

Consideraciones metodológicas para estudiantes de política legislativa mexicana

Sesgo por selección en votaciones nominales

Francisco Cantú, Scott Desposato y Eric Magar*

Resumen: El estudio de la conducta legislativa en la Cámara de Diputados durante el periodo de 1998 a 2006 presenta un problema potencialmente serio: no todos los votos han sido publicados en la *Gaceta Parlamentaria*. Analizamos la naturaleza de estos datos en aras de explorar la representatividad de la muestra de votos disponibles y las posibles repercusiones de este problema en los análisis legislativos existentes. Para esto, aprovechamos la aparición de una página de Internet que registra la totalidad de los votos de la Cámara de Diputados desde el año 2006 en paralelo con el sistema existente desde 1998. Mediante la exploración de los mecanismos que generan la omisión de votos y la comparación de distintas estimaciones del comportamiento legislativo, concluimos que los votos no publicados merman la precisión de estimadores de uso común pero no introducen ningún tipo de sesgo. A la par, hacemos pública una base de datos para el estudio del Congreso mexicano.

Palabras clave: votaciones legislativas, Congreso mexicano, problemas de datos incompletos.

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

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

* Francisco Cantú es profesor asistente del Departamento de Ciencia Política de la Universidad de Houston. Philip Guthrie Hoffman Hall 429, Department of Political Science, The University of Houston, Houston, TX. Tel. 772 04 30 11. Correo electrónico: fcantu10@uh.edu. Scott Desposato es profesor asociado del Departamento de Ciencia Política de la Universidad de California, San Diego. Department of Political Science, University of California, San Diego, 9500, Gilman Drive, MC 0521, La Jolla, CA. Tel. 920 93 05 21. Correo electrónico: swd@ucsd.edu. Eric Magar es profesor del Departamento de Ciencia Política del Instituto Tecnológico Autónomo de México (ITAM). Río Hondo, núm. 1, col. Tizapán San Ángel, C. P. 01000, México, D. F. Tel. 56 28 40 00, ext. 3702. Correo electrónico: emagar@itam.mx. Eric Magar agradece a Mariana Meza Hernández y Lucía Motolinía su ayuda para esta investigación, y a la Asociación Mexicana de Cultura, A. C., por patrocinar parte de esta investigación. Los errores y omisiones son responsabilidad de los autores.

Artículo recibido el 19 de diciembre de 2012 y aceptado para su publicación el 23 de abril de 2013.

Traducción: Danielle Stren y Eric Magar.

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.

Keywords: roll-call votes, Mexican Congress, incomplete data problems.

El partido hegemónico relegó al Congreso mexicano al papel de “apéndice del Ejecutivo” durante la mayor parte del siglo xx. El papel secundario de la rama legislativa se vio reforzado por una falta general de interés de los ciudadanos y académicos por seguir el comportamiento de los legisladores. Durante décadas, las votaciones nominales no fueron publicadas sistemáticamente en el *Diario de los Debates* —el equivalente mexicano del *Hansard* en los sistemas Westminster o el *Congressional Record* en Estados Unidos— sólo la cuenta total de votos al final de las votaciones. Esto opacó el proceso legislativo en beneficio del sistema de control del Partido Revolucionario Institucional (PRI).

La transición gradual a la democracia restableció la autoridad constitucional que el Congreso tiene con respecto al Ejecutivo. Con una mayoría máxima de tres quintas partes de la Cámara Baja y una pérdida de competitividad electoral, el PRI necesitó de la oposición para las enmiendas constitucionales del ambicioso programa de reformas del presidente Carlos Salinas de Gortari. Fue en 1997, cuando el PRI perdió la mayoría de los diputados, cuando los académicos empezaron a formular preguntas acerca de la legislatura y de su papel en el proceso de creación de políticas (por ejemplo, Cantú y Desposato, 2012; Morgenstern, 2003; Morgenstern y Nacif, 2002; Rosas y Langston, 2011; Weldon, 1997 y 2005). La reaparición de datos para el estudio del Congreso mexicano pareciera atestiguar su resurgimiento. Recursos tales como las proposiciones e iniciativas de ley presentadas por los legisladores al pleno, los dictámenes de las comisiones y las votaciones nominales son publicados en la página web de los diputados desde noviembre de 1998. De la noche a la mañana empezó a fluir la clase de información común para el análisis del comportamiento legislativo en el mundo y que proporciona al ciudadano un instrumento para hacer respon-

sables a sus representates. No obstante, el caudal de la nueva información adolece de un problema. Las votaciones nominales publicadas representan sólo una porción, grande pero no completa, de todos los votos realizados en el recinto parlamentario. De hecho, en promedio, desde la aparición de la *Gaceta Parlamentaria* y hasta la LXI Legislatura (2009-2012) inclusive, casi 20 por ciento de los votos no fueron publicados. La cifra alcanzó su máximo, 31 por ciento de las votaciones no publicadas, en la LVIII Legislatura (2000-2003), y ha menguado en cada legislatura subsecuente. Esta discrepancia y aparentemente selectiva publicación de las votaciones nominales plantea una serie de desafíos. Si el subconjunto de los votos disponibles difiere sistemáticamente del universo de los votos, el grueso de la investigación existente sobre el comportamiento legislativo del país está distorsionado, sesgando nuestra comprensión de la política legislativa de México en el periodo. Más aún, si tales votos no fueran una muestra representativa de los emitidos en el pleno, los votantes tendrían en sus manos una herramienta deficiente para dar seguimiento a lo que sus representantes hacen en el Congreso. Los periodos de transición tienen importancia especial para los estudiantes de la democratización en lo general, y de la política mexicana en lo particular. ¿Están quizá siendo retenidos los votos más importantes? ¿Ofrecen acaso las votaciones nominales publicadas una perspectiva capaz de despistar, por incompleta, el examen de la naturaleza de la política legislativa durante la transición? El analista necesita saber si puede o no confiar en estos datos y utilizarlos para hacer inferencias acerca de este importante periodo.

Este trabajo examina la naturaleza de los datos disponibles para los estudios legislativos durante una coyuntura crítica para la democracia mexicana: el periodo comprendido entre 1998 y 2006. Nuestra estrategia de investigación saca provecho de dos características de las votaciones nominales reportadas por la Cámara de Diputados. Desde septiembre de 2006, cuando quedó instalada la LX Legislatura, el Servicio de Información para la Estadística Parlamentaria (Infopal) ha publicado la totalidad de las votaciones nominales y ha coexistido junto con la *Gaceta Parlamentaria*, que publica sólo un subconjunto. Esto nos ha dado la oportunidad de contrastar el subconjunto de votos elegido por la *Gaceta* contra el universo a lo largo de las LX y LXI Legislaturas (2006-2012). Segundo, aunque no contamos con todos los votos nominales para el periodo previo (1997-2006), sí pudimos recopilar el resultado agregado de cada votación —cuántos diputados votaron a favor y cuántos en contra— para todas las mociones sometidas a consideración del pleno, publicadas o no en la *Gaceta*. Esta fuente de datos, que obtuvimos

directamente del *Diario de los Debates*, nos permite comparar las características clave, como la cohesión partidista y la tasa de abstención, de los votos publicados y los no publicados. Recurrimos a varios métodos estadísticos para responder si la muestra de votos publicados de 1998 a 2006 es representativa de todos los votos. Nuestro análisis sugiere que la muestra es estadísticamente indistinguible de la población más amplia de votaciones nominales. De cumplirse el supuesto de que los criterios de selección de votaciones no publicadas por la *Gaceta* se mantuvieron sin cambio antes y después de 2006 —un supuesto fundamental de nuestro análisis y que, también demostramos, se sostiene— podemos afirmar que es posible estudiar el periodo 1998-2006 con los datos incompletos sin temor de sesgar las conclusiones. Las implicaciones de nuestro análisis descartan una omisión selectiva de votaciones nominales y explican el problema como el resultado de una falta de recursos y no de opacidad del sistema de rendición de cuentas.

La *Gaceta* y la omisión de votos

La *Gaceta Parlamentaria* fue creada cuando la Cámara de Diputados instaló un nuevo sistema de votación electrónica al principio del segundo año de la LVII Legislatura. Se trata de un acervo, accesible por internet, de documentos e información pertinentes para las actividades enumeradas en el orden del día de cada sesión plenaria. Empezó a publicarse el 10 de noviembre de 1998. Antes de ello, cuando el PRI controlaba sistemáticamente la mayoría en la Cámara, los diputados iniciaban las sesiones sin mucho más que el orden del día. Quien deseara consultar el dictamen de una comisión, los anexos o cualquier información pertinente a una moción específica tenía que solicitarla al personal de la Cámara. Dichas solicitudes, si eran atendidas, solían llegar tarde, incitando la ira de los legisladores de la minoría. Como consecuencia, los diputados a menudo votaban con los ojos “vendados” en el pleno, con escasa información de la sustancia de la legislación.¹

¹ Una sesión ordinaria de la LVI Legislatura, en el ocaso de la hegemonía del PRI, ofrece un ejemplo elocuente. El *Diario de los Debates* sugiere que el PAN y el PRI habrían acordado enmendar el artículo 105 de la Constitución sin la anuencia del PRD. A la 1:55 pm, el diputado del PAN Francisco José Peniche y Bolio hizo la propuesta para dispensar de la segunda lectura y proceder de inmediato a la votación. El diputado del PRD José Jesús Ortega Martínez fue entonces reconocido para expresar su oposición a la suspensión del procedimiento ordinario. “Lo que va a esperar el pueblo es nuevamente que los diputados estemos legislando sobre las rodillas —dijo— y legislar es un decir, porque sólo se nos presentan las iniciativas y prácticamente se nos llama a que votemos a favor y a que votemos en contra [...] El día de hoy a las 12:00 o a la 1:00 de la tarde [...] la

Además de proporcionar acceso a documentos clave para el trabajo de los legisladores, la *Gaceta* también se convirtió en una fuente invaluable para estudiar el Congreso, ofreciendo información detallada del proceso legislativo, como las votaciones nominales, por primera vez en décadas (Casar, 2002; Nacif, 2002; Rosas y Langston, 2011). A pesar de ser amigable para el usuario, la *Gaceta* no es explícitamente un archivo y de esto se desprenden problemas para la investigación. La *Gaceta* intenta, antes que cualquier otra cosa, tener accesible la información relevante para los trámites del orden del día; esto no siempre es posible. Los coordinadores de las fracciones partidistas negocian el orden del día en la Junta de Coordinación Política y, a menudo, terminan de hacerlo minutos antes de que dé inicio la sesión plenaria. Las iniciativas que son anunciadas con anticipación (o que pueden ser anticipadas por el personal de la Cámara) cuentan con toda la documentación en la *Gaceta* del día. Pero aquellos votos que son programados tarde o después del comienzo de la sesión plenaria, así como votos procedimentales, suelen no aparecer en la *Gaceta* y, por ende, la plantilla utilizada para recibir la votación nominal tampoco está lista. La votación queda excluida de la *Gaceta*.²

Para identificar la totalidad de las votaciones celebradas en el pleno y compararlas con las publicadas, descargamos versiones digitalizadas del *Diario de los Debates* (*Diario*, en adelante) correspondientes a las sesiones regulares y extraordinarias de la Cámara de la LVII a la LIX Legislaturas. Analizamos el texto con un procedimiento automatizado que busca el anuncio formal de cada votación nominal. Este procedimiento y su exactitud se discuten con detalle en el apéndice. Para las LX y LXI Legislaturas, en cambio, el listado completo de las votaciones está disponible en el Infopal. De ahora en adelante adoptaremos la etiqueta “votaciones *Gaceta*” para referirnos a aquellas que sí fueron publicadas en la *Gaceta Parlamentaria*, “votaciones omitidas”, para las excluidas de la *Gaceta*, y “todas las votaciones” para el universo de las votaciones nominales, sin importar si fueron o no publicadas.

gran mayoría de todos nosotros estamos conociendo el contenido de la ley que hoy vamos a discutir y supuestamente vamos a aprobar [...] ¿Cuál es nuestra responsabilidad, compañeras y compañeros, y particularmente mis compañeros de Acción Nacional, que son los que hacen la propuesta de que ahorremos trámites?” (*Diario de los Debates*, segunda sesión ordinaria del primer año de la LVI Legislatura, 24 de abril de 1995).

² Contactamos al *staff* legislativo para recuperar los votos no publicados. Después de varios intentos, nos dijo un director que esos votos habían sido desechados después de las sesiones y que simplemente no existían.

CUADRO 1. Descripción de los votos reportados y sin reportar por la *Gaceta*, LVII-LXI Legislaturas

Tipo de votación	Votaciones reportadas en la <i>Gaceta</i>	Votaciones no reportadas en la <i>Gaceta</i>	Todas las votaciones
En lo general	68.5	8.4	56.8
En lo particular	24.4	12.0	22.0
Procedimental	5.3	26.3	9.4
Toma de posición	0.8	18.0	4.2
Proyecto privado	0.2	30.2	6.0
Viaje presidencial	0.8	5.1	1.6
Total	100%	100%	100%
(N)	(2 439)	(593)	(3 032)

Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*.

El cuadro 1 ofrece una primera comparación de 3 032 votaciones del universo y del subconjunto de 2 439 publicadas en la *Gaceta* en el periodo de 1998 a 2012 —en total, la *Gaceta* omitió casi 20 por ciento de las votaciones nominales. Las votaciones están agrupadas en seis categorías mutuamente excluyentes. Las mociones en el Congreso mexicano son tramitadas con una agenda progresiva (véase Heller y Weldon, 2003). Se votan primero en lo general contra el *statu quo*. Si una mayoría se manifiesta a favor, la moción es entonces sometida a una o más enmiendas en votaciones en lo particular. La categoría de votaciones en lo general se conforma en su casi totalidad de dictámenes de comisión, aunque un puñado fueron propuestas directas al pleno cuando las reglas se suspenden por procedimientos extraordinarios.

La información del cuadro muestra efectos de selección sustanciales en los tipos de proyectos que son publicados por la *Gaceta*. Primero, los votos en lo general están modestamente sobrerrepresentados en la *Gaceta*: alcanzan 68.5 por ciento de las votaciones publicadas contra 56.8 por ciento de la totalidad. Los dictámenes de comisión suelen conocerse de antemano y, por ende, suelen también aparecer en la *Gaceta*: las votaciones en lo general representan sólo 8.4 por ciento de las omitidas por la *Gaceta*, principalmente dictámenes colocados al final en el orden del día, pero también puntos de acuerdo mandados directamente por la Junta de Coordinación al pleno (por ejemplo el presupuesto de la Cámara) y dictámenes negativos para rechazar una o más propuestas simultáneamente.

Los votos en lo particular están apenas sobrerrepresentados en los publicados. Las votaciones en esta categoría inciden sobre iniciativas que ya sortearon el voto en lo general. Las enmiendas introducidas en comisión se conocen *a priori*, dándole tiempo al personal de preparar su inclusión en la *Gaceta*. Pero las enmiendas también pueden ser introducidas cuando la Junta negocia el orden del día y, si esto pasa en el último minuto, pueden ser omitidas por la *Gaceta* del día. Aquellas raras enmiendas introducidas directamente en el pleno, cuando las reglas son suspendidas por una mayoría calificada, también tienden a quedar fuera de la *Gaceta*. Las enmiendas representan casi la cuarta parte de las votaciones publicadas y la mitad de esta proporción entre las no reportadas.

De los votos procedimentales, menos de la mitad fueron publicados en la *Gaceta* durante este periodo, por lo que quedaron sobrerrepresentados en la porción omitida por la *Gaceta* con respecto del universo. La categoría incluye votos que estructuran el proceso legislativo —enmiendas a las reglas permanentes del Congreso, mociones para suspender las reglas y adoptar procedimientos extraordinarios, mociones para aceptar enmiendas no contempladas en el dictamen de la comisión, etc.— y otros asuntos procedimentales de menos fondo. Como éstos tienden a plantearse espontáneamente en el transcurso de la sesión plenaria, suelen ser omitidos por la *Gaceta*.

Las tres categorías restantes son prácticamente excluidas de entre los votos publicados. La categoría “toma de posición” se refiere a votaciones sobre si incluir o no de inmediato un punto de acuerdo en el orden del día. Aunque algunos decretos de importancia de la Cámara (por ejemplo, enmiendas al reglamento de la Cámara o nombramientos para las comisiones permanentes) son puntos de acuerdo, aquellos incluidos en esta categoría son ejercicios de *position-taking* (Mayhew, 1974), como solicitar que la Secretaría de Hacienda conceda créditos libres de intereses a las víctimas de huracanes o lamentarse por un comentario hecho por el presidente que pudo haber herido los sentimientos del régimen cubano. Mientras que el punto de acuerdo en sí mismo se decide por votación económica, la decisión de modificar el orden del día para hacer dicha votación requiere una mayoría calificada y, por lo tanto, una votación nominal. Casi una quinta parte de los votos sin reportar son de este tipo. Los “proyectos privados” y “viajes presidenciales” autorizan a ciudadanos para desempeñar labores o para aceptar honores de gobiernos extranjeros y al Ejecutivo para salir del país, respectivamente. Aunque algunas controversias que definen las líneas de conflicto legislativo pueden aparecer aquí, la *Gaceta* suele ignorarlas por su escasa importancia.

En conjunto, desde una óptica meramente descriptiva, los tipos de votos que no reporta la *Gaceta* son marcadamente diferentes de aquellos que reporta, y son muchos. ¿Acaso la investigación basada solamente en las votaciones publicadas (por ejemplo Rosas y Langston, 2011) está condenada a sacar conclusiones inevitablemente sesgadas? ¿Pueden los estudios legislativos aprender algo sobre el periodo 1998-2006 a pesar del mecanismo de selección empleado por la *Gaceta*? El resto del trabajo investiga en qué medida esto es problemático para el análisis de las votaciones nominales de la Cámara de Diputados.

Análisis

La pregunta empírica central de nuestro trabajo es si las votaciones nominales publicadas por la *Gaceta* son o no una muestra representativa de la población más amplia de votaciones. De serlo, entonces los académicos y analistas pueden estudiar el periodo crítico de transición del Congreso mexicano utilizando los recursos disponibles, sin temor a extraer conclusiones tendenciosas. Si en cambio no lo fuera, será necesaria una recopilación adicional de datos y ajustes a la muestra.³

En un marco más abstracto, cada voto puede ser considerado como un sorteo de una distribución categórica M -dimensional. M en este caso es el número de diputados y las categorías corresponden al número de acciones que puede llevar a cabo un legislador en una votación nominal —votar a favor, en contra, abstenerse, ausentarse—. Las dimensiones no son independientes, porque los legisladores negocian y son presionados por el partido y por otros actores. En este contexto, probar si los votos publicados por la *Gaceta* son una muestra aleatoria de todos los votos requeriría una base de datos inmensa —mucho más grande que la proporcionada por los años de votos nominales que observamos.

Como una alternativa, nos enfocamos en cantidades de interés para los estudios legislativos clásicos: índices Rice de cohesión de los legisladores y los puntos ideales que las técnicas de escalamiento permiten recuperar (Cox, 1987; Poole y Rosenthal, 1997; Rosas y Langston, 2011). Si bien exis-

³ Por supuesto, aun si la muestra de la *Gaceta* fuera representativa, hay pérdida de información en virtud de ser una base de datos más pequeña. Esto conlleva conclusiones menos precisas, pero atinadas. Si estuviera en el futuro disponible el universo de las votaciones para ese periodo, ese censo será preferible a la muestra que consideramos aquí.

ten otras cantidades de interés potencial, es extremadamente raro que quienes analizan votaciones nominales no le presten atención a estas dos cantidades clave. Nuestras pruebas, entonces, plantearán la misma pregunta a estas estadísticas: ¿difieren significativamente las cantidades derivadas a partir de las votaciones reportadas en la *Gaceta* de las cantidades derivadas de todos los votos? ¿Son los votos publicados por la *Gaceta* representativos de todos los votos en términos de cohesión y perfiles de votación?

A continuación abordamos estas preguntas por medio del análisis de permutación no paramétrico. El procedimiento consiste en tomar muchas muestras aleatorias repetidas con remplazo y de tamaño n de las votaciones nominales del universo en cada legislatura. Para cada muestra aleatoria, se calculan cantidades de interés. Comparamos la distribución de las cantidades de interés de las permutaciones con el valor de las mismas medido con las votaciones reportadas por la *Gaceta*. Si los valores de votaciones de la *Gaceta* son típicos de muestras aleatorias de la población más grande, entonces tendremos evidencia de que los votos publicados pueden utilizarse para el análisis. De lo contrario, habrá evidencia de que los mecanismos de selección deberán ser entendidos para contabilizarlos en los estudios legislativos de la Cámara de Diputados.

Procedimos en dos frentes y examinamos por separado los datos de 1998-2006 y los de 2006-2012. Para el periodo crítico 1998-2006 carecemos de casi una cuarta parte de las votaciones que no fueron publicados por la *Gaceta* ni en ningún otro lugar. Sin embargo, aunque para los votos no publicados carecemos de información de cómo votó cada diputado, el *Diario de los Debates* sí enumera estadísticas resumidas del desenlace de cada votación: el número de diputados que votaron a favor, en contra o que se abstuvieron. En consecuencia, aunque calcular la cohesión de las fracciones parlamentarias o los puntos ideales de los diputados está fuera de nuestro alcance, sí resulta posible, por lo menos, preguntar si en sus descriptivos agregados las votaciones de la *Gaceta* tienen patrones similares de divisiones favor-contra y de tasas de participación y de abstención en comparación con todos los votos que detectamos en el *Diario*. Nuestro hallazgo es que, en su mayor parte, las votaciones que aparecen en la *Gaceta* parecen ser una muestra representativa de todas las votaciones. La excepción más importante es que la muestra *Gaceta* contiene más votos unánimes que el *Diario*, sobre todo en 2004. Esto es a primera vista lamentable, ya que los votos unánimes esencialmente no contienen ninguna información para el escalamiento de puntos ideales; sin embargo, la segunda parte de nuestro análisis

proporciona la seguridad de que esto no es un problema. Además, el volumen de votos omitidos en ese último periodo, por debajo de 20 por ciento, fue relativamente menor.

Para el periodo 2006-2012 contamos con la información de todas y cada una de las votaciones nominales. No hubo más datos faltantes porque, desde el inicio de la LX Legislatura, todas las votaciones nominales son listadas en una nueva página web de la propia Cámara, interesada en recopilar cada una de ellas (no como la *Gaceta*, cuya razón de ser es tener lista la información necesaria para tramitar los asuntos de cada día). Al mismo tiempo, la *Gaceta Parlamentaria* continuó la práctica de publicar un subconjunto incompleto de los votos emitidos. Bajo el supuesto de que el proceso de omisión de votaciones de la *Gaceta* no ha cambiado, explotamos la coexistencia de una fuente completa y otra parcial para investigar si ésta afecta la estimación de los puntos ideales de los diputados. De nuevo, nuestro hallazgo más importante es que no afecta. Las votaciones de la *Gaceta* también parecen ser típicas de todas las votaciones desde esta perspectiva.

La comparación de datos agregados 1998-2006

Empezamos con un resumen estadístico, examinando la cohesión anual media del pleno en las votaciones publicadas y las no publicadas. El tamaño de la muestra aleatoria n en este paso del análisis es el número de votaciones nominales publicadas en el año en la *Gaceta*. En otras palabras, el muestreo selecciona un gran número de muestras (10 000) con tantos votos como las sí publicadas en la *Gaceta*. Definimos la cohesión anual media como $c = \frac{1}{m} \sum_{i=1}^m \frac{|y_i - n_i|}{(y_i + n_i)}$ donde m es el número de votaciones nominales celebradas en el año (publicadas o no), y_i es el número de votos a favor de la moción y n_i los votos en contra. En otras palabras, obtenemos primero el índice de cohesión de Rice (1925) para cada moción y lo promediamos para todas las votaciones del año. El índice de Rice está acotado por el intervalo [0,1] y proporciona una medida de conflicto o controversia entre los legisladores. Una moción adoptada o rechazada por unanimidad (todos los diputados que votan lo hacen en el mismo sentido) obtiene un valor de 1 en el índice; aquella que divide la legislatura en mitades exactas, una a favor y la otra en contra, obtiene un valor de 0, y los casos intermedios varían entre estos extremos. El cuadro 2 muestra las estadísticas del resumen de cohesión de todas las votaciones en las LVII-LIX Legislaturas (1998-2006) en contraste con las publicadas y las no publicadas en la *Gaceta*.

CUADRO 2. Comparación de la cohesión y la participación en votaciones publicadas e inéditas, LVII-LIX Legislaturas (1998-2006)

	Todas	Gaceta	Inéditas
Cohesión media	0.79	0.82	0.69
Media de votos emitidos por votación	382	385	375
Número de votaciones nominales	1 702	1 285	417
Porcentaje de votaciones unánimes	45	50	30

Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*.

La naturaleza de la medida de cohesión que aquí calculamos requiere ser elaborada. En general el término cohesión se utiliza para referirse a la conducta de un subgrupo de diputados —esto es, el índice de Rice se calcula por separado para diferentes grupos pertenecientes a una legislatura, con el objeto de comparar la cohesión de los distintos partidos o de los representantes de intereses agrícolas o manufactureros, como en el trabajo original de Rice—. En este caso, calculamos una puntuación de cohesión para toda la legislatura; en otras palabras, con cuánta frecuencia *todos* los legisladores votaron en el mismo sentido, sin importar el partido o pertenencia a otros grupos. Otra forma de pensar esto es que $1 - c$ es la medida de qué tan cerca estuvo el voto total de la división en mitades perfectas ($1 - c = 1$) o de una votación para nada cerrada, con todos los legisladores votando igual ($1 - c = 0$).⁴

De vuelta a los datos, el patrón general que se distingue en el cuadro 2 es que los votos omitidos tienen una medida de cohesión menor, son más controvertidos en promedio que los votos de la *Gaceta*. La cohesión media de los votos publicados fue 0.82, contra 0.69 de los inéditos. Casi la mitad de esta diferencia es atribuible al diferencial de votaciones unánimes. 30 por ciento de las votaciones inéditas fueron unánimes, comparadas con 50 por ciento de las votaciones publicadas por la *Gaceta*. Muchos métodos, como el de puntos ideales que elaboramos más adelante, descartan los votos unánimes, con lo cual la diferencia de cohesión entre las votaciones publicadas y las que no disminuye sustancialmente a 0.08 (0.64 para aquellas, 0.56 para éstas).

Los estudios de la política legislativa también suelen analizar las abstenciones y las ausencias en el pleno, a menudo codificándolas como si se tra-

⁴ Agradecemos a un dictaminador anónimo la sugerencia de esta aclaración.

CUADRO 3. Cohesión media anual por estado de publicación

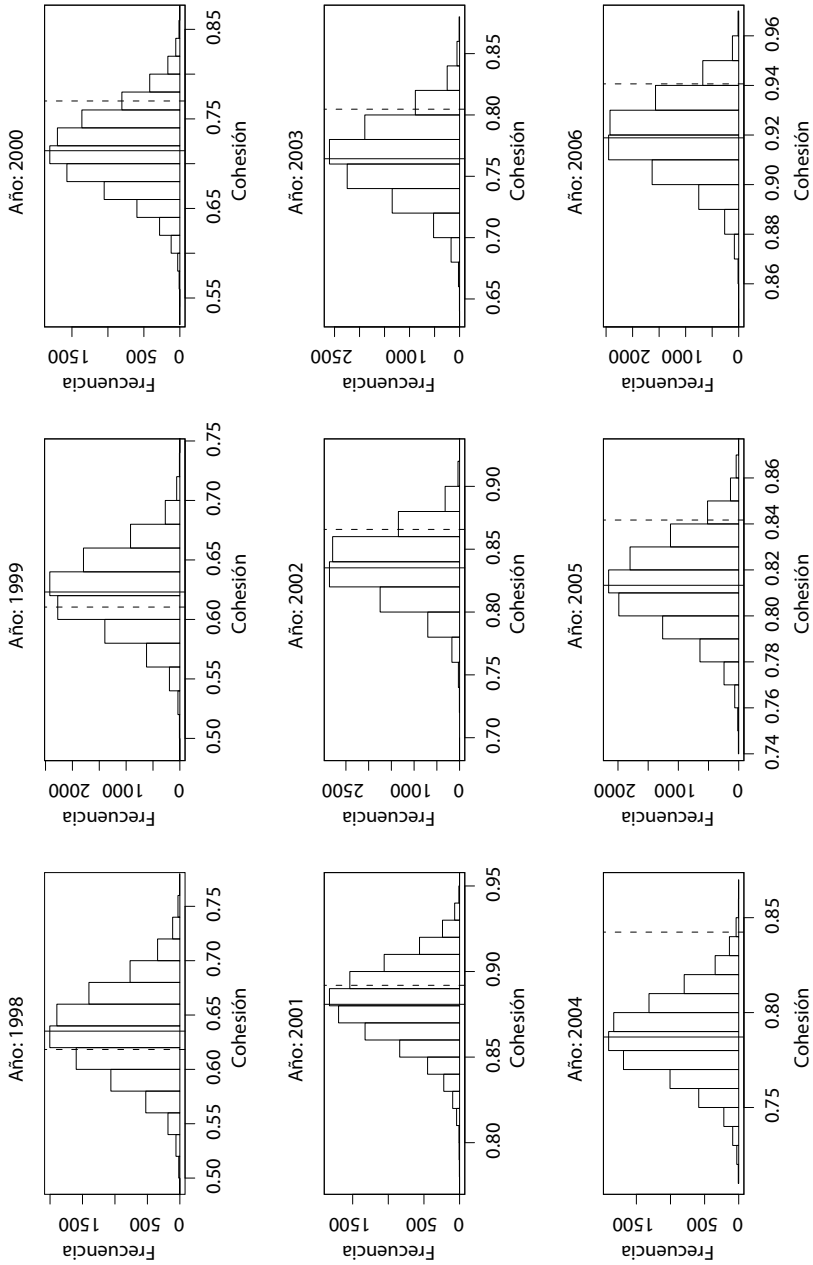
Año	Cohesión			
	Todos los votos	Gaceta	Inéditos	Diferencia
1998	0.64	0.62	0.67	-0.052
1999	0.62	0.61	0.66	-0.052
2000	0.71	0.77	0.60	0.171
2001	0.88	0.89	0.84	0.055
2002	0.84	0.87	0.78	0.086
2003	0.76	0.80	0.69	0.112
2004	0.79	0.84	0.60	0.239
2005	0.81	0.84	0.68	0.163
2006	0.92	0.94	0.73	0.215

Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*.

tase de votos en contra (por ejemplo, Jones y Hwang, 2005). Consideramos dicha codificación con escepticismo (véase Rosas y Shomer, 2008), pero tratándose de una práctica común, sopesamos las diferencias potenciales entre las votaciones publicadas y las no publicadas con respecto a esto. Ambas muestras presentan tasas de participación virtualmente idénticas. El número de abstenciones promedio en las votaciones publicadas por la *Gaceta* es de 6.73 contra 6.79 en los votos inéditos. Las votaciones de la *Gaceta* registraron 104 diputados ausentes en el pleno en promedio, las inéditas 118, modestamente más alto: en otras palabras, las comparaciones simples de cohesión y de tasas de abstención y ausencia revelan diferencias pequeñas entre los votos. Las abstenciones, que son probablemente estratégicas y más costosas para el legislador, tienen tasas casi idénticas. Las ausencias pueden reflejar otras prioridades y obligaciones de los legisladores, presentan diferencias más notables pero a la vez consistentes con esta explicación. Las votaciones *Gaceta* son de mayor perfil, pre-agendadas, y por ende menos probables de llegar al pleno cuando muchos legisladores están ausentes.

El cuadro 3 desglosa las tasas de cohesión anuales de las votaciones según las haya publicado la *Gaceta* o no entre 1998 y 2006. Hay diferencias modestas entre las publicadas y las inéditas, pero los patrones no son consistentes ni en sentido ni en magnitud. La mayor parte de las diferencias son pequeñas, aunque la brecha de 2004 es mayor, en 0.24. Esto significa que la típica moción publicada en ese año tuvo una coalición mayoritaria alrededor de 10 por ciento más grande que la típica moción inédita. Una vez más, la mayor parte de esta diferencia se debe a un número excepcionalmente

GRÁFICA 1. Análisis de cohesión legislativa entre las votaciones *Gaceta* y todas las votaciones de la Cámara de Diputados (1998-2006)



Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*. *Nota:* Las diferencias de cohesión en votaciones de la *Gaceta* y todas las votaciones no son significativas antes de 2004. Los histogramas muestran el rango que adopta la cohesión anual media en 10 000 muestras aleatorias que emulan el subconjunto publicado en la *Gaceta* a partir del total de votaciones. Las líneas punteadas apuntan a la cohesión anual promedio en la *Gaceta*, las líneas continuas a la cohesión anual promedio en todas las votaciones.

elevado de votos unánimes en ese año: 56 por ciento de las votaciones en la *Gaceta* fueron unánimes en 2004, contra sólo 21 por ciento de los votos inéditos, una diferencia mucho mayor que la observada en cualquier otro año.

Aunque modestas, ¿son significativas estas diferencias? La gráfica 1 muestra los resultados del análisis de permutaciones que revelan que las votaciones en la *Gaceta* parecen ser típicas de todas las votaciones nominales. En cada histograma de la figura, la línea vertical continua indica la cohesión media anual de todos los votos, mientras la línea punteada indica la misma cantidad sólo para los votos en la *Gaceta*. Los histogramas muestran la distribución de cohesión anual media obtenida al permutar muestras aleatorias de todos los votos. Se tomaron 10 000 muestras del mismo tamaño que el conjunto de votaciones que sí fueron publicadas en la *Gaceta*. De este modo se consigue emular el proceso de generación de votos no publicados para verificar si los que omite la *Gaceta* parecen ser una muestra aleatoria. Cuando la línea punteada está hacia el medio de la distribución, significa que la cohesión de los votos de la *Gaceta* es típica de una muestra aleatoria de todos los votos. Cuando la línea punteada está a la derecha del histograma, significa que los votos de la *Gaceta* son significativamente diferentes a todo el conjunto de votos.

En todos los años, excepto 2004, la línea punteada se ubica más cerca del centro de la distribución simulada que de un extremo. En otras palabras, aunque presenta diferencias puntuales visibles en 2002, 2003, 2005 y 2006, el promedio de cohesión en la *Gaceta* en esos años se ve como la del típico voto elegido aleatoriamente de la población de todos los votos. No hay evidencia estadística que permita rechazar la hipótesis de que los votos de la *Gaceta* de hecho son una muestra aleatoria de la población completa. En 2004, en cambio, la cohesión media anual de la muestra en la *Gaceta* se ubicó en el percentil 99º de los valores permutados. Sin ser el valor máximo, es más grande que el promedio de cohesión anual de casi todas las muestras permutadas. Esta particularidad parecería resultar del elevado número de votos unánimes de ese año. Cuando restringimos el muestreo de modo que excluyera votaciones nominales unánimes, también obtuvimos un resultado nulo para el año 2004, como en los demás.

El análisis de permutaciones sostiene la afirmación de que los votos publicados por la *Gaceta Parlamentaria* no son fundamentalmente diferentes de una muestra representativa de todos los votos emitidos, pero requiere varias advertencias y observaciones adicionales. Primero, señalamos de nueva cuenta que la escasez de datos para este primer periodo 1998-2006

limita nuestro análisis de cohesión general de votos. Los datos disponibles sólo permiten comparar la medida de controversia agregada de los votos. No podemos descartar que los patrones de votación al interior de cada partido u otros subgrupos difieran entre lo publicado y lo no publicado. Segundo, el hecho de que haya relativamente más votos unánimes en la *Gaceta* significa pérdida de votos controvertidos cuando se trabaja únicamente con el subconjunto *Gaceta*. Tercero, aunque las diferencias en abstenciones y ausencias parecen soslayables entre lo publicado y lo no publicado, es posible que se trate de distintos tipos de diputados en una y otra muestras. Esto podría tener efectos según qué modelo de votación faltante se utilice (Rosas y Shomer, 2008). Por último, repetimos el análisis entero codificando toda abstención y ausencia como si fuera un voto en contra de la moción. Los resultados permanecieron en esencia iguales, sin evidencia de diferencias sistemáticas entre la *Gaceta Parlamentaria* y la totalidad de las votaciones nominales. Con esto en mente, pasamos a buscar diferencias mediante un análisis espacial de las LX y LXI Legislaturas.

Los puntos ideales de los diputados 2006-2012

La segunda parte de nuestro análisis presenta datos del periodo en el que ya está disponible la totalidad de las votaciones nominales a través de la nueva página web del Infopal. Como en las legislaturas anteriores, la *Gaceta* continuó publicando un subconjunto incompleto de votos. Mientras que para el primer periodo contamos sólo con el agregado de votos a favor y en contra, y sólo pudimos comparar la cohesión general de la legislatura, para este segundo periodo tenemos el voto individual de cada legislador; por lo tanto, podemos explorar si el mecanismo de selección de la *Gaceta* tiene alguna influencia en la estimación de puntos ideales.⁵ Con este fin, estimamos los puntos ideales de los diputados dos veces: una utilizando la totalidad de las votaciones nominales celebradas, otra sólo con los votos que publicó la *Gaceta*. En tanto el proceso de selección de la *Gaceta* arroje una muestra no representativa del total de las votaciones, las diferencias se manifestarán en una comparación directa de los puntos ideales estimados con cada grupo de votos.

Se especificó un modelo IRT (*item response theory model*), bidimensional (véase Gelman y Hill, 2007), estimando los puntos ideales de los diputados y otros parámetros de interés con Monte Carlo y cadenas de Markov (MCMC)

⁵Véase Poole y Rosenthal (1997) para una introducción a esta metodología.

CUADRO 4. Votos escalables y no escalables de los votos nominales, 2006-2012

Legislatura	LX		LXI	
	Publicados en la <i>Gaceta</i>	Todos los votos	Publicados en la <i>Gaceta</i>	Todos los votos
Unánimes	259	293	295	321
Casi sin oposición	65	82	138	150
Con oposición (escalables)	160	215	234	267
Participación media en votos escalables	370	365	348	341
Total	484	590	667	738

Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*.

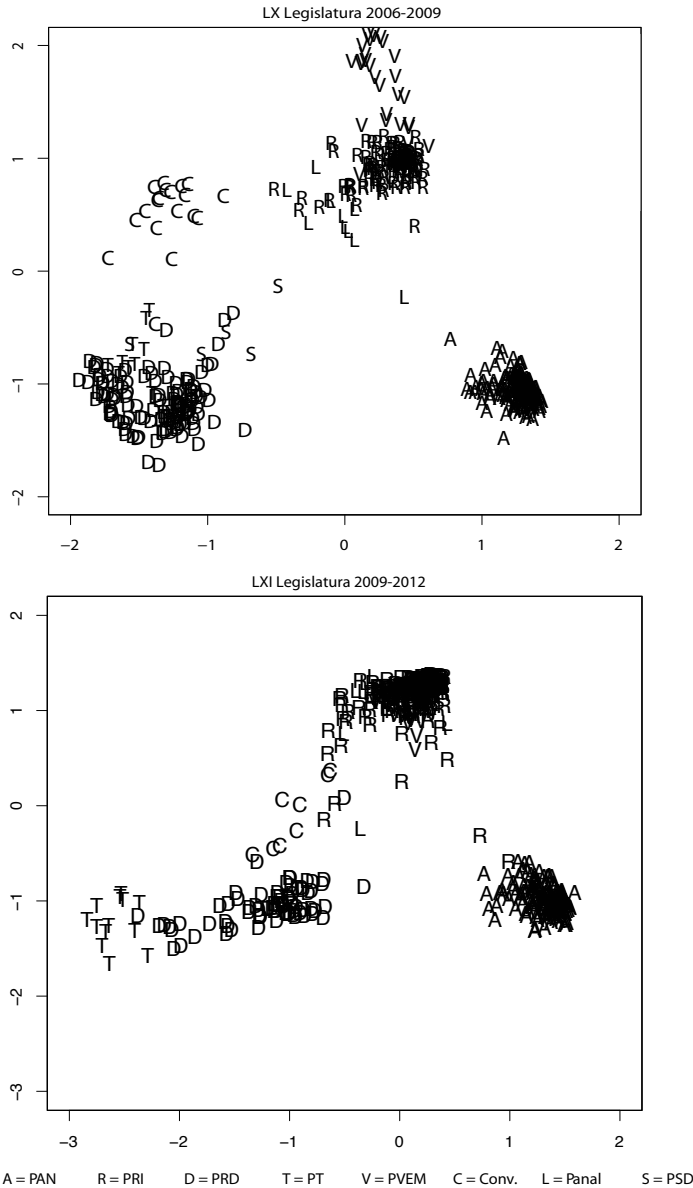
(Clinton *et al.*, 2004; Poole, 2005, 116). Como es estándar, las votaciones nominales unánimes y las mínimamente impugnadas (aquellas con una minoría de 2.5 por ciento o menos, excluyendo las abstenciones y ausencias) fueron eliminadas del conjunto de datos debido a la poca información que ofrecen; el cuadro 4 ofrece un resumen. Por la misma razón, los diputados que nunca votaron o que votaron a favor o en contra de menos de 10 por ciento de todas las mociones consideradas en la legislatura fueron eliminados del análisis. La primera dimensión explica casi 95 por ciento de los votos individuales y describe una segmentación del gobierno contra una oposición en el Congreso. En la segunda dimensión predomina aún la presión continua para la liberalización del sistema político (Robles, 2009).

La identificación del modelo se hizo asignando probabilidades anteriores semiinformativas a los puntos ideales de los miembros de los tres grandes partidos en coordenadas polares: el PRI al norte, el Partido Acción Nacional (PAN) al sureste y el Partido de la Revolución Democrática (PRD) al suroeste.⁶ Tres cadenas se reiteraron 10 000 veces cada una. Los parámetros habían alcanzado una etapa estable después de la primera mitad de todas

⁶ Definimos las distribuciones anteriores de los puntos ideales del siguiente modo (x , y son coordenadas en un plano cartesiano). Los diputados del PAN, PRI y PRD recibieron distribuciones *a priori* bivariadas semi-informativas:

$\begin{pmatrix} x \\ y \end{pmatrix}_{\text{PAN}} \sim N\left(\begin{matrix} 1, \frac{1}{4} \\ -1, \frac{1}{4} \end{matrix}\right)$; $\begin{pmatrix} x \\ y \end{pmatrix}_{\text{PRI}} \sim N\left(\begin{matrix} 0, \frac{1}{4} \\ 1, \frac{1}{4} \end{matrix}\right)$; $\begin{pmatrix} x \\ y \end{pmatrix}_{\text{PRD}} \sim N\left(\begin{matrix} -1, \frac{1}{4} \\ -1, \frac{1}{4} \end{matrix}\right)$. Los demás diputados recibieron distribuciones *a priori* no-informativas. $\begin{pmatrix} x \\ y \end{pmatrix}_{\text{rest}} \sim N\left(\begin{matrix} 0,10 \\ 0,10 \end{matrix}\right)$.

GRÁFICA 2. Puntos ideales estimados con todos los votos en dos legislaturas



Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*.

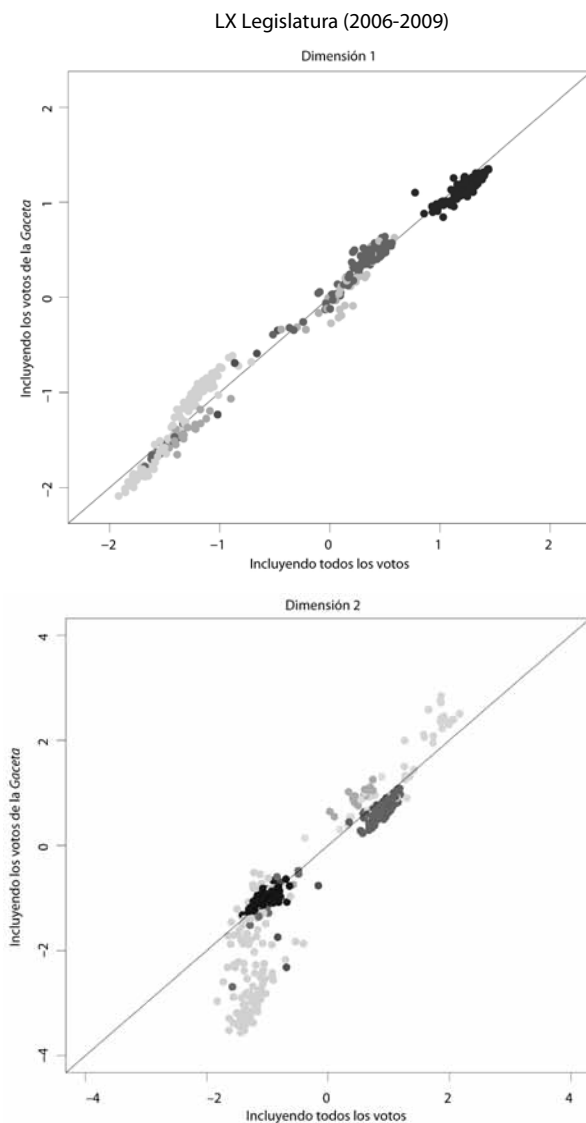
esas repeticiones; se tomó cada 25^a simulación de la segunda mitad como muestra posterior de los parámetros, que reportamos.

Los puntos ideales estimados, que aparecen en la gráfica 2, son consistentes con estimaciones previas de la Cámara de Diputados de México (Rosas y Langston, 2011). La mayoría de los diputados de los principales partidos manifiestan una ligera dispersión en torno a los vértices de un triángulo isósceles con reminiscencia de las probabilidades anteriores asignadas. La excepción es el PRD, alrededor de la mitad del cual se desplazó, junto con el Partido del Trabajo PT, más al oeste en la LX Legislatura. Se trata de la facción de López Obrador, que tras no reconocer el triunfo de Felipe Calderón en la elección presidencial de 2006 comenzó una ruidosa campaña anti-PAN mientras el resto del partido negociaba con los demás partidos. La facción de López Obrador perdió control de la dirección del partido poco después, dejándolo con un disminuido, pero más radical contingente en la LXI Legislatura. En esa Legislatura, instalada después de las elecciones intermedias de 2009, el PRI ganó la pluralidad de la Cámara, y se alió con dos partidos menores —el Partido Verde Ecologista de México (PVEM), su aliado preelectoral común desde 2001, y el Partido Nueva Alianza (Panal)— en pos de una mayoría contra el presidente: alcanzaron un grado de cohesión notable.

La gráfica 3 compara, una dimensión a la vez, los puntos ideales estimados con la totalidad de las votaciones nominales y aquellos estimados con las votaciones de la *Gaceta* únicamente. Con una excepción, los legisladores de todos los partidos aparecen estrechamente alineados con la diagonal de 45° que indica la coincidencia exacta de coordenadas. La excepción es la segunda dimensión de la LX Legislatura, donde un grupo grande del PRD aparece en una posición más radical con sólo los votos *Gaceta* que con la totalidad de votos. Esto sorprende porque la *Gaceta* excluye un buen número de mociones de posicionamiento y de enmienda, y su exclusión debería ejercer una fuerza centrípeta, no centrífuga.⁷ Aun en este caso, la coincidencia general entre los puntos ideales estimados permite sostener que las diferencias entre los votos de la *Gaceta* y la totalidad son pequeñas.

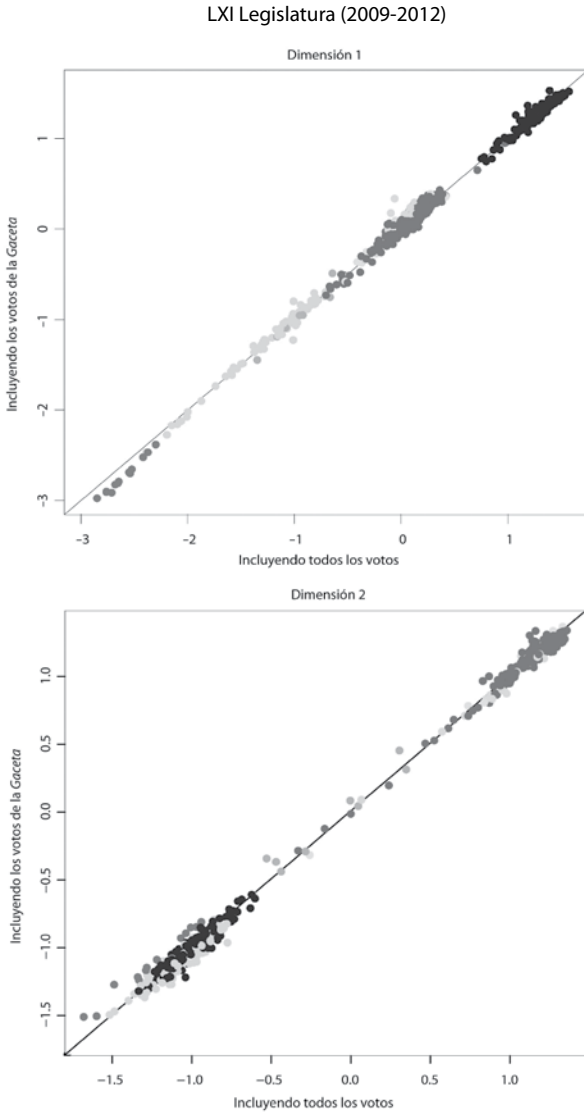
⁷ Como se observó en el análisis de la cohesión del pleno, el mecanismo de selección de la *Gaceta* resulta en una proporción mayor de votos no informativos-votos unánimes o cuasi unánimes. Es posible que las fuerzas moderadoras del PRD ejercieran mayor influencia en votaciones no publicadas por la *Gaceta*, y que sea justamente su exclusión la que mueva este resultado.

GRÁFICA 3A. Comparación coordinada por coordinada de los puntos ideales estimados con todos los votos (eje x) y con los votos publicados en la *Gaceta* (eje y)



Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*. Una versión a color de esta gráfica se encuentra en <http://allman.rhon.itam.mx/~emagar/graphs/pygPaper/>.

GRÁFICA 3B. Comparación coordinada por coordinada de los puntos ideales estimados con todos los votos (eje *x*) y con los votos publicados en la *Gaceta* (eje *y*)



Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*. Una versión a color de esta gráfica se encuentra en <http://allman.rhon.itam.mx/~emagar/graphs/pygPaper/>.

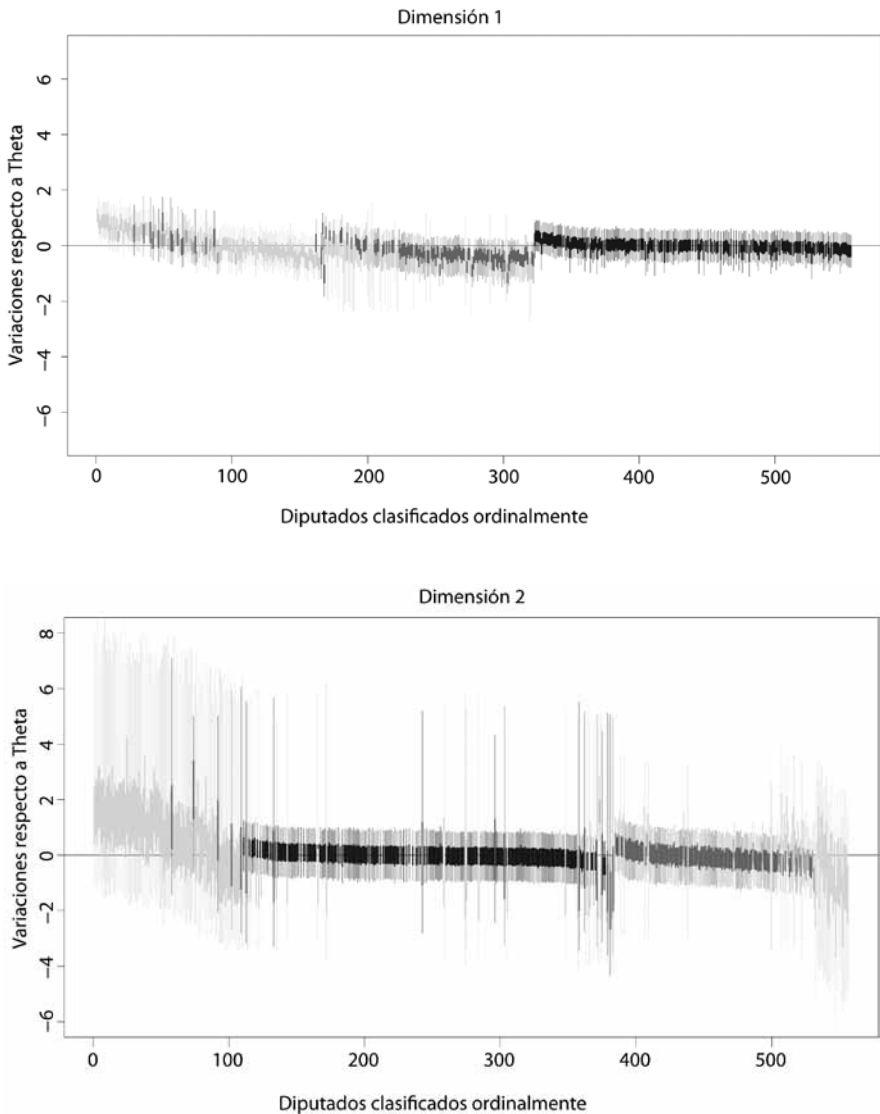
Permutaciones de puntos ideales

El siguiente paso para evaluar los efectos de las votaciones nominales incompletas reproduce el análisis de permutaciones utilizado anteriormente, pero con los puntos ideales en vez de los estadísticos de cohesión agregados. Como la estimación MCMC demanda muchos recursos de la computadora, usar un elevadísimo número de muestras de gran tamaño resultó impráctico. De entre todas las votaciones nominales celebradas en cada legislatura, tomamos 99 muestras aleatorias (en lugar de 10 000) de $n = 50$ votos escalables. Hubiera sido preferible tomar muchas más muestras y que cada una tuviese el mismo tamaño que los votos escalables de la *Gaceta* (o sea $n = 160$ para la LX y $n = 234$ para la LXI Legislaturas, respectivamente), pero el tiempo de cómputo continúa siendo una limitante.⁸ Para realizar la comparación de los puntos ideales estimados con las muestras aleatorias contra los obtenidos con la *Gaceta*, el punto ideal de cada legislador escalado con votos *Gaceta* únicamente fue restado, coordenada por coordenada, de cada uno de los puntos obtenidos con las 99 muestras aleatorias. Si la *Gaceta* es una muestra aleatoria más, los puntos ideales escalados con ella oscilarán estocásticamente en torno de las 99 muestras, sin cambios sistemáticos entre legisladores. Las diferencias estarán centradas en el cero, con raras desviaciones positivas o negativas anormales. Las gráficas 4 y 5 clasifican a los diputados de acuerdo con puntuaciones de la *Gaceta* ascendentes, y el eje vertical reporta la diferencia de coordenadas de las muestras aleatorias. Las líneas verticales gruesas y delgadas señalan los intervalos de 50 y 95 por ciento de la diferencia.

Para el caso de la LX Legislatura, la gráfica 4 muestra que sólo un puñado de los intervalos de 95 por ciento de los diputados excluye el cero, el punto donde la diferencia entre las muestras y la *Gaceta* no se manifiesta. Incluso son muchos los intervalos de 50 por ciento que incluyen el cero. Las discrepancias se concentran en la primera dimensión y en el extremo izquierdo de la distribución —los diputados afines a López Obrador son más vulnerables al muestreo de la *Gaceta*—. El muestreo aleatorio devuelve puntos ideales estimados más volátiles para unos partidos que para otros, notablemente el PVEM en ambas dimensiones, y el PRD en la segunda dimensión. Para el caso de la LXI Legislatura (gráfica 5), la variación de los

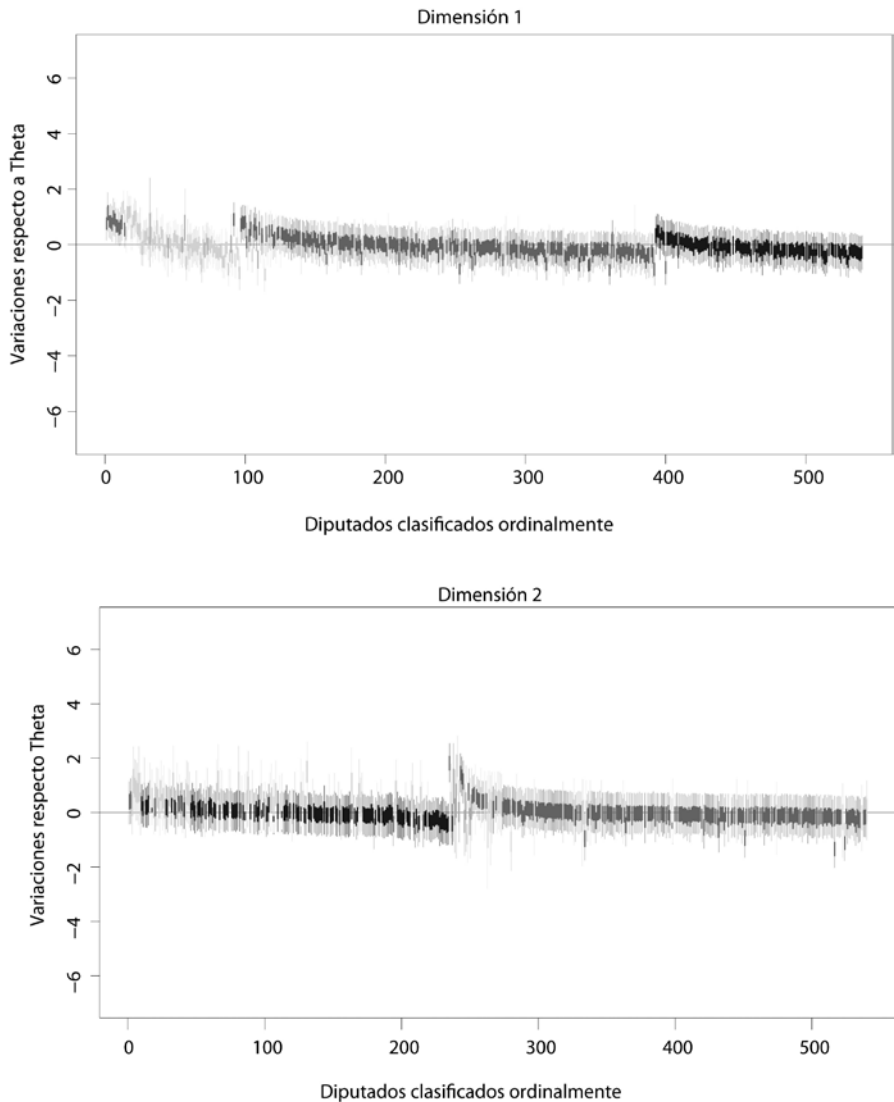
⁸ Aun así, se requirieron cerca de 24 horas para completar la rutina de estimación para cada legislatura usando tres procesadores multinúcleo en paralelo. Con tamaños de muestras entre tres y cinco veces más grandes, el tiempo de computación aumentaría cuatro veces en promedio.

GRÁFICA 4. Puntos ideales estimados con muestras aleatorias de todos los votos y con los de la *Gaceta*, LX Legislatura



Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*. *Notas:* El eje vertical indica el diferencial de las coordenadas obtenidas con las muestras y las obtenidas con la *Gaceta*. Una versión a color de esta gráfica se encuentra en <http://all-man.rhon.itam.mx/~emagar/graphs/pygPaper/>.

GRÁFICA 5. Puntos ideales estimados con muestras aleatorias de todos los votos y con los de la *Gaceta*, LXI Legislatura



Fuente: Elaboración propia con base en los datos del Servicio de Información para la Estadística Parlamentaria (Infopal) y la *Gaceta Parlamentaria*. *Notas:* El eje vertical indica el diferencial de las coordenadas obtenidas con las muestras y las obtenidas con la *Gaceta*. Una versión a color de esta gráfica se encuentra en <http://allman.rhon.itam.mx/~emagar/graphs/pygPaper/>.

puntajes ideales es significativamente más pequeña que en la legislatura previa. Similar a la LX, el grupo de los diputados lopezobradoristas tiene puntuaciones menos radicales en la primera dimensión cuando las estimaciones utilizan una muestra aleatoria de todos los votos, en vez de la muestra de la *Gaceta*. Hay un efecto similar entre los miembros más moderados del PRI en la segunda dimensión. Pueden extraerse algunas generalizaciones de este análisis: primero, los diputados del PAN y del PRI son en general menos sensibles al muestreo que el PRD y los diputados de partidos pequeños. Segundo, los diputados extremistas del flanco izquierdo resienten el muestreo de la *Gaceta*, pero esto no parece suficiente para cambiar su posicionamiento ordinal con respecto a los miembros más moderados de su propio partido. Se debe tener cautela al estimar puntos ideales para este grupo de diputados, pero no encontramos diferencias importantes en sus perfiles de votación usando los votos publicados en la *Gaceta* y el conjunto completo.


Conclusiones

La transición democrática en México ha venido de la mano de un aumento en el estudio académico de las instituciones políticas, incluyendo el Congreso. Desde 1997, ningún partido ha obtenido la mayoría de escaños en la Cámara de Diputados y el dominio del Poder Legislativo por los presidentes mexicanos ha terminado. El resultado es un Congreso fortalecido, cuya capacidad institucional y profesionalismo están en pleno desarrollo. Pero si ha crecido dramáticamente el interés en la política legislativa y el Congreso, la naturaleza de los datos disponibles para su estudio merece un examen serio. En particular, como la publicación de las votaciones nominales entre 1998 y 2006 fue incompleta, surge la posibilidad de que las investigaciones de la conducta legislativa en el periodo hayan extraído conclusiones erróneas, deformadas por el sesgo de selección. Es más, se plantea la interrogante de si la falta de un número cuantioso de información en los registros públicos de la Cámara fue por omisión o por comisión. Este estudio plantea indirectamente ambas interrogantes. No encontramos evidencia de que las conclusiones hasta ahora alcanzadas estén irremediablemente sesgadas ni de que la falta de información sea una política de opacidad intencional del personal de la Cámara.

Mediante un examen de la cohesión legislativa y los puntos ideales de los legisladores, este estudio exploró la posibilidad de considerar las votaciones nominales disponibles en línea a través de la *Gaceta Parlamentaria*

como si se tratase de una muestra aleatoria de todas las votaciones celebradas en el periodo. La evidencia empírica demuestra que, cuando se utilizan solamente los votos publicados, las estadísticas comunes para el estudio del comportamiento legislativo no son significativamente diferentes de cuando se utiliza tanto la información publicada como la inédita. Una implicación clave de este estudio es que pueden usarse los datos disponibles en línea para las LVII, LVIII y LIX Legislaturas sin preocuparse de la validez del análisis. En paralelo a la publicación de este artículo, hemos creado un acervo público que contiene las votaciones divulgadas por la *Gaceta Parlamentaria* entre 1998 y 2006 y por el Infopal de 2006 en adelante.⁹

Es importante reiterar una calificación hecha en el cuerpo del texto. Hemos examinado y descartado el sesgo en las dos herramientas más comunes de los estudios legislativos en la actualidad —la cohesión en el pleno y los puntos ideales de los legisladores—, pero existen muchos otros estudios estadísticos que no hemos analizado. Alentamos a la comunidad académica interesada en otras cantidades a verificar sus propiedades con los datos que hacemos públicos.

El problema de datos incompletos es común en nuevas democracias. El reto de los académicos, consultores, asesores y hasta servidores públicos consiste en evaluar si las inferencias hechas a partir de datos incompletos pueden estar sesgadas o si sólo contienen un mayor margen de error. Las herramientas que hemos propuesto en este estudio contribuyen a achicar las dudas al respecto sobre esta cuestión en los estudios legislativos mexicanos. 

Referencias bibliográficas

- Cantú, Francisco y Scott Desposato (2012), “The New Federalism of Mexico’s Party System”, *Journal of Politics in Latin America*, 2, pp. 3-38.
- Casar, María Amparo (2002), “Executive-Legislative Relations: The Case of Mexico (1946-1997)”, en Scott Morgenstern y Benito Nacif (eds.), *Legislative Politics in Latin America*, Cambridge, Cambridge University Press.
- Clinton, Joshua, Simon Jackman y Douglas Rivers (2004), “The Statistical Analysis of Roll Call Data”, *The American Political Science Review*, 98 (2), pp. 355-70.

⁹ Los archivos digitalizados pueden encontrarse en las páginas de internet de los autores (véase, por ejemplo, <http://allman.rhon.itam.mx/~emagar/rollCall/>).

- Cosío Villegas, Daniel (1981), *El sistema político mexicano*, México, Joaquín Mortiz.
- Cox, Gary W. (1987), *The Efficient Secret*, Nueva York, Cambridge University Press.
- Domínguez, Jorge I. y James A. McCann (1996), *Democratizing Mexico*, Baltimore, The Johns Hopkins University Press.
- Friedl, Jeffrey E. F. (2006), *Mastering Regular Expressions*, Sebastopol, California, O'Reilly Media.
- Gelman, Andrew y Jennifer Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Nueva York, Cambridge University Press.
- González Casanova, Pablo (1965), *La democracia en México*, México, Era.
- Heller, William B. y Jeffrey A. Weldon (2003), “Reglas de votación y la estabilidad en la Cámara de Diputados”, en Luisa Béjar Algazi y Rosa María Mirón Lince (eds.), *El Congreso mexicano después de la alternancia*, México, Asociación Mexicana de Estudios Parlamentarios/Instituto de Investigaciones Legislativas del Senado de la República, pp. 85-119.
- Jones, Mark P. y Wonjae Hwang (2005), “Party Government in Presidential Democracies: Extending the Cartel and CPG Theory Beyond the U. S. Congress”, *American Journal of Political Science*, 49 (2), pp. 267-82.
- Mayhew, David R. (1974), *Congress: The Electoral Connection*, New Haven, Yale University Press.
- Morgenstern, Scott (2003), *Patterns of Legislative Politics: Roll-Call Voting in Latin America and the United States*, Nueva York, Cambridge University Press.
- Morgenstern, Scott y Benito Nacif (2002), *Legislative Politics in Latin America*, Nueva York, Cambridge University Press.
- Nacif, Benito (2002), “Understanding Party Discipline in the Mexican Chamber of Deputies”, en Scott Morgenstern y Benito Nacif (eds.), *Legislative Politics in Latin America*, Cambridge, Cambridge University Press.
- Poole, Keith T. (2005), *Spatial Models of Parliamentary Voting*, Nueva York, Cambridge University Press.
- Poole, Keith T. y Howard Rosenthal (1997), *Congress: A Political-Economic History of Roll-Call Voting*, Oxford, Oxford University Press.
- Rice, Stewart A. (1925), “The Behavior of Legislative Groups: A Method of Measurement”, *Political Science Quarterly*, 40 (1), pp. 60-72.
- Robles Peiro, Gustavo Adolfo (2009), “Dimensiones espaciales de votación legislativa en la Cámara de Diputados de la LX Legislatura del Congre-

- so de la Unión, 2006- 2009”, tesis de licenciatura, Departamento de Ciencia Política, ITAM.
- Rosas, Guillermo y Joy Langston (2011), “Gubernatorial Effects on the Voting Behavior of National Legislators”, *Journal of Politics*, 73 (2), pp. 477-493.
- Rosas, Guillermo y Yael Shomer (2008), “Models of Non-response in Legislative Politics”, *Legislative Studies Quarterly*, 33 (4), pp. 573-601.
- Weldon, Jeffrey A. (1997), “El presidente como legislador, 1917-1934”, en Pablo Atilio Piccato Rodríguez (ed.), *El poder legislativo en las décadas revolucionarias, 1908-1934*, México, Instituto de Investigaciones Legislativas, Cámara de Diputados, LVI Legislatura, pp. 117-45 (Enciclopedia Parlamentaria de México, serie I, vol. I, t. 3).
- Weldon, Jeffrey A. (2005), “The Fall 2004 Term of the Mexican Congress (Mexican Congressional Report Series)”, Washington, D. C., Center for Strategic and International Studies.

Apéndice

Describimos aquí los pasos tomados para implementar una rutina automatizada capaz de rastrear, en el *Diario de los Debates*, las votaciones nominales celebradas en el pleno. Empezamos determinando, a partir de la página web de la Cámara (<http://cronica.diputados.gob.mx/>), las fechas cuando la Cámara estuvo en sesiones regulares o extraordinarias en las LVII, LVIII y LIX Legislaturas. Una conexión al archivo que contiene el *Diario de los Debates* de cada día (o la *versión estenográfica* para las raras ocasiones en que la conexión al *Diario* falla) fue abierta para leer las líneas de código HTML y componer un objeto de R que llamamos dd. Procedimos después al análisis sintáctico de las líneas de texto en el objeto dd, con la ayuda de expresiones regulares (Friedl, 2006), en busca del aviso de que una votación nominal se va a celebrar. Los votos individuales de los diputados no están disponibles en el *Diario*, pero sí la moción que se somete a consideración del pleno y el resultado de la votación con los votos que obtuvo a favor y en contra agregados. Esto nos proporcionó un listado de las votaciones celebradas para después compararlas con las reportadas por la *Gaceta*.

Leímos (esto es, humanamente) una muestra del *Diario de los Debates* para detectar los patrones utilizados comúnmente cuando se anuncia una votación nominal. Después de que los diputados que fueron reconocidos para hablar en pro y en contra de la moción lo han hecho, el presidente lanza la pregunta de si se considera el tema suficientemente debatido, y el secretario de la Cámara convoca a los diputados a votar con un anuncio protocolario que más o menos se lee así: “Ábrase el sistema electrónico de votación y háganse los avisos a que se refiere el artículo 144 del Reglamento de la Cámara de Diputados”. Antes de la enmienda al reglamento de la Cámara de 2010, el artículo en cuestión era el 161. A menudo se presentan ligeras variaciones al protocolo, como cuando el secretario se brinca alguna palabra o agrega algún adjetivo, así que la búsqueda del patrón fue calibrada, a través de prueba y error, en los *Diarios* que leímos humanamente. Tres patrones de búsqueda señalaron todas las votaciones, con una excepción que tratamos a continuación. Los tres patrones utilizados fueron los siguientes (grep es una función que ejecuta expresiones regulares en R):

```
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="\\b[Aa]rt[í]culo 1[64][14].*de.*[Rr]eglamento")
```

Esta rutina detectó muchos eventos que no eran votaciones nominales, tales como mociones de quórum —los más comunes— o cualquier mención al artículo 141, 144, 161 o 164 de cualquier ley o reglamento. Estas fueron eliminadas de la lista, reteniendo sólo las votaciones nominales.

Verificamos la exactitud del procedimiento al bajar también los *Diarios de los Debates* entre el primero de febrero y el 2 de diciembre de 2010, inclusive, para detectar las votaciones celebradas. Estas fueron sesiones que correspondían a la LXI Legislatura, para la cual el universo de 167 votaciones nominales es proporcionado por el Infopal. La rutina automatizada detectó todos los votos salvo uno —una precisión de más de 99 por ciento—. El voto que faltó pertenece al *Diario* del 15 y 16 de noviembre, cuando el presupuesto anual federal fue tramitado y aprobado en una sesión de 21 horas de duración que involucró discusiones eternas, recesos múltiples y muchas votaciones económicas de procedimiento para intentar introducir enmiendas al dictamen de la comisión. En una instancia, la secretaria convocó a una votación nominal sin hacer el anuncio protocolario correspondiente.

