The Geography of Electoral History: A Dataset of Recent Mexican Election Returns and Quantities of Analytical Interest

Eric Magar
ITAM

January 1, 2025

The advent of competition in Mexican politics produced a wealth of government data for the analysis of public policy and politics. Data in distributed at the municipal level and smaller geographic units of aggregation (such as census tracts or similar levels), and in some cases at the individual level. It has sprawned fertile areas of new research in education (Hoyos, Espino and García 2012), public health (Imai, King and Nall 2009; King, Gakidou, Ravishankar, Moore, Lakin, Vargas, Rojo, Ávila, Ávila and Llamas 2007), poverty relief (Díaz Cayeros, Estévez and Magaloni 2016; Molinar Horcasitas and Weldon 1994), legislative politics (Cantú, Desposato and Magar 2014; Rosas and Langston 2011), and electoral regulation (Estévez, Magar and Rosas 2008), to name a few.

This paper's focus are vote returns. Electoral data has been distributed for much longer than the information discussed above. It is also better-known and havs received a good deal of attention since Ames' seminal study of the PRI's support bases in municipalities (Ames 1970). This paper describes a repository...

Some of the distributed data is elementary and available elsewhere, such as the number of valid votes cast for parties in congressional races since 1991 in single-member districts, in municipalities, and in sub-municipal units of aggregation. shares by party and their change since last election) at the municipal and sub-municipal levels. Offers a cross-section time-series of dip fed returns at two levels of aggregation: municipalities and secciones electorales.

More abstract

from blog

This note presents, discusses, and distributes statistics (available here) of party performance in Mexico's competitive era. I elaborate two quantities of interest: *voting forecasts* based on recent electoral history and measures of parties' *core support*. The procedure produces summary measures of recent electoral history in relatively small geographic units, municipalities ($N \approx 2500$) and /secciones electorales/ ($N \approx 66000$) throughout Mexico. I apply the methodology to four federal congressional elections between 2009 and 2018 (I will soon apply it to municipal races too), using results since 1994 as historical input.

The note starts by showing the statistics in action to get a glimpse of their descriptive and analytic potential. By summarizing recent electoral history and its geography, the quantities offer

a scenic view of a critical aspect of contemporary Mexican politics.

Later sections offer methodological detail on the estimation of these quantities of interest and are increasingly technical.

1 The data

	State (abbreviation)	Number of municipios	Years
1	Aguascalientes (ags)	9–11	1977–2024
2	Baja California (bc)	4–7	1971–2024
3	Baja California Sur (bcs)	3–5	1974–2024
4	Campeche (cam)	8–13	1979–2024
5	Coahuila (coa)	38	1978–2024
6	Colima (col)	10	1976–2024
7	Chiapas† (cps)	110–126	1976–2024
8	Chihuahua (cua)	67–68	1974–2024
9	Distrito Federal/Mexico City‡ (df)	16	1997–2024
10	Durango (dgo)	38–39	1971–2024
11	Guanajuato (gua)	46	1979–2024
12	Guerrero† (gue)	75–85	1977–2024
13	Hidalgo (hgo)	84	1981–2024
14	Jalisco (jal)	124–125	1976–2024
15	México (mex)	121–125	1978–2024
16	Michoacán† (mic)	113	1977–2024
17	Morelos† (mor)	33–36	1976–2024
18	Nayarit (nay)	19–20	1972–2024
19	Nuevo León (nl)	51	1973–2024
20	Oaxaca† (oax)	570	1977–2024
21	Puebla (pue)	217	1980–2024
22	Querétaro (que)	18	1973–2024
23	Quintana Roo (qui)	7–11	1978–2024
24	San Luis Potosí (san)	56–58	1970–2024
25	Sinaloa (sin)	17–20	1971–2024
26	Sonora (son)	69–72	1976–2024
27	Tabasco (tab)	17	1976–2024
28	Tamaulipas (tam)	43	1971–2024
29	Tlaxcala (tla)	44-60	1979–2024
30	Veracruz (ver)	203–212	1976–2024
31	Yucatán (yuc)	106	1981–2024
32	Zacatecas (zac)	56–58	1970–2024
	Total by election cycle		

Notes: † Administrative jurisdictions in the Federal District became elected offices since 1997. † Reform in these states withdrew so-called 'usos y costumbres' municipalities from the periodic electoral process: one municipality in Chiapas since 2021, one in Guerrero since 2018 and another since 2024, one in Michoacán since 2011, three in Morelos since 2021, and between 412 and 418 in Oaxaca since 1995.

2 Statistics in action

Table [[fig:1]] presents two pairs of diagrams for the 2015 (below) and 2018 (above) congressional elections. Each dot represents one municipality, colored according to the winning party, with coordinates in the ternary plot according to the relative votes of the PAN, the PRI, and the left in the federal deputy race (other smaller parties are excluded).[1]

The left side shows the /vote forecasts/. The idea behind this statistic is summarizing the evolution of relative votes in the municipality in five previous elections (2003–2015 in the case of 2018) and using the tendency to project a vote forecast for the current year. Plots in the right side show the actual results observed in both elections.

Three features are noteworthy in 2018. The first is the discrepancy between the left and right plots. Either the model does a poor job forecasting, or 2018 was an extraordinary election. History gave license to expect a comfortable PRI victory, both in the number of municipalities won and in margins of victory. Municipalities outside the dotted bands are won by margins of 15 points or more, and the bulk of secure municipalities are red in the forecast, with Morena in a distant second place. In fact, although a significant number of municipalities migrated towards the PAN, it was Morena who showed a clear advantage. PRI was the underachiever. In contrast, the lower left and right plots reveal fewer differences between them—2015 was a more normal election, the past offering much better grounds to forecast.

Second, observed municipalities fled the edges and triangle vertices in 2018. Observations in vertices show a party that has no significant challenger. While those on the edges were bipartisan, whether more (inside dotted bands) or less (outside the bands) symmetric. It is also plain in forecasts that only the PAN–Morena edge was expected to be unpopulated. In practice, however, third party vote rarely collapsed to zero, there was much more dis-coordination than in the past. In fact, the intersection of dotted bands appeared denser and more homogeneous in the right than the left diagram.

Third, the PAN /vs/ left competition was legal tender in 2018. The pattern in competitive municipalities in the last two decades, still visible in the 2015 plots, involved either PRI–PAN or PRI–left rivalries, and rarely PAN–izquierda. It was this pattern that eased electoral alliances between PAN and PRD in sub national races since 2010 that culminated in the Frente they formed in 2018 to nominate a joint presidential candidate.

+CAPTION: Una elección más característica de la partidocracia +NAME: fig:2 | file:../assets/img/triplot2015-vhat-mu.png | [[file:../assets/img/triplot2015-v-mu.png]] |

+CAPTION: Grano más fino: las secciones +NAME: fig:3 | file:../assets/img/triplot2015-v-se.png | [[file:../assets/img/triplot2018-v-se.png]] |

Plots in Table [[fig:4]] report /secciones electorales/ and therefore offer much finer-grained than the previous portraits. They introduce the other quantity of interest in this note: parties' /core support/. The idea behind this statistic is measuring the size of the group that has historically supported the party consistently, in good but also in bad years.

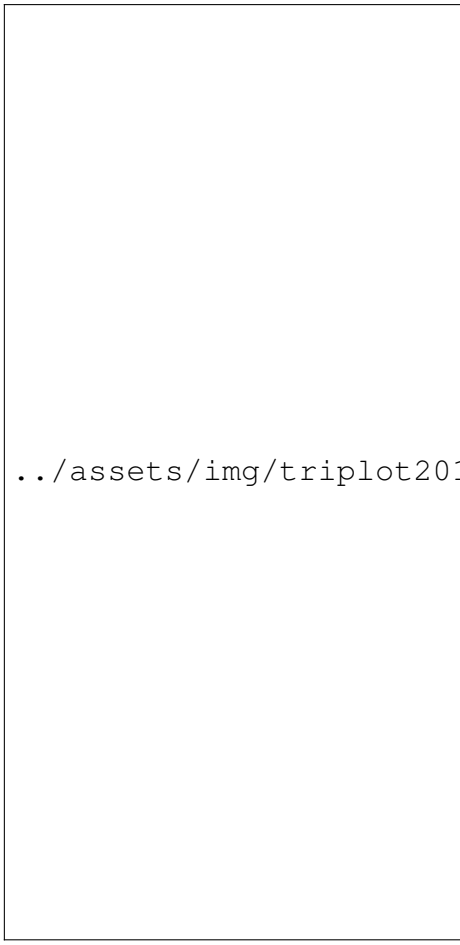
+CAPTION: Support core and party performance in 2018 +NAME: fig:4 | file:../assets/img/resid-pan-2018-vs-pan-core-se.png | [[file:../assets/img/resid-morena-2018-vs-morena-core-se.png]] | file:../assets/img/resid-pri-2018-vs-pri-core-se.png |

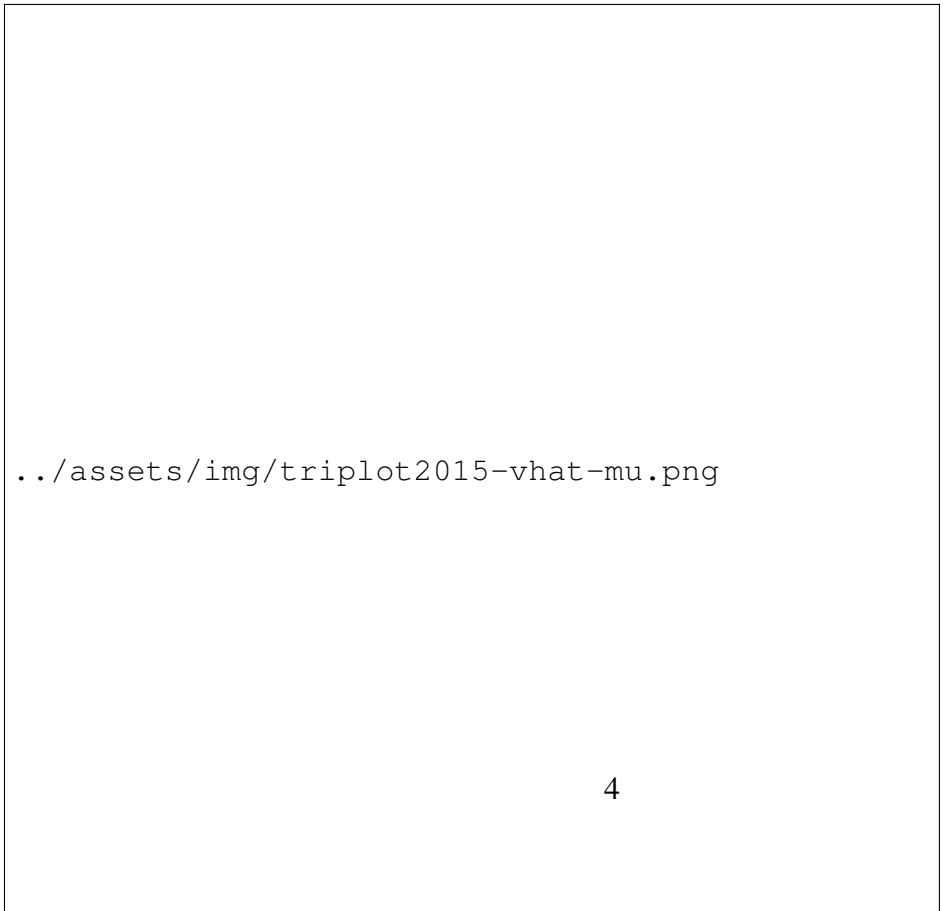
The horizontal axis in each plot measures the size of party core support groups as a proportion of the /sección/'s electorate. The PRI enjoyed a clear edge over the rest of the parties in the period, with sizable cores in /all/ secciones nationwide. The distributions of the PAN and the left, in contrast, appear concentrated towards the zero—they have relatively few secciones with some unconditional support.

The vertical axis reports the three parties' performance in 2018 (the difference between the observed vote and the forecast). Positive values indicate that the party excelled expectations in the /sección/, negative ones that it underperformed it expectation. The PRI's electoral disaster appears all too clearly in the red plot. There were a few secciones with positive performance, but the density concentrates massively below the horizontal zero line, something that Table [[fig:1]] had hinted. What is truly remarkable is that the dismal performance is directly proportional to the size of the PRI core. President Peña and candidate Meade achieved what seemed, if not impossible, extremely improbable: they alienated the PRI's unconditional voters in 2018. The PAN and the left met expectations in secciones where they have enjoyed with groups of support. Both (especially Morena) over-achieved where they lack important cores, taking away PRI voters.

[1] I must note that, for the left's electoral history (which I arbitrarily call "Morena" in the plots and distributed data), I systematically added up the votes of the PRD, PT, and MC up to 2015. I also added Morena's and PES's votes that year. In 2018 the left consisted of Morena, PT, and PES jointly.

Figure 1: Historic expectations and municipal-level vote in two congressional elections





The note elaborates how statistics were prepared (replication code can be found [[https://github.com/emagar/mxDistritos/code/elec-data-for-maps.r][here]]).

[[file:https://github.com/emagar/elecRetrns/raw/master/graph/nytAmloPlusAnayaPlusMeadeNegPenaWon.svg]]

+CAPTION: PAN +NAME: fig:6 $+ATTR_HTML: style = "float: right;" + ATTR_HTML :: width 50 [[file: ../assets/img/resid-pan-2018-vs-pan-core-se.png]]$

* First differences One common approach to study electoral change is through first differences. Denoting v_t the party's vote share in the municipality or the /sección/, in period t, the first difference is simply $d_t = v_t - v_{t-1}$.

d_t is an intuitive quantity, showing the sign and magnitude of change from one election to the next. But, precisely because it compares pairs of consecutive elections only, it misses more dynamic processes in the units. One example, well documented by electoral sociology, is the regression to the mean (Campbell 1991; Segovia 1979). Its detection requires observing the unit through at least three consecutive periods to verify contrary signs in d_{t+1} and d_t. The study of secular change in the Mexican party system in the last quarter of century calls for deeper historical perspective.

(First differences appear in the fields d.pan , d.pri , and d.morena in the distributed data.)

* The recent linear tendency One way of adopting it is with /vote forecasting/ from tendencies discernible in the previous five congressional races (Magar 2012). I summarize the central tendency of the recent historical vote by means of linear estimation in time, fitting a straight line for each year analyzed and each party in the municipality or /sección electoral/.

The slope of the fitted line (the tendency) serves to extrapolate the party's electoral support to the future. For instance, to get the vote share that the recent past predicts for a party in unit u for 2018, I estimate the following equation

$$v_{ut} = a_u + b_u \times t + \text{error}_{ut},\ t = 2003, \ldots, 2015 \tag{1}$$

that I then use to predict $\hat{v}_{u2018} = \hat{a}_u + \hat{b}_u \times 2018$. This is an out-of-sample prediction of the party's vote share, it can be compared to the party's actual vote share in 2018 to gauge whether or not the unit approximates the historical record. For the 2015 forecast the sample shifts one period to become $t = 2000, \ldots, 2012$, and so on an so forth for previous years. I distribute forecasts for 2009, 2012, 2015, and 2018, which involved fitting about 10 thousand municipal regressions and more than 250 thousand sección-level.

(Vote forecasts appear in the fields vhat.pan , vhat.pri , and vhat.morena of the distributed data.)

* The party's support core The other historical statistic is the parties' core support in the unit. Its definition stems from classifying voters in three categories: (1) support groups, that in the past have consistently supported the party; (2) opposition groups, that have consistently supported another party; and (3) swing groups, that have neither consistently supported nor consistently opposed the party (Cox and McCubbins 1986). The party core consists of the support groups.

I use the procedure by Díaz Cayeros /et al/. (2016) to estimate this core. If $\bar{v}_t$ denotes the party's mean support across all units in period t,[2] for each party in each unit I fit

$$v_{ut} = \alpha_u + \beta_u \times \bar{v}_t + \text{error}_{ut},\ t = 1994, \ldots, 2018. \tag{2}$$

β_u measures the effect that national party tides have on the party's vote in unit u. For instance, $\hat{\beta}_u = 1$ estimates that for every percentage point that the party wins or loses nationally in year t, it also wins or loses one percentage point in the unit; $\hat{\beta}_u = 0$, on the other hand, would indicate the unit's full isolation form national swings. It is therefore a measure of party volatility in the municipality or /sección/ (analogous to "beta volatility" in the financial literature).

The α coefficient estimates the core size: expected support in unit u in the hypothetical case that the party receives no vote at the national level. For instance, $\hat{\alpha} = .4$ would indicate that, in the starkest of scenarios, 40

A critique that can be anticipated towards this measure of the party's support core is its extreme counter-factual nature (King and Zeng 2006). It deserves rigorous scrutiny, something I plan doing in the near future.

[2]The analyzed unit u should be dropped from period t's mean in order to not include the dependent variable in the right side of equation 2. I do not drop it due to the large number of municipal or seccional units (each contributing a fraction to the mean) and the use of vote shares (so large units are watered down): this refinement's impact should be negligible in each mean value.

(The parties' core support appears in the fields alphahat.pan , alphahat.pri , and alphahat.morena of the distributed data. Party volatility in betahat.pan , betahat.pri , and betahat.morena .)

* Compositional variables I close with an important feature of the model specifications, associated with the *compositional* nature of electoral returns. Compositional variable are quantitative descriptions of the parts of a whole. They therefore have two characteristics: (1) they are proportions that (2) add up to unity.[3]

When estimating parties separately, the challenge of equations 1 and 2 is to avoid forecasting vote shares less than zero or greater than one; and that the sum of party forecasts equals 1. To achieve this, Aitchison (1986) proposes substituting vote shares by log-ratios in the analysis. Arbitrarily setting the PRI as the reference party, define party p's vote relative to the PRI as

$r_p = \frac{v_p}{v_{\text{pri}}}$.

A value $r_p = 1$ would indicate a tie between the party and the PRI, while $r_p > 1$ that it finished ahead of the PRI in the proportion that the value reveals.

Thus, equation 1 is re-specified as

$\ln r_{put} = a + b \times t + \text{error}$

y equation 2 as

$\ln r_{put} = \alpha + \beta \times \bar{r}_{pt} + \text{error}$.

Applying the natural logarithm attenuates the effect of extreme values of the regressor on the dependent variable, similar as a logit regression would. Models fitting was done with ordinary least squares.

Coefficient estimates requires transformation to collect party vote shares. Illustrating with a three-party caste, it is trivial that

$$\hat{v}_p = \frac{\hat{r}_p}{1 + \hat{r}_{\text{pan}} + \hat{r}_{\text{morena}}} \text{ and } \hat{v}_{\text{pri}} = \frac{1}{1 + \hat{r}_{\text{pan}} + \hat{r}_{\text{morena}}}. \tag{3}$$

These are the quantities that the distributed data report.

Fácil de implementar en R: a = voto partido unidad 1 A = voto efec unidad 1 N = 3 unidades tengo v.bar = 1/3 * (a/A + b/B + c/C) quiero v.bar.sin.aA = 1/2 * (b/B + c/C) hago 1/3 * (a/A + b/B + c/C) = v.bar a/A + b/B + c/C = 3 * v.bar b/B + c/C = 3 * v.bar - a/A 1/2 * (b/B + c/C) = (3 * v.bar - a/A) * 1/2 = v.bar.sin.aA v.bar.sin.aA = (N * v.bar - a/A) * 1/(N-1)

References

Aitchison, John. 1986. *The Statistical Analysis of Compositional Data*. New York: Chapman and Hall.

Ames, Barry. 1970. "Bases of support for Mexico's dominant party." *American Political Science Review* 64(1):153–87.

Campbell, James E. 1991. "The presidential surge and its midterm decline in congressional elections, 1868–1988." *The Journal of Politics* 53(2):477–87.

Cantú, Francisco, Scott Desposato and Eric Magar. 2014. "Consideraciones metodológicas para estudiantes de política legislativa mexicana: sesgo por selección en votaciones nominales." *Política y Gobierno* 21(1):25–54.

Cox, Gary W. and Mathew D. McCubbins. 1986. "Electoral Politics as a Redistributive Game." *The Journal of Politics* 48(2):370–89.

Díaz Cayeros, Alberto, Federico Estévez and Beatriz Magaloni. 2016. *The Political Logic of Poverty Relief: Electoral Strategies and Social Policy in Mexico*. New York: Cambridge University Press.

Estévez, Federico, Eric Magar and Guillermo Rosas. 2008. "Partisanship in Non-Partisan Electoral Agencies and Democratic Compliance: Evidence from Mexico's Federal Electoral Institute." *Electoral Studies* 27(2):257–71.

Hoyos, Rafael E. de, Juan Manuel Espino and Vicente García. 2012. "Determinantes del logro escolar en México. Primeros resultados utilizando la prueba ENLACE media superior." *El Trimerstre Económico* 79(316):783–811.

[3]Formally, the compositional are random variables subject to two constraints: $0 \leq v_p \leq 1 \; \forall \; p \in P$ *and* $\sum_P v_p = 1$..

Imai, Kosuke, Gary King and Clayton Nall. 2009. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science* 24:29–53.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila and Héctor Hernández Llamas. 2007. "A "Politically Robust" Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program." *Journal of Policy Analysis and Management* 26:479–506.

King, Gary and Langche Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14:131–159.

Magar, Eric. 2012. "Gubernatorial coattails in Mexican congressional elections." *The Journal of Politics* 74(2):383–99.

Molinar Horcasitas, Juan and Jeffrey A. Weldon. 1994. Electoral Determinants and Consequences of National Solidarity. In *Transforming State-Society Relations in Mexico: The National Solidarity Strategy*, ed. Wayne A. Cornelius, Ann L. Craig and Jonathan Fox. La Jolla: Center for U.S.-Mexican Studies, University of California, San Diego.

Rosas, Guillermo and Joy Langston. 2011. "Gubernatorial effects in the voting behavior of national legislators." *The Journal of Politics* 73(2):477–93.

Segovia, Rafael. 1979. "Las elecciones federales de 1979." *Foro Internacional* 20(3):397–410.