



Perfiles Latinoamericanos

ISSN: 0188-7653

perfiles@flacso.edu.mx

Facultad Latinoamericana de Ciencias

Sociales

México

Castañeda, Gonzalo; Ibarra, Ignacio

Detección de fraude con modelos basados en agentes: las elecciones mexicanas de 2006

Perfiles Latinoamericanos, núm. 36, julio-diciembre, 2010, pp. 43-69

Facultad Latinoamericana de Ciencias Sociales

Distrito Federal, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=11515386002>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

 redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Detección de fraude con modelos basados en agentes: las elecciones mexicanas de 2006

GONZALO CASTAÑEDA E IGNACIO IBARRA*

Perfiles Latinoamericanos 36
Julio-diciembre 2010

43

Resumen

En este artículo se desarrolla una técnica novedosa para la detección de fraudes electorales a partir del análisis estadístico de los resultados. Con este fin se describe un modelo computacional basado en agentes en el que la dinámica de una campaña electoral condiciona las preferencias partidistas de los individuos. El modelo se construye suponiendo que en un proceso limpio los resultados oficiales dependen, esencialmente, de preferencias partidistas que han sido sujetas a la interacción social y a la influencia de información global. En contraste, para simular la existencia de un fraude se manipula la voluntad de un conjunto de ciudadanos el día de las elecciones. El modelo se calibra con datos preelectorales de la contienda del año 2006 para la presidencia de México, y la presencia del fraude o su ausencia se valida mediante la prueba no-paramétrica de Kolmogorov-Smirnov. Con la información utilizada, el modelo de simulación descarta la presencia de un fraude de gran magnitud en el que no menos del 5-6% del total de votos emitidos son manipulados.

Abstract

In this paper an innovative tool is applied for the detection of electoral fraud by analyzing statistical distributions of the official results. With this aim a computational model is developed to describe the dynamics of an electoral campaign that conditions the individuals' political preferences. The agent-based model is built assuming that in a clean process the electoral outcome depends, essentially, on preferences subjected to the influence of social interaction and global information. In contrast, in order to simulate an electoral fraud the model is modified such that the voting behavior of the citizens is manipulated on election-day. The model is calibrated with data from the 2006 Mexican electoral campaign for the presidency, and the presence or absence of fraud is validated using the Kolmogorov-Smirnov test, a non-parametric procedure. With information aggregated at the electoral district level, the simulation model rejects the existence of a large scale fraud where at least 5-6% of the total vote tally is manipulated in favor of a particular candidate.

Palabras clave: elecciones en México, modelo basado en agentes computacionales, fraude electoral, simulación política.

Key words: Mexican elections, agent-based model, electoral fraud, political simulations.

* Gonzalo Castañeda es doctor en Economía por la Cornell University de Nueva York, actualmente se desempeña como profesor-investigador en El Colegio de México. Ignacio Ibarra es candidato al doctorado en Economía por la Universidad de las Américas-Puebla y actualmente se desempeña como profesor en la Universidad Popular Autónoma del Estado de Puebla.

Introducción

Felipe Calderón resultó ganador en el conteo original de las elecciones del año 2006 para la presidencia de México con un porcentaje de votación que rebasó en 0.58% el obtenido por Andrés Manuel López Obrador (AMLO).¹ Resultado que difícilmente se hubiera podido anticipar a finales de marzo cuando algunos sondeos indicaban una amplia ventaja de diez puntos de AMLO sobre su más cercano perseguidor en una carrera de cinco partidos.² El estrecho margen a favor de Calderón y la presencia de diversos hechos considerados como irregulares hicieron que AMLO y su equipo de campaña impugnarán los resultados anunciados por el Instituto Federal Electoral (IFE).

En los días posteriores a la elección se expresaron todo tipo de acusaciones: un comportamiento anómalo en la dinámica del PREP (Programa de Resultados Electorales Preliminares: conteo en tiempo real de los resultados registrados en las actas de escrutinio), un sinnúmero de inconsistencias en el llenado de las actas, coerción en el voto de burócratas y maestros sindicalizados, inducción del voto a través de programas sociales, la promoción constante de Calderón en los discursos públicos del presidente Fox, y campañas de radio y TV auspiciadas por cámaras empresariales con la intención de ensuciar la imagen de AMLO. Por todas estas razones, y otras más, el Partido de la Revolución Democrática (PRD) y los demás integrantes de su coalición exigieron públicamente un nuevo conteo de todos los votos a nivel de la casilla. Esta petición y la movilización social que paralizó por varios meses la avenida Paseo de la Reforma, una de las más emblemáticas de la ciudad de México, fueron concebidas con la intención de generar presión política a favor de la anulación del proceso electoral.³

Aunque es difícil discernir en un análisis riguroso la influencia que pudieron haber ejercido los discursos de Fox y las campañas negativas de los grupos de cabildo sobre

¹ Felipe Calderón era el candidato del Partido Acción Nacional (PAN), mientras que AMLO era candidato de la Coalición por el Bien de Todos (una alianza de partidos de izquierda: PRD, PT y Convergencia).

² Además del PAN y la Coalición por el Bien de Todos, los otros partidos registrados fueron los siguientes: Alianza por México, una coalición entre el Partido de la Revolución Institucional (PRI) y el Partido Ecologista Verde de México y cuyo candidato era Roberto Madrazo; el Partido Nueva Alianza (Panal) que tenía como candidato a Roberto Campa y el Partido Alternativa Socialdemócrata y Campesina (PASC) que postuló a Patricia Mercado.

³ En los procedimientos legales, el PRD refutó exclusivamente los resultados del 35% de las actas (Ugalde, 2008). Después de que se abrieron 9% de las urnas (11 724) y se descartaron los votos de 743 casillas, el Tribunal Federal Electoral (TRIFE) consideró que la mayoría de las impugnaciones eran infundadas y sólo aceptó la premisa de que Fox arriesgó el proceso electoral al establecer un ambiente de clara confrontación; no obstante, el tribunal concluyó que la intervención del presidente no fue crítica en los resultados electorales. Por lo tanto, la victoria oficial de Calderón fue declarada el 5 de septiembre de 2006 con una diferencia de 0.56 puntos por encima del porcentaje de votos recibidos por AMLO.

los resultados electorales, existen diversas metodologías forenses para detectar fraude al momento de ejercer el voto y en el proceso de escrutinio. Estas pruebas están basadas en el análisis estadístico de los resultados numéricos, ya sea a través de la detección de patrones anómalos en la distribución de los votos o de los errores registrados, validando un comportamiento dinámico irregular del conteo de actas, o bien estableciendo asociaciones con variables que no deberían incidir en la decisión del voto en una contienda limpia.

En este artículo se desarrolla una metodología novedosa para la detección de fraudes a partir de un Modelo Basado en Agentes (ABM, por sus siglas en inglés). Se dice que estos modelos computacionales explican un fenómeno socioeconómico cuando logran “crecer” patrones emergentes (o regularidades estadísticas) de variables agregadas a partir de la interacción de agentes en un sistema. Este ABM, en particular, tiene como objetivo replicar la distribución de la participación del voto de los tres partidos principales en los distritos electorales. La prueba forense consiste en comparar los datos reales de la elección con los datos artificiales generados mediante dos tipos de premisas: un proceso electoral limpio y otro con fraude en el que los votos se manipulan el día de la elección. De esta manera, la hipótesis de fraude electoral se rechaza cuando la distribución simulada con elecciones limpias es estadísticamente más cercana a la distribución real que aquélla que resulta del modelo de elecciones trucadas.

El ABM se construye bajo la presunción de que las preferencias partidistas de los individuos se modifican a lo largo de la campaña electoral por la interacción social y la influencia de la información global. Es decir, el cambio en la decisión de voto se debe, por un lado, a las redes de discusión política de los ciudadanos y, por otro, a la información global que se transmite en noticieros, *spots* publicitarios, debates y encuestas de opinión. Por ende, conviene aclarar que la prueba estadística que aquí se implementa se aplica a una hipótesis conjunta: la limpieza del proceso electoral (entendida exclusivamente como la no manipulación del voto el día de las elecciones) y la relevancia de las influencias locales y globales. Por lo tanto, cuando se rechaza el fraude electoral con los resultados de las simulaciones se está descartando que el fraude haya tenido lugar una vez que se dio un proceso de formación partidista a partir de influencias locales y globales. En otras palabras, la prueba implementada no contempla la posibilidad de que los datos oficiales presenten votos manipulados que modifiquen las decisiones de voto generadas a partir de otro proceso de formación de preferencias.

Si bien el modelo tiene el potencial para analizar fraudes electorales orquestados por cualquier partido a través de diferentes mecanismos, en este trabajo se contempla la posibilidad del fraude por parte del Partido Acción Nacional (PAN). La relevancia de analizar esta opción se debe a que su candidato resultó ser el ganador y a que los seguidores de AMLO nunca aceptaron la validez de los resultados por las razones

arriba mencionadas. Sin embargo, en el análisis de sensibilidad también se plantea la posibilidad de que haya habido errores aritméticos en el conteo de las actas (o fraudes por parte de todos los partidos). Dado que los parámetros se calibran con datos relativamente agregados, como son las encuestas nacionales de opinión y las preferencias preelectorales a nivel estatal, el modelo no tiene el refinamiento necesario para detectar fraudes menores al 5-6% de los votos emitidos; es decir, para niveles de manipulación menores a este porcentaje la prueba no puede discernir si los resultados de la simulación de un proceso limpio son estadísticamente diferentes a los de la simulación de un proceso trucado. De cualquier forma, la evidencia presentada rechaza de manera contundente la existencia de un fraude de grandes dimensiones, como en su momento lo sugirieron AMLO y su equipo.

Las ventajas principales de esta metodología forense son las siguientes: (i) establece con precisión estadística el porcentaje de votos que pueden ser manipulados para ocultar evidencias de fraude; (ii) valida o descarta diferentes mecanismos de manipulación utilizados en los distritos “capturados” por los operadores del fraude; (iii) al plantear un sistema complejo, en el que las variables coevolucionan, no supone relaciones lineales entre variables ni direcciones causales únicas, como se plantea en los modelos estadísticos tradicionales; (iv) detecta la posibilidad de fraude más allá de los errores encontrados en actas inconsistentes (e.g., coerción del voto); en consecuencia, la prueba forense mediante ABM produce evidencias que no pueden obtenerse, inclusive, cuando se lleva a cabo un recuento de votos, ya que éste sólo es capaz de detectar errores en las actas pero no votos inducidos en determinadas casillas o distritos electorales.

El resto del artículo se estructura de la siguiente forma. En la segunda sección se revisa brevemente la literatura que usa pruebas estadísticas tradicionales para detectar la existencia de fraude en el marco de las elecciones de 2006. En la tercera se describe el modelo basado en agentes para explicar la dinámica de una campaña electoral. En la cuarta se generan los resultados electorales simulados con una contienda limpia y un proceso fraudulento, posteriormente, las distribuciones reales y simuladas se comparan con una prueba no paramétrica. En las conclusiones se enfatiza la importancia de elaborar un modelo computacional más detallado para incrementar la potencia de esta prueba forense de resultados electorales.

La evidencia del fraude

La noche de las elecciones, el presidente del IFE, Luis Carlos Ugalde, decidió no declarar un ganador debido al reducido margen de votos entre AMLO y Calderón según los resultados del PREP y el conteo rápido. Esta decisión, el excluir las actas inconsis-

tentes en los cálculos del PREP y otros errores de comunicación propiciaron acusaciones de fraude por parte del candidato de la Coalición. El desconcierto del perredismo era, indudablemente, consecuencia de una derrota no anticipada, después de muchos meses de liderar las encuestas y de una larga historia de elecciones sucias en el sistema político mexicano. No obstante, y como lo plantean Eisenstadt y Poiré (2006), la sugerencia de un fraude electoral de grandes dimensiones nunca fue avalada con evidencia sólida. Ni el IFE ni el Tribunal Federal Electoral (TRIFE) recibieron pruebas bien documentadas de la mayoría de los alegatos. Parecía que, para el grupo de AMLO, las movilizaciones públicas eran evidencia suficiente de la existencia de fraude.

Los reclamos a favor del fraude continuaron una vez que el conteo distrital confirmó el triunfo de Calderón. A la manera de una “teoría de la conspiración” se invocaron todo tipo de maniobras ilícitas por parte de la maquinaria calderonista que, en colusión con el IFE y el gobierno, buscaban impedir su derrota a toda costa. En esta etapa, el enojo fue sustituido por el pensamiento estratégico, por lo que la lógica de acusaciones a partir de evidencia frágil se explica como una forma de desacreditar públicamente el proceso electoral. La incapacidad de la Coalición para aportar pruebas duras de un fraude masivo hacía necesario alentar a un segmento amplio de la población que exigiera el recuento total de los votos para, de esta forma, promover la anulación de las elecciones ante los tribunales.

Tres argumentos principales fueron esgrimidos por los partidarios de AMLO como evidencia del fraude: (i) un escenario improbable en el conteo del PREP en tiempo-real, ya que los votos de Calderón y AMLO nunca se cruzaron una vez que se estableció una brecha a favor del primero, a pesar de que la diferencia de votos entre estos candidatos era mínima;⁴ (ii) un escenario improbable en el conteo oficial de los votos registrados en las actas de los 300 distritos, en donde la brecha inicial a favor de AMLO disminuyó paulatinamente hasta ser revertida en las horas finales; (iii) una gran cantidad de inconsistencias en las actas de escrutinio, algunas de las cuales no fueron incorporadas en el conteo del PREP.⁵

⁴ Una colección de estudios en los que se señalan las “anomalías” del PREP y el conteo oficial de votos se presenta en las páginas de Internet de Luis Mochan <<http://em.fis.unam.mx/~mochan/elecciones/>>, y de Octavio Miramontes <<http://www.fisica.unam.mx/octavio/>>. Los modelos utilizados por estos físicos suponen que el flujo de votos en tiempo-real de un proceso limpio se produce a partir de un conteo aleatorio de las actas provenientes de los 300 distritos dispersados a lo largo del territorio nacional. ¿Supuesto que, indudablemente, resulta ser muy alejado de la realidad!

⁵ Las actas inconsistentes pueden reflejar fraudes o simplemente presentan errores porque los funcionarios de caja se equivocan al contar las boletas, o porque los votantes no depositan la boleta en la urna adecuada. Asimismo, existen errores de reporte de inconsistencias en el llenado de los diferentes campos de referencia: boletas

Como bien dice Aparicio (2006a), estos alegatos no sólo son infundados sino también ilógicos. ¿Por qué molestarse en producir un fraude cibernético con el PREP si éstos no eran los resultados oficiales y el fraude se podía descubrir con el conteo distrital? ¿O acaso la maquinaria era tan eficiente que el PREP, el conteo rápido y el conteo distrital estaban manipulados de forma tal que se evitaban inconsistencias entre ellos? ¿Por qué producir un fraude masivo a través de la manipulación de las actas y dejar evidencia fehaciente de que un ilícito de proporciones mayúsculas había sido orquestado?

De hecho, el patrón descrito en (i) es muy factible cuando los votos registrados en las distintas actas no se contabilizan de manera aleatoria. Como se muestra en Pliego (2007), existe una fuerte correlación entre la naturaleza de los distritos (distancia en relación con las oficinas distritales del IFE, o índice de marginalidad del municipio correspondiente) y el tiempo que tardaban los votos en ser registrados en el conteo del PREP. Por lo tanto, después de un breve periodo de volatilidad, los votos de Calderón obtenidos en las regiones desarrolladas del país le dieron un amplio margen inicial, el cual se redujo considerablemente, pero sin ser eliminado, una vez que los distritos que favorecían a AMLO empezaron a ser contabilizados.⁶

Por otra parte, Aparicio (2006b) presenta evidencia que descarta el argumento (ii), al indicar que el conteo oficial de votos a nivel distrital no era aleatorio por las apelaciones partidistas de los resultados. En otras palabras, los distritos en los que Calderón tenía un gran margen fueron largamente cuestionados por los representantes de AMLO y, por ello, se registraron muy tarde en el conteo global, en contraposición con aquellos distritos en los que AMLO obtenía una clara victoria. En consecuencia, el registro en tiempo-real de los votos seguía una dinámica de interpellación, de aquí que los diferentes tiempos utilizados para el conteo de los distritos ganados por el PAN y el PRD explican por qué la ventaja inicial de AMLO se fue desvaneciendo lentamente hasta perderse.

Si bien es cierto que hubo un gran número de actas inconsistentes (11 184), cuyos votos (2 581 226) no fueron incluidos en el conteo del PREP, las actas con errores aritméticos se redujeron del 51.4% en el año 2000 al 46.7% en 2006, a pesar de que la primera de estas elecciones presidenciales había sido considerada como limpia por los analistas (Aparicio, 2006c); asimismo, el número de votos con errores en las casillas fue bastante similar en promedio (1.26% en 2000 y 1.35% en 2006). Aparicio

recibidas, depositadas, sobrantes, número de votantes de acuerdo con el padrón electoral y el total de votos registrados.

⁶ A este respecto, Aparicio (2006b) argumenta que una división urbano-rural en las actas y la posición favorable de Calderón en los distritos urbanos ayuda a explicar la dinámica observada en el PREP; sin embargo, Romero (2006) apunta que este tipo de división no es una buena explicación para describir dicha dinámica.

(2006d) y Pliego (2007) muestran que los errores aritméticos son neutrales ya que se distribuyen uniformemente en las casillas del PAN y del PRD, por lo que no modifican los resultados oficiales. De hecho, existe un ligero sesgo a favor de AMLO aunque éste no es estadísticamente relevante. Al parecer estos errores se explican simplemente por equivocaciones humanas producidas en un sistema electoral en el que los funcionarios de casilla son voluntarios y tienen bajos niveles de escolaridad. Por otra parte, el problema de las actas “desaparecidas” del conteo del PREP se explica como una falla de comunicación de las autoridades del IFE. Meses antes de la elección los partidos habían aprobado la remoción de dichas actas de los resultados del PREP, y éstos eran conscientes de que dicha remoción estaba teniendo lugar (Ugalde, 2008).

Con el paso de las semanas, AMLO y sus colegas señalaron irregularidades adicionales para respaldar su teoría de la conspiración: (a) votos para AMLO anulados en el conteo de las actas; (b) la decisión del IFE de abrir sólo 2 864 urnas; (c) los vínculos de la líderesa del sindicato de maestros con el partido Nueva Alianza, quien “ordenó” votar por sus representantes en el Congreso (4.55% del voto total) y darle el voto a Calderón en vez de apoyar a su candidato a la presidencia (que obtuvo un magro 0.96% de la votación); (e) funcionarios ausentes que no fueron sustituidos por el primer individuo en la fila de la casilla, el cual debía tener una preferencia partidista aleatoria, sino por emisarios de Calderón dispuestos a manipular el voto; (f) un número inusual de votos a favor de Calderón en las casillas que tenían una alta tasa de participación.

Los analistas electorales emplean tradicionalmente dos tipos de herramientas estadísticas para investigar si las acusaciones de fraude están fundamentadas o no. En un primer enfoque, Aparicio (2006e, 2006f y 2006g) y Pliego (2007) dividen las actas en dos o tres subpoblaciones dependiendo del tipo de hipótesis que se estudia; por ejemplo, en actas en las que Calderón o AMLO ganaron, o en actas que corresponden a casillas en las que no había representantes del PAN o del PRD. De esta manera, las pruebas realizadas son las siguientes: (i) comparación de las medias aritméticas de los errores o de los votos asignados a un candidato entre las diferentes subpoblaciones, y (ii) remoción de casillas atípicas (por ejemplo, con alta participación o muchos errores aritméticos) con el objetivo de verificar si los resultados de las casillas restantes siguen siendo consistentes con los oficiales.

Una segunda alternativa es la empleada por Poiré y Estrada (2006), quienes utilizan análisis de regresión para explicar la diferencia en el número de votos asignados a Calderón y a AMLO en el nivel de casilla. Algunas de las variables independientes incluidas en estos modelos son indicadores de un fraude potencial de acuerdo a los alegatos arriba citados (ausencia de representantes, efecto sindicato de maestros) y otras más son variables de control que explican la decisión de voto bajo condiciones normales (efectos regionales, variables socio-demográficas). Las pruebas estadísticas realizadas por

tales autores no muestran evidencia alguna de fraude. Si bien todos estos estudios no ofrecen certeza plena de que no hubo fraude en 2006, sí se puede afirmar que la evidencia estadística disponible no es favorable a las acusaciones de AMLO y sus seguidores.

Un tercer tipo de prueba electoral forense, de uso reciente, es la implementada por Mebane (2006), quien supone el cumplimiento de la ley de Benford en procesos electorales limpios.⁷ Con 95% de confianza, este autor encuentra que dicha ley se rechaza en los datos para todo el país y para muchos de los estados, de aquí que existe la presunción de fraude electoral. Sin embargo, Castañeda (2009) muestra, mediante un modelo basado en agentes, que esta “ley” no ofrece una prueba robusta para distinguir entre elecciones limpias y las que han sido manipuladas.

Un modelo computacional para procesos electorales

A partir de la evidencia empírica que indica que las redes de discusión política son muy importantes para explicar la formación de opiniones, Castañeda e Ibarra (2009) desarrollan un ABM para las elecciones presidenciales de 2006. En este modelo computacional se establece un mecanismo de contagio social en el que cada agente está sujeto a la influencia de información local y global. Asimismo, el modelo permite combinar diversos elementos institucionales de los procesos electorales, como debates, campañas negativas y sesgo inducidos por la TV u otros medios masivos de comunicación, con los incentivos de los votantes, a través del voto estratégico y el costo-beneficio de la decisión de votar.

Una campaña electoral es considerada un sistema adaptable complejo ya que las posiciones de individuos y partidos afectan las encuestas de opinión y éstas, a su vez, inciden en la formación de opiniones y estrategias partidistas. Estos sistemas son analizados a través de Modelos Basados en Agentes en los que es posible incorporar agentes heterogéneos y mecanismos de decisión condicionados por la interacción local y el contexto social.⁸ A diferencia de los modelos matemáticos tradicionales, la simulación mediante modelos computacionales permite explicar comportamientos agregados que provienen de dinámicas no-lineales causadas por la retroalimentación entre agentes y entre éstos y el entorno de adaptación.

⁷ De acuerdo con esta ley, los dígitos iniciales de un conjunto de números siguen una distribución logarítmica cuando los datos no han sido perturbados.

⁸ Este tipo de modelos se han empleado antes para estudiar sistemas electorales, aunque por lo general se abocan a analizar procesos de adaptación de las estrategias partidistas. Ejemplos clásicos de estas aplicaciones son Kollman, Miller y Page (1992, 1998).

El ABM descrito en este artículo enfatiza el contagio social entre votantes potenciales por lo que se estructura a partir de un autómata celular en donde una retícula bidimensional caracteriza al espacio geográfico de interacción.⁹ La regla de transición de cada agente (o célula) hace que la variable de estado (intención de voto) varíe, esencialmente, en función de opiniones mayoritarias locales (red de discusión política) y globales (encuestas nacionales). De esta forma, la preferencia partidista de un porcentaje de agentes activados aleatoriamente en cada día de la campaña simulada puede cambiar por efecto del contagio social.

Las preferencias políticas iniciales de los ciudadanos y la mayoría de los parámetros del modelo son calibradas con datos agregados de encuestas electorales y con una base de datos panel que le da seguimiento a las preferencias políticas de los encuestados.¹⁰ En particular, el Estudio Panel México 2006, organizado por el Massachusetts Institute of Technology (MIT) y el periódico *Reforma*, ofrece tres oleadas de entrevistas (octubre 2005, abril/mayo 2006 y julio 2006) con 2 400 entrevistados y cerca de 100 preguntas relacionadas a la intención del voto, información sociodemográfica y opiniones de temas económicos y políticos.¹¹

La configuración del ABM

La descripción completa del modelo, bajo la premisa de un proceso electoral limpio, se presenta en el artículo antes referido.¹² Sin embargo, en este apartado se hace una breve exposición del mismo con el propósito de que el lector pueda entender cómo se

⁹ El ABM fue construido con la plataforma *NetLogo* versión 4.0.3 (Wilensky, 1999, <<http://ccl.northwestern.edu/netlogo/>>; Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.), el cual puede ser solicitado en la dirección sociomatica@hotmail.com.

¹⁰ Para la serie de preferencias agregadas se empleó la “encuesta sobre encuestas” calculada por el Centro de Investigación para el Desarrollo A.C. (CIDAC) que utiliza un conjunto de encuestas periódicas levantadas por diferentes compañías (Mitofsky y Gea-Isa) y periódicos nacionales (*Reforma*, *El Universal* y *Milenio*): <<http://www.cidac.org/es/modules.php?name=Content&pa=showpage&cid=98>>.

¹¹ Los cuestionarios para cada oleada, el conjunto de datos, la metodología y algunos estudios relacionados se encuentran disponibles en el portal de la 2006-MPS <<http://web.mit.edu/polisci/research/mexico06>>, proyecto dirigido por las siguientes personas (en orden alfabético): Andy Baker, Kathleen Bruhn, Roderic Camp, Wayne Cornelius, Jorge Domínguez, Kenneth Green, Joseph Klesner, Chappell Lawson (investigador principal), Beatriz Magaloni, James McCann, Alejandro Moreno, Alejandro Poiré, y David Shirk; con el apoyo financiero de la National Science Foundation (SES-0517971) y el periódico *Reforma*; el trabajo de campo fue llevado a cabo por el equipo de investigación y encuestas del periódico bajo la dirección de Alejandro Moreno.

¹² En Castañeda (2009) se explica en detalle la configuración del modelo, su calibración y validación empírica, mientras que el objetivo del presente trabajo es aplicar dicho modelo al estudio forense de los resultados electorales.

simula la formación de preferencias políticas a lo largo de una campaña electoral. El ABM en cuestión está diseñado con módulos específicos que se activan en diferentes períodos de tiempo en atención a la cronología real de la campaña por la presidencia de México. Por ende, el modelo plantea un proceso que dura 240 días de campaña (correspondiente a los ocho meses que van de noviembre a junio); encuestas nacionales que se levantan cada 30 días y el día siguiente de los debates, éstos tienen una amplia cobertura y se efectúan los días 180 y 220; un sesgo-TV que opera de manera continua hasta el día 200; escándalos políticos que se presentan en 2% de los días de campaña; un voto estratégico que es activado diez días antes de la elección y un análisis costo-beneficio sobre la decisión de votar el mismo día de la elección.¹³

El espacio de interacción social se representa a través de una retícula 120 x 120 con fronteras y vecindades tipo Moore (i.e., un agente tiene a lo más ocho vecinos con los que puede interactuar), lo que hace que el espacio en donde se simula el proceso electoral esté acotado al norte, sur, este y oeste. Por lo tanto, en este sistema electoral artificial, el número de votantes potenciales es de 14 400 individuos ubicados en un contexto geográfico y social. El espacio geográfico está dividido en 16 zonas electorales de tal forma que éstas corresponden a regiones del país que tienen, aproximadamente, el mismo número de votantes registrados en el padrón electoral real. Por otra parte, las zonas tienen una dimensión 30 x 30 y están conformadas por nueve distritos electorales de igual tamaño (10 x 10); en estos distritos los agentes ejercen su voto.

Mientras que las preferencias partidistas de los agentes se siembran aleatoriamente en cada zona electoral con datos del primer levantamiento de la encuesta panel a nivel estatal, los valores de otras variables se siembran al inicio de cada corrida a partir de criterios definidos a nivel nacional. Entre estas variables se encuentran las siguientes: compromiso ideológico (voto-duro, voto-débil e indecisos), exposición a la TV (o propensión a ser influenciado), atributos sociodemográficos (género, edad, escolaridad, ingreso y religión), confianza en la equidad del proceso electoral.¹⁴

(i) Módulos de contagio local y global

Durante los primeros 30 días, la información local es la única fuente de contagio en el modelo. Este contagio es posible cuando un agente que ha sido activado en la simulación establece comunicación con uno de sus interlocutores políticos. Los interlocutores son parte de la red de discusión política si pertenecen a la vecindad del agente y

¹³ En la simulación, un día de campaña corresponde a un tic o periodo de activación de la regla de transición de los agentes.

¹⁴ De estas variables, el compromiso ideológico se obtiene de datos de encuestas, la percepción de equidad se calibra mediante un método indirecto y las demás son definidas por el observador.

superan un umbral de similitud, es decir, comparten al menos un cierto número de atributos sociodemográficos. Una vez que la comunicación se produce, el interlocutor elegido al azar logra incidir en la opinión del agente activado cuando la opinión del primero coincide con la expresión mayoritaria de la vecindad.

Al hacerse pública la primera encuesta de opinión, el contagio se produce cuando el interlocutor, elegido al azar entre los miembros de la red del agente activado, tiene una preferencia política que coincide con la opinión mayoritaria reflejada en la encuesta nacional. Estas encuestas se levantan con el 10% de los agentes de cada una de las zonas electorales, por lo que se puede afirmar que las encuestas son representativas del sentir nacional. El observador define al inicio de las corridas el porcentaje de individuos que pueden ser activados diariamente,¹⁵ aunque la influencia solamente es viable en la medida en que el agente activado no sea un votante duro.

(ii) Módulo de campañas negativas

En 2% de los días de campaña definidos al azar, el partido que está en el segundo lugar de la contienda crea un escándalo político en contra del candidato que tiene el liderato según las encuestas. Si el observador especifica que el partido atacante sigue una estrategia agresiva, el escándalo puede ser de gran impacto (o de amplio espectro) y por ello el atacante puede hacerse del 6% de los adherentes no-duros del líder con una probabilidad del 80%; no obstante, también existe el riesgo de alienar al mismo porcentaje de sus adherentes cuando éstos desacreditan el uso de dichas tácticas. En contraste, cuando la estrategia del partido atacante se define como moderada, la posibilidad de ganar/perder adeptos disminuye a sólo 2% de los votantes una vez que se produce el escándalo.

(iii) Módulo de sesgo-TV

La imagen de un candidato puede ser promovida de manera sostenida a través de los medios electrónicos por parte de figuras públicas, noticieros, grupos de cabildeo y gerentes de campaña. Por esta razón, el ABM incluye un módulo que especifica que los partidos tienen la posibilidad de atraer a nuevos adeptos cada día de la campaña hasta 40 días antes de la elección cuando se suspende este tipo de publicidad. Este periodo coincide con la fecha establecida en el “acuerdo de neutralidad” firmado al inicio de la campaña por todos los partidos. El acuerdo también prohibía la transmisión de *spots* promoviendo programas sociales y obras públicas.

¹⁵ Este porcentaje y otros que se describen a continuación son calibrados indirectamente.

El modelo supone que el sesgo-TV favorece al PAN debido al apoyo indiscriminado que Fox le dio a Calderón en sus discursos públicos y a las acusaciones del PRD sobre la injerencia de las cámaras empresariales. El observador especifica el porcentaje de individuos que están sujetos a la influencia de los mensajes de los medios, propensión que se limita a los agentes que no son votantes duros y que tienen una exposición a la TV relativamente alta. A diferencia del módulo de interacción local, en donde el contagio se va dando al interior de las vecindades, el cambio de preferencias políticas a través de este módulo se produce de manera esparsa en el territorio nacional. De tal manera que el sesgo-TV permite introducir agentes con preferencias políticas antagónicas entre grupos de individuos geográfica y socialmente cercanos que mantienen cierta afinidad política.

(iv) Módulo de debates

En cada uno de los debates existe un ganador que se determina probabilísticamente. En el primer debate el candidato victorioso se elige entre los dos contendientes principales, dado que AMLO no participó, y en el segundo entre los tres candidatos punteros de las encuestas de opinión. La victoria ofrece la posibilidad de conseguir el apoyo de ciudadanos indecisos y votantes débiles. El porcentaje de individuos que pueden modificar su opinión bajo estas circunstancias también se especifica antes de correr las simulaciones.

(v) Módulo de voto-estratégico

Cuando la primera opción política de un agente no coincide con alguno de los dos candidatos que van a la cabeza en las encuestas es posible que modifique, de último momento, la intención de voto a favor de su segunda opción. Esta última se define, en el modelo, con el candidato que tiene las preferencias mayoritarias en la vecindad del agente. Aunque el ABM también incorpora la posibilidad de que los adherentes de Calderón vayan con Madrazo, y viceversa, en un escenario en que AMLO es considerado un “riesgo para el país”; en tanto que los votantes indecisos que son activados en este módulo seleccionan al azar entre Calderón y Madrazo. De nueva cuenta el porcentaje de votantes tácticos lo define el observador al iniciar la simulación.

(vi) Módulo de participación

A través de un análisis costo-beneficio los ciudadanos deciden participar o abstenerse de votar el día de la elección. Mientras que el costo de votar es un parámetro definido por el observador, el beneficio se determina usando una probabilidad heurística que mide la posibilidad de ganar que tiene el candidato preferido por el agente. Esta probabilidad se estima al multiplicar la proporción de las intenciones de voto recibidas

por su candidato en las encuestas nacionales, por el porcentaje de individuos que lo apoyan en la vecindad, por un índice binario que indica si el agente confía o no en la elección; obviamente, los votantes duros siempre ejercen su voto.

(vii) Módulo de calibración

Algunos de los parámetros que son definidos por el observador discrecionalmente pueden también ser calibrados indirectamente. Mediante un procedimiento algorítmico de optimización no-lineal, conocido como escalamiento (*hill-climbing*), se estiman los parámetros cuyos valores no se pueden obtener directamente de las encuestas. Con este propósito se especifica una función de adaptación que mide el grado de ajuste de los datos simulados a los datos reales. Esta función se define con una suma de errores cuadráticos promedio, siendo los errores calculados con la diferencia relativa entre las participaciones de la intención de voto de cada partido según las encuestas reales y el número equivalente estimado con los datos artificiales; esta diferencia se calcula para todas las encuestas levantadas a lo largo de la campaña.

Un ABM con fraude electoral

El modelo del apartado previo representa un proceso electoral limpio dado que la decisión individual de voto es respetada; es decir, la votación por un candidato en específico no es inducida con amenazas o premios otorgados por autoridades o partidos, ni el conteo final de votos es alterado por funcionarios electorales. En contraste, en un ABM que simula un proceso fraudulento se supone que las decisiones de los agentes pueden ser manipuladas el día de la elección, ya sea antes de que el voto se deposite o en el proceso de escrutinio. En particular, el módulo de fraude plantea que existen varios distritos electorales “controlados por la maquinaria panista”. En estos distritos, los operadores de Calderón pueden modificar a discreción un cierto porcentaje de votos a través de diversos mecanismos, siendo dicho porcentaje determinado por el observador al inicio de la corrida. Por lo tanto, el conteo final anunciado por las autoridades electorales oculta el hecho de que algunos de los votos obtenidos por el PAN fueron inducidos con “regalos” o alterados por funcionarios distritales o federales.

El módulo de fraude presenta varios procedimientos para definir en qué zonas electorales el voto puede ser manipulado a favor de Calderón. (i) “Autoridades”: cuando en la zona electoral existe al menos un gobernador panista al momento de la elección. (ii) “Enclaves”: cuando el PAN tiene una mayoría relativa en las preferencias políticas de la zona de acuerdo con los sondeos levantados al inicio de la campaña.

(iii) “Inconsistencias”: cuando, en los datos reales, los distritos electorales de la zona tienen un gran número de actas que presentan irregularidades. De acuerdo con estos criterios, la tabla A del “Apéndice” de este artículo muestra las zonas electorales que tienen el potencial de producir un escenario fraudulento.

Asimismo, el modelo supone que en los procedimientos de autoridades y enclaves la posibilidad de un fraude por parte de los operadores del PAN requiere superar dos filtros. En primer término, la zona electoral tiene que estar dominada por el PAN (ya sea por la presencia de gobernantes afiliados al partido o por la mayoría relativa de las preferencias panistas). En segundo término, las preferencias iniciales por Calderón tienen que superar un umbral de relevancia partidista. En otras palabras, para que una zona electoral pueda ser considerada como un reducto del PAN se requiere que el apoyo al partido no esté pulverizado.

En una zona de “control-panista” la fabricación de votos puede ser de dos tipos: “intercambio de votos” o “creación de votos”. En el primer método, un porcentaje f de los votos no-panistas (M) seleccionado al azar se modifican de tal forma que Calderón incrementa su cuenta en $f \times M$ votos en cada uno de los “distritos capturados”. En el segundo método, M corresponde al número de individuos registrados en el distrito que el día de la elección optan por no votar, de tal forma que las elevadas tasas de participación en los “distritos capturados” se deben a la fabricación de votos a favor de Calderón.

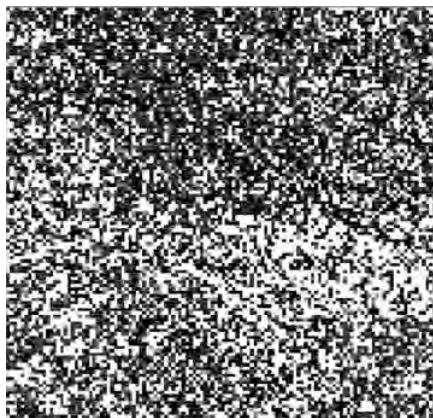
La detección del fraude a través de patrones emergentes

Para un grado de manipulación muy elevado es posible detectar visualmente el fraude ya que en la retícula de preferencias partidistas emergen súbitamente grandes áreas panistas el día del escrutinio; sin embargo, para poder identificar fraudes de menor magnitud es necesario realizar un análisis estadístico. Con este objetivo en mente, el programa calcula el histograma de alguna variable que resume el resultado electoral de los distintos distritos esparcidos por toda la geografía nacional, dicha distribución puede reflejar los votos manipulados que fueron incluidos en el conteo oficial. En este artículo, la variable bajo estudio es la participación de votos obtenidos por cada uno de los tres partidos importantes respecto del total de votos emitidos en cada distrito. El uso de esta variable en particular se debe a que Castañeda e Ibarra (2009) muestran que su distribución es un resultado estable del proceso electoral; es decir, no depende de las realizaciones específicas de los distintos elementos probabilísticos incluidos en el ABM de campañas electorales.

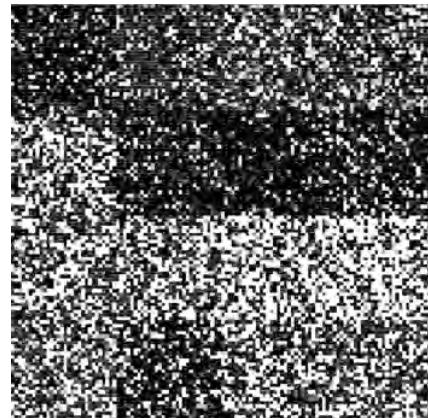
En la figura 1 se presenta la retícula con las 16 zonas electorales (4×4) para el caso de un fraude en el que se alteran 7.01% de las boletas. Cabe notar que la conformación espacial de las preferencias en un proceso trucado es muy diferente a la observada cuando las elecciones son limpias. Para este porcentaje de manipulación, el patrón emergente que se genera con el fraude es fácil de detectar mediante una simple inspección visual. En la figura 2 se ilustran los histogramas de las participaciones de votos para los escenarios limpio y fraudulento y, de nueva cuenta, las diferencias entre estas distribuciones empíricas se aprecian a simple vista. No obstante, para un análisis formal se requiere de un procedimiento estadístico con el que se pueda corroborar si, efectivamente, las dos muestras provienen de la misma distribución teórica. Un procedimiento de este tipo es la prueba no-paramétrica de Kolmogorov-Smirnov (KS), con la que es posible detectar la presencia o ausencia de fraude aun en casos en los que la evidencia visual no es concluyente.

Figura 1
Distribuciones espaciales de las preferencias políticas
según los conteos oficiales

(a) En elecciones limpias



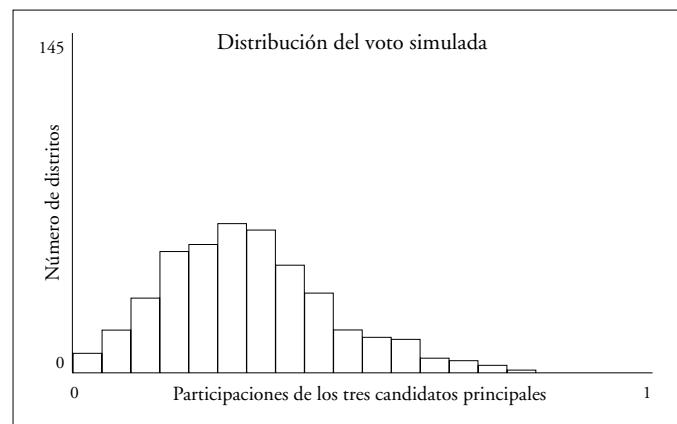
(b) Con fraude



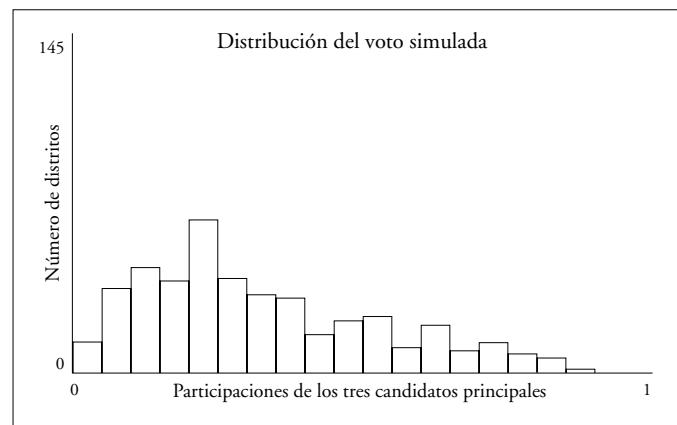
Nota: en la retícula (o espacio) de preferencias partidistas de la interfaz del programa computacional se identifica la opinión diaria de cada votante con un punto de distintas tonalidades. Los puntos de tono oscuro corresponden a Calderón y los de tono claro corresponden a AMLO. El que los puntos oscuros dominen abrumadoramente en determinadas zonas electorales de la pantalla (b) y no así en la pantalla (a) es indicativo de la presencia de fraude electoral.

Figura 2
Distribución de las participaciones de votos al nivel distrital
(con datos de los tres principales contendientes)

(a) En elecciones limpias



(b) Con fraude



Como una primera aproximación al problema se procede a corroborar si, efectivamente, los datos artificiales generados bajo un proceso electoral limpio dan lugar a una distribución diferente a la que se produce en un proceso fraudulento. Por construcción, el observador sabe cuándo los datos provienen de un ABM en el que se simula un fraude, por lo que esta prueba permite validar si dicho fraude genera un patrón emergente singular. A continuación se calcula la prueba KS para una alteración promedio del 8.06% de los votos emitidos. Nivel de manipulación que se produce con el criterio de “autoridades” que se utiliza para determinar cuáles son los distritos “capturados”, y la fabricación de votos a partir del método de “cambio de votos”.¹⁶

La tabla 1 presenta las estadísticas-D y los valores-P de las pruebas KS correspondientes a diez corridas en las que se usan las especificaciones anteriores. En todos los casos, los valores-P son menores a 0.1 por lo que el modelo siempre rechaza la hipótesis de que no existe diferencia entre las dos muestras. En consecuencia, para este grado de manipulación, el ABM es una herramienta poderosa para detectar histogramas “perturbados”, es decir, frecuencias relativas de resultados electorales en los que el voto es inducido o el conteo es manipulado.

Tabla 1
Pruebas KS para datos artificiales en elecciones limpias y fraudulentas

Corrida	1	2	3	4	5	6	7	8	9	10
est-D	0.137	0.111	0.153	0.093	0.132	0.137	0.106	0.130	0.174	0.143
valor-P	0.001	0.009	0.000	0.046	0.001	0.001	0.014	0.001	0.000	0.000

Prueba de Kolmogorov-Smirnov para detectar fraude en las elecciones mexicanas de 2006

Una vez mostrado que el ABM produce regularidades estadísticas diferentes en las simulaciones de procesos electorales limpios y trucados, el modelo se emplea para determinar de qué tipo son los resultados electorales observados en la contienda mexicana. Por este motivo el ABM, con parámetros calibrados, se corre diez veces con

¹⁶ En el programa de cómputo se establecen los parámetros de manipulación-de-voto y control-regional en los valores 15 y 20%, respectivamente. El control-regional define el porcentaje mínimo de adherentes para que una zona electoral tenga la posibilidad de ser considerada un reducto partidista, mientras que la manipulación-de-voto define el porcentaje de votos que son modificados en cada “zona-capturada”.

cada uno de los dos tipos de proceso electoral. La tabla 2 presenta las estadísticas-D y los valores-P de las pruebas KS cuando la distribución de las participaciones del voto obtenidas con distintas muestras artificiales se compara con la distribución observada en las elecciones presidenciales reales.

La misma tabla 2 ilustra el efecto que sobre los resultados de la prueba tiene el criterio con el que se define a los distritos “capturados” (autoridades, enclaves, inconsistencias) en escenarios en que el fraude se produce bajo la mecánica del “cambio de votos”. En estas simulaciones se supone que el porcentaje de votos manipulados en los distritos “controlados” por el PAN es de 12, 10 y 16, y que el valor mínimo para alcanzar el control regional es de 20% (i.e., con las preferencias políticas iniciales dicho porcentaje implica que 8 de las 16 zonas electorales no son consideradas como bastiones panistas); en consecuencia, el porcentaje de votos alterados a nivel nacional con estos tres criterios es de 6.38, 6.36 y 6.83, respectivamente.

Tabla 2
Prueba KS para las distribuciones de las participaciones del voto
(comparaciones entre las muestras real y simuladas)

Corrida	Elecciones limpias		Fraude: autoridades	Promedio 6.38 %	Fraude: enclaves	Promedio 6.36 %	Fraude: Inconsis.	Promedio 6.83 %
	estad-D.	valor-P	estad-D	valor-P	estad-D	valor-P	estad-D	valor-P
1	0.063	0.184	0.113	0.001	0.110	0.002	0.097	0.008
2	0.030	0.946	0.103	0.004	0.122	0.000	0.078	0.055
3	0.042	0.659	0.105	0.003	0.103	0.004	0.109	0.002
4	0.027	0.982	0.134	0.000	0.105	0.003	0.132	0.000
5	0.038	0.782	0.080	0.044	0.098	0.007	0.112	0.001
6	0.025	0.993	0.129	0.000	0.114	0.001	0.062	0.205
7	0.068	0.126	0.100	0.005	0.131	0.000	0.115	0.001
8	0.059	0.253	0.090	0.016	0.100	0.005	0.082	0.038
9	0.038	0.793	0.107	0.002	0.075	0.071	0.093	0.012
10	0.052	0.407	0.119	0.000	0.117	0.001	0.072	0.089

Nota: el cambio de votos opera exclusivamente en los “distritos capturados”. Mientras que en las columnas 2 y 3 se comparan los datos reales con los simulados en un contexto de elecciones limpias, en las columnas 4-9 se comparan los datos reales con los simulados en un contexto de fraude. Los porcentajes de fraude considerados en las columnas 5, 7 y 9 son de un dígito ya que se definen a partir de la interacción de los valores asignados en el programa a dos parámetros: manipulación-de-voto y control-regional. Se trata de promedios ya que estos porcentajes se calcularon con el número de votos manipulados en cada una de las diez corridas. Los valores-P de las columnas 3, 5, 7 y 9 corresponden a la estadística D de los escenarios descritos en la columna previa.

De la columna 3 se desprende que los valores-P son siempre mayores a 0.1 cuando la distribución con datos reales se compara con la distribución simulada que se “hace crecer” en un proceso electoral limpio. En contraste, las distribuciones generadas con un proceso electoral amañado no se parecen al patrón electoral observado (columnas 4-9). Los valores-P de estas últimas pruebas son relativamente pequeños para los tres criterios de fraude utilizados, con la excepción de una corrida en donde el fraude electoral se implementa en distritos con muchas actas inconsistentes. Por lo tanto, las simulaciones muestran que cuando los votos alterados son alrededor del 6.5% de los emitidos, la posibilidad de que los datos reales correspondan a un proceso electoral fraudulento es muy reducida bajo los tres criterios considerados.

Para ilustrar el patrón emergente que resulta de un escenario fraudulento, la figura 3 compara las gráficas de percentiles que describen las participaciones de votos para los datos reales (línea continua) y simulados (línea discontinua). Estas gráficas corresponden a la corrida número 9 del criterio de enclaves presentado en la tabla 2. Como bien lo indica la prueba KS correspondiente, la figura indica que las dos frecuencias acumuladas no parecen provenir de la misma distribución teórica.

La posible detección de un proceso electoral amañado depende del grado de manipulación, por lo que la tabla 3 describe una serie de pruebas KS para escenarios que usan diferentes porcentajes de perturbación. Estas pruebas se aplican con diferentes criterios para definir el control panista pero se mantiene la misma mecánica de “cambio de votos”. En dicha tabla se muestra que la prueba no-paramétrica puede identificar fraude con una probabilidad muy alta (i.e., en 80-90% de las corridas) cuando el porcentaje de votos alterados se encuentra por encima del rango 4-5 en los criterios de autoridades y enclaves. Inclusive, con una probabilidad alta (siete de diez corridas) se observa que la distribución de las participaciones de voto con un fraude promedio cercano al 3.7% no es compatible con una elección limpia bajo el criterio de autoridades. En otras palabras, cuando las modificaciones son del orden del 3% o menos esta prueba forense no resulta ser muy precisa ya que tanto las elecciones limpias como las trucadas “hacen crecer” el mismo tipo de distribución. Por otra parte, las dos últimas columnas de la tabla indican que con el criterio de inconsistencias, la probabilidad de que se produzca un patrón diferente en caso de fraude es baja para niveles de manipulación de 6% o menos.

Figura 3
Comparación de las gráficas de percentiles: datos reales y simulación fraudulenta

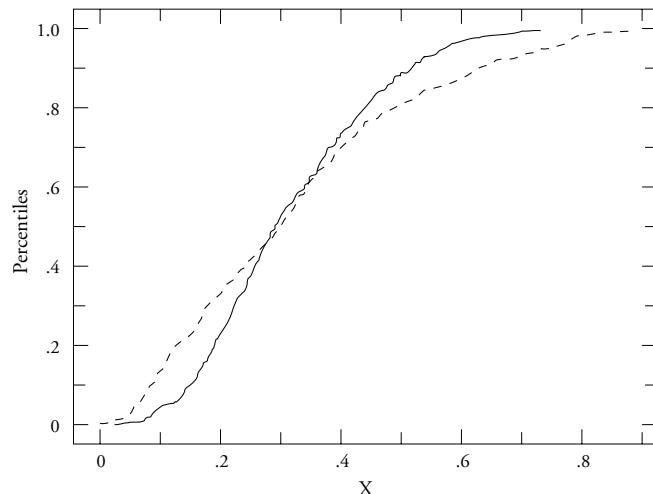


Tabla 3
Grados de manipulación y el poder de la prueba forense (valores-P de la prueba KS)

Criterios de control	autoridades	autoridades	autoridades	enclaves	enclaves	enclaves	Incons.	Incons.
Promedio:	4.25%	3.71%	3.2%	5.13%	3.86%	2.56%	6.4%	5.95%
1	0.010	0.001	0.457	0.017	0.156	0.011	0.010	0.141
2	0.016	0.087	0.176	0.005	0.162	0.506	0.005	0.441
3	0.083	0.246	0.130	0.184	0.001	0.451	0.002	0.527
4	0.068	0.097	0.027	0.012	0.141	0.467	0.004	0.015
5	0.003	0.027	0.222	0.613	0.066	0.104	0.192	0.342
6	0.089	0.016	0.058	0.000	0.038	0.020	0.005	0.006
7	0.048	0.162	0.013	0.008	0.304	0.009	0.407	0.142
8	0.333	0.300	0.160	0.009	0.061	0.655	0.011	0.013
9	0.100	0.002	0.008	0.089	0.000	0.108	0.000	0.005
10	0.021	0.000	0.377	0.006	0.184	0.059	0.337	0.530

Nota: en los escenarios de autoridades y enclaves el control-regional se establece en 20%, y el “cambio de votos” opera exclusivamente en los distritos “capturados”. Las columnas muestran diferentes criterios de control panista (primer renglón) y porcentajes de fraude promedio (segundo renglón). Los promedios para cada columna se calculan a partir del número de votos alterados en las distintas corridas.

La tabla 4 presenta diez corridas más para otros dos porcentajes diferentes de manipulación pero, en este caso, suponiendo que la mecánica del fraude se produce a través de la “creación de votos”; es decir, la voluntad de los individuos que decidieron no ejercer su voto se ve violentada ya que se utilizan sus boletas para fabricar votos a favor de Calderón. Al comparar los datos artificiales provenientes de simulaciones con procesos limpios y amañados, se puede identificar un fraude electoral para las elecciones de 2006 con un grado de manipulación del 6% o superior. Esto es así ya que en siete de las diez corridas se obtuvieron valores-P menores al 0.1, mientras que una relación inversa se da para una manipulación de 5% de los votos.

Tabla 4
Fraude a través de la “creación de votos” en los distritos “capturados”
(valores-P de la prueba KS bajo el criterio de autoridades)

Corrida: % Prom.	1	2	3	4	5	6	7	8	9	10
6.01	0.073	0.286	0.050	0.066	0.036	0.228	0.059	0.196	0.001	0.017
5.07	0.272	0.004	0.034	0.659	0.181	0.053	0.116	0.157	0.395	0.279

Nota: el control regional se establece en 20%. En la primera columna se muestran los porcentajes de fraude promedio para los dos escenarios con los que se realizan las diez corridas.

Análisis de sensibilidad

Resulta prácticamente imposible identificar una perturbación en la distribución de las participaciones del voto cuando la manipulación se dispersa en todo el territorio nacional, especialmente si el grado de fraude es relativamente pequeño. Sin embargo, un fraude de esta naturaleza es muy difícil de justificar cuando hay importantes variaciones espaciales en la militancia partidista de los gobiernos locales, en la fuerza relativa de sus adherentes y en la capacidad de los partidos para monitorear las casillas. Por todas estas razones es importante determinar, en primera instancia, las zonas electorales en las que es más factible que un fraude ocurra. Esto se puede llevar a cabo recolectando diferentes tipos de información: tasas inusuales de votación, falta de representación de los partidos en ciertas casillas, “errores” en el conteo de las actas que superan a la diferencia entre el primer y segundo lugar, entre otros.

En las siguientes simulaciones se muestra que el número de distritos electorales “capturados” es crítico para el funcionamiento adecuado de la prueba forense. La

tabla 5 presenta escenarios limítrofes del grado de manipulación en los que las pruebas KS tienden a rechazar la hipótesis de que las muestras trucadas y limpias provienen de la misma distribución teórica usando el criterio de inconsistencias. Cuando las manipulaciones se implementan exclusivamente en 6, 4, 3 y 2 zonas electorales, la hipótesis se rechaza en 7-8 de las diez corridas realizadas.

Cabe notar que el modelo logra detectar porcentajes de manipulación más pequeños (ver paréntesis de la primera columna) conforme el número de zonas electorales “capturadas” es menor. Por ejemplo, cuando sólo dos zonas electorales están involucradas en la fabricación de votos, la prueba del ABM puede identificar una manipulación promedio del 4.49%, la mitad del porcentaje observado en el caso de seis zonas electorales.

Tabla 5
Fraude a través del “cambio de votos” para diferentes cantidades de zonas “capturadas” (valores-P de pruebas KS bajo el criterio de inconsistencias)

Número de zonas (% de fraude promedio)	1	2	3	4	5	6	7	8	9	10
6 (8.99)	0.008	0.066	0.056	0.202	0.049	0.067	0.016	0.082	0.197	0.047
4 (6.4)	0.010	0.005	0.002	0.004	0.192	0.005	0.407	0.011	0.000	0.337
3 (6.08)	0.008	0.072	0.009	0.039	0.111	0.222	0.015	0.081	0.011	0.007
2 (4.49)	0.005	0.008	0.390	0.018	0.010	0.075	0.451	0.039	0.007	0.173
1 (3.94)	0.043	0.061	0.293	0.190	0.037	0.320	0.507	0.071	0.027	0.267

El análisis de sensibilidad también muestra que el ABM no puede diferenciar estadísticamente entre una elección limpia y una amañada (i.e., la hipótesis sólo se puede rechazar en la mitad de las corridas) cuando se fabrican votos en una sola zona electoral. En este escenario todos los votos no-panistas de la zona se cambiaron a favor de Calderón, lo que implica un grado de manipulación de 3.94% con respecto al total de votos emitidos en el país. Este último resultado implica que el método es inviable cuando el fraude está concentrado en un área muy reducida y el análisis se aplica a nivel nacional. Esto se debe a que no existen los eventos suficientes como para perturbar la distribución de participaciones cuando la mayor parte del fraude sucede en muy pocos distritos. En otras palabras, el poder relativo de la prueba tiene forma de U invertida ya que con muy pocas o con muchas zonas electorales capturadas el fraude no puede

ser detectado aun si éste existe. Debido a que el modelo computacional implementa una hipótesis conjunta que combina la especificación del grado de manipulación y el número de zonas electorales involucradas, una calibración más detallada del modelo es conveniente para mejorar la potencia de la prueba.

Errores aritméticos aleatorios

Diversos analistas políticos han señalado que las inconsistencias observadas en las actas de la elección de 2006 se debieron a errores aritméticos. Por lo tanto, para poder validar esta aseveración formalmente se incluye en el ABM de campañas electorales un módulo adicional para la creación de perturbaciones aleatorias no intencionadas en el conteo de votos. Específicamente, se supone que en las zonas electorales que presentan actas con un gran número de inconsistencias los funcionarios de casillas modificaron por error los porcentajes de votos recibidos por los distintos partidos, pero sin que existiera un sesgo a favor de un candidato en particular. Por lo tanto, para probar que, efectivamente, las equivocaciones neutrales son consistentes con los resultados electorales observados, se requiere que la distribución de las participaciones simulada con el módulo de errores aritméticos logre replicar la distribución calculada con datos electorales reales.¹⁷

La tabla 6 presenta la prueba KS para cinco corridas usando siete niveles de error. Cuando el porcentaje de equivocaciones es del orden de 6.5% o menor, los valores-P son siempre mayores que 0.1 y, por ende, no es posible rechazar la hipótesis de que la distribución real de las participaciones es consistente con la presencia de errores de conteo neutrales. Crespo (2008) encuentra un total de 365 955 votos con irregularidades en una muestra de 150 distritos, si este número se extrae a los 300 distritos electorales en que se divide el país se puede hablar, entonces, de 1.76% de irregularidades en el conteo final. Como se observa en la columna 6 de la tabla 4, este porcentaje de errores de conteo produce una distribución que es consistente con los hechos observados cuando las irregularidades se distribuyen aleatoriamente entre los diferentes candidatos.

¹⁷ Cabe mencionar que estos errores aleatorios son indistinguibles en un escenario en el que todos los partidos por igual incurrieron en diversos fraudes electorales.

Tabla 6
Prueba KS con errores de conteo aleatorios (valores-P)

Fraude promedio	17.16%	14.96%	10.69%	7.76%	6.49%	1.70%
1	0.024	0.352	0.005	0.672	0.892	0.242
2	0.069	0.029	0.237	0.580	0.347	0.848
3	0.190	0.338	0.378	0.056	0.545	0.342
4	0.413	0.100	0.254	0.517	0.501	0.785
5	0.081	0.008	0.613	0.409	0.576	0.152

Nota: incluye sólo las cuatro zonas electorales que registran el mayor número de inconsistencias en las actas de escrutinio.

Conclusiones

Los resultados de la elección de 2006 para la presidencia de México fueron desacreditados por un amplio segmento de la población encabezado por Andrés Manuel López Obrador, candidato opositor de la Coalición por el Bien de Todos. En sus propias palabras “El fraude electoral de 2006 ha sido a la vez el más burdo y el más sofisticado de la historia electoral de México. Se recurrió lo mismo a los métodos tradicionales como el acarreo de votantes, la sustitución de funcionarios de casilla, el relleno de urnas con votos ilegales y la falsificación de actas de escrutinio de las casillas, que a medios más sofisticados como la manipulación de los sistemas de cómputo” (*La Jornada*, 29 de agosto de 2006).

A la fecha este tipo de alegatos no han sido avalados con datos duros y mucho menos con métodos formales de análisis. En democracias jóvenes caracterizadas por instituciones débiles, resulta imprescindible que el desempeño de las autoridades electorales sea respaldado con evidencia sustantiva. Para que una democracia, como la mexicana, se pueda consolidar, es necesario que sus procesos electorales sean nítidos y, en caso de existir duda al respecto, es importante que las acusaciones de fraude sean examinadas seriamente y con transparencia. De aquí la relevancia de desarrollar e implementar pruebas forenses rigurosas para el estudio de procesos electorales.

El objetivo principal de este artículo no es ofrecer una conclusión definitiva sobre si las elecciones del 2006 fueron limpias o fraudulentas, más bien, la intención es presentar una prueba forense novedosa a partir de la modelación basada en agentes. Este enfoque puede explicar resultados electorales en términos de una variedad de elementos que inciden en una campaña (formación de preferencias, factores institucionales e incentivos), lo que hace viable la comparación entre patrones emergentes observados en los procesos reales con aquellos que se derivan de un mundo virtual.

En este último, el observador puede construir artificialmente procesos electorales limpios o trucados. A diferencia de otras pruebas, la simulación de un ABM permite que el observador conozca de antemano las implicaciones que tiene la manipulación del voto sobre ciertas regularidades estadísticas. Asimismo, las pruebas forenses con un ABM detectan no solo la alteración o creación de votos al momento del escrutinio, sino también la posibilidad de que algunos de ellos fueran “comprados” con antelación en distritos capturados por la maquinaria electoral de un determinado partido político.

En este artículo se muestra que el ABM, calibrado con datos mexicanos, es una buena herramienta forense para probar fraude cuando el grado de manipulación es de 5-6% de los votos emitidos. De hecho, este porcentaje se reduce a 3.7 cuando existe evidencia complementaria de que el fraude electoral sólo pudo haber sido orquestado en distritos electorales gobernados por panistas, o inclusive alcanzar una precisión mayor si se tiene información de que el fraude potencial se limita a unas cuantas zonas del país. Por lo tanto, las simulaciones realizadas con el modelo ofrecen una prueba convincente que permite descartar las acusaciones de un fraude masivo.

Dicho lo anterior, cabe advertir que, el modelo no es capaz de distinguir entre una elección limpia y una amañada cuando el fraude se efectúa con precisión micro-quirúrgica, como en el caso en que 0.56% de los votos son alterados. Debido a que éste es, precisamente, el margen que le dio la victoria a Calderón, no es posible sostener que un fraude de esta naturaleza queda descartado. Probablemente, un modelo más grande y detallado sea necesario para que la prueba forense aquí desarrollada tenga más precisión, lo que implicaría la necesidad de contar con información de preferencias políticas a nivel municipal.

Apéndice

Tabla A
Zonas electorales “capturadas” de acuerdo con diferentes criterios

zona	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
autoridades	sí					sí	sí	sí							sí	sí +
enclaves*	sí			sí		sí	sí	sí							sí	
inconsistencias**	7			1		4			2	2	5	5				8

* Fuente Estudio Panel México 2006, las zonas electorales se definen en Castañeda (2009).

** Ranking de acuerdo con el número más alto de inconsistencias en las actas de escrutinio.

Fuente: cálculos propios con datos de Gobierno Legítimo de México <<http://www.amlo.org.mx>>. + La zona 16 no se incluye cuando el control regional se establece en 20%.

Referencias

Aparicio, Javier, 2006a, “La evidencia de una elección confiable”, *Nexos*, núm. 346, México, D.F., octubre.

Aparicio, Javier, 2006b, “PREP y cómputos distritales: análisis de resultados”, México, D.F., CIDE, División de Estudios Políticos, manuscrito reportado en *Excelsior* (29/julio/2006) y disponible en la página de Internet <<http://investigadores.cide.edu/aparicio/elecciones/>> (última fecha de acceso: 15/octubre/2009).

Aparicio, Javier, 2006c, “Errores aritméticos en actas: análisis comparativo para 2000, 2003 y 2006”, CIDE, División de Estudios Políticos, manuscrito reportado en *El Financiero* (28/agosto/2006) y disponible en la página de Internet <<http://investigadores.cide.edu/aparicio/elecciones/>> (última fecha de acceso: 15/octubre/2009).

Aparicio, Javier, 2006d, “¿Por qué hay tantos errores e inconsistencias en las actas de escrutinio?”, CIDE, División de Estudios Políticos, manuscrito reportado en *Excelsior* (16/agosto/2006) y disponible en la página de Internet <<http://investigadores.cide.edu/aparicio/elecciones/>> (última fecha de acceso: 15/octubre/2009).

Aparicio, Javier, 2006e, “Recuento de casillas durante el cómputo distrital: un análisis preliminar”, CIDE, División de Estudios Políticos, manuscrito reportado en *Reforma* 12/agosto/2006) y disponible en la página de Internet <<http://investigadores.cide.edu/aparicio/elecciones/>> (última fecha de acceso: 15/octubre/2009).

Aparicio, Javier, 2006f, “¿Fueron determinantes las casillas atípicas para el resultado electoral?”, CIDE, División de Estudios Políticos, manuscrito disponible en la página de Internet <<http://investigadores.cide.edu/aparicio/elecciones/>> (última fecha de acceso: 15/octubre/2009).

Aparicio, Javier, 2006g, “Representantes de casilla y voto presidencial: un análisis preliminar”, CIDE, División de Estudios Políticos, manuscrito reportado en *Excelsior* (7/agosto/2006) y disponible en la página de Internet <<http://investigadores.cide.edu/aparicio/elecciones/>> (última fecha de acceso: 15/octubre/2009).

Castañeda, Gonzalo, 2009, “La Ley de Benford y su aplicabilidad en el análisis forense de resultados electorales”, manuscrito Colmex-CEE.

- Castañeda, Gonzalo e Ignacio Ibarra, 2009, “Campañas, redes de discusión y volatilidad de las preferencias políticas. Un análisis de las elecciones mexicanas del 2006”, por publicarse en *Foro Internacional*.
- Crespo, José Antonio, 2008, *2006: hablan las actas. Las debilidades de la autoridad electoral mexicana*, México, D.F., Debate, Random House Mondadori.
- Eisenstadt, Todd y Alejandro Poiré, 2006, *Explaining the Credibility Gap in Mexico's 2006 Presidential Election, Despite Strong (Albeit Perfectible) Electoral Institutions*, Cuaderno de Trabajo, American University Center for North American Studies.
- Kollman, Ken, John H. Millar y Scott E. Page, 1998, “Political Parties and Electoral Landscapes”, *British Journal of Political Science*, núm. 28, enero, pp. 139-58.
- Kollman, Ken, John H. Millar y Scott E. Page, 1992, “Adaptive Parties in Spatial Elections”, *American Political Science Review*, núm. 86, diciembre, pp. 929-937.
- La Jornada*, 29/agosto/2006, citado en Tello Díaz (2007) p. 176.
- Mebane, Walter Jr., 2006, *Election Forensics: Vote Counts and Benford's Law*, Working Paper, Departamento de Ciencias Políticas, Universidad de Michigan.
- Poiré, Alejandro y Luis Estrada, 2006, “Allegations of Fraud in Mexico's Presidential Election”, presentación en Power-point preparada para 102avo encuentro de APSA, Filadelfia Penn, septiembre 1.
- Pliego, Fernando, 2007, *El mito del fraude electoral en México*, México, D.F., Editorial Pax.
- Romero, Víctor, 2006, “Voto rural y voto urbano. Una opinión”, Manuscrito Instituto de Física, México, UNAM.
- Tello Díaz, Carlos, 2007, *2 de julio. La crónica minuto a minuto del día más importante de nuestra historia contemporánea*, México, D.F., Editorial Planeta.
- Ugalde, Luis Carlos, 2008, *Así lo viví. Testimonio de la elección presidencial de 2006, la más competida en la historia moderna de México*, México, D.F., Editorial Grijalbo.

Recibido el 1 de julio de 2009.
Aceptado el 26 de noviembre de 2009.