

# Un modelo Monte Carlo para la Cámara de Diputados en México

Javier Márquez<sup>1</sup> y Francisco Javier Aparicio<sup>2</sup>

<sup>1</sup>*Buendía & Laredo*

<sup>2</sup>*Centro de Investigación y Docencia Económicas, A. C.*

Marzo de 2010

## Resumen

La literatura sobre el sistema político electoral en México a menudo se interesa en evaluar el efecto de una reforma electoral, o bien de alguna otra variable contextual, sobre la correlación de fuerzas en la Cámara de Diputados. Desde un punto de vista empírico, estimar estadísticamente este tipo de efectos es una tarea difícil que requiere de cierta experiencia en programación. Este artículo pretende contribuir al estudio del sistema electoral y del Congreso de dos maneras. Por una parte, presentamos un método estadístico para analizar la composición de Cámara de Diputados en México. Por otro lado, facilitamos su implementación con el software **camaradip**, un módulo para **Stata** desarrollado por los autores, que permite estimar cantidades de interés relacionadas con la Cámara de Diputados a partir de simulaciones de Monte Carlo. Para mostrar la aplicación de nuestro método evaluamos el impacto hipotético de dos reformas electorales en la composición de la Cámara: homologar el calendario de las elecciones locales y federales en todo el país y, por otro lado, disminuir el número de diputados plurinominales.

**Palabras clave:** Cámara de Diputados, Reforma electoral, Simulaciones de Monte Carlo.

# 1. Introducción

México es uno de los países que ha modificado su sistema electoral en más ocasiones en el mundo. Como señala [Weldon \(2001\)](#), en cada una de las cinco elecciones federales celebradas entre 1985 y 1997 se utilizó un sistema electoral distinto. Más recientemente, en 2007 el Congreso aprobó una reforma electoral que afectó la elección federal de 2009, y tras la cual ya se discuten diversas propuestas para modificar una vez más las reglas electorales vigentes, tales como las diversas iniciativas de reforma política presentadas tanto por el Ejecutivo como por las principales bancadas legislativas desde finales de 2009.

Este espíritu reformista de la transición democrática en México ha estado acompañado por un creciente interés por parte de los investigadores del Congreso en evaluar los efectos de distintas reglas electorales en la conformación partidista de la Cámara de Diputados. Aunque en la literatura existen varios trabajos que tratan este tema de manera teórica o empírica, los análisis propiamente estadísticos son prácticamente inexistentes.<sup>1</sup> Una de las causas de esta situación radica en la complejidad propia del método de asignación de curules en un sistema electoral mixto que, aunada al número relativamente grande de partidos con registro en México, hacen que la simulación de escenarios sustantivamente interesantes sea una tarea difícil. Además, la composición del Congreso no es un resultado que se pueda estimar directamente por los modelos estadísticos tradicionales y, en el mejor de los casos, es una tarea que requiere de algunas nociones de programación.

Este artículo pretende contribuir al estudio del Congreso en México de dos maneras. Por una parte, desarrollamos el vínculo entre los modelos estadísticos existentes ([Gelman & King, 1994](#); [Honaker et al., 2002](#); [Katz & King, 1999](#); [Tomz et al., 2002](#)) y la Cámara de Diputados en México. Por otro lado, mostramos la implementación de nuestro método en el software **camaradip**, un módulo de **Stata** ([StataCorp, 2009b](#)) desarrollado por los autores para estimar cantidades de interés relacionadas con la composición de la Cámara de Diputados a partir de simulaciones de Monte Carlo.<sup>2</sup> Creemos que **camaradip** tiene varias características deseables para los investigadores del Congreso mexicano, tales como:

---

<sup>1</sup>Una notable excepción es [Diaz-Cayeros \(2005\)](#).

<sup>2</sup>El paquete funciona en Stata versión 10.0 o superior, y está disponible en línea en: <http://www.buendiaylaredo.com/investigacion> y <http://investigadores.cide.edu/aparicio/camaradip>.

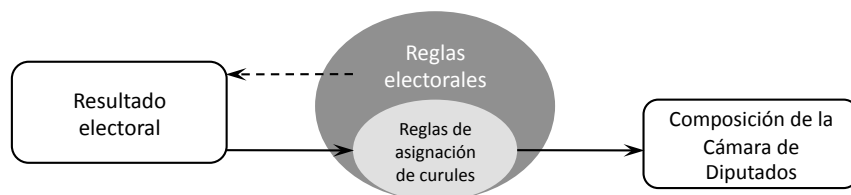
- Está implementado en **Stata**, un paquete estadístico utilizado comúnmente por los científicos sociales.
- Es fácil de usar, pues está basado en unos cuantos comandos sencillos que requieren pocos pasos.
- No requiere aprender nuevos métodos ni modelos estadísticos distintos a los que se usan normalmente en la Ciencia Política contemporánea.
- Está basado en métodos de simulación estadística, una herramienta lo suficientemente flexible como para extraer cantidades de interés relacionadas tanto con los resultados electorales como con la composición de la la Cámara de Diputados en México (por ejemplo, el efecto de los gastos de campaña, métodos de selección de candidatos, cuotas de género, etc.).
- El módulo permite al investigador simular el efecto de un amplio abanico de reglas de asignación de curules (por ejemplo, procesos de redistribución, cambios en el número de curules uninominales y plurinominales, modificaciones en las cláusulas de sobrerrepresentación o bien en los umbrales de representación, etc.).

El artículo está estructurado de la siguiente manera. En la segunda sección abordamos algunos aspectos preliminares de nuestro método estadístico para la Cámara de Diputados, mismos que describimos de manera más formal en la tercera parte. En la cuarta sección mostramos la implementación del paquete **camaradip**, seguida de dos aplicaciones ilustrativas: el efecto de la concurrencia entre elecciones federales y locales y, por otro lado, el efecto de reducir el número de curules plurinominales. Finalmente, discutimos algunas áreas de investigación para estudios futuros.

## 2. Discusión preliminar

El objetivo principal del método estadístico que aquí proponemos es extraer cantidades de interés relacionadas con la conformación partidista de la Cámara de Diputados. Denotemos al número de asientos o curules que le corresponden a cada uno de los partidos políticos con representación en la Cámara como  $\psi = [\psi_1, \psi_2, \dots, \psi_J]$ . El número de curules para cada partido depende fundamentalmente de dos variables: las preferencias de los votantes,

las cuales se observan de manera agregada en el resultado electoral, y las reglas electorales, entre las que destacan las reglas de asignación de curules (Figura 1). Las preferencias de los votantes se manifiestan en los votos  $V = [V_1, V_2, \dots, V_J]$  para cada uno de los  $J$  partidos ( $j = 1, \dots, J$ ), que a su vez se transforman en posiciones en la Cámara a través de las reglas de asignación.



**Figura 1:** Reglas electorales y composición de la Cámara de Diputados

Generalmente, el investigador desea evaluar el cambio en el número de asientos,  $\psi$ , dado un cambio en las reglas electorales. Nuestro modelo requiere diferenciar dos tipos de reglas. Las primeras son aquellas que afectan el resultado electoral al alterar la participación y/o las preferencias de los votantes (como por ejemplo, los gastos de campaña de los candidatos, el método de selección de los candidatos, o cuotas de género en las candidaturas). Las preguntas de investigación relacionadas con estas reglas deben plantearse en términos de “efectos contextuales”, y para fines de estimación deben ser observables o medibles a nivel de los distritos electorales.<sup>3</sup> El segundo tipo de reglas se refieren explícitamente a la asignación de curules, es decir, aquellas reglas que determinan *mecánicamente* el procedimiento mediante el cual los porcentajes de votos se traducen en el reparto de curules entre los partidos; por ejemplo, el porcentaje o umbral mínimo de votación para tener derecho a curules de representación proporcional, o el tope máximo de sobrerrepresentación en la Cámara.

Nuestro método estadístico sirve para elaborar predicciones, explicaciones o estimar contrafactuales de la Cámara de Diputados cuando una reforma

<sup>3</sup>Por otro lado, cuando el objetivo es analizar un comportamiento individual que genera un resultado agregado, es recomendable emplear un modelo de inferencia ecológica; véase King, 1997.

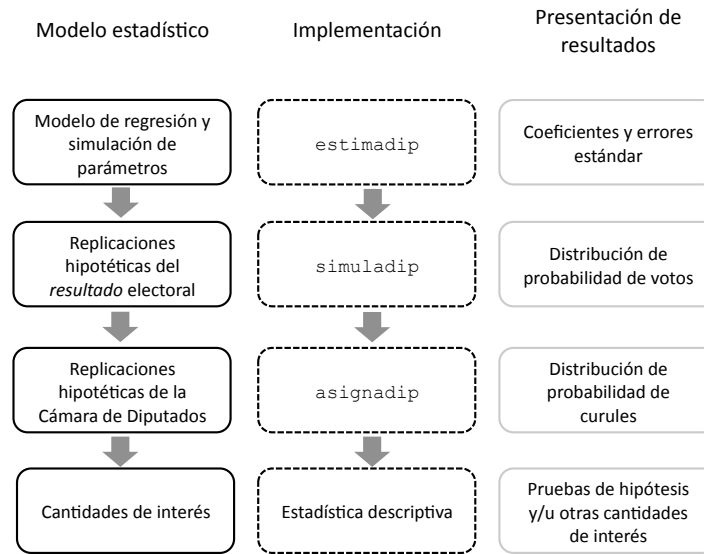
electoral modifica los resultados electorales y/o el método de asignación de curules. Formalmente, el objetivo es describir a la distribución de probabilidad de curules,  $p(\psi)$ .<sup>4</sup> Una manera de hacer esto es a través de simulaciones de Monte Carlo (Jackman, 2000a,b, 2004; Martin, 2008). Es decir, en lugar de encontrar una solución analítica para estimar el efecto de un cambio en las reglas electorales, podemos obtener varios valores de  $p(\psi | V)$  con la ayuda de un generador de números (seudo)aleatorios, agregar dichos valores, y hacer inferencias con ellos mediante estadísticas descriptivas. Con este método se pueden simular diferentes composiciones hipotéticas de la Cámara a partir de una serie aleatoria de resultados electorales. Diaz-Cayeros (2005) utiliza esta técnica para analizar la composición del Senado en México bajo distintas reglas de asignación de escaños.

El problema se torna un poco más complicado cuando se analiza el efecto de una reforma que presumiblemente puede aleterar los resultados de una elección —por ejemplo, al cambiar las preferencias de los votantes o al cambiar las opciones que se le presentan—, pues resulta imposible simular valores aleatorios de  $p(\psi | V)$  de manera directa. No obstante, el método de Monte Carlo sigue siendo una herramienta útil para nuestra aplicación. El algoritmo que aquí se propone incluye estimar un modelo de regresión lineal para explicar los resultados electorales,  $V$ , en función de ciertos parámetros  $\theta$ , de modo que podemos simular valores de la distribución  $p(\theta | V)$  y, a partir de esos valores estimar la distribución de asientos  $\psi = h(\theta)$  (véanse Jackman, 2000a, 2009: cap. 3).

Así pues, en términos generales, el modelo estadístico para la Cámara de Diputados que se propone comprende cuatro etapas que se ilustran en la primera columna de la Figura 2: 1) especificación y simulación de los parámetros de un modelo de regresión, 2) generación de replicaciones hipotéticas del resultado electoral, así como 3) de la conformación de la Cámara de Diputados, y 4) extracción de cantidades de interés. Con ello el investigador puede analizar tanto la tendencia central como la dispersión de las distribuciones de probabilidad de votos, las distribuciones del número de curules de cada partido, o bien otras cantidades que se derivan de la distribución de votos y asientos—por ejemplo, la probabilidad de que la bancada de un partido sea pivotal en una coalición mínima-ganadora.

---

<sup>4</sup>La distribución de probabilidad es el rango de valores que un parámetro o estadístico puede tomar en una muestra aleatoria de cierta población, y las probabilidades asociadas a esos valores.



**Figura 2:** Modelo estadístico de la Cámara de Diputados

En la siguiente sección abordamos brevemente las cuatro etapas del modelo estadístico. En la sección 4 exponemos intuitivamente nuestra implementación del modelo en el paquete **camaradip** de **Stata** (ver segunda columna de la Figura 2). En la sección 5 presentamos dos aplicaciones concretas del método que proponemos para extraer cantidades de interés (tercera columna de la Figura 2).

### 3. Modelo estadístico

#### 3.1. Estimación con datos electorales multipartidistas

Como mencionamos en la sección anterior, el número de asientos que le corresponden a cada partido en la Cámara depende del resultado electoral (la manifestación en votos de las preferencias de los votantes), y de las reglas electorales. En esta parte abordaremos el primero de ellos.

La manera más común de explicar o predecir estadísticamente un resultado electoral es a través de un modelo de regresión (Gelman & Hill, 2007: caps. 3 y 4). Normalmente, el investigador estima el porcentaje de votación de un partido político como una función de diversos factores explicativos a través de un modelo de regresión de Mínimos Cuadrados Ordinarios (MCO).

Desafortunadamente, el modelo estándar de regresión es inapropiado para estimar los resultados electorales de un sistema multipartidista (Katz & King, 1999). Por ejemplo, los resultados de la regresión pueden indicar que un partido podría obtener menos de cero votos, o bien que la suma de los votos de todos los partidos podría ser menor o mayor a 100 %.

Estas inconsistencias surgen porque los resultados electorales multipartidistas violan dos supuestos subyacentes del modelo estándar. En primer lugar, MCO asume que la variable dependiente (en este caso, la proporción de votos de un partido) es una variable continua irrestricta, es decir, que puede tomar cualquier valor. Naturalmente, la proporción de votos de un partido se encuentra necesariamente acotada entre cero y uno. Denotemos a  $V_{ij}$  como la proporción de votos del partido  $j$  ( $j = 1, \dots, J$ ) en el distrito electoral  $i$  ( $i = 1, \dots, n$ ). Formalmente,

$$V_{ij} \in [0, 1] \quad \forall i, j \quad (1)$$

En segundo lugar, el modelo estándar presupone que la proporción de votos de un partido es ortogonal o independiente de la de los demás; sin embargo, es claro que en la composición porcentual de una contienda electoral, el porcentaje de votos de un partido está inversamente relacionado con el de los demás toda vez que la suma de todas las proporciones de votos es igual a uno. Dicho de manera formal,

$$\sum_{j=1}^J V_{ij} = 1 \quad \forall i \quad (2)$$

Para subsanar las inconsistencias del modelo estándar, Katz & King (1999) recomiendan modelar la proporción de votos de los partidos con una transformación logística multivariada (Aitchison, 1982). Esta técnica consiste en transformar a *log ratios* la proporción de votos del partido  $j = 1, \dots, J - 1$ , respecto de un partido base  $J$ . Así, el vector de  $J - 1$  *log ratios* en el distrito  $i$  se representa de la siguiente manera:

$$Y_i = [\ln(V_{i1}/V_{iJ}), \ln(V_{i2}/V_{iJ}), \dots, \ln(V_{i(J-1)}/V_{iJ})]$$

La transformación logística convierte las proporciones de votos en una escala continua irrestricta —como requiere MCO—, y tras la estimación las recupera en su escala original (satisfaciendo (1) y (2)) con una transformación logística inversa. Para fines de estimación, asumimos que el vector de *log*

*ratios* tiene una distribución normal multivariada con su media en el vector  $\mu$  y matriz de varianzas  $\Sigma$ ,  $Y_i \sim N(\mu_i, \Sigma)$  (Tomz et al., 2002).<sup>5</sup> Luego entonces, el resultado electoral esperado en cada distrito,  $\mu_i$ , se puede modelar como una función lineal de un vector de variables explicativas  $X_i$  y los parámetros  $\beta$ , tal que

$$\mu_i = [x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{i(J-1)}\beta_{(J-1)}] \quad (3)$$

Al igual que en cualquier modelo de regresión, la elección de qué variables incluir en el vector  $X_i$  depende de nuestra pregunta de investigación. Si el objetivo es evaluar el efecto causal de una variable en la votación de los partidos (y por lo tanto, en las curules que les corresponden), este vector debe excluir aquellas variables que son en parte consecuencia de nuestra variable explicativa clave (Rosenbaum, 1984; King et al., 1994). En cambio, si el objetivo es hacer pronósticos del resultado electoral (para luego evaluar el efecto de una regla de asignación de curules), la selección de las variables explicativas se rige por otros criterios, siendo el más importante que ayuden a predecir lo mejor posible a la variable dependiente (Gelman & Hill, 2007: 69).

El sistema de ecuaciones en (3) puede estimarse con  $J - 1$  modelos de regresión con MCO; sin embargo, Jackson (2002) y Tomz et al. (2002) recomiendan utilizar Regresiones Aparentemente Inconexas (*Seemingly Unrelated Regression* o SUR en inglés) por varias razones. En primer lugar, este método es útil para estimar sistemas de ecuaciones con errores correlacionados, y los datos multipartidistas tienen esa estructura dado que un mayor *log ratio* para un partido significa un menor *log ratio* para otro. En segundo lugar, si las variables explicativas difieren de una ecuación a otra (como puede ocurrir en las aplicaciones del modelo), SUR incorpora la covarianza de las ecuaciones para obtener estimadores más eficientes que MCO. En tercer lugar, SUR tiene un buen desempeño en muestras pequeñas (cuando el número de distritos

---

<sup>5</sup>Katz & King (1999) argumentan que la distribución normal multivariada no se ajusta correctamente a los datos electorales multipartidistas, y que en su lugar debería emplearse la distribución *t*. Sin embargo, los experimentos de Tomz et al. (2002) muestran que ambas distribuciones arrojan resultados muy parecidos, en particular cuando las cantidades de interés son votos o curules en la legislatura. Como los autores afirman, “[b]y adopting the Normal we lose little of substantive interest but gain quite a bit in ease of use” (p.71). Su investigación también cita los hallazgos de Breusch et al. (1997), quienes sostienen que “although mathematically the two models are different, for purposes of statistical inference they are indistinguishable” (p.269).



es limitado), y es más flexible que otras alternativas cuando el número de partidos es relativamente grande ( $J > 3$ ). Finalmente, la implementación de SUR en los paquetes estadísticos más comunes facilita al investigador estimar las  $J - 1$  ecuaciones con pocos comandos.

### 3.2. Simulación estadística y cantidades de interés

Como normalmente ocurre con otros modelos de regresión, los parámetros en (3) son difíciles de interpretar y están relacionados sólo de manera indirecta con nuestras preguntas de investigación. Los coeficientes del modelo indican el cambio en el *log ratio* de los votos de un partido frente a un incremento de una unidad en las variables explicativas. Como veremos más adelante, para interpretar los resultados en términos de *porcentajes de votos* se requiere aplicar la transformación logística inversa (ecuación 5).

Pero además, para los estudiosos del poder legislativo los votos no son importantes por sí mismos. Los votos cobran relevancia cuando se traducen en curules en la Cámara y éstos, a su vez, tienen un impacto en el control y diseño de políticas públicas concretas. Por ejemplo, para los investigadores del comportamiento electoral puede ser importante conocer el efecto que ejercen las elecciones concurrentes sobre la votación de los partidos; pero a los especialistas del poder legislativo podría resultarles más interesante saber si una reforma encaminada a homologar el calendario electoral alteraría de manera significativa la correlación de fuerzas en el Congreso. Si todas las elecciones locales y federales fueran concurrentes, ¿cuál sería la composición partidista de la Cámara de Diputados? ¿La nueva constelación sobrerrepresentaría a un partido en detrimento de los demás? ¿Aumentaría la probabilidad de que el partido del Presidente tenga mayoría simple en la Cámara? Y si la reforma beneficia a la oposición, ¿sus ganancias serían difusas, o se concentrarían en un solo partido?

Dicho de manera más general, el objetivo del modelo se puede reformular como estimar la distribución de probabilidad de curules  $\psi$  como una función de los parámetros del modelo de regresión de votos  $\theta = [\beta, \Sigma]$ ,

$$\psi = h(\theta)$$

El éxito en la empresa depende de la naturaleza de la cantidad de interés: algunas pueden calcularse fácilmente con métodos analíticos tradicionales (e.g., Gelman & King, 1994: 532); otras requieren métodos más difíciles,

como aproximaciones por series de Taylor (Katz & King, 1999: 25), y algunas son imposibles de obtener con cualquiera de los métodos anteriores. Dada la complejidad de las reglas de asignación, la distribución del número de curules de un partido en la Camara Diputados puede ser difícil o imposible de obtener de manera analítica. Por eso, proponemos calcularla empíricamente con la ayuda de simulaciones. Como King et al. (2000) afirman:

Existe una alternativa basada en simulación para casi cualquier método analítico empleado para calcular cantidades de interés y para realizar pruebas estadísticas, pero no al revés. Así pues, la simulación estadística puede ofrecer respuestas precisas incluso cuando no existen soluciones analíticas (p. 53).

La simulación estadística se basa en el principio de Monte Carlo, según el cual podemos conocer o describir cualquier variable aleatoria obteniendo una muestra de valores que pertenecen a su distribución de probabilidad (Jackman, 2009). De acuerdo con el Teorema del Límite Central, la distribución de probabilidad de los parámetros es (asintóticamente) normal multivarada:<sup>6</sup>

$$\tilde{\theta} \sim N(\hat{\theta}, \hat{v}\hat{a}r(\hat{\theta})) \quad (4)$$

Por lo tanto, podemos aproximarnos a la distribución de probabilidad de  $\psi$  seleccionando o *simulando* aleatoriamente  $t = 1, 2, \dots, T$  valores  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)}$  de (4), y evaluando  $\psi^{(t)} = h(\theta^{(t)})$  con cada uno de ellos.

Específicamente, el algoritmo para simular *un* valor de la distribución de  $\psi$  involucra los siguientes pasos:

---

<sup>6</sup>Este supuesto distribucional es relativamente restrictivo. Como Jackman (2000a) afirma,

If the asymptotic normal approximation is poor, faulty inferences and model predictions can result. This is a real danger for “postestimation” simulation procedures [...], which repeatedly sample from the asymptotic multivariate normal distribution for  $\theta$  to build an approximation to the posterior of an auxiliary quantity  $\psi = h(\theta)$  (pag. 309).

Las simulaciones de Monte Carlo generadas con Cadenas de Markov ofrecen una solución a este problema, pues permiten trazar valores de la distribución finita, en lugar de basarse en el Teorema del Límite Central para justificar la aproximación normal multivarada (Western & Jackman, 1994). El comando `asignadip` resulta útil para los investigadores que prefieren utilizar este método para hallar la distribución posterior del porcentaje de votos de los partidos.

1. Estimar el vector de parámetros  $\hat{\theta}$  mediante SUR.
2. Seleccionar aleatoriamente (*simular*) un vector de parámetros  $\theta$  a partir de la distribución normal multivariada dada en (4).
3. Elegir valores reales (observados) o bien valores hipotéticos para las variables explicativas del distrito  $i$  de acuerdo a la pregunta de investigación. Denotemos al vector de valores como  $X_i^{hip}$ . Si se desea evaluar únicamente el impacto de algún método de asignación de curules,  $X_i^{hip} = X_i$ .
4. Calcular el vector de resultados electorales  $\tilde{\mu}_i^{hip} = [x_{i1}^{hip}\tilde{\beta}_1 + x_{i2}^{hip}\tilde{\beta}_2 + \dots + x_{i(J-1)}^{hip}\tilde{\beta}_{(J-1)}]$  con las simulaciones de  $\tilde{\beta}$  del paso 2 y el vector  $X_i^{hip}$ .
5. Simular  $\tilde{Y}_i^{hip}$  seleccionando aleatoriamente un valor de la distribución  $N(\tilde{\mu}_i^{hip}, \tilde{\Sigma})$ .<sup>7</sup>
6. Transformar  $\tilde{Y}_i^{hip}$  en una proporción de votos con la función logística inversa (Aitchison, 1982):<sup>8</sup>

$$\tilde{V}_{ij}^{hip} = \frac{\exp(\tilde{Y}_{ij}^{hip})}{1 + \sum_{j=1}^{J-1} \exp(\tilde{Y}_{ij}^{hip})} \quad \forall j \neq J \quad (5)$$

La predicción del partido usado como base o referencia en la transformación logística,  $\tilde{V}_{iJ}^{hip}$ , está determinada por la expresión:  $\tilde{V}_{iJ}^{hip} = 1 - \sum_{j=1}^{J-1} \tilde{V}_{ij}^{hip}$ .

7. Repetir los pasos 3 a 6 para cada uno de los  $n$  distritos. Agregar los resultados en el vector  $\tilde{V}^{hip}$ .
8. Evaluar la distribución de curules  $\tilde{\psi}^{hip} = g(\tilde{V}^{hip})$ , donde  $g$  es una función que representa el procedimiento de asignación de curules de acuerdo a la ley electoral, o bien otra regla hipotética de asignación establecida *a priori* por el investigador.

---

<sup>7</sup>Si se desea obtener un valor esperado en lugar de un valor predicho,  $\tilde{Y}_i^{hip} = \tilde{\mu}_i^{hip}$ . Es importante señalar que esta propiedad no es generalizable a otros modelos estadísticos (véase King et al., 2000: 351).

<sup>8</sup>En este paso agregamos el subíndice  $j$  para diferenciar los valores predichos de los partidos en el modelo del valor predicho del partido usado como referencia en la transformación logística multivariada.

Si se repiten los pasos 2 a 8 un número relativamente grande de veces (digamos,  $T = 1,000$ ), obtenemos  $T$  asignaciones hipotéticas de la Cámara de Diputados, cuya distribución se aproxima a la distribución de probabilidad completa de  $\psi^{hip}$ .

## 4. Implementación

El algoritmo de la sección anterior guarda similitudes con el descrito por King et al. (2000), e implementado en *Clarify* por Tomz et al. (2003). De hecho, la implementación de nuestro método se apoya en algunas de las rutinas de ese programa; no obstante, nuestra rutina es más parecida a la del “Porcentaje Correctamente Predicho” formulada por Herron (1999).<sup>9</sup> Las diferencias entre *Clarify* y nuestro método pueden ilustrarse fácilmente con la Tabla 1 (adaptada de Gelman & King, 1994). Cada fila representa un distrito, con el número del distrito en la primera columna y el porcentaje de votos observado del partido  $j$  en la segunda columna. Las demás columnas representan simulaciones de la distribución del porcentaje de votos. Por ejemplo, el voto observado del partido  $j$  en el distrito 1 es  $V_{1j}$ , y la primera simulación es  $\tilde{V}_{1j}^{(hip)1}$ , la segunda es  $\tilde{V}_{1j}^{(hip)2}$ , y así sucesivamente.

Normalmente, el investigador utiliza *Clarify* para analizar el efecto de una variable explicativa bajo escenarios específicos. En términos generales, el procedimiento consiste en simular los parámetros del modelo de regresión, fijar las variables explicativas en valores hipotéticos de interés (digamos, en sus valores promedio), y calcular el valor esperado de la variable dependiente (Katz & King, 1999; King et al., 2000). En otras palabras, *Clarify* suele emplearse para obtener  $T$  valores esperados de *un* distrito promedio (i.e., una fila de la Tabla 1). En cambio, nuestro método utiliza como insumo las  $T$  replicaciones hipotéticas del *resultado electoral agregado*, lo cual implica trazar simulaciones para cada uno de los distritos (Gelman & King, 1994). En la Tabla 1 se observa que simulando  $T$  valores de la distribución de votos para cada distrito (filas), se obtienen  $T$  resultados electorales hipotéticos para cada partido (columnas). Considerando el número de partidos y distritos en la última elección en México, usando *Clarify* se requerirían aproximadamente 5,000 líneas de código (comandos) para obtener una matriz como la de la Tabla 1. El módulo **camaradip** que documentamos en el Apéndice provee un

<sup>9</sup>También véase Tomz et al. (2002).

*wrapper* que automatiza y simplifica notablemente esta tarea y genera, con un solo comando (`simuladip`), una base datos con  $T$  elecciones simuladas para cada distrito y partido similar a la Tabla 1.

**Cuadro 1:** Estructura del modelo estadístico de la Cámara de Diputados (adaptada de Gelman & King, 1994)

Distrito	Resultado observado	Elecciones hipotéticas			
		1	2	...	$T$
1	$V_{1j}$	$\tilde{V}_{1j}^{(hip)1}$	$\tilde{V}_{1j}^{(hip)2}$	...	$\tilde{V}_{1j}^{(hip)T}$
2	$V_{2j}$	$\tilde{V}_{2j}^{(hip)1}$	$\tilde{V}_{2j}^{(hip)2}$	...	$\tilde{V}_{2j}^{(hip)T}$
⋮	⋮	⋮	⋮	⋮	⋮
$n$	$V_{nj}$	$\tilde{V}_{nj}^{(hip)1}$	$\tilde{V}_{nj}^{(hip)2}$	...	$\tilde{V}_{nj}^{(hip)T}$
Cantidad de interés	$\psi_j$	$\tilde{\psi}_j^{(hip)1}$	$\tilde{\psi}_j^{(hip)2}$	...	$\tilde{\psi}_j^{(hip)T}$

Otra diferencia más evidente entre *Clarify* y nuestro algoritmo tiene que ver con la naturaleza de las cantidades de interés. La última fila de la Tabla 1 representa el número de curules del partido  $j$ , denotado por  $\psi_j$  para el resultado observado, y por  $\tilde{\psi}_j^{(hip)t}$  para las replicaciones hipotéticas. Las cantidades de interés se obtienen con la información de su respectiva columna. Para calcular la primera replicación  $\tilde{\psi}_j^{(hip)1}$ , debemos contar el número de distritos en que el partido  $j$  obtuvo más votos que los demás partidos en la primera columna de replicaciones hipotéticas (lo cual representa su número de curules uninominales), y luego asignar las curules plurinominales a las que tendría derecho dada su votación en *todos* los distritos de esa misma columna. Este ejercicio se repite sucesivamente para cada una de las elecciones hipotéticas de manera que la última fila de la tabla está formada por  $T$  simulaciones de la distribución de curules  $\tilde{\psi}_j^{(hip)}$ .

Naturalmente, entre mayor sea el número de simulaciones  $T$ , la tarea de obtener  $\tilde{\psi}^{(hip)}$  se torna tediosa y complicada. Para solucionar este problema resulta práctico generar un programa que realice la asignación de curules por nosotros. El comando `asignadip` del paquete **camaradip** ofrece esa herramienta. Por omisión, `asignadip` aplica el método de asignación ordenado por el COFIPE y que se ha utilizado en México desde la elección federal de 1997. Dicho método consiste en la asignación de 300 diputados uninominales por la regla de mayoría relativa, la asignación de 200 diputados plurinominales

por el método de resto mayor, y la reasignación de curules plurinominales cuando un partido tiene más de 300 curules o su porcentaje de curules por ambos principios excede en 8 puntos o más el porcentaje de votación nacional obtenido.<sup>10</sup> Pero además, las opciones de `asignadip` son lo suficientemente flexibles como para que el investigador pueda experimentar con varios métodos de asignación; por ejemplo, modificar el número de curules uninominales y/o plurinominales, realizar la asignación en varias circunscripciones o en una circunscripción nacional, disminuir el límite máximo de sobrerrepresentación, aumentar el umbral mínimo para tener derecho al reparto de curules plurinominales, etc.

En su forma más simple, el modelo estadístico para la Cámara de Diputados puede implementarse en `Stata` con tres sencillos pasos (Figura 2):

- Transformar los porcentajes de votos de los partidos (`v1`, `v2`, `v3`, `v4`) en *log ratios*; estimar y simular los parámetros del modelo de regresión con el comando `estimadip`:

```
. estimadip (v1 x z) (v2 x z) (v3 x z), base(v4)
```

- Fijar el valor de las reglas electorales (`x`) y de otras variables explicativas (`z`) en sus valores reales o hipotéticos; calcular la distribución predictiva de votos; aplicar la transformación logística inversa; y guardar los resultados en las variables `p1`, `p2`, `p3`, `p4` de la base `simfile`, con el comando `simuladip`:

```
. simuladip using simfile, gen(p1 p2 p3) set(x 1=0, z 0=1)
```

- Calcular la distribución de probabilidad de la composición de la Cámara de Diputados con ayuda de un *loop* (Cox, 2002) y del comando `asignadip`:

```
. use simfile
. forvalues i = 1/1000 {
2.     asignadip p1 p2 p3 p4 if _IDsim == 'i'
3. }
```

---

<sup>10</sup>Por “porcentaje de votación” se entiende la votación nacional emitida. Esta disposición no se aplica cuando la sobrerrepresentación se debe a los triunfos en distritos uninominales.

Con unas pequeñas modificaciones—utilizando el comando dos veces—, `asignadip` se puede adaptar a la reglamentación de las coaliciones electorales vigente a partir de la elección federal del 2009.<sup>11</sup>

Otra de las virtudes de `asignadip` es que el usuario puede acceder a los resultados por medio de escalares y macros guardados en `r()` con la función `return` (StataCorp, 2009a). Esta utilería es una manera práctica de guardar las distribuciones del número de curules en variables que pueden manipularse fácilmente con otros comandos de estadística descriptiva.

Para presentar los resultados del análisis, las distribuciones pueden describirse con medidas de tendencia central (e.g., media, moda, mediana) o medidas de dispersión (e.g., desviaciones estándar o percentiles que delimitan un intervalo de confianza). También pueden efectuarse pruebas de hipótesis contando la fracción de veces que el número de curules de un partido es mayor o menor a una cifra cualquiera. Por ejemplo, para calcular la probabilidad de que un partido tenga mayoría simple en la Cámara, basta con sumar las veces que el número predicho de curules es igual o mayor a 251, y dividirlo entre  $T$ . Además, la distribución del número de curules puede transformarse en otras distribuciones, como la del número efectivo de partidos en la legislatura, índices de poder, coaliciones mínimas ganadoras, etc. En la siguiente sección mostramos cómo el usuario puede extraer éstas y otras cantidades de interés.

## 5. Aplicaciones

### 5.1. Elecciones concurrentes

En esta sección presentamos una aplicación del modelo estadístico de la Cámara de Diputados para evaluar una reforma que homologa el calendario de las elecciones locales y federales en todo el país: es decir, elegir a gobernadores y gobiernos locales al mismo tiempo que a los diputados federales. Las elecciones concurrentes pueden producir efectos de arrastre entre el voto por el presidente o gobernador —cargos unipersonales de alta visibilidad— y el voto por los legisladores federales o locales (Mondak & McCurley, 1994). Este tipo de efectos se han documentado en países como Argentina, Brasil y México (Jones, 1997; Samuels, 2000; Magar, 2006).

---

<sup>11</sup>Véase el archivo de ayuda tecleando `help asignadip` en la ventana de comandos de Stata.

Los datos para el análisis provienen de los cómputos distritales de la elección para Diputados Federales del 2009, agregados por distrito electoral federal ( $n = 300$ ) (IFE, 2009). Las variables dependientes son el *log ratio* de la proporción de votos del PAN, PRI, PRD, PVEM, PT, Convergencia, y Nueva Alianza con respecto a la proporción de votos del PSD. La variable explicativa clave es una variable dicotómica que toma el valor de 1 si en un distrito se llevaron a cabo elecciones para elegir Gobernador y/o Presidentes Municipales, y 0 en caso contrario. Existen razones para suponer que el efecto de la concurrencia también puede verse afectado por el partido político que gobierna el estado; por lo tanto, incluimos en la regresión los términos constitutivos de la interacción entre la variable de concurrencia y otras dos variables que indican si el el Gobernador del estado es de filiación panista o perredista<sup>12</sup> (Brambor et al., 2006). También incluimos la proporción de votos de cada partido o coalición en la elección federal previa.<sup>13</sup> De este modo, podemos estimar el efecto de una elección concurrente para los diferentes partidos en el gobierno estatal, controlando por la fuerza electoral de cada partido a nivel distrital. Es claro que esta especificación podría mejorarse incluyendo otras variables que capturen el posible efecto de arrastre de las elecciones locales; sin embargo, nuestro propósito aquí es meramente ilustrar con un ejemplo la aplicación de nuestro modelo estadístico para un fenómeno de interés.<sup>14</sup>

Con este modelo base podemos estimar el resultado electoral en distritos con y sin elecciones locales concurrentes. El siguiente paso consiste en estimar el resultado electoral bajo un escenario hipotético en que todos los distritos del país hubieran tenido elecciones concurrentes. Para hacer esto, simulamos mil resultados electorales fijando el valor de nuestra variable binaria clave (*concurrencia*) igual a 1 para todos los distritos (y modificando sus respectivas interacciones partidistas), mientras que mantenemos a las demás variables en sus valores observados.<sup>15</sup>

El Panel A de la Figura 3 muestra la distribución del voto distrital simulado para los tres principales partidos políticos, y lo compara con el voto

---

<sup>12</sup>Los gobiernos priístas son la categoría de referencia.

<sup>13</sup>Los detalles de la codificación están disponibles previa solicitud a los autores.

<sup>14</sup>Por ejemplo, podría distinguirse entre la concurrencia de elecciones municipales y federales, por un lado, y la de gobernadores, por otro. Para hacer esto, el número de interacciones partidistas también tendría que duplicarse. Dado el reducido número de casos de elecciones concurrentes del primer tipo en 2009, estimamos el efecto de la concurrencia sin hacer tal distinción.

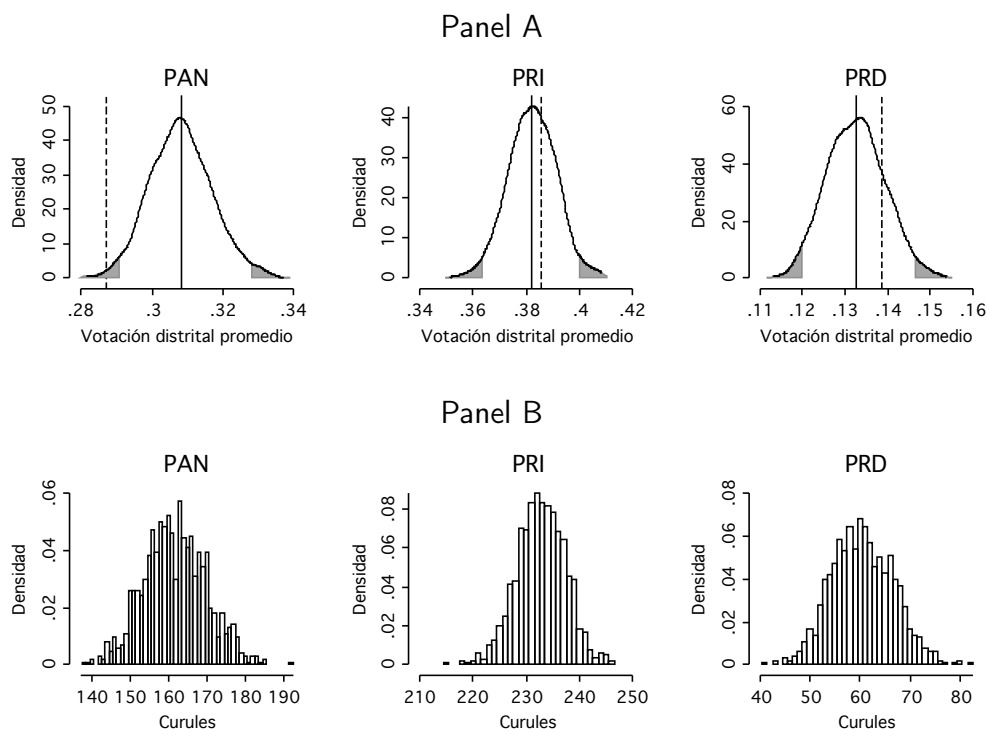
<sup>15</sup>El código para replicar este análisis está disponible previa solicitud a los autores.



realmente observado en 2009. Las líneas sólidas en cada gráfica indican la media de cada distribución simulada bajo concurrencia, mientras que las líneas punteadas representan el voto distrital promedio observado, y las áreas sombreadas indican los intervalos de confianza al 95 %. Como se puede apreciar, el resultado electoral promedio hubiera sido prácticamente el mismo para el PRI y para el PRD (perderían en promedio menos de un punto porcentual en cada distrito) mientras que el PAN obtendría aproximadamente dos puntos porcentuales más en el escenario contrafactual. La distancia entre el voto simulado y observado solamente es estadísticamente significativa en el caso del PAN. Este resultado se debe a que el PAN obtuvo más votos en las entidades con elecciones concurrentes en 2009 que en el resto de los estados.<sup>16</sup>

---

<sup>16</sup>Agradecemos a un dictaminador anónimo por señalar este punto.

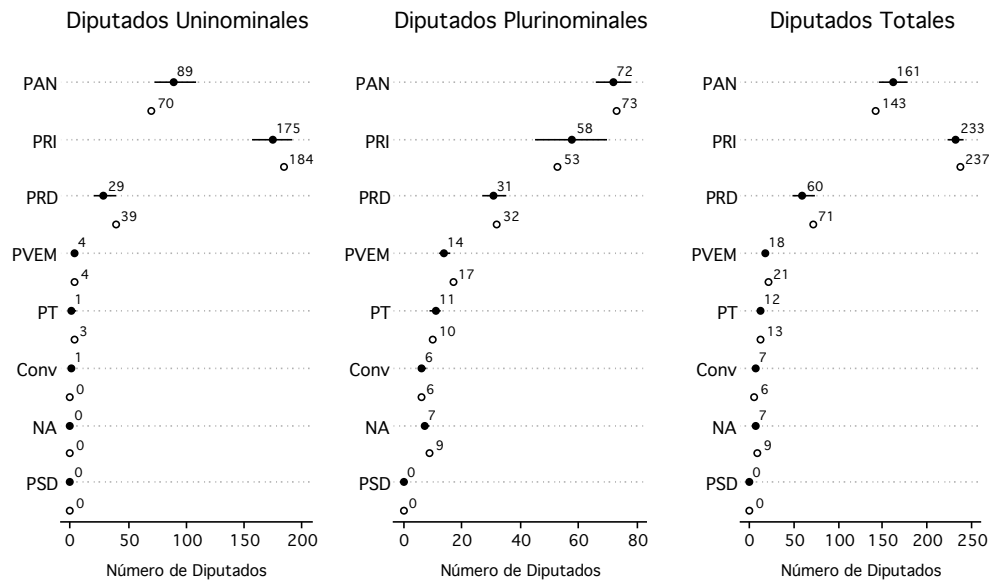


**Figura 3:** Panel A: Distribución de densidad de la proporción distrital de votos por partido político con la reforma que homologa el calendario de las elecciones federales y locales. Las líneas verticales representan la media de la distribución simulada; las líneas punteadas indican la proporción de votos observada en la elección federal de 2009. Panel B: Distribución de densidad del número de curules por partido político con la reforma que homologa el calendario de las elecciones federales y locales.

El efecto que ejercen las elecciones concurrentes sobre el porcentaje de votos tienen a su vez un impacto en el número de curules para cada partido. El Panel B de la Figura 3 muestra las distribuciones de probabilidad del número de curules obtenido por los principales partidos políticos en el escenario hipotético. Estas distribuciones se obtuvieron a partir de las votaciones simuladas en las gráficas del Panel A, y con el método de asignación de curules previsto en la ley electoral vigente. Las líneas sólidas indican la moda de las distribuciones, y las líneas punteadas el número de curules observado. Como se aprecia en la gráfica del PRI, la reforma no afectaría de manera significativa el tamaño de su bancada, mientras que la del PAN aumentaría en aproxi-

madamente 18 diputados y la del PRD disminuiría en aproximadamente 11 diputados.

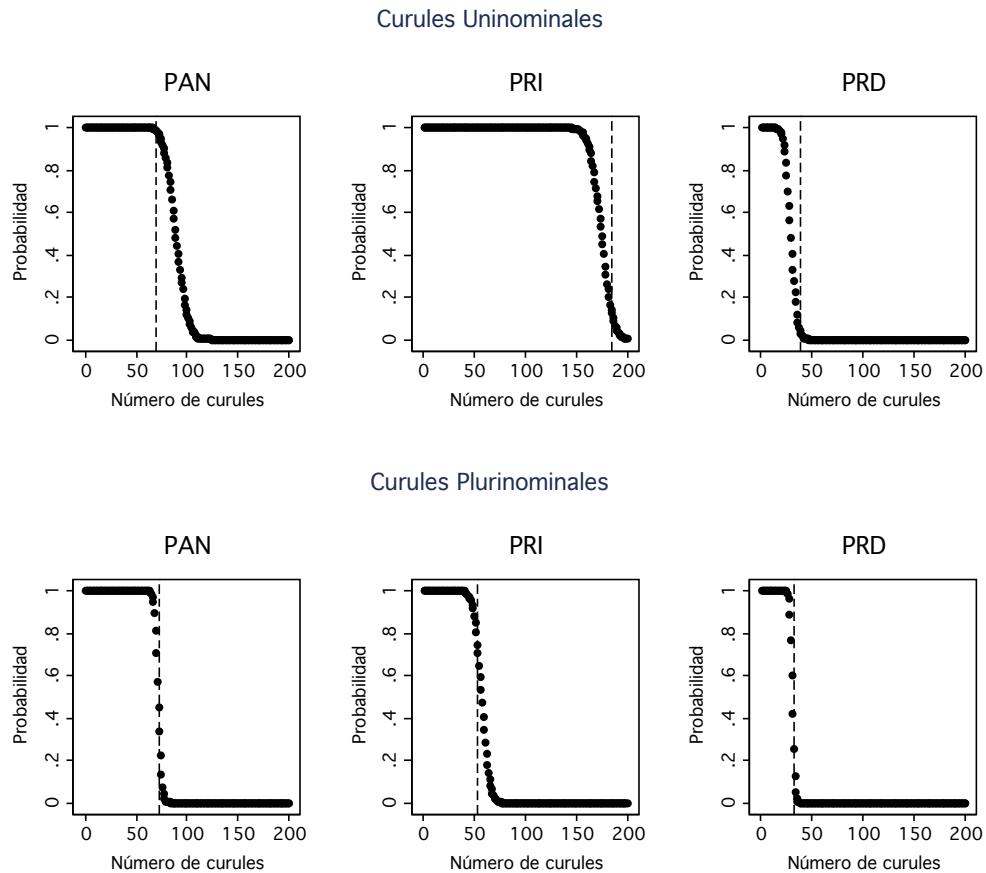
La Figura 4 ofrece algunos elementos para entender mejor estos resultados. Las gráficas muestran las mismas distribuciones del Panel B de la Figura 3, bajo un formato que enfatiza el valor promedio de cada distribución y distingue por separado el resultado para los diputados uninominales y plurinominales. En primer lugar, la gráfica de las curules de mayoría muestra que los cambios que produce la reforma se registran principalmente en el componente mayoritario del sistema electoral. El PAN obtendría 19 diputados uninominales más de los que consiguió en 2009, el PRD 10 diputados menos, y el PRI 9 diputados menos. Estas cifras sugieren que los cambios relativamente pequeños en el porcentaje de voto esperado para cada partido en el Panel A de la Figura 3 o bien provienen de cambios abruptos en unos cuantos distritos, o bien afectan a varios distritos altamente competidos donde el partido ganador cambia dado el estrecho margen de victoria. Toda vez que el método realiza simulaciones para cada distrito, podemos distinguir entre uno u otro tipo de resultado.



**Figura 4:** Número de Diputados por partido político. Los puntos representan la media de la distribución de probabilidad del número de curules con la reforma que homologa el calendario de las elecciones federales y locales. Los círculos indican el número de curules observado en la elección federal de 2009. Las líneas horizontales representan intervalos de confianza al 95 %.

En segundo lugar, la Figura 4 muestra que el número de diputados plurinominales es muy parecido al observado para casi todos los partidos. Esto es consistente con el principio de representación proporcional y el hecho de que la reforma no modifica sustancialmente la votación total de los partidos. Sin embargo, el intervalo de confianza del PRI es considerablemente más amplio que el de los demás partidos. Esto sucede debido a la cláusula de sobrerrepresentación: en 2009, al PRI le fueron deducidos 27 diputados para que su porcentaje de curules en la Cámara no excediera en más de 8 puntos su porcentaje de votación nacional efectiva. De la misma manera, en 99.4 % de las elecciones simuladas, el PRI se ubicó dentro de los supuestos de la cláusula de sobrerrepresentación. Por lo tanto, la estimación del número de curules plurinominales del PRI reproduce la incertidumbre del resultado en los distritos de mayoría relativa: cuando obtiene menos curules uninominales,

recibe más de representación proporcional, y viceversa. Por eso, su número *total* de curules es muy parecido al observado, y sus intervalos de confianza son relativamente estrechos.



**Figura 5:** Probabilidad de que cada partido político obtenga más de cierto número de curules con la reforma que homologa el calendario de las elecciones federales y locales. Las líneas verticales indican el número de curules observado en la elección federal de 2009.

Otra manera de presentar los resultados del modelo es a través de pruebas de hipótesis. En lugar de utilizar estimaciones puntuales e intervalos de confianza, las gráficas de la Figura 5 nos permiten transmitir nuestros resultados en términos probabilísticos. Cada gráfica indica la probabilidad de que el número de curules de un partido sea mayor o igual a cierta cifra o

proporción de interés. La construcción de estas gráficas es muy sencilla; por ejemplo, para averiguar la probabilidad de que un partido tenga al menos 100 curules con la reforma, basta con registrar el número de simulaciones en que ese partido tiene más de 100 curules, y luego dividirlo entre el total de las simulaciones.

Cabe mencionar que la distribución del número de curules también se puede transformar fácilmente en otras distribuciones de interés. Por ejemplo, podemos aproximar la distribución del índice de [Banzhaf \(1964\)](#), o del número efectivo de partidos, calculándolos para cada simulación de la conformación hipotética de la Cámara.

## 5.2. Disminución del número de curules plurinominales

En esta sección ilustramos una aplicación de nuestro método estadístico para evaluar una reforma que disminuye el número de curules plurinominales en la Cámara de Diputados; es decir, esta reforma modifica el método de asignación de curules. Lo anterior significa que el modelo de regresión sirve únicamente para separar a los factores sistemáticos (preferencias de los votantes) de los aleatorios (eventos fortuitos, como errores humanos en el conteo de los votos) ([Gelman & King, 1994](#); [King et al., 1994](#)). También significa que el criterio principal para seleccionar a las variables explicativas es que ayuden a predecir lo mejor posible las variables dependientes ([Gelman & Hill, 2007](#)).

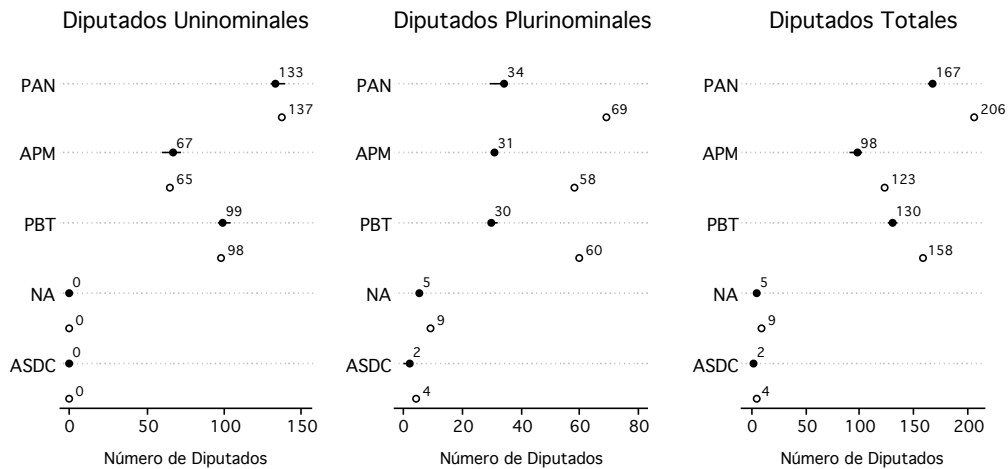
Los datos para el análisis provienen de los cómputos distritales de la elección de Diputados Federales del 2006, agregados por distrito electoral federal ( $n = 300$ ) ([IFE, 2006](#)). Las variables dependientes son los *log ratios* de la proporción de votos del PAN, la Alianza por México (APM = PRI + PVEM), la Coalición por el Bien de Todos (PBT = PRD + PT + Convergencia) y Nueva Alianza (NA), con respecto a la proporción de votos del partido Alternativa (ASDC). Las variables explicativas son los *log ratios* de la proporción de votos que esos mismos partidos obtuvieron en la elección presidencial de 2006, y dos variables dicotómicas que indican si el PAN o el PRD obtuvieron más votos que el resto de los partidos en la elección para Diputados Federales del 2003.<sup>17</sup> Puesto que el objetivo de la regresión es predecir el resultado electoral, al simular valores de la distribución  $\tilde{V}_{ij}^{hip}$  fijamos los valores de las

---

<sup>17</sup>Los distritos ganados por el PRI en 2003 son la categoría de referencia. Los datos consideran el proceso de redistribución que ocurrió entre 2003 y 2006.

variables explicativas en sus valores reales,  $X_i^{(hip)} = X_i$ .

¿Qué efecto tiene esta reforma en la composición de la Cámara de Diputados? La Figura 6 muestra la asignación de curules que ocurrió en 2006 y la compara con la distribución del número de curules en un escenario contrafactual con sólo 100 asientos de representación proporcional. Los puntos indican la media de las distribuciones de probabilidad, las líneas horizontales denotan los intervalos de confianza al 95 %, y los círculos el número de curules observado en 2006. Como se puede apreciar, esta reforma sólo afecta a las diputaciones plurinominales, por lo que las diferencias en el número de curules uninominales entre el resultado observado y nuestro escenario hipotético son muy pequeñas, y teóricamente se deben a eventos aleatorios. En cambio, todos los partidos muestran una reducción en el número de curules plurinominales a consecuencia de la reducción del tamaño del congreso.



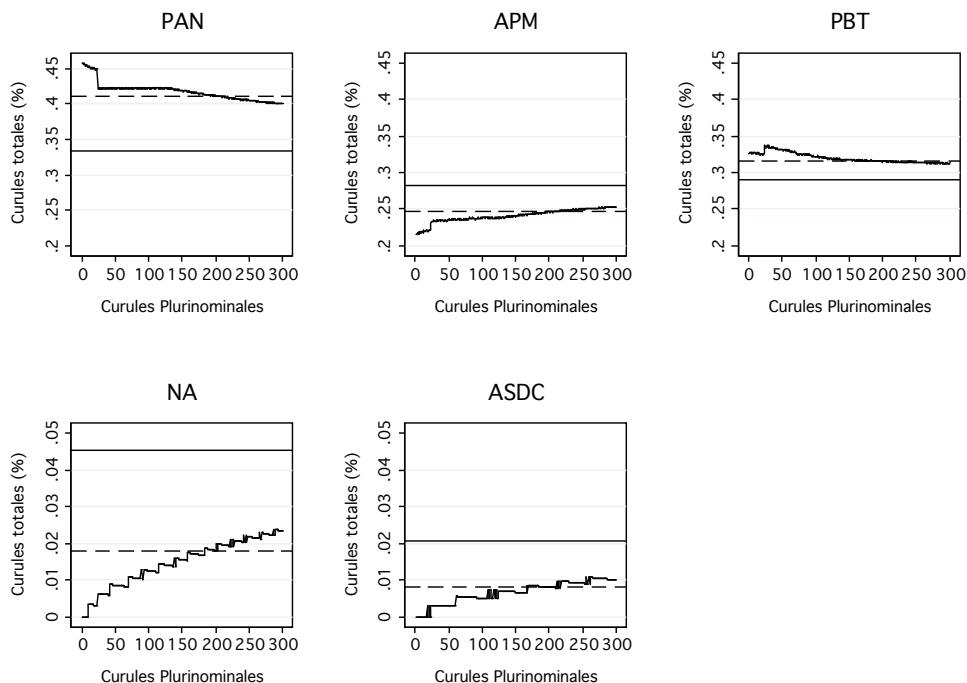
**Figura 6:** Número de Diputados por partido político. Los círculos indican el número de curules observado en la elección federal de 2006. Los puntos representan la media de la distribución de probabilidad del número de curules con la reforma que disminuye en 100 el número de diputados plurinominales. Las líneas horizontales simbolizan intervalos de confianza al 95 %.

Como se puede observar, la reforma que reduce el número de curules plurinominales en 50 % se traduce en una disminución casi proporcional de

diputaciones plurinominales para todos los partidos. Esto significa que en términos porcentuales, prácticamente no existen diferencias entre la composición observada de la Cámara y el escenario hipotético. La Tabla ?? muestra el porcentaje de curules que los partidos recibieron en la elección de 2006, y lo compara tanto con el porcentaje de votos obtenidos como con el porcentaje de curules que hubieran obtenido con la reforma. La diferencia entre votos y curules ilustra el sesgo mayoritario del sistema electoral vigente. Sin embargo, los porcentajes de curules con y sin reforma son muy parecidos: en ningún caso la diferencia entre el escenario observado y el hipotético es mayor a un punto porcentual.

Este ejercicio puede generalizarse para diferentes tamaños de la Cámara de Diputados. La Figura 7 ilustra el tamaño relativo de las bancadas como una función del número de curules plurinominales en un rango de 0 a 300. Las simulaciones utilizan los resultados electorales de 2006 y asumen que los resultados de los 300 distritos de mayoría relativa permanecen sin cambio. Es decir, estamos simulando una Cámara mixta que va desde 300 a 600 asientos con una composición cada vez más proporcional. Como se aprecia, incluso en este rango considerablemente amplio, el porcentaje de curules totales no varía más de 4 puntos porcentuales con respecto al porcentaje observado en realidad (líneas punteadas). Casar (2009) llega a las mismas conclusiones cuando analiza una reforma similar con los resultados de las elecciones federales de 1997 a 2009. La razón de que estas variaciones sean pequeñas es que 100 curules de representación proporcional es un número suficientemente grande como para mantener resultados razonablemente proporcionales (Rae, 1967; Taagepera & Shugart, 1989). De hecho, el número de curules plurinominales en Israel—un país al que se recurre con frecuencia para ilustrar los sistemas electorales de representación proporcional “pura”—es de 120.





**Figura 7:** Porcentaje de curules por bancada en la Cámara de Diputados en función del número de curules plurinominales, manteniendo constantes los resultados de la elección federal de 2006. Las líneas punteadas indican el porcentaje de curules totales observado (con 200 curules plurinominales). La línea sólida indica el porcentaje de votos obtenido por cada partido o coalición.

Conforme la Cámara se hace más proporcional, la bancada del PAN y de la Coalición Por el Bien de Todos (PBT) –los partidos punteros en 2006– disminuye en términos relativos. Por otro lado, las bancadas de la Alianza por México (PRI+PVEM), NA y ASDC, aumentan relativamente conforme crece el número de plurinominales–tal y como es de esperarse en una Cámara menos mayoritaria. Vale la pena destacar los picos que se observan en el rango de 0 a 25 curules plurinominales. Con cero curules de representación proporcional, el PAN hubiera conseguido 45.3% de la Cámara en 2006 (pues ganó en 135 de 300 distritos uninominales). Sin embargo, al tener 24 curules plurinominales, el PAN hubiera alcanzado el tope de sobrerrepresentación (8% entre votos y curules totales), razón por la cual su proporción de curules hubiera disminuido

bruscamente y se mantendría constante hasta que la Cámara tuviera más de 124 curules de representación proporcional. Al haber suficientes curules plurinominales, el tope de sobrerrepresentación deja de ser vinculante para el PAN y su porcentaje de curules se aproximaría paulatinamente a su porcentaje del voto nacional (33.41 %). Una lógica similar explica la pendiente negativa de la proporción de curules del PBT, segunda fuerza en la elección de 2006. Por otro lado, conforme la Cámara se hace más proporcional, la coalición APM consigue un mayor porcentaje de curules.

En diciembre de 2009, el Presidente de la República envió al Congreso una iniciativa de reforma que propone disminuir en 100 asientos la Cámara de Diputados. A diferencia de nuestro ejemplo, la propuesta busca reducir proporcionalmente las curules de mayoría (de 300 a 240) y de representación proporcional (de 200 a 160). Es probable que la motivación de mantener el ratio de curules uninominales y plurinominales sea mantener el equilibrio de fuerzas en la legislatura. Paradójicamente, una reducción en los asientos de mayoría, y no tanto los de representación proporcional, modificaría la correlación de fuerzas en la Cámara de Diputados. Esto se debe a que la reducción de los diputados uninominales implica necesariamente un proceso de redistribución, el cual puede modificar notablemente el número de curules de los partidos políticos, en particular de aquellos que tienen una base electoral regionalmente concentrada ([Gudgin & Taylor, 1979](#); [Grofman & King, 2007](#); [Taagepera & Shugart, 1989](#)).

## 6. Conclusiones

Hace casi veinte años, [King \(1989\)](#) observó que dada la creciente cantidad de datos disponibles para los investigadores en Ciencia Política, el desarrollo de nuevos métodos estadísticos para analizarlos tendría un impacto “desproporcionado” en la disciplina. Sin lugar a dudas, la literatura sobre el poder legislativo en México se encuentra en esa fase de expansión. Por una parte, existen excelentes fuentes de información sobre datos electorales (IFE, atlas electorales locales) e indicadores censales desagregados al nivel de las unidades electorales. Además, en los últimos meses se han presentado diversas iniciativas de reforma del sistema electoral cuyos efectos potenciales constituyen, por sí mismas, una amplia agenda de investigación tanto teórica como empírica. Finalmente, la metodología política ha tenido enormes avances en la generación de técnicas y herramientas para analizar los datos electorales

de países como México. El modelo Monte Carlo para la Cámara de Diputados que presentamos en este artículo pretende incorporar algunos de estos avances para coadyuvar al desarrollo de los estudios sobre el sistema político mexicano en general y del Congreso en particular.

Entre las propuestas de reforma concretas que nuestro modelo permite analizar se encuentran: cambiar el número de asientos del Congreso, modificar la proporción de asientos uninominales y plurinominales, cambiar tanto el umbral de representación como el tope de sobrerrepresentación. Por otro lado, el modelo es lo suficientemente general y flexible como para adaptarse fácilmente al estudio de otras instituciones legislativas, como el Senado. Además, debido a que las legislaturas estatales tienen un sistema electoral semejante al de la Cámara de Diputados, nuestro modelo también puede utilizarse para analizar el efecto de diversas reformas electorales a nivel local.

El marco en que se desarrolla el modelo también puede adaptarse para analizar otras consecuencias de los resultados electorales, más allá de la conformación propia del Congreso, como pueden ser la identificación de bancadas pivotales, o la asignación del financiamiento público y el acceso a medios para los partidos políticos. Esto, toda vez que desde un punto de vista estadístico, no existen grandes diferencias entre analizar cantidades de interés tales como el número de curules, el tamaño relativo de las bancadas, la probabilidad de ganar o perder un distrito, o incluso estimar las prerrogativas que corresponderían a cada partido de acuerdo los resultado electorales.

En el ámbito metodológico, el modelo estadístico también puede complementarse con otros métodos de estimación de los resultados electorales distintos al desarrollado en este artículo. Por ejemplo, las simulaciones pueden ser obtenidas con métodos bayesianos o de cadenas markovianas (MCMC), adaptarse para bases de datos longitudinal (TSCS, por sus siglas en inglés), o bien extenderse para la especificación de modelos jerárquicos que combinen datos con diferentes niveles de agregación.

## Referencias

- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 139–177).
- Banzhaf, J. (1964). Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.*, 19, 317.
- Brambor, T., Clark, W., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1), 63.
- Breusch, T., Robertson, J., & Welsh, A. (1997). The emperor's new clothes: a critique of the multivariate t regression model. *Statistica Neerlandica*, 51(3), 269–286.
- Casar, M. (2009). Reformas en el aire. *Nexos*, (12).
- Cox, N. (2002). Speaking stata: How to face lists with fortitude. *The Stata Journal*, 2, 202–222.
- Diaz-Cayeros, A. (2005). Endogenous institutional change in the mexican senate. *Comparative Political Studies*, 38(10), 1196.
- Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press Cambridge, UK:.
- Gelman, A. & King, G. (1990). Estimating the electoral consequences of legislative redistricting. *Journal of the American Statistical Association*, 85(410), 274–282.
- Gelman, A. & King, G. (1994). A unified method of evaluating electoral systems and redistricting plans. *American Journal of Political Science*, 38(2), 514–554.
- Greene William, H. (2000). Econometric analysis. *Handbook of Econometrics*, 3.
- Grofman, B. & King, G. (2007). The future of partisan symmetry as a judicial test for partisan gerrymandering after LULAC v. Perry. *Election Law Journal*, 6(1), 2–35.

- Gudgin, G. & Taylor, P. (1979). *Seats, votes, and the spatial organisation of elections*. Routledge Kegan & Paul.
- Herron, M. (1999). Postestimation uncertainty in limited dependent variable models. *Political Analysis*, 8(1), 83.
- Honaker, J., Katz, J., & King, G. (2002). A fast, easy, and efficient estimator for multiparty electoral data. *Political Analysis*, 10(1), 84.
- IFE (2006). *Computos Distritales de la Eleccion de Diputados Federales*. Instituto Federal Electoral.
- IFE (2009). *Computos Distritales de la Eleccion de Diputados Federales*. Instituto Federal Electoral.
- Jackman, S. (2000a). Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation. *Political Analysis*, 8(4), 307.
- Jackman, S. (2000b). Estimation and inference via Bayesian simulation: An introduction to Markov chain Monte Carlo. *American Journal of Political Science*, 44(2), 375–404.
- Jackman, S. (2004). Bayesian analysis for political research. *Annual Review of Political Science*, 7, 483–505.
- Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Wiley.
- Jackson, J. (2002). A seemingly unrelated regression model for analyzing multiparty elections. *Political Analysis*, 10(1), 49.
- Jones, M. P. (1997). Federalism and the number of parties in argentine congressional elections. *The Journal of Politics*, 59(02), 538–549.
- Katz, J. & King, G. (1999). A statistical model for multiparty electoral data. *The American Political Science Review*, 93(1), 15–32.
- King, G. (1989). *Unifying political methodology*. Cambridge University Press.
- King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton Univ Pr.

- King, G., Keohane, R., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton University Press Princeton, NJ.
- King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2), 347–361.
- Magar, E. (2006). Gubernatorial coattails and Mexican congressional elections since 1979. In M. S. Shugart & J. Weldon (Eds.), *What Kind of Democracy has Mexico? The Evolution of Presidentialism and Federalism (por publicarse)*. Stanford University Press.
- Martin, A. (2008). Bayesian Analysis. In J. Box-Steffensmeier & D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford University Press.
- Mondak, J. & McCurley, C. (1994). Cognitive efficiency and the congressional vote: The psychology of coattail voting. *Political Research Quarterly*, 47(1), 151.
- Rae, D. (1967). *The political consequences of electoral laws*. Yale University Press.
- Rosenbaum, P. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, 147(5), 656–666.
- Samuels, D. J. (2000). The gubernatorial coattails effect: Federalism and congressional elections in brazil. *The Journal of Politics*, 62(01), 240–253.
- StataCorp (2009a). *Stata Programming Reference Manual, Release 11*. College Station, TX: STATA Press.
- StataCorp (2009b). *Stata: Release 11. Statistical software*. College Station, TX: STATA Press.
- Taagepera, R. & Shugart, M. (1989). *Seats and votes: The effects and determinants of electoral systems*. Yale University Press.
- Tomz, M., Tucker, J., & Wittenberg, J. (2002). An easy and accurate regression model for multiparty electoral data. *Political Analysis*, 10(1), 66.

- Tomz, M., Wittenberg, J., & King, G. (2003). Clarify: Software for interpreting and presenting statistical results. *Journal of Statistical Software*, 8(1), 1–30.
- Weldon, J. (2001). The Consequences of Mexico's Mixed-Member Electoral System, 1988-1997. In M. Shugart & M. Wattenberg (Eds.), *Mixed-Member Electoral Systems: The Best of Both Worlds?* Oxford University Press.
- Western, B. & Jackman, S. (1994). Bayesian inference for comparative research. *The American Political Science Review*, 88(2), 412–423.