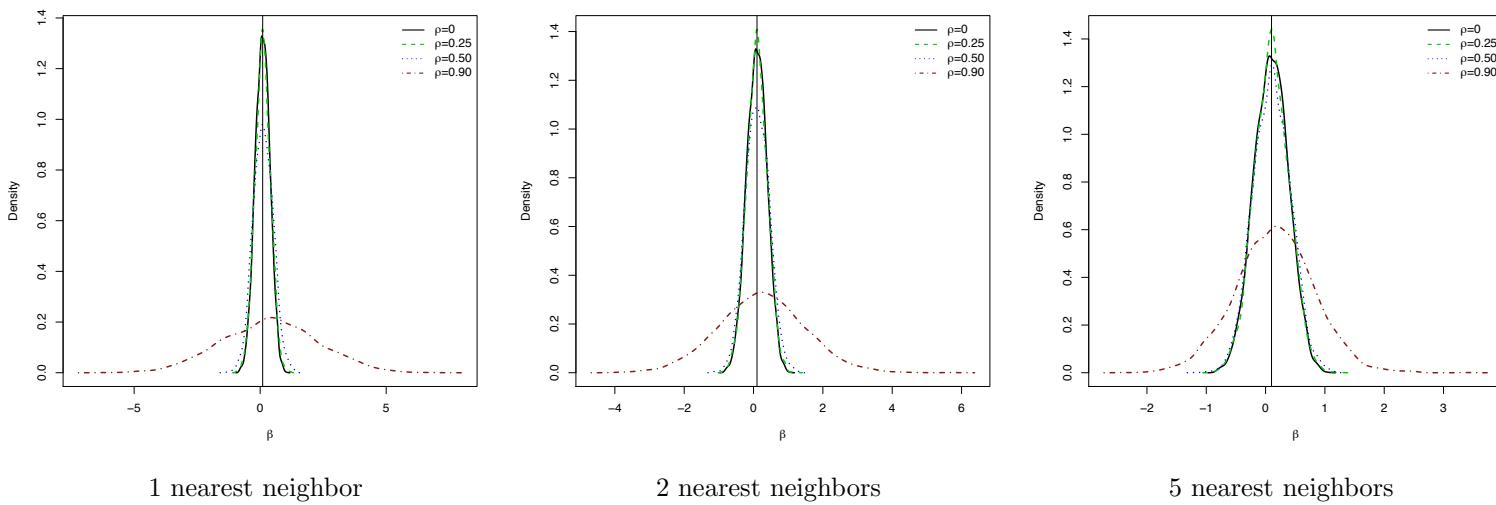
Figure 4  
Simulated Spatial Networks,  $V = 100$



The graph shows a random draw of  $V = 100$  villages randomly assigned into treatment and control status as described in Section 4.2 and the assumed links with their nearest neighbor(s). The interactions between simulated villages are not symmetric.

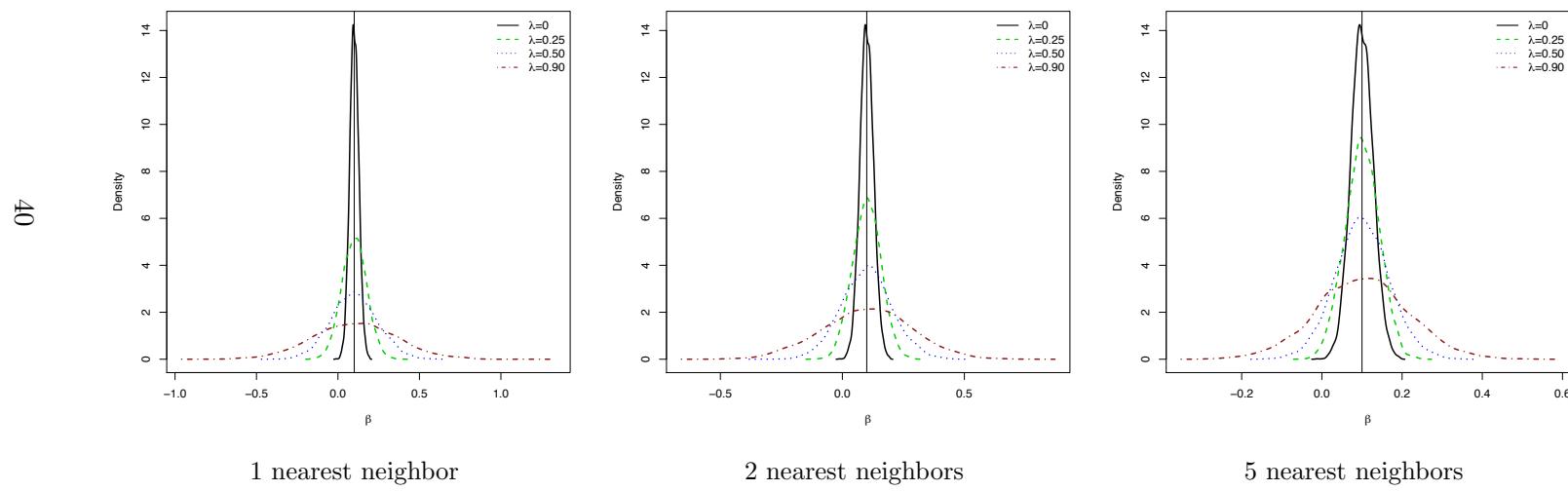
68

Figure 5  
Simulated treatment effect distributions: Spatially correlated outcome



Source: Authors' calculations from 5,000 Monte Carlo Simulations.

Figure 6  
Simulated treatment effect distributions: Spatially correlated unobservable



Source: Authors' calculations from 5,000 Monte Carlo Simulations.