ONLINE APPENDIX

The Quality of Vote Tallies: Causes and Consequences

Cristian Challú, Enrique Seira, and Alberto Simpser[1]

May 2020

[1]Carnegie Mellon (cchallu@andrew.cmu.edu), ITAM (enrique.seira@itam.mx), and ITAM (alberto.simpser@itam.mx).

HAND COUNTING IN OTHER COUNTRIES

FIGURE A1. List of counties with hand counting

Votes counted by hand			No mention of hand counting
Albania	Germany	Romania	Armenia
Algeria	Guadeloupe	Saint Lucia	Brazil
American Samoa	Guyana	Saint Vincent / Grenadines	Costa Rica
Argentina	Hungary	Samoa	Cote d'Ivoire
Australia	Iceland	Seychelles	French Polynesia
Austria	India	Sierra Leone	Ghana
Bahamas	Iran	Slovakia	Israel
Bahrain	Ireland	Slovenia	Jordan
Bangladesh	Italy	Somalia	Kenya
Barbados	Japan	Spain	South Korea
Belgium	Kiribati	Sri Lanka	Lithuania
Belize	Macedonia	Sudan	Luxembourg
Benin	Malaysia	Sweden	Netherlands
Bolivia	Marshall Islands	Switzerland	New Caledonia
Bosnia and Herzegovina	Martinique	Timor-Leste	Niger
Bulgaria	Mexico	Tonga	Philippines
Cameroon	Micronesia	Trinidad and Tobago	Russia
Canada	Monaco	Turkey	Tunisia
Cape Verde	Morocco	Tuvalu	Venezuela
Chile	Nepal	Ukraine	Wallis and Futuna
Colombia	New Zealand	United Kingdom	
Djibouti	Nigeria	Uruguay	
Dominica	Niue	Vanuatu	
Ecuador	Norway	Yemen	
Egypt	Pakistan	Zanzibar	
El Salvador	Palestine		
Fiji	Papua New Guinea		
Finland	Paraguay		
France	Portugal		
French Guiana	Reunion		

Note: Source: ACE Project (aceproject.org, accessed April 10 2018) and Mexican electoral law. Out of 105 countries with information on how votes are counted, in 85 (or 81%) votes are counted by hand. Hand counting was not mentioned in the ACE Project information for the rest of the countries with information on how votes are counted (based on a keyword search). This table shows the names of the countries in both subsets.

RELATED LITERATURE

A large fraction of existing research on the quality of vote tallies studies the relative performance of different voting and tallying technologies (Allers and Kooreman 2009; Alvarez and Hall 2008; Alvarez et al. 2009; Ansolabehere and Reeves 2004; Lott 2009; Dee 2007; Garner and Spolaore 2005;

Mebane 2004). We take technology (hand counting) as given, and instead focus on sociodemographic, workload, and cognitive causes of variation in tally quality. A few papers investigate sociodemographic correlates of voting technologies (Garner and Spolaore 2005; Lott 2009) and poll worker performance (Atkeson et al. 2014), but not of tally quality. While some literature has studied voter fatigue (e.g. (Augenblick and Nicholson 2015)) we know of no causal estimates of the effect of PW fatigue on tally quality. We know of no research on the downstream electoral/political consequences of variation in tally quality, an issue we investigate.

The election forensics approach, meanwhile, tests for statistical fingerprints of fraud in official voting results (Beber and Scacco 2012; Cantú 2014; Kobak, Shpilkin, and Pshenichnikov 2016; Mebane 2010; Myagkov, Ordeshook, and Shakin 2010). The fingerprints of fraud often correlate strongly with patterns of partisan advantage/disadvantage. Myagkov, Ordeshook, and Shakin 2010, for example, find that precinct-level turnout correlates with percent vote for the incumbent party, but does not correlate with votes for opposition parties. The inconsistencies we study, in contrast, bear no association to partisan performance and are not properly viewed as fingerprints of partisan malfeasance.

Research on trust in electoral results/institutions studies its relationship with voting technology (Alvarez et al. 2009), electoral fraud (Wellman, Hyde, and Hall 2018), individual traits (Alvarez et al. 2004), or the quality of service provided by poll workers (Claassen et al. 2008; Hall, Quin Monson, and Patterson 2009) but not with the quality of tallies. Finally, our study joins a few recent data-intensive studies of Mexican elections, including Cantú (2018) on the fraudulent 1988 Mexican elections, Larreguy, Marshall, and Querubín (2016) on partisan monitoring of vote-buying brokers, and Cantú and Ley (2017) on determinants of citizen participation as poll workers. None of these examine the quality of vote tallies.

CONTEXT: THE COUNTING OF VOTES IN MEXICAN ELECTIONS

Mexico experienced electoral authoritarian government for most of the 20th century. After a series of crises, in the 1990s the major political parties negotiated a set of profound reforms to the electoral system that turned Mexico's regime into a democracy. The reformed system was designed to render partisan manipulation of elections very difficult. Its features included a transparent and reliable list of registered voters, a highly regulated process to select citizens to function as poll workers responsible for counting votes, a procedure to aggregate voting results quickly after polls close, a system of public financing and campaign spending rules that govern electoral campaigns, and an independent electoral tribunal of last resort to resolve electoral controversies. Perhaps chief among the reforms was the creation of an independent bureaucracy charged with organizing elections and producing official electoral results—the *Instituto Federal Electoral* (now called the *Instituto Nacional Electoral* or INE). Previously, all aspects of elections had been under the direct control of the executive branch of government.

Precincts (*Secciones*) and polling stations

The basic unit of Mexico's electoral geography is the *sección electoral* (subsequently precinct). Every precinct contains one or more polling stations (henceforth PS), depending on the number of voters registered in the precinct. To provide a sense for the magnitudes: there were 62,692 precincts and 129,238 PSs in the 2012 presidential election. The average precinct covers about 1,200 registered voters. A strictly-enforced maximum of 750 registered voters can be assigned to vote at any given PS. This maximum determines the total number of polling stations needed in an election. Registered citizens are apportioned equally across the PSs in a precinct. For example, in a precinct containing 752 citizens, two PSs will exist, with 376 citizens assigned to each. The first PS in a precinct is known as the *básica* (basic) PS, the second PS is called *contigua 1* (contiguous 1), the third is *contigua 2*, and so on. Within a given precinct, polling stations are expected to have about the same number of total votes and vote shares across parties because voters are assigned to them in a quasi-random way.

Assignment of citizen poll workers to polling stations

Mexico is one of the largest democracies in the world.Mexican law requires that votes be tallied by hand by randomly-selected citizens who function as polling-station workers (henceforth PW). This is a challenging logistical feat. Each PS is allocated 4 acting PWs and 3 substitute PW.[2] INE recruits PWs from the same area where the corresponding PS is located following a very detailed, transparent, and rigorous invitation procedure enshrined in the law. This procedure involves inviting a random set of 13% of registered voters in every precinct to function as PWs. To achieve this, the INE hires a large team of professional recruiters to visit citizens at their home and assess their eligibility to function as PWs. To be eligible, a citizen must not work for a political party and must be able to read and write, among other things. A second lottery then selects a randomly-chosen subset of the eligible citizens in a precinct, and these are designated to staff each of the PS in that precinct.

The assignment of designated citizens to PS within a precinct proceeds according to educational attainment. The citizen with the highest educational attainment is designated President of the first PS. The one with the next highest educational attainment is designated President of the second PS. Once every PS in the precinct has a President, the person with the next highest educational attainment is designated Secretary of the first PS. In the same manner, citizens are next designated to the positions of First Counter (*primer escrutador*) and Second Counter (*segundo escrutador*). The remaining citizens are designated as first, second, and third substitutes (*suplentes*). This assignment rule implies that the average educational attainment of poll workers is generally higher in the *básica* PS as in the *contigua 1* PS in the same precinct, a fact that we exploit to identify the causal effect of education on inconsistencies.

The general functions of the PW team for a given PS are to staff the PS during Election Day, to make sure only those eligible to vote at the PS do so, to count the votes by hand after the close of voting, and to fill out the *acta* that same evening. PWs are trained by thousands of INE employees (*capacitadores-asistentes electorales* or CAE) hired for that purpose.

[2]Substitutes are also trained in case one of the four acting PWs drops out or fails to show up on Election Day. When local elections are concurrent with national ones, the number of PWs allocated to a PS increases to 6 acting and 3 substitutes.

Political party representatives

Political parties are entitled to send representatives to sit at the PSs along with PWs. These representatives are registered with INE by the political parties prior to the election. In general they are registered for a specific PS. They can observe the work of the PW, but they have no formal role in the ballot counting or in the filling out of the *actas*. There were 600,743, 743,263, and 846,336 party representatives in the elections of 2009, 2012 and 2015, respectively.

DATA

We provide here additional details on the data used in the analysis.

Data on PW characteristics

For each PW we observe age, gender, and years of education completed. We use data on those PWs who attended their polling station on Election Day.[3] Part C of Table A1 displays polling-station averages. On average, PWs are in their late thirties, close to 42 percent are male, and completed about 11 to 12 years of education—that is, almost completed high school. Variation across PSs is significant, with standard deviations of 7 years of age, 2.7 years of education, and 25 percentage points (pp) in male fraction.[4]

Data on political party representatives

For elections in the years 2012 and 2015, we observe which party sent representatives to which PS. The main political parties—PAN, PRI and PRD—respectively sent monitors to 73%, 93%, and 55% of

[3]Some fraction of PWs do not show up at their PS on Election Day.

[4]We also have data on PW training. There is about one CAE for every 6 polling stations, on average. We observe which citizen PW was trained by which CAE. We also observe CAE characteristics including education level, score in the interview and job exams, and experience as CAE in previous elections. We use these data as controls in regressions.

PSs in 2012. The equivalent figures for 2015 are 81%, 94%, and 63%. Party representatives have to be registered at INE at prior to the election, and their identities are verified by PWs on Election Day. Party representatives can vote in a PS even if they are not on the voter list for that PS or precinct.

Data on recounts

Our data indicate whether the votes in a given PS in a particular election were recounted, for all elections in 2009, 2012 and 2015. There were 34,795, 198,007, and 77,113 *actas* recounted in 2009, 2012 and 2015 respectively. These amount to 27.6%, 51.1%, and 62.5% of the total number of *actas* in the corresponding election years.

Data on voting results

Aggregate vote results are public information published by INE in the SICEEF (Federal Elections Statistical System of Queries). For each PS we observe the number of votes received by each party for the lower house of congress in 2009, 2012, and 2015, and also for president and the upper house of congress in 2012. From other data sources we have information on the number of registered voters as well as on the PS and precinct to which each registered voter was assigned to vote. We were able to verify that the rule that no PS can have more than 750 registered voters allocated to it is strictly followed.

Survey of poll workers

The survey was implemented in the 2017 local elections in Coahuila, Estado de Mexico, Nayarit and Veracruz (these states were the only ones to hold elections in that year.) The survey was targeted to a random sample of about 80,000 PWs and implemented by the CAEs during the second stage of the PW recruitment process. Randomization was conducted at the level of ARE (an ARE is a set of contiguous precincts assigned to a CAE). Of the 6,690 CAE, 4,014 were selected to implement the survey to all the PWs in the precincts under their charge. We collected 85,006 surveys in 7,161 different precincts, of which roughly 60,000 are usable (some of the surveys lacked the information needed in order for us to merge them with the rest of the data). Figure A14 in shows the precincts in which at least one PW completed a survey. In some of the analysis, we use 12 covariates obtained from these surveys as control variables. These include proxies of wealth, prosocial behavior, political activity, and satisfaction with democracy. Further details are provided in the notes to Table 4 in the body of the paper.

Socioeconomic data

The national statistical agency (INEGI) and the electoral authority (INE) provide a publicly-accessible version of the 2010 Population Census for which data are aggregated at the electoral precinct level. These data cover 66,740 precincts. The basic set of control variables we use in the analysis consists of 15 variables from this data source. The specific variables are listed in the note to Table 4 in the body of the paper.

Summary statistics

Table A1 displays summary statistics. Part A describes the number of precincts, polling stations, poll workers, registered voters, and votes cast for each of the elections of 2009, 2012 and 2015. It provides a glimpse of the breadth of the data we use. For each election our data cover approximately to 60,000 precincts, about 125,000 PS, half a million PW, more than 70 million registered voters, and between 31 million and 45 million votes cast, depending on the election. About 40,000 precincts contained more than one PS.

In the analysis, we do not use the full universe of PS. Of the 429,765 PS installed in the three electoral years, we drop 38,493 PS located in precincts with at least one *extraordinaria* PS. This category of polling station is quite different from the rest (e.g. located in remote places, places with operational problems, and related issues). To retain comparability, therefore, we restrict our data to exclude these. Additionally, we drop 12,725 PS located in precincts with at least one PS did not merge with one or more of the datasets we use in the analysis (described above in the Appendix as well as in the Data section in the body of the paper).

TABLE A1. Summary Statistics by Election

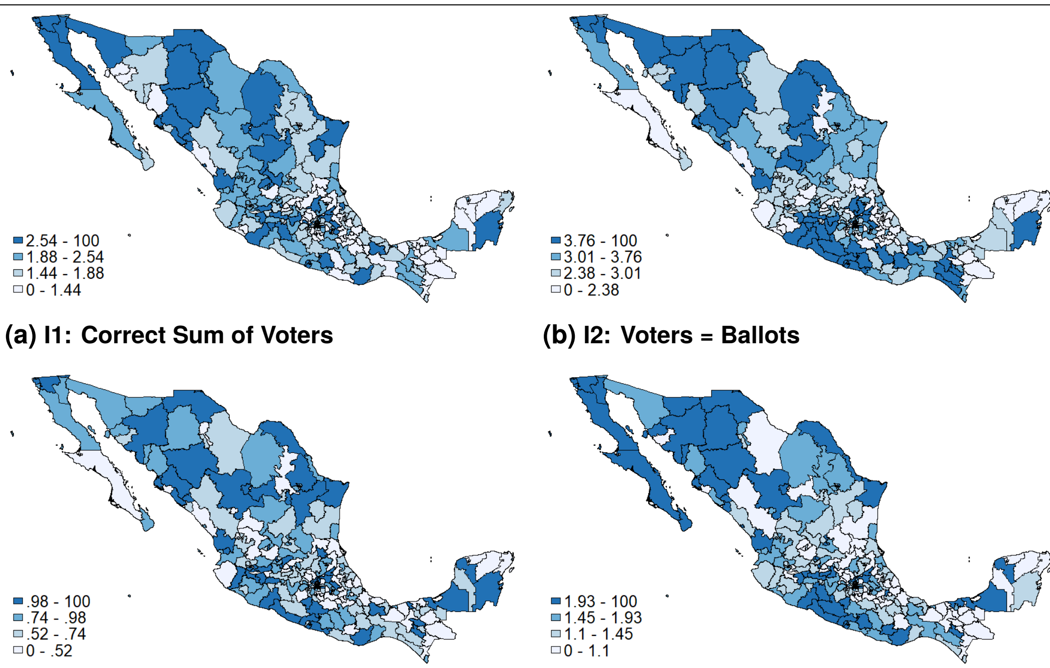
	2009		2012						2015	
	Congressional		Congressional		Presidential		Senatorial		Congressional	
Part A: Election-level variables										
Num. precincts	61,089				62,692				58,797	
Num. PS	126,198				129,238				123,319	
Num. registered voters	71,319,536				72,925,360				70,111,928	
Num. precincts with contiguous PS	42,136				41,523				39,184	
Num. PW	499,489				514,742				623,714	
Num. votes cast	31,671,852		45,584,376		45,577,568		45,572,336		33,368,108	
Turnout (%)	44.41		62.51		62.50		62.5		47.59	
Part B: Inconsistencies in actas										
Num. I1: Correct Sum of Votes	13.4	(64.9)	7.2	(44.9)	8.1	(47.8)	7.3	(45.3)	6.9	(45.9)
Num. I2: Voters = Ballots	17.6	(74)	8.5	(44.6)	10.2	(49.2)	8.6	(44.6)	10.4	(54.4)
Num. I3: Votes Cast = Ballots	2.4	(24.6)	2.8	(25.7)	2.9	(26.1)	3.0	(27.1)	2.7	(26.5)
Num. I4: Ballot Balance	5.4	(34.8)	7.3	(37.5)	7.2	(38.5)	7.1	(37)	NA	NA
% of PS with incons. 1	12.5		7.7		9.0		7.7		8.0	
% of PS with incons. 2	24.9		32.0		32.7		32.0		27.6	
% of PS with incons. 3	4.4		7.5		8.1		7.7		6.9	
% of PS with incons. 4	29.1		38.2		38.3		38.0		ND	
% of PS with at least one inconsist.	41.6		43.6		44.5		43.2		30.5	
Part C: Poll worker traits										
Age	34.8	(7)			36.0	(7.1)			37.2	(7.1)
% male	42.7	(25.9)			42.4	(25.3)			41.0	(22.9)
Years of education	11.6	(2.7)			11.9	(2.5)			11.1	(2.7)

Note: Figures in parentheses are standard deviations. See text for further details. Part C describes traits of PW. For these, we compute the average value for all PWs within a PS and then average across PSs.

ADDITIONAL DETAILS ON TALLY INCONSISTENCIES DATA

Figure A2 displays the geographical distribution of inconsistencies across electoral districts. Each degree of shading represents a quartile, with darker shading indicating a greater extent of inconsistencies. The main thing to note is that there are inconsistencies in almost every area of Mexico. There seems to be some geographical clustering, suggesting possibly common determinants.

FIGURE A2. Geographical Distribution of Inconsistencies



(a) I1: Correct Sum of Voters **(b) I2: Voters = Ballots**

(c) I3: Votes Cast = Ballots **(d) I4: Ballot Balance**

Note: The maps display the geographical distribution of inconsistencies by electoral district, pooling together all election years and election types in our data. Specifically, for each inconsistency of type $j = 1, .., 4$, the figure plots the ratio of the absolute number of type-j inconsistencies ($AbsNumInc^j$) in the district over the total number of votes cast. Shading represents quartiles.

Overall, the different measures of inconsistencies are not strongly correlated (Table A3). Inconsistency types 1 and 2 are weakly correlated with types 3 and 4. The stronger correlation (.75) between types 1 and 2 is likely due to the fact that they share the term SV (total number of votes cast)[5], which is the result of a sum computed by the PWs and therefore prone to mistakes.

Now we present correlations between the different measures of tally inconsistencies that we use.

[5]Some types of inconsistencies share inputs, which induces correlation between them.

We define the presence of inconsistences as a dummy variable indicating the presence of a given inconsistency type. We define the extent of inconsistencies as the size, in absolute value, of the discrepancy between the two sides of the corresponding equality (the equalities are described in the Data section in the body of the paper).

TABLE A2. Presence of inconsistencies correlations

	Inconsistency 1	Inconsistency 2	Inconsistency 3	Inconsistency 4
Inconsistency 1	1.00			
Inconsistency 2	0.11	1.00		
Inconsistency 3	0.06	0.17	1.00	
Inconsistency 4	0.05	0.49	0.23	1.00

Note: The numbers are pairwise correlations across the four PS-level measures of *presence* of tally inconsistencies. Presence of inconsistency is a dummy variable indicating that the extent of tally inconsistencies of type $j = 1, .., 4$ (defined in the Data section in the body of the paper) was greater than zero.

TABLE A3. Extent of inconsistencies correlations

	Inconsistency 1	Inconsistency 2	Inconsistency 3	Inconsistency 4
Inconsistency 1	1.00			
Inconsistency 2	0.75	1.00		
Inconsistency 3	0.07	0.22	1.00	
Inconsistency 4	0.07	0.17	0.39	1.00

Note: The numbers are pairwise correlations across the four PS-level measures of *extent* (i.e., the absolute value of the discrepancy between the sides of the corresponding equality, in number of votes) of tally inconsistencies (defined in the Data section in the body of the paper).

DO INCONSISTENCIES REFLECT PARTISAN TAMPERING?

We first study the correlation between the party vote and the extent of inconsistencies. For each type of inconsistency $j \in \{1,2,3,4\}$ and for each of the major political parties $k \in \{\text{PRI, PAN, PRD}\}$ we estimate the following regression:

$$PartyVotes^{k}_{pste} = \alpha + \beta^{kj} AbsNumInc^{j}_{pste} + \gamma x_{pst} + n_{st} + \epsilon_{pste} \tag{5}$$

where $PartyVotes^{k}_{pste}$ denotes the number of votes for party k in polling station p within precinct s in election type e (presidential, congressional, senatorial) in election-year t, while $AbsNumInc^{j}_{pste}$ denotes the extent of inconsistencies of type j in absolute value as defined in Section 3.1, and n_{st} are precinct-by-year fixed effects. Including fixed effects in the regression implies that we are only making use of the variation across polling stations in a given precinct in a given year.[6] This has the advantage of controlling for all time-invariant factors that may drive inconsistencies (e.g. number of votes cast in the precinct, or average income) but it throws out all variation across precincts. Therefore we also estimate regressions without fixed effects. We estimate 24 separate regressions (3 parties × 4 inconsistency types × with vs. without FE) and report the estimates for β^{kj} in Table A4.

[6]More precisely, this is true for 2009 and 2015, while for 2012 we also make use of variation across election types (president, congress, and senate).

TABLE A4. Votes and inconsistencies

	(1) PRI	(2) PAN	(3) PRD	(4) PRI	(5) PAN	(6) PRD
Inconsistency 1	0.0070***	-0.0021	0.0020	0.00088	-0.00073	0.0022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 2	0.0062***	-0.0036***	0.0013	0.00085	-0.000071	0.0025***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 3	0.00092	-0.0022	-0.00049	-0.0026	0.0013	0.00099
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 4	-0.0024	-0.0087***	0.019***	-0.0015	-0.00060	0.0018*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PS controls	Yes	Yes	Yes	Yes	Yes	Yes
Precinct controls	Yes	Yes	Yes			
Precinct FE				Yes	Yes	Yes

Note: This table shows the correlation between votes and inconsistencies. We estimate this (partial) correlation via the following regression: $PartyVotes_{kpste} = \alpha + \beta_{kj} AbsNumInc^{j}_{pste} + n_s + xpst'\gamma + \epsilon_{kpste}$ where $PartyVotes_{kpste}$ counts the number of votes for party $k \in \{PRI, PAN, PRD\}$ in poll booth p of precinct s, in election year t, of election-type e (presidential, congressional, senatorial). $AbsNumInc^{j}_{pste}$ measures the inconsistencies of type $j = 1, .., 4$ (as defined in Section 3.1). We estimate 24 separate regressions (3 parties $\times$ 4 inconsistencies $\times$ with and without n_s.) and each cell displays β_{kj} for each regression. For columns 4,5 and 6 precinct FE were included. For regressions corresponding to columns 1, 2 and 3 we use precinct characteristics instead of precinct fixed effects n_s: mean age, education and gender pf PW in each PS and sociodemographic characteristics obtained from the 2010 Census carried out by INEGI. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The estimated association between party vote and inconsistencies is substantively tiny. For ease of interpretation, we transform the estimated coefficients into the number of inconsistencies of each type j associated with 1 additional vote for party k (Table A5). For example, if an estimated coefficient were equal to .01, that would imply that 100 inconsistencies are needed to "generate" a single additional vote for the party.[7] The actual estimated correlations imply that, generally speaking, hundreds or thousands of inconsistencies would be needed to generate a single vote for any of the major political parties. For instance, 500 additional inconsistencies type 1, or 2,401 fewer inconsistencies of type 3, would be required to generate one more vote for the PRD (column 3). Given that on average there are about 8.6 type-1 inconsistencies and 2.8 type-3 inconsistencies per PS, and that the average number of voters per PS is 565 and the maximum 750, it is very difficult to argue that these inconsistencies are associated with a substantively-important misallocation of votes to parties.

The previous analysis deals with national averages over all the election years in our data. It remains

[7]We emphasize that these associations are not necessarily causal.

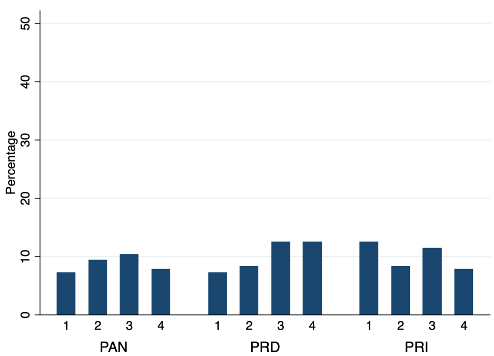
TABLE A5. Inconsistencies needed to add one vote to the respective party

	(1) PRI	(2) PAN	(3) PRD	(4) PRI	(5) PAN	(6) PRD
Inconsistency 1	143	-483	494	1,134	-1,374	456
Inconsistency 2	161	-274	767	1,179	-14,163	398
Inconsistency 3	1,088	-461	-2,048	-391	761	1,008
Inconsistency 4	-417	-115	54	-660	-1,664	568
PS constrols	Yes	Yes	Yes	Yes	Yes	Yes
Precinct controls	Yes	Yes	Yes			
Precint-year FE				Yes	Yes	Yes

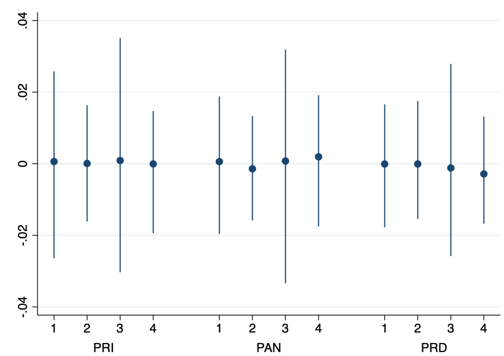
Note: The table displays the extent of the inconsistency (i.e., the number of votes by which an accounting identity must differ) needed to add one vote to party k, on average. Columns 1-3 control for precinct characteristics, while 4-6 contain fixed effects n_s. We estimate 24 separate regressions (3 parties × 4 inconsistency types × with vs. without FE) and report the estimates for β^{kj} in Table A4 in the Appendix. Each cell in the current table displays the inverse of the estimated coefficients: $1/\hat{\beta^{kj}}$. Controls are as in Table A4

possible that inconsistencies could be related to party votes in particular regions and years but not in others, and that such effects could wash out in the pooled average. We therefore repeat the analysis at the state-by-election-year and district-by-election-year levels. Since there are 32 states and we have information on 3 election years, this implies that we estimate 96 coefficients for every combination of political party k and inconsistency type j, for a total of 1,152 coefficients for the state level, and 10,788 coefficients for the district level. As the table below shows, as in the national analysis, the correlation between inconsistencies and party vote is negligible: the median coefficient is zero, less than 1 percent of the coefficients are statistically significant (accounting for multiple testing), and 95 percent of the coefficients are smaller than .02 in absolute value. Figure A3 below shows the fraction of coefficients that are significant at the 5% level, as well as the distribution of their magnitudes. It turns out that 8.8% are significant at the 5% level when no multiple testing correction is done on average, but only 2.6% are significant when using a Benjamini-Holberg correction at the state level.

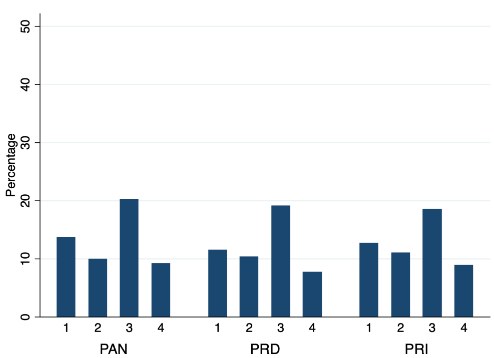
FIGURE A3. Dis-aggregated inconsistencies and votes for political parties: State and District level regressions



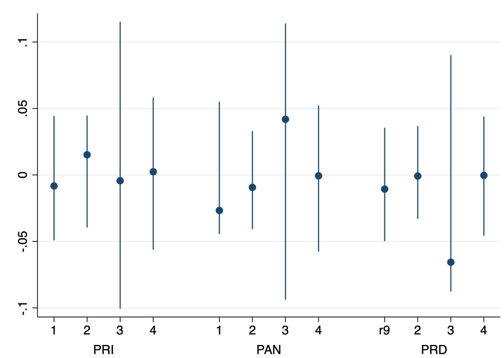
(a) Percentage of significant coefficients, state level



(b) 95% Interval of coefficients, state level



(c) Percentage of significant coefficients, district level



(d) 95% Interval of coefficients, district level

Note: This figure shows the correlation between inconsistencies and votes for political parties. We estimate separate regressionsfor each party, inconsistency, state (or district) and electoral year of the form $PartyVotes^k_{pste} = \alpha + \beta^{kj} NumInc^j_{pste} + x'_{pst}\gamma + \epsilon_{pste}$ where $PartyVotes^k_{pste}$ counts the number of votes for party $k \in \{PRI, PAN, PRD\}$ in poll booth p of precinct s, in election year t, of election-type e (presidential, congressional, senatorial) and x_{pst} are precinct and poll station controls. $NumInc^j_{pste}$ measures the inconsistencies of type $j = 1, .., 4$ (as defined in Section 3.1). Part (a) shows the percentage of significant coefficients β_{kj} for each party and inconsistency. Part (b) shows the interval [percentile 2.5, percentile 97.5] of the empiric distribution of significant coefficients for each party and inconsistency. Parts (c) and (d) are analogous to parts (a) and (b) at the district-year level.

Political party representatives and inconsistencies

We next study whether the presence of party representatives is associated with the extent of inconsistencies. We conjecture that if party representatives can exercise pressure to count or not count specific ballots then we should observe a correlation between their presence and the extent of inconsistencies. In our data, the major political parties covered a large fraction of the approximately-140,000 PS. The PRI, PAN, and PRD respectively covered 93%, 73%, and 55% of PS, on average.

For each $j \in \{1, 2, 3, 4\}$, we regress the extent of inconsistencies of type j (in absolute value) on a dummy for whether party k had representatives in polling station p within precinct s (in a given year t and election e). The model includes one such dummy for each of the major political parties:

$$AbsNumInc^{j}_{pste} = \alpha + \sum_{k} \beta^{j}_{k} RepresentativePresence_{kpste} + n_{st} + \epsilon_{pste} \tag{6}$$

Table A6 displays the results. Our preferred specification includes precinct-by-election year fixed effects (columns 5 to 8), but for completeness we also present the results without fixed effects. Most coefficients in either set of specifications are substantively small and statistically indistinguishable from zero, despite the very large number of observations.

Focusing on the specifications with fixed effects, the presence of a PAN representative, for example, is associated with an additional .21 inconsistencies of type 1 and .25 inconsistencies of type 3, but neither estimate is statistically significant (columns 5 and 7).[8] We also investigate the possibility that the presence of party representatives might moderate the relationship between inconsistencies and party votes. To test this possibility, we interact inconsistencies with an indicator for the presence of party representatives (Table A7). We find no evidence for the moderation hypothesis. In sum, the presence party representatives is not associated with the incidence of inconsistencies.

[8]Because parties choose where to send representatives, these results do not have a straightforward causal interpretation.

TABLE A6. Inconsistencies and party representatives

	(1) Incon. 1	(2) Incon. 2	(3) Incon. 3	(4) Incon. 4	(5) Incon. 1	(6) Incon. 2	(7) Incon. 3	(8) Incon. 4
PRI representatives	0.429	-0.228	-0.317*	-0.015	0.794	-0.391	-0.679*	0.106
	(0.34)	(0.36)	(0.17)	(0.27)	(0.83)	(0.79)	(0.36)	(0.52)
PAN representatives	-0.697***	-1.402***	0.133	0.051	0.210	-0.038	0.248	0.293
	(0.23)	(0.23)	(0.10)	(0.17)	(0.55)	(0.53)	(0.24)	(0.38)
PRD representatives	-0.168	-0.230	0.126	0.120	0.199	0.367	0.422*	0.644*
	(0.19)	(0.19)	(0.09)	(0.14)	(0.48)	(0.47)	(0.23)	(0.34)
Constant	9.433***	10.510***	3.311**	1.712	0.148	37.961	9.166	24.920
	(3.33)	(3.47)	(1.51)	(2.62)	(23.35)	(26.84)	(13.73)	(27.04)
N	437782	521736	517122	424481	437939	521915	517304	424633
R-sq	0.004	0.007	0.002	0.004	0.462	0.460	0.374	0.353
Precinct Controls	Yes	Yes	Yes	Yes				
Precinct FE					Yes	Yes	Yes	Yes

Note: Each column corresponds to a different regression. Standard errors (clustered by precinct-year) are shown in parentheses below estimated coefficients. The unit of observation is an *acta*. Columns 1-4 include precinct-level controls, while 5-8 control for precinct fixed effects. We do not have information of inconsistency type 4 for the year 2015. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A7. Votes and party representatives interaction with inconsistencies

	(1) PRI	(2) PAN	(3) PRD	(4) PRI	(5) PAN	(6) PRD
Inconsistency 1	-0.005	0.004	0.006*	-0.001	-0.001	-0.001
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 2	-0.001	-0.005**	0.007**	0.000	-0.001	-0.002*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inconsistency 3	-0.020**	0.004	-0.007	-0.007	0.000	-0.002
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Inconsistency 4	-0.016**	0.006	-0.008	-0.002	-0.001	-0.000
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
PS controls	Yes	Yes	Yes	Yes	Yes	Yes
Precinct controls	Yes	Yes	Yes			
Precinct FE				Yes	Yes	Yes

Note: This table shows estimates for the interaction between the presence of political parties representatives at the PS and inconsistencies. We estimate this (partial) correlation by estimating the following regression:: $PartyVotes^k_{pste} = \alpha + \gamma^{kj} PartyPresent^k_{pst} + \delta^{kj} AbsNumInc^j_{pste} + \beta^{kj} PartyPresent^k_{pst} * AbsNumInc^j_{pste} + n_{st} + \epsilon_{pste}$. $AbsNumInc^j_{pste}$ measures the inconsistencies of type $j = 1, .., 4$, party is indexed by $k \in \{PRI, PAN, PRD\}$, poll booth by p, precinct by s, election year by $t = 2009, 2012, 2015$, and election-type by e (presidential, congressional, senatorial). The unit of observation is a PS in one election process. Each cell presents the estimated β^{kj} for a different regression, for a different party and a different inconsistency, and a different specification (24=3x4x2). Columns 1 to 4 use precinct level controls, while columns 5 to 8 use precinct year level fixed effects. Standard errors in parentheses are clustered at the precinct-year level. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Persistence of inconsistencies

If inconsistencies reflected partisan tampering, and if party machines varied geographically in their capacity to tamper with the vote, then one would observe that the extent of inconsistencies persists over time (and over consecutive elections) in a given geographical unit. We test for this possibility by estimating an AR(1) model for each type of inconsistency $j \in \{\text{all}, 1, 2, 3, 4\}$, where geographical unit of analysis is the precinct. We estimate the following equation:

$$\frac{AbsNumInc^j_{s,t}}{votes_{s,t}} = \alpha + \gamma^j \frac{AbsNumInc^j_{s,t-1}}{votes_{s,t-1}} + \phi_t + \nu_{s,t} \qquad (7)$$

where ϕ_t are year fixed effects. The closer γ^j is to 1, the greater the persistence of inconsistency type j. For each inconsistency of type $j = 1, .., 4$, we compute the ratio of inconsistency type-j ($AbsNumInc^j$) in the precinct over the total number of votes. Table A8 presents these results. The dependent variable in column 1 is the average extent of all inconsistencies per ballot cast in precinct s in election year t. The dependent variables in columns 2-5 is the average extent of each type of inconsistencies. We find no evidence of persistence.

TABLE A8. Persistence of inconsistencies

	(1) All Inc.	(2) Inc. 1	(3) Inc. 2	(4) Inc. 3	(5) Inc. 4
Lag	0.001	0.002	0.001	0.000	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.072***	0.016***	0.029***	0.007***	0.020***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
N	119140	119140	119140	60477	119140
R-sq	0.001	0.000	0.000	0.000	0.000

Note: This table shows coefficients of the model: $\frac{AbsNumInc^j_{s,t}}{votes_{s,t}} = \alpha + \gamma^j \frac{AbsNumInc^j_{s,t-1}}{votes_{s,t-1}} + \phi_t + \nu_{s,t}$, where ϕ_t are year fixed effects. The closer γ^j is to 1, the greater the persistence of inconsistency type j. For each inconsistency of type $j = 1, .., 4$, we compute the ratio of inconsistency type-j ($AbsNumInc^j$) in the precinct over the total number of votes. The dependent variable in column 1 is the average extent of *all* inconsistencies per ballot cast in precinct s in election year t. The dependent variables in columns 2-5 is the average extent of each type of inconsistencies. We do not have data for inconsistency 4 in 2015. Standard errors shown in parentheses below coefficient estimates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

All the estimates of γ^j are very close to zero and none are statistically significant, indicating essentially no over-time persistence. This result suggests that inconsistencies have an important random component and that fixed characteristics of localities can explain only a limited amount of the variation in inconsistencies.

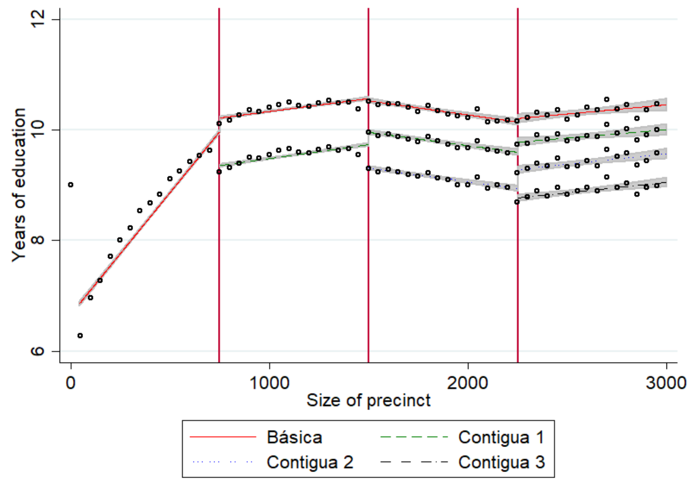
In sum, all the evidence we examined (on party votes, party representatives, and persistence), together with the scholarly consensus on the state of contemporary Mexican elections, suggest that inconsistencies do not arise out of partisan manipulation, but instead reflect primarily honest mistakes. However, as with any forensic voting analysis, some types of manipulation are rejected but not all of them.

CAUSAL EFFECT OF THE EDUCATIONAL ATTAINMENT OF PW ON INCONSISTENCIES

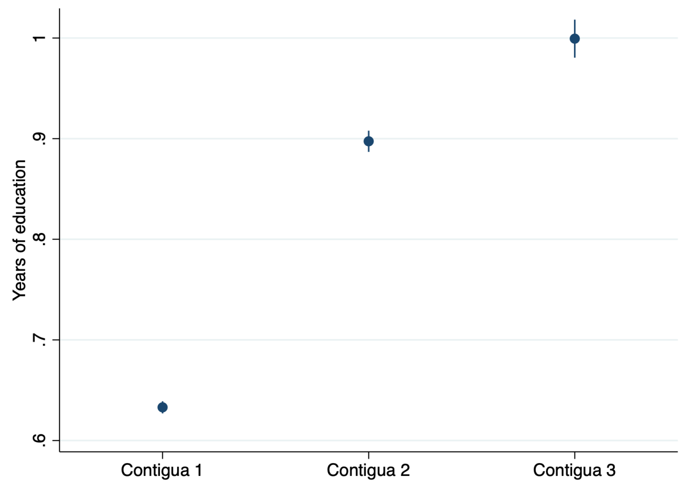
We complement the explanation of the role of educational attainment in the allocation of PWs to PSs in the body of the paper with an additional explanation of the allocation process. Consider a precinct where two polling stations are needed to serve the amount of citizens registered to vote at that precinct. Of the list of citizens eligible to serve as poll workers, 14 citizens are selected. Next, this list of 14 citizens is sorted in descending order of educational attainment, so that spot 1 on the list corresponds to the citizen with the greatest educational attainment, and spot 14 to the citizen with the lowest educational attainment. The next step is to assign citizens from this list to the two polling stations. The citizen in spot 1 on the list (i.e., the most highly-educated citizen) is assigned to be President of the Basica polling station. The citizen in spot 2 of the list (i.e., the second most highly-educated one) is assigned to be President of the Contigua polling station. Then the citizen in spot 3 is assigned to be Secretary in the Basica polling station, and the citizen in spot 4 is assigned to be Secretary in the Contigua polling station. This procedure is followed to assign each of the 7 poll-worker spots for the two polling stations. Because of this assignment procedure, it is necessarily the case that the President in the Basica polling station has equal or greater educational attainment in comparison with the President in the Contigua polling station. The same is true for the Secretary, for the First Counter, the Second Counter, and the three Replacements—i.e., for each of the poll workers. It necessarily follows that the average educational attainment of poll workers in the Basica polling station is equal to or greater than that of poll workers in the Contigua polling station. In practice, the difference in average educational attainment of poll workers in Basica vs. the first Contigua polling station in a precinct is approximately 0.7 years, as illustrated in Figure A4.

In other words, the reason that the basica casilla has more-educated poll workers is that there is a single list of poll workers for a whole precinct. The list is then ranked by education, and then the respective positions in the different polling stations within the precinct are sequentially filled as explained. This means that position for position, the education of workers in the basica polling station is larger than or equal than the education of those in the contigua polling station. Figure A4 verifies that this is the case in practice.

FIGURE A4. Difference of years of education



(a) Years of Education by type of PS



(b) Within Precinct Differences across PS types

Note: This figure shows the relation between average years of education of the PW designated to each PS, and type of PS (basica vs contigua). Part (a) is a bin scatterplot (bins of size 30) of average years of education against the size of the precinct. Recall that if a precinct has between 750 and 1500 of registered voters, it must have two PS: the first is called Basica while the second is called contigua 1, and so on. The dots indicate the average education for the bin, and different line colors/patterns indicate if the average is taken for basicas, contigua 1, contigua 2, etc. As we would expect from the allocation rule, PW at basicas have more education that at contigua 1, whereas contiguas themselves are ranked. Part (b) avoids comparisons across precincts of different sizes, and computes the average *difference* in education across basica and contigua 1 *within* the same precinct. It does the same comparing Basica vs contigua 1, etc. It then averages this difference across all precincts and plots this difference along with a 95% confidence interval.

TABLE A9. First Stage Years of education

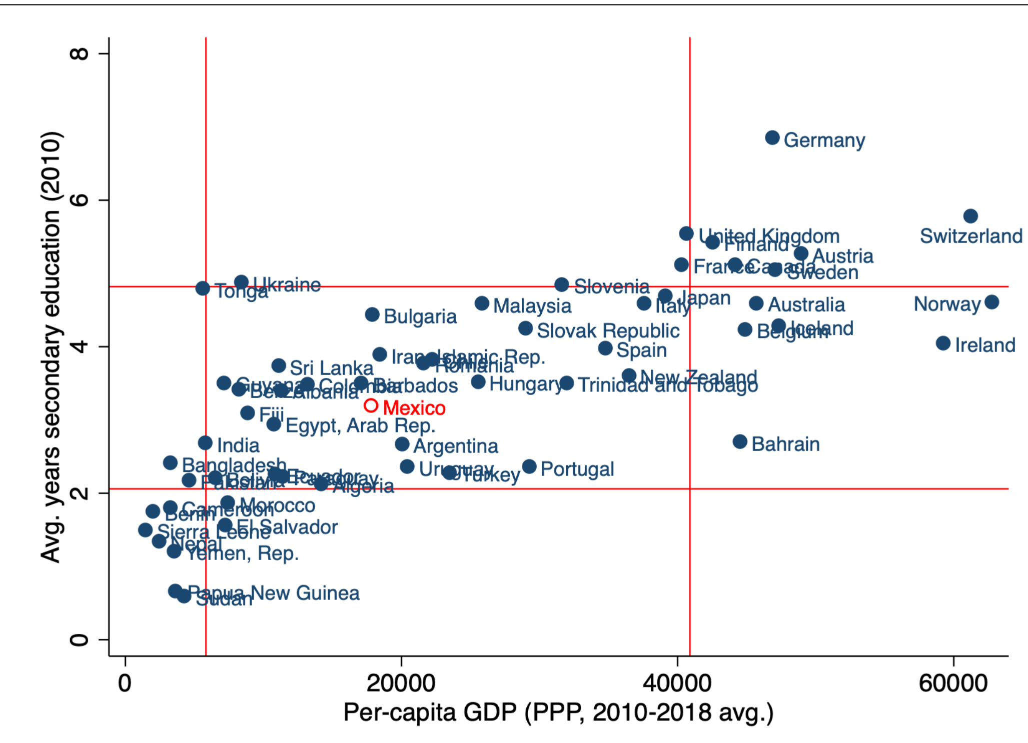
	(1) Years of education
Dummy PB basica	0.743***
	(0.00)
Age	-0.029***
	(0.00)
Male (%)	0.168***
	(0.01)
CAE age	-0.000
	(0.00)
CAE male	-0.011
	(0.04)
CAE score	0.056
	(0.08)
CAE education	-0.009
	(0.01)
Constant	10.885***
	(0.68)
N	494061
R-sq	0.923
F	5871.545

Note: This table shows the first stage of the IV estimate of inconsistencies and years of education. Taking advantage of the allocation rule of PW to PS, we instrument the average education of workers in a poll booth by an indicator of whether the poll booth is Basic or Contiguous, while controlling for precinct fixed effects and other PW average characteristics. precincts with less than 2 PS do not have contiguous polling booths and are excluded. We also exclude 994 PS which are placed in precinct with more than 14 PS. Standard errors level in parentheses. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. Note that the first stage is powerful for two main reasons: the first is that we have hundreds of thousands of observations. The second is that the allocation of PWs with different levels of education across different PS is a legal rule that is strictly followed, and therefore the instrument is sharp. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

EXTERNAL VALIDITY: MEXICO IN COMPARATIVE PERSPECTIVE

Figure A5 displays 58 countries where votes are counted by hand. The countries are placed on a "level of development" space, with the horizontal axis denoting per-capita GDP and the vertical axis the average years of secondary education (the latter two variables are taken from the World Development Indicators; data on hand counting are from the ACE Project). The red lines denote an interval of plus/minus one standard deviation from the mean, for each of the two development variables. As it happens, Mexico is close to the means of both variables. The graph suggests that Mexico is quite a typical middle-income country in terms of education, and therefore the issues with hand counting that Mexico experiences are likely to be present in many other countries.

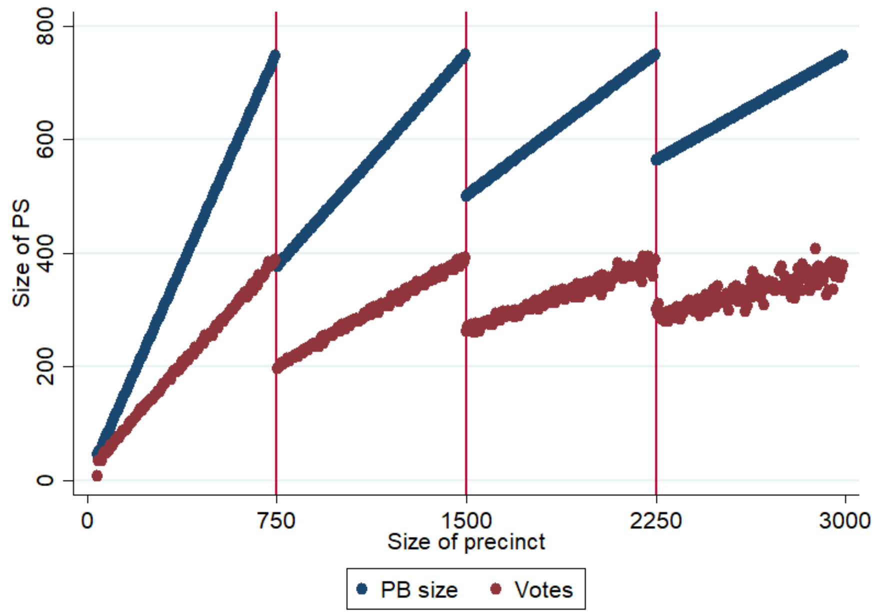
FIGURE A5. Years of secondary education: international comparison



Note: This figure shows years of education against per capita GDP for countries with hand counting for which we could also get data on per capita GDP and secondary education, a total of 58 countries. Red lines display 1 standard deviation away from the mean values of education and per-capita GDP for each axis.

CAUSAL EFFECT OF WORKLOAD ON INCONSISTENCIES

FIGURE A6. Precinct Size and votes cast



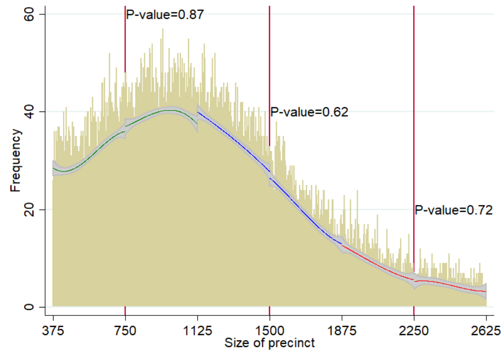
Note: This figure shows the relation between PS size, votes cast and precinct size. Each point corresponds to the average of PS sizes and votes in a neighborhood of size 5 of precinct size. The vertical red lines correspond to precinct sizes where an additional PS is required. It shows that the legal rule of precinct size is strictly followed.

TABLE A10. RD: Tests of Quasi-Random Assignment of Pre-Determined Characteristics

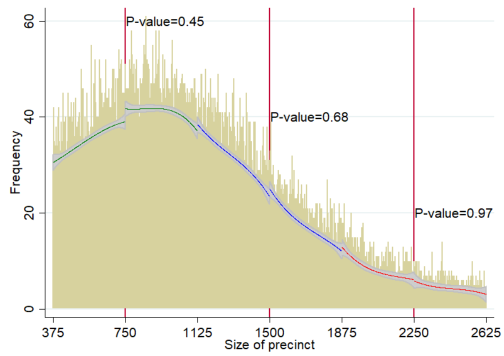
	(1) Population	(2) Houses without goods	(3) Employment Rate	(4) Male PW	(5) PW Education	(6) PW age	(7) CAE evaluation	(8) Difficulty Dummy
				Panel A: 2009 Federal Election				
Above Cutoff 751	30.127*	0.229	-0.281	0.852	0.005	-0.551	-0.007	0.023
	(18.18)	(0.28)	(0.20)	(1.26)	(0.14)	(0.35)	(0.03)	(0.03)
Above Cutoff 1501	-8.083	-0.082	-0.070	3.083**	-0.203	-0.094	0.049	-0.016
	(35.59)	(0.21)	(0.17)	(1.47)	(0.14)	(0.40)	(0.03)	(0.03)
Above Cutoff 2251	84.674	-0.414	-0.294	-0.363	-0.133	-0.000	-0.021	0.015
	(144.54)	(0.31)	(0.36)	(3.32)	(0.27)	(0.82)	(0.07)	(0.07)
N	51287	51265	51260	51281	51274	51281	47538	39452
				Panel B: 2012 Federal Election				
Above Cutoff 751	-5.952	-0.338	0.215	-1.156	0.292**	0.272	-0.002	0.041
	(15.42)	(0.26)	(0.21)	(1.19)	(0.12)	(0.34)	(0.02)	(0.03)
Above Cutoff 1501	-5.531	-0.197	-0.174	-0.459	0.209	0.603	0.015	0.016
	(30.99)	(0.24)	(0.18)	(1.48)	(0.14)	(0.42)	(0.02)	(0.03)
Above Cutoff 2251	-12.429	0.371	0.401	2.367	0.374	-1.035	-0.129***	-0.019
	(84.19)	(0.41)	(0.31)	(2.95)	(0.24)	(0.78)	(0.04)	(0.06)
N	51621	51604	51599	51618	51615	51618	48378	38340
				Panel C: 2015 Federal Election				
Above Cutoff 751	-4.332	-0.237	0.015	1.399	0.120	0.253	0.012	0.007
	(14.28)	(0.24)	(0.21)	(1.11)	(0.13)	(0.35)	(0.02)	(0.03)
Above Cutoff 1501	14.291	0.213	0.168	-1.383	-0.129	0.180	-0.048	0.026
	(23.04)	(0.24)	(0.18)	(1.41)	(0.16)	(0.43)	(0.03)	(0.03)
Above Cutoff 2251	49.495	0.181	-0.152	4.295	-0.156	-0.532	-0.003	0.036
	(71.27)	(0.36)	(0.32)	(2.75)	(0.31)	(0.81)	(0.06)	(0.07)
				Means at cutoff [-100,-1]				
Cutoff 751	911.843	2.283	95.245	41.052	11.229	37.718	8.549	0.388
Cutoff 1505	1847.132	1.684	95.237	41.395	11.785	37.208	8.582	0.348
Cutoff 2251	2988.767	1.749	95.309	39.833	11.589	36.311	8.609	0.338
N	48481	48467	48462	48481	48475	48442	46088	37011

Note: This figure shows the balance tests for pre-treatment characteristics, using regression discontinuity methodology where the running variable is precinct size. Three cutoffs at which precincts are split into one more PS are used: 751, 1501 and 2251. More concretely, we estimate by OLS the following regression $Y_{ste} = \alpha + \beta_1 I(precinct_s >= 751) + \beta_2 I(precinct_s >= 1501) + \beta_3 I(precinct_s >= 2251) + Poly(s, interval_k) + \epsilon_{ste}$, where $Poly(size, interval_j)_p$ is a polynomial function of the size of precinct s, a different one at each interval $j \in [0, 750], [751, 1500], [> 1500]$ of the precinct s size. Part A shows the RD estimation for the 2009 federal election, Part B for the 2012 federal election and Part C for the 2015 federal election. Part D shows the average of the dependent variable for the interval [-100,-1] from each cutoff. For the RD, a linear model with a bandwidth of plus/minus 374 voters was used. Clustered standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

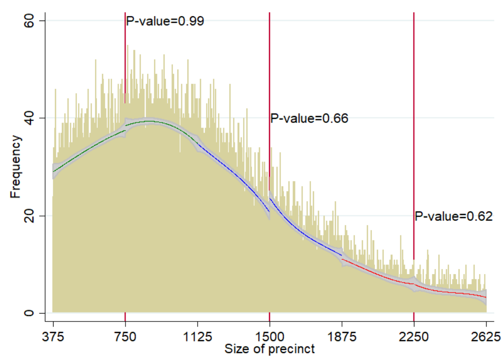
FIGURE A7. Distribution of precinct size (McCrary Tests)



(a) 2009 Federal Election



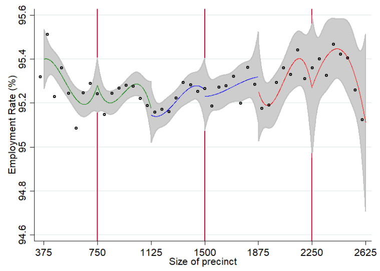
(b) 2012 Federal Election



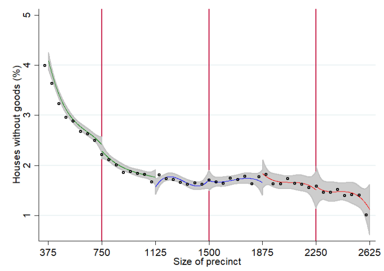
(c) 2015 Federal Election

Notes: To assess the likelihood of sorting across thresholds, this figure shows the distribution of precinct size for each electoral year. Lines represent third degree polynomial approximations. We show p-values of McCrary tests for the null hypotheses that treatment effect is equal to 0 is reported for each cutoff. We fail to reject that there is no jump in the density, consistent with no manipulation of the running variable.

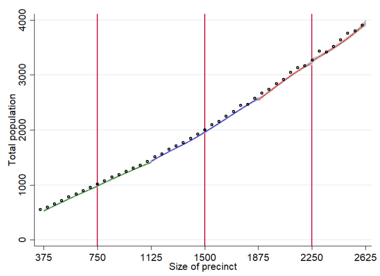
FIGURE A8. Pre-Treatment Characteristics



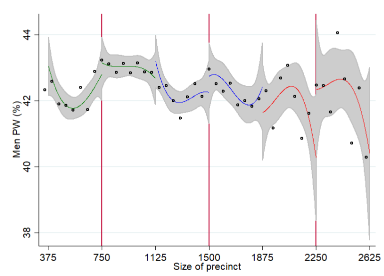
(a) Employment Rate (%)



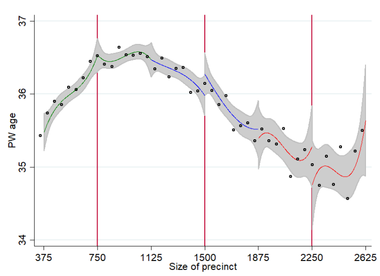
(b) Houses without goods (%)



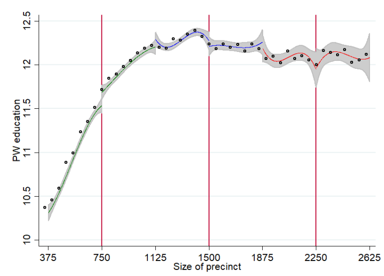
(c) Total population



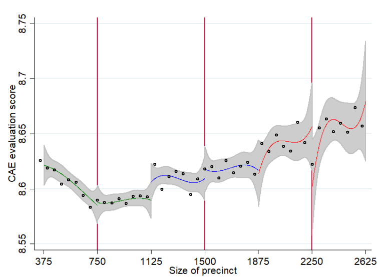
(d) % Men PW



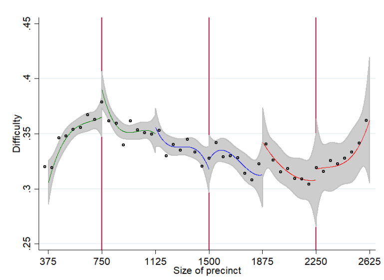
(e) PW age



(f) PW education



(g) CAE evaluation



(h) Difficulty

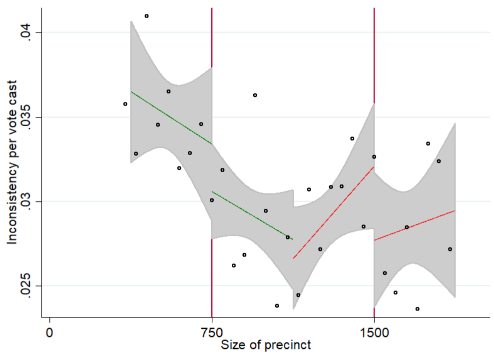
Notes: This figure shows the balance tests for pre-treatment characteristics of the RD methodology with the precinct size as running variable The unit of observation is a PS in one election process. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections and INEGI census. Each point represents the average in a 30 precinct size neighborhood. Lines represent the RD methodology estimation and the 95% confidence interval. The RD estimate consists of a linear model with a 375 bandwitdh of each cutoff (red lines). Vertical lines represent cutoffs.

TABLE A11. RD estimates of the effect of the PS-size-ceiling on inconsistencies

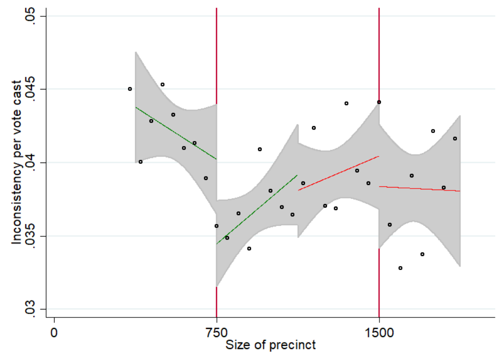
	(1) Inconsistency 1	(2) Inconsistency 2	(3) Inconsistency 3	(4) Inconsistency 4
		Part A: RD (OLS)		
Above Cutoff 751	-5.493***	-7.607***	-1.923***	-4.014***
	(0.74)	(0.73)	(0.34)	(0.54)
Above Cutoff 1501	-4.524***	-5.449***	-1.689***	-4.194***
	(0.93)	(0.95)	(0.45)	(0.67)
Above Cutoff 2251	-3.706*	-4.504**	-1.642*	-3.652**
	(1.92)	(1.97)	(0.88)	(1.57)
		Part B: IV		
Above Cutoff 751	0.027***	0.037***	0.009***	0.019***
	(0.00)	(0.00)	(0.00)	(0.00)
Above Cutoff 1501	0.034***	0.040***	0.012***	0.030***
	(0.01)	(0.01)	(0.00)	(0.00)
Above Cutoff 2251	0.037*	0.045**	0.017*	0.036**
	(0.02)	(0.02)	(0.01)	(0.02)
		Part C: Means[-10,-1] from cutoff		
751	10.165	13.578	4.987	6.860
1501	13.477	17.504	5.783	13.851
2251	10.575	8.639	1.699	8.787
N	189237	225581	223440	184995
		Panel D: Testing (p-values)		
RD 751 = 1501 = 2251	0.559	0.106	0.896	0.944
RD 751 = 1501	0.415	0.072	0.678	0.835
RD 751 = 2251	0.385	0.140	0.765	0.827
RD 1501 = 2251	0.702	0.666	0.962	0.751

Note: The figures shown correspond to regression discontinuity estimates of the causal effect on the extent of tally inconsistencies of the jump in precinct size at the various cutoffs. The cutoffs are: 751, 1501 and 2251 registered voters. Part A displays RD estimates with precinct size as the running variable. In Panel A we estimate by OLS the following regression $Inconsistency^k_{ste} = \alpha + \beta_1 I(precinct_s >= 751) + \beta_2 I(precinct_s >= 1501) + \beta_3 I(precinct_s >= 2251) + Poly(s, interval_k) + \epsilon_{ste}$, where $Poly(size, interval_j)_p$ is a polynomial function of the size of precinct s, a different one at each interval $j \in [0, 750], [751, 1500], [> 1500]$ of the precinct s size. $Inconsistency^k_{ste}$ is the inconsistency in a given precinct in a given election, in a given year. This gives the effect of precinct size on the extent of inconsistencies. Part B uses and IV design to estimate the effect of the number of votes on inconsistencies, using the threshold dummies (e.g. $I(precinct_s >= 751)$) as instrumental variables for the number of votes in the PS. Note that if a precinct has 751 votes it will have half as many enrolled voters in each of its two PS, than if it had 750 registered voters. So the threshold dummies mechanically affect the number of registered voters, which affect the number of actual voters. Part C shows the average extent of inconsistencies for the interval [-100,-1] from each cutoff. Part D reports p-values for tests of equal coefficients at different cutoffs. For the RD, a linear model with a bandwidth of plus/minus 374 registered voters from each cutoff was used. Clustered standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

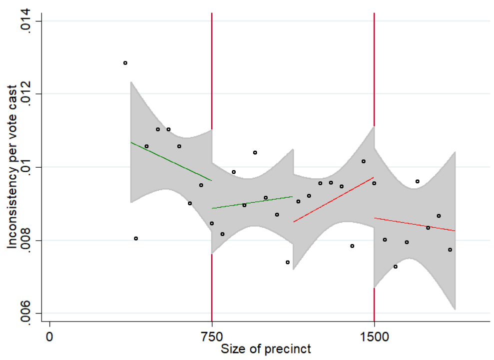
FIGURE A9. Effect of splitting precincts on inconsistencies per vote cast



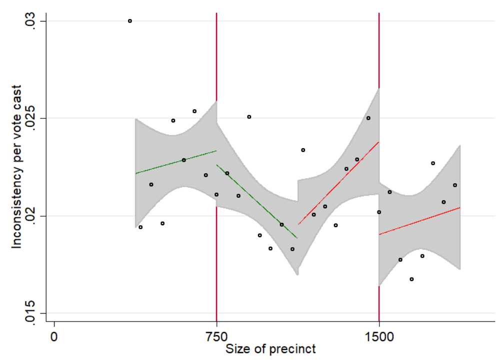
(a) Inconsistency 1



(b) Inconsistency 2



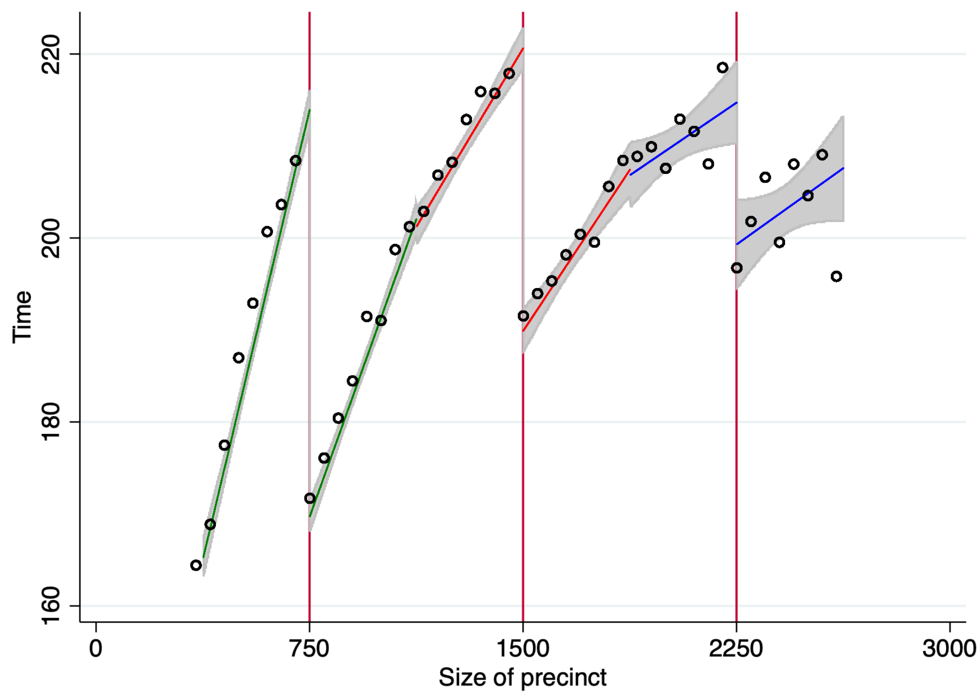
(c) Inconsistency 3



(d) Inconsistency 4

This figure shows the effect of splitting PS on inconsistencies **per vote cast** using a RD methodology with the precinct size as running variable and split PS as treatment dummy. That is, it measures the proportion of inconsistencies as a function of votes as the dependent variable. The rest of the specification is as the tables before. Data employed for this estimation comes from INE administrative information for 2009, 2012 and 2015 federal elections. Each point represents the average inconsistency in a 30 precinct size neighborhood. Lines represent the RD estimate and the 95% confidence interval. The RD estimate consists of a linear model with a 374 bandwidth of each cutoff. It is important to note that **the scale on the Y axes shown here is very different from that in Figure 2 in the body of the paper**, implying that the rate of inconsistencies is essentially flat as a function of precinct size.

FIGURE A10. Effect of splitting PS on counting time (minutes)

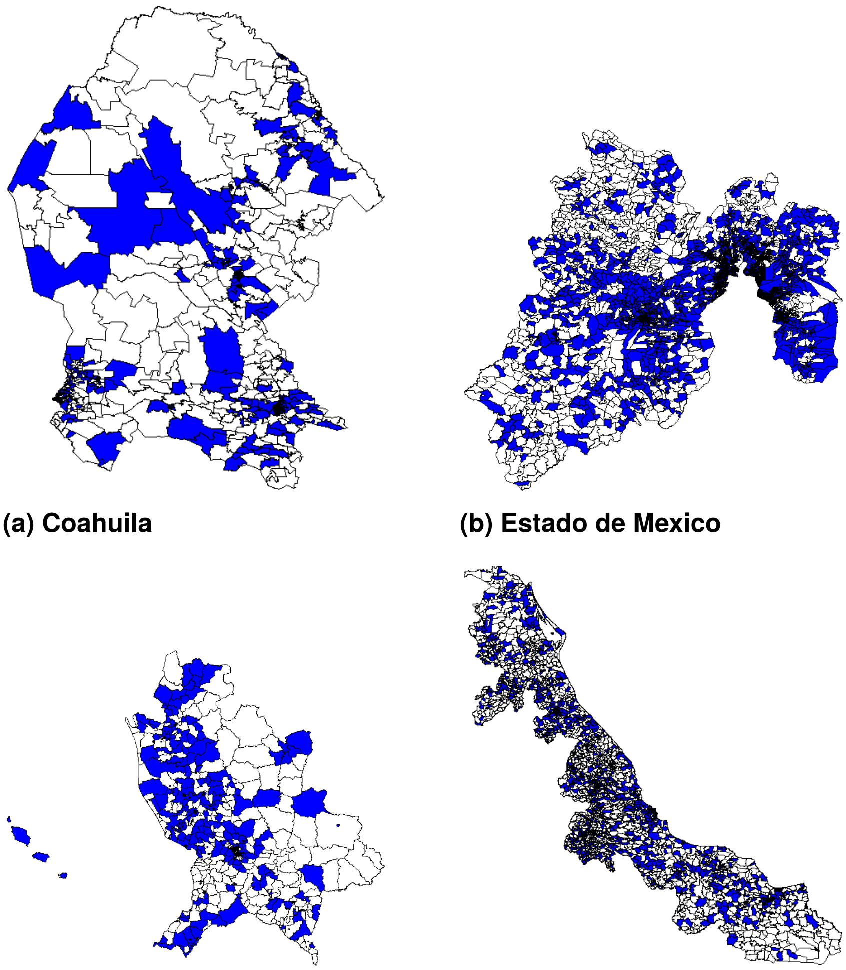


Note: This figure shows the effect of splitting PS on time of counting using a regression discontinuity methodology where precinct size is the running variable and the cutpoints correspond to precinct sizes at which an additional PS is added, as explained in the body of the paper. The specification is as in the other graphs explained above. Time is measured in minutes from when the PS is closed to voting to when the numbers of the acta are delivered, thus it is just tallying time.

TALLY QUALITY AND BALLOT RECOUNTS

The following Figure shows a map of recounts for 4 selected States.

FIGURE A11. Precincts with at least one acta recounted



(a) Coahuila **(b) Estado de Mexico**

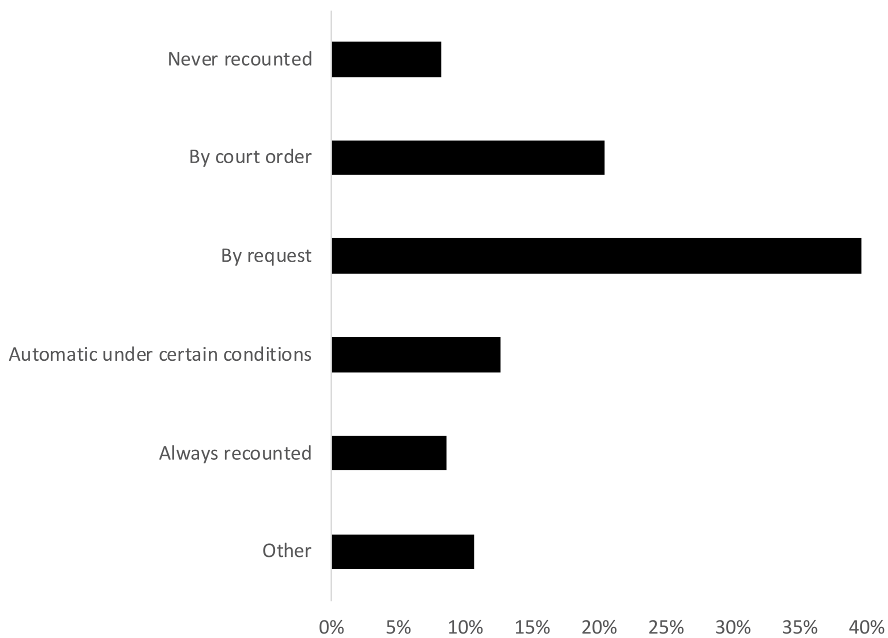
(c) Nayarit **(d) Veracruz**

Notes: This figure shows in blue the precincts were at least one acta was recounted in 2015.

Inconsistencies as causes of recounting ballots internationally

Recount legislation is very prevalent. Figure A12 shows what fraction (out of 177 countries) have vote recounting legislation by type of recount trigger. Since a country's law may consider several recount triggers, an observation is a country-recount type pair (not just a country). Only about 1 in 10 countries do no provide for vote recounts in their electoral laws. Within those that do, the most common trigger is an explicit request—as is the case in Mexico, where inconsistencies in the tally can lead to a recount if a political party representative makes a request.

FIGURE A12. Share of countries with vote recount Legislation



Note: Source: ACE Project (aceproject.org, accessed November 2019). The figure shows the fraction of countries (out of 177 countries for which the ACE Project provides information on laws governing recounts) where the law contemplates a particular provision for recounting votes. The codification of laws into the different kinds of recount provisions is taken from the ACE Project. Since a country may consider several recount triggers, an observation is a country-recount type pair (not just a country). "Never recounted" means that the law does not contemplate vote recounts. The rest of the bars correspond to the ways vote recounts can legally take place. The total number of country-provision observations in the graph is 245, corresponding to 121 countries with only one provision, 45 with two provisions, 10 with three provisions, 1 with four provisions.

The following table cites specific articles in the law for a selected sample of countries where inconsistencies in the vote tally constitute legal grounds for recounting ballots.

TABLE A12. Inconsistencies and recounting ballots

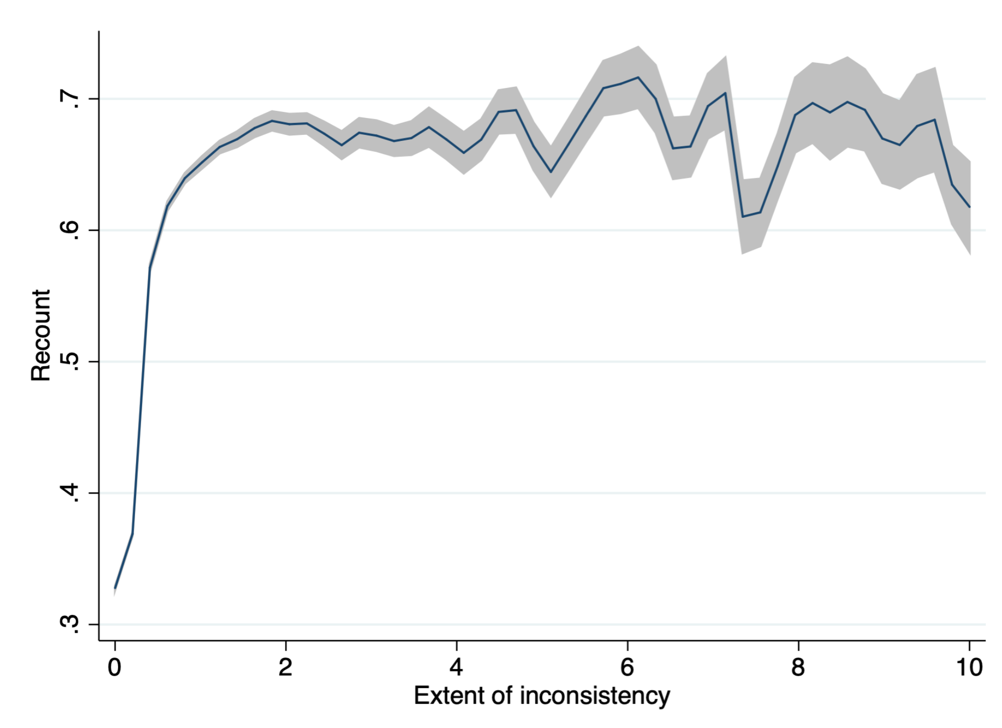
Country	Law
Argentina	Ley 19945, Código Nacional Electoral, art. 118: (2012)
Austria	Federal Law on Parliamentary Elections (1992), Art. 110
Brazil	"Código Eleitoral - Lei 4.737", art. 179 (II §8), 180 (II), 181: (2018); "Lei N 9.504 - das eleições", art. 88: (2018)
Chile	Electoral Law: "Ley Orgánica Constitucional sobre Votaciones Populares y Escrutinios (2016)", art. 96 & 97
Colombia	"Código Electoral", art. 122, 163, 164, 182, 189, 192: (2016)
Denmark	Parliamentary Elections and Election Administration in Denmark, Ch.5, Art. 5.2
Ecuador	"Ley Orgánica Electoral y de Organizaciones Políticas. Código de la Democracia", art. 139, 145: (2017)
Honduras	"Reglamento del Escrutinio General Definitivo en las Elecciones Generales 2009", art. 40
Mexico	"Ley General de Instituciones y Procedimientos Electorales", art. 311, 313 & 314: (2014)
Spain	Electoral Law 5/1985 of 19 June: "Ley Orgánica del Régimen Electoral General", art. 95 & 106: (2016)

Note: This table shows a partial list of countries where inconsistent vote tallies are a legal reason for recounting ballots. A complete list of electoral laws per country can be found at `https://tinyurl.com/y7b4eo23`.

Relationship between inconsistencies and recounts: presence vs. extent

The following figure A13 shows a kernel regression of the relationship between the fraction of PS recounted and the extent of inconsistencies. The relationship is not linear. It seems that having a few inconsistencies (say from 1 to 20) is enough to sharply increase the likelihood of inconsistencies, with further inconsistencies beyond those not increasing the likelihood. It is also consistent with Mexican electoral law, which specifies that it is the presence, and not the extent, of inconsistencies in an *acta* that provides grounds for a recount. Our working paper version shows that the *extent* of inconsistencies are uncorrelated with recounts once we control for a dummy of their *presence*.

FIGURE A13. Recounts and extent of inconsistencies



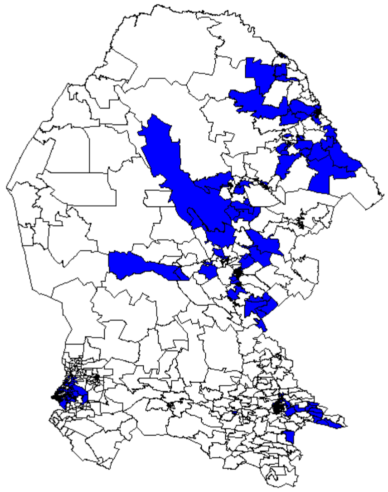
Note: This figure shows the Kernel-weighted local polynomial smoothing of the probability of recounts and average extent of inconsistencies. The average extent of inconsistencies corresponds to the average of the four types of inconsistencies at each PS. The gray area shows the 95% confidence interval.

INCONSISTENCIES AND TRUST IN THE ELECTORAL AUTHORITY

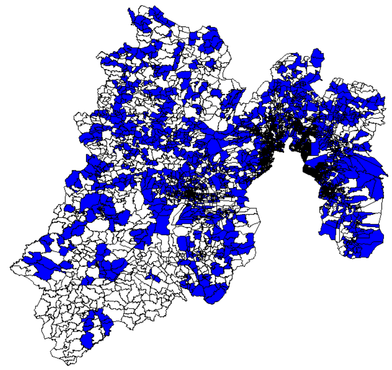
Poll worker survey (2017) The survey was implemented in the 2017 local elections in Coahuila, Estado de Mexico, Nayarit and Veracruz. The survey was targeted to a random sample of almost 80,000 PW of the election during the second stage of the recruitment process. The survey was carried out by the CAE in the PW house during the first visit of the second stage. This visit corresponds to the notification of acceptance as PS with an specific role. At the moment of the survey the PW were not trained yet, so their attitudes towards INE and democracy should not have been affected.

The implementation of surveys were randomized at the CAE level. Of the 6,690 CAE, 4,014 were selected to implement the survey to all their PW. In total we have 85,006 answered surveys in 7,161 different precincts. Figure A14 shows the precincts in which at least one PW did the survey.

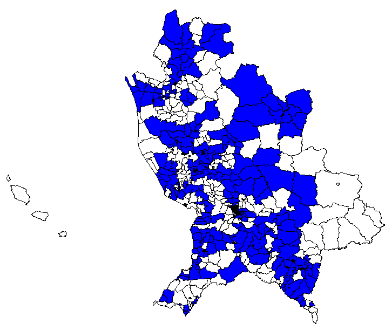
FIGURE A14. Precincts with surveys by state



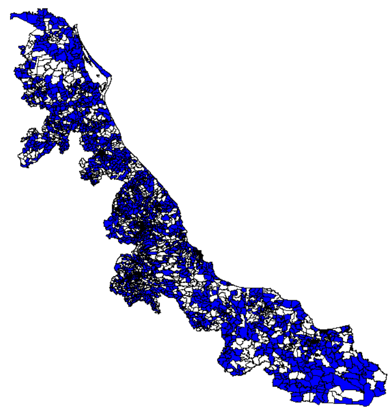
(a) Coahuila



(b) Estado de Mexico



(c) Nayarit



(d) Veracruz

Notes: This figure shows in blue the precincts were the workers survey was implemented in each of the four states of the 2017 Federal Election.

Recounts and trust: robustness analyses

This section presents robustness checks and supporting information pertaining to the findings in the paper's section: "Inconsistencies and trust in the electoral authority."

News about recounts in social media: We checked 7 major media outlets directly in their webpages: El Universal, Reforma, La Jornada, La Razon, Excelsior, Milenio, and Animal politico (this later one is an influential web-based investigative news outlet). We searched for articles containing the following keywords: recounts, recounting, "fraude electoral" (election fraud), "urnas" (ballot boxes), "voto por voto" (vote by vote), "casilla por casilla" (polling station by polling station), "paquetes electorales" (electoral packets, referring to the box containing ballots and the tally corresponding to a polling station). We found 157 articles covering recounts. The titles themsevles give an idea of just how widespread geographically this issue is. Titles of the news articles included:"INE pide modificar ley para evitar los amplios recuentos de votos", "Inicia recuento de más del 70% de paquetes electorales", "Recuento fue jaque mate a la elección dice Morena", "Inicia recuento de votos en BCS", "Acusa PES conteo erróneo en elección", "Monterrey: recontarán oootra vez 800 paquetes", "Reserva TEPJF resultado de recuento", "Van por recuento total en Ciudad Juárez", "Habrá recuento en la elección para gobernador en Puebla", "Hubo fraude en Puebla; el IEE, cómplice: Morena", "Inicia INE recuento de votos en todo el país", "Voto por voto decidirá resultado en Ciudad del Carmen", "Candidato del PAN por Monterrey pide recuento total de votos", "Anular y recontar votos, las estrategias del Panal y del PES para conservar el registro".

We also checked select local media, with similar results. Local media presented the complication that in smaller municipalities many outlets do not publish their news online and, even when they do, it is difficult to search their archives. We view our article count as a lower bound on the information on recounts that citizens likely are exposed to in printed paper or TV in their municipalities. To narrow down the search, we first identified the 30 municipalities with more recounts in 2018. We then searched the local media with an online presence in those municipalities. We found a total of 210 news articles in 44 local newspapers in those municipalities using the same search criteria as above. That amounts to an average of 5 articles per outlet.

We believe these findings provide clear evidence that vote recounts are indeed covered by the news media. We take our findings as a lower bound to the overall dissemination of information about vote recounts for the following reasons. First, social media (WhatsApp and Facebook primarily) played a key role in 2018 in the dissemination of political propaganda and partisan information, but we were unable to collect systematic information on coverage of recounts in such media. Second, we were unable to scan the hundreds of local newspapers that would have covered recounts associated with the results of locally-relevant (but not always nationally-relevant) legislative elections. Third, the 2018 election is a hard case for finding coverage of vote recounts because the election resulted in the most overwhelming victory, by far, for any political party since Mexico became a democracy in 2000. Morena's victory was thorough both at the local and the national levels, in accordance with virtually all pre-electoral polling. The overwhelming and pre-announced nature of the election's outcomes almost certainly dampened incentives for political parties to push for recounts, as there was very little hope of either reverting the results or of successfully calling into question the election's outcomes in the public eye. In comparison, the 2012 elections (also presidential, like those in 2018) resulted in forceful political posturing about their legitimacy, in media spectacle featuring various political parties, and in large numbers of vote recounts.

Figure A15 shows more than a million attendees to a rally whose purpose was to de-legitimize the electoral authority's counting of votes in 2006.

Comparing precincts with vs. without recounts: Figure A16 compares characteristics of precinct where no PSs were recounted vs. precincts where some or all PSs were recounted. While balance on observables provides neither necessary nor sufficient grounds to determine whether estimates in Table 4 are confounded by unobservables[9], we followed a reviewers' suggestion and provide a "balance test" in Figure A16. Before we explain the figure a note is in order. It is entirely possible for a covariate to exhibit imbalance between the treatment and control cases and yet not be a source of bias. In order

[9]It is not sufficient since there may still be unobservable counfounds. It is not necessary since in the regression of Table 4 we control for 27 observables which (by construction) in the balance table we are not, and because for the effect to be biased it has to be that unobservables are correlated with the error term and with the fraction of PS recounted.

FIGURE A15. "Voto por Voto" recount protest (2006)

Note: Notes: Mexico City center.

to cause bias, a predetermined covariate would have to be omitted from the analysis, to cause the treatment, and to cause the outcome. Balance only probes the association between the covariate and the treatment. A causal interpretation for the analysis in Table 4 does not require that recounted and not-recounted precincts be the same in all respects.

Figure A16 provides an extensive set of "balance tests", probing balance on the a large battery of predetermined covariates. Specifically, we regress each predetermined covariate (Y_{pt}) on an indicator for whether or not at least one recount took place in the precinct, $Y_{pt} = \alpha + \beta I(\text{precinct p had recounts})_{pt} + \epsilon_{pt}$. An observation is a precinct. In order to have the same units across dependent variables in the figure, we normalize β by the standard deviation of the respective Y, and plot θ_Y defined as $\theta_Y = (\beta_Y/\sigma_Y)\sigma_X$, where X is the explanatory variable. The figure reports θ_Y and its confidence interval obtained by the delta method.

We find that differences across recounted and non-recounted precincts are substantively small in magnitude and half are statistically indistinguishable from zero. The modal imbalance is 1% of a standard deviation; the median is 3% (both in absolute value). Despite the fact that imbalances are remarkably small, in the body of the paper we show results adjusting for these covariates, and as explained in the main text, the effect of vote recounts is robust to the inclusion of the full set of covariates.

FIGURE A16. Comparing recounted vs not recounted precincts

Note: This figure reports balance test. Each dot is obtained by a bivariate regression of the respective variable (e.g. Years of Education) on an indicator of the precinct having at least one PS recounted: $Y_{pt} = \alpha + \beta I(\text{precinct p had recounts})_{pt} + \epsilon_{pt}$. Observations are at the precinct level, and therefore errors are not clustered. There are 27 balance tests reported. In order to have the same units across dependent variables in the figure, we normalize β by the standard deviation of the respective Y, as $\theta_Y = (\beta_Y/\sigma_Y)\sigma_X$, where X is the explanatory variable. The figure reports θ_Y and its confidence interval obtained by the delta method.

The effect of recounts on trust is larger in close elections:

One anonymous referee suggested the hypothesis that recounts are more likely to impact public trust in close elections, because the losing party should have greater incentives to publicize the existence of inconsistencies and recounts and to frame these as evidence of malfeasance (possible reasons for this include appeasing the politician's supporters, deflecting blame for the loss, and creating political pressure for recounting or anulling the results). To test for this hypothesis, we estimated the effect of recounts on trust as a function of the margin of victory. One implication of this hypothesis is that we should estimate a negative coefficient on the interaction between a dummy indicating whether the race was close and the presence of recounts (controlling also for the baseline effects, that is, for the dummy of close elections and the dummy for recounts) in a model analogous to equation 4 in the body of the paper.

Table A13 reports estimates of that regression. The difference between Parts A, B and C is the set of geographical fixed effects included (none, district, or municipality). The 5 different columns use different definitions for what counts as a close election: 1% margin between the 1st and 2nd place, 2% margin,..., up to 5% margin. The rest is analogous to Table 4 in the main text. The estimates lend strong support to the referee's prediction: the effect of recounts on trust in INE is larger in tighter races. This result is also consistent with our causal interpretation of our estimates in Table 4, and it provides evidence consistent with a specific mechanism for how recounts are translated to lower trust: via the efforts of politicians. Even though this new result does not settle the issue of causality, it certainly raises the bar for alternative stories. Any endogeneity story would now have to be consistent with these findings as well.

TABLE A13. The effect of recounts on trust is larger in close elections

Panel A: Trust in electoral authority (INE), no FE

	(1)	(2)	(3)	(4)	(5)
	1%	2%	3%	4%	5%
Recount	-0.060***	-0.057***	-0.058***	-0.058***	-0.060***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Margin	0.26	0.22*	0.046	-0.040	-0.056
	(0.16)	(0.12)	(0.07)	(0.06)	(0.05)
Recount*Margin	-0.28	-0.27**	-0.067	0.014	0.042
	(0.18)	(0.13)	(0.08)	(0.07)	(0.05)
Constant	3.46***	3.46***	3.46***	3.46***	3.46***
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
N	6186	6186	6186	6186	6186
R-sq	0.175	0.175	0.174	0.174	0.175
P value: full controls	0.000	0.000	0.000	0.000	0.000
Pct observations	0.70%	2.22%	4.47%	7.13%	9.97%

Panel B: Trust in electoral authority (INE), FE por distrito

	(1)	(2)	(3)	(4)	(5)
	1%	2%	3%	4%	5%
Recount	-0.039***	-0.035**	-0.036**	-0.035**	-0.037**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Margin	0.26	0.22*	0.043	-0.046	-0.061
	(0.16)	(0.12)	(0.07)	(0.06)	(0.05)
Recount*Margin	-0.31*	-0.29**	-0.079	0.0067	0.036
	(0.17)	(0.13)	(0.08)	(0.06)	(0.05)
Constant	3.59***	3.59***	3.58***	3.58***	3.58***
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
N	6186	6186	6186	6186	6186
R-sq	0.195	0.196	0.195	0.195	0.195
P value: full controls	0.000	0.000	0.000	0.000	0.000
Pct observations	0.70%	2.22%	4.47%	7.13%	9.97%

Panel C: Trust in electoral authority (INE), FE por municipio

	(1)	(2)	(3)	(4)	(5)
	1%	2%	3%	4%	5%
Recount	-0.032**	-0.028*	-0.032**	-0.032**	-0.033**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Margin	0.27*	0.23*	0.050	-0.043	-0.054
	(0.16)	(0.12)	(0.07)	(0.06)	(0.05)
Recount*Margin	-0.35**	-0.31**	-0.073	0.020	0.039
	(0.17)	(0.13)	(0.08)	(0.07)	(0.05)
Constant	3.56***	3.56***	3.55***	3.55***	3.55***
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
N	6186	6186	6186	6186	6186
R-sq	0.229	0.228	0.229	0.229	0.000
P value: full controls	0.000	0.000	0.000	0.000	0.000
Pct observations	0.70%	2.22%	4.47%	7.13%	9.97%

Note: This table reports estimates of a regression analogous to equation 4 but adding two regressors: a dummy for the margin being close, and the interaction of this dummy with the fraction of PS recounted in the precinct. The difference between Parts A, B, and C is the set of geographical fixed effects (none, district, municipality). The 5 different columns use different definitions for what counts as a close election: 1% margin between the 1st and 2nd place, 2% margin,..., up to 5% margin.

Further robustness checks: trust analysis

Table **??** is analogous to Table 4 with two main differences. Part A weights every observation by the number of surveys completed in the corresponding precinct. This gives greater weight to precincts with more information on the attitudes variables. Results are robust. Part B runs regressions at the polling station level. The explanatory variable in this case is a dummy variable indicating whether the PS was recounted (=1) or not (=0). The dependent variable continues to be the precinct-level attitudes measure (repeated for all PS within the precinct since there are too few surveys to construct this variable at the PS level). Note that to the extent that there are spillovers from one PS to the other within a precinct the coefficient estimates in the PS level regressions would be smaller. The effect of recounts is somewhat lower as expected, but not much.

Sensitivity analysis

In the paper we asses the robustness of the effect of recounts on trust to the magnitude of unobserved confounders. This methodology has been fruitfully and influentially applied by Altonji et al (2005) and by Oster (2019)[10] and consists of accepting the possibility that there are unobserved confounders, and asking how large would they have to be in order to generate (or eliminate) the posited causal effect. Even if after the large battery of controls the reader thinks there are still unobserved counfounders, it does not follow that there is no causal effect.

We use Oster's aproximation under the assumption of proportional selection (Oster 2019, 8):

$$\beta^* \approx \tilde{\beta} - \delta(\beta^o - \tilde{\beta}) \frac{R^{max} - \tilde{R}}{\tilde{R} - R^o} \tag{8}$$

where $\tilde{\beta}$ is the coefficient on recounts in the controlled regression (column 1 in Table 4), β^o is the coefficient on recounts in a bivariate regression of trust on recounts (=-0.097), $\tilde{R}$ and R^o respectively denote the R-squared associated with each of the regressions ($R^o = 0.015$), and R^{max} is set to $1.3\tilde{R}$ as recommended by Oster's calibration exercise. Setting $\beta^* = 0$ and solving for δ yields the result in the text.

The result from this sensitivity analysis is that unobserved variables would have to have at least 6.1 times more explanatory power than observables in order for all the effect to be driven by them –i.e. in order for there to be no causal effect. We just find this too large to be reasonable.[11] Therefore we think we there is a causal effect from recounts to trust.

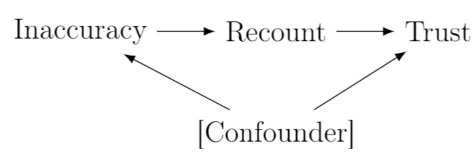
[10]Altonji, J., Elder, T., & Taber, C. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy, 113(1), 151-184, and Emily Oster, 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence", Journal of Business & Economic Statistics, vol. 37(2).

[11]Using the regression estimates that condition on the full set of controls (column 2 in Table 4) instead of the basic set of controls (column 1 in Table 4), the equivalent number is 5.2. As a point of comparison, in their seminal paper Altonji et al (2005) find effects that are *less* robust than ours. They find that "*selection on unobservables would need to be 3.55 times stronger than selection on observables in the case of high school graduation,... and 1.43 times stronger to explain the entire college effect" which they find "highly unlikely."*

Front-door adjustment

Finally we also implemented the front-door criterion to estimate the relationship between tally inaccuracies and trust in electoral authorities, without relying on the assumption that tally inaccuracies are exogenous. The front-door criterion formalizes the intuition that if a treatment causes an outcome via a mediator, it is possible to estimate the effect of the treatment on the outcome by breaking the causal chain into two segments: the effect of the treatment on the mediator, and the effect of the mediator on the outcome. The main assumption that is needed for this to work is that the mediator be unconfounded with the outcome.[12] The front-door criterion produces consistent estimates of the causal relationship even if inaccuracies and trust are confounded. Hence, if the IV estimates were endogenous due to an omitted variable (as suggested by Reviewer 1 and depicted in Figure A17), there would be no reason to expect the IV and front-door estimates to yield similar results. The following diagram illustrates the respective identifying assumptions of the IV and the front-door approaches.

FIGURE A17. Causal logic of IV vs. front-door approaches



The diagram depicts a structure where the instrument (tally inaccuracies) is endogenous because it is confounded with the outcome variable (trust in electoral authorities). This is the confounding scenario that concerns Reviewer 1. The front-door criterion is robust to this kind of confounding, as mentioned previously. To see why, note that stratifying on inaccuracies kills the causal effect of the confounder on recounts. The key assumption is that the confounder does not directly influence recounts.[13]

Following Pearl (1995), the front-door adjusted estimate of the causal effect of inconsistencies on

[12]For further details see Morgan and Winship (2015), pp.331-338.

[13]In contrast, the IV restriction requires that *inaccuracies*—not recounts— should be unrelated with the counfounder as depicted in the figure. Thus, the identifying assumptions for the IV vs. front-door approaches are different.

trust is:

$$\text{ATE} = \sum_m (P(rec = m|inc = 1) \sum_a (E[trust|inc = a, rec = m] P(inc = a)) - \sum_m (P(rec = m|inc = 0) \sum_a (E[trust|inc = a, rec = m] P(inc = a))$$

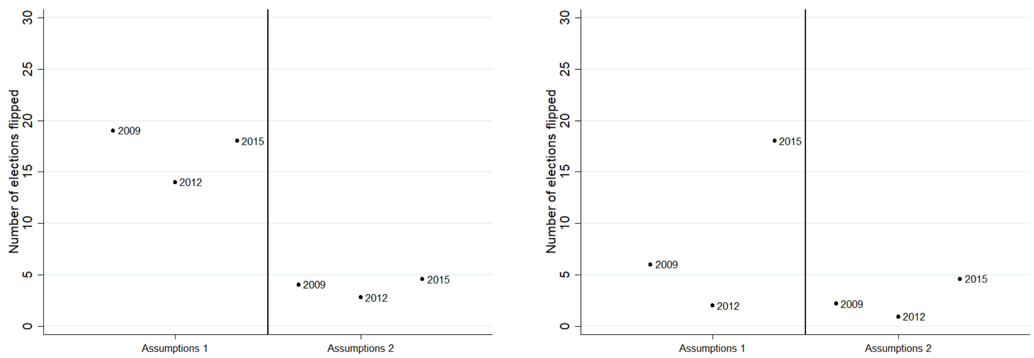
where inc stands for inconsistencies and rec for recounts. The validity of front-door adjustment relies on the alternative assumptions that: (a) the effect of inconsistencies on trust is fully mediated by recounts, and (b) recounts are unconfounded. More precisely, that the association between inconsistencies and recounts is unconfounded, and the association between recounts and trust is unconfounded after stratifying on inconsistencies. Glynn and Kashin (2018), moreover, show that the front-door adjustment can yield good results even when these assumptions do not hold exactly. Importantly, the front-door adjustment is valid even in the presence of unobserved common causes of inconsistencies and trust. Consistent with the rest of our analyses with trust in INE as the dependent variable, the front-door estimates are negative and statistically significant. This adds to our confidence that the negative relationship between tally inconsistencies and trust in INE is not simply the product of endogeneity.

BENCHMARKING THE CONSEQUENCES OF INCONSISTENCIES ON ELECTORAL OUTCOMES

Another consequence of mistakes is that that vote counts may misrepresent who was really elected to office, especially if the margin of victory is small. Here we simulate simple back of the envelope scenarios where the the number of vote inconsistencies are assigned to parties under different assumptions. This way we explore whether the extent of tallying mistakes might conceivably have flipped elections to the national legislature. In order to implement the simulations we are forced to make some (ad-hoc) assumptions. In particular, (a) we focus on type-3 inconsistencies, since this is the only type that involves votes assigned to parties; (b) we assume that the number of inconsistencies is translated into votes one-to-one; (c) we assume that the process generating inconsistencies is independent across PS; (d) in simulation 1 we allocate the inconsistencies of a PS as votes for the first runner up; (e) in simulation 2 we allocate inconsistencies to the winner and the runner up in the PS with a 50-50 chance. In both simulation exercises we focus on congressional elections in 2009, 2012, and 2015.

In the first simulation we find that inconsistencies are prevalent enough to flip up 19 out of 300 elections in 2009. That is, 19 electoral districts in 2009 were won by a margin smaller than the total sum of inconsistencies of type 3. We view this as a worst-case scenario, since in reality it is unlikely that all the non-partisan inconsistencies, were they to be corrected, would favor just one party (the first runner up). The second simulation is a bit more realistic, and randomly assigns inconsistency-3 related votes of each PS to either the winner or the first runner up with 50-50 probability. We repeat this procedure 100 times with a different randomization for each of the 900 elections and compute the probability that each of the 900 elections is flipped. With these probabilities in hand, we estimate the expected number of flipped elections in each election year. This number is 4, 3, and 5 for 2009, 2012, and 2015 respectively. Figure A18 reports these results. We take these simulations as suggestive that, in tight races, inconsistencies could violate a basic tenet of democratic elections, namely plurality rule. But as the previous two sections show, even when inconsistencies do not flip races they can yield equally worrisome outcomes, such as spurring recounts and reducing trust in the electoral system.

FIGURE A18. Simulation of Inconsistencies and Elections Flipped



(a) Inconsistency 3 **(b) Inconsistency 3 | Inconsistency 4 = 0**

Note: This figure shows simulations of the impact of inconsistencies in the number of elections of federal deputies election flipped, for 2009, 2012 and 2015 separately. There are two panels, both focus on type-3 inconsistency, but Panel (b) restrict the sample to Actas that had zero type-4 inconsistencies. Within each panel we present simulation 1 (left of the vertical line) and simulation 2 (right of the vertical line), which correspond to simulations under different assumptions as defined in section 13. Each dot represents the expected number that would be flipped under the simulation.