Appendix

A Additional quantitative evidence on Chávez's rhetoric **29**

A.1 Method for estimating the prevalence of key topics 29

A.2 Topic model results 30

B Details on Public Opinion Surveys **33**

C Correlation with past anti-system votes **36**

D Additional tables and figures **38**

E Scatter plots for electoral realignment result **41**

A Additional quantitative evidence on Chávez's rhetoric

In the main text, we estimate the prevalence of two topics—*institutional change* and *poverty*—in campaign-trail interviews of the two principal candidates in the 1998 presidential election (Hugo Chávez and Henrique Salas Römer). In this appendix, we: (1) provide details of the method and (2) estimate a topic model to help evaluate whether our approach misses other important topics.

A.1 Method for estimating the prevalence of key topics

We propose a method that leverages pre-trained word embeddings to guide the construction of topic-specific dictionaries. Using embeddings trained on huge collections of Spanish-language text,[22] we learn which words tend to co-occur with two key seed words—"poverty" and "constituent assembly" *constituyente*)—we then use this information to build a dictionary of words associated with each topic. The method is as follows:

1. Given a vocabulary V, construct a weighted lexical graph wherein links are weighted by the cosine distance between words on the (pre-trained) embedding space. In our case, the vocabulary V comprises all words that appear in our corpus –interview transcripts– minus a standard set of stopwords.

2. For a given seed s (or set of seeds) that define the topic of interest, propagate a 'topic label' using the random walk algorithm proposed by Zhou et al. (2004). This method yields a 'topical relevance score' for every word $w \in V$ equivalent to the probability that a random walk initiated at the seed word lands on w. We use the seed *poverty* for the economic policy topic and *constituyente* (constituent assembly) for the institutional change topic.

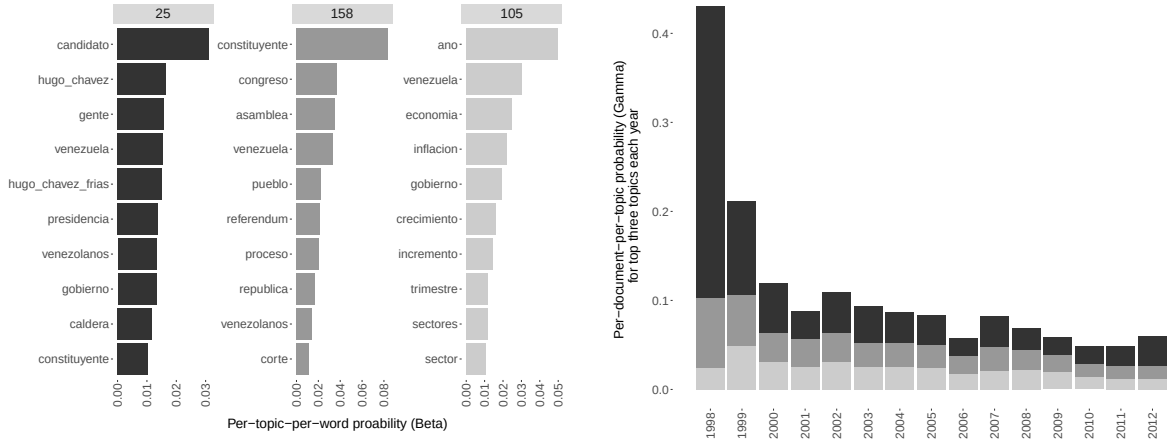
3. To approximate a measure of score uncertainty, iterate this process for N randomly selected seeds (we use $N = 100$). For each word $w \in V$, there are now 101 topical relevance scores: one for the seed s (*poverty* or *constituyente*), and 100 for the randomly selected comparison words. Candidate words for the dictionary are those words w that have a higher topical relevance score for the seed s than for some threshold proportion of the randomly

[22] Word embeddings are dense vector representations of words learned from local co-occurrence statistics in huge collections of text. Unlike traditional distributional semantic models, the co-occurrence statistics used to train embedding models come from small—usually symmetric—windows of text around each word (see Spirling and Rodriguez (2019) for a useful introductory discussion). They have been shown to be capture well 'human' semantics (Mikolov et al., 2013; Pennington et al., 2014).

Figure A.5: Top Themes According to Topic Model

Fig. (a) lists the words associated with the top-three most-prevalent topics in Chávez's speech in the year 1998; Fig. (b) plots the prevalence of these three topics in Chávez's speech over time.

(a) Top Three Topics in 1998 (b) Weight of Top 3 Topics



selected words; we use 0.95.

4. Have human coders validate candidate words for inclusion in the dictionary.[23]

5. If necessary (e.g., in the case that there are too few candidate terms) repeat the process using the set of validated words as seeds.

For our pre-trained embeddings, we use `GloVe`, trained on the Spanish Billion Word Corpus (Cardellino, 2019). These embeddings have been shown to correlate highly with embeddings trained on a subset of political texts (Spirling and Rodriguez, 2019).

A.2 Topic model results

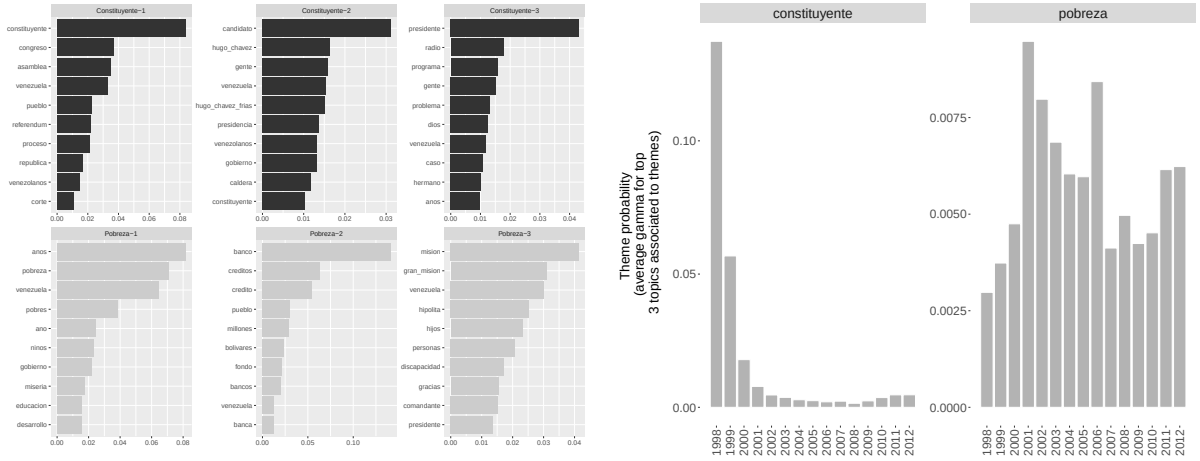
By restricting our analysis to two themes—institutional change and economic policy—we may miss other important topics. To evaluate this possibility, we use the full corpus of Chávez's speeches and interviews to estimate a topic model

[23]We validated and selected the top 20 words for each topic. The selected words for the *institutional change* topic are: constituyente, constitucion, constitucional, legislativa, constituyentes, referendum, legislativo, democraticamente, congreso, sufragio, electo, reelecto, electos, convocar, suprema, senado, independentista, senadores, parlamentaria, organica. The selected words for the *poverty* topic are: pobreza, extrema, pobres, erradicar, desempleo, inseguridad, milenio, mortalidad, marginalizacion, desarrollo, viven, injusticia, reconociendo, reduccion, globalizacion, combatir, ignorancia, violencia, economica, metas.

Figure A.6: Institutional Change Dominates Election; Economy Dominates Later

Fig. (a) lists the words associated with the top-three most-prevalent topics in Chávez's speech in the year 1998; Fig. (b) plots the prevalence of these three topics in Chávez's speech over time.

(a) Top Topics by Theme (b) Theme Prevalence Over Time



(Blei et al., 2003). The model has 165 topics, the number that maximizes the pairwise information divergence across all topics (Deveaud et al., 2014). Figure A.5a plots the 10 terms associated with the three most-prevalent topics in 1998. We might label the first topic *election*, with mentions of Hugo Chávez (then candidate), Caldera (then incumbent), Venezuelans, and *presidency*; the second topic is clearly the constituent assembly (*constituyente*; and the third topic appears to be the economy, with terms like *economy*, *inflation*, and *growth*. Figure A.5b plots the prevalence of these three topics over time, showing that they collectively accounted for more than 40% of Chávez's words in 1998. Confirming conventional wisdom, this analysis suggests that the 1998 campaign did not focus on a topic other than those we consider in our primary analysis.

We also use the topic model to check our intuition that Chávez campaigned on institutional change but then turned his focus to economic policy in later years, perhaps consistent with the vote-choice analysis of Section 2.2 in the main text. To do so, we estimate the prevalence of the *institutional change* theme by combining the three (of 165) topics that assign the highest probability to the term *constituyente* (constituent assembly); we estimate the prevalence of the *economic policy* theme by combing the three (of 165) topics that assign the highest probability to the term *poverty*. Figure A.6a lists the terms associated with each of the six selected topics; the vast majority of the terms appear clearly linked to the two themes of interest. Figure A.6b plots the prevalence of these two themes over time. Consistent with our claims, we observe that the *institutional change* theme had

a very high prevalence during the campaign (nearly 15% of words were generated from this theme) but declined quickly over time; indeed, Chávez talked about the constituyente in 1998 (during the campaign) than in 1999, it actually took place. The *economy* theme, in contrast, tripled in prevalence during Chávez's first years in office.

In our view, these results contradict the notion that Chávez merely exploited class cleavages or left-right polarization in order to get away with dismantling checks and balances. Instead, Chávez's speeches are consistent with the proposal that the 1998 election was less a referendum on economic policy than "a referendum on [rewriting the constitution]" (Handlin, 2017, 88).

B Details on Public Opinion Surveys

Table B.1 reports educational attainment in the Venezuelan adult population (ages 18+). Over the decades we study, the proportion of adults who did not finish primary school (6th grade) fell from the vast majority in the 1950s–60s to approximately 40% by the early 1980s and less than 10% by the 2000s.[24] These figures are consistent with previously reported changes in adult literacy (Ortega and Rodríguez, 2008). The proportion of adults with a college degree also increased during this period, from approximately 1% in the 1950s–60s to more than 10% by 2001.[25]

Table B.1: Educational Attainment in the Venezuelan Population, Ages 18+

This table compares educational attainment for the 18+ population as measured in two sources: decennial censuses, and the semi-annual national household survey (analogous to the U.S. Current Population Survey). The latter began in 1967, but is only available as of 1975.

Censuses					Household Surveys				
Year	Less than Primary	Primary	High School	College	Year	Less than Primary	Primary	High School	College
1961[1]	87%	10%	2%	1%					
1971[2]					1975	48%	39%	8%	2%
1981	44%	42%	12%	2%	1981	36%	45%	12%	3%
1990	19%	59%	13%	9%	1990	26%	48%	19%	5%
2001	9%	51%	28%	12%	2001	8%	60%	21%	11%
2011	5%	24%	47%	20%					

[1]The printed volumes of the 1961 census only report attainment for the population 25+, not 18+; the 18+ population was likely more educated.
[2] The 1971 census did not measure attainment, due to an error in the questionnaire.

The public opinion surveys capture these changes. Table B.2 reports educational attainment among respondents in the public opinion surveys; as in the population, the share with less-than-primary education falls from more than 40% in the 1970s to 20% by the early 1990s and then less than 10% by the 2000s (the 1983 survey appears to slightly over-represent educated respondents). And as in the population, the share with college degrees increases from 1% to more than 10%.

The weights included in the 1993–2006 surveys actually render the sample *less* representative of the true distribution of educational attainment, which is why we report unweighted results in the main text (though applying the weights makes

[24]Neither the 1950 census nor the 1961 census clearly reported attainment in the 18+ population, but the 25+ numbers from the 1961 census strongly suggest that a majority of adults 18 had not completed primary school.

[25]The 2011 census figure of 20% is likely overstated, and the household surveys are not publicly available after 2006.

Table B.2: Educational Attainment in Public Opinion Surveys

This table reports educational attainment as recorded in the nine public opinion surveys analyzed in the main text, for comparison with the population proportions in Table B.1.

	Unweighted				Weighted			
Year	Less than Primary	Primary	High School	College	Less than Primary	Primary	High School	College
1973	41%	41%	17%	1%	·	·	·	·
1983	20%	49%	25%	6%	·	·	·	·
1988	22%	47%	25%	5%	·	·	·	·
1993	19%	52%	22%	6%	17%	47%	20%	16%
1998	12%	41%	24%	23%	19%	46%	20%	16%
2000	9%	42%	34%	15%	20%	43%	20%	17%
2004	14%	32%	37%	17%	16%	40%	22%	22%
2006	13%	50%	26%	12%	16%	41%	23%	20%
2012	6%	35%	37%	21%	6%	35%	37%	21%

little difference, as we show in this appendix). Ideally, we would weight the samples not to make them nationally representative but rather to make the college graduates in the sample representative of college graduates in the population and to make the least-educated in the sample representative of the least-educated in the population. However, there are many characteristics one could potentially target and no theory-driven approach to selecting among them.

In Figure 3 in the main text, we report the vote-intention gap between respondents with a college degree and respondents who did not graduate from primary school—except for the first four elections (1958–1973), for which we pool high-school and college graduates because the latter make up such a tiny portion of the sample (see Tables B.1 and B.2). In the second row Table B.3, we instead consider college graduates only for these four elections (despite the small sample size). The results are qualitatively similar: for all four elections, there is a large gap between the fractions of the most- and least-educated who voted for AD. In the third row of Table B.3, we use survey weights where available (namely, for 1993–2006). The weights make little difference: there remains a large vote-intention gap in 1993 and 2006 but not in 1998 or 2000 (if anything, the gap in 1998 is in the opposite direction, with more-educated voters *more* likely to state that they intend to vote for Chávez).

Table B.3: Education–Voting Gradient, Alternative Specifications

This table reports the difference visualized in Figure 3 in the main text (i.e., $\widehat{\theta}_t$ from Equation 1), compared against two alternative specifications. The first alternative specification compares college graduates to respondents with less-than-primary-school education even for years <= 1973, rather than pooling high school and college graduates for these years. The second alternative specification applies survey weights where available (namely, for the 1993–06 surveys).

Specification	'58	'63	'68	'73	'78	'83	'88	'93	'98	'00	'06	'12
Reported in paper	-34.4	-30.6	-16.8	-22.2	-8	-13.7	-13.7	-9.6	2.9	-2.2	-17.5	-24.5
College only for $<=' 73$	-45.1	-22.2	-15	-34								
W/ survey weights								-9.6	6	-1	-17.9	

C Correlation with past anti-system votes

If, as we and Handlin (2017) propose, Chávez's 1998 coalition drew voters from the left and the (much smaller) right who sought wholesale institutional change, we would expect that his vote share would correlate with those of previous anti-system challengers—even right-wing challengers. If, on the other hand, Chávez won in 1998 primarily by taking up the mantle of the neglected left (as Ellner (2003), among others, claim), we would expect his vote share to correlate with those of left challengers (whether pro- or anti-system); we would *not* expect that his performance would resemble that of right-wing anti-system parties.

Venezuela's 1968 congressional election provides an opportunity to evaluate these predictions. That contest featured both an anti-system far-right party—the Cruzada Cívica Nacionalista, or CCN—and a pro-system far-left party, the Movimiento Electoral del Pueblo, or MEP, a splinter faction that had separated from AD the previous year (1967).[26]

CCN and MEP held radically different visions for Venezuela. CCN was the vehicle of former military dictator Marcos Pérez Jimenez, who controlled Venezuela from 1948 through January 1958. He had been convicted of profiting from public office just four months before the election; during his trial, Pérez Jimenez publicly extolled the virtues of his dictatorship.[27] According to one historian who consulted the trial documents, "it appeared to the ex-dictator that dictators, even when they misused funds, accomplished more than democrats who did not steal" (Ewell, 1977, 312). In other words, CCN was unabashedly anti-system. MEP, on the other hand, had split from AD the previous year primarily because of conflict over policy issues.

CCN won 11% of the vote, earning four seats in the Senate and 21 in Congress, the fourth-best performance after AD, Copei, and MEP; MEP won 13% of the vote, five senators, and 25 Congressional representatives.[28] The court ultimately barred Pérez Jimenez from becoming a Senator, but other CCN politicians took their seats.

As it turns out, many districts that voted for CCN in 1968 also voted for Chávez in 1998. In fact, as Table C.1 reveals, Chávez's 1998 vote share is more correlated with CCN's 1968 vote share than with MEP's. In our view, this suggests that the same types of voters—perhaps some of the same individuals—who supported

[26]The labels "far-right" and "far-left" for CCN and MEP, respectively, come from Baloyra and Martz (1979); they also concord with our subjective assessment.

[27]Pérez Jimenez was sentenced to less time than he had already served, so at the time of the election he was living in exile in Spain.

[28]Both houses of congress were elected using closed-list proportional representation.

the far-right authoritarian CCN 1968 also supported Chávez in 1998; indeed, Velasco (2015) documents as much in his study of one Caracas neighborhood. We interpret this finding as evidence in favor of the notion that many voters were drawn to Chávez in 1998 because of their preference for institutional change, not exclusively because of their identification with the left or their preference for specific distributional policies.

Table C.1: Left *and* Right Anti-System Votes Predict 1998 Chávez Vote

ρ reports the bivariate correlation between (a) Chávez's vote share in 1998 and (b) the vote share (or log vote share)† of the party listed in the first column, in the election listed in the second column.

Party	Election	Ideology†	ρ
MEP	1968, Congressional	Left	0.15
CCN	1968, Congressional	Far right	0.24
Convergencia	1993, Presidential	Center-left	0.33
LCR	1993, Presidential	Left	0.47

†We take the log of CCN and MEP vote shares because the relationship between the raw shares and Chávez's share is nonlinear. There are no municipalities with zero MEP votes and four with zero CCN votes; for these, we take ln(0.002).

††Baloyra and Martz (1979, 118) estimated the ideological placement of MEP and CCN in the 1960s and 1970s; the placement of Convergencia and LCR in 1993 is based on our subjective assessment, as discussed in the main text.

The 1993 presidential election is also instructive. That contest featured Andrés Velásquez, a candidate with several striking similarities to Hugo Chávez. Like Chávez, Velásquez did not come from an elite family: he entered politics through involvement with the union at SIDOR, the state steel corporation, where he had been an electrical worker; during the presidential campaign, journalists even asked whether he knew how to wear a tie (López-Maya, 1994). Like Chávez, Velásquez was a political outsider; he had been governor of the state of Bolívar, but he had never held national office. Like Chávez, Velásquez led a (previously) small left-wing party called La Causa R (R for radical). Like Chávez, Velásquez's party proposed a *constituyente* to rewrite the Venezuelan constitution. In the 1993 presidential contest, he earned 22% of the vote. It was the best performance of any third party since 1968, but it was not enough to win. Table C.1 reveals that Velásquez's 1993 vote share is highly correlated with Chávez's vote share five years later.

D Additional tables and figures

Figure D.1: The AD and Post-2006 Chávez Coalitions were Rural

Using an original panel data set of municipal election returns, this figure plots the bivariate correlation (across municipalities) between (i) AD's or Chávez's vote share and (ii) population density in each year indicated on the x-axis.

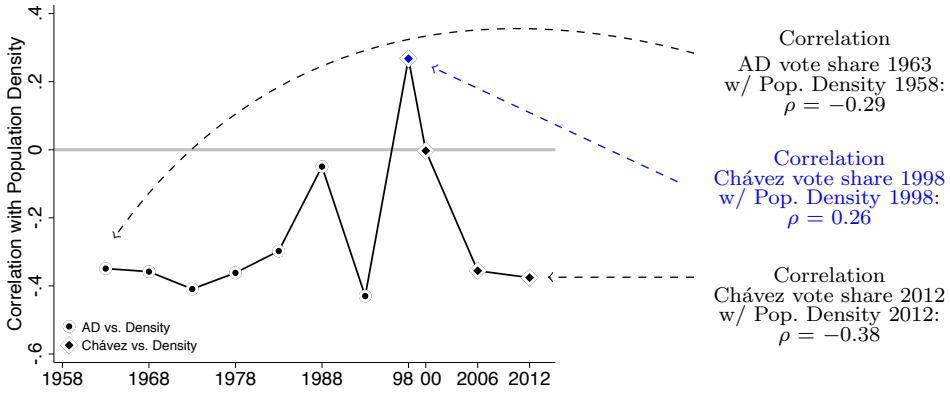
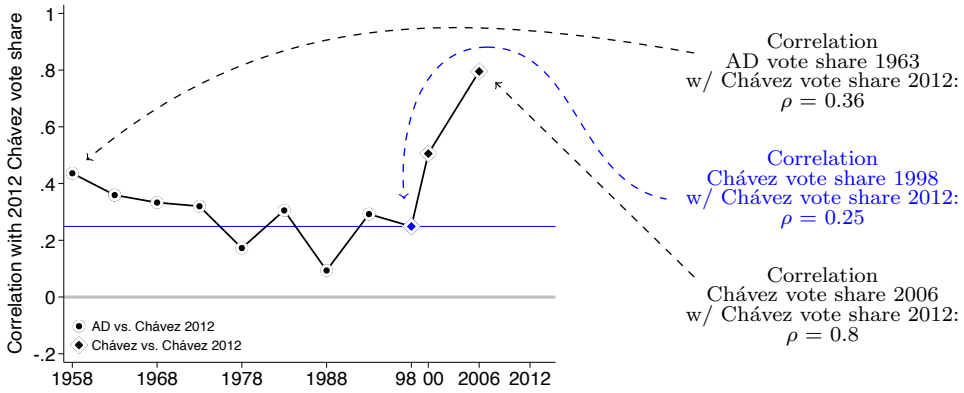


Figure D.2: Alternate Views of Inter-Election Correlations

This figure presents bivariate correlations analogous to those of Figure 1 in the main text, but using 2012 (top figure) or 1998 (bottom) rather than 1958 as the base comparison year. The takeaway is similar. The top figure shows that Chávez's 2012 vote share us more correlated with many historical AD vote shares than with Chávez's own 1998 vote. The bottom figure reveals that Chávez's inter-election correlations deteriorated more in 14 years (to $\rho = 0.25$) than AD's did in 35 years (per Figure 1, AD's 1958–1993 correlation was $\rho = 0.56$).

(a) Base Year 2012



(b) Base Year 1998

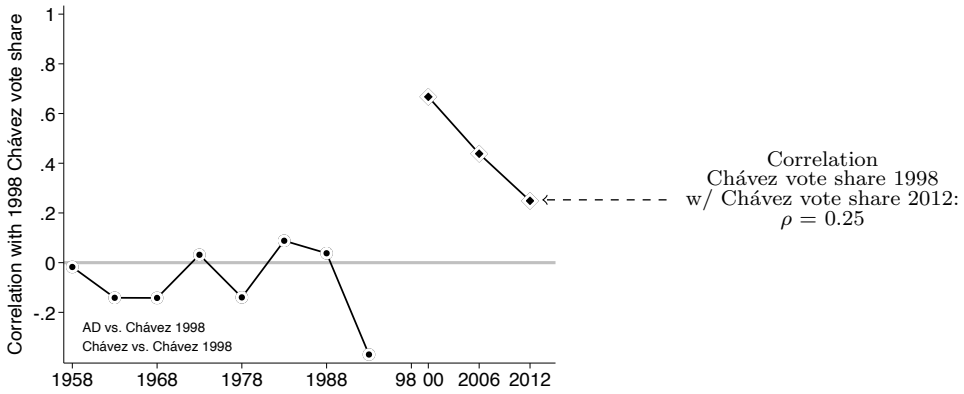
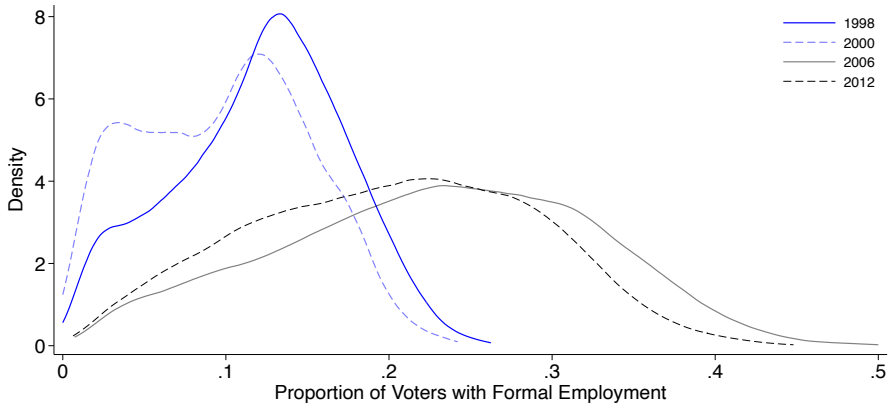


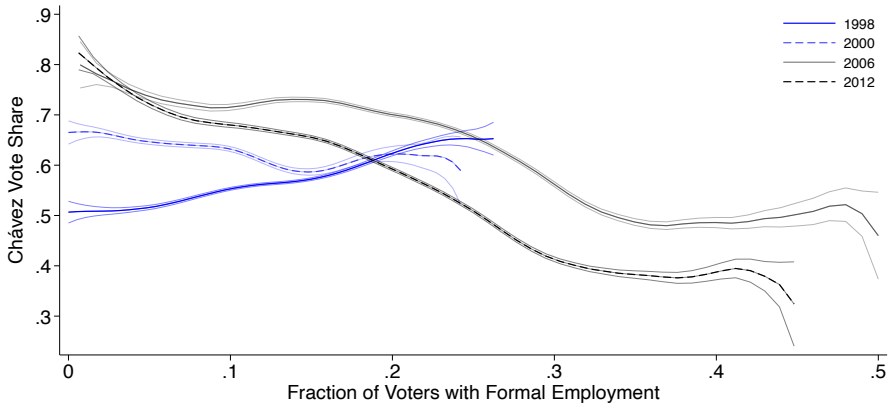
Figure D.3: Alternate View of Voting-SES Gradient

Using voting-booth-level data, Figure 4 in the main text plots Chávez's vote share in each of four presidential elections against each voting booth's *percentile* in the distribution of the fraction of voters with formal employment. Here we present those distributions (1), as well as the vote choice–SES gradient, but without transforming the x-axis (b).

(a) Distribution of % of Voters with Formal Employment in Each Election Year



(b) Vote Choice – % Formal Gradient in Each Election Year

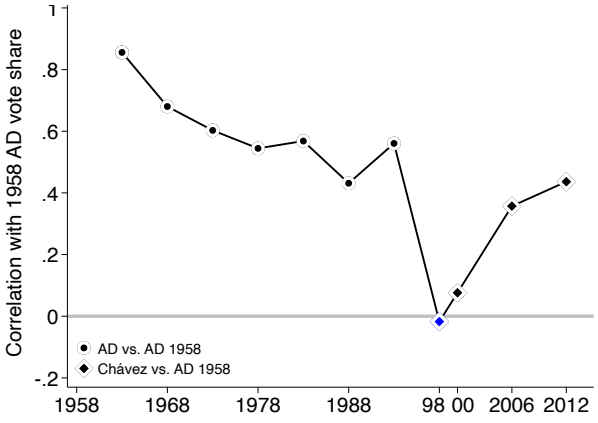


These figures exclude the top and bottom one-tenth of one percentile.

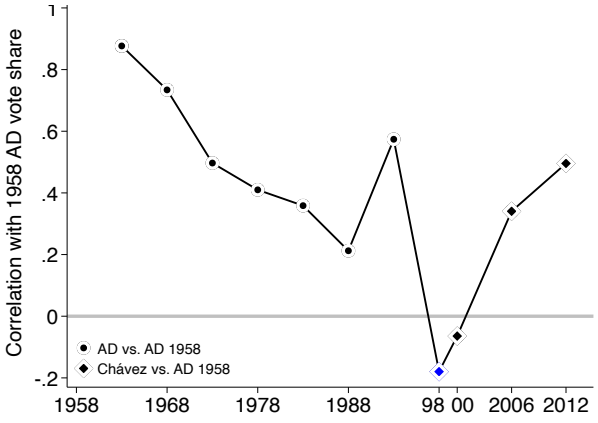
E Scatter plots for electoral realignment result

Figure E.4: Chávez's Election Scrambled Venezuela's Electoral Map

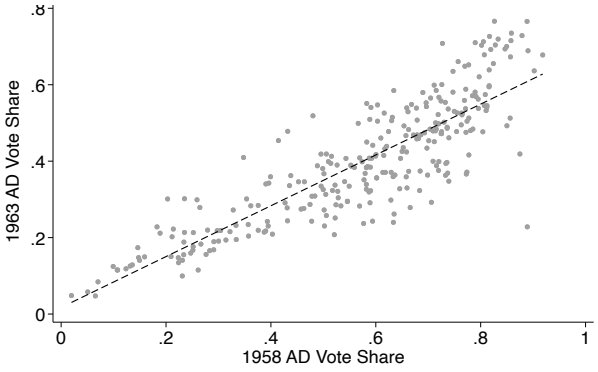
(a) Correlation Between AD or Chávez ('98–) Vote Share and 1958 AD Vote Share



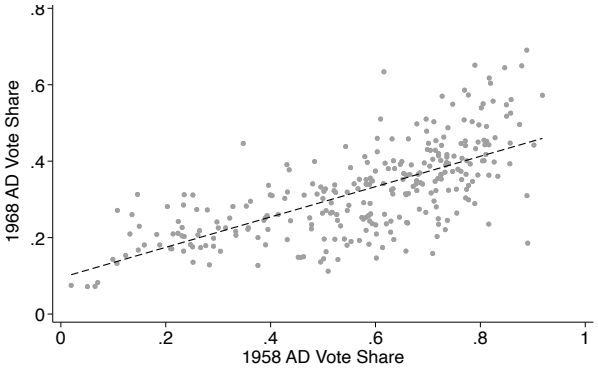
(b) **State-Level** Correlation with 1958 AD Vote Share



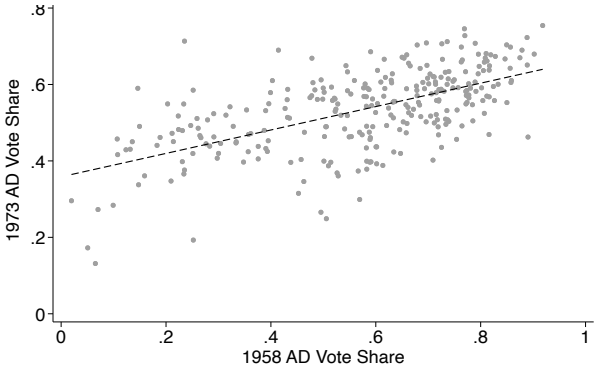
(c) 1963 v. 1958



(d) 1968 v. 1958



(e) 1973 v. 1958



(f) 1978 v. 1958

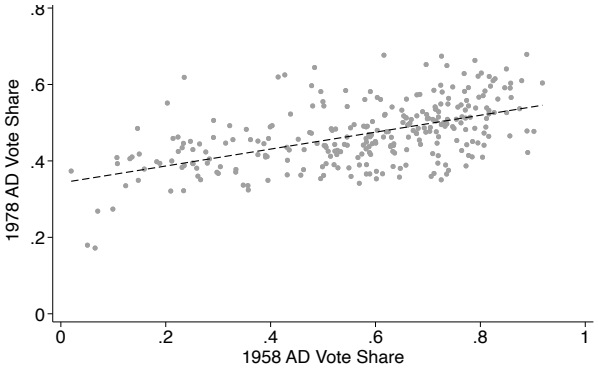
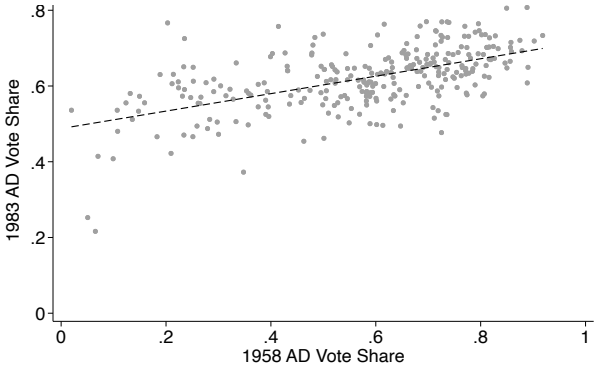
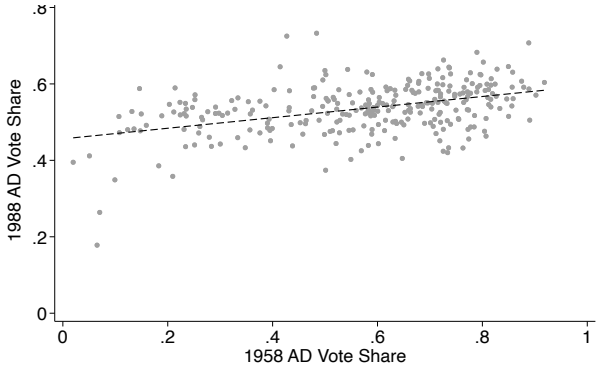


Figure C.2 Con't: Chávez's Election Scrambled Venezuela's Electoral Map

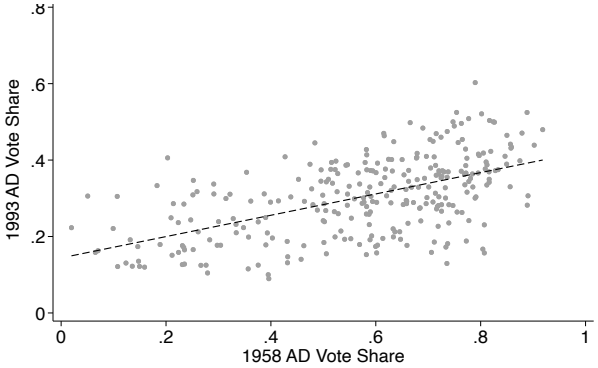
(g) 1983 v. 1958



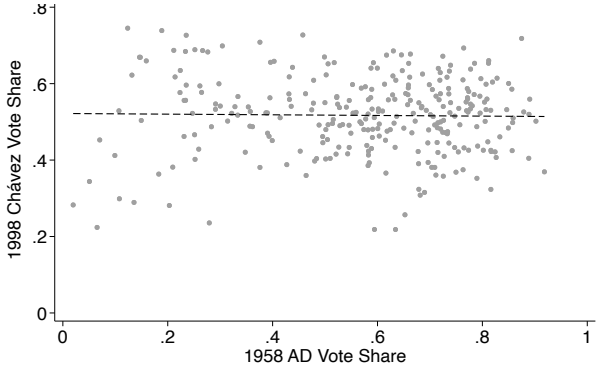
(h) 1988 v. 1958



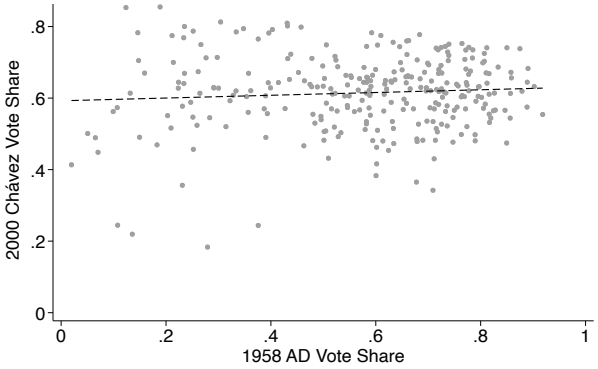
(i) 1993 v. 1958



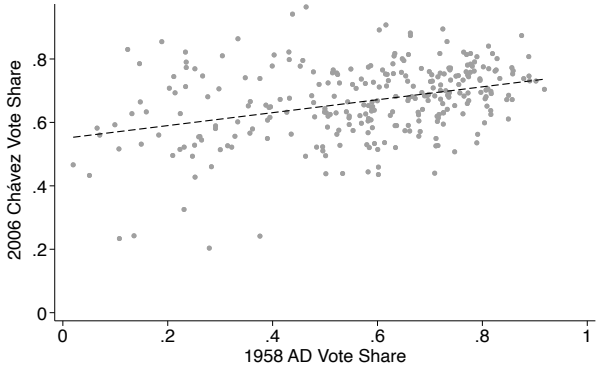
(j) 1998 v. 1958



(k) 2000 v. 1958



(l) 2006 v. 1958



(m) 2012 v. 1958