# Do Nonpartisan Programmatic Policies Have Partisan Electoral Effects? Evidence from Two Large Scale Experiments*

Kosuke Imai† Gary King‡ Carlos Velasco Rivera§

June 13, 2018

**Abstract**

A vast literature demonstrates that voters around the world who benefit from their governments' discretionary spending cast more ballots for the incumbent party than those who do not benefit. But contrary to most theories of political accountability, some suggest that voters also reward incumbent parties for implementing "programmatic" spending legislation, over which incumbents have no discretion, and even when passed with support from all major parties. Why voters would attribute responsibility when none exists is unclear, as is why minority party legislators would approve of legislation that would cost them votes. We study the electoral effects of two large prominent programmatic policies that fit the ideal type especially well, with unusually large scale experiments that bring more evidence to bear on this question than has previously been possible. For the first policy, we design and implement ourselves one of the largest randomized social experiments ever. For the second policy, we reanalyze studies that used a large scale randomized experiment and a natural experiment to study the same question but came to opposite conclusions. Using corrected data and improved statistical methods, we show that the evidence from all analyses of both policies is consistent: programmatic policies have no effect on voter support for incumbents. We conclude by discussing how the many other studies in the literature may be interpreted in light of our results.

*We are grateful to Tina Green and Ana De La O for data and replication information, and for help with followup questions; to Isadora Antoniano (formerly of the Mexican Electoral Institute), Miguel Rojano (Director of Electoral Cartography) and Luis Ruvalcaba (Deputy Director of Geographical Electoral Systems Development) for discussing with us the intricacies of electoral cartography in Mexico; to Wangyal Shawa for his helpful GIS technical advice; and to Rikhil Bhavnani, Graeme Blair, Chris Blattman, Ben Fifield, Miriam Golden, Ken Greene, Guy Grossman, Macartan Humphreys, John Londregan, Gabriel López-Moctezuma, Will Lowe, Horacio Larreguy, John Marshall, Grigore Pop-Eleches, Jake Shapiro, Tara Slough, and Cesar Zucco Jr. for helpful comments.

†Professor, Department of Politics and Center for Statistics and Machine Learning, Princeton University, Princeton NJ 08540; http://imai.princeton.edu, kimai@princeton.edu, (609) 258-6601.

‡Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge MA 02138; GaryKing.org, King@Harvard.edu, (617) 500-7570.

§Research Fellow, Institute for Advanced Study in Toulouse, 21 allée de Brienne, 31000 Toulouse, France; http://www.cvelasco.org, carlos.velasco@iast.fr.

# 1 Introduction

Political scientists in American and comparative politics have amassed considerable support for the theory that office-holders target *discretionary* government spending to gain votes, and voters reward them for doing so. As the magisterial literature review by Golden and Min (2013, p.12) summarizes: "Studies overwhelmingly find that incumbent politicians are rewarded by voters for distributive allocations, and in particular for those that are clientelistic and from which recipients can be excluded." Yet, perhaps paradoxically, some argue that voters react in the same way to *programmatic policies*, for which incumbents have little or no discretion in delivering benefits because citizens receive services based on well known, publicly stated rules (see Hicken, 2011; Kitschelt and Wilkinson, 2007; Stokes, Dunning, Nazareno, and Brusco, 2013, for a definition). This is all the more puzzling in those situations where programmatic policies are passed with support from every major political party, including those who will be hurt electorally by their very action.

Although important efforts have been made to address the "programmatic incumbent support hypothesis" via qualitative argument (Cornelius, 2004; Diaz-Cayeros, Estevez, and Magaloni, 2009, 2016), and via the possibility of publication bias (Golden and Min, 2013), neither the formal theoretical nor quantitative empirical literatures has offered a resolution. We address this puzzle by analyzing two large programmatic policies, each designed to reduce poverty and its effects — *Seguro Popular de Salud* (SPS) and *Progresa*. Studying SPS and *Progresa* together is also advantageous because the literature views the two as highly visible, highly impactful, programmatic policies passed and implemented in manner close to the pure theoretical form (Camp, 2013; Diaz-Cayeros, Estevez, and Magaloni, 2009, 2016).

We are fortunate in being able to study the electoral effects of these policies with two very large scale randomized experiments and a natural experiment, which in total what appears to be the best evidence ever brought to bear on this question. The first, which we designed and implemented originally for the purpose of evaluating SPS in Mexico, is one of the largest social experiments ever conducted, and the largest randomized health policy experiment to date (King, Gakidou, Ravishankar, Moore, Lakin, Vargas, Téllez-Rojo, J. E. H. Ávila, M. H. Ávila, and Llamas, 2007). The second, which includes the only randomized social experiment previously used to study the electoral effects of programmatic policies along with a novel natural experiment, left its authors' conclusions in disagreement (De La O, 2013, 2015; Green, 2006). Like the first experiment, it was also very large, originally designed as a policy evaluation, and in Mexico.

We thus have two cases relatively close to the theoretical ideal for a programmatic policy. We are able to analyze each of these policies with a unusually strong experimental frameworks. Our results resolve some disagreements in the literature and wind up not supporting the programmatic incumbent support hypothesis. We show that neither SPS nor *Progresa* has a causal effect on voter turnout or electoral support for the incumbent party.

Sections 2 and 3 describe our two experiments. Section 4 reinterprets the literature in light of our results. We also provide an appendix regarding some coding and analysis issues that led to prior disagreements in the literature, and also an extensive online Supplementary Appendix with supporting technical information, measurement issues, statistical analyses, robustness checks, alternative measurement strategies, analyses of a formal theory, and other details.

# 2 Experiment 1: *Seguro Popular de Salud*

## 2.1 Background

Although *Seguro Popular de Salud* (SPS) translates literally to "universal health insurance," the Spanish word for "insurance" does not appear in the authorizing legislation, as it is a social welfare (income redistribution) program, not a self-sustaining insurance program. The program was designed to build or improve medical facilities, and provide medical services, preventive care, pharmaceuticals, and financial health protection to the 50 million Mexicans with no regular access to health care (constituting about half the population of the country). SPS was aimed at those with low incomes; its main purpose is to reduce the devastating effects of catastrophic health expenditures, when, due to illness or injury, greater than 30% of a family's annual disposable income is spent on health care in one year. Before SPS, about 10% of the poor had catastrophic health expenditures each year. SPS was designed to eventually spend an additional one percentage point in GDP; in 2005, expenditures totaled a substantial $795.5 million. As it turned out, SPS was the most visible accomplishment of the Vicente Fox Quesada administration. By all accounts, it was designed, passed, and implemented in a nonpartisan, programmatic fashion.[1]

Because of Mexico's term limits, President Fox, along with Health Minister Julio Frenk Mora, decided they needed a way to convince whoever would succeed them to keep SPS in place. How one democratically elected government can "tie the hands" of the next democratically elected government is a classic question of normative political theory (Klarman, 1997; Posner and Vermeule, 2002; Sterk, 2003; Thompson, 2005), formal theory (Alesina and Tabellini, 1990), and empirical political science (Franzese, 2002). In this case, Fox and Frenk's approach was to commission an independent scientific evaluation by a team we led at Harvard's Institute for Quantitative Social Science (see `http://j.mp/ExpMex`). The idea was that if the evaluation favored SPS, it would at least have been more difficult for the next government to eliminate. And to make the evaluation worthwhile, we committed publicly and in print to say so if the evaluation was not favorable in any way.

The Mexican government signed legal contracts that gave us free rein to design and implement whatever evaluation, and spend whatever funds, we judged appropriate. The government gave us unfettered access to government officials, the ability to influence how SPS was implemented so we could more easily evaluate it, and convening power to speak with the numerous local officials across the country in charge of implementation. We retained the legal right to publish without prior review.

We developed a new experimental research design robust to the interventions by politi-

[1]Julio Frenk, one of SPS's main architects said "From the beginning I suggested that this [SPS] had to be a proposal from all parties, in other words, that it was not only a PAN, PRI, or PRD project, but a project for the country, for which everyone could claim credit. In a conversation with President Fox, after introducing the bill to Senate, I told him: 'If we want the reform to move forward, it is extremely important that the project is handled as a shared project and that we give credit to everyone.' The president agreed, and promised that credit would be given to everyone, which is what happened" (Ortiz, 2006, p.81, our translation). Whether we judge by public statements from minority parties, or from the votes of legislation, there exists near consensus on this point.

cians who regularly — indeed usually — derail large public policy experiments, as they choose to be more attentive to the short term desires of their constituents than any longer term benefit of scientific evaluation (King, Gakidou, Ravishankar, Moore, Lakin, Vargas, Téllez-Rojo, J. E. H. Ávila, M. H. Ávila, and Llamas, 2007). Details of the experiment include discussions of the background and design, published prior to any data analysis (King, Gakidou, Ravishankar, Moore, Lakin, Vargas, Téllez-Rojo, J. E. H. Ávila, M. H. Ávila, and Llamas, 2007), novel statistical methods we developed for this design (Imai, King, and Nall, 2009b), empirical results (King, Gakidou, Imai, Lakin, Nall, Moore, Ravishankar, Vargas, Téllez-Rojo, J. E. H. Ávila, M. H. Ávila, and Llamas, 2009a), and publicly available replication data (Imai, King, and Nall, 2009a; King, Gakidou, Imai, Lakin, Nall, Moore, Ravishankar, Vargas, Téllez-Rojo, J. E. H. Ávila, M. H. Ávila, and Llamas, 2009b).

We began by defining 12,284 *health clusters*, which are new, continuous geographic areas we defined that tile all of Mexico's 31 states in which fall one health clinic (or potential clinic that we could decide would be built) along with its catchment area (defined as less than a day's travel time to the clinic, using locally available methods of transportation rather than "as the crow flies"). We recruited 13 of Mexico's 31 states to participate, including 7,078 health clusters. We matched these health clusters in pairs based on background characteristics and then selected 74 pairs (based on the closeness of the match, likelihood of compliance with the experiment, and necessary political criteria). The experiment thus included 534,457 research subjects, in 118,569 households, within 1,380 localities, nested within 148 health clusters.

Within each of the 74 matched pairs, we randomly selected one health cluster to receive treatment and the other, as the control, to receive no change. "Treatment" included new (or upgraded) hospitals and other medical facilities, doctors, access to medicines and other medical interventions, advertising campaigns to encourage affiliation with the program, individual insurance, and funds to pay for it all. The treatment was applied August-September, 2005, coincident with a baseline survey of 32,515 respondents randomly selected from 50 of the health cluster pairs; the outcome was measured the same way 10 months later, July-August 2006. We measured an extensive array of variables including individual opinions, attitudes, health status, and financial spending; household level variables, such as assets, wealth, demographics, and others; and physical health measures via three separate blood draws for each person.

The 2006 Mexican presidential election was held on July 2, coinciding with the start of the follow-up survey, which happens to be perfect timing for studying the effects of SPS on the election (see Figure 4 in the online Supplemental Appendix). To merge federal election results reported by one governmental office, with census data reported by another, we define a new unit of analysis for this study within the confines of the experimental randomization. This is the *precinct cluster*, which we define as the largest possible geographic subset of a single health cluster in the SPS experiment for which we can accurately merge all relevant electoral and census information (for details, see the Supplementary Appendix, Section 2). When insufficient information is available to identify a precinct cluster, we retained all the benefits of this matched pair cluster randomized design by removing it and the health cluster (and corresponding precinct cluster) to which it was pair matched (as suggested by Imai, King, and Nall, 2009b).[2] This left us in the end with 57 matched-pairs (47 rural and 10 urban) out of the original 74 and all the benefits of matched pair randomization. Table 1 and Figure 7 in the Supplemental Appendix present full details and descriptive statistics of the sample we use and to which our inferences apply. (As another robustness check, we repeated our analyses with all available precinct clusters even when no match was available, trading off more model dependence for less inefficiency, and found no substantive change in the results we present below. See Tables 2 and 3 in the Supplementary Appendix.)

Especially useful for the present paper is that our evaluation indicated that SPS had a massive effect on its intended outcome variable of financial assistance to the poor — a fact which is all the more impressive because government programs designed to help the poor in most countries typically have no measurable impact on the poor (Gwatkin, Wagstaff, and Yazbeck, 2005). In only 10 months, SPS eliminated about a *third* of the catastrophic health expenditure problem among the poor in Mexico, and about 60% among experi-

[2]This decision is not affected by which unit received the treatment within a given pair.

mental compliers (those who would affiliate to SPS when in the treatment group and not when in the control group). Perhaps even more important for our purposes, the people of Mexico clearly thought the program would help them: Fully 44% of those eligible in treatment areas enrolled in the program in the first month, which for each family involved taking a trip that could last as long a full day to formally affiliate to SPS. Moreover, those who enrolled liked the program a great deal: 69% of those enrolled rated the quality of health services as good or very good, and 97% planned to enroll again in the follow-up period after the experiment.

Some aspects of the program did not have the intended effects, as we reported (King, Gakidou, Imai, Lakin, Nall, Moore, Ravishankar, Vargas, Téllez-Rojo, J. E. H. Ávila, M. H. Ávila, and Llamas, 2009a), but for our purposes this large financial impact makes this an unusually strong test of the electoral effects of programmatic policies.

## 2.2 Results

We present two sets of outcome variables for the SPS experiment based on analyses of actual electoral results (using aggregation procedures described in the Supplementary Appendix, Section 2) and on retrospective survey evaluations (which require no aggregation), respectively. The statistical methods used in both cases are fully nonparametric, enabling us to reap the benefits from our randomized experiment without making modeling assumptions (for a full description, see Imai, King, and Nall, 2009b).

First, we estimate the total causal effect of SPS on electoral outcomes in Figure 1. (This analysis and all other types of causal effects on electoral outcomes are based on the 57 precinct matched cluster pairs, resulting from mapping the 74 evaluation health cluster pairs into precincts.) A point estimate (black dot) and 95% confidence interval (vertical line) appear for each causal effect, with a horizontal line at zero, indicating no effect. If the literature's hypothesis is correct, that nonpartisan programmatic policies increase support for incumbents, the points and confidence intervals would appear above the horizontal line. Instead, all the confidence intervals cross the zero, no effect, line, and thus none confirm the hypothesis.

The estimated total causal effect of SPS on the percent voting for the incumbent party

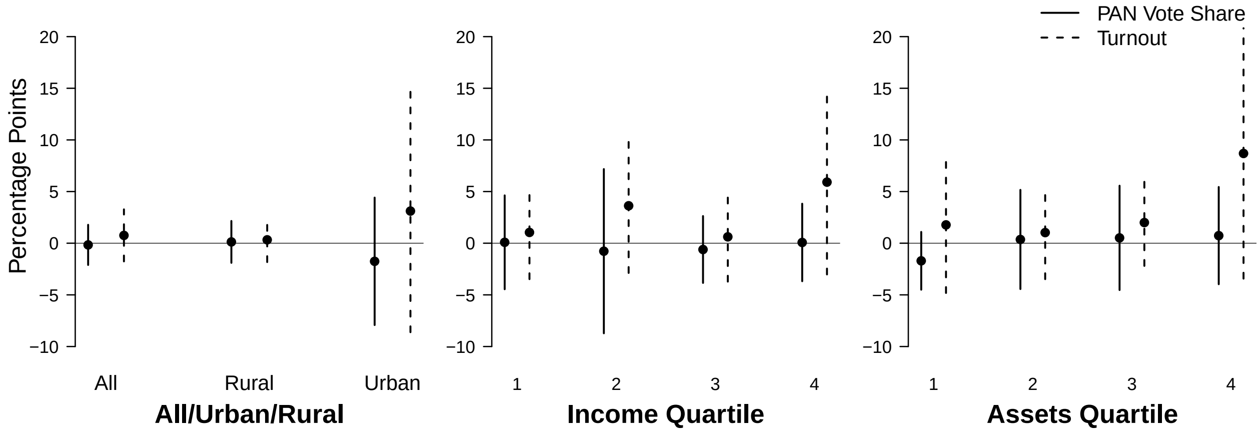


Figure 1: **Intention to Treat Estimates of SPS Effect on Voter Turnout and Incumbent Party Vote.** This figure gives point estimates and 95% confidence intervals for the total causal effect of SPS on voter turnout (solid line) and the incumbent (PAN) vote share (dashed line), overall (far left of left panel), and in three partitions of the data, including urban/rural (left panel), income quartile (middle), and asset quartile (right).

(the PAN) is the first result, at the far left of Figure 1 (solid vertical line); the estimated effect on federal voter turnout appears next to it (dashed vertical line). In both cases, the estimated total causal effect is not distinguishable from zero. Indeed, the confidence intervals are quite narrow for the estimated overall effects, indicating that the experiment is well powered for this hypothesis and so we should have high confidence that the effect is negligible if not exactly zero.

To further search for possible support for the hypothesis, we partition our sample in three different ways to examine subgroup effects. We give the total causal effect within rural and within urban precinct clusters at the right side of the left panel. All results are statistically indistinguishable from zero. The right two panels estimate total causal effects by quartile of the proportion of individuals in the first two income deciles (center) and household assets (right). These may be especially relevant since households that report less income will pay less (or nothing if in the first two income declines) for SPS services. These results can also be viewed as predicting compliance with receiving the experimental treatment, since less poor communities were expected to (and actually did) sign up for SPS services less often. This analysis follows Calvo and Murillo (2004), whose account suggests that the largest electoral effect of policies could be found among individuals who benefit most from a given policy. Nevertheless, the estimated causal effect of SPS on voter turnout and on incumbent party vote within every one of these segments of the public is not distinguishable from zero.

Second, we avoid all issues involved in aggregating precincts to precinct clusters (discussed in the Supplementary Appendix, Section 2) by estimating the total causal effect of SPS on individual survey evaluations. (This analysis, and all causal estimates on individual evaluations, are based on the 50 matched health clusters participating in the survey.) Recall that the beginning of our follow-up survey coincides with the election and hence we are measuring voters' opinion right after the election. We do this for how well respondents thought the country was doing on economic, political, and social issues, compared to five years before (we provide descriptive statistics in Figure 8 of our Supplemental Appendix).[3] This analysis also allows us to test the channels through which retrospective voting may work, one of the mechanisms hypothesized for why incumbents may benefit from programmatic politics (C. Pop-Eleches and G. Pop-Eleches, 2012).

These results appear in Figure 2 (in a format parallel to Figure 1). For the overall effect (at the far left) and for each of the three domain areas (economic as a solid line, political as dashed, and social as dotted), the effect is estimated at near zero with a confidence interval that overlaps zero. We find the same essentially zero effect of SPS, and failure of the programmatic policy hypothesis, for individual effects within different subgroups, including rural, urban (left panel), income quartile (center panel), and asset quartile (right panel).

In our Supplementary Appendix, we also present a wide array of other analyses of the same data. For example, we reran the analysis from Figures 1 and 2 with different coding rules and statistical techniques. These and many other analyses reveal no noticeable effect. (To be specific, in the Supplementary Appendix, Figure 9 analyzes PAN votes as a share of registered and eligible voters, as alternative measures of incumbent support. Figure 10 measures turnout with total votes cast as a share of eligible voters. Tables 2 and 3 report estimates for the impact of SPS on incumbent support and turnout

[3]The exact wording of the retrospective evaluation question in the survey was as follows: "Compared to five years ago, do you think Mexico is better, the same, or worse today economically/politically/socially?"

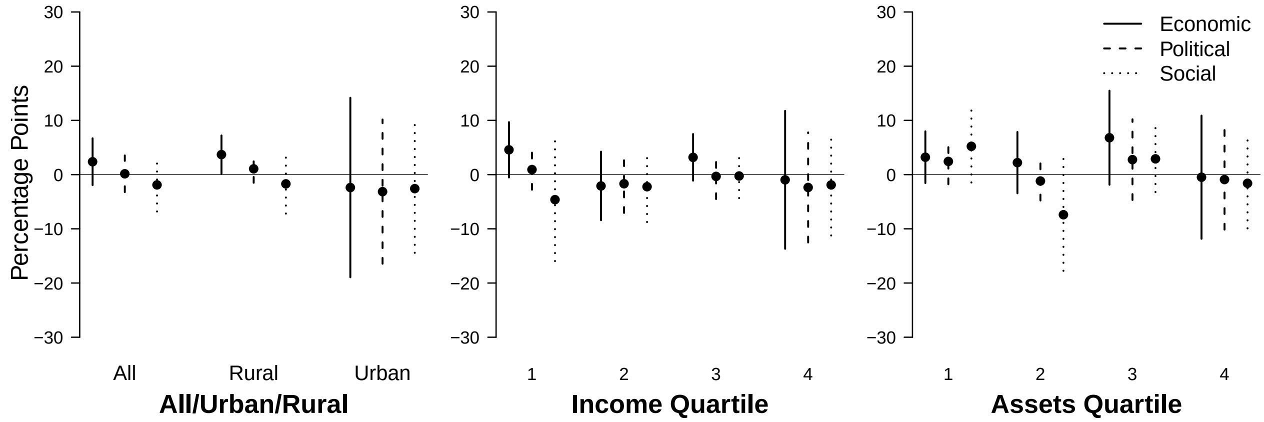


Figure 2: **Itention to Treat Estimates of SPS Effect on Retrospective Survey Evaluations.** The figure reports point estimates and 95% confidence intervals of the total causal effect of SPS on economic (solid vertical lines), political (dashed), and social (dotted) retrospective evaluations of whether the country was doing better today than it was five years ago ($n = 32,515$ individuals in 50 matched health cluster pairs). Results are reported for all respondents and by urban/rural breakdown (left panel), income quartile (center), and asset quartile (right).

when we break up precinct clusters pairs and analyze the data in a regression framework. Tables 4 and 5 do the same but for the sample of rural precinct clusters, for which we know the population measurement error is minimal. Figure 11 shows that the re-allocation of opposition resources following the introduction of SPS does not explain the policy's null effect. Figure 12 reports intent-to-treat (ITT) estimates by the share of evaluation population in precinct clusters to address concerns of attenuation bias in rural areas. Finally, Figure 13 reports estimates for the impact of SPS on individual retrospective economic, political, and social evaluations of the country when we control for the responses of the same people in a survey we conducted at baseline.)

Overall, the results are unambiguous: the highly successful SPS programmatic policy had little effect on turnout or the vote.

# 3 Experiment 2: *Progresa*

## 3.1 Background

We are fortunate to be in the possibly unprecedented position for political science of having a second large scale randomized experiment, along with a natural experiment, to study the same substantive question. This analysis evaluates one of the largest poverty-alleviation programs in Mexico. The program consists of nutritional, educational, and health components. The policy's key feature consists of cash transfers that eligible households receive on the condition that they attend regular health check-ups, and children enroll and attend school. Program benefits vary according to household composition. The average level of benefits is about 35 USD per-month, which represents about 25% of income in poor rural households (Levy, 2006, p. 23).

The origins of the program date back to the mid 1990s, when the country experienced one of the worst economic crises in its history. The government had previously relied on a myriad of food subsidies to alleviate poverty. However, government officials in the administration of then president Zedillo concluded that subsidies benefited urban centers at the expense of the poorest rural areas in the country, were regressive, and too costly to administer (Levy, 2006, ch. 2). The motivation underlying the new administration's alternative approach was a recognition that in order to break the cycle of poverty, one had to recognize the relationship between the educational, health and nutritional components of human capital.

The architects of the program were committed to the eradication of poverty in the country, and as a result they implemented a design that would increase its long-term viability. *Progresa* defined a target population and operated under clear, programmatic rules. In the eyes of the policy-makers, this would ensure the political neutrality of the policy.[4] Political scientists have also arrived at the same conclusion about the nonpartisan nature of

[4] Santiago Levy, one of *Progresa*'s main architects, emphasized the nonpartisan goal of the program: "Congress's role in *Progresa-Oportunidades* has also contributed to its continuity in yet another way: it has established strong legal provisions against the 'political' use of the program. More particularly, it has sought to separate the program from the public image of the president and to provide information directly to beneficiaries about the nature of the benefits that they receive, their rights, and their obligations. . . . These factors, along with the program's positive results, have contributed to the program's transit through three shifts in the composition of the House of Representatives since 1998" (Levy, 2006, pp.107–108).

*Progresa*.[5] Analogous to the experience with SPS, the government hired the International Food Policy Research Institute as a trusted third-party to evaluate *Progresa* and bolster the program's credibility (Levy, 2006, p. 43).

The main evaluation of *Progresa* exploited the phasing of the policy across the country's localities. The sample consists of 506 rural localities in the country distributed across seven states, with 320 villages drawn from a population of localities eligible to receive the program by November 1997. Eligible localities were then assigned to the treatment group. The remaining 186 villages were randomly drawn from populations that would receive the policy in one of the later phases (November-December and March-April of 2000) as the control group. Although *Progresa's* design was "completely randomized", as distinct from SPS's more powerful "matched pair randomized" design, both have the advantages of large scale experimental randomization. Then, the government carried out several surveys in each of these villages, first to determine household eligibility and then to measure outcomes over a period of two years. Behrman and Todd (1999) find that treatment and control villages are fairly similar across a large battery of socio-economic indicators.

The results of the evaluation based on this sample of villages shows that *Progresa* increases school enrollment, improves the health and nutrition of children and adults, and increases household consumption (largely on food) (E. Skoufias, 2005, Ch. 5). However, as first discussed in Green (2006), the evaluation potentially poses a challenge for estimating the effects on electoral variables because the 2000 Mexican presidential election was held on July 2. This means that for the purposes of studying the electoral effects of *Progresa*, the "treatment group" is defined as having received the program for 31–32 months

[5] Diaz-Cayeros, Estevez, and Magaloni (2016) conducted interviews across communities in Oaxaca – a state with a long history of clientelism – about the experience of voters regarding the provision of *Progresa*. In their interviews, one voter noted that "One can be PANísta, PRIísta or PRDísta and still receive benefits from Oportunidades ... before you had to be with the PRI to get anything from the government." Another respondent noted that "the governor controls everything in Oaxaca. However, here you can be PANísta, be with the governor [back then from the PRI], and still get benefits from Oportunidades." A final interviewee asserted that "although sometimes people who do not really need it get Oportunidades, it is less corrupt because benefits arrive regardless of which party you like," On the basis of evidence like this, Diaz-Cayeros, Estevez, and Magaloni (2016, p. 195) conclude: "In our interviews in villages in Oaxaca, it became clear that the poor perceive big differences between *Oportunidades* and other social programs, and that they are generally most satisfied with the former because they perceive it as an entitlement rather than a political favor that comes and goes according to the waves of elections."

before the election whereas the "control group" is defined as receiving the program for only 3–8 months (see the time line in Figure 5 of our Supplementary Appendix). This is not as clean a test of the programmatic hypothesis as with our SPS experiment, since those who received the program more recently in the control group may be as or more grateful as those who received it earlier, although the distinction between the two groups remains unambiguous.

Fortunately, the way in which *Progresa* was rolled out across the country offers a different identifying assumption in the form of a natural experiment. Government officials relied on a poverty index to determine which communities ( *localidades*) would be enrolled in the program.[6] Authorities first enrolled in the program all localities reporting high and very high levels of poverty (and meeting other criteria, such as having access to a health center and educational facilities, and having a threshold population level). Once authorities completed this phase, they proceeded to progressively incorporate localities in the poorest quintile, the second poorest quintile, etc., among the set of localities reporting medium levels of poverty (Green, 2006, p.67). As Figure 5 in the next section shows this procedure generated two large exogenous discontinuities in the proportion of communities (and households) enrolled in the program. [7] Following Green (2006), we exploit these discontinuities to estimate the impact of *Progresa* under a Regression Discontinuity Design (RDD).

As with the SPS experiment, analyzing federal election outcomes in Mexico requires a procedure for merging or matching the boundaries of electoral precincts with often overlapping census geography, as the two are generated by different administrative offices that

[6]The index (*índice de marginación*) classified 105,749 localities in the country across the following five categories of poverty: very low, low, moderate, high, and very high. To distribute the localities across the five categories, officials used factor analysis to create a latent measure of poverty. Once this measure was obtained authorities then implemented an optimal classification algorithm.

[7]Green (2006, p. 74) reports that the exact cutoffs are located at the value of the index separating localities reporting low and moderate levels of poverty (Threshold 1) and at the level separating localities in the three poorest quintiles from those in the two richest quintiles among the set of localities reporting moderate levels of poverty (Threshold 2). However, as shown in Figure 45, we find that Threshold 2 is slightly lower than the value separating the quintiles of interest. For our estimation we set Threshold 2 equal to $-0.96$ (instead of $-0.932$), which corresponds to the largest effect of the encouragement on *Progresa* enrollment. Figures 45 and 46 show, however, that our main estimates are robust to different values of Threshold 2. Finally, we note that Threshold 1 and 2 are not deterministic because government authorities, in addition to the poverty index, took into account access to health and educational facilities for program enrollment.

typically do not coordinate. The same issue exists in almost all analyses of electoral data around the world, but the method of dealing with it is crucial. As we explain below, errors in this merging process for *Progresa* explains certain prior results in the literature.

## 3.2 Results

The literature analyzing *Progresa* is divided over whether the data support the programmatic incumbent support hypothesis. The first analysis in this literature, based on the natural experiment, finds no effect on either incumbent support or voter turnout — with estimated effects close to zero and small confidence intervals (Green, 2006). In contrast, the second analysis, based on the randomized design, reports strong positive effects (De La O, 2013, 2015). We show here that the reason for this discrepancy is not related to the differing identification strategy, but rather faulty data merging procedures (affecting about 70% of the observations) that happened to induce misleading results for the randomized experiment and not the natural experiment. Unconventional data analysis choices in the randomized experiment also contributed to the incorrect conclusions. Correcting either (or both) in the analysis of the randomized experiment generates results that mirror those from the natural experiment, which are also consistent with our analysis of SPS in Section 2 — both indicating little or no effect of programmatic policies on voter turnout or incumbent support.

Appendix A reveals these errors and shows how to correct them. The rest of this section replicates the original results reported in Green (2006) and De La O (2013, 2015). For the first analysis we exactly replicate the results using data that was incorrectly merged, with incumbent voting and turnout measured as a function of all people, and then with corrected data based on accurate GIS coordinates along with more appropriate statistical techniques. We also add a new data source, with outcome variables based on voting as a function solely of those officially registered. This alternative coding has two advantages. First no merging is necessary and so no corrections are needed. And second, we offer a much stronger test of the hypothesis by excluding those who cannot vote and thus have zero causal effects — such as those underage, not citizens, not registered, etc. — from the denominator of the outcome variables. For the second analysis, we report results relying on the correctly merged data using GIS coordinates, and show that the original conclusions of the study hold.[8] (In our Supplementary Appendix we provide numerous other analyses, tests, and other evidence, all of which yield the same conclusions as that offered here.)

We begin our analysis with Figure 3, which replicates the regression estimate and 95% confidence intervals from Green (2006) and De La O (2013) for the total causal effect of *Progresa* on turnout (left panel) and incumbent vote (right panel). Results for De La O (2013) are based on the sample as originally coded with errors and including those who cannot vote in the denominator of the outcomes (squares). We also analyze official measures of the outcomes (turnout and incumbent support) with the incorrectly merged sample (diamonds) and with a corrected GIS sample, without merging problems (dots). Results for Green (2006) are based on official outcomes for the sample of precincts, each of which has only one village (*localidad*) obtained with the correct GIS procedure for merging (triangle). The horizontal line marking no effect appears at zero.

Reading the left panel in Figure 3 from left to right, the different specifications we tried for De La O (2013) include the linear regression in the original article and book; a simple difference-in-means; a matching estimator[9]; a regression controlling for log-population; a regression with lag turnout on the same scale as the outcome; and a regression, under the original specification, after removing the two observations with the highest leverage. For Green (2006), we report the original pooled sharp RDD results.[10] The panel on the right repeats all the analyses for incumbent (PRI) vote share, including for Green (2006) the original estimates for the total effect of *Progresa* on PRI support in the Proportional Representation (PR) Senate election (triangle) and in the presidential election (inverted triangle).

[8]We did not have access to the exact sample analyzed in Tina Green's unpublished dissertation and so are unable to report numerical estimates obtained in her analysis. However, this information would not change our conclusions or her's.

[9]For matching, we did Coarsened Exact Matching (CEM), adjusting the coarsening to deal with the presence of high leverage observations among the pre-treatment covariate. The distributions of covariates before and after matching are reported in Figures 18– 19 and 23–24 in the Supplementary Appendix, with full information in our replication data set.

[10]Figures 39 and 40 in the Supplementary Appendix report additional results across the different thresholds for locality enrollment to *Progresa*, implementing different kernels (uniform and triangular), and employing different RD estimators (standard and bias-correcting).

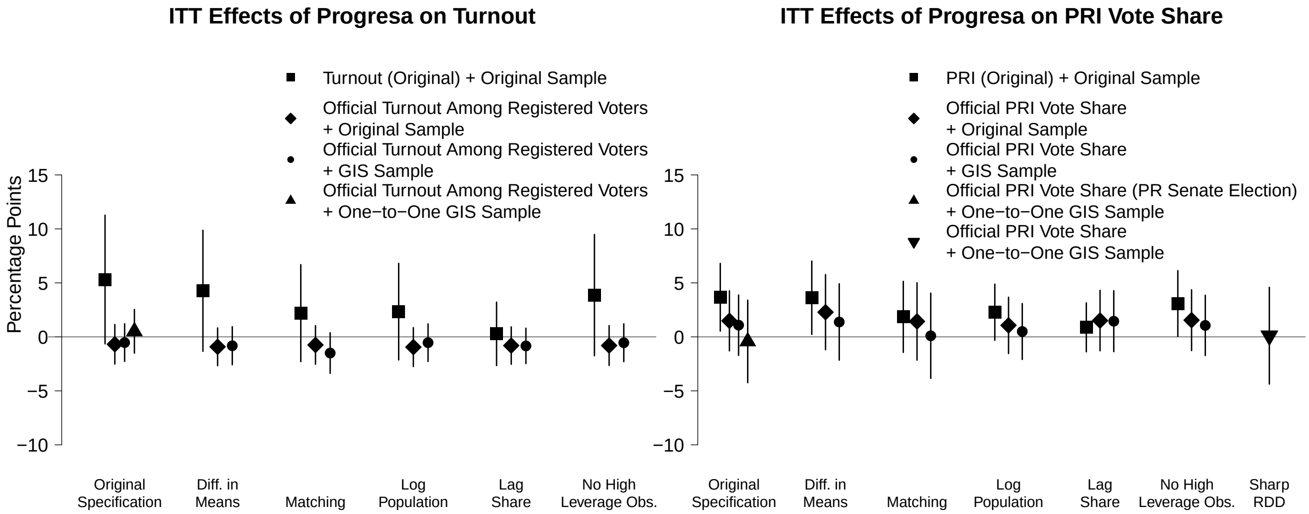


Figure 3: **Intention to Treat Estimates of *Progresa* Effect on Turnout and Incumbent Party Vote.** The left panel reports point estimates and 95% confidence intervals for the total causal effect of *Progresa* on turnout in the 2000 presidential election as originally, and incorrectly, measured in the De La O (2013) sample (squares), for official turnout among registered voters in the same sample (rhombuses), and for official turnout among registered voters in the correct GIS sample (dots). The panel also replicates Green (2006)'s total causal effect of *Progresa* on turnout in the sample of precincts with only one village under a sharp RD design (triangle). The right panel repeats the same analyses for incumbent (PRI) vote share, and the effect of *Progresa* under sharp RDD on both PRI support in the 2000 Proportional Representation (PR) Senate election (triangle), as in Green (2006), and in the presidential election (inverted triangle). Every estimate is indistinguishable from zero, except when using the flawed original measure used in De La O (2013) and without controls (first two lines with squares representing point estimates in the right panel).

The results in Figure 3 exactly replicate results in De La O (2013, 2015), with positive point estimates for turnout and vote share for the incumbent party, and a 95% confidence interval that excludes 0 for vote share but is insignificant for turnout.[11] Using the original variable (with errors uncorrected) reveals the same basic results, even using a simple difference in means estimator. However, once we use any of the four alternative approaches, each of which control for the large imbalance induced by the data errors, the positive effects vanish with no statistically significant evidence for the effect of *Progresa* on either turnout or vote share. Moreover, rerunning any of the six analyses, while dropping the original incorrectly coded variable and switching to official registration data (which

[11]Figures 16 and 17 in the Supplementary Appendix display all the point estimates reported in this section but with 90% confidence intervals.

has no possibility of data merging errors and higher probability of revealing an effect if present), reveals no evidence of for the effect of this nonpartisan programmatic policy on either partisan outcome, regardless of how the data are analyzed. Moreover, with the clean registration data, the confidence intervals are much narrower, and all twelve include zero as a causal effect. Our reanalysis of Green (2006) strongly confirm the substantive conclusions reported in that study, namely that *Progresa* did not have a substantial positive impact on either turnout or incumbent support.

Figure 4 repeats the same analyses, with the same robustness checks, for the instrumental variable analysis estimate of the causal effects in De La O (2013, 2015) and Green (2006).[12] The results here tell essentially the same story, with no statistically significant effect of nonpartisan programmatic policies on voter turnout or vote for the incumbent party. Although again, only the official turnout and vote figures (all point estimates except squares) offer valid causal estimates, and these are not statistically different from zero.

To study the possibility that the null electoral effect of *Progresa* is due to attenuation bias, resulting from the presence of program beneficiaries and non-beneficiaries across precincts, we estimate the total causal effect by the average precinct poverty level and the share of experimental population in precincts. The findings, reported in Figures 21-22 and 25-32 in the Supplementary Appendix, show that the effect of Progresa is not increasing in poverty levels across precincts or share of experimental population, and so attenuation bias does not seem to be a concern. Numerous other alternative specifications and analyses of the results, studying effects in every alternative way we could think of, appear in our Supplementary Appendix. All lead to the same conclusion as with SPS: *Progresa* has little or no effect on either voter turnout or the incumbent vote.

[12]The Fuzzy RDD estimates reported in Figure 4 are based on a regression specification where the treatment is a binary indicator for whether a locality was enrolled in *Progresa* as originally defined in (Green, 2006). Figures 43 and 44 in the Supplementary Appendix report results when the treatment is instead the proportion of families in a locality enrolled in *Progresa*.

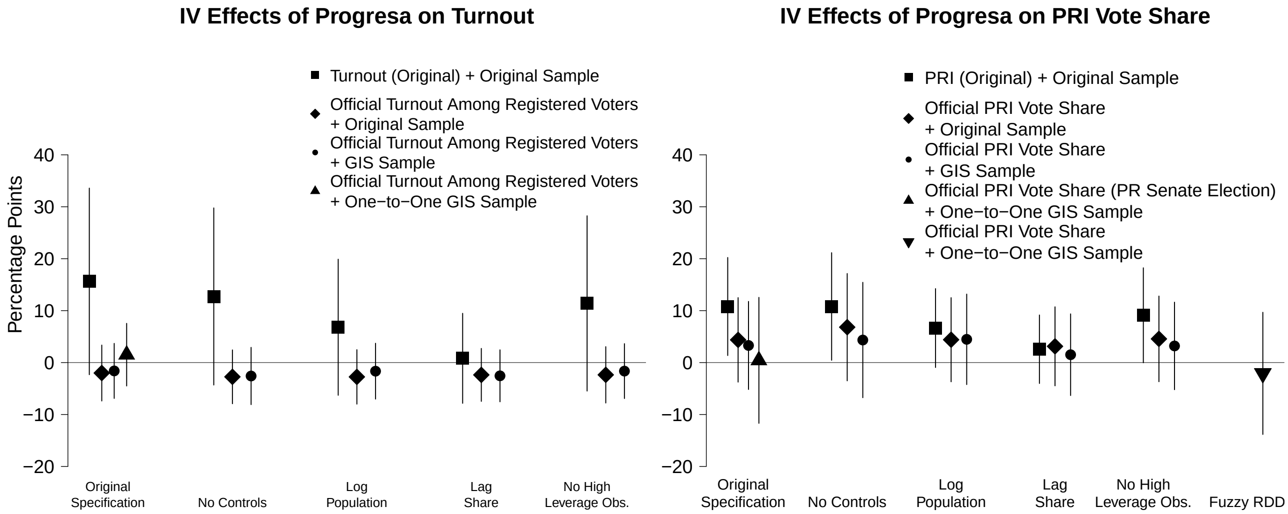


Figure 4: **Complier Average Treatment Estimates of *Progresa* Effect on Turnout and Incumbent Party Vote.** In a manner directly parallel to Figure 3, this figure replicates the instrumental variable estimation from De La O (2013) and the fuzzy RD design from Green (2006). Every estimate is indistinguishable from zero, except when using the wrong measure without controls (first two lines with squares representing point estimates in the right panel).

## 3.3 Robustness of our substantive conclusion

Two important shortcomings of the experimental evaluation of *Progresa* where first noted in the doctoral dissertation of Green (2006, fn.26), and mostly ignored thereafter. First, all villages (localities) in the evaluation study, including those in the control group, had received the treatment by the time of the 2000 election. Second, some treated localities share precincts with other localities that were not part of the evaluation, leading to a small number of households in the experiment who enrolled in *Progresa*. These two shortcomings of the randomized experiment may have contributed to the evidence that the program had little impact on election results.

Below, we address these two shortcomings of the randomized experiment by following Green (2006) and employing an alternative identification strategy based on a regression discontinuity design (RDD). To do this, we exploit the arbitrary cutoffs government officials used to phase-in localities to *Progresa*. These results strongly confirm those of Green (2006) and our substantive conclusion that *Progresa* had little impact on electoral results.

Although the original data from this dissertation were not available to us, we requested and received directly from the Mexican government information about the number of families incorporated to *Progresa* across all localities in the country for each of the expansion phases of the program. We thus focus on the $105,749$ localities reporting a value of the poverty index used to enroll localities in the program. For these localities we create two versions of the treatment: (1) an indicator variable (`Progresa`) that takes the value of 1 if at least one family was enrolled in the program by the 11th phase of program expansion (the last one prior to the year 2000 election) and; (2) the proportion of families receiving *Progresa* within a given locality by the same program expansion phase.[13] Relying on GIS we then we merged these localities with the set of precincts in the 2000 election. Following Green (2006), we examine precincts containing ony one locality to avoid all issues of aggregation and merging.[14] This process left us with a total of $7,865$ precincts for our RDD analysis.[15]

Finally, the Mexican authorities relied on a poverty index to determine which localities were given priority to be enrolled in *Progresa* (Emmanuel Skoufias, Davis, and Behrman, 1999, Section 3). Authorities first enrolled in the program all localities reporting high levels of poverty (and fulfilling other criteria, such as having access to a health center, educational facilities, and with a certain level of population). Once authorities completed this phase, they proceeded to progressively incorporate localities in the poorest quintile, the second poorest quintile, etc., among the set of localities reporting medium levels of poverty (Green, 2006, p.67).

Figure 5 shows that this procedure generated two large discontinuities in the propor-

[13]Because we do not have data on the total number of families per locality, we use instead the total number of inhabited households as the denominator to compute the proportion of *Progresa* families across localities. This results in 231 localities reporting a value greater than one. The likely reason for these values, according to INEGI documentation, is that two or more families may share a household.

[14]Chiapas and Mexico City are excluded from the sample. We exclude Chiapas because over 1500 localities lacked geographic coordinates in the 1995 population count. We exclude Mexico City because it was not incorporated to *Progresa* by the program's 11$^{th}$ phase of expansion.

[15]The total number of precincts analyzed in Green (2006) is $3,379$. The reason for the discrepancy between our sample and the sample analyzed in Green (2006) is the name-matching procedure the latter study used to merge localities with precincts. As we have shown, this procedure is unreliable, and in this particular case may have led to a underestimate of the number of precincts with only one locality. This may have happened because a large number of precincts reporting two (or more) localities in the files from the electoral authority may in fact contain only one according to the way census authorities aggregate population at the local level.

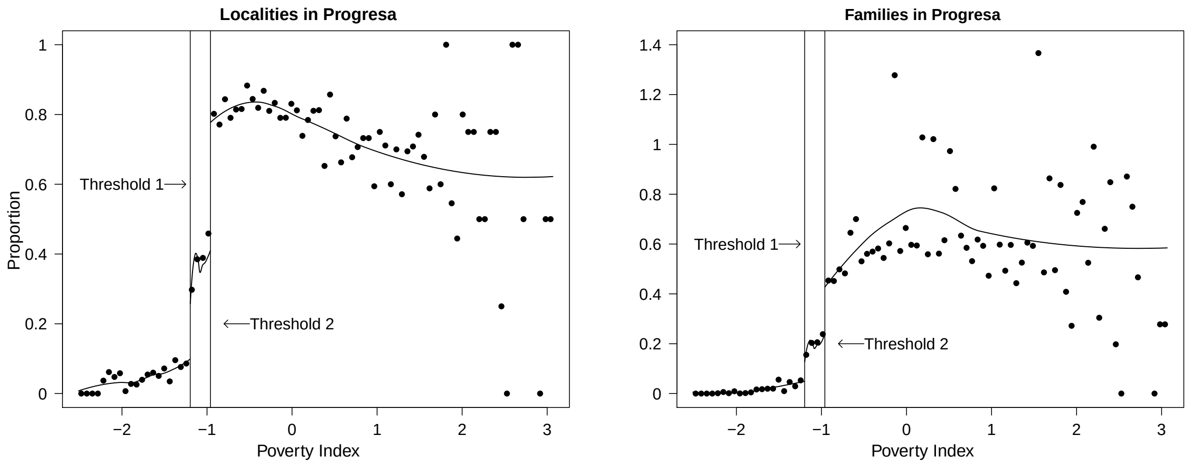


Figure 5: **Enrollment of Localities and Families in Progresa by Poverty Index.** The panels display the proportion of localities phased into *Progresa* (left panel) and the proportion of Progresa-beneficiary families per locality (right panel) as a function of the census poverty index. The panels show two large discontinuities in the proportion of localities incorporated to *Progresa* and in the proportion of *Progresa*-beneficiary families at the cutoffs the government used to phase-in localitites to the anti-poverty program.

tion of communities and households enrolled in the program. At Threshold 1, located at the value of the index separating localities reporting low and moderate levels of poverty, we can see a substantial 28 percentage point increase in the the proportion of localities incorporated to *Progresa*, and 15 percentage points for households. Similarly, at Threshold 2, we find a 40 percentage point increase the proportion of localities enrolled in *Progresa*, and 23 percentage points for households. In Section 3.3 of the Supplementary Appendix, we examine several pre-treatment covariates including previous election results and show that there is no such discontinuities in these variables, giving a strong support for the validity of the RDD analysis.

Figure 6 presents the ITT effects of *Progresa* on PRI vote-share and turnout in the 2000 presidential election. The figure reveals the absence of any discontinuity at either of the two thresholds, indicating that *Progresa* had no discernable effect on turnout or PRI voteshare. Indeed, the point estimate which is almost exactly zero which, along with 95% confidence intervals, appear in in Figure 3. We also present the IV estimate based on the fuzzy RDD analysis as shown in Figure 4, again showing that the estimated effect is essentially zero. In Section 3.3 of the supplementary appendix, we also examine other election results examined by Green (2006) — Proportional Representation (PR) senate

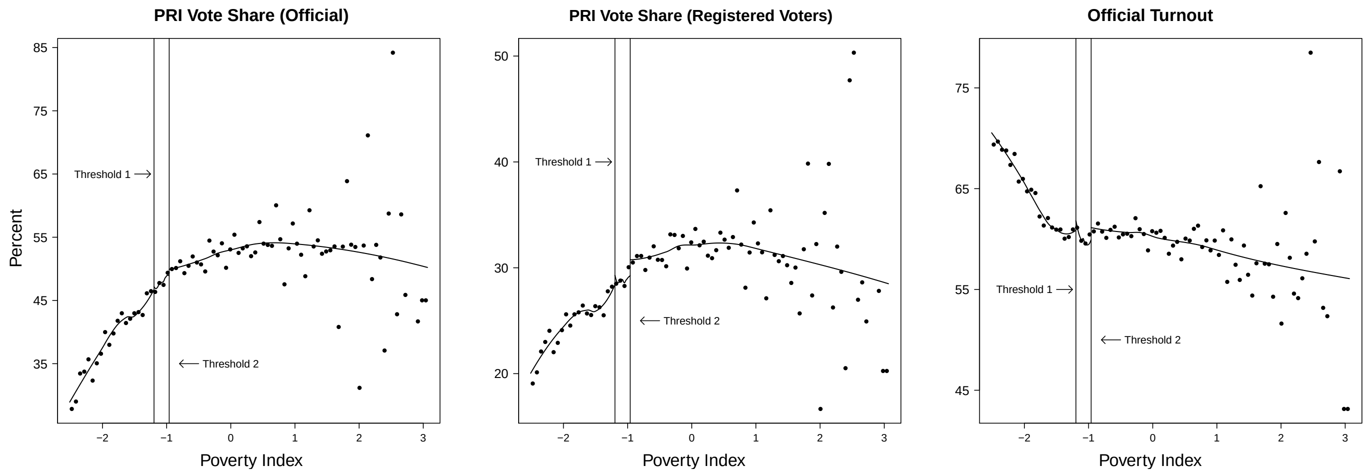


Figure 6: **ITT Effects of Progresa on PRI Vote Share and Turnout in the 2000 Presidential Election.** The panels display average official PRI vote share (left), PRI vote as a share of registered voters (center), and official turnout (right) in the 2000 presidential election as a function of the poverty index. There is no discernible discontinuity in the outcomes at either of the government cutoffs used to phase-in localities to *Progresa*, indicating that the program had no effect on electoral outcomes.

and simple plurality deputies elections. We find again that *Progresa* had little impact on any of these elections.

# 4 Prior Research

Our results, which clearly reject the programmatic incumbent support hypothesis, differ at least superficially from the largely positive conclusions reported in the literature (Labonne, 2013; Larreguy, Marshall, and Trucco, 2015; Manacorda, Miguel, and Vigorito, 2011; C. Pop-Eleches and G. Pop-Eleches, 2012; Zucco, 2013). After some study, we conjecture that the key factor leading to these divergent results is the nature of the politics under which each policy was passed, and the ways in which the policies were implemented (Layton and Smith, 2015). Whereas SPS and *Progresa* are essentially pure forms of programmatic policies, most policies studied in prior research are partly programmatic and partly clientelistic. To be specific, programmatic policies in general, including those we study and all others in the literature, (a) impose objective rules that give incumbents no discretion over implementation. This condition implies that programmatic policies should (b) lead voters to expect to receive the same services regardless of the party in office. In addition, SPS and *Progresa* are distinctive in that they also (c) were passed with broad support in the legislature from all major parties and, perhaps as a consequence, (d) did not lead any single party to attempt to claim exclusive credit for its implementation. Our hypothesis is that Conditions (c) and (d) play a major role in explaining the differing results.

It is true that studies of other programmatic policies have been almost exclusively observational. In contrast, randomized experiments enable researchers to estimate causal effects without risky modeling assumptions that are necessary in observational studies. Yet, experimental studies share with observational studies the problem of not being automatically representative outside the (political) context in which they are conducted (Imai, King, and Stuart, 2008). This makes understanding the political context essential for any general understanding of the programmatic incumbent support hypothesis, experimental or observational. We thus now discuss the differences in the politics under which policies are adopted and implemented in relatively pure-form and under mixed programmatic-clientelistic policies. We first consider differences the politics of adoption and implementation, then discuss broader theoretical explanations, and finally make suggestions about future research.

**Politics of Policy Adoption** Clientelism has historically been the norm across developing countries because it yields large electoral benefits to incumbents (Golden and Min, 2013). Incumbent chief executives generally prefer clientelistic policies, because they are likely to yield large electoral benefits for themselves and their parties (Stokes, 2005). However, adopting clientelistic policies is only possible with unified partisan control of government (De La O, 2015). Under divided partisan control, the only policies that have a chance of passing are those supported by all parties (or at least all veto players). This explains why programmatic policies — SPS, *Progresa*, and every other major programmatic anti-poverty program in the developed world — have only arisen under divided control. For example, in ensuring congressional approval for SPS and *Progresa*, policymakers took great pains to ensure no political actor could claim credit distinct from other parties, and could not use the policies for their own electoral advantage (Levy, 2006; Ortiz, 2006).

[16] Voters were also well aware of these facts, and as a result widely believed they would receive program benefits regardless of the incumbent's partisan identity (Diaz-Cayeros, Estevez, and Magaloni, 2016).

**Politics of Implementation** As relatively pure programmatic policies, SPS and *Progresa* reduce the chances of partisan credit claiming via a clear policy design and resulting nonpartisan bureaucratic implementation. *Progresa* is known for having the strongest design and implementation protocol among the universe of conditional cash transfers in Latin America (De La O, 2015), and SPS was quite similar in the period we examine. In contrast, none of the observational studies of the programmatic incumbent support hypothesis analyze pure programmatic policies (Larreguy, Marshall, and Trucco, 2015; Manacorda, Miguel, and Vigorito, 2011; C. Pop-Eleches and G. Pop-Eleches, 2012; Zucco, 2008, 2013), and as such none exactly fit criteria (a)–(d) outlined above. These observational studies also analyze different policies, political situations, and political contexts than from each other and for our two experiments. Of course, *every* pair of studies that analyze a different policy or time period differ in an infinite number of ways, any one of which might be the lynch pin that determines any differences in results. As such, future researchers have an important opportunity to help teach us more about how far the programmatic incumbent support hypothesis, even though apparently inapplicable to pure programmatic policies, may work for more mixed programmatic-clientelistic government programs.

Consider that PANES, the temporary emergency relief program analyzed in Manacorda, Miguel, and Vigorito (2011), was a one-time effort. As such, and as discussed by the authors, voters may have associated the policy with the party responsible for its enactment and implementation. Similarly, in the case of Romania, C. Pop-Eleches and G. Pop-Eleches (2012) explain that the Euro 200 program had a well known partisan intent. In Brazil, Hunter and Power (2007) and Zucco (2008) argue that one of the reasons for

[16]Recent work in lab experiments shows individuals reward actors with proposal power for collective decisions (Duch, Przepiorka, and Stevenson, 2015). A key condition for this finding is that individuals know the actor with such power. This is unlikely to hold in the Mexican case as the political bargain required for the adoption of *Progresa* and SPS prevented the party in power from advertising its role as the policy initiator.

*Bolsa Família*'s electoral success was the president's ability to claim credit for the program. In addition, A. Hall (2006) notes that the implementation of *Bolsa Família* was decentralized to municipalities, which led to charges of clientelism in the delivery of the program's benefits. Finally, Larreguy, Marshall, and Trucco (2015) study an interesting urban titling program crafted in 1973, before Mexico was a democracy. Under this program, incumbents organized events, claiming exclusive credit for the number of land titles granted during a given time period.

Similarly, in a field experiment in Uganda, Blattman, Emeriau, and Fiala (2016) find that a program transferring cash to groups of unemployed youth increased support for and work on behalf of the *opposition*, among survey respondents. The policy was mostly funded by the World Bank, and a large proportion of survey respondents credited this institution for the creation of the program instead of the country's autocratic government. The authors hypothesize that this aspect of the program freed voters from clientelistic ties, thereby increasing their propensity to support and work for the opposition. Frey (2015) finds a similar result in Brazil, where local mayors have a limited ability to claim credit for *Bolsa Família*.

In the Phillippines, Labonne (2013) finds, based in a randomized evaluation, that a conditional cash transfer (CCT) program modeled after Brazil's *Bolsa Família* and Mexico's *Progresa* increased support for local incumbents only in competitive (i.e., non-dynastic) municipalities with small federal budgetary transfers. The author does not find similar electoral benefits in competitive municipalities with large transfers. The explanation given for this difference is that in competitive municipalities with large transfers, mayors distribute funds among individuals not receiving the CCT, making the voting behavior of program beneficiaries and non-beneficiaries indistinguishable from each other.

**Theoretical Explanations** The well developed formal theory literature in this area is almost solely concerned with the electoral effects of discretionary, rather than programmatic, spending policies. For example, Dixit and Londregan (1996, p.1132–1133) devote the first two pages of their article to clarifying this point. Theoretical results suggest that targeting tactical or pork barrel spending occurs most often to benefit swing voters (Dixit and Londregan, 1996) and marginal constituencies (Weingast, Shepsle, and Johnsen, 1981), and this is indeed what the evidence shows (see Dahlberg and Johansson 2002 on the former and Keefer and Khemani 2009; Primo and Snyder 2010 on the latter).

Studies have also found that such spending has substantial electoral payoffs (Evans, 2006; Levitt and Snyder, 1997). One clever exception, proposed in De La O (2015), holds that minority parties will rationally choose to support a programmatic policy that will hurt them electorally if opposing it imposes even larger costs. Unfortunately, as we show in Section 4 of our Supplementary Appendix, the proposed formalization of this argument is consistent with both the programmatic incumbent support hypothesis and the opposite, and so the theory cannot be used to explain either. We also show there that the parameters of the theory are impossible to test from the *Progresa* (or SPS) experiment and so either way the theory cannot explain the empirical results. However, the foundations of the model in De La O (2015) are of considerable value: when governments try to pass major public policies to cope with poverty in the presence of divided government they are much more likely to adhere to the programmatic policy ideal.

A variety of other perspectives might also help explain apparent divergent conclusions of different studies. For example, as we describe above, the *Progresa* experiment compared those who received the program recently and more distantly, unlike the SPS experiment which had a clean control group that did not receive the program at all.[17] Or, voters may simply reward incumbents for the implementation of programmatic policies as a result of reciprocity (Finan and Schechter, 2012; Manacorda, Miguel, and Vigorito, 2011). Incumbents may also signal their commitment to the poor by adopting certain policies, making it rational for voters to re-elect them (Diaz-Cayeros, Estevez, and Magaloni, 2009). Another possibility is that voters are retrospective in their voting behavior. Therefore, welfare improvements associated with government programs prompts them to reward incumbents (C. Pop-Eleches and G. Pop-Eleches, 2012). Finally, it may simply be that while incumbents pledge to implement a policy in a non-clientelistic manner, bureaucrats and local-level politicians fail to fulfill this promise (A. Hall, 2006; Rocha-Menocal,

[17]However, Green (2006), using a fuzzy regression discontinuity design, also finds that *Progresa* does not have an impact on incumbent support or turnout.

2001).[18]

**Future Research** A valuable area for future research, then, would involve deriving theories that could shed light on when and under which political contexts and policy proposals the programmatic incumbent support hypothesis may or may not hold. Empirical research would also benefit from studies that seek to study the programmatic incumbent support hypothesis in policy and political contexts that are partially programmatic and partially clientelistic (Camp, 2013; Diaz-Cayeros, Estevez, and Magaloni, 2009, 2016).

# 5 Concluding Remarks

When incumbent parties exercise discretion over government spending to benefit specific groups of voters, an impressive literature shows that these voters reward the incumbents with electoral support in subsequent elections. The reason for this powerful result would seem to be called into question by claims that nonpartisan programmatic policies — policies over which incumbents possess no discretion — have similar effects on voters making partisan decisions. We sought to contribute to an explanation of why this would happen and, also, why minority party politicians would, in the first place, support policies that would lead voters to oppose them in the next election.

We are in the unusual position for our discipline of being able to test an important hypothesis with two extremely large scale, comparatively high quality randomized experiments and a compelling natural experiment. We use these experiments to study two relatively pure form programmatic policies, and find no evidence in support of the programmatic incumbent support hypothesis. Our results may suggest why it is difficult for a country to break the cycle of perverse accountability (Stokes, 2005): Since only clientelistic policies yield electoral payoffs, incumbents are willing to adopt programmatic policies only in a context of high political competition.

The evidence we offer may raise other issues as well, such as the fact that parties in

[18] Blind retrospection may yet be another mechanism accounting for a positive effect of programmatic policies on incumbent support (Achen and Bartels, 2004). We consider this possibility unlikely, as other studies have questioned the existence of such phenomenon (see, for example, Fowler and A. B. Hall, 2016).

Latin America do not seem to have converged to political competition on a left-right dimension as is common in Western democracies. Indeed, the cases of *Progresa* and *Seguro Popular* illustrate that certain policy-making domains, which may be of intense political contestation in other countries, have become de-politicized. The exact political context in which policies are adopted and implemented may have a major role in explaining differences of results in different settings. We look forward to future researchers shedding additional light on this important question in other national and policy contexts.

# Appendix A Fixing Data and Analysis Issues

We summarize here the coding errors created by faulty merging, evidence that these errors are sufficient to bias the causal effects, and the unconventional analysis techniques in De La O (2013, 2015). We also show how to fix these problems. More information is available in our Supplementary Appendix. Although we reach the unavoidable conclusion that these errors invalidate the claims in this work, De La O deserves credit for highlighting this important issue, thinking of the idea of repurposing a randomized experiment, gathering the necessary data, and making available a replication data set so that further discoveries became possible.

**Coding Errors** In Mexico, as in many countries, electoral and population data are generated by government agencies that do not coordinate and so wind up with inconsistent geographies. For instance, while electoral authorities assign population to a precinct, census officials may aggregate part of the same population with a neighboring village that happens to be outside that precinct. Neither office is necessarily correct or incorrect, as they have differing goals, but the result is difficult to analysts trying to merge incompatible data sources. This lack of inter-agency coordination is a common problem in the analysis of elections in many countries, but the issue here caused particular difficulties here when De La O (2013, p. 5) "overlayed the 506 experimental villages to the smallest unit of outcome measure for which census, precinct, and electoral data roughly coincide: the *sección electoral* (precinct)." These precincts were "name matched" based on the textual names of villages in different data files, from organizations that assigned different geographic locations and meanings to the same names. Unfortunately, the result turns out to be that 71.3% of the observations (villages) were incorrectly matched to areas with similar names but from places outside of designated treatment and control precincts (Supplementary Appendix, §2.2). We confirmed these facts via formal Mexican Freedom of Information Act requests we filed, and conversations with officials at the Dirección Ejecutiva de Organización Electoral (DEOE) and Dirección Ejecutiva del Registro Federal de Electores at Mexico's National Electoral Institute (INE).[19]

Of the incorrect matches due to this procedure, two caused especially large biases. Correcting only these two units (or all of them) leads to the conclusion that *Progresa's* programmatic policies have little or no impact on turnout or voting for the incumbent. The large coding errors are apparent in the first column, for voter turnout, and the fourth, for the incumbent (PRI) vote, in the distributions portrayed in Figure 7. Since it is impossible for more than 100% of people to vote, or to vote for any one party, every observation above the dashed line at 100 is mistaken. Moreover, these include some large errors that extend into the impossible region beyond 50% of the range of the original data, and for turnout more than three times the range. To choose the extreme point, turnout obviously cannot be 375%, as in the data analyzed in De La O (2013, 2015) indicate. In addition to these known errors, the large number of observations with nearly zero turnout (left column, bottom) in a presidential election with 65% overall turnout (IFE, 2013) are of dubious validity.

Before turning to the consequences of these data errors, we analyze turnout and incumbent (PRI) vote among those officially registered (using precincts in De La O (2013) and an alternative GIS-determined sample). These alternative variables appear in Figure 7 in columns 2 and 3 for turnout and 5 and 6 for incumbent vote. These data come from

[19] An important clue comes when De La O (2013, fn.17) tries to account for some of the data with turnout and vote percentages greater than 100 by noting that 10 precincts "contained special voting booths where by law out-of-precinct voters can cast a ballot." Although such precincts do exist, none are in the author's data set — a fact which can be verified in data made publicly available by the Mexican government (http://j.mp/mxife). To be more specific, of the four types of polling stations in Mexico — básicas, contiguas, extraordinarias, and especiales — only especiales allow out-of-precinct voters. Yet, none of the polling stations included in the sample analyzed are of this type.

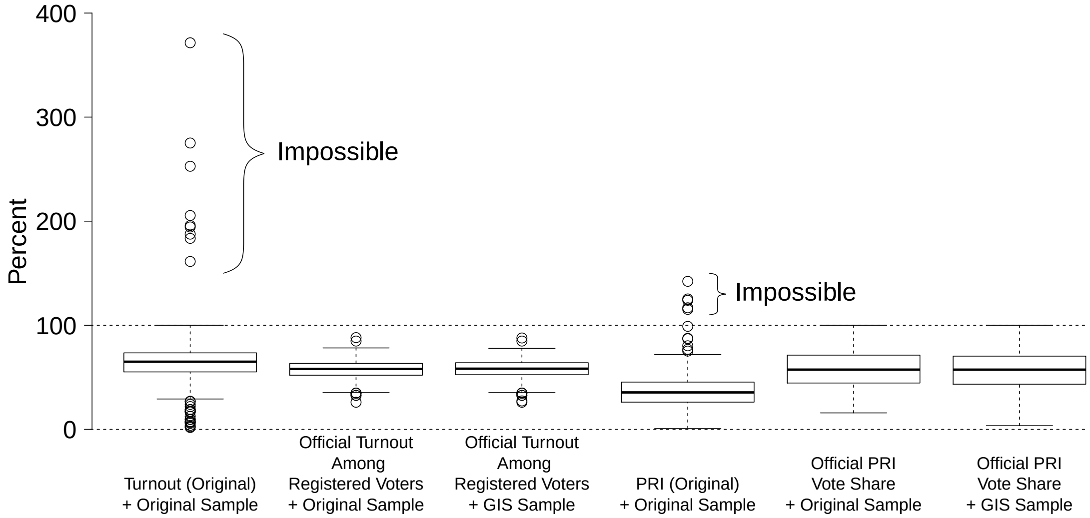


Figure 7: **Univariate Distribution of Turnout and Incumbent Party Vote in 2000.** This figure compares the variables originally constructed in De La O (2013) via name matching (in columns 1 and 4), with the official turnout among registered voters and PRI vote share in the name-matching sample (columns 2 and 5) and in the GIS sample (columns 3 and 6).

one source, with the same geographic boundaries, and no possibility of coding errors. As a result, we can see in the figure that all the observations naturally fall within the possible region between 0 and 100%.

Our alternative turnout and incumbent vote share variables, based on officially registered voters, have no such measurement error bias problem since all those in the denominator would appear in the numerator if they choose to vote. If *Progresa* works in part by increasing levels of registration — as hypothesized in De La O (2013, p.7–8) — then the interpretation of the turnout and vote variables would change, but no post-treatment or other bias or inefficiency would be induced.[20] Since no evidence has previously been offered for this hypothesis, we directly test it and present our results in Tables 14–17 and Tables 24–25 of our Supplementary Appendix. The results, with small confidence intervals around zero, demonstrate that the program had little or no impact on registration rates. These alternative codings for turnout and vote are considerably cleaner tests of the

[20] To see why no post-treatment bias is introduced, let $V(t)$ and $R(t)$ represent the potential number of those who turned out and registered under the treatment condition $t = 0, 1$, respectively. This notation allows for the possibility that the treatment can affect both turnout and registration. Our quantity of interest, i.e., turnout rate, is denoted as $Y(t) = V(t)/R(t)$, which is a well-defined potential outcome. Since we do not condition on the realized registration rate, no post-treatment bias is introduced.

programmatic hypothesis, even if the original data had no coding errors.

**Conditions for Bias** Thus far, we have revealed the existence of data errors, but that alone is insufficient to change any conclusion; the real question is whether these errors make a difference in the results. We find that this is indeed the case, especially for two huge outliers. The particular impact of these outliers is their correlation with the control variable "lagged population," which we graph in Figure 8. In the left panel, we graph the raw data, which reveals the two outliers (precincts 266 and 1502), each 90 to 100 times the size of the median and both of which appear in the treated group, with nothing comparable in the control group. De La O (2013, p. 6) was right to check balance between the means of the treated and control groups, but checking solely based on the means led to missing the massive imbalance in the tails of the two distributions evident in this figure. The consequence of imbalance is model dependence (Ho, Imai, King, and Stuart, 2007). Since the most important advantage of a correctly analyzed randomized experiment is the absence of model dependence, introducing these coding errors eliminated this important benefit of randomization. Below, we illustrate and correct this problem. As we show in the next section, regardless of the reason for these outliers, when we correct for the imbalance they create, all evidence of significant effects of *Progresa* on turnout or incumbent support vanish.

We also go a step further and formally evaluate the bias these (incorrect) outlier observations have on the least squares analysis in De La O, 2013. We do this by computing the "statistical leverage" of these observations, with and without the lagged population variable. Observations with larger values of leverage have more influence on the magnitude of coefficient estimates in linear regression.[21] As the right panel in Figure 8 makes clear, the statistical results in De La O, 2013 are driven largely by these two observations. Indeed, when controlling for lagged population, the two precincts have leverage that is 15 to 20 times larger than the median (see the vertical axis of the right panel). In contrast, when population is not included as a covariate, the leverage of these precincts is a mod-

[21]The statistical leverage of observation $i$ is defined as $x_i^\top (X^\top X)^{-1} x_i$, where $X$ is an $n \times k$ matrix of pre-treatment covariates with $k \times 1$ row $x_i$.

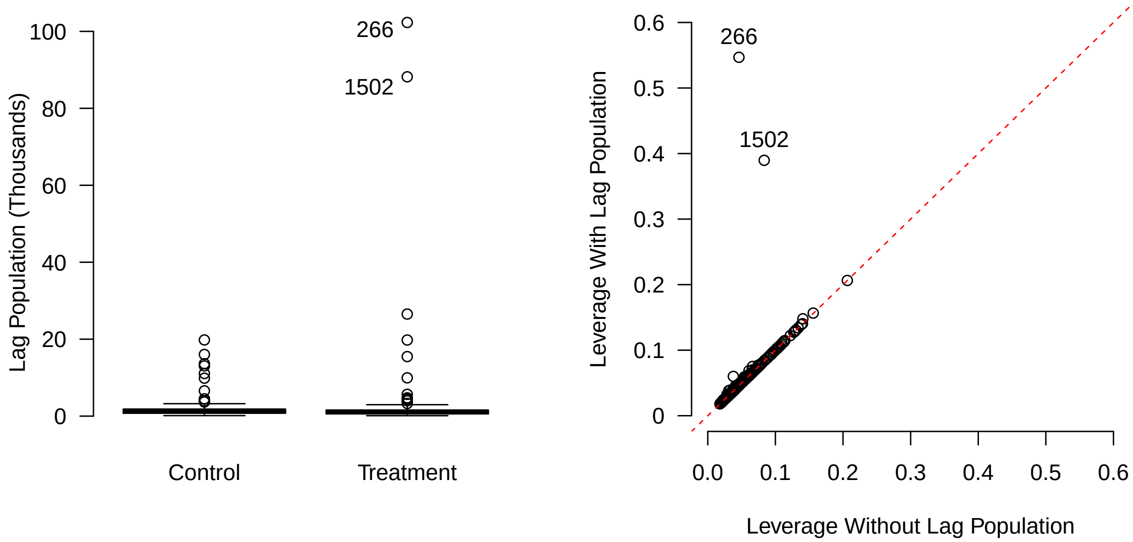


Figure 8: **Population Outliers and Statistical Leverage.** This figure reveals the extreme degree to which two of the miscoded observations are outliers (left panel) with extremely high statistical leverage (right panel). The left graph is raw data; the right is computed from the least squares analysis in De La O (2013).

est 1.5 to 2.75 times larger than the median (see the horizontal axis). Since leverage is computed solely from the explanatory variables, this result applies identically to both outcome variables, turnout and incumbent vote. Particularly unfortunate is that both of these extremely high leverage observations happen to also have extremely (and unrealistically) low reported turnout rates (1.78 and 2.59 percentage points, respectively) and PRI vote shares (0.74 and 1.08 percentage points). Judging from how the data were constructed, the considerably larger vote and turnout results based on official registration, the much larger vote and turnout national figures, or by comparisons with similar precincts, it is likely that these extremely low turnout and vote percentages are incorrect.

We have thus demonstrated all the conditions for bias in the main results: errors in the data that matter, substantial imbalance in the treated and control distributions, huge outliers highly correlated with a covariate included in the regression, and extreme outcome values for the same outlier units.

**Unconventional Model Specifications** Finally, we consider the unusual model specifications in De La O (2013, 2015). Across electoral studies in American and comparative politics, researchers almost always measure the vote for an incumbent party as the *incumbent vote share* — the number of ballots cast for the incumbent party divided by the total number of ballots cast. Researchers from most countries then typically model vote share by including a lagged value of vote share as a control variable.

In contrast, De La O (2013, 2015) measured the incumbent vote for a party as the number of voters for that party divided by the *voting age population* — including voters, nonvoters, noncitizens, those ineligible to vote for other reasons, etc. The article then includes a lagged control variable which was not the incumbent vote share, and not the voting age population, but instead the number of people in the *total population* — including the voting aged population as well as all those under 18. As a result of this unusual decision, the outcome variable in this analysis does not add to 100% across parties, or across parties plus nonvoters, and the outcome and control variables are not logically or necessarily related. Although the model is theoretically possible — more noncitizens or infants in an area could in principle lead to more incumbent voting — this idea was not suggested in De La O (2013, 2015); indeed, to our knowledge, no other published statistical model of the electoral politics in any country and election has ever made these assumptions or chosen this type of specification.

More importantly, because counts, unlike percentages, are unbounded, outliers on this scale greatly exacerbate the bias that results from influential outliers. This is why, as we will show, either correcting the coding errors or switching to the dominant method of constructing vote variables eliminates any support for the effects of programmatic policies on partisan outcomes.

# References

Achen, Christopher H. and Larry M. Bartels (2004): *Blind Retrospection Electoral Responses to Drought, Flu, and Shark Attacks*. Working Paper. Princeton University. URL: https://goo.gl/C4te4H.

Alesina, Alberto and Guido Tabellini (July 1990): "A Positive Theory of Fiscal Deficits and Government Debt". In: *The Review of Economic Studies*, no. 3, vol. 57, pp. 403–414.

Behrman, Jere R. and Petra E. Todd (Mar. 1999): *Randomness in the Experimental Samples of Progresa (Education, Health, and Nutrition Program)*. Research Report. http://goo.gl/YLWz8x. 2033 K Street, NW Washington, DC 20006: International Food Policy Research Institute.

Blattman, Chris, Mathilde Emeriau, and Nathan Fiala (Jan. 2016): *Do employment programs buy votes? A field experiment in Uganda*. Working Paper. http://chrisblattman.com/research. Columbia University.

Calvo, Ernesto and Maria Victoria Murillo (Oct. 2004): "Who Delivers? Partisan Clients in the Argentine Electoral Market". In: *American Journal of Political Science*, no. 4, vol. 48, pp. 742–757.

Camp, Roderic A. (2013): "The 2012 Presidential Election and What It Reveals about Mexican Voters". In: *Journal of Latin American Studies*, no. 451-481, vol. 45.

Cornelius, Wayne A. (2004): "Mobilized Voting in the 2000 Elections: The Changing Efficacy of Vote Buying and Coercion in Mexican Electoral Politics". In: *Mexico's Pivotal Democratic Election: Candidates, Voters, and the Presidential Election of 2000*. Ed. by Jorge I. Dominguez and Chappell Lawson. Stanford University Press, pp. 47–65.

Dahlberg, Matz and Eva Johansson (Mar. 2002): "On the Vote-Purchasing Behavior of Incumbent Governments". In: *American Political Science Review*, no. 1, vol. 96, pp. 27–40.

De La O, Ana L. (Jan. 2013): "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from A Randomized Experiment in Mexico". In: *American Journal of Political Science*, no. 1, vol. 57, pp. 1–14.

— (2015): *Crafting Policies To End Poverty In Latin America: The Quiet Transformation*. Cambridge University Press.

Diaz-Cayeros, Alberto, Federico Estevez, and Beatriz Magaloni (2009): "Welfare Benefits, Canvassing, and Campaign Handouts". In: *Consolidating Mexico's Democracy: The 2006 Presidential Campaign in Comparative Perspective*. Ed. by Jorge I. Dominguez, Lawson Chappell, and Alejandro Moreno. Johns Hopkins University Press, pp. 229–245.

— (2016): *The Political Logic of Poverty Relief: Electoral Strategies and Social Policy in Mexico*. Cambridge University Press.

Dixit, Avinash and John Londregan (1996): "The Determinants of Success of Special Interests in Redistributive Politics". In: *Journal of Politics*.

Duch, Raymond, Wojtek Przepiorka, and Randolph Stevenson (2015): "Responsibility Attribution for Collective Decision Makers". In: *American Journal of Political Science*, no. 2, vol. 59, pp. 372–389.

Evans, Thomas (2006): "The Effects of Discretionary Federal Spending on Parliamentary Election Results". In: *Economic Inquiry*, no. 2, vol. 44, 234:248.

Finan, Frederico and Laura Schechter (2012): "Vote-Buying and Reciprocity". In: *Econometrica*, no. 2, vol. 80, pp. 863–881.

Fowler, Anthony and Andrew B. Hall (2016): *Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent FInding on Voter Competence*. Working Paper. University of Chicago. URL: j.mp/FHsharks.

Franzese, R.J. (2002): *Macroeconomic policies of developed democracies*. New York: Cambridge University Press.

Frey, Anderson (Oct. 2015): *Cash Transfers, Clientelism, and Political Enfranchisement: Evidence from Brazil*. Working Paper. http://goo.gl/RQlJzB. University of British Columbia.

Golden, Miriam and Brian Min (2013): "Distributive politics around the world". In: *Annual Review of Political Science*, vol. 16, pp. 73–99.

Green, Tina (2006): "Do Social Transfer Programs Affect Voter Behavior? Evidence from PROGRESA 1997-2000". PhD thesis. UC Berkeley.

Gwatkin, Davidson R., Adam Wagstaff, and Abdo S. Yazbeck, eds. (2005): *Reaching the Poor*. j.mp/reachpoor. Washington, D.C.: The World Bank.

Hall, Anthony (2006): "From *Fome Zero* to *Bolsa Família*: Social Policies and Poverty Alleviation under Lula". In: *Journal of Latin American Studies*, vol. 38, pp. 689–709.

Hicken, Allen (2011): "Clientelism". In: *Annual Review of Political Science*, no. 289-310, vol. 14.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart (2007): "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". In: *Political Analysis*, vol. 15, pp. 199–236. URL: j.mp/matchP.

Hunter, Wendy and Timothy J. Power (2007): "Rewarding Lula: Executive Power, Social Policy, and the Brazilian Elections of 2006". In: *Latin American Politics and Society*, no. 1, vol. 49, pp. 1–30.

IFE (Oct. 2013): *Estudio Censal de la Participación Ciudadana en las Elecciones Federales de 2012*. Report. http://goo.gl/cYPqML. Dirección Ejecutiva de Capacitación Electoral y Educación Cívica. Instituto Federal Electoral.

Imai, Kosuke, Gary King, and Clayton Nall (2009a): "Replication Data for: The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation". In: hdl:1902.1/11047 [UNF:3:jeUN9XODtYUp2iUbe8gWZQ==] Harvard Dataverse [Distributor].

— (2009b): "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation". In: *Statistical Science*, no. 1, vol. 24, pp. 29–53. URL: j.mp/essrole.

Imai, Kosuke, Gary King, and Elizabeth A. Stuart (2008): "Misunderstandings Among Experimentalists and Observationalists about Causal Inference". In: *Journal of the Royal Statistical Society, Series A*, vol. 171, part 2, pp. 481–502. URL: j.mp/misunEO.

Keefer, Philip and Stuti Khemani (2009): "When do Legislators Pass on Pork? The Role of Political Parties in Determining Legislator Effort". In: *American Political Science Review*, no. 1, vol. 2009, pp. 99–112.

King, Gary, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Clayton Nall, Ryan T. Moore, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas (2009a): "Public Policy for the Poor? A Randomised Assessment of the Mexican Universal Health Insurance Programme". In: *The Lancet*. http://j.mp/lgZF4R.

— (2009b): "Replication Data for: Public Policy for the Poor? A Randomized Ten-Month Evaluation of the Mexican Universal Health Insurance Program". In: hdl:1902.1/11044 [UNF:3:jeUN9XODtYUp2iUbe8gWZQ==] Harvard Dataverse [Distributor].

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas (2007): "A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program". In: *Journal of Policy Analysis and Management*, no. 3, vol. 26. http://j.mp/lXnta2, pp. 479–506.

Kitschelt, Herbert and Steven Wilkinson (2007): "Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition". In: ed. by Herbert Kitschelt and Steven Wilkinson. Cambridge University Press. Chap. Citizen-politician linkages: an introduction.

Klarman, Michael J. (1997): "Majoritarian Judicial Review: The Entrenchment Problem". In: *The Georgetown Law Journal*, vol. 85, pp. 491–554.

Labonne, Julien (2013): "The local electoral impacts of conditional cash transfers". In: *Journal of Development Economics*, vol. 104, pp. 73–88.

Larreguy, Horacio, John Marshall, and Laura Trucco (Nov. 2015): *Breaking Clientelism or Rewarding Incumbents? Evidence from an Urban Titling Program in Mexico*. Working Paper. http://goo.gl/P4YXky. Harvard University.

Layton, Matthew L. and Amy Erica Smith (2015): "Incorporating Marginal Citizens and Voters: The Conditional Electoral Effects of Targeted Social Assistance in Latin American". In: *Comparative Political Studies*, no. 7, vol. 48, pp. 854–881.

Levitt, Steven and James M. Snyder (1997): "The Impact of Federal Spending on House Election Outcomes". In: *Journal of Political Economy*, no. 1, vol. 105, pp. 30–53.

Levy, Santiago (2006): *Progress Against Poverty: Sustaining Mexico's Progresa-Oportunidades Program*. Brookings.

Manacorda, Marco, Edward Miguel, and Andrea Vigorito (July 2011): "Government Transfers and Political Support". In: *American Economic Journal*, no. 3, vol. 3, pp. 1–28.

Ortiz, Mauricio (2006): *El Seguro Popular: Una Cronica de la Democracia Mexicana*. Fondo de Cultura Economica.

Pop-Eleches, Cristian and Grigore Pop-Eleches (2012): "Targeted Government Spending and Political Preferences". In: *Quarterly Journal of Political Science*, no. 3, vol. 7, pp. 285–320.

Posner, Eric A. and Adrian Vermeule (May 2002): "Legislative Entrenchment: A Reappraisal". In: *The Yale Law Journal*, no. 7, vol. 111, pp. 1665–1705.

Primo, David M. and James M. Snyder (Apr. 2010): "Party Strength, the Personal Vote, and Government Spending". In: *American Journal of Political Science*, no. 2, vol. 54, pp. 354–370.

Rocha-Menocal, Alina (2001): "Do Old Habits Die Hard? A Statistical Exploration of the Politicisation of Progresa, Mexico's Latest Federal Poverty-Alleviation Programme, under the Zedillo Administration". In: *Journal of Latin American Studies*, vol. 33, pp. 513–538.

Skoufias, E. (2005): *PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico*. International Food Policy Research Institute. http://goo.gl/ROJeRM. Washington.

Skoufias, Emmanuel, Benjamin Davis, and Jere R. Behrman (June 1999): *An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition*

*Program (PROGRESA) of Mexico*. Research Report. 2033 K Street, NW Washington, DC 20006: International Food Policy Research Institute. URL: goo.gl/AVamve.
Sterk, Stewart E. (Apr. 2003): "Retrenchment on Entrenchment". In: *The George Washington Law Review*, no. 2, vol. 71, pp. 231–254.
Stokes, Susan C. (2005): "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina". In: *American Political Science Review*, no. 3, vol. 99, pp. 315–325.
Stokes, Susan C., Thad Dunning, Marcelo Nazareno, and Valeria Brusco (2013): *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge University Press.
Thompson, Dennis F. (June 2005): "Democracy in Time: Popular Sovereignty and Temporal Representation". In: *Constellations*, no. 2, vol. 12, pp. 245–261.
Weingast, Barry R., Kenneth Shepsle, and Christopher Johnsen (Aug. 1981): "The Political Economy of Benefits and Costs: A Neoclassical Approach". In: *Journal of Political Economy*, no. 4, vol. 89, pp. 642–664.
Zucco, Cesar (2008): "The President's 'New' Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Elections". In: *Journal of Latin American Studies*, no. 1, vol. 40, pp. 29–49.
— (2013): "When Payouts Pay Off: Conditional Cash Transfers and Voting Behavior in Brazil 2002-10". In: *American Journal of Political Science*, no. 4, vol. 57, pp. 810–822.