WPS 3386

An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA

Hielke Buddelmeyer

**Melbourne Institute of Applied
Economic and Social Research
and IZA Bonn
The University of Melbourne
Parkville Victoria 3010 - AUSTRALIA
Email:** hielkeb@unimelb.edu.au

Emmanuel Skoufias

**The World Bank
1818 H Street NW
Washington DC 20433-USA
Email:** eskoufias@worldbank.org

Abstract

While providing the most reliable method of evaluating social programs, randomized experiments in developing and developed countries alike are accompanied by political risks and ethical issues that jeopardize the chances of adopting them. In this paper we use a unique data set from rural Mexico collected for the purposes of evaluating the impact of the PROGRESA poverty alleviation program to examine the performance of a quasi-experimental estimator, the Regression Discontinuity Design (RDD). Using as a benchmark the impact estimates based on the experimental nature of the sample, we examine how estimates differ when we use the RDD as the estimator for evaluating program impact on two key indicators: child school attendance and child work.

Overall the performance of the RDD was remarkably good. The RDD estimates of program impact agreed with the experimental estimates in 10 out of the 12 possible cases. The two cases in which the RDD method failed to reveal any significant program impact on the school attendance of boys and girls were in the first year of the program (round 3). RDD estimates comparable to the experimental estimates were obtained when we used as a comparison group children from non-eligible households in the control localities.

1. Introduction

In recent years the evaluation of social programs in developing countries has been receiving increasing recognition as an essential component of social policy. Credible and reliable program evaluations can not only lead to better efficacy of programs but can also contribute to the more cost effective use of public funds as well as increase social and political accountability in counties with limited experience in such practices. By general consensus, the most credible evaluation designs are experimental, involving the randomized selection of the set of individuals or communities or geographic areas receiving the intervention and those that serve the role of comparison group not receiving the intervention (Newman, Rawlings and Gertler, 1994; Heckman, 1992; Heckman and Smith, 1995; and Heckman, LaLonde and Smith, 1999). However, while providing the most reliable method of evaluating social programs, randomized experiments in developing and developed countries alike are accompanied by risks that jeopardize the chances of adopting them. For example, there are considerable political difficulties in justifying and practical difficulties in maintaining a group of "untreated" or comparison households simply for the purposes of an evaluation of the program's impact. To make matters worse, the delay or denial of program benefits to individuals or communities randomized out of the program serving the role of a comparison group provides plenty of opportunities for opponents to the program to manipulate for their own political purpose the ethics of withholding benefits from "equally deserving" households.[1]

The toolkit of alternative designs for the evaluation of social programs includes a variety of non-experimental methods (e.g., Shadish, Cook and Campbell, 2001; Heckman, LaLonde and Smith, 1999; Baker, 2000). Most popular among such methods are propensity score matching

---

[1] A case in point is the criticism raised against the PROGRESA experimental design during the

and more recently regression discontinuity design. As in social experiments, both of these "quasi-experimental" methods attempt to equalize the selection bias present in treatment and comparison groups so as to yield unbiased estimates of parameters measuring program impact, such as the average treatment effect, the treatment of the treated effect and the local average treatment effect (Blundell and Costa-Diaz, 2002). One critical question associated with these non-experimental approaches is the extent to which they result in impact estimates that are comparable to those that would be obtained with an experimental approach.

With these considerations in mind, this paper uses a unique data set from rural Mexico collected for the purposes of evaluating the impact of the PROGRESA poverty alleviation program to examine the performance of the Regression Discontinuity Design (RDD). The PROGRESA program and its impact on a variety of welfare indicators have been extensively evaluated using the experimental design of the sample. [2] In this paper, we evaluate a non-experimental estimator rather than the impact of the program itself, by examining how different impact estimates would be if one were to use the RDD as the estimator for evaluating program impact on two key indicators: child school attendance and child work.

The RDD estimator in the economic literature has been recently used by Van Der Klaauw, 2002; DiNardo and Lee, 2002; Lee, 2001; Black, 1999; Pitt and Khandker, 1998; and Angrist and Lavy, 1999, while the identification and estimation of treatment effects with an RD design are discussed in Hahn, Todd and Van der Klaauw, 2001. In the case of PROGRESA the particular method of selecting households eligible for the program benefits provides the basis for the RD design. As discussed in more detail later, all households within a region are ranked

---

media campaign of opposition parties in Mexico prior to the elections of 2000.
[2] Skoufias (2001) provides a detailed discussion of PROGRESA, the evaluation design and the experimental estimates impacts of the program.

by their (predicted) probability of being below the regional cut-off point (or poverty line). Eligibility then is determined based on whether the predicted probability of a household being poor is above some pre-determined cut-off point. Relying on the experimental design adopted for the evaluation of the program, we measure the performance of RDD by comparing impact estimates using RDD against the benchmark impact estimates obtained using experimental methods. The nature of the sample for the evaluation of PROGRESA offers two additional advantages. Firstly, we can investigate whether the non-experimental estimator is subject to evaluation bias, i.e., whether individuals from non-eligible households in treatment localities provide an adequate comparison group for program beneficiaries and hence for impact evaluation. Secondly, we can also identify some of the possible sources of evaluation bias by testing whether spillover or anticipation effects of the program contaminate comparison groups.

Although in recent years there has been increasing attention to non-experimental estimators (Friedlander and Robbins, 1995) most of it, if not all, has been focused on the performance of propensity score matching (e.g., Heckman, Ichimura and Todd, 1997; Dehejia and Wahba, 2002; Smith and Todd, 2001). To our knowledge, our study is the first one providing evidence on the performance of the RD design. Taking into consideration the fact that a number of countries apply methods similar to those of PROGRESA to determine household eligibility for social programs, but are either too averse to adopting experimental designs or too far advanced in their implementation of the program and coverage of households, the findings reported herein are particularly useful for determining whether non-experimental approaches can provide a reliable alternative to social experimentation.[3] Moreover, the RDD approach requires relatively little information and produces precisely the variable of interest when one is

---

[3] Examples include the *Bolsa Escola* and *Bolsa Alimentacao* programs in Brazil and the *Familias en*

3

interested in scaling up (down) the program by lowering (raising) the eligibility threshold[4].

Our paper is organized as follows. Section 2 describes in brief the PROGRESA program and the process used to select beneficiaries. Section 3 outlines the RDD approach and how it applies to the PROGRESA evaluation sample. Section 4 describes the empirical strategy adopted for the analysis of the performance of RDD and presents our various findings. Section 5 concludes.

2. Some background on PROGRESA

PROGRESA (recently renamed *Oportunidades*) is one of the major programs of the Mexican government aimed at developing the human capital of poor households. Targeting its benefits directly to the population in extreme poverty in rural and recently expanded to urban areas with less than 1 million in inhabitants, it aims to alleviate current poverty through monetary and in-kind benefits, as well as reduce future levels of poverty by encouraging investments in education, health and nutrition. By the end of 2002, the program included nearly 4.24 million families in more than 72,000 localities in all 31 states. This constitutes around 20 percent of all Mexican households and 77 percent of those households considered to be in extreme poverty. The total annual budget of the program in 2003 is approximately US$2.3 billion, equivalent to 20% of the Federal poverty alleviation budget or 0.2% of GDP.

Although the program consists of a variety of interventions in health and nutrition, the features most related to our purposes are those regarding schooling. The education component of PROGRESA is designed to increase school enrollment among youth in Mexico's poor rural

---

Accion program in Colombia.
[4] This is different from the type of scaling up in Attanasio et. al. (2003) that discusses in depth the issues related to predicting the effect of exporting an existing program from a population

communities by providing cash transfers to mothers in the households on the condition that their children attend school on a regular basis. In localities where PROGRESA currently operates, households that have been characterized as poor, and have children enrolled in grades 3-9, are eligible to receive these educational grants every two months. The levels of these grants were determined taking into account, among other factors, what a child would earn in the labor market or contribute to family production. The educational grants are slightly higher at the secondary level for girls relative to boys, given their propensity to drop out at earlier ages. Every two months, confirmation of whether children of beneficiary families attend school more than 85% of the time is submitted to PROGRESA by school teachers and directors, and this triggers the receipt of the bi-monthly cash transfer for school attendance. The average monthly payment (received every two months) by a beneficiary family amounts to 20% of the value of monthly consumption expenditures prior to the initiation of the program.

Because of budgetary constraints, the program was introduced in phases. The necessity to introduce the program in phases was capitalized upon by phasing in the program along the lines of a randomized experiment where localities were either randomized to be in (treatment localities) or out (control localities). The resulting experimental data was used to evaluate program impacts regarding different outcomes related to schooling, health and nutrition[5]. Specifically, the sample used in the evaluation of PROGRESA consists of repeated observations (panel data) collected for 24,000 households from 506 rural localities in the seven states of Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi and Veracruz. Of the 506 localities, 320 localities were assigned to the treatment group and 186 localities were assigned as controls.

---

where its effects were evaluated to a different population.

The selection of households as PROGRESA beneficiaries was accomplished by first identifying the communities to be covered by the program (geographic targeting) and then selecting the beneficiary households within the chosen communities.[6] The household selection within communities covered by the program involves conducting a census of all the households in the communities to be covered by the program and collecting detailed information on household demographic composition, assets and income. Adult per capita income is then compared with the cost of the *Standard Food Basket* (which is roughly equivalent to two times the minimum wage) of 320 pesos per capita per month in order to generate a new binary variable taking the value of 1 for poor households (if income is less than the cost of the standard food basket) and 0 for nonpoor. Finally, discriminant analysis is applied, <u>separately for each geographical region</u>, in order to identify the variables that discriminate best between poor and nonpoor households and compute an index (discriminant score) that represents parsimoniously the differences between the poor and nonpoor households.[7]

In the early stages of the program (i.e. during 1998) the PROGRESA beneficiary selection method led to approximately 52% of the households in the evaluation sample to be classified as eligible for the program benefits. By July 1999, PROGRESA had added new households to the list of beneficiaries since it was felt that the original selection method was biased against the elderly poor who no longer lived with their children.[8] As a result of the revised selection process the fraction of households classified as eligible for program benefits increased from 52%

---

[5] See Skoufias (2001) for a synthesis of all the available results of PROGRESA's evaluation.
[6] For a more detailed discussion see Skoufias et al. (2001).
[7] Discriminant analysis may be considered as analogous to a logit probability model for a binary dependent variable. Its main difference from the logit model is that it allows a variety of ways for classifying observations into groups (for more details see Sharma, 1996).
[8] The Spanish term used to describe this revised selection process is *densificacion*. The revised selection process did not simply increase the region-specific thresholds but instead it revised the

of the evaluation sample to 78% of the sample. However, after the release of the payment records in late August 2000, it was discovered that in the evaluation sample, many of the households that were supposed to be added to the updated list of beneficiaries had not received any cash benefits since the start of the distribution of program benefits in these localities.

Specifically, in the treatment localities 27% of the total eligible population had not received any benefits by March 2000. After crosschecking this with the PROGRESA administration it was confirmed that the explanation for this was due to the fact that the majority (85.7%) of the households not receiving any benefits, due to some administrative error, had never been incorporated into the program. All of these "forgotten" households were households with a revised eligibility status from non-beneficiary to eligible beneficiary as a result of the revision of the selection process (densification). Given the intricacy of what constitutes an eligible household we have decided to exclude from our sample those households that were later reclassified as eligible and received nonzero cash transfers during the three years covered by the evaluation surveys (October 1997-November 1999). [9] Thus our set of eligible households includes the original households classified as such using PROGRESA's early selection methods while the set of non-eligible includes all of the "forgotten" households with revised eligibility status that never received any cash benefits.

3. The Regression Discontinuity Design applied to PROGRESA

The RD design applied in this paper is based on the discontinuity in the eligibility

---

way household-specific discriminant scores were calculated.
[9] These households were identified from the administrative records of PROGRESA that contained information on the bi-monthly payments sent out to beneficiary households by the headquarters of the program in Mexico City.

criterion.[10] In order to be eligible for PROGRESA services one needs to have a low discriminant score. The localities for which data were collected were grouped into seven broad geographical regions. For each region a separate discriminant analysis was performed to calculate the discriminant score which resulted in a situation were different regions have different threshold scores to determine if one is eligible or not. Figure 1 presents kernel estimates of the density of the discriminant score in each region along with the threshold score applied in each region. These kernel densities are estimated by pooling all the households in treatment and control localities in the region. As it can be seen in most regions the threshold scores fall frequently at the mean of the distribution of the discriminant scores which implies that there is a considerable number of households just to the right of the threshold score that have a discriminant score very close to that of the eligible households. This is precisely what the RDD approach relies upon to estimate program impact: by comparing individuals just below and above the threshold score.

Using more formal notation, a child's participation in the labor market or his/her school attendance Y_i may be modeled by an equation such as:

$$Y_i = \alpha_i + \beta^{RDD} B(DS_i) + \varepsilon_i \tag{1}$$

where α_i and β^{RDD} are parameters to be estimated, and B is the treatment indicator that equals one if child i is eligible for PROGRESA benefits and zero otherwise. In our case, the treatment indicator is a function of DS_i, the discriminant score. Of course, if one were only

---

[10] An alternative and increasing popular method is that of propensity score matching. A key prerequisite for the application of propensity score matching is that individuals with equal probability of being selected as beneficiaries of the program are left out of the program (Heckman, Ichimura and Todd, 1997). Given that this "overlapping support" condition is not satisfied by PROGRESA's methods of selecting eligible households we are unable to apply this method and test its performance using the PROGRESA sample. See Diaz, Handa and Orozco (2003) for an analysis of the evaluation bias of matching estimators using comparison households outside the PROGRESA evaluation sample.

interested in the program effects for those who are treated, the randomized experiment setup to evaluate PROGRESA would allow us to easily compare mean enrollment rates for eligible children in treatment and control localities. But the availability of the experimental data in combination with the discontinuity in eligibility offers the opportunity to examine the performance of the RDD estimator.

The RDD literature distinguishes between the so-called sharp and fuzzy designs. Our case is one of a sharp design since treatment B is known to depend on DS_i in a deterministic way[11]. Denoting the (region specific) threshold score by COS, we know families with a discriminant score above COS are excluded from receiving PROGRESA benefits. The RDD approach relies on the maintained, albeit untested, hypothesis that individuals with a discriminant score just below the threshold score are similar in their observed and unobserved characteristics to individuals with score just above the threshold score. Comparing a sample of individuals within a very small range around the threshold score will be analogous to conducting a randomized experiment at the threshold score. This is why the RDD approach is often referred to as a quasi-experimental design. More formally, children with a discriminant score just to the left and just to the right of the threshold score may be considered as identical in the sense that in the absence of the treatment, the unconditional mean values of Y are the same, i.e.,

$$\mathrm{E}(\alpha_i \mid DS_i = COS - \Delta) \cong \mathrm{E}(\alpha_i \mid DS_i = COS + \Delta) \tag{2}$$

where Δ denotes an arbitrarily small number.

Then, assuming that $E(\alpha_i \mid DS_i = DS)$ and the conditional mean function $E(\varepsilon_i \mid DS)$ are

---

[11] See for instance Van der Klaauw (2002) and Hahn et al (2001) for a detailed discussion on the sharp and fuzzy designs. Table 1 provides more supporting evidence on the extent to which the

both continuous at $DS_i = COS$, it can be shown that for the case of a constant treatment effect, the average treatment effect can be identified by a simple comparison of the mean values of Y between those individuals to the left (eligible) and to the right of the threshold score COS (non-eligible), i.e.,[12]

$$\beta^{RDD} = Y^- - Y^+ = \lim_{DS\uparrow COS} E(Y_i \mid DS_i = DS) - \lim_{DS\downarrow COS} E(Y_i \mid DS_i = DS). \tag{3}$$

In the less restrictive case where the treatment effect is allowed to be heterogeneous across individuals (i.e., β varies across individuals) then β_i identifies the local average treatment effect (LATE) for the subgroup of individuals at the threshold point COS.

In our application of the RDD estimator, we use only information on the outcome measures (school enrollment and work incidence), the discriminant scores, and the six qualifying threshold discriminant scores that vary from region to region. In order to obtain estimates of the unconditional means of the outcome measures of interest, denoted by Y^- and Y^+ we use one-sided kernel regressions, which are given by the expressions

$$Y^- = \frac{\sum_{i=1}^{n} Y_i * \omega_i * K(u_i)}{\sum_{i=1}^{n} \omega_i * K(u_i)} \quad and \quad Y^+ = \frac{\sum_{i=1}^{n} Y_i * (1-\omega_i) * K(u_i)}{\sum_{i=1}^{n} (1-\omega_i) * K(u_i)} \tag{4}$$

where, ω_i denotes the indicator function $I(DS_i \le COS_R)$, where COS_R is the region-specific threshold value of the discriminate score, $K(u_i)$ denotes the kernel and $u_i = \frac{DS_i - COS_R}{h}$,

---

where h is an appropriate bandwidth.[13]

As Hahn et al (2001) demonstrate this procedure is numerically equivalent to an instrumental variable estimator for the regression of Y_i on B_i (our treatment indicator), which uses ω_i as an instrument, applied to the sub-sample for which $COS_R - h < DS_i < COS_R + h$.

Table 1 presents the number of households in the PROGRESA sample by region and the cut-off points that can be inferred by examining the maximum and minimum values of the discriminant scores of eligible and non-eligible households respectively. Using the maximum value of the discriminant score observed eligible households in each region, we can see that the cut-off points vary from region to region. The incidence of discriminant scores less than the region-specific threshold point for some non- eligible households suggests that there may have been some additional criteria used to determine eligibility. The PROGRESA central administration, for example, claims that it did not adopt a purely mechanical approach in the sense that selected households were reclassified from one category to another based on an additional set of filters such as age, as well as feedback from local authorities and personnel with better information about household assets and their "true" poverty status. However, the low numbers of non-eligible households with scores lower than the threshold (see column 3 in table 1) suggest that these instances are rare in most regions. This also implies that the PROGRESA selection process is better approximated by a "sharp design" especially in regions

---

Typically $c(DS_i)$ is specified as a higher order polynomial.

[13] An alternative approach consists of evaluating the experimental and the RDD estimates of program impact separately by region. This however, tends to confound the general question of how well the RDD performs due to the additional complications introduced by differences in sample sizes across regions and differences in the operational efficiency of the program across regions. Because of the poor boundary performance of standard kernel estimators we also explored local linear regressions (LLR) as suggested by Fan (1992). Our results showed very similar estimates based on LLR versus RDD and hence these are not reported.

3, 4, 5, 6 and 27.

4. Empirical Strategy and Findings

To provide the reader with a better understanding of the empirical strategy we adopt to evaluate the performance of the RDD, Table 2 provides a schematic decomposition of the PROGRESA evaluation sample and the various alternatives available for selecting comparison groups for an evaluation of the impact of the program. Within any survey round before ($t = 0$) or after the start of the program ($t = 1,2,3…$), the total population surveyed can be divided into 4 different groups depending on whether an individual child or adult belongs in a household classified as eligible to receive PROGRESA benefits ($B=1$ for eligible households and $B=0$ for non-eligible households) and according to whether the household that the individual belongs to resides in a locality where PROGRESA is in operation (treatment locality or $T=1$) or not (control locality or $T=0$).

A social experiment, randomly assigning individuals or communities into treatment and control groups, solves the evaluation problem by using information from individuals or households in the control group to construct an estimate of what participants would have experienced had they not participated in the program (i.e. individuals from group B or D if needed).[14] Non-experimental and quasi-experimental estimators on the other hand, assuming that the program has already covered all of the targeted localities, are constrained to estimating program impact based on comparisons of households or individuals between groups and A and C. Specifically, the RDD estimator evaluates impact using children from households who are eligible for PROGRESA benefits (i.e. with household discriminant scores just below the

[14] In fact all of the evaluation of the PROGRESA program has relied exclusively on comparison

12

threshold value in group A) and children from households who are ineligible (i.e., with household discriminant scores just above the threshold value in group C).

Our analysis is limited to the school attendance and work activities of boys and girls between 12 and 16 years of age. We focus on this age group for the primary reason that the programs' impact is more likely to be found among children transitioning from primary to junior high school.[15] As noted earlier, we also exclude from our sample households in the treatment villages (group C) that were later reclassified by PROGRESA as beneficiaries, as well as all children that reside in another place outside the household for the purposes of studying or working. The adult member responding to the interviewers is less likely to have accurate information about their school attendance or work activities.

School attendance is defined according to those who respond that the child attends school. This question is identical over the different rounds of analysis. Our definition of working is defined to include all children who report that they worked over the previous week (whether paid or unpaid). There is also a follow-up question to capture individuals who may engage in informal activities that the respondent may not have initially considered as work. This question asks about participation in a) selling a product, b) helping in family business c) making products to sell, d) washing, cooking or ironing and e) working in agriculture activities or caring for animals. We also include as working individuals who respond that they engage in any of these activities.

All households were initially surveyed in October/November 1997, and based on the first survey the eligibility status of households was determined. Based on PROGRESA's

between the A and B groups of households.
[15] An additional reason for not including 17 year old boys and girls in our analysis is the fact that in round 3, the evaluation survey collected information on the school attendance and

beneficiary selection method, all households in both treatment and control communities were classified as eligible or non-eligible for participation in the program. A second survey took place in March 1998 before the initiation of payments in July 1998. The third round of the survey took place in October 1998, which was well after most households received some cash transfers from the program. The next round of the survey took place in June 1999, and the fifth round took place in November 1999. Soon after November 1999, the benefits of the program began to be distributed to eligible households in the control communities. We limit our analysis to rounds 1, 3 and 5 of the survey, since these three surveys took place at the same months in time within each calendar year.[16]

Performance of the RDD estimator

Tables 3a and 3b present the RDD and experimental estimates of program impact by round and gender on the sample of boys and girls between 12 and 16 years of age. Our benchmark experimental impact estimates are the cross-sectional difference (CSDIF) estimates obtained by comparing the post program differences in the means between treatment and control groups. Using the sample of beneficiary/eligible households (*B=1*) in treatment and control localities (groups A and B in Table 2), the program impact on the binary outcome indicator Y_i pooling together the three rounds, is specified by a linear probability model (LPM) of the form:

$$Y_i = \alpha + \beta_0 T_i + \gamma_3 R3 + \beta_3 \left(T_i * R3\right) + \gamma_5 R5 + \beta_5 \left(T_i * R5\right) + \sum_{j=1}^{J} \theta_j X_{ij} + \eta_i , \qquad (5)$$

---

market work of children only up to age 16 (inclusive).
[16] The means of all the variable used in our analysis for boys and girls in groups A,B, C and D separately are provided in Appendix table A.

where T_i represents a binary variable equal to 1 if child i lives in a treatment community and 0 otherwise, R3 and R5 are binary variables that take the value of one (zero otherwise) for observations in the third (October 1998) and fifth (November 1999) rounds of the survey, respectively, X_{ij} represents the vector of J control variables for individual, household and locality characteristics and η is an error term summarizing the influence of random disturbances.[17] The vector X of control variables consists of parental characteristics, such as the education level of the mother and father of the child, the age of the mother and father, whether parents speak an indigenous language and whether they also speak Spanish.[18] We also include a number of variables measuring the demographic composition of the household. These variables include the number of children aged 0 to 2 and aged 3 to 5, boys and girls aged 6-7, 8-12, and 13 to 18, men and women aged 19 to 54 and men and women over the age of 55.

As control variables at the community level, we include an index variable constructed by the PROGRESA administration as a means of summarizing the infrastructure and the level of development of the locality (otherwise known as the marginality index) and a variable measuring distance from the locality to the "*cabecera municipal*" which is an indicator of distance to the governing center of the municipality (and likely the largest locality of the municipality). This may be taken to be an indicator of the availability of local labor markets. It may, nevertheless, have different impacts on both school and work. Closer available labor markets may make (paid) work more attractive and reduce schooling or, in fact, it may make school

---

[17] We have also estimated regressions separately by round and the estimated impacts were practically identical.
[18] Missing variable dummies are also included in the regressions for the cases in which data are not available (for instance, because the father no longer lives in the household).

more attractive by providing more information about the expected returns to schooling.[19] We

also include a variable measuring distance to the closest secondary school from the locality.

This provides an indicator of the cost of attending school and thus is likely to affect the relative

time spent in both school and work. Finally, the value of the discriminant score assigned to the

household by PROGRESA's beneficiary selection method is used as an additional explanatory

variable.

With this specification, given that we use only eligible households from treatment and

control villages, an estimate of the cross-sectional difference in the conditional mean of Y

between children in treatment and control communities, in the third and fifth rounds of the

survey is provided by the sum of the regression coefficients

$$\beta^{CSDIF}(R3=1) = \beta_0 + \beta_3 \qquad\qquad (6a)$$

$$\beta^{CSDIF}(R5=1) = \beta_0 + \beta_5 \qquad\qquad (6b)$$

Before going any further it is necessary to clarify that the extent to which our benchmark

experimental impact estimates β^{CSDIF} provide an estimate of the "true" program impact

depends on the quality of the randomization. At least two studies have investigated in detail

the extent to which the randomization was successfully implemented (Behrman and Todd,

1999; and Skoufias and Parker, 2001). A comparison of the means of key variables transformed

into locality means in control and treatment localities could not reject the hypothesis that the

means are equal suggesting that the randomization was quite successful at the locality level

(Behrman and Todd, 1999). However, some significant differences were detected when the

means of key variables were compared at the individual level. Skoufias and Parker (2001), for

---

[19] We do not attempt to construct at the individual level predicted wages for children given the
large number of children who do not work for an income.

example, noted that observed individual or household characteristics in the first pre-program round had some significant role towards predicting the assignment of an individual or a household into the treatment sample. For example, boys who attend school or who are working are more likely to be in the treatment sample than in the control sample. Also, boys (girls) whose father speaks Spanish are less (more) likely to be in the treatment (control) sample. Mindful of these considerations, we also report the double difference (2DIF) estimates of program impact in rounds 3 and 5. These are obtained directly from the coefficients β_3 and β_5 from regression (5) above. [20]

While the RDD estimation methods we employ are known as non-parametric estimators, they do depend on the choice of a kernel function and the bandwidth. In the paper we have chosen to report RDD estimates obtained with a bandwidth of 50. One of the kernel functions used is the uniform (or rectangular) kernel function which assigns equal weights to all observations falling within the band of +/- 50 discriminant points away from the region-specific threshold value and zero weight to observations outside the band (i.e. scores less or more than 50 points away from the region-specific threshold). Alternative kernel functions, such as the biweight, triangular, quartic, and Epanechnikov kernels, allow one to adjust the assignment of weights within the band accordingly by placing more weight on observations inside the band closer to the threshold and less weight on observations that are also inside the band but further away from the threshold (e.g. see Deaton, 1997). In order to examine the sensitivity of the RDD estimates, we also present estimates using these alternative kernel functions.

In addition, we also report RDD estimates using the Gaussian kernel that does not use a

---

[20] The coefficient β_0 in regression (5) provides and estimate of the pre-program differences that may exist between eligible households in treatment and control villages.

17

discrete band, but instead assigns some weight to each of the households below (or above) the threshold. Given the shape of the normal density, the Gaussian kernel ends up assigning higher weights to observations that are closer to the threshold value and very low weights to observations that are far away from the threshold. The standard errors of all the RDD estimates are estimated based on 500 bootstrapped samples. The standard errors reported for the CSDIF and 2DIF estimates are robust to heteroskedasticity by clustering at the locality level.

We begin with a brief discussion of the 2DIF and CSIDF estimates in tables 3a and 3b that use households in group B as a comparison group. The 2DIF estimates of the program impact on boys (table 3a) suggest that the program increased their school attendance but had no significant effect on their work activities. Specifically school attendance increased by 5 percent in round 3 and this increase was maintained at approximately the same level in round 5. It is also the case that program had a bigger impact on the schooling and work activities of girls. In round three the school attendance of girls from beneficiary households in the treated localities is 8.5 percent higher than that of similar girls in the control localities and this impact increases to 9.9 percent by round 5 (table 3b). At the same time the program seems to more than eliminate the pre-existing higher participation of poor girls in work activities. These estimates overall confirm the findings of earlier studies evaluating in more detail the impact of PROGRESA with the same data set but slightly different age groups (e.g. Schultz, 2003; Skoufias and Parker, 2001).

The absence of any significant pre-program differences in school attendance between beneficiary households in treatment and control villages, results in CSDIF estimates that are close to the 2DIF estimates of the program impact on the school attendance and work activities

of boys and the school attendance of girls. The significant differences between the CSDIF and 2DIF estimates of program impact on the work activities of girls can be attributed to the pre-existing differences in the work activities of girls in beneficiary households between treatment and control villages.

One complication involved in the comparison of the experimental and the RDD estimates relates to the possibility of the program having heterogeneous impacts. The RDD provides an estimate of the Local Average Treatment Effect (LATE) for the subgroup of individuals around the cut-off point, whereas the experimental 2DIF and CSDIF estimates discussed so far yield the Treatment of the Treated Effect (TTE) that is an average effect of the program on the treated population.[21] To control for possible differences between RDD and experimental estimates arising from heterogeneity in the impacts of the program we also present cross-sectional estimates of program impact based on the experimental nature of the sample by re-estimating equation (5) on the sub-sample of beneficiary households with a discriminant score within a range of 50 points from the threshold score.[22] These "local" experimental estimates, denoted by CSDIF-50 and presented in columns (3) of tables 3a and 3b, provide estimates of program impact on households that are close to the threshold score used for their region. A comparison of the estimates in column (3) with the CSDIF estimates in

---

[21] The TTE may be defined as $TTE = E(Y_1 \mid X, B = 1) - E(Y_0 \mid X, B = 1)$ where $E(Y_1 \mid X, B = 1)$ is what PROGRESA participants experience by participating in the program $E(Y_0 \mid X, B = 1)$ is the counterfactual term summarizing what PROGRESA participants would have experienced had they not participated in the program.

[22] This means that when we compare groups A and B in tables 3a and 3b (or groups C and D in table 4 below) we use the sub-sample of children in households with discriminant scores within a band of 50 points below (above) the threshold. When we compare groups B and C, groups B and D, and groups A and D as in tables 5-7, the local CSDIF estimates are based on the sub-sample of children in households with discriminant scores within a band of +/-50 points around the threshold. These are denoted by CSDIF+/-50.

column (2) reveals that there is some heterogeneity in the impacts of the program. Therefore a fair evaluation of the performance of the RDD should be based on a comparison of the RDD estimates with the local CSDIF-50 estimates of column (3).

A thorough inspection of the RDD estimates using children from households in group C (see table 2) as a comparison group yields a number of remarkable patterns. First of all, the RDD estimates confirm the absence of any pre-program (round 1) differences in the schooling attendance of either boys or girls that are also revealed by the CSDIF and CSDIF-50 estimates. Secondly, the RDD estimates also confirm the absence of a program impact on the work activities of boys and girls in the post-program rounds 3 and 5 (compare columns 4-9 with column 3 in tables 3a and 3b). There is only one instance where the RDD estimates suggest a significant difference between work activities in treatment and control households (girls in round 1 in table 3b) but this can be easily explained by the fact that this significant difference is obtained only with the Gaussian kernel. As mentioned earlier the Gaussian kernel does not use a discrete band but assigns some weight to all the observations below (or above) the threshold point. This feature makes the RDD estimates using the Gaussian kernel more comparable to the CSDIF estimates of column (2) than to the local CSDIF-50 of column (3). The similarity of the RDD estimates with the Gaussian kernel with the CSDIF estimates in column (2) combined with the absence of any significant pre-program differences using RDD with kernel function using a band suggest that much of the pre-program differences in the work activities of girls are due to differences among girls that are from households further away from the threshold.

Thirdly, the RDD estimates of the impact of the program on the school attendance of boys and girls in the third round of the survey (the first round after the start of the program benefits) suggest that the program had no significant impact. This is in sharp contrast to the significant program impact estimates obtained in the same round for both boys and girls using

20

CSDIF-50 (or CSDIF). However, in spite of the apparent poor performance of the RDD approach in the third round of the survey, in the fifth round of the survey the RDD estimates of program impact on the school attendance of boys and girls appear to be quite similar to those obtained by CSDIF. Specifically, in the fifth round the RDD estimates for boys are lower than the CSDIF-50 estimates, while the estimates for girls are practically identical to the local CSDIF estimates.

All in all, if one were to put aside, for the moment, the discrepancies observed in round 3, the performance of the RDD appears to be remarkably good. The RDD estimates of program impact agree with the "local" experimental estimates (CSDIF-50) in 10 out of the 12 possible instances. However, the apparent failure of the RDD method to detect any impact in round 3 for either boys or girls raises some serious concerns. For this reason, it is necessary to investigate in more depth some of the possible explanations for these discrepancies between the RDD and the experimental estimates.

Choice of Bandwidth and Inter-Regional Differences in the Threshold

One of the possible explanations may lie with the choice of bandwidth. The analysis so far presented estimates using a bandwidth of 50. Perhaps a different bandwidth may provide a more reliable estimate of the impact of the program. For this purpose we have also re-estimated the RDD estimates of program impact using a bandwidth of 75 and a bandwidth of 100 (see tables A1a, b and A2a, b in the appendix). However, increasing the bandwidth of the kernel functions appears to provide only a partial explanation of the weak performance of RDD. For example, with a bandwidth of 75 and 100 the third round estimates of program impact on the schooling attendance of boys continue to be insignificant. In contrast, the estimates for girls' school attendance now become significantly different from zero, albeit somewhat lower than the local experimental estimates. Also, the same general patterns continued to hold when we

repeat the analysis (with a bandwidth of 50) on the sub-sample of regions with very similar threshold scores, i.e. regions 3, 4, 5 and 6 (see appendix tables A3a, b).

Spillover Effects and Evaluation Bias

Another potential explanation for the observed patterns of impact obtained using the RDD method may be due to the inadequacy of the comparison group used by the RDD.[23] It is conceivable, for example, that after the start of the PROGRESA program in the treatment villages, non-eligible households in these villages may have altered their behavior by enrolling their children to school or withdrawing them from the labor market either due to "peer effects" within these small rural communities or due to expectations that this behavior may increase their chances of becoming eligible for the program's benefits. Whatever the reason, as long as non-eligible households change their behavior because of the presence of the program in their community they may cease to provide an appropriate comparison group for the evaluation of the program. This may also impact on the performance of the RDD relative to the benchmark experimental estimates.

In order to investigate this latter possibility in more detail, we conduct two tests based on two different comparison groups. Firstly, we examine whether the program had any impact on children from non-eligible households just above the threshold point in group C using as a comparison group children from non-eligible households that are also just above the threshold score in the control villages (group D in table 2). Secondly, we examine program impact on group C using as a comparison group children from group B. Since none of these groups of

---

[23] Bobonis and Finan (2002), for example, in their analysis of the spillover effects of PROGRESA, find that the program did have an impact on the non-eligible households, and that this impact occurred primarily in round 3 the first year after the start of the program.

households are benefiting from the program, the estimated impact of the program on group C should be zero independently of the comparison group used. Then, under the maintained hypothesis that the comparison group is totally unaffected by the presence of the program in surrounding localities, any evidence of a non-zero impact of the program in the sample of non-eligible households can be interpreted as evidence of spillover effects. This could also provide an explanation as to why the RDD estimates of using children from group C as a comparison group do not reveal any impact and suggest that the weak performance of the RDD may not be due to the method itself, but the compromised integrity of the comparison group.

Table 4 presents estimates of the program impact using 2DIF and CSDIF as well as local CSDIF estimates for a households with 50 and 75 points just above the threshold (CSDIF+50 and CSDIF+75). The CSDIF estimates appear to indicate that the program had some spillover effects on the school attendance of boys during the first year of its implementation. However, the 2DIF estimates for round 3 and the CSDIF estimates for round 1 combined suggest that these differences are more of a reflection of pre-existing differences between these groups of households than a significant increase in the school attendance of boys due to the presence of the program. Moreover these differences seem to disappear by the fifth round of the survey. Overall the estimates presented in table 4, suggest that spillover effects compromising the integrity of the comparison group could be a plausible explanation, at least for boys, for the poor performance of the RDD method during the October 1998 round of the survey.

Another possible explanation for the observed differences between RDD and experimental estimates in the PROGRESA sample is that the RDD estimator applied to the sample of non-eligible households in the treatment localities is subject to "evaluation bias". Borrowing the notation of Heckman, LaLonde, and Smith (1999), and ignoring for the moment the possibility that the RDD estimator may be a local estimate, the RDD estimate may be subject

to evaluation bias arising from the fact that the counterfactual term $E(Y_0 \mid X, B = 1)$ summarizing what PROGRESA participants would have experienced had they not participated in the program is approximated non-experimentally from the experience of households that are not eligible for the program. In the context of our sample, the evaluation bias associated with a non-experimental estimator such as the RDD may be defined as[24]

$$BIAS(X) = E(Y_0 \mid X, B = 1) - E(Y_0 \mid X, B = 0). \tag{7}$$

One advantage derived from the experimental nature of the evaluation sample of PROGRESA is that it offers the opportunity to investigate whether the size of this bias is significant. Following Smith and Todd (2001) we test for evaluation bias estimating program impact on the sample of non-eligible households in treatment localities (group C) using the sample of eligible households in the control villages (group B) as a comparison group (see table 5).

Irrespective of whether one uses CDIF or RDD as a method for evaluating program impact, the estimates in table 5 reveal that there is no significant impact of the program on non-eligible households near the threshold in the treatment localities. Thus evaluation bias cannot account for the weak performance of RDD in the earlier tables 3a and 3b. These estimates also suggest that there are no significant spillover effects compromising the integrity of the comparison group C in the third round of the survey.

Testing the integrity of the controls

Our investigation so far was conducted under the maintained hypothesis that the

---

[24] Note that by definition the bias associated with an experimental design is equal to zero since they use directly information from individuals or households in the control group to construct

control communities are immune from contamination. However, given that the PROGRESA program covered the control localities after November 1999, it is also conceivable that households in the control communities might alter their behavior in anticipation of coverage by the program.[25] If that were to be the case then the RDD estimates of program impact may be the ones that are closer to the true impact of the program instead of the experimental estimates used so far as the benchmark of "true" program impact. One way to assess the validity of maintained hypothesis that households in the control localities were unaffected by the presence of the program in neighboring villages is to focus on the control localities and look for possible impacts on households that would be eligible for program benefits in the case their villages were to be covered by the program. If anticipation effects were to contaminate the control communities then one might expect this to occur in the later rounds that are closer to the date of coverage of the control communities by PROGRESA.

Table 6 presents the 2DIF, CSDIF (average and local) and RDD estimates of program impact in the control localities using the eligible households as a treatment group and the non-eligible households from the same control localities as a comparison group (groups B and D, respectively). The estimates suggest that the existence of the program in nearby communities did not have any significant effect on the school attendance or work activities of boys and girls from eligible households. Moreover, it is important to note that both the RDD estimates which are local estimates of impact around the threshold as well as the CSDIF and 2DIF estimates, which are average estimates, yield the same general answer. Leaving aside some of the pre-

---

an estimate of the counterfactual term.
[25] It has been impossible to establish whether the households in the control localities knew in advance of their eventual coverage by the PROGRESA program. However, Attanasio, Meghir and Santiago(2001) provide evidence that such "announcement" effects induced in the PROGRESA control villages are important.

existing differences between the school attendance and work activities of eligible and non-eligible children, there does not appear to be any evidence of a significant differences between these households in the later rounds of the surveys.

Re-estimating program impacts with a different comparison group

Before making a final judgment on the performance of the RDD methods it is worthwhile to take advantage of one last comparison allowed by PROGRESA's evaluation design. The comparison group used by the RDD in Tables 3a and 3b consists of children in non-eligible households in the treatment villages with discriminant scores that are just above the region-specific threshold value. The various tests conducted so far were unable to reveal any significant problems with this comparison group or any significant differences between this group and the corresponding group (group D) in the control localities. If group C is indeed not affected by any spillover effects of the program then using group D as a comparison group in its place should also yield impact estimates for the third round that are similar to those presented in tables 3a and 3b. Any evidence to the contrary would imply that the weak performance of the RDD lies not with the method itself but with the comparison group.

In Table 7 we re-estimate the impact of the program using as a comparison group children from group D. The program impact estimates using 2DIF or CSDIF are remarkably close to the experimental impact estimates of tables 3a and 3b. For example, the original 2DIF estimates using group B as a comparison group suggest that the program increased the school attendance of boys (girls) by 5 (8.6) percent in round 3. The 2DIF estimates using group D as a comparison group suggest that the effect of the program in the same round was 6.8 (7.0) percent. Similarly, in the last round the original 2DIF estimates using group B as a comparison group suggest that the program increased the school attendance of boys (girls) by 4.8 (9.9)

percent. Also, the CSDIF estimates of program impact are remarkably close independently of the comparison group used. The change in the comparison group also seems to imply that the program had a significant effect in reducing the work activities of boys and girls.

However, some notable differences begin to emerge when comparing the local CSDIF estimates using the different comparison groups. The estimates of impact on schooling on households around PROGRESA's threshold are much higher when group D is used as a comparison group. For example, in the third round the program seems to have increased school attendance by 15.6 percent for boys and by 14.1 percent for girls around the threshold (see table 7). In contrast, the corresponding impact estimates are 7.2 (boys) and 7.8 (girls) percent. In addition, during the third round the RDD estimates of program impact on schooling (table 7) turn out to be high and significant for both boys and girls. This striking change of program impact with the RDD method attests that it is the comparison group rather than the method itself that is primarily responsible for the poor performance of the RDD in round 3 in tables 3a and 3b.

5. Conclusions

In this paper we investigated the performance of a quasi-experimental estimator, the Regression Discontinuity Design (RDD). Using as a benchmark the impact estimates based on the experimental nature of the sample, we examined how estimates differ when we use the RDD as the estimator for evaluating program impact on two key indicators: child school attendance and child work.

Overall the performance of the RDD was remarkably good. The RDD estimates of program impact agreed with the experimental estimates in 10 out of the 12 possible cases. The two cases in which the RDD method failed to reveal any significant program impact on the

school attendance of boys and girls were in the first year of the program (round 3). In this round the experimental methods detected significant program impacts for both boys and girls.

The nature of the PROGRESA sample allowed us to investigate more deeply three potential explanations for these discrepancies between the RDD and experimental methods. Specifically, we tested whether spillover effects contaminate the comparison group, whether the RDD estimator is subject to evaluation bias, and whether there are any contamination problems with the control group due to announcement effects. Although none of these tests were able to reveal any problems with the comparison group used, it did turn out that the RDD method was able to yield significant impact estimates in both post-program rounds, comparable to the experimental estimates, when we used as a comparison group children from non-eligible households in the control localities. In conclusion, it would be fair to say that the RDD approach, using only information on the outcome variables of interest, the household-specific discriminant score and the region-specific threshold values, is a valuable approach to evaluating program impacts as it has shown to generate estimates that are remarkably close to conventional experimental methods that require much richer data.

One additional key finding from our analysis is that the quality of the control group is very important. In combination, the variety of tests we conduct suggests that the reliability of the estimated program impact depends more on the integrity of the comparison group used and less on whether an experimental or quasi-experimental estimator is used to measure impact. In the case of PROGRESA, a quasi-experimental method such as the RDD would have yielded program impacts that were comparable to the "ideal" experimental impact estimates as long as the comparison group of households did not come from localities where the program was in operation. This finding has potentially critical implications since it implies that the evaluation of social programs at the national scale covering large segments of the poor may be extremely

difficult, if not impossible, to evaluate with quasi-experimental methods because of difficulties

in finding/constructing adequate comparison groups.

REFERENCES

Angrist, D. Joshua. and Victor Lavy (1999) "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement" *Quarterly Journal of Economics* 114(2), 533-575

Attanasio, Orazio, Costas Meghir, and Ana Santiago (2001) "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA," Unpublished Manuscript, University College London, UK.

Attanasio, Orazio, Costas Meghir, and Miguel Szekely (2003) "Using Randomized Experiments and Structural Models for 'Scaling Up'. Evidence from the PROGRESA Evaluation," Unpublished Manuscript, University College London, UK.

Baker, Judy (2000) "Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners" The World Bank, Washington D.C.

Behrman, Jere, and Petra Todd (1999) "Randomness in the Experimental Sample of PROGRESA (Education, Health, and Nutrition Program)" Mimeo, IFPRI. March.

Black, S.E. (1999) "Do 'Better' Schools Matter? Parental Valuation of Elementary Education" *Quarterly Journal of Economics* 114(2), 577-599

Blundell Richard, and Monica Costa-Dias (2002) "Alternative Approaches to Evaluation in Empirical Microeconomics," London: CEMMAP Working Paper CWP10/02.

Bobonis, Gustavo, and Frederico Finan (2002) "Do Transfers to the Poor Increase the Schooling of the Non-Poor: The Case of Mexico's PROGRESA Program, Manuscript, The University of California at Berkeley.

Deaton, A. 1997. *The Analysis of Household Survey Data*. Baltimore: Johns Hopkins University Press.

Dehejia, R. and S. Wahba (2002) "Propensity Score Matching Methods for Non-experimental Causal Studies" *The Review of Economics and Statistics*, Vol. 84 No. 1 (February), pp. 151-61.

Diaz, Juan-Jose, Sudhanshu Handa, and Monica Orozco (2003) "Estimating the Evaluation Bias of Matching Estimators using Randomized-out Controls and Non-Participants from PROGRESA" Working paper, Inter-American Development Bank.

DiNardo, J. and Lee, D.S. (2002) "The Impact of Unionization on Establishment Closure: A Regression Discontinuity Analysis of Representation Elections" NBER Working paper Series No. 8993, NBER, Cambridge

Fan, J. (1992) "Design-adaptive Nonparametric Regression" *Journal of the American Statistical Association* 87, 998-1004

Friedlander, Daniel, and Philip K. Robbins (1995) "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods" *American Economic Review* Vol. 85 No. 5 (September), pp. 923-937.

Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw (1999) "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design" NBER Working paper Series No. 7131, NBER, Cambridge.

Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw (2001) "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design" *Econometrica* 69(1), 201-209

Heckman, James J. 1992. Randomization and social policy evaluation. In *Evaluating welfare and training programs*, ed. C. Manski and I. Garfinkel. Cambridge, MA: Harvard University Press.

Heckman, James J., Hidehiko. Ichimura, and Petra Todd (1997) "Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review*

of Economic Studies, Vol. 64, pp. 605-654.

Heckman, James J. and Jeffrey Smith. 1995. Assessing the case for social experiments. *Journal of Economic Perspectives* 9(2) Spring: 85-110.

Heckman, James. , Robert La Londe, and Jeffrey. Smith. 1999. The economics and econometrics of active labor market programs. In *Handbook of labor economics*, vol. 3A, ed. O. Ashenfelter and D. Card. Amsterdam, The Netherlands: North Holland.

Lee, D.S. (2001) "The Electoral Advantage to Incumbency and Voter's Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the U.S. House" NBER Working paper Series No. 8441, NBER, Cambridge

Newman John., Laura Rawlings, and Paul Gertler (1994). "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *The World Bank Research Observer*, 9 (2): 181-201.

Pitt, Mark and Sahidur Khandker (1988) "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy*, Vol. 106, pp. 958-96.

Schultz, T. Paul (2003) "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program" *Journal of Development Economics*, (forthcoming).

Shadish, William, Thomas Cook, and Donald Campbell (2001): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin.

Sharma, Subbash (1996) *Applied Multivariate Techniques*, J. Wiley & Sons, Inc.

Skoufias, Emmanuel (2001) "PROGRESA and its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI" Mimeo, IFPRI. December.

Skoufias, Emmanuel, Benjamin Davis, and Sergio de la Vega (2001): "Targeting the Poor in Mexico: An Evaluation of the PROGRESA Selection Mechanism," *World Development*, Vol. 29, No. 10, October 2001, pp. 1769-1784.

Skoufias, Emmanuel and Susan W. Parker (2001) "Conditional Cash Transfers and their Impact on Child Work and Schooling: Evidence from the PROGRESA program in Mexico", *Economia,* Vol.2, No. 1, Fall 2001, pp. 45-96.

Smith, A. Jeffrey, and Petra E. Todd (2001) "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods, *American Economic Review Papers and Proceedings*, Vol. 91, No. 2, pp. 112-118.

Van der Klaauw, Wilbert. (2002) "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach", *International Economic Review*, Vol. 43, No. 4 (November), pp. 1249-87.

Figure 1: Kernel Densities of Discriminant Scores and Threshold points by region
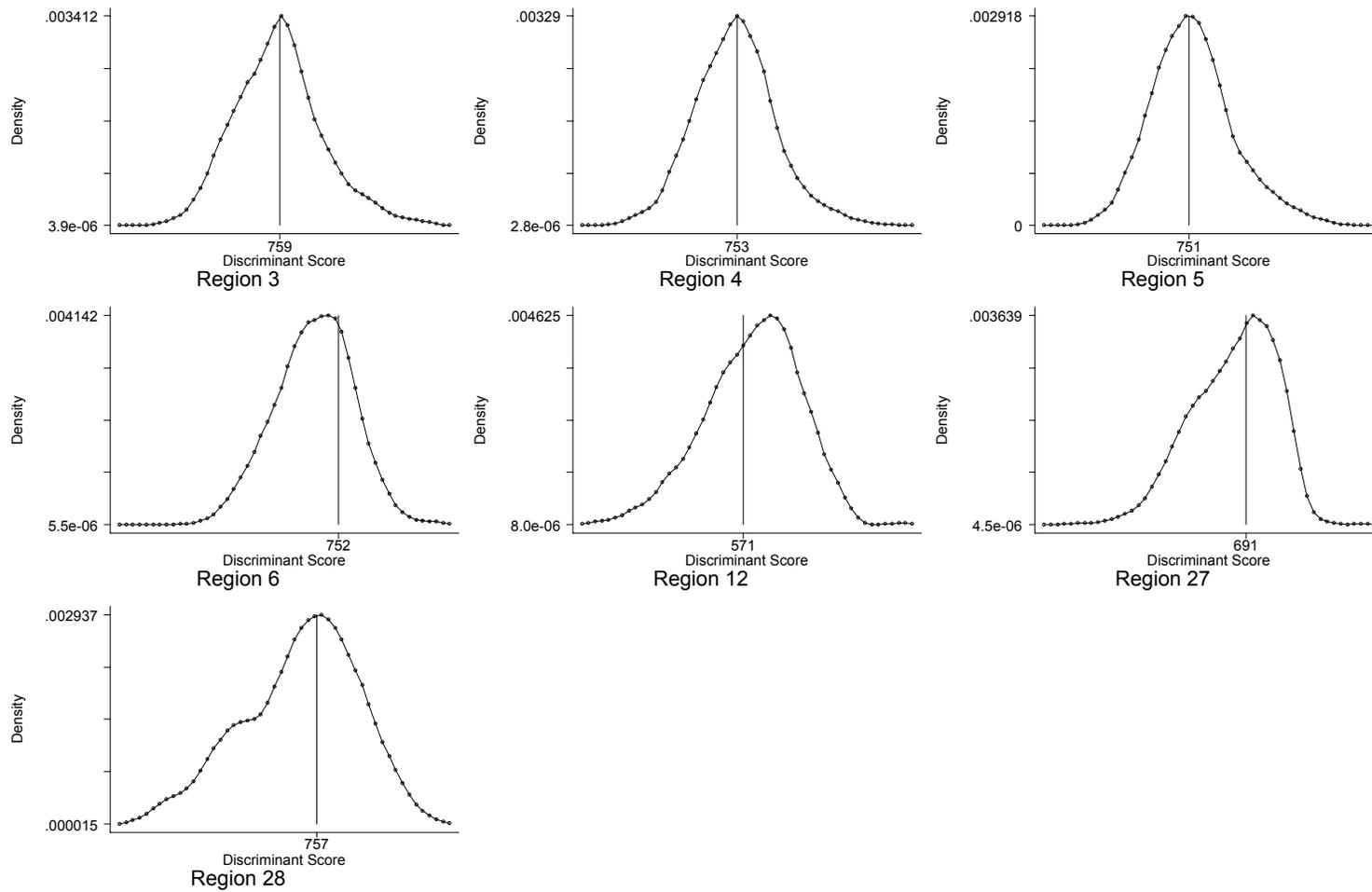
Table 1: Sampled Households and Discriminant Scores by Region

Region	(2) Number of households in sample	(3) Maximum value of discriminant score among eligible households	(4) Minimum value of discriminant score among non-eligible households	(5) Number of non-eligible households with discriminant score less than value in column (3)
Sierra Negra-Zongolica-Mazateca (code=3)	3,031	759.36	576	14
Sierra Norte-Otomi Tepehua (code=4)	4,559	753.14	653	15
Sierra Gorda (code=5)	10,790	751.5	610	14
Montana (Guerrero) (code=6)	1,907	752	693	3
Huasteca (San Luis Potosi) (code=12)	383	571	573	0
Tierra Caliente (Michoacan) (code=27)	2,934	691	546	213
Altiplano (San Luis Potosi) (code=28)*	472	856	757	116

Note: * In region 28 there were only 15 eligible households with a discriminant score greater than the minimum value of the score among non-eligible households

Table 2: A Decomposition of the Sample of All Households in Treatment and Control Villages

Household Eligibility Status	Discriminant Score (*'puntaje'*)	Localities: 320 Households:14,856 TREATMENT LOCALITY where PROGRESA is in operation ($T=1$)	Localities: 186 Households: 9,221 CONTROL LOCALITY where PROGRESA operations are delayed ($T=0$)
Eligible for PROGRESA benefits ($B=1$)	Low Below Threshold	**A** *B=1, T=1*	**B** *B=1, T=0*
Non-Eligible for PROGRESA benefits ($B=0$)	Above Threshold High	**C** *B=0, T=1*	**D** *B=0, T=0*

Estimates of Program Impact By Round (BOYS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-50	Uniform	Biweight	Epanechnik	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a	0.013	-0.001	-0.053	-0.016	-0.031	-0.018	-0.016	**-0.050**
st. error		0.018	0.028	0.027	0.031	0.029	0.031	0.031	0.021
Round 3	**0.050**	**0.064**	**0.071**	0.020	0.008	0.010	0.008	0.008	0.005
st. error	0.017	0.019	0.028	0.028	0.034	0.031	0.033	0.034	0.022
Round 5	**0.048**	**0.061**	**0.099**	0.052	**0.072**	**0.066**	**0.069**	**0.072**	**0.057**
st. error	0.020	0.019	0.030	0.028	0.032	0.030	0.032	0.032	0.021
Nobs	16331		4279						
R-Squared	0.21		0.25						
WORK									
Round 1	n.a.	0.018	0.007	0.012	-0.016	-0.004	-0.013	-0.016	0.025
st. error		0.019	0.029	0.027	0.032	0.029	0.031	0.032	0.021
Round 3	-0.037	-0.018	-0.007	0.007	-0.004	0.002	0.001	-0.004	0.005
st. error	0.023	0.017	0.029	0.024	0.028	0.026	0.028	0.028	0.019
Round 5	-0.046	-0.028	-0.037	-0.031	-0.029	-0.030	-0.029	-0.029	-0.028
st. error	0.025	0.017	0.025	0.024	0.028	0.026	0.027	0.028	0.019
Nobs	16331		4279						
R-Squared	0.16		0.19						

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group for Experimental & RDD Estimates: Beneficiary Households in Treatment Villages (Group A)

Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)

Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

TABLE 3b
Estimates of Program Impact By Round (GIRLS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-50	Uniform	Biweight	Epanechnik.	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a.	-0.001	0.000	-0.027	-0.025	-0.026	-0.025	-0.025	-0.035
st. error		0.020	0.030	0.029	0.036	0.033	0.034	0.036	0.023
Round 3	**0.086**	**0.085**	**0.082**	0.038	0.039	0.041	0.039	0.039	**0.054**
st. error	0.017	0.020	0.029	0.030	0.036	0.033	0.034	0.036	0.024
Round 5	**0.099**	**0.098**	**0.099**	**0.078**	**0.114**	**0.097**	**0.107**	**0.114**	**0.084**
st. error	0.020	0.019	0.028	0.031	0.036	0.033	0.035	0.036	0.025
Nobs		15046	3865						
R-Squared		0.22	0.23						
WORK									
Round 1	n.a.	**0.034**	0.000	0.033	0.026	0.027	0.027	0.026	**0.030**
st. error		0.017	0.024	0.019	0.022	0.020	0.021	0.022	0.015
Round 3	**-0.034**	0.000	0.001	0.005	0.001	0.003	0.002	0.001	-0.008
st. error	0.017	0.009	0.016	0.015	0.018	0.016	0.017	0.018	0.012
Round 5	**-0.042**	-0.008	-0.025	-0.019	-0.034	-0.029	-0.033	-0.034	-0.025
st. error	0.019	0.009	0.018	0.015	0.018	0.017	0.018	0.018	0.013
Nobs		15046	3865						
R-Squared		0.05	0.07						

NOTES:
Estimates in **bold** have t-values >=2
Treatment Group for Experimental & RDD Estimates: Beneficiary Households in Treatment Villages (Group A)
Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)
Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

TABLE 4

Program Impacts By Round on Non-Eligible Households in Treatment Localities using Group D as a comparison

	Experimental Estimates-Boys 12-16 yrs old				Experimental Estimates-Girls 12-16 yrs old			
	2DIF	CSDIF	CSDIF+50	CSDIF+75	2DIF	CSDIF	CSDIF+50	CSDIF+75
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Round 1	n.a	0.041	0.067	**0.063**	n.a.	0.033	0.043	0.037
st. error		0.022	0.034	0.030		0.022	0.035	0.030
Round 3	0.017	**0.058**	**0.071**	**0.067**	-0.013	0.020	0.041	0.006
st. error	0.020	0.024	0.033	0.030	0.021	0.023	0.036	0.029
Round 5	-0.021	0.020	0.065	0.035	-0.012	0.021	0.015	0.016
st. error	0.023	0.024	0.036	0.032	0.024	0.023	0.036	0.031
Nobs	7935		2762	3933	7314		2636	3694
R-Squared	0.24		0.25	0.25	0.23		0.23	0.24
WORK								
Round 1	n.a.	0.026	0.062	0.033	n.a.	0.006	-0.022	-0.008
st. error		0.022	0.032	0.028		0.014	0.026	0.020
Round 3	-0.033	-0.007	-0.010	0.001	-0.003	0.003	-0.026	-0.004
st. error	0.027	0.020	0.031	0.026	0.018	0.012	0.019	0.016
Round 5	-0.025	0.001	-0.009	0.004	0.005	0.011	0.017	0.031
st. error	0.028	0.019	0.032	0.027	0.020	0.013	0.021	0.020
Nobs	7935		2762	3933	7614		2636	3694
R-Squared	0.19		0.18	0.19	0.04		0.61	0.57

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group: Non-Eligible Households in Treatment Villages (Group C)

Comparison Group: Non-Eligible Households in Control Villages (Group D)

Table 5
Program Impacts By Round on Non-Eligible Households in Treatment Localities using Group B as a comparison

| | BOYS 12-16 yrs old | | | | | | GIRLS 12-16 yrs old | | | | | |
| | 2DIF | CSDIF | CSDIF+/-50 | RDD Uniform | RDD Triangular | RDD Gaussian | 2DIF | CSDIF | CSDIF+/-50 | RDD Uniform | RDD Triangular | RDD Gaussian |
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Round 1	n.a.	**0.081**	0.060	0.042	0.044	**0.055**	n.a.	0.048	-0.010	0.000	-0.034	0.021
st. error		0.029	0.051	0.031	0.036	0.026		0.029	0.049	0.034	0.038	0.027
Round 3	-0.001	**0.081**	0.049	0.024	0.033	**0.056**	-0.002	0.046	0.001	0.022	0.031	0.014
st. error	0.019	0.028	0.050	0.031	0.035	0.024	0.021	0.030	0.050	0.031	0.037	0.025
Round 5	-0.033	0.048	0.030	0.033	0.026	0.014	-0.023	0.025	-0.011	0.010	0.011	0.013
st. error	0.022	0.028	0.053	0.032	0.037	0.025	0.023	0.028	0.049	0.033	0.037	0.026
Nobs	10378		2950				9867		2971			
R-Squared	0.21		0.25				0.20		0.21			
WORK												
Round 1	n.a.	-0.014	-0.014	-0.002	0.004	-0.030	n.a.	0.001	-0.028	-0.016	-0.021	-0.012
st. error		0.025	0.039	0.030	0.034	0.023		0.018	0.030	0.020	0.023	0.016
Round 3	-0.015	-0.030	-0.023	0.002	-0.005	-0.019	0.007	0.007	0.000	0.007	0.005	0.013
st. error	0.026	0.023	0.038	0.028	0.031	0.022	0.018	0.014	0.026	0.015	0.017	0.012
Round 5	-0.015	-0.030	-0.001	0.001	-0.021	0.007	-0.004	-0.003	-0.006	-0.002	-0.003	0.004
st. error	0.028	0.021	0.037	0.026	0.030	0.021	0.020	0.016	0.031	0.017	0.021	0.015
Nobs	10378		2950				9867		2971			
R-Squared	0.16		0.19				0.04		0.06			

NOTES:
Estimates in **bold** have t-values >=2
Treatment Group: Non-Eligible Households in Tretament Villages (group C)
Comparison Group: Eligible Households in Control Villages (group B)

Table 6
Testing the Integrity of the Control Groups: Program Impacts on Eligible Households By Gender and by Round in the Control Villages

	BOYS 12-16 yrs old						GIRLS 12-16 yrs old					
	2DIF	CSDIF	CSIDF+/-50	RDD Uniform	RDD Triangular	RDD Gaussian	2DIF	CSDIF	CSIDF+/-50	RDD Uniform	RDD Triangular	RDD Gaussian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Round 1	n.a.	-0.028	0.016	0.030	0.014	0.009	n.a.	0.009	**0.104**	**0.070**	**0.113**	0.045
st. error		0.023	0.041	0.030	0.035	0.024		0.026	0.048	0.033	0.037	0.026
Round 3	0.017	-0.011	0.026	0.050	0.063	0.025	-0.011	-0.002	**0.092**	0.056	0.063	0.041
st. error	0.018	0.022	0.042	0.031	0.035	0.025	0.019	0.024	0.044	0.032	0.036	0.027
Round 5	0.013	-0.016	0.039	0.063	0.044	0.035	0.012	0.021	0.075	0.025	0.014	0.012
st. error	0.021	0.024	0.045	0.032	0.035	0.025	0.023	0.026	0.047	0.031	0.036	0.025
Nobs	9837		2986				9459		2937			
R-squared	0.21		0.25				0.21		0.23			
WORK												
Round 1	n.a.	**0.050**	0.068	**0.060**	0.062	**0.068**	n.a.	0.003	-0.022	-0.009	-0.022	-0.003
st. error		0.023	0.038	0.029	0.033	0.023		0.016	0.034	0.019	0.023	0.016
Round 3	-0.021	0.029	0.005	-0.018	-0.014	0.002	-0.008	-0.006	-0.049	-0.031	**-0.042**	-0.021
st. error	0.024	0.019	0.037	0.027	0.032	0.022	0.017	0.013	0.030	0.017	0.021	0.014
Round 5	-0.011	**0.039**	-0.014	-0.044	-0.022	-0.029	0.010	0.012	0.000	0.024	0.029	**0.029**
st. error	0.024	0.018	0.038	0.028	0.031	0.023	0.017	0.012	0.031	0.015	0.018	0.012
Nobs	9837		2986				9459		2937			
R-squared	0.16		0.18				0.04		0.08			

NOTES:
Estimates in **bold** have a t-value >=2
Treatment Group: Eligible Households in Control Villages (Group B)
Comparison Group: Non-Eligible Households in Control Villages (Group D)

Table 7

Estimates of Program Impact Using Non-Elligible Households in Control Villages as a Comparison Group

	BOYS 12-16 yrs old						GIRLS 12-16 yrs old					
	2DIF	CSDIF	CSDIF+/-50	RDD Uniform	RDD Triangular	RDD Gaussian	2DIF	CSDIF	CSDIF+/-50	RDD Uniform	RDD Triangular	RDD Gaussian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Round 1	n.a.	-0.002	**0.080**	0.020	0.039	0.014	n.a.	-0.011	0.080	0.044	0.053	0.031
st. error		0.024	0.040	0.028	0.031	0.022		0.026	0.046	0.029	0.033	0.023
Round 3	**0.068**	**0.066**	**0.161**	**0.094**	**0.105**	**0.086**	**0.070**	**0.059**	**0.142**	**0.117**	**0.132**	**0.108**
st. error	0.018	0.025	0.041	0.029	0.032	0.023	0.017	0.025	0.043	0.029	0.033	0.023
Round 5	**0.060**	**0.059**	**0.196**	**0.147**	**0.139**	**0.107**	**0.105**	**0.094**	**0.149**	**0.112**	**0.131**	**0.108**
st. error	0.021	0.025	0.041	0.027	0.030	0.022	0.020	0.026	0.046	0.032	0.037	0.024
Nobs	13888		4121				12793		3541			
R-Squared	0.22		0.25				0.23		0.24			
WORK												
Round 1	n.a.	**0.067**	0.026	**0.070**	0.053	**0.063**	n.a.	**0.031**	-0.045	0.009	-0.015	0.015
st. error		0.024	0.036	0.025	0.028	0.020		0.017	0.036	0.020	0.023	0.016
Round 3	**-0.059**	0.008	-0.057	-0.009	-0.018	-0.012	**-0.041**	-0.010	**-0.070**	-0.019	-0.034	-0.016
st. error	0.024	0.021	0.037	0.024	0.026	0.019	0.017	0.013	0.030	0.017	0.020	0.013
Round 5	**-0.059**	0.008	**-0.099**	**-0.074**	**-0.072**	**-0.050**	-0.032	-0.001	-0.048	0.004	-0.007	0.008
st. error	0.024	0.021	0.034	0.023	0.026	0.018	0.019	0.012	0.028	0.014	0.015	0.011
Nobs	13888		4121				12793		3541			
R-Squared	0.18		0.18				0.06		0.08			

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group: Beneficiary Households in Treatment Villages (group A)

Comparison Group: Non-Eligible Households in Control Villages (group D)

APPENDIX TABLES
for
An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA

In this appendix we present seven tables. The first table (table A) contains the means of all the variables used in our analysis by gender for groups, A, B, C, and D, separately. The other six tables can be compared with tables 3a and 3b in the body of the paper. Tables A.1a and A.1b re-estimate program impact using a bandwidth of 75 points. Tables A.2a and A.2b use a bandwidth of 100 points. The last two tables (tables A.3a and A.3b) use a bandwidth of 50 applied to the sub-sample of regions 3, 4, 5, and 6, where the threshold scores are practically the same.

	Group A		Group B		Group C		Group D	
	Boys	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**
Round 1 (October 1997)	N=3,301	N=2,941	N=1,952	N=1,863	N=1,563	N=1,378	N=1,326	N=1,265
Age=12 yrs (1=Yes 0=No)	0.24	0.24	0.25	0.23	0.19	0.20	0.19	0.17
Age=13 yrs (1=Yes 0=No)	0.21	0.22	0.21	0.23	0.19	0.18	0.20	0.20
Age=14 yrs (1=Yes 0=No)	0.22	0.20	0.21	0.21	0.19	0.19	0.17	0.20
Age=15 yrs (1=Yes 0=No)	0.19	0.19	0.19	0.19	0.19	0.21	0.21	0.21
Age=16 yrs (1=Yes 0=No)	0.14	0.15	0.14	0.15	0.24	0.21	0.22	0.21
Missing Mother Characteristics (1=Yes 0=No)	0.06	0.08	0.06	0.10	0.08	0.09	0.09	0.11
Mother speaks Indigenous language (1=Yes 0=No)	0.36	0.36	0.38	0.35	0.25	0.23	0.19	0.17
Mother speaks Spanish (1=Yes 0=No)	0.28	0.29	0.27	0.24	0.22	0.20	0.16	0.14
Mother's Age	37	36	37	36	40	39	40	39
Mother is Literate (1=Yes 0=No)	0.53	0.54	0.53	0.51	0.64	0.62	0.63	0.60
Mother completed primary School (1=Yes 0=No)	0.55	0.55	0.55	0.53	0.61	0.59	0.62	0.59
Mother completed secondary School (1=Yes 0=No)	0.01	0.02	0.02	0.02	0.05	0.05	0.04	0.03
Missing Father Characteristics (1=Yes 0=No)	0.14	0.14	0.13	0.16	0.15	0.17	0.16	0.19
Father speaks Indigenous language (1=Yes 0=No)	0.33	0.34	0.36	0.32	0.24	0.22	0.18	0.17
Father speaks Spanish (1=Yes 0=No)	0.30	0.30	0.34	0.30	0.23	0.21	0.17	0.16
Father's Age	38	38	39	37	41	40	40	39
Father is Literate (1=Yes 0=No)	0.60	0.60	0.59	0.61	0.68	0.68	0.66	0.63
Father completed primary School (1=Yes 0=No)	0.58	0.57	0.57	0.60	0.59	0.60	0.59	0.57
Father completed secondary School (1=Yes 0=No)	0.02	0.02	0.03	0.02	0.07	0.06	0.05	0.04
Discriminant Score assigned to household	633	622	625	627	851	851	839	841
Marginality Index	0.56	0.55	0.64	0.63	0.03	0.05	0.16	0.15
Distance of Municipality center	9.14	9.32	10.01	10.01	9.72	10.16	10.40	10.65
Distance from Secondary School	2.26	2.26	2.37	2.41	1.86	1.92	1.84	1.77
Children between 0 and 2 yrs of age	0.46	0.47	0.44	0.45	0.17	0.16	0.21	0.23
Children between 3 and 5 yrs of age	0.58	0.63	0.63	0.62	0.24	0.24	0.29	0.28
Boys between 6 and 7 yrs of age	0.26	0.25	0.27	0.24	0.09	0.11	0.12	0.11
Girls between 6 and 7 yrs of age	0.23	0.25	0.23	0.26	0.10	0.11	0.12	0.13
Boys between 8 and 12 yrs of age	0.94	0.70	0.94	0.69	0.59	0.42	0.63	0.40
Girls between 8 and 12 yrs of age	0.64	0.92	0.69	0.90	0.39	0.57	0.39	0.59
Boys between 13 and 18 yrs of age	1.36	0.58	1.30	0.56	1.43	0.61	1.41	0.66
Girls between 13 and 18 yrs of age	0.53	1.29	0.54	1.30	0.53	1.34	0.58	1.34
Males between 19 and 54 yrs of age	1.08	1.07	1.07	1.09	1.26	1.27	1.28	1.32
Females between 19 and 54 yrs of age	1.16	1.17	1.18	1.16	1.26	1.25	1.31	1.26
Males 55 yrs old or older	0.17	0.17	0.18	0.19	0.28	0.28	0.26	0.31
Females 55 yrs old or older	0.14	0.12	0.14	0.15	0.21	0.20	0.20	0.21
Region 3	0.11	0.10	0.13	0.15	0.14	0.13	0.16	0.12
Region 4	0.17	0.16	0.21	0.21	0.13	0.15	0.17	0.19
Region 5	0.41	0.43	0.41	0.39	0.54	0.53	0.49	0.51
Rgeion 6	0.13	0.12	0.08	0.07	0.05	0.05	0.01	0.02
Region 12	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Region 27	0.13	0.15	0.15	0.14	0.10	0.11	0.13	0.12
Region 28	0.03	0.03	0.01	0.01	0.02	0.03	0.01	0.01
Attending School? (1=Yes 0=No)	0.62	0.52	0.60	0.52	0.66	0.60	0.61	0.54
Working? (1=Yes 0=No)	0.33	0.12	0.32	0.09	0.31	0.10	0.30	0.09
Round 3 (November 1998)	N=3,454	N=3,169	N=2,056	N=1,968	N=1,422	N=1,442	N=1,230	N=1,202
Attending School? (1=Yes 0=No)	0.68	0.62	0.62	0.53	0.68	0.60	0.61	0.56
Working? (1=Yes 0=No)	0.21	0.06	0.23	0.06	0.21	0.08	0.23	0.07
Round 5 (November 1999)	N=3,436	N=3,080	N=2,132	N=2,025	N=1,280	N=1,191	N=1,141	N=1,136
Attending School? (1=Yes 0=No)	0.69	0.68	0.63	0.58	0.66	0.62	0.62	0.58
Working? (1=Yes 0=No)	0.18	0.05	0.21	0.06	0.18	0.06	0.20	0.05

Estimates of Program Impact By Round (BOYS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-75	Uniform	Biweight	Epanechnik	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a	0.013	0.004	**-0.063**	-0.042	**-0.053**	-0.042	-0.042	**-0.050**
st. error		0.018	0.024	0.022	0.027	0.024	0.026	0.027	0.019
Round 3	**0.050**	**0.064**	**0.086**	0.007	0.012	0.013	0.012	0.012	-0.001
st. error	0.017	0.019	0.025	0.023	0.027	0.025	0.026	0.027	0.019
Round 5	**0.048**	**0.061**	**0.102**	**0.051**	**0.063**	**0.059**	**0.062**	**0.063**	**0.053**
st. error	0.020	0.019	0.026	0.023	0.028	0.025	0.027	0.028	0.019
Nobs		16331	6198						
R-Squared		0.21	0.25						
WORK									
Round 1	n.a.	0.018	-0.009	0.049	0.010	0.024	0.013	0.010	0.028
st. error		0.019	0.025	0.022	0.026	0.024	0.025	0.026	0.019
Round 3	-0.037	-0.018	-0.028	0.006	0.002	0.004	0.003	0.002	0.006
st. error	0.023	0.017	0.025	0.020	0.024	0.022	0.023	0.024	0.017
Round 5	-0.046	-0.028	-0.039	-0.030	-0.035	-0.035	-0.034	-0.035	-0.020
st. error	0.025	0.017	0.022	0.020	0.023	0.021	0.022	0.023	0.016
Nobs		16331	6198						
R-Squared		0.16	0.20						

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group for Experimental & RDD Estimates: Eligible Households in Treatment Villages (Group A)

Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)

Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-75	Uniform	Biweight	Epanechnik.	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a.	-0.001	0.020	-0.035	-0.023	-0.025	-0.024	-0.023	**-0.043**
st. error		0.020	0.027	0.025	0.029	0.027	0.028	0.029	0.020
Round 3	**0.086**	**0.085**	**0.092**	**0.060**	0.050	**0.056**	0.052	0.050	**0.050**
st. error	0.017	0.020	0.026	0.024	0.029	0.027	0.029	0.029	0.020
Round 5	**0.099**	**0.098**	**0.108**	**0.076**	**0.092**	**0.085**	**0.092**	**0.092**	**0.079**
st. error	0.020	0.019	0.025	0.027	0.030	0.028	0.029	0.030	0.022
Nobs	15046		5554						
R-Squared	0.22		0.23						
WORK									
Round 1	n.a.	**0.034**	0.010	**0.034**	0.029	0.031	0.030	0.029	**0.029**
st. error		0.017	0.022	0.016	0.018	0.017	0.018	0.018	0.013
Round 3	**-0.034**	0.000	-0.012	-0.006	-0.001	-0.003	-0.002	-0.001	-0.014
st. error	0.017	0.009	0.014	0.013	0.015	0.014	0.014	0.015	0.011
Round 5	**-0.042**	-0.008	-0.029	-0.025	-0.029	-0.028	**-0.030**	-0.029	-0.022
st. error	0.019	0.009	0.015	0.014	0.015	0.015	0.015	0.015	0.012
Nobs	15046		5554						
R-Squared	0.05		0.06						

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group for Experimental & RDD Estimates: Eligible Households in Treatment Villages (Group A)

Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)

Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

APPENDIX TABLE A.2a--Bandwidth=100
Estimates of Program Impact By Round (BOYS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-100	Uniform	Biweight	Epanechnik	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a	0.013	0.006	**-0.051**	**-0.052**	**-0.055**	**-0.049**	**-0.052**	**-0.049**
st. error		0.018	0.023	0.020	0.023	0.022	0.023	0.023	0.017
Round 3	**0.050**	**0.064**	**0.090**	-0.007	0.009	0.004	0.006	0.009	-0.005
st. error	0.017	0.019	0.023	0.021	0.024	0.023	0.023	0.024	0.018
Round 5	**0.048**	**0.061**	**0.080**	**0.059**	**0.058**	**0.056**	**0.059**	**0.058**	**0.049**
st. error	0.020	0.019	0.023	0.021	0.024	0.022	0.023	0.024	0.018
Nobs		16331	7937						
R-Squared		0.21	0.24						
WORK									
Round 1	n.a.	0.018	-0.004	0.023	0.025	0.030	0.022	0.025	0.027
st. error		0.019	0.024	0.020	0.023	0.021	0.022	0.023	0.018
Round 3	-0.037	-0.018	-0.028	0.012	0.005	0.007	0.006	0.005	0.007
st. error	0.023	0.017	0.022	0.018	0.021	0.020	0.020	0.021	0.016
Round 5	-0.046	-0.028	-0.020	-0.023	-0.033	-0.029	-0.030	-0.033	-0.015
st. error	0.025	0.017	0.020	0.017	0.020	0.019	0.020	0.020	0.015
Nobs		16331	7937						
R-Squared		0.16	0.19						

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group for Experimental & RDD Estimates: Eligible Households in Treatment Villages (Group A)
Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)
Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

Estimates of Program Impact By Round (GIRLS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-100	Uniform	Biweight	Epanechnik.	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a.	-0.001	0.016	**-0.048**	-0.029	-0.034	-0.031	-0.029	**-0.049**
st. error		0.020	0.025	0.022	0.025	0.023	0.025	0.025	0.019
Round 3	**0.086**	**0.085**	**0.081**	**0.054**	**0.056**	**0.060**	**0.056**	**0.056**	**0.045**
st. error	0.017	0.020	0.024	0.022	0.026	0.024	0.025	0.026	0.018
Round 5	**0.099**	**0.098**	**0.096**	**0.081**	**0.084**	**0.083**	**0.087**	**0.084**	**0.076**
st. error	0.020	0.019	0.025	0.024	0.027	0.026	0.027	0.027	0.021
Nobs		15046	7221						
R-Squared		0.22	0.24						
WORK									
Round 1	n.a.	**0.034**	0.012	**0.033**	0.030	**0.030**	0.030	0.030	**0.028**
st. error		0.017	0.020	0.014	0.016	0.015	0.016	0.016	0.012
Round 3	**-0.034**	0.000	-0.007	-0.014	-0.005	-0.008	-0.006	-0.005	-0.016
st. error	0.017	0.009	0.012	0.012	0.013	0.012	0.013	0.013	0.010
Round 5	**-0.042**	-0.008	**-0.028**	**-0.026**	-0.027	**-0.026**	-0.027	-0.027	-0.019
st. error	0.019	0.009	0.014	0.013	0.014	0.013	0.014	0.014	0.011
Nobs		15046	7221						
R-Squared		0.05	0.06						

NOTES:

Estimates in **bold** have t-values >=2

Treatment Group for Experimental & RDD Estimates: Eligible Households in Treatment Villages (Group A)

Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)

Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

APPENDIX TABLE A.3a--Regions 3,4,5 & 6, Bandwisth=50
Estimates of Program Impact By Round (BOYS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-50	Uniform	Biweight	Epanechnik	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1-coeff	n.a	0.013	0.011	**-0.067**	-0.047	-0.055	-0.050	-0.047	**-0.060**
st. error		0.020	0.031	0.030	0.037	0.034	0.036	0.037	0.023
Round 3-coeff	**0.042**	**0.054**	**0.070**	-0.001	-0.023	-0.017	-0.022	-0.023	-0.013
st. error	0.019	0.021	0.030	0.032	0.036	0.034	0.035	0.036	0.024
Round 5	**0.042**	**0.055**	**0.117**	**0.062**	0.070	0.066	**0.068**	0.070	**0.057**
st. error	0.022	0.021	0.033	0.031	0.038	0.034	0.037	0.038	0.023
Nobs		13509	3529						
R-Squared		0.42	0.24						
WORK									
Round 1	n.a.	0.020	0.006	0.032	0.008	0.019	0.010	0.008	0.038
st. error		0.022	0.032	0.030	0.036	0.033	0.034	0.036	0.023
Round 3	-0.037	-0.017	-0.001	0.016	0.020	0.020	0.022	0.020	0.009
st. error	0.026	0.019	0.031	0.026	0.031	0.028	0.030	0.031	0.020
Round 5	-0.046	-0.027	-0.033	-0.043	-0.049	-0.047	-0.048	-0.049	-0.037
st. error	0.028	0.018	0.027	0.027	0.034	0.031	0.032	0.034	0.021
Nobs		13509	3529						
R-Squared		0.39	0.20						

NOTES:

Estimates in **bold** have t-values >=2

Treatment group for Experimental & RDD Estimates: Eligible Households in Tretament Villages (Group A)
Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)
Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)

APPENDIX TABLE A.3b--Regions 3,4,5 & 6, Bandwisth=50
Estimates of Program Impact By Round (BOYS 12-16 yrs old)

	Experimental Estimates			RDD Impact Estimates using different kernel functions					
	2DIF	CSDIF	CSDIF-50	Uniform	Biweight	Epanechnik.	Triangular	Quartic	Guassian
SCHOOL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Round 1	n.a.	0.012	0.023	-0.003	-0.002	-0.003	-0.003	-0.002	-0.011
st. error		0.022	0.033	0.035	0.042	0.038	0.041	0.042	0.028
Round 3	**0.066**	**0.079**	**0.079**	0.034	0.032	0.032	0.032	0.032	**0.054**
st. error	0.018	0.022	0.032	0.031	0.037	0.034	0.035	0.037	0.024
Round 5	**0.087**	**0.099**	**0.121**	0.059	**0.093**	**0.075**	**0.086**	**0.093**	**0.072**
st. error	0.022	0.021	0.032	0.032	0.038	0.034	0.037	0.038	0.026
Nobs		12344	3248						
R-Squared		0.44	0.22						
WORK									
Round 1	n.a.	0.034	0.011	0.025	0.016	0.018	0.017	0.016	0.023
st. error		0.019	0.027	0.022	0.026	0.024	0.026	0.026	0.017
Round 3	-0.035	-0.002	-0.001	0.008	0.005	0.008	0.006	0.005	-0.005
st. error	0.020	0.010	0.018	0.016	0.020	0.017	0.019	0.020	0.014
Round 5	-0.043	-0.009	-0.028	-0.024	**-0.044**	-0.039	**-0.042**	**-0.044**	**-0.030**
st. error	0.023	0.011	0.020	0.018	0.022	0.020	0.021	0.022	0.014
Nobs		12344	3248						
R-Squared		0.26	0.07						

NOTES:

Estimates in **bold** have t-values >=2

Treatment group for Experimental & RDD Estimates: Eligible Households in Tretament Villages (Group A)

Comparison Group for Experimental Estimates: Eligible Households in Control Villages (Group B)

Comparison Group for RDD Estimates: NonEligible Households in Treatment Villages (Group C)