# WHO *REALLY* VOTES? RETHINKING THE EFFECTS OF SOCIODEMOGRAPHIC CONDITIONS ON VOTING IN MEXICO IN LIGHT OF TURNOUT OVERESTIMATION

David Crow
Dept. of Government, UT Austin
dbcrow@mail.utexas.edu

Edmundo Berumen
Berumen y Asociados
direccion@berumen.com.mx

Abstract: One of the most constant relationships to emerge from the systematic study of politics is the relationship between sociodemographic variables (such as age, education, income, and employment) and voting. However, as in other countries, these results in Mexico are based on self-reported voting behavior. Post-electoral surveys invariably overestimate the proportion of people who vote either because the sample includes too many non-voters and not enough voters or, more likely, respondents claim to have voted when they really did not (i.e., they *overreported*). Under reasonably general conditions, unit non-response and overreporting bias our estimates of predictors' effects on turnout. We parse the problem of turnout overestimation in Mexico's 2000 presidential election and test three corrective methods: (A) survey "filters" to reclassify as non-voters respondents unlikely to have voted, (B) sample weights derived from aggregate data, and (C) a sample selection model with Heckman-type correction. The results of all three methods suggest that failing to account for turnout overestimation exaggerates the effects of sociodemographic variables on turnout.

I. Introduction

One of the most constant causal relationships to emerge from the systematic investigation of politics is that between sociodemographic conditions and voting. Numerous studies in established democracies have found that citizens' social and economic circumstances wield decisive influence on their choice of whether or not to cast ballots in popular elections. Education, income, age, and employment status (among other variables) repeatedly appear as determinants of electoral turnout. By virtually all accounts, better educated, higher-earning, older, employed potential electors are likelier to vote than their less educated, lower-earning, younger, unemployed counterparts. Recent research on Mexican elections indicates that the same relationship appears to obtain south of the Rio Grande.

In this study, we call the received wisdom into question. As in other countries, Mexican surveys that rely on self-reported voting data invariably overestimate the percentage of people who vote. Survey overestimation of turnout typically occurs either because non-voters refused to answer the survey at a greater rate than did voters (i.e., because non-response bias exists) or, more likely, because respondents report voting when they did not actually do so—that is, respondents *overreported* turnout. Under reasonably general conditions, turnout overestimation may bias our estimates of predictors' effects on the decision to vote. If, say, lower-income earners as a group are likelier to inflate their electoral participation or to enter the sample than their better-off peers,[1] turnout models may underestimate the effects of income on the probability of voting.

Here, we argue that turnout overestimation does, indeed, present an inaccurate picture of who really votes in Mexico. Using data from the Comparative Study of Electoral Systems

[1] In contrast to the United States, Mexican surveys tend to over-represent those who earn less. This may be true because higher earners are simply too busy to make themselves available to interviewers, especially given that most interviews are conducted face-to-face in respondents' homes.

(CSES) 2000 post-electoral survey, as well as aggregated vote and census data, we examine overestimation of electoral participation in Mexico and test three methods to correct for it: 1) survey "filters", 2) reweighting, and 3) Heckman-type sample selection models. Since we believe inaccurate responses to be the main culprit of post-electoral surveys' inflation of turnout figures, we focus on overreporting. However, for the latter two corrective techniques, it is immaterial whether overestimation is caused by turnout overreporting or non-response bias.

This paper proceeds as follows. We begin with a discussion of turnout studies on Mexican elections and the magnitude of the overreporting problem there. This discussion invokes voting research in the United States as a foil to Mexico. In particular, the National Election Survey's (NES) validated vote studies—in which researchers checked self-reported voting against official records to determine who had really voted—led to a large body of work on *who* overreports, whether overreporting affects statistical inference, and how to correct for overreporting bias.

Next, we explain how turnout overreporting may distort our knowledge of factors that influence the individual decision to vote. We begin with a heuristic explanation of overreporting bias, which we then formalize mathematically. Though there are conditions under which self-reported voting data do *not* bias our estimates of the determinants of turnout, these conditions appear to be rare and exceedingly difficult to meet in any country, including, of course, Mexico.

We then explain and test three methods to correct for overreporting bias. Since validated voting data are unavailable in Mexico, we must turn to methods other than those derived from the NES studies. The first method uses survey "filter" items to reclassify as non-voters respondents people unlikely to have voted and redoes the analysis on the reclassified sample. In effect, this method amounts to a sort of poor man's vote validation. The second and third methods both use auxiliary information gleaned from the known population values contained in aggregate data to adjust estimates based on individual-level data. We reweight the sample using weights derived from official voting data and census information that contains known population quantities. Finally, we estimate a sample selection model with Heckman-type correction.

We find that correcting for turnout overestimation using all three methods reduces our estimates of sociodemographic variables' effects on voting. Age, education, and, to a lesser extent, employment become insignificant predictors of electoral participation after taking into account the effects of turnout overestimation. On the other hand, political attitudes and beliefs demonstrate remarkable perseverance, across all models, in predicting voting. Specifically, confidence that the ballots will be counted fairly and a belief that parties represent their constituents well weigh more in the decision to vote than does respondents' social background.

In our concluding evaluation of the exercise, we sketch several tentative explanations for this surprising result, including the special foundational nature of the 2000 Mexican elections and the possibility that sociodemographic factors continue to influence turnout *indirectly* through attitudes and beliefs. Yet the implication of our results is clear: turnout studies, in Mexico and perhaps generally, may unduly emphasize the role of sociodemographic attributes in getting people to the polls.

## II. Voting and Turnout Overestimation in Mexico and the United States

### *Determinants of Voting*

Political scientists have recently begun to study the social and attitudinal factors that shape the individual decision to vote in Mexico. Several studies have highlighted significant relationships between sociodemographic conditions and the proclivity to vote. Higher income earners vote more than those who earn less; employment and greater education also leads to higher voting rates; and age is positively related to the tendency to vote (Buendía 2002: 449; Buendía and Somuano 2003: 320; Moreno 2003: 152, 160).[2]

Beyond sociodemographic characteristics, however, attitudes and psychological attributes play an important role in orienting an individual's voting behavior. For example, those with stronger affective affinity with any political party are likelier to vote than those without partisan ties (Moreno 2003: 160). Similarly, citizens who exhibit higher levels of political trust and connection to the political system vote more than the mistrustful and alienated (Buendía 2002: 449), as do those who believe that political parties are responsive to social needs (Buendía and Somuano 2003: 320). Greater information and political sophistication also lead to more voting (Moreno 2003: 160).

Finally, electoral campaigns are effective in getting people to the polls. Electors who received some sort of campaign communication—flyers, letters, and door-to-door stumping—showed up at the polls in greater proportion than those not contacted by campaigns (Buendía and Somuano 2003: 320).

These results are broadly consistent with studies on turnout in the United States. Voting research there has regularly pointed to socioeconomic status as a predictor of voting. A great many studies have shown that education, age, income, and social status (e.g., occupational prestige) are positively related to voting (see, e.g., Milbraith and Goel 1977, Wolfinger and Rosenstone 1980). As in Mexico, psychological factors such as interest in politics and affective identification with a political party also make U.S. citizens likelier to vote (Silver *et al.* 1986).

The same explanations for these relationships have been marshalled in both countries. Higher education and income levels entail greater access to information and increased cognitive ability to process that information, which, in turn, translates to an enhanced sense of political

[2] Moreno uses a pre-electoral survey and bases his conclusions on "probable voters".

efficacy. Those higher up on the social ladder perceive they have more opportunity to influence political events, making them likelier to vote. Furthermore, the better off feel a greater "stake in society" (Silver *et al.* 1986: 614), which redounds in heightened emotional attachment to the political system and, consequently, higher voting rates.

One important difference between the Mexican and U.S. electorates, however, is the existence in Mexico of widespread citizen distrust of elections formed under the country's dominant-party authoritarian regime. For almost seven decades, elections in Mexico were legitimizing window-dressing rather than a real opportunity to select political leadership. Despite general agreement that the midterm election of 1997 was largely clean, suspicion about the counting and reporting of votes from the days of single-party rule lingered inertially in 2000. Mexicans who believe that elections are rigged and electoral authorities, biased are understandably less likely to vote (Moreno 2003: 320). In Buendía and Somuano's words, "[A]lthough electoral participation in Mexico may be explained through some of the same factors that explain this phenomenon in established democracies, it is worrisome that political attitudes developed in the *ancien régime* continues to have such a large impact on participation rates" (2003: 318).[3] In contrast, there is no comparable legacy of electoral distrust in the United States, notwithstanding the conflict following the controversial 2000 presidential election.

*The Magnitude of Overreporting*

In both Mexico and the U.S., survey estimates of turnout based on self-reported behavior invariably inflate the percentage of voters, sometimes wildly. This problem has long been noted

[3] Unfortunately, the controversy surrounded the 2006 general election has resuscitated distrust in Mexico's electoral authorities. In the months following the elections, about half of the electorate believed there were "irregularities" (*El Universal*, "Comicios dañaron la imagen del IFE, revelan", November 10, 2006) and the majority called Mexico "undemocratic" (Parametría poll, September, 2006, http://www.parametria.com.mx/es_cartaext.php?id_carta=165). Even after political passions had subsided, a year later one-third of Mexicans persist in the belief the elections were fraudulent (Consulta Mitofsky, Bulletin #221, August 2, 2007). For many Mexicans, the "*ancien régime*" is not as *ancien* as they would like.

in the context of U.S. national elections. In the National Election Survey (NES), undertaken after every presidential and congressional election since 1948, the percentage of respondents who report having voted is always higher than turnout figures provided by the Federal Election Commission (FEC). In an early study on overreporting, Wolfinger and Rosenstone (1980) found that the difference was never less than five percent and sometimes climbed to nearly 20 percent (1980: 115). The gap in the 1984 presidential election was 18.3% and in 1988, 19.1% (Deufel and Kedar 2000: 24). Based on "validated vote data",[4] Silver *et al*. give even higher figures: 27.4% in 1964, 31.4% in 1976, 22.6% in 1978, and 27.4% in 1980 (1986: 613). In short, a lot more people said they had voted than actually did.

Mexican citizens are not immune to the temptation to claim they voted when they did not. A summary recitation of figures from polls undertaken after the 2000 presidential elections gives an idea of the degree of overreporting. Though the Federal Electoral Institute (IFE) reports an official turnout rate of 64.0%,[5] four post-electoral surveys yield higher figures. First, the Mexican 2000 Post-Electoral Survey's self-reported turnout rate was 84.2%.[6] Second, a 2000 post-electoral survey undertaken by Víctor Manuel Durand Ponte, of the National University of Mexico's (UNAM) Institute for Social Research, registers an electoral participation rate of 83.9%.[7] Third, the Mexican version of the 2000 Comparative Study of Electoral Systems (CSES) poll yields a rate of 81.8%.[8] Finally, the first *Encuesta Nacional de Cultura Política* (ENCUP, undertaken in November, 2001, by the Mexican Interior Ministry), informs us that

---

[4] The NES validated vote studies compare survey respondent's self-reported voting to actual voting records kept by county registrars.

[5] Information taken from the "Resultados Electorales" section of the IFE Web site (http://www.ife.org.mx).

[6] Principal Investigator Chappell Lawson, NSF Grant No. SES-9905703. Data available on-line at http://web.mit.edu/polisci/faculty/C.Lawson.html#books.

[7] See Durand Ponte (2003) for survey details. The authors thank Prof. Durand Ponte for furnishing the database of his 2000 survey.

[8] This poll was undertaken by the Office of the Presidency during the Zedillo administration. It is stored in the Banco de Información de Opinión Pública of the Centro de Investigación y Docencia Económicas (CIDE, http://biblioteca.cide.edu/ encuestas.htm) and may be purchased for a nominal fee.

73.5% of those eligible cast a ballot. Thus, Mexicans overreported their participation at the polls in 2000 by a margin of 9.5% to 20.3%.

*The Causes of Overreporting*

Why do survey respondents say they voted when they did not? Scholars have found the most credible explanation in so-called "social desirability theory." That is, survey respondents wish to represent themselves to interviewers as good citizens. As Silver *et al.* put it, "Overreporting is seen to result from the respondent's desire to please the interviewer and to appear to engage in socially desirable behavior" (1986: 613). Humans naturally want to leave others with a good impression, even in the relatively impersonal interaction that occurs during a survey interview, and this desire frequently overcomes the countervailing wish to be truthful.

Complementary accounts of overreporting focus on memory lapses. Belli *et al.* demonstrate that the greater the time elapsed between elections and the post-electoral survey,[9] the greater the probability of vote *misreporting*—which includes underreporting, or denying having voted when one did in fact vote, as well as overreporting (1999: 99-103). However, misreporting is almost entirely unidirectional in favor of overreporting. For example, using validated vote data, Sigelman reports that just 1.8% of survey respondents underreported voting in 1978 (1982: 49). In 1988, just 12 respondents who said they did not vote were confirmed as voters, a fact Katz suggests may be attributable to sloppy record keeping (2000: 10). The lopsidedness of misreporting has been (probably rightly) interpreted as buttressing social desirability theory (Katz 2000: 2; Presser and Traugott 1992: 78).

But defective memory and social desirability are not incompatible explanations. Interviewees may genuinely confuse instances of where they actually voted with occasions

[9] Not to be confused with the "exit poll", in which pollsters interview voters as they exit the voting booths (as opposed to a face-to-face or phone interview several weeks or months after the election).

where they only thought or talked about voting, especially when the survey interview is long after the elections. In turn, a sort of "internal social desirability", the need to conceive of oneself as a responsible citizen, may slant the mnemonic confusion in favor of overreporting (Belli *et al.* 1999: 91-92). We believe the same reasons that lead U.S. citizens to say they voted although they really did not—wanting to look good to interviewers and inaccurate recall—also cause overreporting in Mexico.

*Correcting for Turnout Overreporting in the U.S.*

Whatever the reasons, overreporting poses problems not only for estimating turnout[10] but also, more perniciously, for determining its causes. Aware of this problem, U.S. scholars began to investigate two related questions. First, are overreporters distributed evenly among different social segments or, on the contrary, do certain classes of survey respondents tend to overreport more than others? Second, if the latter, does this alter previous research's substantive conclusions about who votes and what factors affect turnout?

To answer these questions, the NES undertook "validated vote" studies in 1964 and in every election from 1972 to 1990 (with the exception of 1982). Since whether an individual voted in a given election (although *not* for whom she voted) is a matter of public record in the United States, it was possible for survey researchers to register respondents' identities and compare self-reported vote data against official records. The validated voting data allowed students of voting behavior to determine who the overreporters were and, by comparing results from models using self-reported voting with those using validated data, reassess findings of prior turnout studies (see, e.g., Sigelman 1982, Abramson and Claggett 1984, and Silver *et al.* 1986).

[10] Although official turnout figures in Mexico provide better estimates of turnout than do post-electoral surveys, the former are not exempt from problems. Failure to remove ineligible voters from the rolls means that the denominator in the fraction that gives turnout (votes cast over eligible voters) is too high and, consequently, the turnout proportion too low.

The results suggest that overreporting did, in fact, distort estimation of turnout predictors' effects. Silver *et al.* find that overreporters tended to be better educated, feel a greater sense of civic duty, identify more with one party or another, and exhibit greater interest in politics than the general population (1986: 615-617). As a result, analyses based on self-reported voting overestimated the strength of the relationship between education, partisanship, and political interest (on the one hand) and voting (on the other).[11] In other words, overreporters tended to be people who, because of their education, interest in politics, and attachment to a political party, feel they *should* have voted, but for some reason did not.

Furthermore, African Americans appear to overreport more than members of other ethnic groups (Abramson and Claggett 1991). Consequently, most U.S. analyses *underestimate* the effect of ethnicity on turnout (Deufel and Kedar 2000: 3-5). This suggests that social desirability may exert greater pressure on some social and ethnic groups than others. In this case, the powerful legacy of the Civil Rights movement and the 1965 Voting Rights Act may have inculcated in African Americans a sense that they had a special responsibility to vote.

Despite the hazards of self-reported voting data, only one of the studies on turnout in Mexico (Buendía and Somuano 2003) attempts to correct for overreporting. Noting that the CSES survey inflates turnout, Buendía and Somuano effect a partial correction and recategorize as non-voters those self-reported voters whose electoral ID cards were not marked to reflect that they voted. As we suggest below, this recoding procedure probably does not affect their analysis substantially compared to the results they would have obtained had they simply ignored the problem of overreporting.

[11] There are dissenters from this view, however. For instance, Sigelman avers that overreporting does not affect substantive conclusions about factors associated with voting: "[A]nalysis indicates that no major differences emerge when identical discriminant models of voting are fitted using respondent-reported and officially validated voting data" (1982: 54).

Unfortunately, validated voting data are unavailable in Mexico. Electoral law keeps individual voting information strictly confidential and mandates destruction of balloting material and voter rosters six months after the election (if no challenges are pending).[12] Thus, we must resort to other methods to correct for the effects of overreporting (or non-response bias), which we set forth in the Section IV.

## III. Overreporting and Statistical Inference

In this section, we explain how overreporting may bias inferences about predictors in turnout models. We start with a heuristic presentation of the problem and then develop it formally. If the propensity to overreport is systematically related to characteristics that influence voting—if, for example, overreporters are on the whole wealthier or poorer, more or less educated, and so on, than the average survey respondent and if wealth, education, etc., also ostensibly influences voting—we may over- or underestimate the true effects of factors associated with turnout. In other words, non-random overreporting among survey respondents will bias parameter estimates if the correlates of overreporting overlap with the putative correlates of turnout. If, on the contrary, the subsample of overreporters mirrors that of the total sample in characteristics related to voting, we may be reasonably confident in the magnitude of predictors in turnout models (provided, of course, that the other usual statistical assumptions hold).

A simple example will suffice to see how overreporting might bias our parameter estimates upward or downward. Let us suppose that voters earn $7,500 pesos per month; non-voters, $2,000; overreporters, $3,500; and that income is the only variable that affects the

[12] Some election authorities are aware of the desirability of validated vote data. In a phone conversation with Edmundo Berumen, one former IFE official said that making individual voting information available to scholars had been proposed but vetoed by a political party's representatives on the election board.

likelihood of voting (i.e., there is no specification error). Further suppose that these types of respondents are distributed as follows: voters, 50%; non-voters, 25%; and overreporters, 25%.

Here, relying on the self-reported data underestimates the effect of income on the chances of voting. Figure 1 presents the results of two linear probability models (LPM's) for the simulated data, one for the self-reported voting data (the dashed line), which counts overreporters as voters, and another for the "real" voting data (the solid line), which correctly classifies overreporters as non-voters. For the "real" data, the LPM estimates an increase of 20.1% in the probability of voting for every $1,000 pesos that monthly income goes up. For the self-reported data, the rate of increase is just 13.2% for every $1,000 increment, as reflected in the dashed line's flatter slope. [13]

[Figure 1 here]

Because overreporters' mean monthly income ($3,500 pesos) is below that of the combined mean for voters and non-voters ($5,667 pesos), erroneously counting overreporters as voters (represented by the triangle at [X = $3,500, Y = 1]) pulls up the left end of the regression line from where it would be if overreporters were rightly reckoned as non-voters (represented by the circle at [X = $3,500, Y = 0]). The more even regression line underestimates the effect of income on the chances of voting.[14]

On the other hand, if overreporters' mean income is equal to the sample mean, overreporting doesn't bias parameter estimates. (In our example, this happens when overreporters earn $5,667 per year.) Figure 2 shows LPM's for both the self-reported and the

[13] Though we use linear probability models for ease of exposition, logit and probit models would produce essentially the same results.

[14] Technically, random overreporting requires that in this case, in addition to overreporters' mean income being the same as that for the entire sample, overreporters be drawn from the same probability distribution as other respondents. This implies that the mean, variance, and shape (or functional form) of the income distribution are the same for overreporters as for other respondents. Nonetheless, a difference in mean income between overreporters and other respondents is a condition sufficient to conclude that overreporters are a non-random subset of all respondents.

"real" data when overreporters' income is distributed similarly to that of all respondents. Although the point estimates for the probability of voting are different, their slopes are identical. Both models estimate the same effect of income on voting: for every $1,000-peso increment in income, the chances of voting go up (linearly) by 18.2%.

[Figure 2 here]

This logic may be readily extended to the case where overreporters' mean income is *above* that of the sample mean. Here, counting overreporters as voters would pull up the right end of the regression line for self-reported data, giving it a steeper slope than that for the actual voting data. In other words, if overreporters earn more than the average respondent, overreporting will *overestimate* the effect of income on voting.

There is a hypothetical case in which all respondents report voting or abstaining truthfully (i.e., no misreporting occurs) and parameter estimates are nonetheless biased. This happens when someone selected into the survey sample refuse (or otherwise fail) to answer a survey. Including a disproportionate number of voters in a survey sample may also distort the impact of qualities such as education and income on voting, if the sample is not representative of the population on these variables. "Sample selection bias" to econometricians (Heckman 1976, 1979), and "unit non-response" or "sample truncation" to survey researchers. The second and third methods we propose below to correct parameter estimates in the presence of turnout overestimation are also appropriate when non-random refusals to participate in surveys, rather than overreporting, is inflating turnout estimates.

Having presented the problem of overreporting intuitively, we now formalize it mathematically. In turnout prediction models, the dependent variable is the probability a citizen will vote conditional upon her socioeconomic circumstances, political attitudes, and so on.

Unfortunately, we cannot observe whether a respondent voted or not, only whether she *says* did. Table 1 is a cross-classification of self-reported voters with actual voters, or so-called "truth table":

[Table 1 here]

The column marginals represent the unconditional probability that a respondent *voted* (taken from official turnout figures) and the row marginals represent the unconditional probability that a respondent *reported voting* (taken from a post-electoral survey). Of course, without validated voting data we cannot know what percentage of those who said they voted really did vote (upper left-hand cell) and what percentage fibbed (upper right-hand cell). In other words, we don't know the *cell* or *inner* percentages, as indicated by the question marks in the cells.

Following Katz (2000: 9-10), we assume a "0" in the lower left cell (see Table 2). This amounts to a belief that all survey respondents who admitted not voting are telling the truth—i.e., that there is no underreporting (respondents who deny having voted when they really did so). For any given distribution of marginal probabilities, the "structural zero" also determines the cell values. The inner probabilities for our example are shown in Table 2:

[Table 2 here]

Besides simplifying the arithmetic, postulating a structural zero for the frequency of underreporting is justified on three grounds. First, it is consistent with the expectation derived from social desirability theory that a respondent who can report voting truthfully will do so. Second, the NES validated vote studies show that the number of respondents who underreport is empirically negligible—a condition we believe may hold in Mexico as well. Finally, there is evidence that, in contradistinction to overreporting, underreporting *is* random (see Silver *et al.*

1986, Presser and Traugott 1992). Therefore, it does not affect estimation of coefficients for turnout predictors.

To predict the probability of really voting based on self-reported data, we need a model that adjusts for the probability of misreporting. We may write the probability of voting:

$$\Pr(V=1 \mid \mathbf{x}) = \Pr(V=1 \mid RV=1, \mathbf{x}) * \Pr(RV=1 \mid \mathbf{x}) + \Pr(V=1 \mid RV=0, \mathbf{x}) * \Pr(RV=0 \mid \mathbf{x}),$$

(Eq. 1)

where "V" is "voted", "RV" is "reported voting", "1" is "yes", "0" is "no", and **x** is a vector of regressors. Making use of our assumption that there is no underreporting (i.e., that Pr(V=1| RV=0) = 0) and rearranging the terms of the equation, we may express the conditional probability of reporting truthfully as:

$$\Pr(V=1 \mid RV=1, \mathbf{x}) = \frac{\Pr(V=1 \mid \mathbf{x})}{\Pr(RV=1 \mid \mathbf{x})}$$

(Eq. 2)[15]

Adapting Katz, we rewrite Eq. 1 to express the probability of reporting voting as some function of the probability of voting and an adjustment for misreporting:

$$\Pr(RV=1 \mid \mathbf{x}) = f([1 - \Pr(RV=1 \mid V=0, \mathbf{x}) - \Pr(RV=0 \mid V=1)] * \Pr(V=1 \mid \mathbf{x}) + \Pr(RV=1 \mid V=0, \mathbf{x}))$$

(Eq. 3) [16]

Katz gives two conditions under which parameter estimates in turnout models will be the same for self-reported and actual voting behavior (2000: 3-4):

---

[15] We provide the algebraic derivations of these formulas for interested readers in the Appendix A. Note that since we assume no underreporting, $\Pr(V=1 \mid \mathbf{x}) \le \Pr(RV=1 \mid \mathbf{x})$; thus, $\frac{\Pr(V=1 \mid \mathbf{x})}{\Pr(RV=1 \mid \mathbf{x})} \le 1$. That is, if no people report abstaining when they really voted, the probability of voting will always be equal to or less than the probability of reporting voting, ensuring that the right-hand side of Eq. 2 remains in the interval [0,1].

[16] Under our assumption of no underreporting, the term Pr(RV=0 | V=1,x) would drop out of Eq. 3. We retain this term temporarily, however, to specify conditions under which regressions using self-reported voting yield the same results as if accurate voting data had been used.

1. $\Pr(RV=1 \mid V=0,\mathbf{x}) = \Pr(RV=0 \mid V=1,\mathbf{x}) = 0$

2. $\begin{array}{l} \Pr(RV=1 \mid V=0,\mathbf{x_i}) = \pi^1, \forall_i \\ \Pr(RV=0 \mid V=1,\mathbf{x_i}) = \pi^2, \forall_i \end{array}$, where $\pi^1$ and $\pi^2$ are constants.

In words, estimates of the independent variables' effects using self-reported data will equal those using real data when (1) there is no over- or underreporting or (2) the probability of both over- and underreporting is the same for all respondents (i.e., when the propensity to misreport is distributed randomly among survey respondents).

Alternatively, we can give the probability of voting as a function of the probability of self-reported voting and a correction for the probability of reporting truthfully:

$$\Pr(V=1 \mid \mathbf{x}) = f[\Pr(V=1 \mid RV=1,\mathbf{x}) * \Pr(RV=1 \mid \mathbf{x})] \qquad \text{(Eq. 4)}$$

In Eq. 3, the probability of voting is essentially weighted by the probability of misreporting to obtain the probability of *reporting* voting. Eq. 4 effects the reverse operation: weighting the probability of *reporting* voting by the probability of reporting accurately to get the probability of voting.

In both Eqs. 3 and 4, we need an estimate of the likelihood of voting to approximate the probability of misreporting. Since we have individual-level data only on reporting voting, but not voting, we get this estimate from aggregated data. We explain how in the second and third parts of the following section (on reweighting and Heckman-type correction, respectively), in which we propose and test several methods to correct for turnout overreporting.

## IV. Correcting for Turnout Overreporting

In this section, we use the CSES data to test out three methods to correct for turnout overreporting: (1) using survey "filter" questions to recode some respondents as non-voters, (2) reweighting samples with estimates of voting probabilities obtained from aggregated data, and (3) sample selection models with Heckman-type estimators.

*Survey "Filters"*

The first method uses what might be called "survey filters", questions that attempt to distinguish *real* voters from those who reported voting although they had not. In Mexico, a mid-1990's election reform established measures to ensure that citizens could not vote more than once, including marking the voter identifications (*credencial electoral*) of those who cast ballots. The CSES 2000 interview protocol used subterfuge to inspect respondents voter ID's to verify that they had indeed voted. Checking voter ID's is thus an ersatz validation: respondents whose credentials were not marked probably did not vote.

The ruse worked like this. The first question in the survey asked respondents if they had voted. Then, 68 questions later (to forestall the rapport-busting suspicion that interviewers may have been calling respondents' word into doubt), a four-item battery of questions on voter ID's appears (see Appendix B). Respondents are asked if they have a voter ID. Those who responded yes were then asked to produce the credential, ostensibly so they could "point out what they liked" about it and indicate what could be improved. Finally, on the pretext of verifying what sort of mark was used on voter ID's to indicate a citizen had voted (the letter "v" or a circle around the election year), interviewers checked whether the ID bore the corresponding voting mark.

Cognizant of the potential effects of turnout overreporting, Buendía and Somuano reclassify as non-voters those who reported voting in the 2000 presidential election but whose voter cards did not show the corresponding mark. The authors note that this procedure reduces the self-reported turnout rate from around 82% to 76%—still 12% above the official rate, but within the overreporting rates in the NES studies (2003: 291, 318).

We include the same variables used by Buendía and Somuano, but use two different implementations of the survey filter to recode reported voters as non-voters. First, Model A1 (Table 3) follows Buendía and Somuano in classifying non-voters as respondents whose voter ID's did not indicate they had voted. However, in contrast to Buendía and Somuano, we also recode as non-voters cases with "could not see [marking]" or "other" as non-voters, reasoning that had their credentials displayed the requisite markings, respondents would have been eager to make clear to interviewers that the that they had indeed voted (consistent with social desirability theory). This procedure reclassified 103 voters as non-voters (8.5% of the sample, N=1,205) for an estimated turnout rate of 72%.

Model A2 tests another, stricter use of the filter items. Buendía and Somuano give the benefit of the doubt to respondents who reported voting but could not show their ID's, neither reclassifying nor removing them from the sample. We are not so charitable. Since Mexican voter cards bear photos (one of several fraud prevention features), they are widely used as personal identification in the same way that driver's licenses are in the United States. Thus, it is highly probable that someone who has a voting credential will also be carrying it on him or her or at least have it handy. Model A2 deems non-voters those who claimed to have voted but were not carrying their voter ID's, said they didn't have a voting card, or refused to show the voting card. These more rigorous criteria resulted in 207 reassignments (17.2% of the sample), reducing the turnout estimate to 64.6%—close to the official figure.

Table 3 presents Models A1 and A2 together with a baseline model in which none of the respondents was reclassified (Baseline Model I). The 95% confidence interval for the baseline model parameters is provided to facilitate comparisons across models. We constructed the samples so that all three models are estimated using the same observations, deleting cases with any missing values.[17]

The first two columns also present the coefficients and p-values, respectively, from Buendía and Somuano's results. Though the authors do furnish an account of how they construct their variables, the information was not quite sufficient to enable us to replicate their research. While our Model A1 (reclassifying as non-voters those interviewees whose credentials did not indicate they had voted, but leaving in respondents who failed to produce voter ID's) is not, therefore, directly comparable to that of Buendía and Somuano, these authors' parameters provide a point of comparison for our estimates.

Unfortunately, Buendía and Somuano do not provide coefficients for their model using the unfiltered sample. It is therefore impossible to know what corrective effect, if any, reclassifying self-reported voters as non-voters had on their model. Nonetheless, an eyeball comparison of their model with our baseline model reveals substantively very similar results, suggesting that the corrections wrought by the Buendía and Somuano procedure were minimal.

[Table 3 here]

In contrast, our employment of survey filters yielded significantly different results. The most noteworthy effect, especially in the model using the stricter reclassification criteria (Model A2), is to reduce significance levels for our estimates of the effect of age on voting. Comparing Model A1 to Baseline Model I (BM-I), we see that the parameter for the age category "41-60

[17] We are aware of the dangers posed by list-wise deletion of incomplete cases (see King *et al.* 1998). If respondents who complete surveys differ on variables related to voting from respondents who withhold some information from interviewers, list-wise deletion may bias parameter estimates. Nonetheless, including the same respondents in all three models at least maintains any bias constant across them. Although we cannot be sure of the magnitude of the models' parameter estimates in the presence of list-wise deletion, we can be sure that *differences* in magnitude between the filtered models and the baseline model are due solely to the corrective techniques and not to the effects of missing data.

years old" declines dramatically from .603 to .016 (lying outside the 95% confidence interval for BM-I) and exceeds the threshold of p = .05. Model A2 presents even more striking results. The two age categories "41-60" and "Over 60" are now rendered completely insignificant as predictors of voting. It would appear that older citizens are not, in fact, more inclined to vote than their younger compatriots.

On the other hand, psychological variables generally retained or *increased* in importance relative to the baseline model. Partisanship, in particular, achieves a much stronger effect in Models A1 and A2. Both models reinforce the idea of PRI adherents as more electorally mobilized (consistent with Buendía and Somuano's findings). On the other hand, whereas sympathizing with the center-left Party of the Democratic Revolution (PRD) is weakly predictive of voting in the baseline model, affective identification with the PRD exerts an important, significant effect in Model A2. The same result obtains for those who identify themselves with smaller parties (denominated "Other"), though volatility in estimates for party identification and unconventional political participation may owe to the small number of respondents who identify with the PRD or small parties.

Furthermore, political sophistication (a composite variable comprising political knowledge, exposure to news sources, and interest in politics) is an important correlate of voting in the baseline and both filtered models. Also, the perception that parties and politicians are responsive to citizen needs leads citizens to vote more often across all models. Thus, the significance of cognitive and evaluative dimensions in the decision to vote remain unaffected by the recategorization of self-reported voters as non-voters in Models A1 and A2.

Taken together, the chief corrective effects of the survey filters is to cast doubt upon the propensity of citizens over 40 to vote proportionally more than younger Mexicans. On the other hand, political attitudes and assessments continue to figure importantly among the factors that increase citizens' propensity to vote.

*Sample Reweighting*

Another technique to correct for bias caused by turnout overestimation is reweighting the sample. The intuition behind the weighting schemes employed here is simple: give more weight to people who admit not voting—by assumption, they are telling the truth—and less to those who claim to have voted, most of whom will be telling the truth, but some of whom will not. Another way to think of weighting is as a correction for underrepresentation of truthful non-voters in the sample. Raghunathan notes that sample weighting has long been used to compensate for unequal probabilities of selection (2004: 105).

As shown in Eq. 4 above, to get the probability someone voted, we weight her observation of reported voting (RV=1) by the probability she is reporting truthfully (i.e., $Pr(V = 1 \mid RV = 1)$). When the percentage of respondents who reported voting (RV) is greater than the percentage that actually voted (V; i.e., $RV > V$), this weight will adjust RV down to get V.

To carry out this operation, we need an estimate of the probability of voting, $Pr(V=1)$. But the survey gives us only an estimate of *claiming* to have voted, $Pr(RV=1)$. Both Katz (2000) and Brehm (1999), whose method we adapt in the next section on "Heckman-type estimators", suggest that we can use aggregated data, taken as known population values, and use them to correct the estimates of turnout predictors derived from self-reported voting. The data we use here are official nationwide and state-level turnout figures, as reported by the IFE. We also use state-level gender (% male), age (median), education (average years studied), employment (% working over 33 hours per week), and GDP per capita growth between 1999 and 2000, taken from National Institute for Geography, Statistics, and Informatics (INEGI) numbers.

We are well aware of the dangers posed by using aggregate data. They are two-fold in this case. First, as we hint above, the IFE turnout estimates surely *underestimate* real turnout. This is so because voter lists contain many names of people who are ineligible to vote. Death, emigration (especially to the United States), disqualification because of criminal conviction (or another reason), and illegal acquisition of the voter ID (by illegal immigrants in Mexico, especially Central Americans) are the main causes of ineligibility. Audits undertaken after the massive 1994 voter registration drive showed the national voter list[18] to be highly reliable. Nonetheless, that reliability has steadily deteriorated since then, partly because of the time required to get records from the various vital statistics and justice departments to the IFE.[19] Even so, the official numbers are almost certainly closer to the truth than are post-electoral polls.

The second pitfall of aggregated data is that they may lead to different conclusions than those based on individual-level measurements of the same variables. Inferring individual behavior from aggregated data has been deemed the "ecological fallacy", an examination of which is beyond what we can accomplish here.[20] However, we do note that the finer the unit of analysis (voting districts instead of states, say)—that is, the more observations—and the more the variables examined, the closer aggregated data will approximate individual data. Here, our most complex scheme uses five variables for each of Mexico's 32 *entidades federales* (31 states plus the Federal District).

We compare parameter estimates generated by sample weights (Models B1 to B3) to those produced by the unweighted sample (Baseline Model II, or BM-II). Since we are no longer

---

[18] In Mexico, a newly registered voter is put on a nationwide list of eligible voters (*el padrón*). When he picks up his credential, he is subsequently put on a list of names with photo ID's (*lista nominal*) used at the actual polling site.

[19] One source privy to the closed National Voter Registry meetings says the Registry acknowledges an inaccuracy rate of 15% or higher as of 2005.

[20] See Robinson (1950) for an influential early formulation of the slippery relationship between aggregated and individual data and King (1997) for ways to "reconstruct individual behavior from aggregate data".

attempting to replicate Buendía and Somuano, we modified the baseline model. It now comprises five sociodemographic variables typically associated with voting[21] and six political and economic attitudes.[22]

The first weighting scheme (Model B1) uses the marginal totals of the national turnout rate and the CSES sample-wide self-reported turnout rate to arrive at the probability of voting, Pr(V=1). In other words, we use the simple, unconditional probabilities of voting and of saying you voted, as shown in Tables 1 and 2. We derive our weights from the following identities:

$$\Pr(V=1) = \left(\frac{\Pr(V=1)}{\Pr(RV=1)}\right)\Pr(RV=1)$$

and

$$\Pr(V=0) = \left(\frac{\Pr(V=0)}{\Pr(RV=0)}\right)\Pr(RV=0)$$

Thus, observations in which (RV=1) are weighted by 0.78 $\left(\frac{\Pr(V=1)=.64}{\Pr(RV=1)=.82} \approx 0.78\right)$, and observations where (RV=0) are weighted by 2.01 $\left(\frac{1-\Pr(V=1)=.36}{1-\Pr(RV=1)=.18} \approx 2.01\right)$.

As we see from Table 4, the main effect of reweighting using national-level turnout estimates is to increase the significance of political evaluations. Two variables already significant at conventional levels, the perception that parties concern themselves with ordinary

[21] These are gender, age, education (categorical variable "linearized" using midpoint scoring, e.g., "secondary incomplete" = 7.5), income, and employment (dummy variable where working half-time or more = 1 and missing item responses are imputed the value of 0).

[22] The six attitudes are (1) Satisfaction with Democracy in Mexico (4-pt. scale), (2) 2000 Elections Were Clean (dummy with Yes = 1), (3) Elections Influence What Happens in Country (Yes = 1), (4) Political Parties are Concerned About People (*se preocupan de la gente*, Yes = 1), (5) Intensity of Party Identification (5-pt. scale), and (6) Evaluation of National Economy (3-pt. scale created from categories "Worsened", "Stayed the Same", and "Improved").

people and partisanship, vastly increase in significance. One variable that appeared meaningless in the baseline model, satisfaction with democracy, acquires substantive relevance in Model B1.

Otherwise, however, the simplest weighting scheme does little to change parameter or significance estimates. No variable changes sign or lies outside the 95% confidence intervals for BM-II. Model B1 paints with too broad a brush to be effective in correcting overreporting bias.

[Table 4 here]

We thus turn to a smaller unit of analysis, the state. Instead of the single, nationwide truth table represented in Tables 1 and 2, imagine a truth table for each state. Tables 5a to 5c present some examples:

[Tables 5a to 5c here]

The state-by-state marginals, coupled with our assumption of no overreporting, lead to the sample weights for Model B2.[23] Substituting Eq. 2 into Eq. 4 gives an expression for sample weights for Pr(V=1) and Pr(V=0), respectively.

$$\Pr(V=1\,|\,\mathbf{x}) = \left(\frac{\Pr(V=1\,|\,\mathbf{state})}{\Pr(RV=1\,|\,\mathbf{state})}\right)\Pr(RV=1\,|\,\mathbf{state})$$

and

$$\Pr(V=0\,|\,\mathbf{x}) = \left(\frac{1-\Pr(V=1\,|\,\mathbf{state})}{1-\Pr(RV=1\,|\,\mathbf{state})}\right)\Pr(RV=0\,|\,\mathbf{x})\,.$$

We get our estimates of Pr(RV=1) from disaggregating the full CSES sample (1,766, with no fewer than 45 per state) to the state level. Thus, voters from Aguascalientes who reported

[23] Note that absent the assumption of no overreporting, the expected cell probabilities correspond to the independence hypothesis, with marginals (unconditional probabilities of voting) fixed by the IFE and CSES estimates. If probabilities are conditioned upon independent variables, expected probabilities are conditionally independent—i.e., independence after controlling for explanatory factors.

voting, (RV=1), get a weight of .947, $\left( \frac{\Pr(V=1)=.667}{\Pr(RV=1)=.705} \approx 0.947 \right)$, and those who copped to not voting, (RV=0), are weighted by 1.129, $\left( \frac{1-\Pr(V=1)=.333}{1-\Pr(RV=1)=.295} \approx 1.129 \right)$.

Again, a combination of larger parameter magnitudes and smaller standard errors makes three variables already significant at conventional levels—parties' concern for people, party ID, and employment—even more significant. But model B2 does not change parameters appreciably, either: no signs change and no corrected parameters fall outside the 95% CI.

Finally, we test a far more refined weighting scheme, Model B3, in which Pr(V=1) is estimated by state according to gender, age, education, income, and employment. The state-level estimates are obtained via the "dose-probit" model described by Brehm (1999), essentially a probit model for grouped data.[24] The dependent variable is the turnout rate, expressed as the absolute number of voters divided by the number of names on state voting lists. Table 6 reports the results of the dose-probit model.

[Table 6 here]

We recover probabilities for each state from the dose-probit model and use them for our estimate of Pr(V=1) in the numerator of Model B3's sample weights. Whereas our estimate of Pr(V=1) derives from aggregated data, our estimate of the probability of reporting having, Pr(RV=1), is based on individual-level data. The individual predicted probabilities from BM-II are saved and inserted in the denominator of the sample weight:

$$\Pr(V=1 \mid \mathbf{x}) = \left( \frac{\Pr(V=1 \mid \mathbf{state}, \mathbf{gender}, \mathbf{age}, \mathbf{education}, \mathbf{income}, \boldsymbol{\Delta}\mathbf{GDP})}{\Pr(RV=1 \mid \mathbf{x})} \right) \Pr(RV=1 \mid \mathbf{x}).$$

[24] The terminology comes from pharmacological experiments in which different doses of some substance are administered to subjects. The rate of change from one substantive state to another (i.e., of recoveries, deaths, etc.) is disaggregated to some unit of analysis and the resulting observations are regressed on dosage to get the transition probabilities at different dose levels.

In contrast to the previous two models, the finer reweighting scheme tested in Model B3 effects significant changes in our parameter estimates. Most drastically, it nullifies the effects of several sociodemographic predictors of voting. The effects of age, education, and employment on turnout go from significance at $p < .05$ (and $p = .000$, in the case of age) to complete insignificance. The corrected estimates thus yield no significant differences in turnout rates between the old and the young citizens, the well- and the lesser-educated, and the employed and unemployed. It is also striking that in *no* model, neither the baseline nor the reweighted models, does turnout appear to increase with income.[25]

In contradistinction, attitudinal variables maintain their importance in determining turnout, consistent with results obtained by applying survey filters. Parties' concern for citizens and intensity of party ID went from significant at $p = .034$ and $p = .061$, respectively, to highly significant at $p = .001$ and $p = .016$. In brief, the most finely tuned weighting scheme deprives sociodemographic variables of significance, buts adds to that of political attitudes.

*Sample Selection Model with Heckman-Type Estimator*

The final corrective technique we essay is a Heckman-type sample selection model. This class of statistical models was designed to correct for "selection bias", which arises in the context of survey data when a sample over- or underrepresents certain classes of respondents. That is, the chances of inclusion in the sample differ systematically across members of the population. In formal terms, selection bias exists when the probability of observing a given case is a function of the value of the dependent variable (whereas in a completely random, unbiased sample each

[25] The lack of a presumed effect for income, however, may be due to the notorious difficulty of achieving reliable estimates of income. Since over 40% of Mexico is employed in the informal sector, incomes and employment status may vary widely from month to month. Thus, it may be difficult for a respondent to know with certainty how much she earned in any given month. In addition, those in the informal sector may be reluctant to report their true income.

member of the population has an equal probability of entering the sample and this probability is independent of values for the endogenous variable).

Underrepresentation of non-voters (and, consequently, overrepresentation of voters) in a sample can occur with either overreporting or "unit non-response", when an individual chosen in a sample fails to answer the survey because she is unreachable, refuses to answer the survey, or for any other reason. With unit non-response, or "stochastic truncation", non-voters may be underrepresented in the sample because proportionally more non-voters than voters refuse to answer a political opinion survey. Perhaps non-voters are embarrassed at not having voted, perhaps simply because they are not interested in politics. In the case of overreporting, non-voters are underrepresented because many of them dissemble their voting behavior to comply with social norms. Social desirability may lead to both unit non-response *and* overreporting. Heckman-type correction thus applies regardless of the mechanism responsible for turnout overestimation.

Heckman's formulated selectivity as a two-equation system (1976, 1979). The first equation, the *outcome* model, is a regression equation that represents the substantive phenomenon we wish to explain:

$$Y_{1i}^{*} = X_i \beta + \varepsilon_i \qquad \text{Eq. 5}$$

where, $Y_{1i}^{*}$ is an unobserved continuous latent variable, $X$ is an $n$ x $k$ matrix of values for $k$ explanatory variables and $n$ respondents, β a vector of parameters to be estimated and ε an error vector. [26]

[26] For an accessible introduction to sample selection models, see Winship and Mare (1992), whose presentation of the basic model we liberally adapt here.

In the second equation, the *selection* model, the dependent variable is binary: being selected into the sample or not. This leads most commonly to probit and logit models of the underlying probability of getting in the sample:

$$Y_{2i}^{*} = Z_i \alpha + v_i \qquad \text{Eq. 6}$$

$$Y_{2i} = 1 \text{ if } Y_{2i}^{*} > 0 \qquad \text{Eq. 7}$$

$$Y_{2i} = 0 \text{ if } Y_{2i}^{*} \leq 0 \qquad \text{Eq. 8}$$

where $Y_{2i}^{*}$ is the latent probability of selection; $Z_i$, a matrix of independent variables; $\alpha$, a parameter vector, υ, an error vector; and $Y_{2i}$, a binary realization of the underlying probability—i.e., whether the respondent was selected (or chose to respond) or not.

The two equations, Eq. 5 and Eq. 6 are related because when the dependent variable for the selection equation (Eq. 6) is above the threshold where the respondent is selected into the sample, the latent variable in the outcome equation $Y_{1i}^{*}$ (Eq. 5) is equal to its observed value, $Y_{1i}$. However, when a respondent is below the selection threshold and does not enter the sample, the latent variable $Y_{1i}^{*}$ (Eq. 5) is unobserved:

$$Y_{1i}^{*} = Y_{1i} \quad \text{if } Y_{2i}^{*} > 0 \qquad \text{Eq. 9}$$

and

$$Y_{1i}^{*} = Y_{1i} \quad \text{if } Y_{2i}^{*} > 0 \qquad \text{Eq. 10}$$

A problem arises when some of the same variables are in both Eqs. 5 and 6—when, say, income affects both the chances of voting and of responding to a pollster. Here, the disturbances of the outcome equation (ε in Eq. 5) depend on the disturbances of the selection equation (υ in Eq. 6). As a consequence, the independent variables in Eq. 5 are not independent of the error (in violation of the classical regression assumption) and the regressors of Eq. 5 are, in effect, subject to omitted variable bias.

Heckman proposed a two-stage correction technique that (1) estimates the selection model and (2) includes the expected errors from the selection model as a regressor in the outcome model. This process accounts for the bias produced by correlated errors in the simultaneous equation system. First, we rewrite Eq. 5 to decompose the error into a systematic component (the expected error of ε depending on υ, or $E(\varepsilon \mid \upsilon_i > -Z_i\alpha)$), and a random component , $\eta_i$ :

$$Y_{1i}^* = X_i\beta + E(\varepsilon \mid \upsilon_i > -Z_i\alpha) + \eta_i \qquad \text{Eq. 11}$$

Assuming that $\varepsilon_1$ and $\upsilon_i$ have a bivariate normal distribution, and standardizing the variance of $\upsilon_i$ to 1, $Var(\upsilon_i) = 1$, the expected value of the outcome model errors, ε, given the errors of the selection equation, υ , is the product of the selection equation errors and the covariance between the errors: $E(\varepsilon_1 \mid \upsilon_1) = \sigma_{\varepsilon\upsilon}\upsilon_1$. This implies:

$$E(\varepsilon_i \mid \upsilon_1 = -Z_1\alpha) = \sigma_{\varepsilon\upsilon}\frac{\phi(-Z_1\alpha)}{1-\Phi(-Z_1\alpha)} = \sigma_{\varepsilon\upsilon}\lambda(-Z_1\alpha) \qquad \text{Eq. 12}$$

where $\phi$ is the normal probability density function and $\Phi$ is the cumulative normal probability density function. The term $\lambda(-Z\alpha)$ is known as the inverse Mills' ratio (IMR).

The final step is inserting the IMR into the substantive model, achieved by substituting Eq. 12 into Eq. 11:

$$Y_{1i}^* = X_i\beta + \sigma_{\varepsilon\upsilon}\lambda(-Z\alpha) + \eta_i \qquad \text{Eq. 13}$$

Supplying the excluded variable corrects for the bias caused by its omission.

In our case, the outcome equation models the probability of *reporting* voting and the selection equation models the probability of voting. In a random sample, each citizen has an equal probability of being chosen to participate. Thus, if turnout is 64%, each case in a random sample will have a .64 probability of having cast a ballot. This implies that a model of the probability of voting is the functional equivalent of a selection model. Here, then, we use a Heckman-type technique to correct the self-reported voting parameters with a model for the probability of voting.

Normally, we would use individual-level data from a selection model to get the IMR. Since these do not exist, we take Brehm's suggestion (1999: 190-192) and use aggregated data to derive a "pseudo-IMR" from the selection/voting probability model. (Since this method involves a "pseudo-IMR" rather than a true one, we qualify ours a "Heckman-type", rather than "Heckman", approach.) Specifically, we estimate a dose-probit model with age, education, and employment as the predictors.

Table 7 shows the results of the Heckman-type corrections. Baseline Model III (BM-III) is a linear probability model of the propensity to vote, using the same variables and number of cases as did the logit models of the previous section.[27] Model C1 is BM-III corrected by the pseudo-IMR.

[Table 7 here]

Consistent with the changes effected by the previous two techniques, the Heckman-type corrector reduces greatly the significance of sociodemographic variables' effects. The effect of age plummets from a .04 increase in the probability of voting for every decade of life, highly significant at $p < .000$, to an effect statistically indistinguishable from zero ($p = .518$) after correction. Similarly, the influence of education on the decision to vote goes from highly significant ($p = .001$) to insignificant at conventional levels ($p = .111$). Though the coefficient

[27] A logit or probit model could also be estimated for the outcome equation, but OLS estimation is much more tractable technically. We note that in both of the models presented here, only 18 (of 862) observations predicted out-of-range, slightly over 1.00.

for employment remains virtually unchanged at .06 (that is, being employed at least part-time increases the probability of voting by 6%), it becomes less significant as the p-value increases from .042 to .06.

In contradistinction to sociodemographic conditions, political attitudes and evaluations remain important after accounting for turnout overestimation. An assessment of elections as clean continues to raise the probability of voting. Those who believe political parties are responsive to citizen concerns are still likelier to vote. Finally, those with more intense psychological affinities for one or another political party vote more frequently than those with weak partisan attachments, even after correcting for turnout overestimation. As is the case with applying survey filters and reweighting the sample, Heckman-type correction attenuates the effects of sociodemographic variables, but leaves intact those of psychological factors that predispose citizens to vote.

## V. Discussion: Are Previous Turnout Studies Wrong?

Two broad results, consistent across all three corrective techniques, stand out. Correcting for turnout overestimation does not appreciably diminish the effects of political beliefs and attitudes on turnout. However, in these models, it weakens or *eliminates* the effects of sociodemographic variables. It appears that older voters are no likelier to vote than younger ones; that the more educated are no likelier to vote than their less educated counterparts: and that those with jobs feel no more compunction to vote than those without them. Also contrary to conventional wisdom, greater income is not associated with a higher propensity to vote in either the corrected or uncorrected turnout models. These results contrast starkly with not only other turnout studies on Mexico, but also what studies on electoral participation in the U.S. and other established democracies might lead us to expect in the Mexican case.

On the other hand, political attitudes and beliefs are strong predictors of turnout across all models. In particular, those who believe that elections are generally clean and that parties are concerned with the needs of average citizens are likelier to vote than those who don't hold these beliefs. The intensity of one's identification with a given political party is also a strong predictor of electoral turnout. These factors' effects are consistent over corrected and uncorrected models, and in some cases intensify after correction.

This leaves several important questions. Do turnout studies generally exaggerate the effects of sociodemographic variables, as some of the NES studies suggest? Or is Mexico (specifically, perhaps, the 2000 election) an unusual case? If so, why? Or is it possible that sociodemographic variables still affect turnout indirectly, through political attitudes, in Mexico?

It may be that the 2000 Mexican elections were *sui generis*. They were foundational democratic elections, the first in which voters had a realistic chance of turning out the Revolutionary Institutional Party (PRI) after seven decades of one-party rule. The 2000 presidential election, thus, represented the culmination of the country's decade-long transition to electoral democracy, in which opposition victories at the local and state levels snowballed and led to 1997 congressional elections after which no single party held an outright majority and, ultimately, the first non-PRI president in 70 years. Given the singular importance of the 2000 elections, and a strong, pent-up longing for change, it is highly conceivable that the desire to vote was more or less equal across all social segments.

Another possible explanation for these surprising results is that one's social background does, indeed, continue to influence the decision to vote, but indirectly via political attitudes and beliefs. In this scenario, what one thinks about politics, parties, and politicians is itself determined by variables like education and employment status. Thus, attitudinal variables would absorb the effects of sociodemographic factors, which, insignificant in and of themselves, acquire relevance only through their impact on attitudes and beliefs.

Finally, studies of electoral participation might generally inflate the importance of income, education, age, and employment on voting. This conclusion is consistent with results from the United States, where Deufel and Kedar observe that, in the context of the NES studies, "[w]e can expecting overreporting to bias our results so that estimated relationships between voting and explanatory variables appear stronger than they are" (2000: 11; see also Silver *et al.*1986). It is also consonant with results from Latin America. In their multi-country, longitudinal study using aggregate data, Fornos *et al.* found that "what would appear to be the most intuitive socioeconomic variable, a country's aggregate level of wealth, has no discernible impact on electoral participation in Latin America." (Fornos *et al.* 2004: 931.) The authors conclude that features of the electoral system, such as compulsory voting and concurrent elections, are more determinative.

Of course, absent validated data in Mexico, we do not know the truth.[28] However, it is believable that the people traditionally considered most likely to vote—the well-educated, the well-off, and so on—are precisely those who are most embarrassed to admit not voting when they fail to do so. If this is true, then our estimates of the effects of income, wealth, and education on voting will be overstated. The results of correcting for turnout overestimation in Mexico, which obtain consistently over the three corrective techniques tested here, suggest that much of what we think we know about who votes stands in need of revision.

[28] Of course, sloppy record keeping and NES vote validation studies' disagreement over results indicate that validated voting data offer no fail-safe guaranties of truth even when they do exist.

## REFERENCES


Abramson, Paul R., and William Claggett, 1984. "Race-Related Differences in Self-Reported and Validated Turnout", *Journal of Politics*, Vol. 46, pp. 719-738.

Abramson, Paul R., and William Claggett, 1991. "Racial Differences in Self-Reported and Validated Turnout in the 1988 Presidential Elections", *Journal of Politics*, Vol. 53, No. 1 (February), pp. 186-197.

Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle, 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring", *Public Opinion Quarterly*, Vol. 63, No. 1 (Spring), pp. 90-108.

Brehm, John, 1999. "Alternative Corrections for Sample Truncation: Applications to the 1988, 1990, and 1992 Senate Election Studies", *Political Analysis*, Vol. 8, No. 2 (December), pp. 183-199.

Buendía Laredo, Jorge, 2002. "Determinantes de participación electoral", in *Deconstruyendo la ciudadanía. Avances y retos en el desarrollo de la cultura democrática en México* (Mexico City: Segob and Miguel Ángel Porrúa), pp. 439-454.

Buendía, Jorge, and Fernanda Somuano, 2003. "La participación electoral en la elección presidencial de 2000 en México", *Política y gobierno*, Vol. X, No. 2 (2nd Semester), pp. 289-324.

Deufel, Benjamin, and Orit Kedar, 2000. "But Do They Really Vote? Correcting for Overreporting of Turnout", paper prepared for presentation at the November, 2000, Political Psychology and Behavior Workshop.

Durand Ponte, Víctor Manuel, 2003. *Ciudadanía y cultural política. México 1993-2001* (Mexico City: Siglo XXI).

Fornos, Carolina, Timothy J. Power, and James C. Garand, 2004. "Explaining Voter Turnout in Latin America, 1980 to 2000", *Comparative Political Studies*, Vol. 37, No. 8 (October), pp. 909-940.

Heckman, James J., 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimation for Such Models", *Annals of Economic and Social Measurement*, Vol. 5, No. 4, pp. 475-492.

Heckman, James J., 1979, "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, No. 1, pp. 153-161.

Katz, Jonathan, 2000. "Correcting for Survey Misreports using Auxiliary Information", http://jkatz.caltech.edu/papers/misreport.pdf (originally presented at the 1998 American

Political Science Association and the 1999 Midwest Political Science Association Meetings).

King, Gary, 1997. *A Solution to the Ecological Inference Problem. Reconstructing Individual Behavior from Aggregate Data* (Princeton University Press).

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve, 1998. "Listwise Deletion is Evil: What to Do About Missing Data in Political Science", paper presented at the 1998 Annual Meeting of the Political Methodology Society, San Diego, CA.

Milbraith, Lester W., and M.L. Goel, 1977. *Political Participation: How and Why Do People Get Involved in Politics?* 2nd ed. (Chicago: Rand McNally).

Moreno, Alejandro, 2003. *El votante mexicano. Democracia, actitudes políticas y conducta electoral* (Mexico City: Fondo de Cultura Económica).

Presser, Stanley, and Michael Traugott, 1992. "Little White Lies and Social Science Models: Correlated Response Errors in a Panel Study of Voting", *Public Opinion Quarterly*, Vol. 56, No. 1 (Spring), pp. 77-86.

Presser, Stanley, Michael W. Traugott, and Santa Traugott, 1990. "Vote 'Over' Reporting in Surveys: the Records or the Respondent", paper prepared for the International Conference on Measurement Errors, Tucson, Arizona.

Raghunathan, Trivillore E., 2004. "What Do We Do with Missing Data? Some Options for Analysis of Incomplete Data", *Annual Review of Public Health*, Vol. 25, pp. 99-117.

Robinson, W.S., 1950. "Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, Vol. 15, No. (June), pp. 351-357.

Sigelman, Lee, 1982. "The Nonvoting Voter in Voting Research", *American Journal of Political Science*, Vol. 26, No. 1 (February), pp. 47-56.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson, 1986. "Who Overreports Voting?", *American Political Science Review*, Vol. 80, No. 2 (June), pp. 613-624).

Winship, Christopher, and Robert D. Mare, 1992. "Models for Sample Selection Bias", *Annual Review of Sociology*, Vol. 18, pp. 327-350.

Wolfinger, Raymond, and Steven J. Rosenstone, 1980. *Who Votes?* (New Haven: Yale University Press).

## APPENDIX A

Eq. 2

$$\Pr(V=1\,|\,\mathbf{x}) = \Pr(V=1\,|\,RV=1,\mathbf{x}) * \Pr(RV=1\,|\,\mathbf{x}) + \Pr(V=1\,|\,RV=0,\mathbf{x}) * \Pr(RV=0\,|\,\mathbf{x})$$

$$= \Pr(V=1\,|\,RV=1,\mathbf{x}) * \Pr(RV=1\,|\,\mathbf{x}) + 0$$

$$= \Pr(V=1\,|\,RV=1,\mathbf{x}) * \Pr(RV=1\,|\,\mathbf{x})$$

$$\Pr(V=1\,|\,RV=1,\mathbf{x}) = \frac{\Pr(V=1\,|\,\mathbf{x})}{\Pr(RV=1\,|\,\mathbf{x})}$$

The derivation of Eq. 2 takes as its starting point the so-called “accounting identities” from basic probability theory, which say that the (marginal) probability of voting is equal to the probability of voting conditional on having reported voting multiplied by the marginal probability of reporting voting, plus the probability of voting conditional on having reported not voting multiplied by the probability of reporting not voting. The second step makes use of our assumption of no underreporting (i.e., that Pr(V=1 | RV=0) = 0).

Eq. 3

$$\Pr(RV=1\,|\,\mathbf{x}) = \Pr(RV=1\,|\,V=1,\mathbf{x}) * \Pr(V=1\,|\,\mathbf{x}) + \Pr(RV=1\,|\,V=0,\mathbf{x}) * \Pr(V=0\,|\,\mathbf{x})$$

$$= \Pr(RV=1\,|\,V=1,\mathbf{x}) * \Pr(V=1\,|\,\mathbf{x}) + [1-\Pr(V=1\,|\,\mathbf{x})] * \Pr(RV=1\,|\,V=0,\mathbf{x})$$

$$= \Pr(RV=1\,|\,V=1,\mathbf{x}) * \Pr(V=1\,|\,\mathbf{x}) - \Pr(RV=1\,|\,V=0,\mathbf{x}) * \Pr(V=1\,|\,\mathbf{x}) + \Pr(RV=1\,|\,V=0,\mathbf{x})$$

$$= [\Pr(RV=1\,|\,V=1,\mathbf{x}) - \Pr(RV=1\,|\,V=0,\mathbf{x})] * \Pr(V=1\,|\,\mathbf{x}) + \Pr(RV=1\,|\,V=0,\mathbf{x})$$

$$= [1-\Pr(RV=0\,|\,V=1,\mathbf{x}) - \Pr(RV=1\,|\,V=0,\mathbf{x})] * \Pr(V=1\,|\,\mathbf{x}) + \Pr(RV=1\,|\,V=0,\mathbf{x})$$

Starting from the accounting identities, step two makes use of the complement of Pr(V=0), that is: Pr(V=0) = 1 – Pr(V=1). The fifth step invokes another complement, i.e., Pr(RV=1 | V=1) = 1 – Pr(RV=0 | V=1).

**Eq. 4**

$$\Pr(V=1\mid \mathbf{x}) = \Pr(V=1\mid RV=1,\mathbf{x}) * \Pr(RV=1\mid \mathbf{x}) + \Pr(V=1\mid RV=0,\mathbf{x}) * \Pr(RV=0\mid \mathbf{x})$$

$$= \Pr(V=1\mid RV=1,\mathbf{x}) * \Pr(RV=1\mid \mathbf{x}) + 0$$

$$= \Pr(V=1\mid RV=1,\mathbf{x}) * \Pr(RV=1\mid \mathbf{x})$$

Like Eq. 2, Eq. 4 also makes use of the assumption of no underreporting.

# APPENDIX B

**68.-** Do you have a voter ID?

Yes (1) No (2) → **72** DK (3) → **72** NR (4) → **72**

**69.-** Some people don't like the voter ID's while others think they're fine. We'd like to know your opinion about how the voter ID could be improved. Do you have your voter ID with you so you can show me what you like most about it?

| | | | |
|---|---|---|---|
| Doesn't have a voter ID | (1) → **72** | Claims ID is handy, but didn't show it | (4) |
| Yes, has ID with him/her | (2) | NR | (5) |
| No, doesn't have ID with him/her | (3) → **72** | | |

**70.-** What do you like most and what could be improved?

_______________________________________________________________

DK (99) NR (98)

**71.-** At the polling places in some states, voters get their ID's marked with the letter "v", while in other states they get a circle to indicate that you already voted. Which did you get, a "v" or a circle? Would you let me see what mark you got? **(INTERVIEWER, NOTE WELL IF THE ID IS MARKED AS VOTING IN THE 2000 FEDERAL ELECTIONS)**

| | |
|---|---|
| Yes, ID is marked for 2000 FEDERAL | (1) |
| No, ID is not marked for 2000 FEDERAL | (2) |
| Couldn't see | (3) |
| Other _________________________________ | |

DK (99) NR (98)

Figure 1: Linear Probability Models for Self-Reported and "Real" Vote Data When Overreporters' Income is Below the Sample Mean

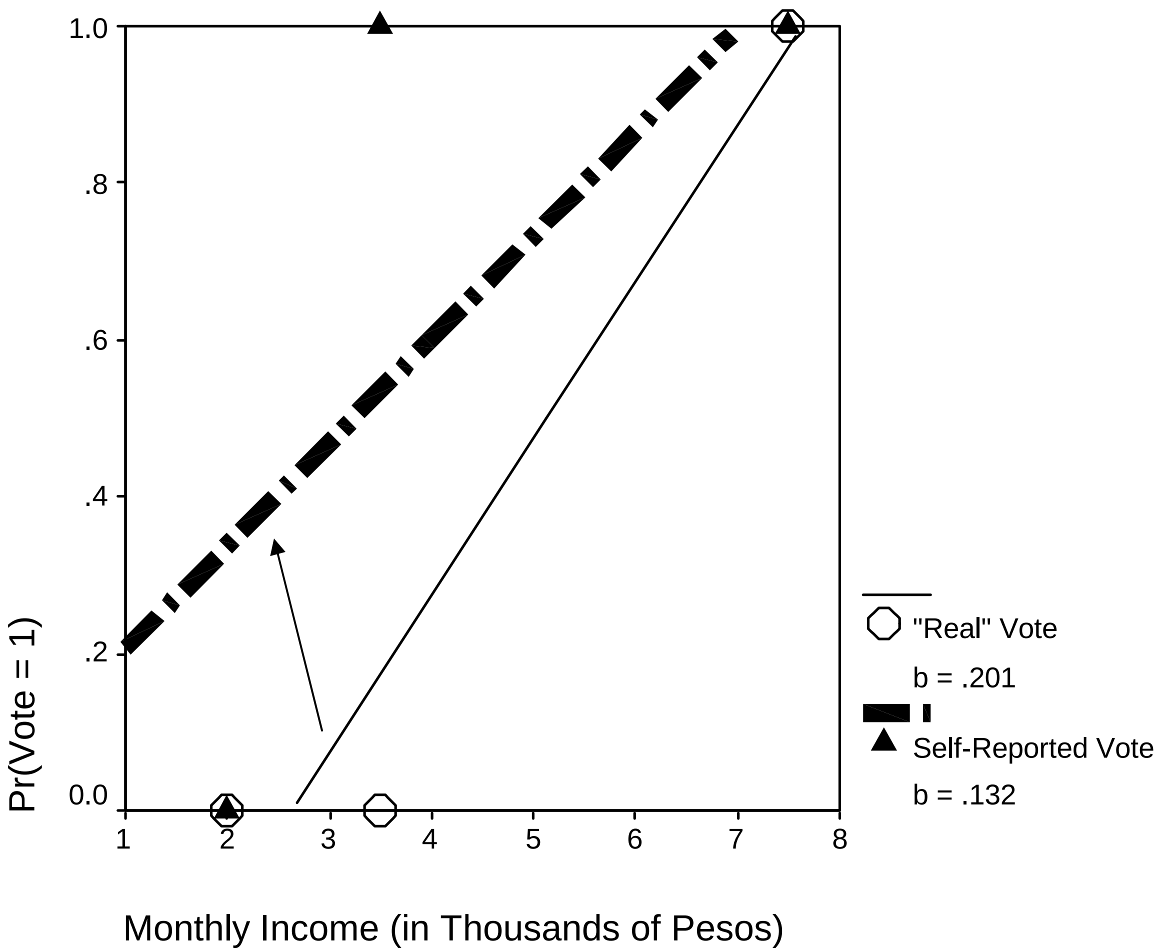

Figure 2: Linear Probability Models for Self-Reported and “Real” Vote Data When Overreporters’ Income is Equal to the Sample Mean

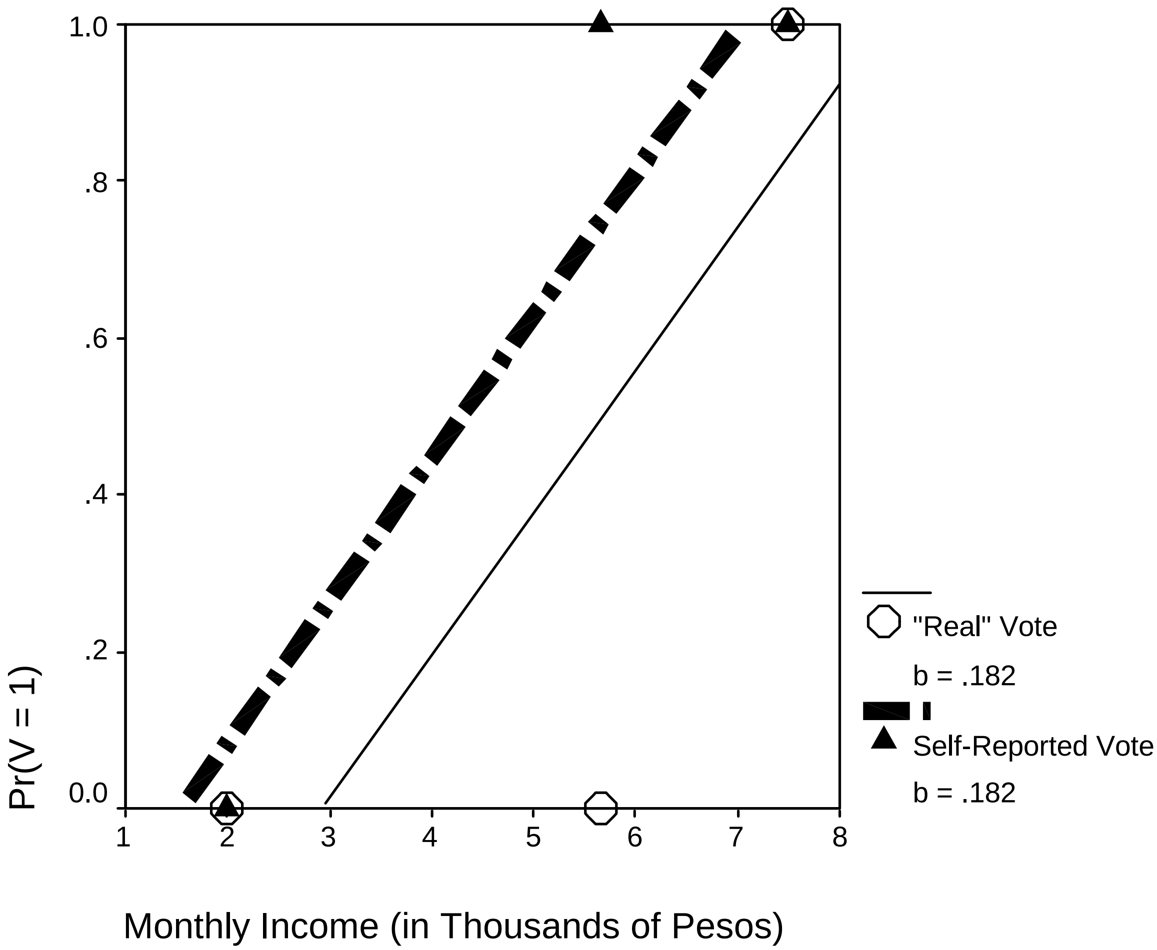

Table 1: Cross-Classification of Self-Reported Voting with Actual Voting Behavior, or "Truth Table" (Mexican 2000 Presidential Elections)

| | | **Voted** | | |
|---|---|---|---|---|
| | | Yes | No | |
| **Reported Voting** | Yes | $Pr(V=1 \mid RV=1) = ?$ | $Pr(V=0 \mid RV=1) = ?$ | 82% |
| | No | $Pr(V=1 \mid RV=0) = ?$ | $Pr(V=0 \mid RV=0) = ?$ | 18% |
| | | 64% | 36% | |

Column marginals taken from IFE data, row marginals taken from CSES poll.

Table 2: Cross-Classification of Self-Reported Voting with Actual Voting Behavior, Assuming No Underreporting (Mexican 2000 Presidential Elections)

| | | **Voted** | | |
|---|---|---|---|---|
| | | Yes | No | |
| **Reported Voting** | Yes | .64 | .18 | 82% |
| | No | .00 | .18 | 18% |
| | | 64% | 36% | |

Column marginals taken from IFE data, row marginals taken from CSES poll.

Table 3: Logit Regressions for Turnout Models Using Survey "Filters", Mexican 2000 Presidential Elections

| | **Buendía & Somuano** | | **Baseline Model I** | | | 95% CI | | **Model A1** | | | **Model A2** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent (Vot = 1) | **β** | p | **β** | se | p | Lower | Upper | **β** | se | p | **β** | se | p |
| **Constant** | **-1.033** | .004 | **.079** | .238 | .740 | -.388 | .546 | **.052** | .208 | .804 | **-.500** | .194 | .010*** |
| **Political Attitudes** | | | | | | | | | | | | | |
| Parties' and Pols' Responsiveness | **.219** | .003*** | **.447** | .087 | .000*** | .275 | .617 | **.334** | .073 | .000*** | **.202** | .066 | .002*** |
| Perceptions of Democracy and Elections | **.219** | .002*** | **.125** | .086 | .144 | -.043 | .295 | **.120** | .072 | .093* | **.086** | .066 | .195 |
| Unconventional Political Participation | **-.093** | .012*** | **-.836** | .552 | .130 | -1.919 | .247 | **-.724** | .457 | .113 | **-1.317** | .418 | .002*** |
| **Group Identity** | | | | | | | | | | | | | |
| PAN | **.085** | .642 | **-.009** | .207 | .966 | -.415 | .398 | **.136** | .176 | .439 | **.204** | .158 | .196 |
| PRI | **.408** | .040** | **.204** | .235 | .384 | -.256 | .664 | **.362** | .198 | .067* | **.564** | .180 | .002*** |
| PRD | **.771** | .008*** | **.221** | .115 | .055* | -.005 | .447 | **.098** | .087 | .262 | **.181** | .081 | .025** |
| Other | **-.009** | .983 | **.373** | .579 | .519 | -.762 | 1.509 | **-.315** | .423 | .456 | **.741** | .436 | .089* |
| Social Group Membership | **.102** | .051** | **.121** | .096 | .206 | -.067 | .309 | **.054** | .074 | .462 | **.073** | .066 | .275 |
| **Campaign Effects** | | | | | | | | | | | | | |
| Home Visits | **-.006** | .972 | **.117** | .345 | .557 | -.274 | .508 | **.191** | .166 | .250 | **-.059** | .148 | .692 |
| Mailer | **.418** | .007*** | **.464** | .185 | .012** | .101 | .827 | **.294** | .153 | .054* | **.247** | .139 | .075* |
| Gift | **-.520** | .015*** | **-.625** | .253 | .014** | -1.120 | -.129 | **-.661** | .213 | .002*** | **-.447** | .198 | .024** |
| **Individual Resources** | | | | | | | | | | | | | |
| Political Sophistication | **.150** | .000*** | **.298** | .063 | .000*** | .175 | .420 | **.246** | .051 | .000*** | **.162** | .046 | .001*** |
| 26-40 years old | **.019** | .918 | **-.070** | .219 | .736 | -.481 | .340 | **-.260** | .187 | .165 | **.018** | .169 | .916 |
| 41-60 years old | **.392** | .066* | **.603** | .276 | .029** | .061 | 1.145 | **.016** | .389 | .944 | **.302** | .201 | .132 |
| Over 60 years old | **1.235** | .002*** | **.911** | .477 | .057* | -.025 | 1.846 | **.389** | .357 | .276 | **.442** | .312 | .156 |
| Married | **.334** | .035** | **.741** | .198 | .000*** | .354 | 1.130 | **.446** | .171 | .009*** | **.560** | .156 | .000*** |
| **N** | 1,502 | | 1,205 | | | | | 1,205 | | | 1,205 | | |
| Pseudo R2 (Cox & Snell)† | .074 | | .096 | | | | | .075 | | | .073 | | |
| -2LogL | 1,402.2 | | 931.3 | | | | | 1,244.9 | | | 1,437.4 | | |
| Deviance | | | 923.0 | | | | | 1,236.6 | | | 1,426.3 | | |

† Buendía and Somuano do not specify which pseudo-$R^2$ measure they report.
*  $p \leq .10$
** $p \leq .05$
*** $p \leq .01$

Table 4: Logit Regressions for Turnout Models Using
Sample Weights, Mexican 2000 Presidential Elections

| | Baseline Model II | | | | | Model B1 | | | Model B2 | | | Model B3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 95% CI | | | | | | | | | | |
| Dependent (Vot = 1) | **β** | se | p | Upper | Lower | **β** | se | P | **β** | se | p | **β** | se | p |
| **Constant** | **-2.135** | .576 | .000*** | | | **-3.076** | .483 | .000*** | **-2.846** | .487 | .000*** | **-1.486** | .473 | .002*** |
| **Sociodemographic** | | | | | | | | | | | | | | |
| Sex (M=1) | **-.193** | .2114 | .3611 | -.608 | .221 | **-.165** | .170 | .331 | **-.116** | .171 | .494 | **.054** | .164 | .743 |
| Age | **.032** | .008 | .000*** | .015 | .048 | **.032** | .007 | .000*** | **.032** | .007 | .000*** | **.002** | .006 | .749 |
| Education | **.056** | .028 | .042** | .002 | .110 | **.062** | .023 | .069* | **.054** | .023 | .020** | **.016** | .022 | .484 |
| Income | **.000** | .000 | .584 | .000 | .000 | **.000** | .000 | .465 | **.000** | .000 | .307 | **.000** | .000 | .583 |
| Employed (Works at least part-time = 1) | **.424** | .214 | .025** | .004 | .844 | **.420** | .173 | .015** | **.477** | .174 | .003*** | **-.027** | .170 | .876 |
| **Political and Economic Attitudes** | | | | | | | | | | | | | | |
| Satisfaction w/ Democracy in Mexico (4-pt.) | **.159** | .100 | .114 | -.038 | .355 | **.148** | .081 | .068* | **.066** | .083 | .427 | **.114** | .080 | .156 |
| 2000 Elections were Clean (Yes=1) | **.181** | .068 | .008*** | .047 | .320 | **.176** | .057 | .002*** | **.174** | .058 | .003*** | **.198** | .056 | .000*** |
| Elections Influence what Happens in Country (Yes=1) | **.042** | .073 | .567 | -.101 | .184 | **.032** | .060 | .588 | **.030** | .061 | .618 | **.024** | .060 | .693 |
| Political Parties are Concerned About People (Yes=1) | **.157** | .074 | .034** | .012 | .307 | **.159** | .059 | .007*** | **.162** | .060 | .007*** | **.191** | .059 | .001*** |
| Intensity of Party ID (5-pt.) | **.152** | .081 | .061* | -.007 | .184 | **.153** | .065 | .019** | **.172** | .067 | .010*** | **.155** | .064 | .016** |
| Evaluation of National Economy in Last Yr. (3-pt.) | **.007** | .068 | .910 | -.125 | .140 | **.007** | .056 | .894 | **.031** | .056 | .576 | **.018** | .056 | .740 |
| | | | | | | | | | | | | | | |
| **N** | 862 | | | | | 862 | | | 862 | | | 862 | | |
| Pseudo R2 (Cox & Snell) | .065 | | | | | .065 | | | .093 | | | .063 | | |
| -2LogL | 751.3 | | | | | 751.2 | | | 1,013.8 | | | 1,056.8 | | |
| Deviance | 855.8 | | | | | 855.7 | | | 842.3 | | | 862.5 | | |

*  p ≤ .10
** p ≤ .05
*** p ≤ .01

Tables 5a-c: Cross-Classifications of Self-Reported Voting with Actual Voting Behavior By State, Assuming No Underreporting (Mexican 2000 Presidential Elections)

## Aguascalientes

| Reported Voting | Voted: Yes | Voted: No | |
|---|---|---|---|
| Yes | .667 | .038 | 70.5% |
| No | .00 | .295 | 29.5% |
| | 66.7% | 33.3% | |

## Baja California Sur

| Reported Voting | Voted: Yes | Voted: No | |
|---|---|---|---|
| Yes | .672 | .106 | 77.8% |
| No | .00 | .122 | 22.2% |
| | 67.2% | 22.8% | |

•
•
•

## Yucatán

| Reported Voting | Voted: Yes | Voted: No | |
|---|---|---|---|
| Yes | .720 | .080 | 80.0% |
| No | .00 | .100 | 20.0% |
| | 72.0% | 18.0% | |

Note: the CSES sample of 1,766 did not include all states, but adhered to a minimum of 45 or so for the states it did include.

Table 6: Dose-Probit Model for Turnout, with State as Sampling Unit

| **DV = voters / registered voters (by state)** | **$\beta$** | **se** |
|---|---|---|
| Intercept | 3.739 | .014 |
| Sex | -9.011 | .029 |
| Age | .011 | .000 |
| Education | .005 | .000 |
| Employment | 1.007 | .007 |
| $\Delta$GDP 1999-2000 | .567 | .010 |

Table 7: Linear Probability Models for Turnout with Heckman-Type Correction, Mexican 2000 Presidential Elections

| | **Baseline Model III** | | | | | **Model C1** | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 95% CI | | | | |
| Dependent (Vot = 1) | **β** | se | p | Upper | Lower | **β** | se | p |
| **Constant** | **.287** | .080 | .000 | | | **.342** | .370 | .356 |
| **Sociodemographic** | | | | | | | | |
| Sex (M=1) | **-.026** | .029 | .372 | -.082 | .031 | **-.025** | .029 | .383 |
| Age | **.004** | .001 | .000*** | .002 | .006 | **.003** | .005 | .518 |
| Education | **.008** | .004 | .001*** | .000 | .015 | **.007** | .005 | .111 |
| Income | **.000** | .000 | .658 | .000 | .000 | **.000** | .000 | .658 |
| Employed (Works at least part-time = 1) | **.060** | .029 | .042** | .002 | .118 | **.059** | .031 | .060* |
| **Political and Economic Attitudes** | | | | | | | | |
| Satisfaction w/ Democracy in Mexico (4-pt.) | **.022** | .014 | .122 | -.006 | .049 | **.022** | .014 | .122 |
| 2000 Elections were Clean (Yes=1) | **.030** | .010 | .004*** | .010 | .051 | **.030** | .010 | .004*** |
| Elections Influence what Happens in Country (Yes=1) | **.007** | .011 | .497 | -.014 | .029 | **.007** | .011 | .494 |
| Political Parties are Concerned About People (Yes=1) | **.021** | .010 | .036** | .001 | .040 | **.021** | .010 | .035** |
| Intensity of Party ID (5-pt.) | **.021** | .010 | .051** | .000 | .042 | **.021** | .011 | .052** |
| Evaluation of National Economy in Last Yr. (3-pt.) | **.000** | .009 | .984 | -.018 | .019 | **.000** | .009 | .984 |
| **N** | 862 | | | | | 862 | | |
| Adj. R2 | .054 | | | | | .053 | | |

* p ≤ .10
** p ≤ .05
p ≤ .01