On-line appendix for "Components of Partisan Bias Originating from Single-Member Districts in Multi-Party Systems: The Case of Mexico"

E. Magar   A. Trelles   M. Altman   M.P. McDonald
ITAM   *Pitt*   *MIT*   *UF*

June 13, 2016

A1   Introduction

We offer here a sketch of the on-line appendix. Code to replicate the analysis is available

here. All will be polished/fully written if manuscript is accepted for publication.

A2   The Mexican estimation procedure step by step

Step 1: Aggregate sección-level into district-level votes

Table A1 summarizes five Diputado elections analyzed. Part A reports official returns

almost straight from the source (we have aggregated a residual others category for conve-

nience). Data includes the number of plurality districts that parties/coalitions contested, the

national vote share v, and the share of plurality seats, out of 300, won s. Statistics on partial

coalitions (e.g., between the PRI and Green in all but one year) appear here: the number of

districts where it happened or not, and the share of votes and seats won in tandem.

Part A: before partial coalitions allocated to major party

	2003			2006			2009			2012			2015		
	dis.	v	s	dis.	v	s	dis.	v	s	dis.	v	s	dis.	v	s
PAN	300	.32	.27	300	.34	.45	300	.29	.24	300	.27	.17	300	.22	.18
PRI	203	.24	.40	—	—	—	237	.28	.46	101	.11	.17	50	.04	.08
PRI coal.	97	.14	.15	300	.28	.22	63	.12	.17	199	.27	.41	250	.33	.53
PRD	300	.18	.18	—	—	—	300	.14	.13	—	—	—	200	.05	.02
PRD coal.	—	—	—	300	.31	.34	—	—	—	300	.28	.24	100	.08	.10
Green	203	.04	—	—	—	—	237	.06	—	101	.02	.01	50	.01	—
MC	300	.02	—	—	—	—	—	—	—	—	—	—	300	.06	.03
MC coal.	—	—	—	—	—	—	300	.07	.01	—	—	—	—	—	—
Morena	—	—	—	—	—	—	—	—	—	—	—	—	300	.09	.05
Other	300	.06	—	300	.07	—	300	.04	—	300	.05	—	300	.12	.01

Part B: partial coalitions given to major party, shares of effective vote and district population

Party	2003			2006			2009			2012			2015		
	v_1	v_2	s	v_1	v_2	s	v_1	v_2	s	v_1	v_2	s	v_1	v_2	s
PAN	.32	.08	.27	.34	.11	.45	.29	.07	.24	.27	.11	.17	.22	.07	.18
PRI	.38	.10	.55	.28	.09	.22	.40	.09	.63	.38	.16	.58	.38	.11	.62
PRD	.18	.05	.18	.31	.10	.34	.14	.03	.13	.28	.12	.24	.13	.04	.12
Green	.04	.01	—	—	—	—	.06	.01	—	.02	.01	.01	.01	.00	—
MC	.02	.01	—	—	—	—	.07	.02	.01	—	—	—	.06	.02	.03
Morena	—	—	—	—	—	—	—	—	—	—	—	—	.09	.03	.05
Abtention	—	.75	—	—	.71	—	—	.77	—	—	.60	—	—	.72	—

Table A1: Summary of five elections. Entries in part A indicate the number of plurality districts contested by each party/coalition, the national vote share won (v), and the share of plurality seats won (s). In part B, the national vote share relative to effective votes with coalitions allocated to the corresponding major party (v_1) and the same including non-voters compared to total population (v_2). Prepared with data from www.ine.mx and www.inegi.org.mx

Part B aggregates the partial coalitions by allocating their votes in those districts to the major party involved, as discussed in the text. It reports two metrics for national vote shares: v_1 has effective votes in the denominator (total votes minus voided ballots, votes for write-in candidates, and votes for small parties dropped from analysis[1]); v_2 has total district population in the denominator (so that abstention is the share of the total population not voting for one of the parties/coalitions in the analysis). The separation method relies upon relative district populations and turnout statistics to compute partisan bias.

Data discussed in the last two paragraphs are national-level aggregates. Analysis in fact proceeds from much smaller aggregates: district- and sección-level returns. Sección-level votes are used to reproduce district returns using hypothetical maps (e.g., the 2003 votes had it been held with the 2006 district map)—secciones are the basic building blocks of redistricting. District-level returns are then used as input for the simulation of national elections, as detailed in step 3.

Step 2: Produce inter-census population estimates

Population figures for years 2000, 2005, and 2010 prepared by Mexico's census bureau (INEGI) for the purpose of redistricting. Data is available at the sección-level, making the aggregation of district populations possible. With elections taking place between census counts, and after, we proceeded to estimate inter-census populations by linear projection of 5-year rates of growth. So the 2000–2005 growth rate was used to estimate year 2003 district populations, and the 2005–2010 rate for years 2006, 2009, 2012, and 2015. Population projections for different maps were done after sección census populations had been

----

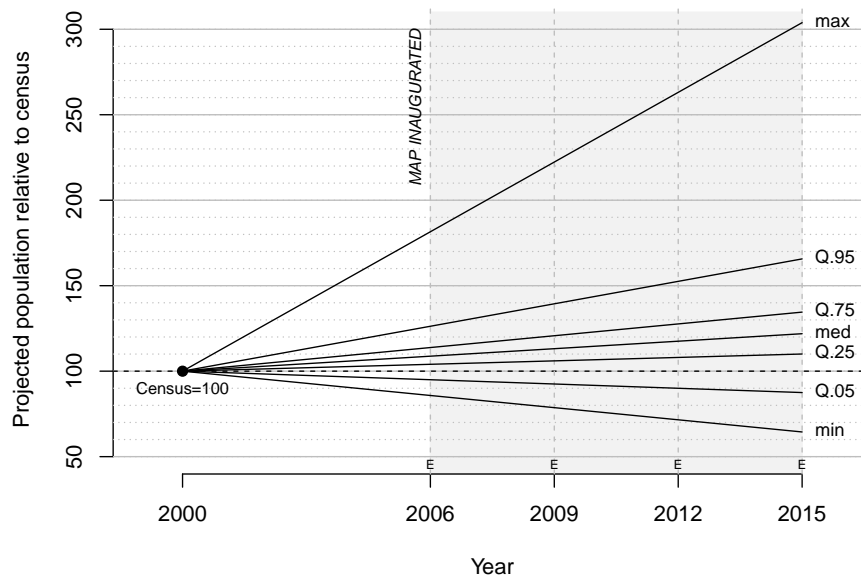[1]All will be listed in this footnote.

Figure A1: The 2006 map and demographic change. Plotted population projections relied on the 2000–2010 censuses rate of change (see text for details). Letters E in the horizontal axis indicate elections using this map. Source: prepared with data from www.inegi.org.mx and www.ine.mx.

aggregated into actual or hypothetical districts. Performing linear projection on secciones before any aggregation might have been preferable (because they are much smaller geographic units), but a fair amount of over-populated secciones are routinely split into new ones between elections, complicating the projection exercise.

Figure A1 illustrates the projection exercise, and how it puts creeping malapportionment within reach to assess its impact on representation. If every district had experienced the same rate of population change, the redistricting census lag would be inconsequential. With variable rates of population growth, malapportionment creeps in. Compared to the 2000 census, projected district populations in 2006, when the map was inaugurated, are off by 9.7% in absolute value on average, with a standard deviation of 10.6%. Indexing 2000 census district populations at 100, as the figure does, reveals how different the most demo-

graphically dynamic units were on paper and in reality. The fastest-growing district was 88% larger in 2006 than what census data otherwise suggested (the line labeled 'max'). The district shedding most population was 16% smaller (the 'min' line). These are out-liers, but central tendencies reflect sizable lags as well. The inter-quartile range (lines Q.25 and Q.75) covered 1%–13% above census in the 2006 election, and expanded to 4%–20%, 7%–27%, and 10%–35% in the three subsequent Diputado elections using the same map.

Estimating intercensal populations non-linearly is also preferable to our linear approach. The key problem on this front appears to be the choice of a functional form that both smoothes the inter-census rate of population growth while also taking the values actually observed on three census years (2000, 2005, and 2010). An exponential form between pairs of census counts (Das Gupta 1978) does a good job for years between observations, but not before and after, nor does it treat transitions from one pair to the next smoothly. A polino-mial form would allow work with all three census counts, but also seems problematic for proecting estimates beyond 2010. We therefore abandoned the attempt to refine population estimates non-linearly.

Step 3: Simulate national-level elections

In order to simulate national elections (i.e., a vote-seat share for each party), the Linzer method infers the distribution of district-level returns from every pair of parties' correla-tions. Some parties do not field candidates in every district, so finding district patterns of party contestation starts the process (Linzer 2012:405). For example, two patterns were ob-served in 2009: the PAN, PRI, PRD, and MC contested 63 districts, while the same plus the
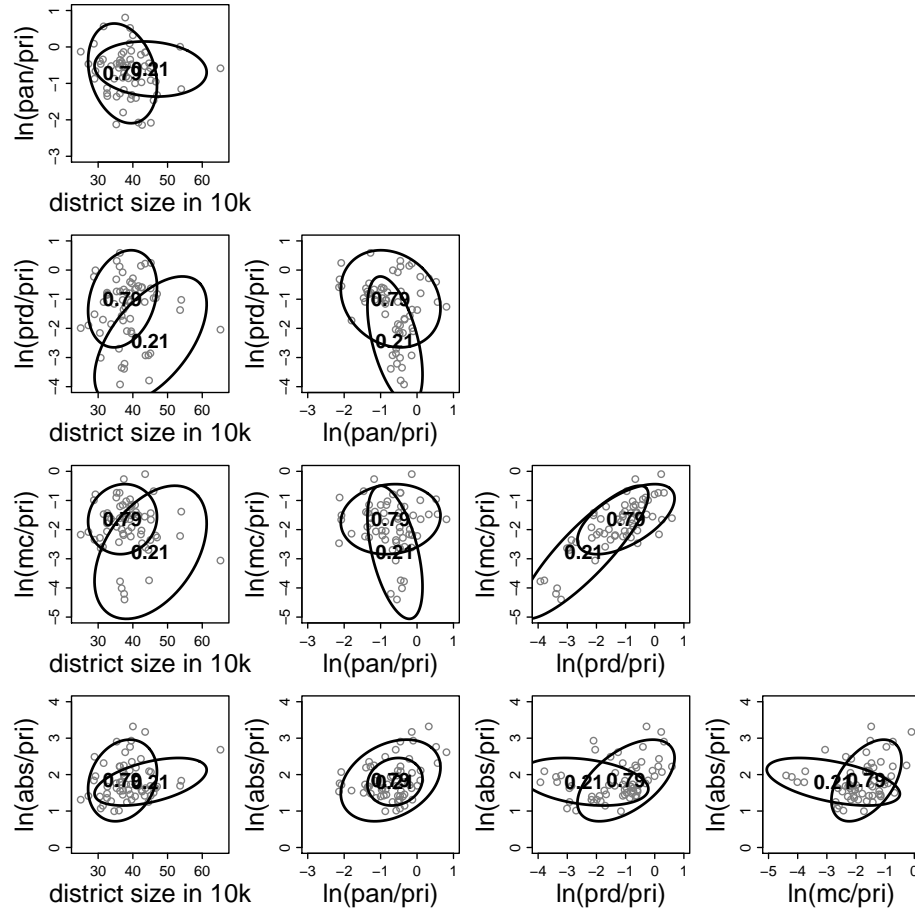
Figure A2: How pairs of parties' votes correlated (vis-à-vis the PRI) in 2009: 63 districts that the Green party did not contest

Green party contested the remainder 237 districts. Subsets of districts in each pattern are analyzed separately. To also take district population and turnout into account, the patterns of contestation also include the abstention rate, as discussed in step 1.

The density of district returns is unlikely to conform to a textbook statistical distribution. But a mix (superimposition) of multi-variate normals will capture much of the observed variance. Exploration of district outcomes by plotting the relative performance of party pairs (including abstention) relative to the PRI in 2009 (first pattern of contestation) reveals two apparent clusters of districts. In one, including 21 percent of districts in the

pattern, MC won/lost votes relative to the PRI ($\ln(\frac{mc}{pri})$ increases) without hurting PAN's performance relative to the PRI ($\ln(\frac{pan}{pri})$ constant); in the other cluster, accounting for 79 percent of districts, PAN won/lost votes relative to the PRI without much affecting how MC performed relative to the PRI. The same is generally true of the PRD and the PAN (as seen in the $\ln(\frac{prd}{pri})$ and $\ln(\frac{pan}{pri})$ plot), but not of the PRD and MC (districts where one grew relative to the PRI also had the other growing relative to the PRI). Choosing the number of components to use is impressionistic (Linzer 2012:405): we proceeded with a single component for the first pattern of 2003 (hypothetical map), for both patterns of 2015 (status quo map), and for the second pattern of 2015 (hypothetical map); two components for the second pattern of 2003 (hypothetical map), for the first pattern of 2015 (hypothetical map), and for both patterns of 2003 (status quo map), 2009, and 2012; and three components for the sole pattern of 2006.

Figure A3 compares observed 2009 frequencies with the mixture model's marginal density to verify that there is a good match between the two. Other than a slight bimodality in the abstention rate, marginals follow the contour neatly. Monte Carlo sampling with a weighted mix of the bi-variate normals generated one thousand simulated elections for 2009 in the analysis. The procedure is identical for all other years/maps.

Step 4: Estimate responsiveness and partisan bias

We use MCMC to estimate equation 2, implemented with Jags (Plummer 2003), called from R (R Dev. Core Team 2011). Preparation involves generating data objects first: i indexes $I = 100$ simulated observations for a given year; p indexes $P = 6$ parties in the
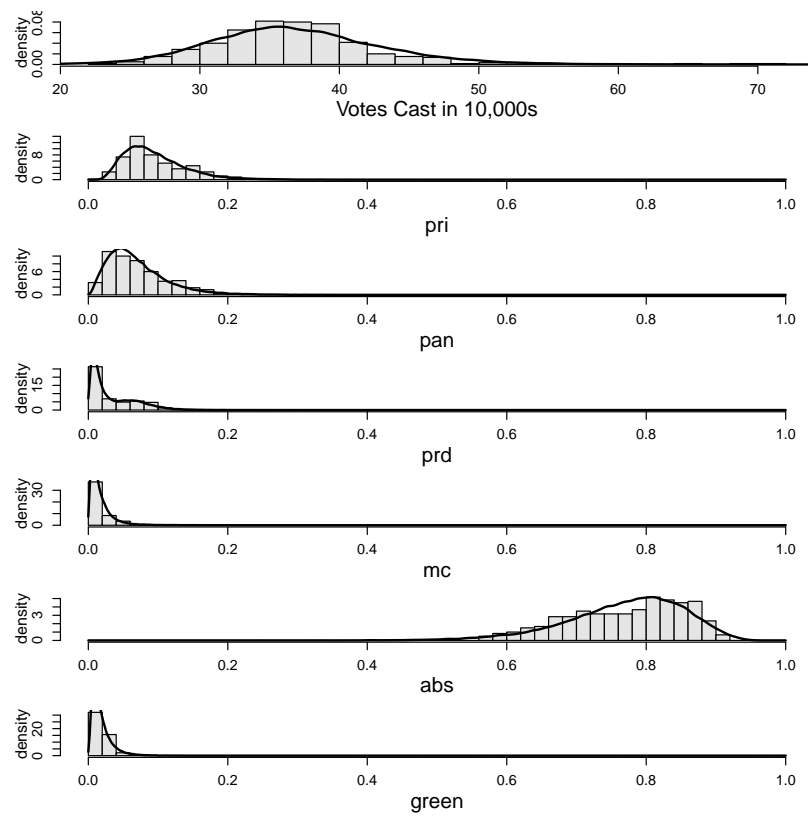
Figure A3: Mixture model's marginal densities in 2009

analysis; and $D = 300$ is the number of plurality districts; votes and seats $I \times P$ matrices $\mathtt{v}$ (where $\mathtt{v}_{i,p}$ is party p's vote share in the ith simulated election) and $\mathtt{S}$ (where $\mathtt{S}_{i,p}$ are the seats that party p won in the ith simulation); and a P-length vector $\mathtt{dummy}$ (where $\mathtt{dummy}_p$ equals 1 if party p contested the election in the given year, 0 otherwise). Our single-year research design avoids the obstacle posed by elections with varying sets and sizes of candidatesin in MCMC estimation (the analyst could adapt the Bugs model in Table A2 to the number of parties in the simulated elections). The code, however, is prepared to tackle a multi-year estimation with different contestation patterns: $\mathtt{dummy}$ indicates parties contesting each election and fed to the MCMC process. Each additive component of the numerator and denominator (i.e., the party's $\lambda * v^{\rho}$) is multiplied by the corresponding dummy, so that parties not contesting drop from the likelihood function. All data objects are bundled in a list

$$\text{lrdata} \; \textbf{<−} \; \textbf{list} \; (\text{"S", "v", "I", "J", "D", "dummy"}).$$

We next randomize a set of initial values for the model parameters (λ needs only $J - 1 = 5$ initial values because we restrict $\lambda_{\mathrm{PRI}} = 0$)

$$\text{lrinits} \; \textbf{<−function}\,()\{ \; \textbf{list} \; (\text{lambda=}\textbf{rnorm}(J{-}1), \text{rho=}\textbf{rexp}(1)) \}$$

and define objects to store parameters' posterior distribution samples

$$\text{lrparameters} \; \textbf{<−c}(\text{"lambda", "rho"}).$$

The formal model estimated appears in Table A2. Reliance on $\mathtt{dummy}$ is convenient as is allows to proceed with six-party votes and seats matrices regardless of the year selected— otherwise the code would need to be adapted each year for variable parties, which is trivial.

9

```
lambda.rho <- function() {
  for (i in 1:I){    # loop over observations (simulations)
    for (p in 1:P){ # loop over parties (dummy selects who ran)
      S[i,p] ~ dbin(pi[i,p], D[i])   # D is the number of SMD seats
    }
    numerator[i,1] <- dummy[i,1] * exp( lambda[1] ) * v[i,1]^rho
    numerator[i,2] <- dummy[i,2]                      * v[i,2]^rho
    for (p in 3:P){
      numerator[i,p] <- dummy[i,p] * exp( lambda[p-1] ) * v[i,p]^rho
    }
    for (p in 1:P){ # loop over parties
      d1[i,p] <- dummy[i,1] * exp( lambda[1] ) * v[i,1]^rho
      d2[i,p] <- dummy[i,2]                     * v[i,2]^rho # reference
      d3[i,p] <- dummy[i,3] * exp( lambda[2] ) * v[i,3]^rho
      d4[i,p] <- dummy[i,4] * exp( lambda[3] ) * v[i,4]^rho
      d5[i,p] <- dummy[i,5] * exp( lambda[4] ) * v[i,5]^rho
      d6[i,p] <- dummy[i,6] * exp( lambda[5] ) * v[i,6]^rho
      denominator[i,p] <- d1[i,p] + d2[i,p] + d3[i,p]
                       + d4[i,p] + d5[i,p] + d6[i,p]
      pi[i,p] <- numerator[i,p] / denominator[i,p]
    }
  }
  ### priors
  for (q in 1:5){ # P=6 party labels minus reference party is 5
    lambda[q] ~ dnorm( 0, tau.lambda )
  }
  tau.lambda <- pow(.25, -2)
  rho ~ dexp(.75) # has positive range, med about 1, mean 1.25, max 4.5
}
```

Table A2: Code for Bugs model

Parties not contesting an election (Morena in 2003–12, Green in 2006, MC in 2006 and 2012), with zero votes and seats columns, need not be dropped from matrices. Instead, multiplying the corresponding $\lambda * v^\rho$ term by contestation `dummy` annuls them from equation 2.

Our multinomial logistic regression type model satisfies the independence of irrelevant alternatives assumption in the same way that King does. Quoting at length:

> [T]he implied assumption of independence of irrelevant alternatives is satisfied ... since the entire stochastic component is conditional on all parties and votes. The only random choice being made is by the electoral system in assigning seats to parties. Therefore, I use the multinomial probability distribution for the number of seats allocated to the J political parties, a straightforward generalization of the binomial (King 1990:168).

The only difference is our use of P seat-allocating binomial distributions instead of one multinomial (line 4 of the code in Table A2).

R package `r2jags` (Su and Yajima 2012) is a suite of functions to bind all these objects into a Jags call for estimation (and do post-estimation evaluations):

```
results <- jags (   data = lrdata,
                    inits = lrinits,
                    parameters.to.save = lrparameters,
                    model.file = lambda.rho,
                    n.chains = 3,
                    n.iter = 50000,
                    n.burnin = 25000,
                    n.thin = 50   ).
```

We stopped the algorithm after 50 thousand iterations had been completed. The first 25 thousand were discarded to let the model adapt by updating the prior distributions of parameters to the data. Non-informative priors were used across the board: $\lambda_q \sim \mathcal{N}(0,4)$, $q =$

1,...5) and $\rho \sim \exp(\frac{3}{4})$ (the exponential distribution has positive range and, with this parameterization, a median of about 1, a mean of 1.25, and a maximum value of 4.5). Every fiftieth iteration of the final 25 thousand was saved as sample of parameters' posterior distributions (of size 3 chains $\times 25000/50 = 1500$).

To verify convergence, Tables A3–A9 plot Gibbs sampler iterations (x-axis) and sampled values of model parameters (y-axis). One election/map is reported per Table, and one parameter per plot, with columns selecting three ways to measure votes: raw vote shares (v), mean district vote shares ($\bar{v}$), and population-weighted mean district votes ($\bar{\bar{v}}$). With three chains utilized in each Jags call, plots allow verification of two things: that chains had reached a steady state; and that all chains (one in green, one in red, one in blue in the plots) had converged. With no exception, the parameter samples we report in the text had all converged.

A3 The effect of alternative coalition specifications

As recognized in the text, the solution to deal with partial coalitions—situations where two parties present a joint candidate in less than all districts, competing against each other in the rest—is less than perfect. We offer some perspective of how alternative approaches affect the bias estimates reported. We illustrate with the case of the PRI-Green partial coalition in the 2015 election (partners were rivals in 50 plurality districts, see Table A1). Partial coalitions of this kind occurred every year in the period except 2006. Two cases in official vote returns need consideration.

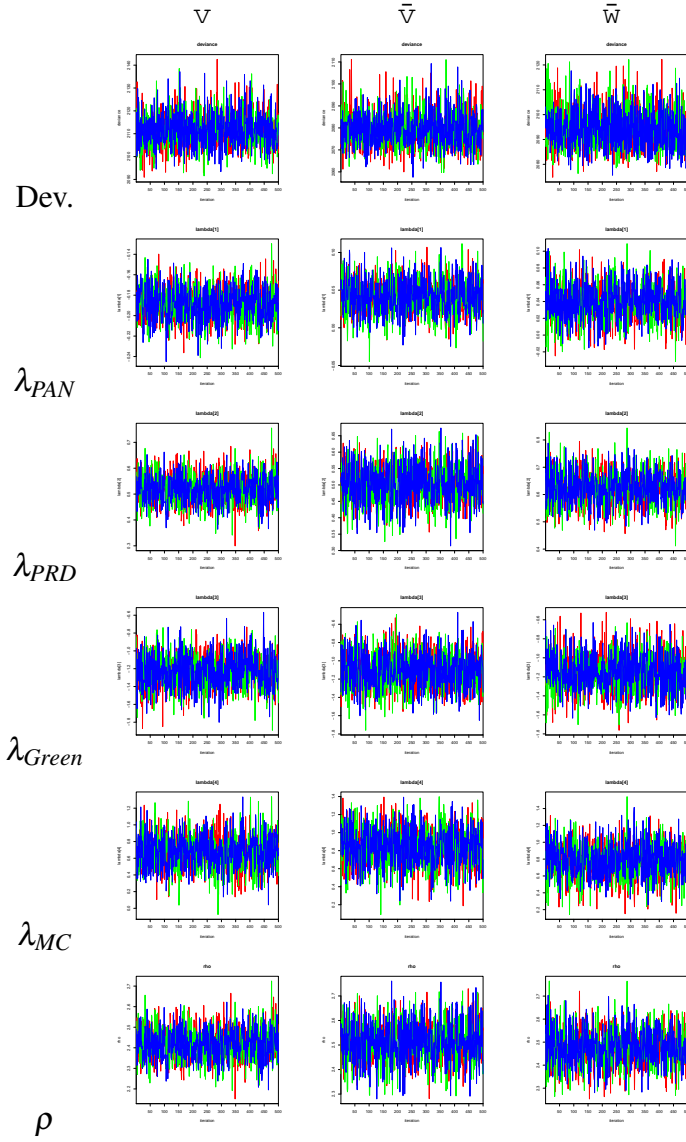(A) In districts where the partners fielded a joint candidate, coalition supporters had

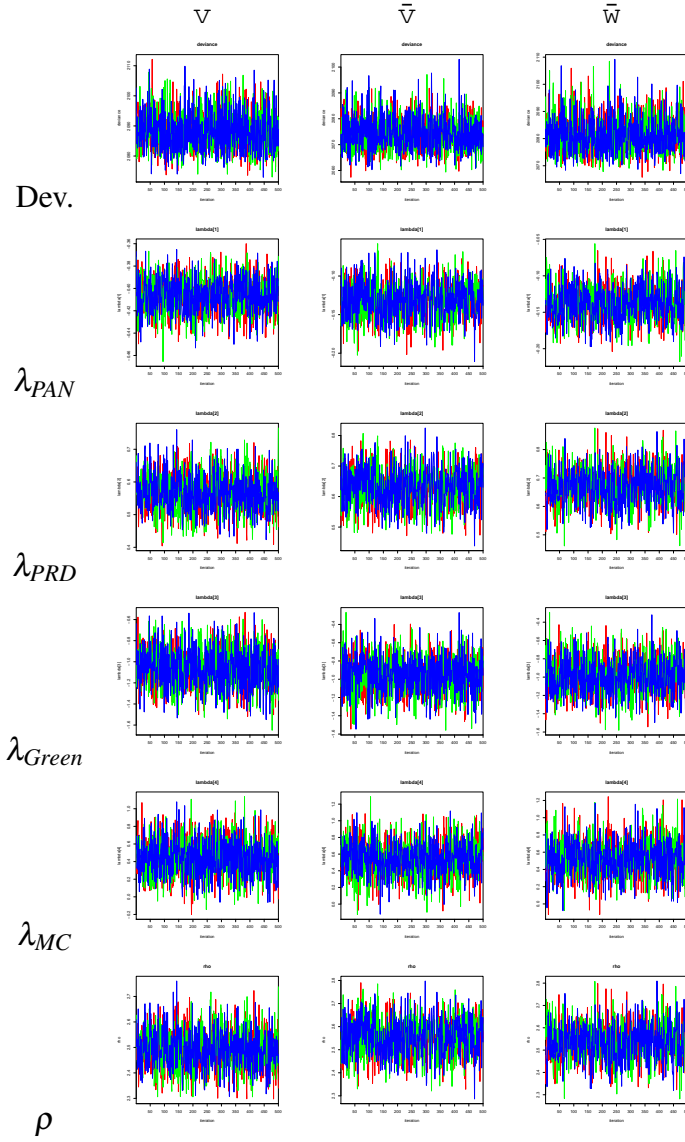Table A3: Three-chain traceplots for 2003 with status quo map

	V	V̄	w̄
Dev.			
λ_{PAN}			
λ_{PRD}			
λ_{Green}			
λ_{MC}			
ρ			

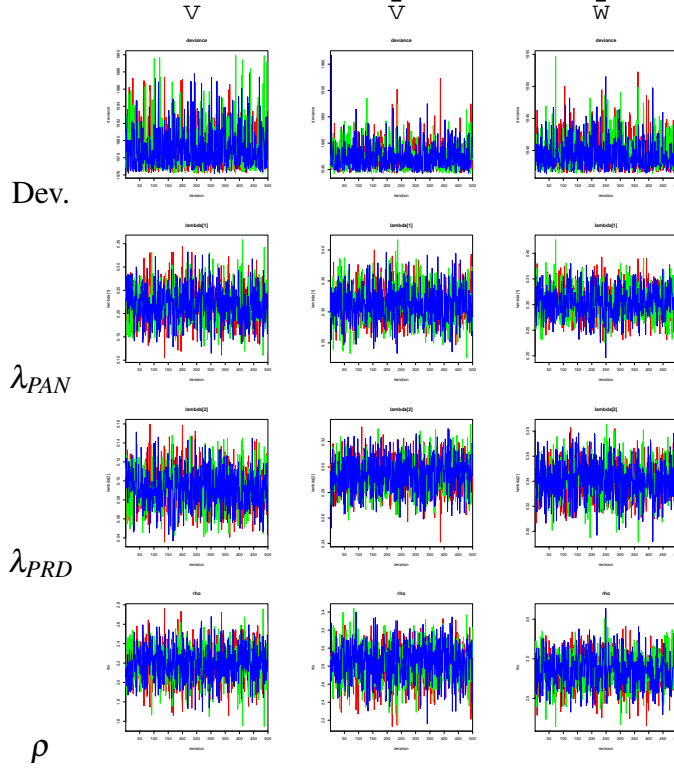Table A4: Three-chain traceplots for 2003 with hypothetical map

14

Table A5: Three-chain traceplots for 2006 with status quo map

three valid voting options: vote PRI, vote Green, or vote both. Official returns therefore include three raw votes figures for the coalition, which were tallied to determine whether or not the coalition candidate won the district. Seats, therefore, are all allocated to the coalition.

(B) In districts where PRI and Green were rivals, supporters could validly pick only one of them. Official returns therefore include two independent raw votes figures, and each party won seats on its own.

The disconnect between party votes and seats in type A districts is an obstacle when inferring national votes and seats in the Linzer simulation method. In the analysis, we removed the disconnect by giving the PRI the full coalition total votes and seats in type A districts, leaving the Green party with zero votes and seats. Type B districts required no
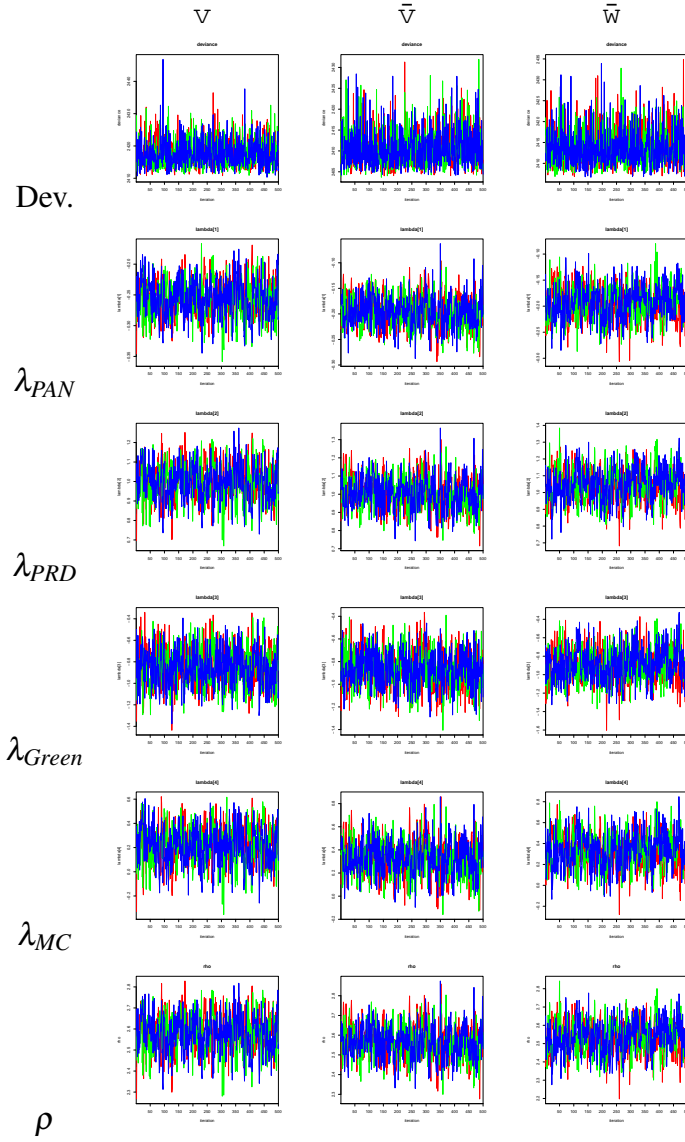
15

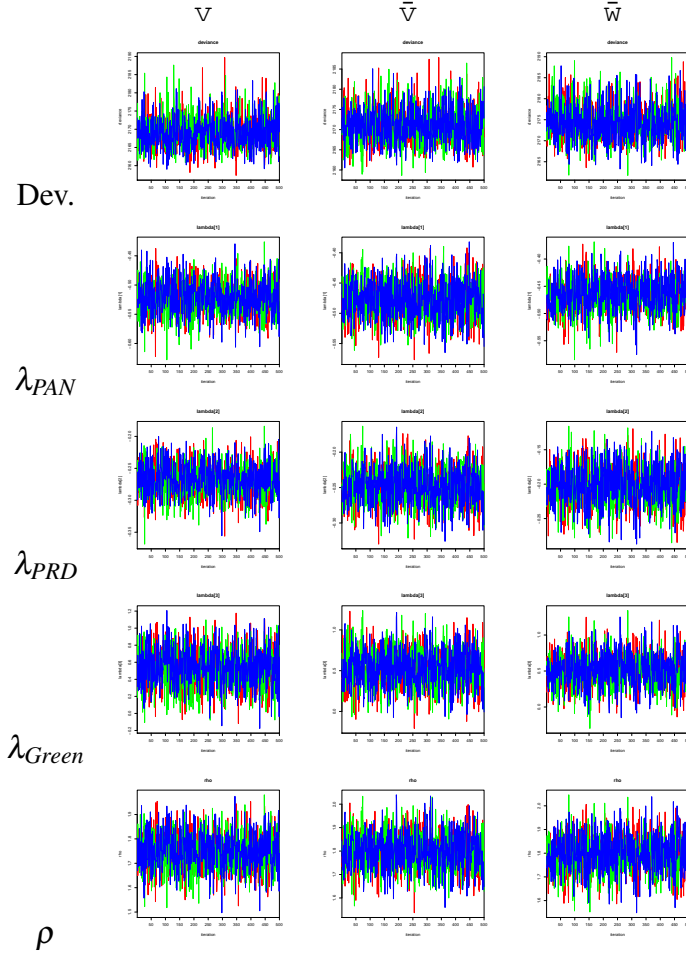Table A6: Three-chain traceplots for 2009 with status quo map

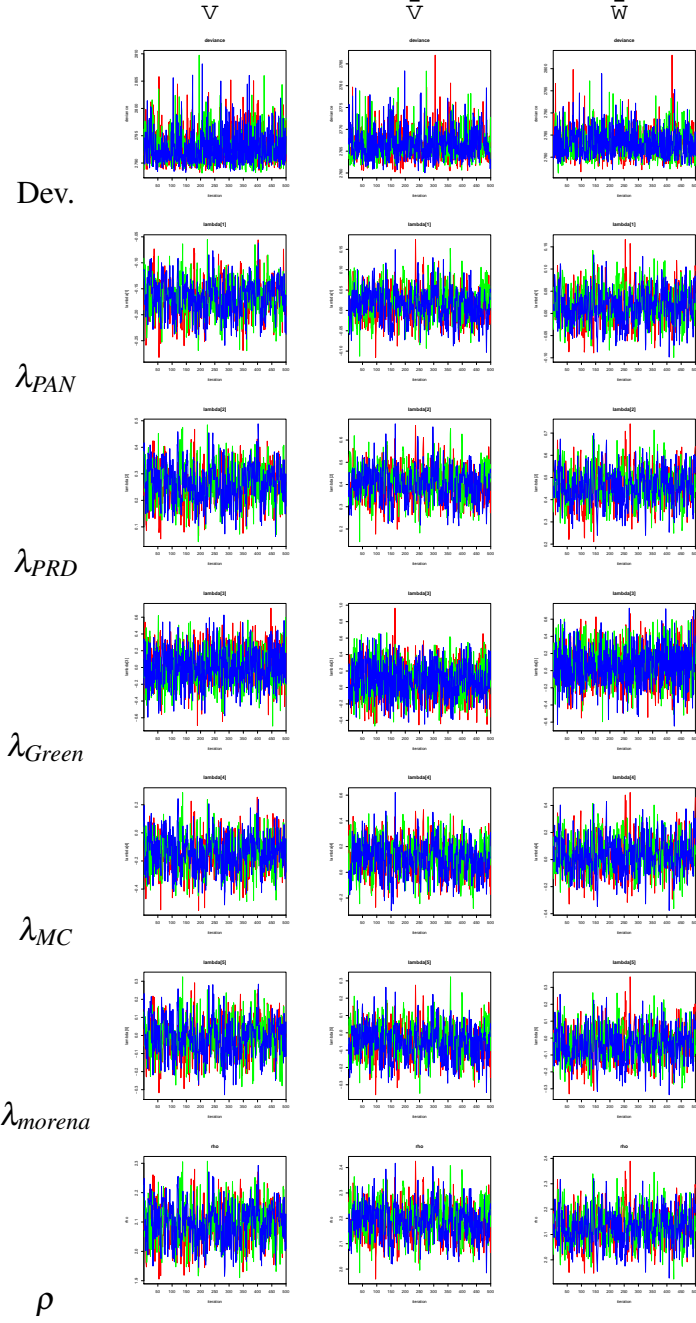Table A7: Three-chain traceplots for 2012 with status quo map

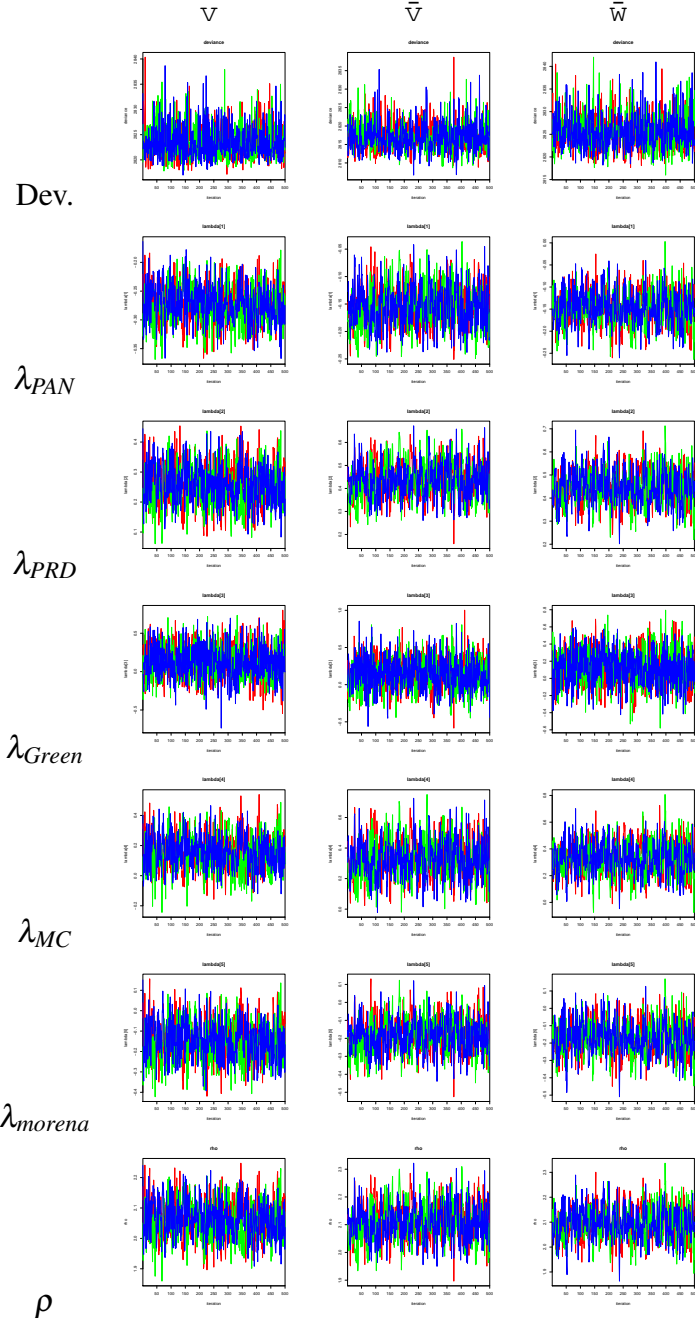Table A8: Three-chain traceplots for 2015 with status quo map

	$\mathbb{V}$	$\overline{\mathbb{V}}$	$\overline{\mathbb{W}}$
Dev.			
λ_{PAN}			
λ_{PRD}			
λ_{Green}			
λ_{MC}			
λ_{morena}			
ρ			

Table A9: Three-chain traceplots for 2015 with hypothetical map

19

manipulation, so both the PRI and the Green retained the votes and the seats they won.

We consider two alternative methods to handle partial coalition results. One we call the 'Green as a PRI Faction': in every district, the PRI receives all votes and seats won by the Greens solo or jointly with the PRI. The Green party is therefore dropped from the analysis. The other is the 'Three Way Split': in type A districts, all votes and seats are allocated to the "PRIcoal." entity, as if an offspring of the parent parties. In type B districts, the PRI and the Green keep the votes and seats each won separately. So nationwide, the partial coalition gives rise to three vote- and seat-winning parties. The Green as PRI Faction approach makes the PRI artificially stronger than it actually is. The approach in the text does the same, but to a smaller extent (by not manipulating districts with rival partners). The Three Way Split does the opposite, a weaker than true PRI.

Table A10 reports 2015 partisan bias breakdown estimates with the alternative partial coalition manipulations. Part A transcribes the estimates reported in the text, for reference. Part B reveals bias estimates that deviate from the reference in predictable ways: by swelling the PRI votes and seats, the Green as PRI Faction approach slightly magnifies total relative bias against the PAN, and cuts relative bias in favor of the left by more than half. The distributive component follows a pattern similar to total bias for both parties. But the turnout component, which operates in similar degrees against both major parties relative to the PRI, also shrinks to about half with the artificially stronger PRI. The coalition appears to not have amalgameted as succesfully in lower-turnout districts as it dis in the higher-turnout. Changes in the malapportionenment component are mild.

The Three Way Split in part C underscores how total partisan bias relative to the PRI operates against every party reported, including the Green and the PRIcoal entity, but also

20

partisan bias	PAN–PRI	PRD–PRI	Green–PRI	PRICoal–PRI
Part A: 2015 election as reported				
total	−.17 (0)	+.26 (0)	+.01 (.47)	
distrib.	+.02 (.32)	+.40 (0)	+.10 (.32)	
turnout	−.19 (0)	−.19 (0)	−.04 (.46)	
malapp.	+.00 (.38)	+.05 (0)	−.05 (.42)	
Part B: 2015 election, Green as a PRI Faction				
total	−.23 (0)	+.11 (.11)		
distrib.	−.11 (.01)	+.20 (.02)		
turnout	−.10 (.05)	−.10 (.23)		
malapp.	−.02 (.41)	+.01 (.46)		
Part C: 2015 election, Three Way Split				
total	−.97 (0)	−.99 (0)	−2.36 (0)	−.36 (.01)
distrib.	−1.09 (0)	−1.03 (0)	−2.11 (0)	−.60 (0)
turnout	+.06 (.35)	−.06 (.28)	−.22 (.16)	+.17 (.18)
malapp.	+.07 (.32)	+.10 (.15)	−.04 (.42)	+.06 (.36)

Table A10: Result sensitivity to how partial coalitions are handled

the left. Estimates are also much less precise with this manipulation.

In sum, results are robust to the the Green as PRI Faction across the board. This approach is not too different from the one in the text. But, by breaking the PRI into two entities, the results reported in the text are much harder to spot in the Three Way Split. This digression should ease understanding of the implications associated with the handling of partial coalitions in the text—a Mexico specific feature that should pose no problem for analysis elsewhere.

A4   How the PR tier mitigates under-representation

Table A11 portrays how the mixed component mitigates the votes-seats distorsions of the plurality tier.

References

Das Gupta, P. 1978. "A general method of decomposing a difference between two rates into several components." *Demography* 15(1):99–112.

King, G. 1990. "Electoral responsiveness and partisan bias in multiparty democracies." *Legislative Studies Quarterly* 15(2):159–81.
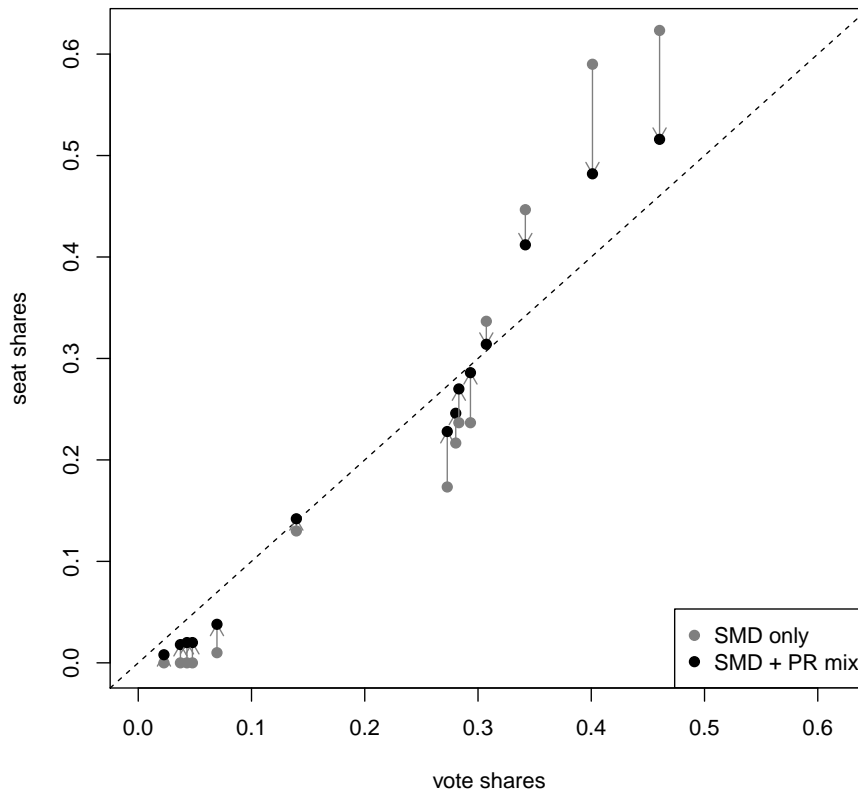
Table A11: Representativeness of the plurality tier and the mixed system compared, 2003–2015. Plot reports votes and seats won by each party nationwide in plurality districts only (gray points) and after the PR compensation is applied (black points).

Linzer, D. A. 2012. "The relationship between seats and votes in multiparty systems."

Political Analysis 20(3):400–16.

Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using

Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Sta-*

tistical Computing. Vienna, Austria: .

URL: *http://www.ci.tuwien.ac.at/Conferences/DSC-2003*

R Dev. Core Team. 2011. "R: A Language and Environment for Statistical Computing." R

Foundation for Statistical Computing `http://www.R-project.org`.

Su, Y.-S. and M. Yajima. 2012. *R2jags: A Package for Running jags from R*. R package

version 0.03-08.

URL: *http://CRAN.R-project.org/package=R2jags*