Scaling-up Effects in Experimental Policy Evaluations. The case of *Progresa* in Mexico*

Matteo Bobba†

November 2008

Preliminary Version. Please do not quote.

Abstract

The scaling-up process of social experiments is likely to generate interaction effects among units that would challenge the validity of the experimental results. In this paper, we attempt to evaluate the extent to which the potential indirect effects associated with the geographical expansion of an educational and health policy in a developing country may affect behavioral responses to the policy itself. We do so in a reduced-form framework by exploiting the experimental design and the concomitant nation-wide expansion of the *Progresa* program in Mexico. More specifically, we use variations in the proportion of treated localities across regions generated by the uneven process of geographical expansion of the program in order to estimate average treatment effects at different degrees of within-region treatment intensity. Results show that average behavioral responses to the program appear to be systematically affected by the local geographical dimension of the treatment.

*I thank Francois Bourguignon and Marc Gurgand, for many comments and insightful conversations during the preparation of this draft. I have also benefited from helpful comments by Vincenzo Di Maro and Marc Ferracci. In addition, I acknowledge the help of Vera Chiodi, Gabriela Ormenio Vasquez and Raul Perez for answering many questions on the Progresa/Oportunidades database. All remaining errors are my own.

†Paris School of Economics (PSE), 48 bd Jourdan, 75014, Paris, France. E-mail: bobba@pse.ens.fr.

1 Introduction

In recent years, social experiments have been strongly advocated as a tool for evaluating a wide range of public policies in both developed and developing countries. In such a context, small subsets of the population are randomly selected and assigned to treatment and control groups and the resulting sample-mean differences in outcomes are considered to draw inferences on the effectiveness of those policies.

In particular, the last 10 years have witnessed a spread of randomized evaluations in development economics. Topics are as diverse as the effects on school inputs on learning (Glewwe and Kremer [2006]), the adoption of new technologies in agriculture (Duflo et al. [2006]), corruption in driving licenses administration (Bertrand et al. [2006]) or moral hazard and adverse selection in consumer credit market (Karlan and Zinman [2005])[1].

Compared to non-experimental and quasi-experimental methodologies, social experiments have two important advantages. First, the experimental design allows to solve the inherent selection bias problem by rendering the estimation of the counterfactual straightforward. Second, by relying on small random samples and hence on policies that are small in scope, internal validity is assured since the crucial assumption of no interaction among units (SUTVA) is likely to be satisfied[2]. However, such estimates may be of limited usefulness if the policy experimented on a small number of individuals for evaluation purposes is implemented for the whole eligible population in the economy.

In principle one could consider two types of extrapolation or scaling-up. First, one might want to predict the effect of exporting an existing program from a population where its effects were evaluated (evaluation population) to a different population (implementation population). Attanasio et al (2004) consider both the conceptual and technical issues involved in this type of exercise. They employ the Progresa program and its experimental evaluation design in order to compare extrapolation results based on a structural model with the actual ex-post evaluation.
A second dimension of scaling-up consists in the increase in the number of treated units within the given area or region from which the random evaluation sample is randomly drawn. In such a type of setting, the raise in the density of treated units that are geographically close with each other is likely to generate interaction effects,

[1]Comprehensive reviews include Kremer [2003] for randomized evaluations in education; Banerjee and Duflo [2006] on ways to improve teacher's and nurse's attendance in developing countries and Duflo [2005] on the lessons on incentives, social learning, and hyperbolic discounting.

[2]More precisely, average outcomes of the non-treated individuals are not likely to be affected by the policy implementation:

$$E\left[y_i^0|D_i=0\right]=E\left[y_i^0\right]$$

where y_i^0 is a given outcome of interest for non-treated individuals and D_i is the treatment indicator. This condition is a direct implication of the *Stable-Unit-Treatment-Value Assumption* (SUTVA), which is in turn the crucial identification assumption embedded in the standard micro-econometric framework for policy evaluation. For a discussion on SUTVA and its implication see Rubin [1974], Rubin [1986] and Heckman and Smith [1998] among others.

either through prices or in the form of spillovers/externalities, that may be in turn related to program outcomes. In the context of a deworming program in Kenya, Miguel and Kremer [2004] found that deworming interferes with disease transmission and thus reduces worm infection among both children in the program schools who did not receive the medicine and children in neighboring schools. Angelucci and De-Giorgi [2007] find evidence of within-villages (positive) effects of Progresa benefits on non-eligible households consumption through insurance and credit markets. Gignoux (2008) studies spillover effects across villages in the Progresa experiment by comparing the rates of enrollment of children living in control localities with access to secondary school in beneficiary localities with those of children in other control localities and finds evidence of negative spillovers on the progression into the first secondary grade, suggesting some constraints in the supply of secondary education.

One potential solution to the above mentioned concerns is to redefine the unit of interest. If the interactions between individuals are at an intermediate level, say a village, a local labor market, or a classroom, rather than global, one can analyze the data using the local labor market or classroom as the unit and changing the no-interaction assumption to require the absence of interactions among local labor markets or classrooms. Such aggregation is likely to make the no-interaction assumption more plausible, albeit at the expense of reduced precision. For the case of evaluating training programs for unemployed workers, a recent paper by Ferracci et al. [2008] points to this direction by suggesting an identification and estimation method that distinguishes between treatment effects at the individual and at the group (the local labor market) level. Their estimation results based on observational data in a non-experimental setting show that the outcomes of both treated and non treated workers depend on the proportion of treated in a local labor market.

Other studies find little if no evidence of the potential scaling-up effects of program/policies previously implemented on a small-scale. Blundell et al. [2004] evaluate the effects of a major targeted active labor market program in the UK. The program was piloted in certain areas before it was rolled out nation-wide. The authors did not find much differences in the employment responses between the two program implementation stages. However, for identification reasons, the control group used for evaluation purposes changes in the two phases, therefore it is not possible to evaluate whether area-specific general equilibrium effects due to the changes in the wage pressure from the increased supply of workers and job subtitution (spillover) effects between treated and non-treated are not very strong or, alternatively, whether they compensate with each others.

The scaling-up of policies evaluated in an experimental setting is likely to be a non-neutral phenomenon from a pure economic point of view, regardless the statistical validity of the experimental results. In this paper, we attempt to evaluate the extent to which the potential indirect effects associated with the geographical expansion of an educational and health policy in a developing country may affect behavioral responses to the policy itself. We do so in a pure reduced-form framework by exploiting the experimental design and the concomitant nation-wide expansion of the *Progresa* program in Mexico. More specifically, we estimate average treatment effects of localities situated in different economic environments generated by the uneven process of geographical expansion of the program across regions in order to evaluate the potential scaling-up (macro) implications of such a (micro) policy. Results show that average behavioral responses to the program appear to be systematically affected by the local geographical dimension of the treatment.

This paper is organized as follows. Section 2 describes the Progresa program and depicts the data used for our analysis. Section 4 deals with the identification strategy and its empirical implementation; while Section 5 presents the main results. Section 6 concludes.

2 Scaling-up Patterns within an Experimental Setting

2.1 The Progresa Program and its Nation-wide Expansion

Launched in Mexico in 1997, Progresa is a program whose main aim is to improve the process of human capital accumulation in the poorest communities by providing households with conditional cash transfers (CCT) targeted on specific types of behavior in three key areas: nutrition, health and education. The program grants were initially promised for only 3 years, since the election in the fall of 2000 would lead to a change in government which might decide to change the program. For the educational component, grants were provided for children less than 17 years old of age enrolled in school between third grade of primary and the third grade (grade 9) of secondary school[3].

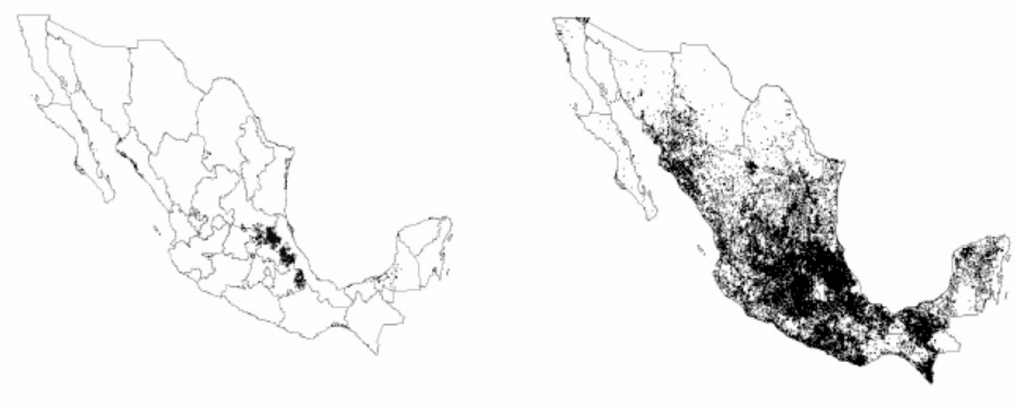
Progresa was noticeable and remarkable not only for its original design that has lately germinated in several other countries in Latin America and elsewhere but also because, when the program begun, the Mexican government started a rigorous evaluation of its effects. The transfers have been initially randomly assigned to a small sample of 506 rural localities (villages) belonging to 191 municipalities in the seven Mexican states of Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz[4] and the effectiveness of the policy has been evaluated in accomplishing with the main educational and health outcomes of the program.

[3]In order to provide incentives for human-capital accumulation, benefits are contingent on fulfillment of certain obligations by the beneficiary families. Grants are linked to school attendance of children: if a child unjustifiably misses more than 15 percent of school days in a month, the family will not receive the grant that month. All of the benefits are given directly to the mother of the family, with a maximum monthly limit of 750 pesos per family. Average monthly benefits are currently 255 pesos, equivalent to about 22 percent of the monthly income of beneficiary families. After three years, families may renew their status as beneficiaries, subject to a reevaluation of their socioeconomic conditions. For more extensive and detailed descriptions of the program and the experimental results see Skoufias (2001), Coady (2004) Schultz (2004) among others.

[4]This implies that the average number localities per municipality within this random sample is 2.7. Whereas the average number of program incorporated (i.e. treated) localities per municipality is 42.6 in 1997, 64.4 in 1998 and 76.7 in 1999. This in turns suggests that in the evaluation sample we observe on average roughly 4-6% of the whole eligible population of localities in each municipality.

During this experimental evaluation phase, the program has been progressively expanding in its geographical scope at the national level. Figure (1) below depicts the dramatic increase in the beneficiary localities during the period of the experimental evaluation.

Figure 1: Proliferation of Treated Villages. Aug 1997-Sep 1999



The program begun in 1997 where the first 2,390 localities situated in 154 municipalities in the 7 states of Campeche, Coahuila, Guanajuato, Hidalgo, Queretaro, San Luis Potosì and Veracruz were incorporated. The greatest expansion occurred in 1998 when nearly 1.3 million families in 30,192 localities were incorporated. By the end of 1999 the program included 46,233 localities in all 31 states. This constitutes almost 60 percent of all rural localities in Mexico[5]. From 2000 onwards, the program has carried on with the new Administration under the name of Oportunidades and continued its expansion until it reached the whole eligible rural Mexican population by year 2006.

2.2 Program-driven Demand for Schooling

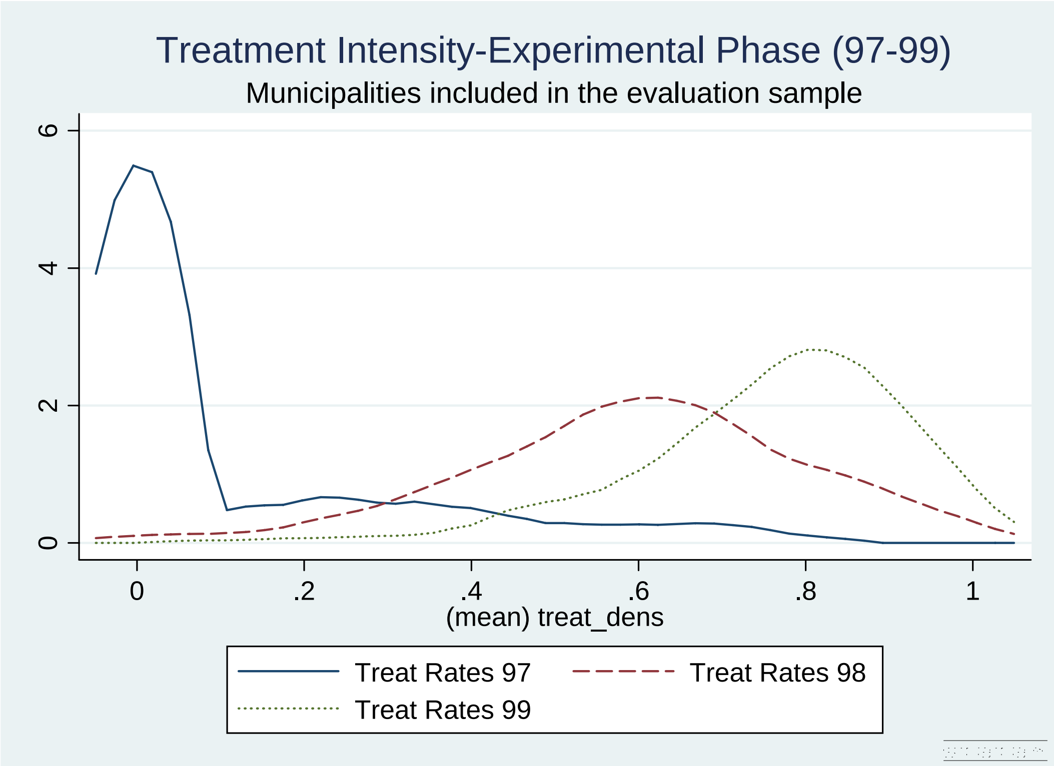
In this paper, we study the potential implications of the scaling-up process outlined above on the experimental evaluation of the program. We therefore restrict our attention to the 191 municipalities from which the random sample of the experimental evaluation has been drawn. In other words, we focus on the scaling-up *'within'* the regions that participated to the social experiment and hence we abstract from potential differences in the effects of the program in the case it is implemented in a different population[6] (i.e. scaling-up *'across'* regions). Figure (2) displays the empirical distributions of the extent of geographical penetration of the program across municipalities over the period 1997-1999.

The evolution of the empirical densities over time highlights two important features of the scaling-up of the program. First, average treatment intensity at the municipality level has been increasing more sharply in the regions from which the

[5]We thank the Oportunidades Staff for making these data available to us.

[6]Appendix A.1 displays the evolution in the proportion of treated localities during all program years and decompose the overall scaling-up process into a 'within' and 'between' dimension. The former has witnessed a steeper increase of treated localities during the initial experimental phase.

Figure 2: Scaling-up 'Within' the Evaluation Sample



random sample has been drawn in comparison with the national data: from 10% of treated localities in 1997 to 76% in 1999. Since the effect of the school subsidy is both to decrease the price of schooling and increase the familys income, to the extent that schooling is a normal consumption good for which demand increases with income, or income relaxes a credit constraint that allows the poor family to invest more in the schooling of their child, both the income effect and the income-compensated price of schooling effect of the school subsidy will increase the households demand for schooling. This dramatic increase in program recipients in such a short period of time is then likely to have key implications for the demand of both primary and secondary schooling in these areas.
Second, there seems to be a great deal of variation in treatment intensity across municipality or, in other words, the process of incorporation of program-eligible localities has been to a large extent heterogeneous across the regions that have participated to the experimental evaluation. As this source of variation will turn out to be crucial in the analysis that follows, we will come back to this issue in greater detail in section 3.2.

2.3 Supply of Schools

Concomitant with the monetary transfers of Progresa, there is also a supply-side component of the program with resources allocated toward improving school quality (e.g., more teachers, health clinic staff, higher salaries) and access (e.g. construction

of new schools).
We have access to data on school characteristics from the Secretary of Public Education (SEP), which collects every year information on all schools nationwide and therefore we are to able to track the evolution of relevant school supply characteristics over the period 1997-1999 for the 191 municipalities from which the experimental evaluation has been drawn.
Two important features of the geographical distribution of basic education in rural Mexico are worth mentioning. First, the great majority of the rural localities has a primary school whereas a much lower proportion (only 18% in our random sample) has a secondary school. Second, inter-municipality migration is almost absent as 98% of secondary school children declare to attend a school within the same municipality. Taken jointly, these stylized facts provide a straightforward rationale for a negative externalities mechanism across localities induced by the scaling-up of the program in the form of crowding-out of secondary schools in common within the same municipality. We will therefore focus our analysis on the secondary school educational component of the program and consider the potential interrelelationships between villages in the same municipalities.

Table 1 below depicts sample means, standard deviations and percentages changes over the period 1997-1999 of several secondary school supply indicators at municipality-level. Consistent with the increasing intensity of the program in those areas, we observe a clear average growing pattern of secondary school enrollment which is partly compensated by an increase in the number of secondary schools per municipality. This is confirmed by the fact that both student-teacher and student-classrooms ratios have been gradually increasing over the period. Interestingly, indicators of school quality such as educational attainment of teachers and the percentage of students reported as failing between one and five classes during the previous year appear pretty much stable, whereas the number of classroom with more than one grade has been declining steadily over the period. All in all, the general picture is one of increasing demand being only partly compensated for by matching supply-side resources. Nevertheless, there appear to exist sustantial differences across municipalities in those broad trends as confirmed by the large standard deviations of each supply-side indicator in our sample.

TABLE 1. Summary of Supply-Data at Municipality-level, 1997-1999 (Means and Standard Deviations)

	1997	1998	1999	% change ('99-'97)
School Enrollment	109.68 (61.14)	114.88 (62.17)	118.09 (64.31)	7.67%
# of Sec Schools	13.08 (12.55)	13.66 (12.56)	14.06 (12.79)	7.53%
Student-Teacher Ratio	21.81 (5.15)	23.35 (5.97)	23.37 (5.55)	7.17%
Student-Classroom Ratio	23.89 (5.49)	25.99 (6.04)	26.20 (6.60)	9.67%
% Teachers with Higher Educ	0.22 (0.21)	0.20 (0.19)	0.20 (0.19)	-6.38%
% Student Failing	0.05 (0.06)	0.05 (0.05)	0.05 (0.05)	-5.64%
Multiple Classroom	0.30 (0.40)	0.21 (0.30)	0.20 (0.25)	-33.06%

Our objective is to evaluate the potential spillovers/externalities induced by the scaling-up of the program within the same geographical areas in which the experimental evaluation took place. As both randomization strategy and program expansion have been conducted at the locality (village) level, we focus on locality average educational outcomes and conduct our analysis at the village - rather than at the individual - level[7]. To this aim, we combine the two administrative database described above with the random sample employed for the experimental evaluation of the program.

2.4 The Experimental Evaluation Sample

The experimental evaluation sample of Progresa consists of socio-economic characteristics at the individual level collected for a random sample of 506 localities - among the whole population of rural localities that met the inclusion criteria for the Program - which participated to the social experiment from October 1997 until the end of 1999. Within this group, 320 locality was randomly allocated to the treatment group and 186 localities to the control group. The ENCASEH 97 survey represents the baseline or original sample. This has been followed by Household Evaluation Surveys (ENCEL) collected every 6 months for 2 years (1998-1999) for a total of five time periods including the baseline. However, for some outcomes of interest, we only have information for two out of four post-program waves - namely October 1998 and November 1999. This allows building a panel of three subsequent surveys spaced by periods of roughly one year each of roughly $24,000$ households.

[7]This would also take into account potential spillovers within villages induced by the program . See Angelucci and De Giorgi (2007).

Some information was also collected at the level of localities and provides their distance to the nearest school[8], the index of marginalization, access to public services, infrastructures and local labor market.
Our primary interest lies in locality-level educational outcomes for children exposed to junior secondary school (grades 7-9) attendance. We therefore build population-weighted locality averages for enrollment rates, attendance rates, years of education and labor supply for children aged less than 17 years old and who have been completed 6 to 8 years of education in each year during the period 1997-1999[9]This sample restriction implies loosing 5 localities (3 treated and 2 controls) in 1997, 3 localities (2 treated and 1 control) in 1998 and 9 localities (7 treated and 2 controls) in 1999. Because of random treatment assignment, we believe such attrition bias can be assumed to be negligible and not affecting our results..

3 Some Descriptive Trends

3.1 Randomization Strategy

Behrman and Todd [1999] formally test whether the distributions for all the individual and households-level socio-economic characteristics for treated and control localities are statistically different prior to the administration of the program (i.e. 1997) in order to determine whether the two groups truly appear to have been randomly assigned. They conclude that while distributions of various characteristics are not different on locality means, there appear to be some differences on household level[10]. In our setting in which localities are the units of interest, we shall further test whether the distribution of both the locality and municipality level observable characteristics that are potentially related to outcomes are balanced between the two groups. Table 1 below provides means, mean differences and the relative test of equality for all the outcomes and the covariates we employ in our analysis observed during the pre-program period (1997). Beyond very few exceptions[11], the great majority of the variables do not appear, on average, to be different across treated and control localities. This suggests that treatment assignment at the locality level can be considered as random.

[8]Distance data have been constructed by Coady and Parker (2004) using a Geographic Information System (GIS) based on a national census of schools. We thank Jeremie Gignoux for giving us access to this database

[9]We focus on years of education attained rather than age-specific restrictions since, as noticed by Shultz (2004), in contrast with high-income countries variation in age appears to be less important for explaining enrollments than years of schooling completed.

[10]This suggests the use of a Difference-in-Difference (DD) estimator that takes into account preexisting program differences across groups when dealing with household level data.

[11]Namely, Child Labor, enrollment in schools and the percentage of student failing in the previous academic year.

TABLE 2. Covariate-balance Test Across Treated and Control Localities.
Pre-program (1997) Data

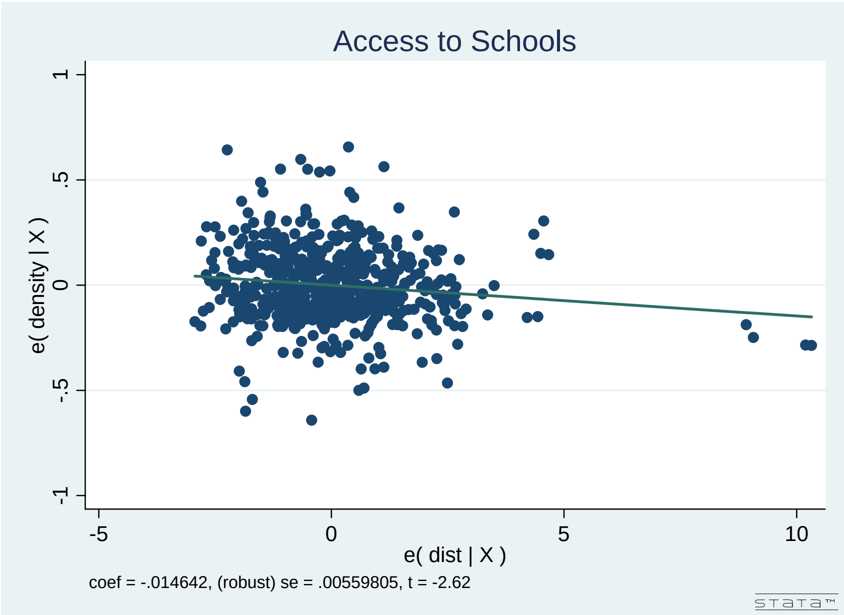
	Treated Localities		Control Localitites		Test of Means	
Variable	Obs	Mean	Obs	Mean	Mean Diff	Pr(|T| > |t|)
Outcomes						
Enrollment	317	0.54	184	0.53	0.01	0.79
Years of Education	317	6.46	184	6.47	0.00	0.87
Child Labor (Mkt)	317	0.18	184	0.17	0.01	0.55
Child Labor (Mkt+HH)	317	0.25	184	0.21	0.04	0.01*
School Attendance	301	0.98	177	0.98	0.00	0.76
Municipality-Level Treatment						
% of Treated Localities	317	0.084	184	0.087	-0.003	0.85
Access to Public Services						
Marginalization Index	317	0.45	184	0.46	-0.01	0.89
Dist to Nearest Sec School (Km)	317	2.44	184	2.48	-0.04	0.85
Health Centre Index	316	0.88	184	0.86	0.02	0.62
Health Person index	316	0.71	184	0.70	0.00	0.91
Locality Infrastructure						
Water	316	1.91	184	2.07	-0.16	0.15
Trash	316	1.56	184	1.61	-0.05	0.56
Light	316	0.49	184	0.53	-0.04	0.28
Drain	316	0.07	184	0.06	0.01	0.51
Telephone	316	1.69	184	1.61	0.08	0.06
Post Office	316	0.02	184	0.01	0.00	0.63
Local Labor Market						
Dummy for No Agricolture	316	0.05	184	0.07	-0.01	0.51
Dummy for Multiple Activities	316	0.29	184	0.33	-0.03	0.46
Agricoltural Products Index	316	3.38	184	3.39	-0.02	0.94
Avg Wage in Agricolture (Adult)	316	26.99	184	27.49	-0.50	0.56
Avg Wage in Agricolture (Child)	316	7.47	184	6.53	0.93	0.34
# of Natural Shocks	316	1.47	184	1.52	-0.05	0.34
School Supply (Municipality Level)						
School Enrollment	317	102.29	184	118.876	-16.59	0.01*
# of Sec Schools	317	17.558	184	18.7554	-1.20	0.33
Student-Teacher Ratio	317	21.219	184	21.2043	0.02	0.97
Student-Classroom Ratio	317	22.822	184	23.3304	-0.51	0.29
% Teachers with Higher Educ	317	0.2078	184	0.22992	-0.02	0.28
% Student Failing	317	0.045	184	0.05644	-0.01	0.04*
Multiple Classroom	317	0.3019	184	0.27833	0.02	0.48

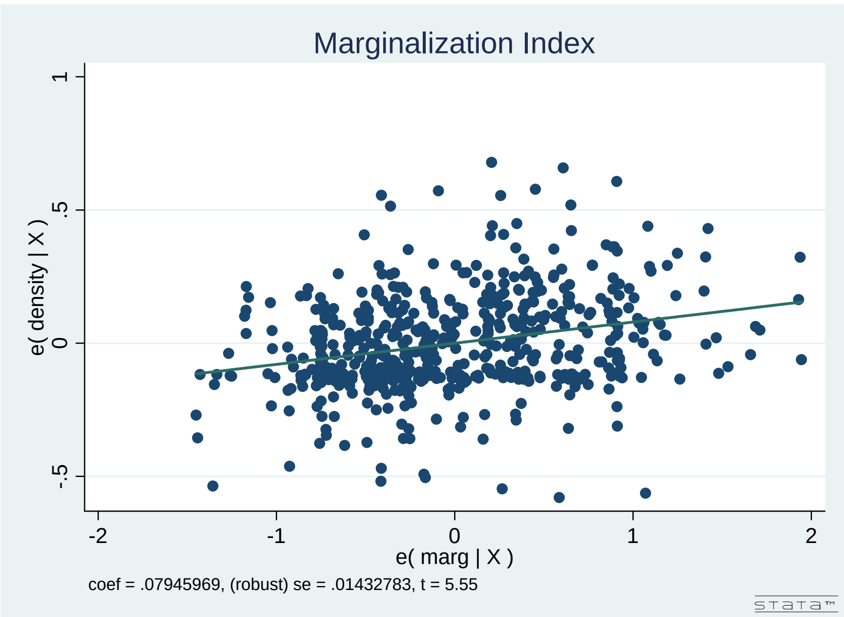
*Significant at 5% level

3.2 Determinants of the Scaling-up

As noticed earlier, the process of scaling-up across the municipalities from which the random sample was drawn has been far from homogeneous during the experimental phase of the program. At the same time, municipality-level school supply patterns over the same period reveal a wide dispersion. In this section, we try to provide some correlations between these two closely related phenomena in order to shed light on the possible underlying determinants of the scaling-up process during the experimental phase of the program.
The program administration (Progresa, 1999) puts forward two main reasons underlying this disparity in the incorporation of program-eligible localities during this period. Localities are initially selected on the basis of their *degree of marginalization* and their access to educational and health services. At first glance, this looks a bit counterintuitive as the two objectives seem to be difficult to be jointly satisfied (i.e. the more marginalized a given municipality the more likely it lacks access to public services). Hence, within this trade-off, we can envisage that other unobservable factors may have driven the actual process of geographical expansion depicted in Figures 1-2. For identification purposes, it is key to understand the underlying factors driving the differences in the proportion of treated localities (i.e. the density of treatment) across municipalities in the period 1997-1999. In particular, we shell check whether observable municipality-level characteristics may explain the allocation patterns of the density of treatment. As descriptive evidence, Figure 3 below shows scatter plots of the conditional relationship between the proportion of treated localities in each municipality and our proxy of access to educational services - the municipality-average distance from each locality to the nearest school - (top panel) and the average index of marginalization in each municipality (bottom panel). In both cases, we condition upon other municipality (observable) characteristics such as the average number of schools in each municipality, proxies for access to health services and time dummies to control for the common increasing trend.

Figure 3: Conditional Relationship between Municipality Treatment and Characteristics.





Both coefficients are significant, have the expected signs and reasonable magnitudes. Namely, a one standard deviation increase in distance to schools is associated with roughly 3% less localities treated in that municipality. Similarly an increase in the marginalization index implies a 6% increase in treatment at municipality level. Nevertheless, there seems to be an adequate amount of noise in those relationship and it is likely that unobserved heterogeneity at the municipality-level may play a significant role in explaining the pattern of allocation of treatment densities across municipalities. The crucial point though is whether this heterogenity component is also correlated with locality-level program outcomes and whether it can be assumed to be time-invariant over the period 1997-1999 or not. In the next section we explore more formally those issues.

4 Empirical Strategy

4.1 Individual and Market-level Treatment Effects

The economy is composed of M isolated markets (municipalities) populated by N_m program-eligible villages (localities). The program starts at time t_0 and progressively expands its scope till the period t_{Max} in which it incorporates the whole eligible set of villages in each market m. For each village $l \in m$ we consider the standard *individual* treatment $T_{l,t} = \{0, 1\}$ and, as an indicator of the treatment at the *market* level, we also define $P_{m,t}$ to be the proportion of treated villages in a given market $\forall t \geq t_0$:

$$P_{m,t} = E_{l \in m}(T_{l,t}) = \frac{1}{N_m} \sum_{l \in m} T_{l,t} \tag{1}$$

We observe individual outcomes and a set of pre-determined village characteristics $X_{l,97}$ for a random draw of 506 localities within 191 municipalities from the economy described above. Within this sample, village-level treatment assignment has been randomly determined[12] and constant over a given period of time $t = (t_0, t_1)$, where $t_1 < t_{Max}$. Any given village-level outcome of interest is then allowed to depend on both treatment variables T and P:

$$y_{l,t}^{T,P}(X_{l,97}) = y(T_{l,t}, P_{m,t} | X_{l,97}), \; l \in m \tag{2}$$

According to this formulation of the evaluation problem, it is immediate to see that the validity of the crucial assumption of no interaction among units (SUTVA) implicitly embedded in the impact evaluation literature would imply that:

$$\partial y(1,p)/\partial p = \partial y(0,p)/\partial p = 0, \; \forall p \in P_{m,t} \tag{3}$$

The remaining of this section is devoted to develop a reduced-form framework able to evaluate the identifiability of restriction (3) in the context of the scaling-up of Progresa over its experimental phase.

4.2 Parameters of Interest

Let us define $S_{m,t}$ to be a vector of relevant market-level school supply characteristics that may affect both outcomes and the allocation of $P_{m,t}$ across markets. For any

[12] See section 3.2 and in particular Table 2 that shows evidence of covariate-balance of $X_{l,97}$ across the two locality groups.

$p \in P_{m,t}$, the parameters of interest are the following:

$$E\left[y_{l,t_1}^{1,p}|T_l = 1, P_{m,t_1}, S_{m,t_1,X_{l,97}}\right] - E\left[y_{l,t_1}^{0,p}|T_l = 1, P_{m,t_1}, S_{m,t_1,X_{l,97}}\right] \tag{4}$$

In words, we are interested in computing the average treatment effect of the program evaluated at different degrees of the *market* treatment. In order to do so, we simply make use of the randomness of treatment assignment induced by the experimental design of our sample in order to postulate the following identification assumption:

$$E\left[y_{l,t_1}^{0,p}|T_l = 1, P_{m,t_1}, S_{m,t_1,X_{l,97}}\right] = E\left[y_{l,t_1}^{0,p}|T_l = 0, P_{m,t_1}, S_{m,t_1,X_{l,97}}\right] \tag{5}$$

Where the expectation in (5) considers simultaneously the two levels of disaggregation of our data and can thus be interpreted as a Conditional Independece Assumption (CIA) at both village and market level. It implies that conditionally on a set of locality $X_{l,97}$ and market $S_{m,t}$ characteristics, the allocation of both village treatment T_l across localities and market treatment $P_{m,t}$ across markets does not depend on the average potential outcomes of treated non-treated villages both within and across markets. In other words, condition (5) implies that after controlling for $S_{m,t}$ the allocation of P across markets is random. By using (5), we can re-write equation (4) as

$$\Delta(p) = E\left[y_{l,t_1}^{1,p}|T_l = 1, P_{m,t}, S_{m,t_1,X_{l,97}}\right] - E\left[y_{l,t_1}^{0,p}|T_l = 0, P_{m,t}, S_{m,t_1,X_{l,97}}\right] \tag{6}$$

The identification condition (5), whilst trivially satisfied within each municipality due to random assignment of T_l, may not hold across municipalities since, as outlined in section 3.2, despite our conditioning vector $S_{m,t}$, there may exist unobserved market specific factors that simultaneously determine the allocation of $P_{m,t}$ across markets and locality-level potential outcomes. Hence, we need further assumptions in order to correctly identify (6). If this sort of selection bias can be assumed to be constant over the very short time span $t = (t_0, t_1)$, then we can exploit the longitudinal dimension of our data and assume that (5) holds in first-differences:

$$E\left[y_{l,t_1}^{0,p} - y_{l,t_0}^{0,0}|T_l = 1, P_{m,t}, S_{m,t}\right] = E\left[y_{l,t_1}^{0,p} - y_{l,t_0}^{0,0}|T_l = 0, P_{m,t}, S_{m,t}\right] \tag{7}$$

This condition says that, at any given proportion of post-program market treatment $P_{m,t_1} = p$, the average *change* in outcomes between post and pre-program periods of non-treated villages provide a meaningful counterfactual for the average *change* in outcomes of treated villages if they wouldn't have been treated. Under assumption (7), in order to evaluate the relative importance of the market dimension of the treatment vis-a-vis the standard locality-level treatment effect we can identify the following quantities of interest from our data:

$$DD(p) = E\left[y_{l,t_1}^{1,p} - y_{0,t_0}^{0,0}|T_l = 1, P_{m,t}, S_{m,t}\right] - E\left[y_{l,t_1}^{0,p} - y_{l,t_0}^{0,0}|T_l = 0, P_{m,t}, S_{m,t}\right] \tag{8}$$

4.3 Empirical Implementation

Without loss of generality and for the sake of clarity, we restrict our analysis to the dichotomous version of (1), so that $P_{m,t} \rightarrow p_{m,t} \in \{0, 1\}$ according to whether a small or large proportion of villages are treated in market m at any point in time. A somehow natural candidate for the cut-off point is the median of the distribution of $P_{m,t}$ in each year[13]. Then the empirical model writes as follows:

$$y_{l,m,t} = \alpha_0 + \alpha_1 T_l + \alpha_2 D_{t_1} + \alpha_3 (T*D)_{l,t_1} + \alpha_4 (T*D*p)_{m,t} + \alpha_5 p_{m,t} + \mathbf{s}_{m,t}\phi + \epsilon_{l,m,t} \quad (9)$$

Where α_3 and α_4 capture the average behavioral responses to the program in a low and high density municipality respectively, net of any direct effects of $p_{m,t}$ on potential outcomes (measured by the parameter α_5).
In this parametric framework, the sample counterpart of the parameters of interest (8) is:

$$\widehat{DD(p)} = \widehat{\alpha_3} + \widehat{\alpha_4} \quad (10)$$

5 Results

5.1 Non-parametric Relationships

At first glance, it is useful not to restrict ourselves to the dichotomous version of the continuos market treatment variable. Hence, we begin by allowing it to vary continuosly in the 0-1 interval so as to evaluate its relationship with potential outcomes in a fully non-parametric way. By doing so, we focus on this pure univariate relationship without conditioning on any other market-specific observable characteristics. Figure 3 below displays Local Linear regression (Lowess) estimates of the relationship between the local density of market treatment ($P_{m,t}$) and outcomes for both treated ($y^{1,p}_{l,t_1}$) and control ($y^{0,p}_{l,t_1}$) localities in the post-program year 1999. We consider three educational outcomes, namely locality average enrollment rates, child labor rates and number of years of education. Figure 4 below depicts the results.

[13] In the regression analysis of Section 5 we will perform some robustness checks for the choice of this cut-off.

The charts reveal a systematic pattern. The educational outcomes of treated and non-treated localities tend to diverge after a given level of market treatment. The cut-off point lies between 60% and 80% in the proportion of treated localities within the municipality.

Interestingly, while treated localities do not appear to be sensitive to market treatment, the divergence in outcomes is due to a sharp deterioration of average outcomes (i.e. lower enrollment rates and years of education completed and higher child labor rates) of non-treated localities at high frequencies of the distribution of market treatment. Moreover, at low frequency of the distribution, both treated and non-treated average outcomes follow a parallel trend. This is consistent with the crowding-out mechanism we have outlined earlier. Due to common secondary schools resources within a municipality, the higher is the proportion of treated localities the lower the quality of learning and the expected return of schooling and hence the lower the incentive for choosing secondary school education vis-a-vis the best available alternative, i.e. working for pay. For treated localities, there are indeed two effects at play: the positive effect provided by the Progresa grant which is strictly conditional on school enrollment and attendance and the negative effect induced by the market dimension of the treatment. It is then possible to interpret those findings as evidence that common school resources are easily available in the left tail of the distribution and hence the observed increasing parallel trend. On the other hand, in the high frequency domain, the two effects are likely to compensate with each other for treated localities and hence the lack of sensitivity for this group, whereas non-treated localities face only the negative spillover induced by the program expansion and therefore are the ones that appear to be (negatively) affected by the scaling-up of the program.

While informative, this evidence is nevertheless only suggestive since a) we are not conditioning on supply-side school data that may simultaneously explain the uneven patterns of market treatment and potential outcomes and b) there may be unobserved factors that counfound the relationship we are estimating. The next section deals specifically with these issues.

5.2 Parametric Difference-in-Differences

As outlined in Section 4.2, any cross-sectional estimator of the differences in outcomes between treated and control localities may be biased due to the non-random allocation of market treatment across municipalities that may in turn be related with potential outcomes at the locality level. As long as we assume that this unobserved heterogeneity bias is constant over the period 1997-1999, we can consistently estimate equation (9) using a Double-Difference estimator and test whether behavioral responses to the program are affected by crowding-out effects due to its scaling-up. Table 3 below displays estimation summary results for our parameters of interest using different educational outcomes across alternative specifications and estimators[14]. The full specifications are reported in Appendix 7.2. Here we focus on the estimated average treatment on the treated effects (ATT) at the locality level at dif-

[14] For enrollment and attendance rates we use both OLS and Tobit ML estimators. Results are similar across the two, see Appendix A.2 for further details.

ferent degrees of market treatment. Column 1 abstracts from the market dimension of the treatment and displays the program average behavioral responses of treated villages vis-a-vis the non-treated villages, which provide a natural benchmark for the evaluation of such a policy. For example, The estimated effect of receiving the Progresa benefits is to increase locality-level secondary school enrollment rates by 4%. Models in columns 2-5 include the additional treatment variable at the market (municipality) level in the form of a dummy variable indicating whether the relative municipality in which the locality is situated displays high density of treatment or not in the post-program period 1998-1999[15]. This variable alone represents the negative spillover effect of the market density of the treatment on both treated and non-treated villages outcomes, whereas the relative interaction term with the ATT coefficient provides the effect of the Progresa grant at high densities of market treatment. Localities that reside in a high treatment density municipality appear to be associated with systematically lower enrollment rates, years of education attained and school attendance than localities situated in less intensively treated areas, as the estimated coefficient of the high-density dichotomous indicator suggests. On the other hand, treated localities in highly treated areas seem nevertheless responsive to the educational incentives embedded in the program grants, as suggested by the estimated coefficients of the relative interaction term shown in the table which is systematically bigger in magnitude to its countepart for the low density case (ATT). This confirms the previous non-parametric results and implies that the estimated difference between treated and control villages tend to diverge at high frequencies of market treatment. As expected, the choice of a more stringent cut-off -i.e. from the median to the the 75th percentile of the distribution- tend to increase the magnitude and the statistical precision of the relevant parameters. The introduction of school supply-side variables (see Appendix A.2 for the full specification for each outcome) do not seem to affect these findings in a sensible way. Among this set of municipality characteristics, access to education facilities appear to be systematically associated with lower educational outcomes. For example, a one standard deviation increase in the municipality average distance from each locality to the nearest school is associated with a decrease in secondary school enrollment rates of 9 percentage points. Similarly, quality of the education provided by each municipality, proxied by the average across schools in the number of classroom with more than one grade, appears inversely related to the program education responses.

[15] We have selected as plausible and alternatives cut-off points the median and the 75th percentile of the distribution of market treatment in each year. Results are nevertheless robust to other cut-offs.

Figure 4: Non-parametric Relationships Between Outcomes and Market Treatment

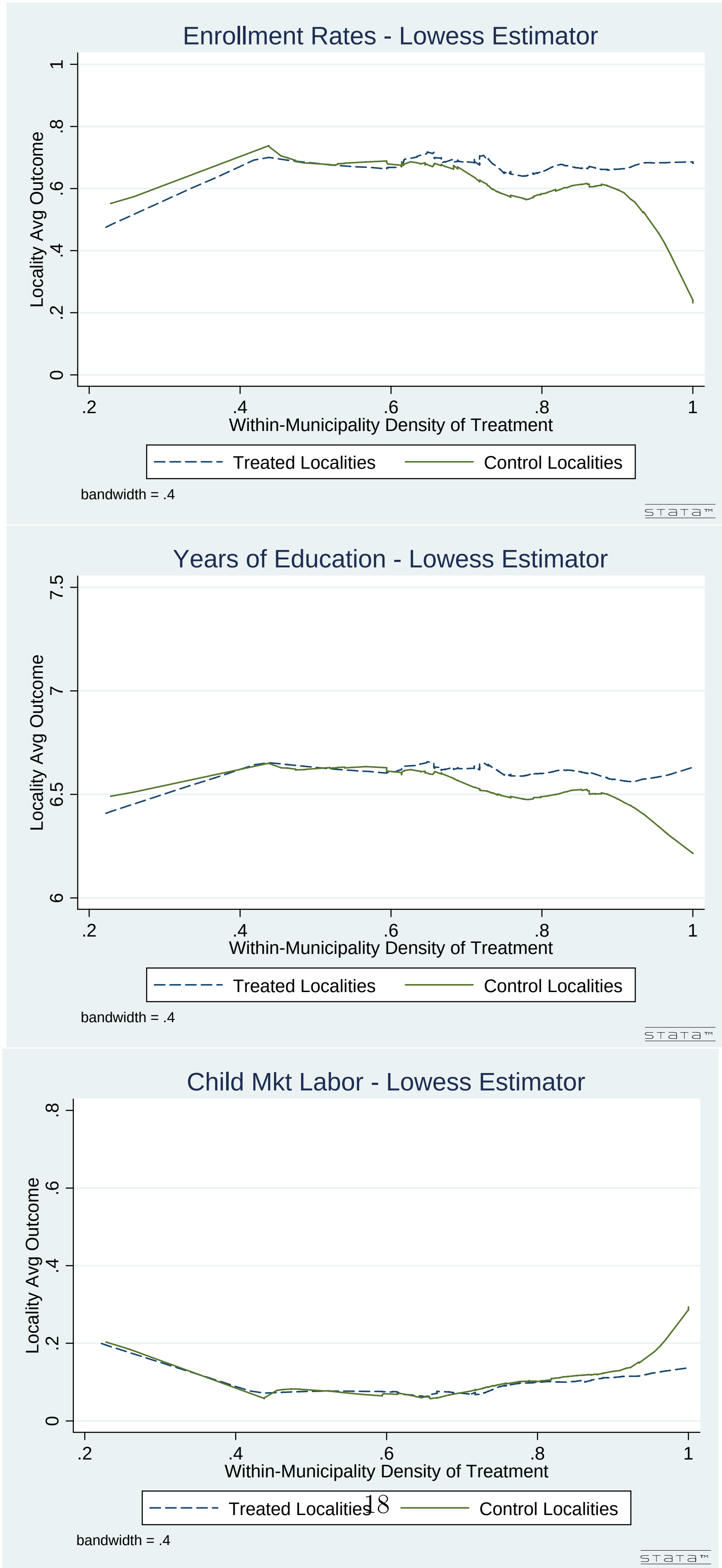


TABLE 3: Treatment Effects with Varying Density of Market Treatment. Secondary School Education

Outcome: Enrollment Rate

	(1)	(2)	(3)	(4)	(5)
		Cut-Off at 50th perc.		Cut-Off at 75th perc.	
ATT	0.044	0.032	0.034	0.035	0.034
	(2.77)***	(1.99)**	(2.12)**	(2.17)**	(2.06)**
ATT*(High Density)		0.047	0.035	0.073	0.065
		(2.18)**	(1.86)*	(2.52)**	(2.39)**
High Density		-0.013	-0.026	-0.048	-0.064
		(-0.54)	(-1.17)	(-2.26)**	(-2.88)***
School-supply controls	No	No	Yes	No	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes
Number of Obs	1501	1501	1501	1501	1501
R-squared	0.096	0.098	0.223	0.101	0.229
Number of Municipalities	191	191	191	191	191

Outcome: Years of Education Attained

	(1)	(2)	(3)	(4)	(5)
		Cut-Off at 50th perc.		Cut-Off at 75th perc.	
ATT	0.044	0.020	0.023	0.031	0.031
	(2.04)**	(0.91)	(1.05)	(1.42)	(1.43)
ATT*(High Density)		0.107	0.090	0.109	0.095
		(4.08)***	(3.92)***	(2.81)***	(2.66)***
High Density		-0.054	-0.056	-0.067	-0.074
		(-2.14)**	(-2.27)**	(-2.88)***	(-2.92)***
School-supply controls	No	No	Yes	No	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes
Number of Obs	1501	1501	1501	1501	1501
R-squared	0.059	0.067	0.179	0.067	0.181
Number of Municipalities	191	191	191	191	191

Outcome: School Attendance Rate

	(1)	(2)	(3)	(4)	(5)
		Cut-Off at 50th perc.		Cut-Off at 75th perc.	
ATT	0.007	0.004	0.004	0.006	0.005
	(0.89)	(0.51)	(0.44)	(0.70)	(0.62)
ATT*(High Density)		0.015	0.016	0.015	0.016
		(2.04)**	(2.22)**	(1.56)	(1.71)*
High Density		-0.009	-0.011	-0.008	-0.010
		(-1.81)*	(-2.18)**	(-1.80)*	(-2.21)**
School-supply controls	No	No	Yes	No	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes
Number of Obs	1451	1451	1451	1451	1451
Number of Municipalities	191	191	191	191	191

Note: Standard errors clustered at municipality-level.
* significant at 10%; ** significant at 5%; *** significant at 1%

5.3 Interpretation of the Results and Implications for the Evaluation

Both the parametric and non-parametric results point toward the importance of both treatment variables as far as behavioral responses to such educational policy are concerned. In this context, the dichotomous village-level treatment indicator is not fully informative about the several mechanisms through which the geographical expansion of the program may affect outcomes. The market-level treatment indicator is then ought to capture these spillover effects across villages that share the access to and the use of some common public services that are key for the effectivness of the program.

Overall, the main message embedded in these empirical regularities is that the market dimension of the treatment matters for the evaluation of large in scope social programs such as Progresa in Mexico. More precisely, whenever social programs are rolled-over on a large scale during the experimental evaluation attention should be paid to the potential interactions among units within a delimited geographical area. Programs of this type should be evaluated by taking into account the geographic dimension of the local market in which the unit of analysis (villages, households, individuals, etc) is situated, so as to incorporate both the direct and indirect effects that these policies induce.

6 Conclusions

In this paper, we have attempted to evaluate the extent to which the experimental results of a small-scale randomized experiment may be informative in capturing behavioral responses whenever programs are simultaneously implemented on a larger scale.
Our approach is fairly atheoretical, in the sense that we pose minimal assumptions about behaviors and the structure of the economy we ought to analyze. In fact, we use the experimental design of the *Progresa/Oportunidades* CCT program recently implemented in Mexico and its contemporaneous nation-wide expansion as a 'natural experiment' in order to evaluate policy responses of localities facing different exposure to market treatment.

Overall, the results show that average educational responses are indeed sensitive to the local density of the treatment of the relative market. This is interpreted as supporting evidence for crowding-out mechanism across localities that share common resources within specific geographical regions. This and perhaps other types of spillovers across units (villages) are unlikely to be captured by the standard experimental evaluation framework that considers small random samples within the whole eligible population.

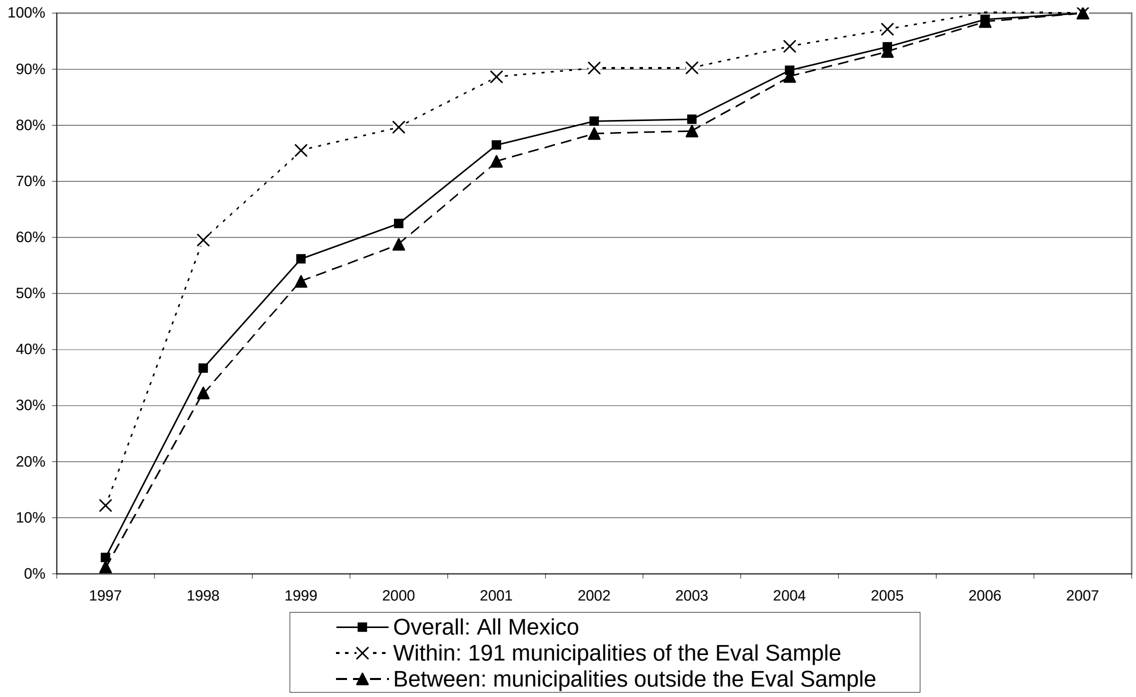
This study opens several avenues for future research on the topic. Beyond these reduced form evidence, a potentially fruitful and complementary approach points toward building a structural model that incorporates the main features of the CCT program under study. In particular, a dynamic behavioral model of work, schooling and occupational choice seems particularly suited for simulating the policy and its counterfactuals at different degrees of geographical expansion of the program[16]. Modeling would allow for a neat and more rigorous understanding of the possible indirect effects at play induced by the scaling-up of the program.

[16] Two recent papers - Todd and Wolpin [2006] and Attanasio et al. [2005] - have employed structural models to evaluate the partial equilibrium or direct effects of Progresa during its experimental phase 1998-2000. In particular, both studies use the experimental results to validate the model predictions. However, none of them extend the rich structure of the model to encompass the potential indirect effects of the program so as to compare, disentangle and assess the differential behavioral responses at different implementation stages of the program.

A Appendix

A.1 Scaling-up. Decomposing the Various Dimensions

Figure 5: Evolution in the Proportion of Treated Localitites in Different Regions. All Program Years (1997-2007)



A.2 Difference-in-Differences Models. Full Specifications.

Enrollment Rates

Model	(1) OLS	(2) OLS	(3) OLS	(4) Tobit
Treat	0.008	0.007	0.005	0.001
	(0.32)	(0.28)	(0.23)	(0.03)
Year (1998-1999)	0.066	0.066	0.048	0.046
	(5.13)***	(5.03)***	(3.24)***	(2.00)**
Treat*Year	0.044	0.035	0.034	0.042
	(2.77)***	(2.17)**	(2.06)**	(1.45)
Treat*Year*(High Density)		0.073	0.065	0.073
		(2.52)**	(2.39)**	(2.18)**
High Density		-0.048	-0.064	-0.071
		(-2.26)**	(-2.88)***	(-4.09)***
Distance to nearest school			-0.049	-0.051
			(-7.77)***	(-14.01)***
# of Sec Schools			0.000	0.000
			(0.35)	(0.80)
Student-Classroom Ratio			-0.002	-0.003
			(-0.94)	(-1.34)
Student-Teacher Ratio			0.010	0.011
			(3.19)***	(5.00)***
% Teachers with Higher Educ			0.056	0.046
			(0.62)	(0.87)
% Student Failing			0.287	0.325
			(1.07)	(1.92)*
Multiple Classroom			-0.043	-0.047
			(-1.71)*	(-1.99)**
State Dummies	Yes	Yes	Yes	Yes
Number of Obs	1501	1501	1501	1501
R-squared	0.096	0.101	0.229	0.462
Number of Localities	191	191	191	191

Note: Standard errors clustered at municipality-level.
* significant at 10%; ** significant at 5%; *** significant at 1%

Years of Education Attained

Model	(1) OLS	(2) OLS	(3) OLS
Treat	-0.005	-0.006	-0.008
	(-0.19)	(-0.23)	(-0.33)
Year (1998-1999)	0.045	0.045	0.036
	(2.51)**	(2.45)**	(1.87)*
Treat*Year	0.044	0.031	0.031
	(2.04)**	(1.42)	(1.43)
Treat*Year*(High Density)		0.109	0.095
		(2.81)***	(2.66)***
High Density		-0.067	-0.074
		(-2.88)***	(-2.92)***
Distance to nearest school			-0.059
			(-8.37)***
# of Sec Schools			0.000
			(0.35)
Student-Classroom Ratio			-0.004
			(-1.36)
Student-Teacher Ratio			0.005
			(1.50)
% Teachers with Higher Educ			0.065
			(0.73)
% Student Failing			0.278
			(0.95)
Multiple Classroom			-0.063
			(-2.41)**
State Dummies	Yes	Yes	Yes
Number of Obs	1501	1501	1501
R-squared	0.059	0.067	0.181
Number of Localities	191	191	191

Note: Standard errors clustered at municipality-level.
* significant at 10%; ** significant at 5%; *** significant at 1%

School Attendance Rates

Model	(1) Tobit	(2) Tobit	(3) Tobit
Treat	-0.001	-0.001	-0.001
	(-0.11)	(-0.14)	(-0.15)
Year (1998-1999)	0.004	0.004	0.003
	(0.61)	(0.58)	(0.46)
Treat*Year	0.007	0.006	0.005
	(0.89)	(0.70)	(0.62)
Treat*Year*(High Density)		0.015	0.016
		(1.56)	(1.71)*
High Density		-0.008	-0.010
		(-1.80)*	(-2.21)**
Distance to nearest school			0.005
			(3.96)***
# of Sec Schools			-0.001
			(-1.73)*
Student-Classroom Ratio			0.001
			(0.02)
Student-Teacher Ratio			0.001
			(0.75)
% Teachers with Higher Educ			0.013
			(0.86)
% Student Failing			-0.033
			(-0.68)
Multiple Classroom			-0.009
			(-1.40)
State Dummies	Yes	Yes	Yes
Number of Obs	1451	1451	1451
R-squared	-0.026	-0.032	-0.069
Number of Localities	191	191	191

Note: Standard errors clustered at municipality-level.
* significant at 10%; ** significant at 5%; *** significant at 1%

References

Angelucci and De-Giorgi (2007). Indirect effects of and aid program: how do liquidity injections affect non-eligibles consumption? *American Economic Review, Forthcoming*.

Attanasio, Meghir, and Santiago (2005). Education choices in mexico: Using a structural model and a randomized experiment to evaluate progresa. Technical report, UCL Working Papers.

Banerjee and Duflo (2006, Winter). Addressing absence. *Journal of Economic Perspectives 20*(1), 117–132.

Behrman and Todd (1999). Randomness in the experimental sample of progresa. Technical report.

Bertrand, Djankov, Hanna, and Mullainathan (2006, June). Does corruption produce unsafe drivers? NBER Working Papers 12274, National Bureau of Economic Research, Inc.

Blundell, C. Dias, C. Meghir, and V. Reenen (2004). valuating the employment impact of a mandatory job search program. *Journal of the European Economic Association 2*, 569–606.

Duflo (2005, June). Field experiments in development economics. BREAD Policy Papers 12, BREAD.

Duflo, Kremer, and Robinson (2006, April). Understanding. technology adoption: Fertilizer in western kenya: Evidence from field experiments. Mimeograph, mit, MIT.

Ferracci, Jolivet, and V. de Berg (2008). Joint evaluation of treatment effects at the individual and population levels. Mimeo.

Glewwe and Kremer (2006, November). *Schools, Teachers, and Education Outcomes in Developing Countries*, Volume 2 of *Handbook of the Economics of Education*, Chapter 16, pp. 945–1017. Elsevier.

Heckman and Smith (1998, May). Evaluating the welfare state. NBER Working Papers 6542, National Bureau of Economic Research, Inc.

Karlan, D. S. and J. Zinman (2005, May). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. Working Papers 911, Economic Growth Center, Yale University.

Kremer (2003). Randomized evaluations of educational programs in developing coun- tries: Some lessons. *American Economic Review 93*(2), 102–106.

Miguel and Kremer (2004, 01). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Rubin (1974). Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology 66*(5).

Rubin (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association 81*(396), 961–962.

Todd and Wolpin (2006, December). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review 96*(5), 1384–1417.