Methodological Considerations for Students of Mexican Legislative Politics: Selection Bias in Roll-Call Votes*

Francisco Cantú
UCSD
fcantu@ucsd.edu

Scott Desposato
UCSD
swd@ucsd.edu

Eric Magar
ITAM
emagar@itam.mx

August 12, 2013

Abstract

This paper examines the nature of the data available for studying legislative behavior in Mexico. In particular, we evaluate a potentially serious problem: only a subset of roll-call votes have been released for the critical transition period of 1998–2006. We test whether this subset is a representative sample of all votes, and thus suitable for study, or whether it is biased in a way that misleads scholarship. Our research strategy takes advantage of a partial overlap between two roll call vote reporting sources by the Chamber of Deputies: the site with partial vote disclosure, created in 1998 and still in place today; and the site with universal vote disclosure since 2006 only. An examination of the data generation and publication mechanisms, comparing different estimations of legislative behavior, reveals that omitted votes reduce the precision of estimates but do not introduce bias. Scholarship of the lower chamber can therefore proceed with data that we make public with the publication of the paper.

Hegemonic party rule in Mexico relegated Congress to the role of rubber-stamp for most of the twentieth century. The secondary role of the legislative branch was parallel with a general lack of resources and interest for citizens and scholars to track the behavior of legislators.[1] In recent decades, roll call votes were not published in

---

[1]For example, classic depictions of Mexican politics, such as ? or Cosío Villegas (1981), include no chapter on Congress. González Casanova (1965) addresses the question, but can only muster yearly aggregate figures showing a very high frequency of bills passed unanimously for selected years in support for the claim of presidential predominance.

the *Diario de los Debates*—the chamber's equivalent to the Hansard in Westminster systems or the Congressional Record in the U.S.—only total vote counts at the end of roll calls, producing opacity in the legislative process which benefitted the Institutionalized Revolutionary Party's (PRI) institutions of ruling.

The gradual transition to democracy reestablished constitutional authority that Congress has vis-à-vis the executive. With maximum majority size capped at three-fifths of the lower chamber and a loss of electoral competitiveness, the PRI required opposition for constitutional amendments required by the Salinas administration's ambitious reform program. And it was in 1997, when the PRI lost the majority of seats in the lower house of Congress, that scholars started to formulate questions about the legislature and its systematic role in the policy-making process (e.g., Cantú and Desposato 2012; Morgenstern 2003; Morgenstern and Nacif 2002; Rosas and Langston 2011; Weldon 1997, 2005). Accompanying this congressional revival has been a rise in the available data for its study. Resources like the roll-call votes are published in the Congress' website since November 1998. This information is the most common tool to analyze legislative behavior across the world, and provides citizens an instrument for holding deputies accountable.

One problem for citizens and scholars, however, is the published roll-call votes for the transition period 1998–2000 only represent a subset of all floor votes. Indeed, on average, 22 percent of all votes are not released to the public. During the 57th legislature (1997–2000), this figure peaked at 28 percent. This discrepancy and apparently selective release of roll-call votes poses a series of challenges. If the subset of available votes systematically differs from the universe of votes, most of the extant research on legislative behavior of the country is distorted, biasing our understandings of Mexican legislative politics during this period. Moreover, if such votes fail to represent the universe of roll call votes held in the floor of the Chamber of Deputies, voters possess a deficient tool to monitor what their representatives do in Congress.

This transition period is especially important for students of Mexican politics and of democratization more generally. Are the most important votes being held back? Do published roll-call votes provide a misleading perspective on the nature of legislative

politics during the transition? Scholars of Mexican legislative politics need to know whether these data can be trusted and utilized to make inferences about this critical period.

This paper examines the nature of the data available for studying this critical juncture in Mexican politics. Our research strategy takes advantage of two features of roll call vote reporting by the Chamber of Deputies. Beginning in 2006, the Parliamentary Statistics Information Service (INFOPAL) has released all roll call votes, while the Gaceta Parlamentaria continues to release only a subset. Thus, we can compare the universe of all roll-call votes with the subset chosen by Gaceta for the period 2006–2012. Second, although we do not have all roll call votes for the critical earlier period (1997–2006), we do have published vote totals—the total number of yea and nay votes—for every roll-call, published or not. Using this data source, we can compare key features of released votes with all votes, including overall cohesion, abstention, and participation.

Using a number of statistical methods, we analyze the released votes for evidence that they are biased subsets of the population of all votes. Our analysis suggests that these data are statistically indistinguishable from the broader population of roll-call votes. If the assumption that the pre- and post-2006 non-reporting criteria remains unchanged, we claim that scholars can study the 1998–2006 period without concern for the quality and nature of the sample mechanisms. The implications of our analysis rule out a selective omission of the roll-call votes and explains this problem as a result of lack of resources rather than opacity in the accountability system.

1   The *Gaceta* and missing votes

The *Gaceta Parlamentaria* was created in tandem with a new electronic voting system in the second half of 1998. It is a web-accessible repository of documents and information relevant to the day's business. Before that, when the PRI routinely controlled majorities in the chamber, deputies would not have access to much more than the order (*Orden del día*) at the start of the day. Anyone wishing a committee report,

attached documents or information pertinent to a specific motion, would have to request it from the chamber staff. Requests would often arrive late, if ever, prompting vocal complaints by minority legislators. As a consequence, deputies would often vote essentially "blindfolded" on the floor, with little information about the substance of the legislation.[2]

Besides providing access to key legislative documents, *Gaceta* also became a resource for congressional scholars, archiving detailed information including roll-call votes for the first time in decades (e.g., Casar 2002; Nacif 2002; Rosas and Langston 2011). Despite user-friendliness, it is not explicitly an archive and that has raised problems for research. *Gaceta* attempts, first and foremost, to have information relevant to the day's session readily available, something that is not always possible. The order, formalizing the actual business of the day, is negotiated by party leaders at the Junta de Coordinación Política and, quite often, finalized just minutes before the plenary session begins. Order votes that are announced (or can be anticipated by chamber staff) with enough time in advance have all documentation ready in the day's *Gaceta*. But votes that are either scheduled late, or scheduled under special rules after the start of the plenary session, as well as procedural votes, often do not appear in the *Gaceta* at all. As a consequence, the template used by chamber staff to receive the roll call vote after it is taken in the floor will not be ready either, and the vote remains unreported.[3]

To identify all roll call votes for comparison with the published subset, we downloaded text-only versions of the *Diario de los Debates* (Diario, for short) corresponding

---

[2]The waning days of an ordinary session of the 56th Legislature offer one example. The record suggests that PAN and PRI had agreed to amend article 105 of the constitution against the PRD's opposition. At 1:55pm, PAN deputy Francisco José Peniche y Bolio requested a vote to dispense the second reading and proceed to final passage instead. PRD deputy José Jesús Ortega Martínez was then recognized to state his opposition to the special rule. "We are once again asked to legislate on our knees" he accused, "and legislate is just a manner of speaking because we are merely presented with proposals and practically called to vote in favor or against (...) [T]oday at noon or 1pm you all, or the great majority of you all, became aware of the contents of the law we are about to discuss (...) What is our responsibility, my peers, especially my Acción Nacional peers, who are requesting to skip procedural steps?" (*Diario de Debates*, second ordinary session of the 56th Legislature's first year, April 24, 1995.)

[3]We attempted to recover these unpublished votes, contacting legislative staff and repeatedly requesting information. After multiple efforts, we were told by one director that these votes had been discarded after sessions and simply did not exist.

	Votes reported in *Gaceta*	Votes not reported in *Gaceta*	All votes
Vote type			
Passage	68.5	3.5	58.2
Amendment	24.4	9.2	22.0
Procedural	5.3	30.1	9.2
Position taking	0.8	21.2	4.0
Private bill	0.2	29.9	4.9
Presidential trip	0.8	6.1	1.6
Total	100%	100%	100%
(N)	(2,439)	(458)	(2,897)

Table 1: *Gaceta*-reported and unreported vote breakdown, 57th–61st Legislatures. See text for categorization criteria.

to every regular or extraordinary session of the lower chamber of the 57th (1997–00) through 59th (2003–06) Legislatures. We parsed the text with a machine-assisted procedure that searches for the formal announcement of a roll call. This procedure and its accuracy are described with detail in the appendix. For the 60th (2006–09) and 61st (2009–12) Legislatures, the full list of votes was available from the Parliamentary Statistics Information Service (INFOPAL), which was created in 2006. To simplify discussion in the rest of the paper, we henceforth adopt the label "Gaceta votes" for those roll-calls only published in *Gaceta*, "missing votes" for those votes excluded from Gaceta, and "all votes" for all roll call votes, regardless of where or whether they were published.

Table 1 compares the universe of 2,897 votes in the 1998–2012 period to the 2,439 *Gaceta*-reported—*Gaceta* missed nearly 16 percent of all roll calls. Votes are grouped in six mutually exclusive categories. Every new bill in the Mexican Congress is scheduled by forward-agenda (see Heller and Weldon 2003). It is first paired against the status quo in a general vote or *voto en lo general*. If a majority votes in favor in general, the new bill may then be subject to amendment in particular votes, or *votos en lo particular*. The 'passage' category includes general votes on bills, nearly always committee reports, but sometimes direct proposals when rules are suspended by extraordinary procedures.

The information on Table 1 reveals substantial selection effects in the types of bills that are published by *Gaceta*. First, final passage votes are modestly overrepresented in the *Gaceta*: they represent 68.5% of published votes versus 58.2% of all votes. Committee reports tend to be known beforehand, so they also tend to be present in the *Gaceta*: passage votes represent only 3.5 percent of the *Gaceta*-missing, mostly committee reports placed at the last minute in the day's order, but also cameral decrees (e.g. the chamber's budget) that are directly sent by the Junta de Coordinación to the floor, and negative reports to reject one or more proposals simultaneously.

Amendment votes are also overrepresented in published votes. The 'amendment' category includes particular votes on bills that cleared the general vote stage. Amendments introduced in committee are known ex-ante when the *Gaceta* is prepared. But amendments can also be introduced when the Junta negotiates the day's order, and when this happens at the last minute, they can be missed by the day's *Gaceta*. Rare amendments introduced directly in the floor, when the rules are suspended by supermajority vote, also tend to be missed. Amendments are a sizeable group of votes (22 percent) yet most are known ex-ante, making them roughly one-tenth of the unreported.

Only about half of procedural votes were published in the *Gaceta* during this period. The 'procedural' category includes votes on how to structure the legislative process—amendments to standing congressional rules, motions to suspend rules and adopt extraordinary procedures, motions to accept amendments not listed in the committee report, and so forth—and other less substantive procedural matters. Since these tend to occur spontaneously in the floor, most are missed by the *Gaceta*.

The remaining three categories are severely underrepresented in published votes. 'Position-taking' are votes about whether or not to immediately include a point of agreement (*punto de acuerdo*) in the day's order. While points of agreement may involve some important cameral decree, those included in this category are position-taking exercises, such as asking the Finance Ministry to grant interest-free credits to hurricane victims or lamenting a remark by the president that may have hurt the Cuban regime's feelings. While the point of agreement itself is decided by voice vote,

whether or not to modify the order to take such a vote requires a super-majority and, therefore, a roll-call. More than a fifth of votes unreported are of this type. 'Private bill' and 'presidential trip' authorize citizens to work for and accept honors from foreign governments or presidents to leave the country, respectively. These issues tend to be ignored by the *Gaceta*. For scholars, these last categories might seem less important—but they may well be items of contention that define legislative conflict lines.

Overall, from a descriptive standpoint, the types of votes unreported by the *Gaceta* are markedly different from those it reports. They are also numerous. Is analysis relying solely on *Gaceta* reports inevitably biased? Will the selection mechanisms employed by *Gaceta* affect conclusions of legislative scholars? The remainder of the paper investigates the extent to which this is problematic for roll-call analyses of the Chamber of Deputies.

2 Analysis

Our key empirical question is whether or not *Gaceta*-published roll call votes are a representative sample of the broader vote population. If that sample is representative of the larger population, then scholars can study the critical transition period of the Mexican Congress using these resources. If the sample is not representative, additional data collection and/or sample adjustments may be needed.[4]

In the most abstract framework, each vote can be considered a draw from an M-dimensional categorical distribution. M in this case is the number of deputies, and the number of categories is the number of possible actions available to legislators on a roll-call vote—yea, nay, abstain, absence, and so on. The dimensions are not independent, except under odd assumptions[5]. Using this framework to test whether *G*aceta votes are a subset of all votes would require a massive dataset—far larger

---

[4]Of course, even if the *Gaceta* sample is representative, there is still information lost by virtue of having a smaller dataset. Should the universe of all votes for that period ever become available, that census would be preferred to any sample.

[5]Independence would mean that there are no spatial or coalitional patterns to behavior

than that provided by several years' of roll-call votes.

As an alternative, we focus on quantities of interest to legislative scholars: cohesion and estimated ideal points. While there are other potential quantities of interest, it is extremely rare that users of roll call votes do not focus on these two key quantities. Our tests, then, will all ask the same question: do quantities derived from *Gaceta* votes differ significantly from quantities derived from the universe of all votes? Or are *Gaceta* votes representative of the universe of all votes?

The analyses that follow address our central questions using nonparametric permutation analysis. We take repeated random samples of size n of votes from the universe of N roll call votes in the legislature. For each of these random samples, we calculate a quantity of interest, compare the distribution of permutations with the actual value of the same measure using *Gaceta* votes. If values calculated from *Gaceta* votes are typical of random samples from the broader population, then we have evidence that the published votes can be used for analysis. If results from *Gaceta* votes are not typical, then we have evidence that the selection mechanisms at work should be accounted for in scholarship on the Mexican chamber of deputies.

We proceed on two fronts, separately examining data from 1998–2006, and 2006–2012. For 1998–2006, we compare aggregate vote statistics from the *Gaceta* with the universe of all votes identified in the *Diario*. This was the critical time period from which votes are missing—votes that were not published by *Gaceta* were not published anywhere else. However, although complete details of missing votes were never published, the *Diario* does list summary statistics for these votes: the number of deputies that cast votes, and the total numbers of yeas, nays, and abstentions. Consequently, although we cannot calculate party cohesion or deputy ideal points, we can at least ask whether the typical *Gaceta* votes have similar patterns of yea-nay divisions, participation, and abstention when compared with all votes. Our results find that, for the most part, *Gaceta* votes appear to be a representative sample of all votes. The most important exception is that the *Gaceta* sample contains more unanimous vote than the *Diario*, especially in 2004. This is unfortunate, since these unanimous votes contain essentially no information for ideal point scaling. However,

	All	Gazeta	Unpublished
Mean Cohesion	0.77	0.8	0.69
Mean Votes Cast	388	392	377
Number of Votes	1511	1116	395

Table 2: Comparison of Published and Unpublished Votes

the second part of our analysis provides assurance that this is not a problem. In addition, the number of missing votes during this period is relatively small—under 20% of all votes.

In the second time period, we do have information on every roll-call vote. During this period there are no missing data, because all roll-call votes are available on the Camara's new informational website. At the same time, *Gaceta* continued its previous practice of publishing only the subset of scheduled votes. We exploit this difference to see whether estimated ideal points from Gaceta votes are different from the universe of all votes. Again, our most important finding is that Gaceta votes appear to be typical of all roll-call votes, and scholars conduct research on the critical transition period without fear of biased results.[6]

2.1 Comparing vote tallies 1998–2006

We begin with some summary statistics, examining overall cohesion on *Gaceta*-published and unpublished votes. Mean cohesion here is defined as $c = \frac{1}{m}\Sigma_{i=1}^{m}\frac{|y_i - n_i|}{(y_i + n_i)}$, where m is the number of roll-calls, y_i is the number of yes votes cast on bill i, and n_i is the number of no votes cast on bill i.[7] This provides a measure of conflict or controversy on votes. Bills with unanimous opposition or unanimous support have a cohesion score of "1"; those that divide the legislature in half will have a cohesion score of "0" and intermediate cases range between these extremes. Table 2 shows summary statistics for cohesion of all votes, Gaceta votes, and missing votes.

---

[6]Obviously, a weakness in our strategy is that for the earlier period in question—when only a subset of votes is available—we cannot compare available data with the true and complete population of votes. At the same time, there would be no problem or issue to resolve if we could make that comparison—it would imply that we had the complete dataset. We did attempt to collect such data directly from the National Congress, with no success.

[7]See Rice (1925) for more details.

We draw attention to the nature of the cohesion scores we are calculating herein. Note that the term "cohesion" is often used to refer to subgroup cohesion - that is, the Rice cohesion score is often calculated separately for different groups within a legislature, comparing cohesion across parties, or across legislators with farming or labor constituencies, as in Rice' original application. In this case, we are calculating an overall cohesion score for the entire legislature, in other words, how often did *all* the legislators vote together, regardless of party or other group memberships. Another way to think about this is that $1 - C$ would measure how close the vote total was to a perfect split ($1 - C = 1$) or do a vote that was not close at all, with all legislators voting the same way ($1 - C = 0$).[8]

Returning to the data, the overall pattern in Table 2 is that missing votes have lower overall cohesion, or are more controversial on average, than *Gaceta* votes. Mean cohesion for *Gaceta* votes is 0.80; mean cohesion among missing votes is 0.69. About half of this difference is due to a difference in the number of unanimous *Gaceta* votes when compared with missing votes. Twenty-nine percent of missing votes are unanimous, compared with 46% of *Gaceta* votes. Many methods would require discarding these unanimous votes, and were one to do so, the difference between *Gaceta* and missing votes shrinks substantially to 0.07 (0.63 for *Gaceta*, 0.56 for missing votes, not shown). This difference is substantively small, and may reflect increased party-line voting on less publicized bills.

Many scholars of Mexican legislative politics also analyze abstentions and absences, often coding them as if they were "no" votes. We disagree with this coding, but given this common practice, we must consider potential differences between *Gaceta* and *Diario* votes in this regard. As it turns out, the two samples of votes have virtually identical rates of abstention and absences. The mean number of abstentions on *Gaceta* votes is 7.0; on *Diario* votes it is 6.9. The mean number of absences on *Gaceta* votes is 97; on *Diario* votes it is modestly higher, 116. In other words, simple comparisons of cohesion, rates of abstention and absences reveal small differences between *Gaceta* and *Diario* votes. These patterns of difference further suggest no

---

[8]We thank an anonymous reviewer for suggesting this clarification.

| | | Cohesion | | |
Year	Overall	*Gaceta*	Missing	Difference
1998	0.63	0.62	0.64	-0.022
1999	0.62	0.61	0.66	-0.052
2000	0.71	0.77	0.60	0.171
2001	0.88	0.89	0.84	0.055
2002	0.84	0.87	0.78	0.086
2003	0.76	0.80	0.69	0.112
2004	0.79	0.84	0.60	0.239
2005	0.81	0.84	0.67	0.170

Table 3: Mean Cohesion by Sample and Year. Diarios de debates for 2001 and 2002 have not been scanned by authors. If the paper is accepted, we will complete the data for these years.

important differences between *Gaceta* and *Diario* votes. On abstentions, which are more likely to be strategic and costly, the rates are virtually identical. On absences, which may just reflect legislators' other priorities and meetings, the difference is consistent with the scheduling differences: *Gazeta* votes are higher-profile, pre-scheduled votes, and thus less likely to come to the floor when many legislators are absent.

Table 3 shows mean cohesion for *Gaceta*, Missing, and All votes by year for the critical period of 1998–2005, as well as their difference. There are modest differences between Missing and *Gaceta* votes, but patterns are not consistent in terms of direction. Most differences are small, though the gap in 2004 is larger—.24. This means that the typical published bill in 2004 had about a 10% larger majority coalition than missing bills; most of this difference is due to an exceptionally large number of unanimous votes that year. Fully 56% of *Gaceta* votes were unanimous in 2004, versus just 21% of missing votes—this difference is much larger than that observed in any other year.

Are these modest differences significant? Figure 1 shows results from a permutation analysis, and shows that Gaceta votes appear to be typical of all roll-call votes. In the figure, the solid line shows overall cohesion for all votes; the dashed line shows average cohesion just for the *Gaceta* votes, and the histogram shows the distribution of cohesion random samples of roll-call votes of size n from all votes. In other words,
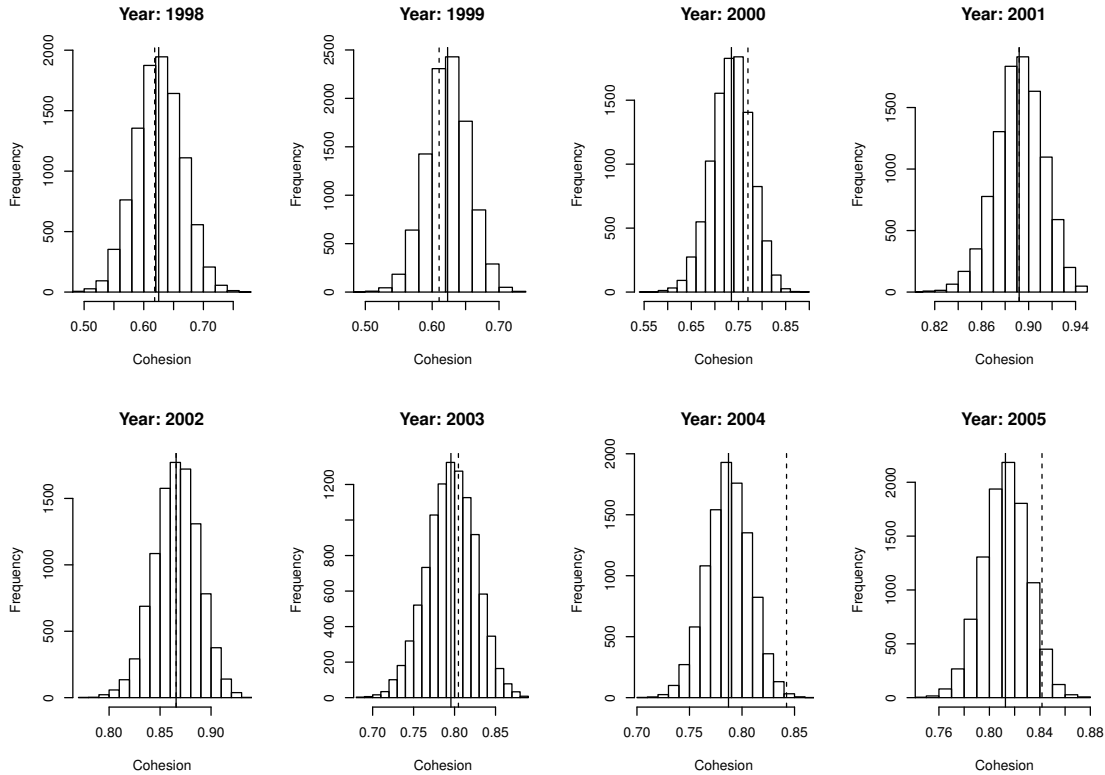
Figure 1: The histograms shown the range of mean overall cohesion on all votes in the respective year when taking a random sample of size equal to the number of reported *Gaceta* votes that year. The dashed line shows the actual *Gaceta* cohesion, and the solid line shows overall cohesion on all votes.

the distribution shows what kind of cohesion scores we would see if we were just randomly choosing votes to publish and other votes to exclude. When the dashed line is right in the middle of the distribution, it means that the cohesion of Gaceta votes is typical of a random sample from all votes. When the dashed line is above or below the histogram, it means the Gaceta votes are significantly different from the typical random sample.

In every year except 2004, the *Gaceta* votes' cohesion average is right in the heart of the simulated distribution. In other words, in terms of overall cohesion of votes, the *Gaceta* votes look exactly like a typical randomly chosen vote from the population of published and unpublished votes, and we cannot reject the null that these votes are indeed a random sample from that population. For 2004, the actual *Gaceta* sample is in the 99th percentile of permuted values—it is not the maximum, but is larger than

the majority of all permuted samples. This particular result appears to stem from a high number of unanimous votes in 2004; restricting the sample to non-unanimous roll-calls leads to a null result for 2004, as for all other years.

The permutation analysis supports the position that the published *Gaceta* votes are not fundamentally different from a typical representative sample from all votes, but requires several caveats and additional comments. First, we again note that the limited data for this first period (1998–2005) limits our analysis to overall cohesion on votes—essentially a measure of how controversial votes are—and it is of course possible that voting patterns across parties or other subgroups are not so consistent with the aggregate patterns. Second, the fact that there are relatively more unanimous votes in the *Gaceta* sample means that some controversial votes are being lost by using the *Gaceta* subset. Third, the differences in abstentions suggests that it is possible that different types of deputies are abstaining or not attending *Gaceta* sessions when compared with missing sessions. Finally, we repeated this entire analysis, coding all abstentions and absences as if they were "no" votes. When treating non-votes as if they were "no" votes, the results were essentially unchanged, with no evidence of any systematic differences between *Gaceta* and all votes[9]. With this in mind, we turn now to a spatial analysis of *Gaceta* versus all votes from the 2006–2011 period.

2.2 Ideal point estimates 2006–12

The second part of our analysis uses data from 2006–2012. During this period, the *Gaceta* continued to publish a subset of votes, but the new INFOPAL website began publishing the vote universe. Whereas for the first period we only had vote aggregates and could only compare overall cohesion, for this second period we have all roll call votes. We can therefore explore the impact of *Gaceta*'s selection mechanisms on estimated ideal points. We begin our inspection of roll call votes with a relatively straightforward approach. For legislatures with all votes available (the 60th and 61st,

---

[9]Of course, raw total cohesion scores were affected by coding all abstentions and absences as "no's", but trends and significance did not change, with one exception. After recoding abstentions and absences as "nay" votes, the difference between *Gaceta* and *Diario* votes had an effective p-value just under .05 when including non unanimous votes.

	60th (2006–09)		61st (2009–12)	
	Gaceta-	All	*Gaceta-*	All
Vote type	reported	votes	reported	votes
Uncontested	259	293	295	321
Near-uncontested	65	82	138	150
Scalable	160	215	234	267
Total	484	590	667	738

Table 4: Scalable and unscalable roll call votes, 2006–12

spanning the 2006–12 years), we estimate deputies' ideal points twice: once using all minimally-contested votes; then using *Gaceta*-published, minimally-contested votes. To the extent that the *Gaceta*'s selection process returns a non-representative vote sample, differences should manifest in a direct comparison of legislator ideal points estimated with each set of votes.

A two-dimensional item-response model was specified, estimating deputies's ideal points and other parameters of interest via Markov chain Monte Carlo (MCMC) (Clinton, Jackman and Rivers 2004; Poole 2005:116). As is standard, unanimous and minimally-contested roll calls (those with a minority of 2.5 percent or less, excluding abstentions and absences) were dropped from the dataset due to the little or no relevant information they offer; table 4 offers a summary and figure 6 in the Appendix shows the proportion of contested votes by vote type. For the same reason, deputies who either never pledged or who voted 'aye' or 'nay' in less than 10 percent of all roll calls in the legislature were dropped from analysis. The first dimension explains almost 95 percent of the individual votes and describes a government / opposition split in Congress. The second dimension is a combination of a continuing pressure for liberalization of the political system and occasional issues exclude at least one of the small parties from the winning coalition (Robles Peiro 2009).

Ideal points were identified by assigning semi-informative priors to major party deputies at polar coordinates: PRI at the North, PAN at the Southeast, and PRD at the Southwest.[10] Two chains were iterated ten thousand times each. Parameters

---

[10]Prior distributions were defined thus. PAN, PRI, and PRD deputies were given the following semi-informative priors: $\begin{pmatrix} x \\ y \end{pmatrix}_{\texttt{PAN}} \sim \mathcal{N}\begin{pmatrix} 1, \frac{1}{4} \\ -1, \frac{1}{4} \end{pmatrix}$; $\begin{pmatrix} x \\ y \end{pmatrix}_{\texttt{PRI}} \sim \mathcal{N}\begin{pmatrix} 0, \frac{1}{4} \\ 1, \frac{1}{4} \end{pmatrix}$; and $\begin{pmatrix} x \\ y \end{pmatrix}_{\texttt{PRD}} \sim \mathcal{N}\begin{pmatrix} -1, \frac{1}{4} \\ -1, \frac{1}{4} \end{pmatrix}$. All other
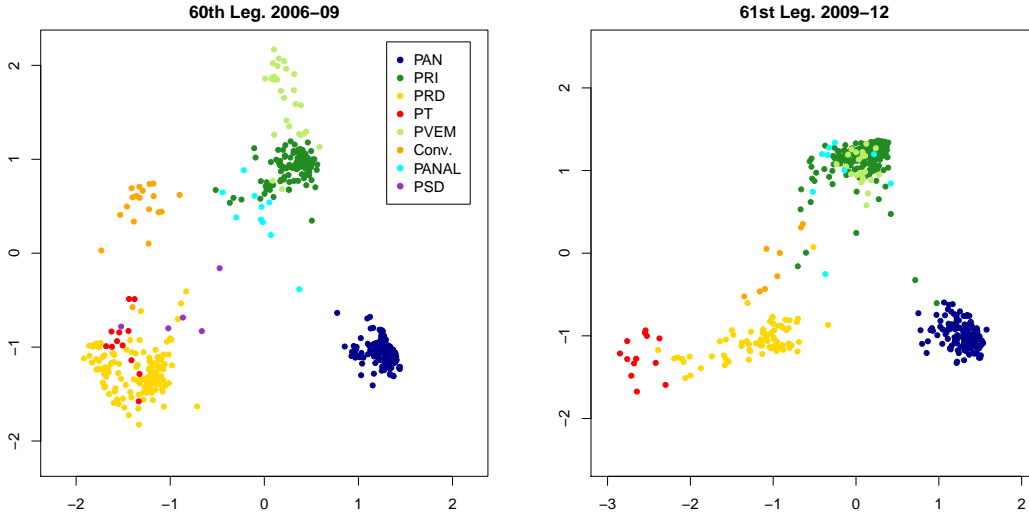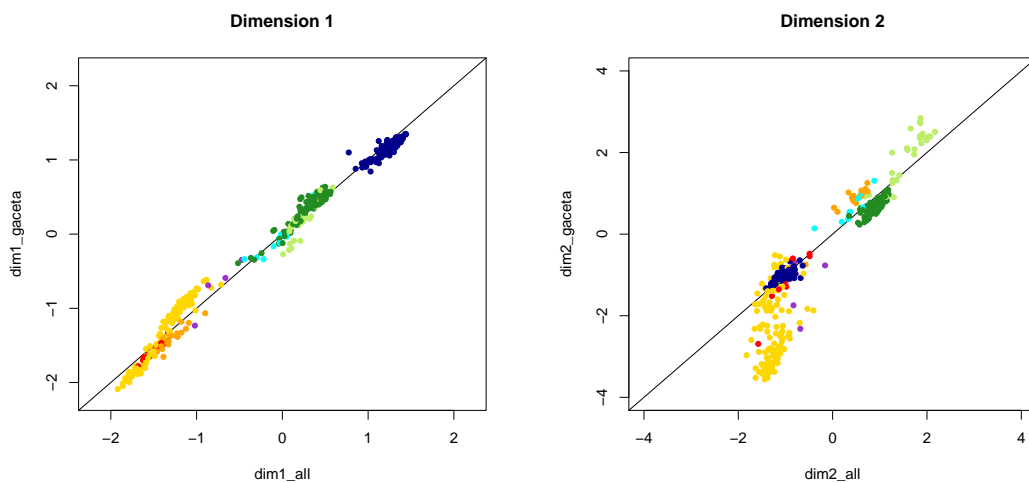
Figure 2: Ideal point estimates using all votes in two legislatures

had reached a stable stage after the first half of those iterations; every twenty-fifth simulation of the second half was sampled in order to estimate parameters' posterior distributions, which we report here.

Estimates, portrayed in Figure 2, are consistent with previous ideal point estimations of the Mexican Chamber of Deputies (Rosas and Langston 2011), major party deputies manifesting slight dispersion around the vertices of an approximate isoceles triangle reminiscent of the priors assigned. The exception is the PRD, about half of which pulled further west in the 60th Legislature, en bloc with most of the minor PT: this was the López Obrador faction, who did not concede after the 2006 presidential election and began a vocal anti-PAN campaign while the rest of the party often negotiated with other parties. The López Obrador faction lost control of the party leadership, leaving it with a diminished, but more radical contingent in the 61st Legislature. That legislature, installed after Felipe Calderón's midterm election, also saw the PRI become the plurality in the chamber, allying with two minor parties, the Green (PVEM, their common pre-electoral ally since 2001) and PANAL, to form a majority against the president. They achieved remarkable cohesion.

Figure 3 compares, one dimension at a time, ideal points estimated with all votes

---

deputies were given non-informative priors $\binom{x}{y}_{\texttt{rest}} \sim \mathcal{N}\binom{0,10}{0,10}$.

Part A. 60th Legistature, 2006–09

Dimension 1

Dimension 2

Part B. 61st Legistature, 2009–12
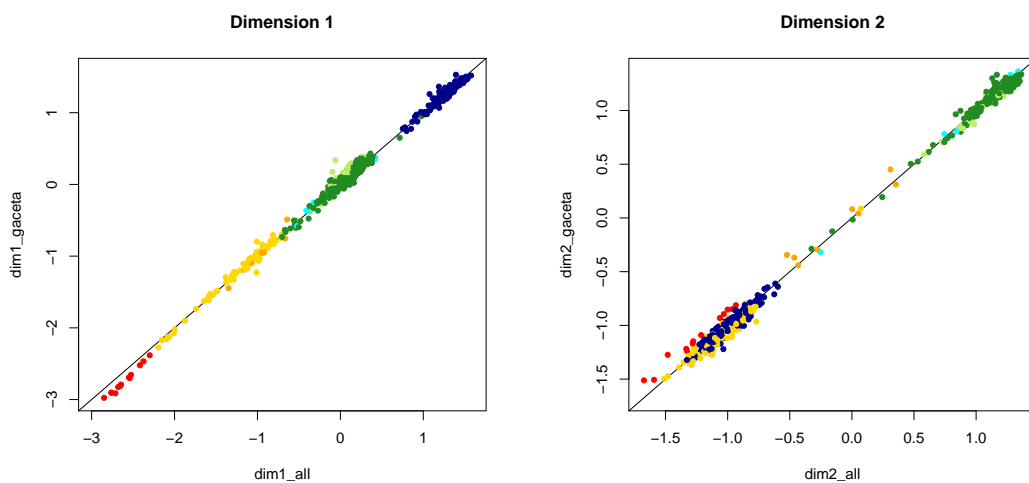
Dimension 1

Dimension 2

Figure 3: Dimension-by-dimension comparison of ideal points estimated with all votes (x-axis) and *Gaceta*-reported votes (y-axis)

to those estimated with *Gaceta*-reported votes only. With one exception, the match is near perfect, legislators from all parties closely aligned with the 45° line indicating exact coordinate match. The exception is the second dimension of the 60th Legislature, where some parties, notably a large group of the PRD, appear somewhat more radical when estimated with *Gaceta*-only votes than when estimated with all votes. This is odd, given that *Gaceta* excludes a large number of position taking as well as amendment votes, which ought to exert centrifugal, not centripetal force.[11] Yet even in this case, the general match between ideal points estimated with *Gaceta*-reported votes only is not very different from those estimated with all votes.

2.3 Permutation analysis for summary statistics

Our next approach to evaluate the effects of incomplete roll calls replicates the permutation analysis we used above with ideal points instead of cohesion scores. We took 99 random samples of size 50 from the whole population of scalable votes. It would have been preferable to take samples the same size as the *Gaceta*'s scalable votes—160 and 234 for the 60th (2006–09) and 61st (2009–12) legislatures, respectively—but computation time was a major issue.[12] To compare random sample estimates with *Gaceta*'s, each legislator's ideal point scaled with *Gaceta*-only votes was subtracted, dimension-by-dimension, from each of the 99 coordinates. If ideal points scaled with the random and *Gaceta* samples were identical, this difference would be nil, positive (negative) values indicating sample coordinates larger (smaller) than *Gaceta* coordinates. Figures 4 and 5 rank deputies according to their *Gaceta* scores, and the horizontal axis reports the comparison across random samples. The thick and thin lines report the 50 and 95 percent intervals.

For the case of the 60th legislature, Figure 4 shows that only a handful of deputies'

---

[11]As observed in the cohesion analysis, above, *Gaceta*'s selection mechanism results in a higher proportion of uninformative votes—votes that are unanimous, or nearly unanimous. It is likely that the PRD's more moderate positions were on non-*Gaceta* bills, and dropping these drives the results.

[12]About 24 hours on average were required to complete the routine for each legislature using parallel computation, one third of the random samples at a time, in a multicore processor. With sample sizes between three and five times larger, computation time would increase four-fold on average.
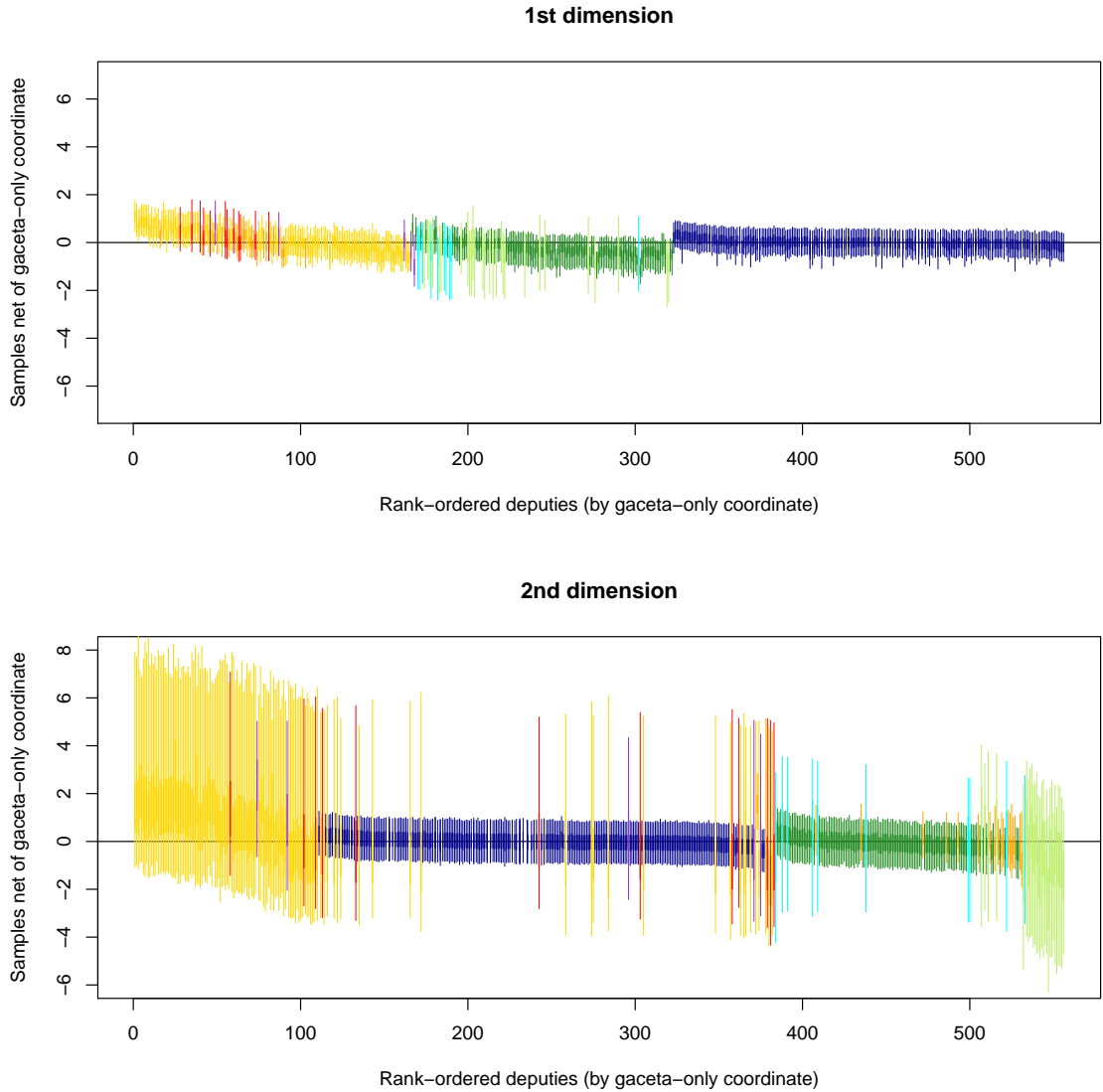
1st dimension

2nd dimension

Figure 4: Samples of *Gaceta*-reported votes against theta, the point prediction of deputies's ideal scores using all votes, 60th Legislature (2006–09)
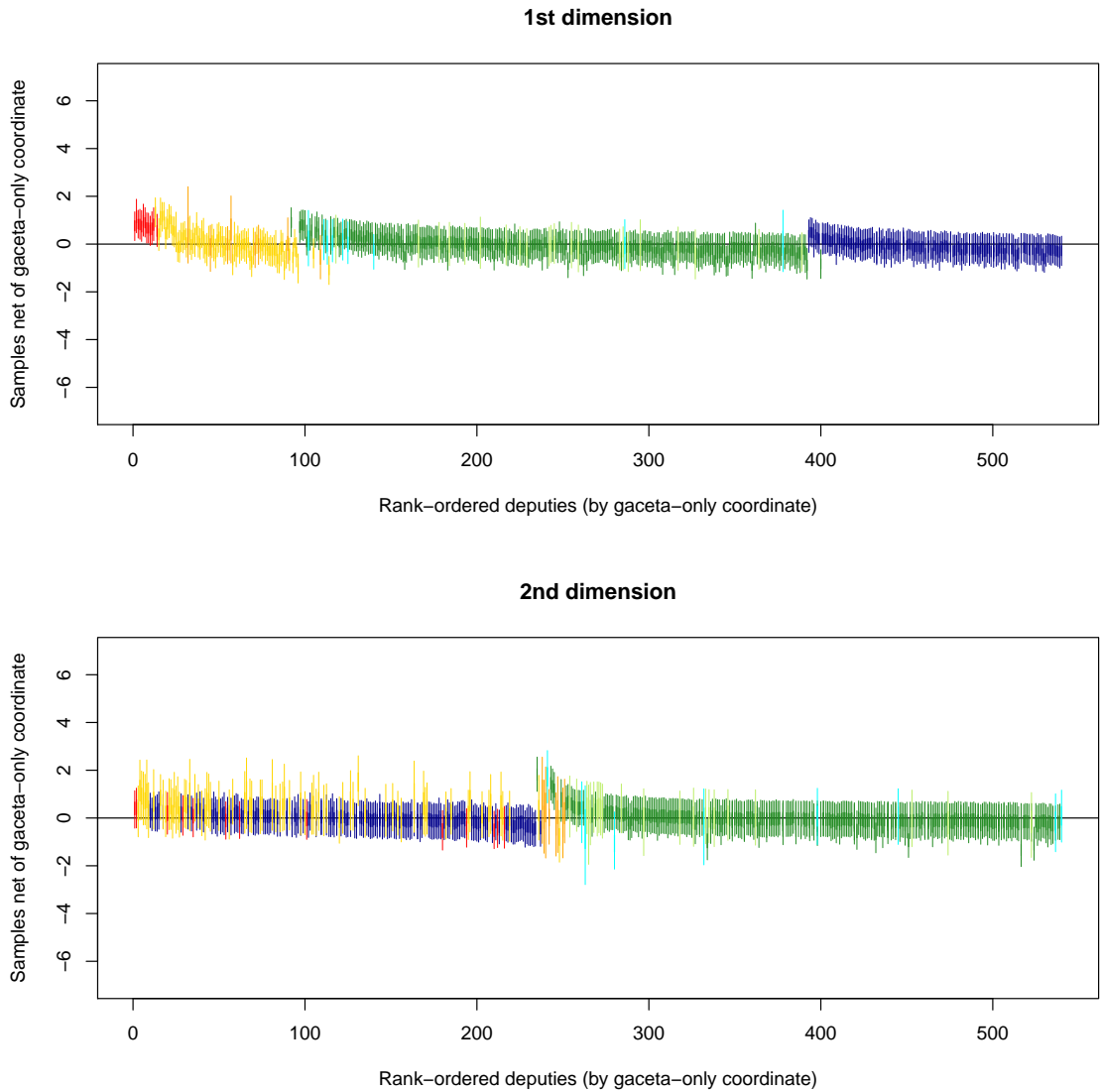
1st dimension

2nd dimension

Figure 5: Samples of *Gaceta*-reported votes against theta, the point prediction of deputies's ideal scores using all votes, 61th Legislature (2009–12)

50 percent intervals exclude zero—the point where no difference between the sample and theta is manifest as significant—in the first dimension. More such cases are observable in the second dimension, but they remain relatively few and only marginally spread from zero. Random sampling returns more volatile ideal point estimates for some parties, notably the Green Party (PVEM) in both dimensions, small parties and—notably—the PRD in the second dimension. Furthermore, estimates for the PRD deputies aligned with López Obrador are more vulnerable to the sampling of votes in the first dimension. As the top panel reveals, deputies on the left end of the spectrum have a more moderate ideal scores when using the sample of votes from a random sample than when estimations from the whole universe of votes. For the case of the 61st legislature (Figure 5), the variance for the ideal scores on the second dimension is significantly smaller than in the previous legislature. Similar to the 60th legislature, the group of radical deputies identified with López Obrador has more conservative scores in the first dimension when the estimations use a random sample of all votes instead of the *Gaceta* sample. There is a similar effect among the most moderate PRI members on the second dimension.

Some generalizations can be drawn from this analysis. First, PAN and PRI deputies are generally less sensitive to sampling than PRD and small party deputies. Second, extremist deputies appear more moderate when using sample of votes in comparison than when using the whole set of votes. Besides the additional caution when estimating the ideal score for this group of deputies, we find no significant differences in the estimations between the *Gaceta* deputies and the whole set of votes.

3 Conclusions

Mexico's democratic transition has been accompanied by a rise in scholarship on political institutions, including the Congress. Since 1997, no single party has held a majority of seats in the Chamber of Deputies, and Mexican Presidents' dominance of the legislative branch has ended. The result is a newly empowered Congress, which is in the process of developing institutional capacity and professionalism.

Along with learning how to defy the executive power and create procedural instruments for negotiating the enactment of the budget, the public release of the roll-call votes signaled yet another role of the Mexican Congress in the policy-making process. But while the study of congressional politics has increased dramatically, the nature of the available data for analysis deserves examination. In particular, given that not all the roll-call votes from the critical transition period of 1998–2005 have been released to the public, legislative studies using this subset of all roll-call votes may yield biased estimates. The omission of many votes in publicly available data raises the question on whether these omissions are due to a lack of resources within the Congressional staff or to a lack of transparency in Congress. This paper indirectly tests both hypotheses and concludes that there is no evidence of bias in the set of publicly available votes.

Examining legislative cohesion and ideal scores, this study explored whether we can consider the roll-call votes available online to be a random sample of the broader population of votes. The empirical evidence shows that, when using the subset of published roll-call, the most useful summary statistics of legislative behavior are not significantly different than when using both published and unpublished information. Briefly, the our permutation analyses show that for cohesion scores and for ideal points, estimates are very similar using either set of roll-call votes. A key implication of this paper is that scholars can use the data available online without worrying about the construct validity of their analyses. The quality and variety of data for legislative behavior in Mexico is a growing field. The replacement of the old electronic board and the development of INFOPAL bring an opportunity to scholars to start analyzing the new role of Congress in the Mexican policymaking process.

We repeat, however, our earlier qualification: we have examined evidence from estimated ideal points, which are the most widely used took for legislative studies today. We also examined aggregate roll-call cohesion for the period where data was more limited. There are, however, many other potential quantities of interest that we have not—and in some cases—could not explore. We encourage scholars interested in other secondary quantities of interest to verify the properties of their subsets using

the data we have released on our websites.

Incomplete data problem is a common event that many developing democracies face while making accessible electoral, legislative, or budgetary information. The problem of scholars, consultants, and government workers when using incomplete data is that inferences may be biased and inaccurate from the reality. Moreover, it is difficult to assess whether the origins of the missing data are due to lack of resources rather than of intention. The tools proposed in this paper may be useful when assessing the representative nature of the available data and the type of inferences that can be made from such data.

Appendix

We describe the steps taken to implement a machine-assisted routine to trace roll call votes taken in the floor. We began by determining, from the chamber's web page (`http://cronica.diputados.gob.mx/`), the dates when the chamber was in regular or extraordinary session in the 57th to 59th Legislatures . A connection to the file containing each day's *Diario de los debates* (or the *versión estenográfica* for the rare instances when the connection to the Diario failed) was opened, its lines read into an R object we named `dd`. We then parsed the text lines in `dd` with the help of regular expressions (Friedl 2006) in search for the announcement of a roll call. Deputies's individual votes are not available in the Diario, but the motion under consideration and aggregate ayes and nays are always reported, providing us with the list of floor votes that could then be compared to those reported by the *Gaceta*.

A sample of Diarios was read humanly to detect patterns in use when calling a roll call vote. After recognized deputies have spoken and the matter is deemed sufficiently debated, the Chamber Secretary summons deputies to vote by making a protocollary announcement reading, more or less, as: "Ábrase el sistema electrónico de votación y háganse los avisos a que se refiere el artículo 144 del Reglamento de la Cámara de Diputados". Before the 2010 amendment to the chamber's rules, the article in question used to be 161. Slight variations to the statement, such as skiping some

word or adding some adjective, are frequent, so the search pattern was calibrated through trial-and-error. The search for three patterns pointed to all votes with one exception discussed below. The three patterns used are the following:

```
pat1 <- grep(dd, pattern="\\b[ÁAáa]bra\\w*\\b la votaci[oó]n")
pat2 <- grep(dd, pattern="\\b[ÁAáa]bra\\w*\\b el sistema [ed][le][e ][cv]
                         [to][rt][óoa][nc]i[coó][on]")
pat3 <- grep(dd, pattern="[Aa]rt[íi]culo 1[64][14].*de.*[Rr]eglamento")
```

The routine returned hits for many events that were not roll call votes, such as quorum motions—the most common—or any mention to article 141, 144, 161, or 164 of any law. These were dropped from the list, retaining roll call votes only.

We verified the accuracy of the procedure by downloading Diarios between February 1 and December 2, 2010, inclusive, and running the regular expression in search for votes. These were sessions correspond to the 61st Legislature, for which the roll call universe of 167 votes is given by INFOPAL. The machine-assisted routine detected all votes but one—more than 99 percent accuracy. The vote missed belongs to the the November 15–16 Diario, when the yearly federal budget was discussed and passed in a 21-hour long session involving endless discussions, recesses, and procedural voice votes to attempt to introduce amendments to the committee report. In one instance, a procedural roll call vote was called without the Chamber Secretary voicing the protocolary announcement.
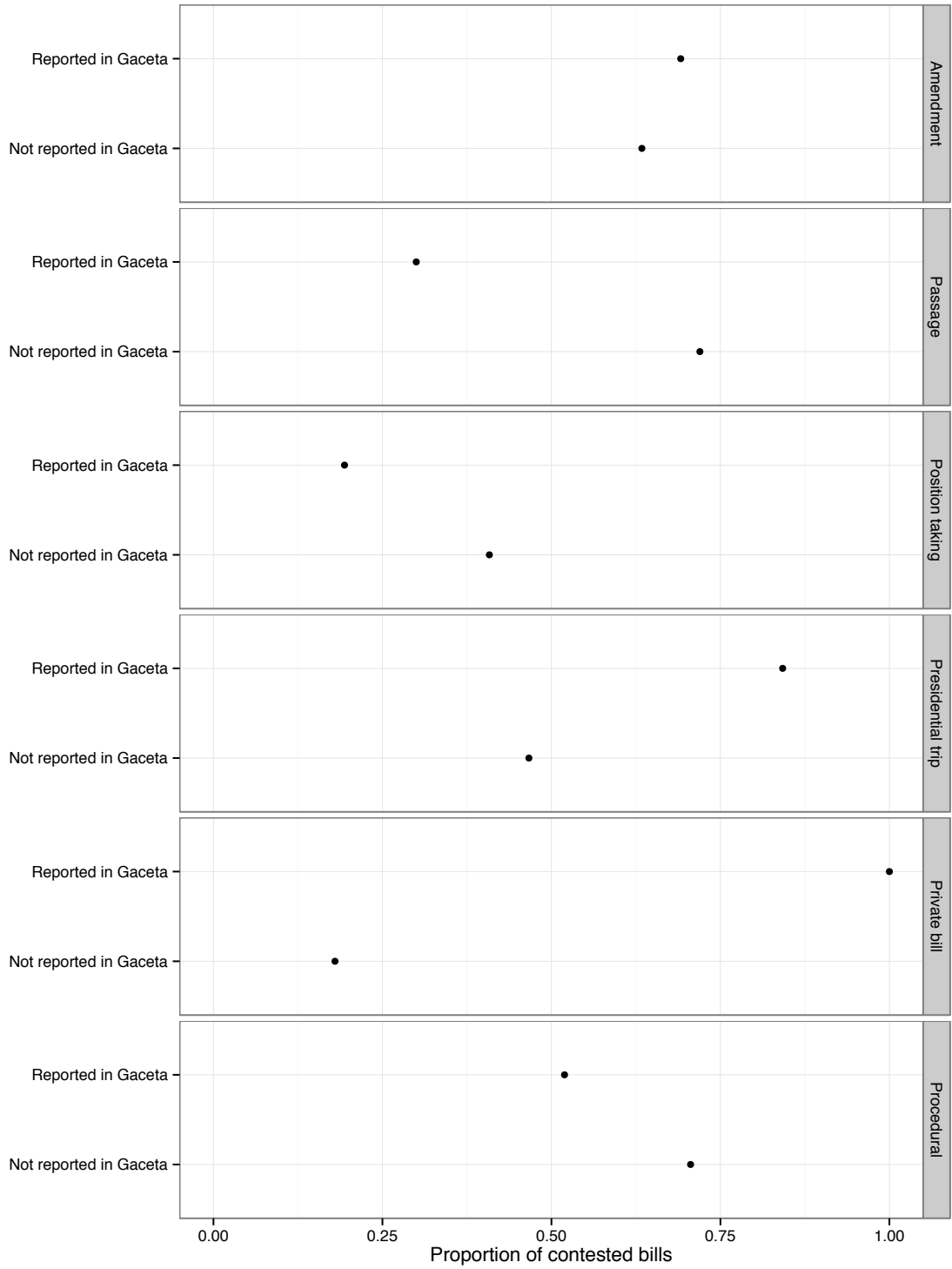
Figure 6: Proportion of contested bills in Congress by publication in *Gaceta* and type of bill

References

Cantú, Francisco and Scott Desposato. 2012. "The New Federalism of Mexico's Party System." *Journal of Politics in Latin America* (2):3–38.

Casar, Mario Amparo. 2002. Executive-Legislative Relations: The Case of Mexico (1946-1997). In *Legislative Politics in Latin America*, ed. Scott Morgenstern and Benito Nacif. Cambridge University Press.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *The American Political Science Review* 98(2):355–70.

Cosío Villegas, Daniel. 1981. *El sistema político mexicano.* Mexico City: Joaquín Mortiz.

Domínguez, Jorge I. and James A. McCann. 1996. *Democratizing Mexico.* Baltimore: The John Hopkins University Press.

Friedl, Jeffrey E.F. 2006. *Mastering regular expressions.* Third ed. Sebastopol, CA: O'Reilly.

González Casanova, Pablo. 1965. *La democracia en México.* Mexico City: Era.

Heller, William B. and Jeffrey A. Weldon. 2003. Reglas de votación y la estabilidad en la Cámara de Diputados. In *El Congreso Mexicano después de la alternancia*, ed. Luisa Béjar Algazi and Rosa María Mirón Lince. Mexico City: Asociación Mexicana de Estudios Parlamentarios and Instituto de Investigaciones Legislativas del Senado de la República pp. 85–119.

Morgenstern, Scott. 2003. *Patterns of Legislative Politics: Roll-Call Voting in Latin America and the United States.* New York: Cambridge University Press.

Morgenstern, Scott and Benito Nacif. 2002. *Legislative Politics in Latin America.* New York: Cambridge University Press.

Nacif, Benito. 2002. Understanding Party Discipline in the Mexican Chamber of Deputies. In *Legislative Politics in Latin America*, ed. Scott Morgenstern and Benito Nacif. Cambridge University Press.

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting.* New York: Cambridge University Press.

Rice, Stewart A. 1925. "The Behavior of Legislative Groups: A Method of Measurement." *Political Science Quarterly* 40(1):60–72.

Robles Peiro, Gustavo Adolfo. 2009. Dimensiones Espaciales de Votacion Legislativa en la Camara de Diputados de la 60 Legislatura del Congreso de la Union, 2006-2009. Master's thesis Departamento de Ciencia Politica, ITAM Mexico City: .

Rosas, Guillermo and Joy Langston. 2011. "Gubernatorial Effects on the Voting Behavior of National Legislators." *Journal of Politics* 73:477–493.

Weldon, Jeffrey A. 1997. El presidente como legislador, 1917–1934. In *El Poder Legislativo en las décadas revolucionarias, 1908–1934*, ed. Pablo Atilio Piccato Rodríguez. Enciclopedia Parlamentaria de México (Serie I, vol. I, tomo 3) Mexico City: Instituto de Investigaciones Legislativas, Cámara de Diputados, LVI Legislatura pp. 117–45.

Weldon, Jeffrey A. 2005. "The Fall 2004 Term of the Mexican Congress (Mexican Congressional Report Series)." Center for Strategic and International Studies, Washington DC.