

Estimación de la intención de voto a nivel municipal: Modelo MRP en las elecciones del Edomex 2023

Estimated Voting Intention at the Municipal Level: MRP Model in the State of Mexico's 2023 Elections

Daniel Pérez*
Rodrigo Castro Cornejo**

Recibido el 1 de mayo de 2024
Aceptado el 27 de junio de 2024

Resumen

Este artículo estudia la campaña electoral de 2023 para elegir al gobernador del Estado de México, con un enfoque en la importancia del nivel local de las elecciones. A pesar de que las encuestas electorales tradicionales solo proporcionaron información a nivel estatal, este estudio ofrece estimaciones detalladas de la intención de voto por municipio por medio de un modelo MRP. Utilizando un modelo multinivel con post estratificación (MRP), se generaron estimados de la intención de voto para cada candidata en los 125 municipios del Edomex. Los resultados muestran una competencia heterogénea, reflejando cambios significativos en la dinámica electoral, particularmente en regiones históricamente dominadas por un partido político. Además, se evaluó la precisión del modelo en comparación con los resultados reales de la elección, demostrando una alta correlación. Se discute la relevancia de permitir la divulgación de encuestas en los días previos a la elección para mejorar la calidad de la información disponible para los votantes y candidatos.

Palabras clave

Elecciones, encuestas, candidatas, municipios, referencias electorales, modelo multinivel.

* Egresado de la licenciatura en Ciencia Política y Relaciones Internacionales en el Centro de Investigación y Docencia Económica (CIDE A.C.). Investigador en materia electoral y opinión pública en la firma encuestadora Áltica Research. daniel@altica.net. ORCID: 0009-0009-6956-7164.

** Doctor en Ciencia Política por la Universidad de Notre Dame. Profesor Asistente en la Universidad de Massachusetts-Lowell y Director Asociado del Centro de Opinión Pública de la misma universidad. rodrigo_castrocornejo@uml.edu. ORCID: 0000-0001-7232-9622.

Abstract

This article studies the 2023 gubernatorial campaign in State of Mexico, focusing on the importance of the local level in elections. Despite traditional electoral polls providing only state-level information, this study offers detailed estimates of voting intentions by municipality through an MRP model. Using a multilevel model with post-stratification (MRP), estimates of voting intentions were generated for each candidate in the 125 municipalities of the State of Mexico. The results show heterogeneous competition, reflecting significant changes in electoral dynamics, particularly in regions historically dominated by one political party. Additionally, the model's accuracy was evaluated compared to actual election results, demonstrating a high correlation. The relevance of allowing the disclosure of surveys in the days leading up to the election to enhance the quality of information available to voters and candidates is discussed.

Keywords

Elections, surveys, candidates, municipalities, electoral preferences, multilevel model.

Gracias a las encuestas fue posible conocer la intención de voto en momentos clave de las campañas a gobernador en el Estado de México. Tan solo en el portal que el Instituto Electoral del Estado de México (IEEM) dedicó para recopilar estos estudios, podemos contabilizar hasta antes del día de la elección, cerca de 148 encuestas preelectorales, de las cuales 102 fueron realizadas durante el periodo oficial de campaña.¹

Lamentablemente, la mayoría de estos estudios solo se centran en reportar un dato: la intención de voto a nivel estatal sin ofrecer un desglose detallado de las preferencias por municipio. Ante la ausencia de esta información, este texto pone énfasis en la importancia del nivel local de las elecciones al ofrecer estimados de la intención de voto por municipio. En el caso del Edomex, conocer el dato municipal de la intención de voto es relevante por dos razones. En primer lugar, por el número de municipios que conforman el estado, 125 en total, y cuya posición geográfica nos indica que cada municipio mexiquense tiene características particulares que pueden influir en el voto estatal. La segunda razón tiene que ver con los patrones de voto en el estado. En Edomex, hasta antes de la elección, donde el Partido Revolucionario Institucional ha gobernado más de 60 años, a nivel local se han conformado regiones que agrupan a municipios que históricamente han votado a favor de un partido, las más importantes son el Corredor Azul, de hegemonía panista, y el *cinturón amarillo* de dominio perredista.

Aunado al contexto político-geográfico del Edomex, el motivo de este artículo es aplicar una técnica estadística considerada en la literatura de opinión pública el ‘método estándar’ para estimación en áreas pequeñas: el modelo multinivel con post-estratificación o MRP.² Esta técnica permite, por un lado, obtener estimaciones de áreas donde es difícil tener información al aplicar una encuesta, ya sea por un tema de acceso o por diseño de la muestra. Y, por otro lado, ayuda a mejorar la información reportada por una encuesta al incluir fuentes de datos externas, como datos censales o históricos de votación, información que, según la literatura en comportamiento electoral, está relacionada con el contexto específico de los encuestados y que puede mover el voto mexiquense en cada municipio.³

La estructura de este artículo es la siguiente. Presentamos primero las razones por las cuales una estimación usando MRP es relevante en el contexto de las elecciones locales en Edomex. Seguido de los fundamentos teóricos que sostiene el uso de este método estadístico aplicado a datos de encuestas, además de ejemplos donde un modelo con postestratificación ha sido usado en México y otros países. En

1 Portal IEEM | Encuestas Electorales

2 Bisbee, J. (2019). BARP: Improving Mister P Using Bayesian Additive Regression Trees. *American Political Science Review*, 113(4), 1060–1065.

3 Bowler, Shaun. “Party identification” en *The Routledge Handbook of Elections, Voting Behavior and Public Opinion* editado por Justin Fisher, Edward Fieldhouse, Mark N. Franklin, Rachel Gibson, Marta Cantijoch y Christopher Wlezién, 146-157. Abingdon, Oxon ; New York, NY: Routledge, 2018.

la segunda parte del artículo exponemos los datos que usamos para este modelo, explicamos el manejo que le dimos a los datos para instrumentalizar nuestras variables, presentamos la estructura formal del modelo, su implementación durante la campaña y el contraste con los resultados de la elección. Después, mostramos los resultados de nuestro modelo, la intención de voto en los municipios, en el estado y nos concentraremos en describir las estimaciones en las regiones con más peso político en Edomex. Finalmente, abrimos la discusión a las mejoras a nuestro modelo, su uso futuro en diferentes elecciones, los datos que se pueden incorporar y sugerimos cambios a la regulación de las encuestas, especialmente eliminar la prohibición de publicar encuestas los días previos a una elección.

Contexto: La diversidad política del Edomex

El Estado de México es un estado complejo en términos políticos y geográficos. No solo porque concentra una parte considerable del padrón electoral del país sino también porque su extensión geográfica hace que tenga frontera con otros nueve estados, lo que genera dinámicas particulares dentro y fuera de sus municipios. Evidencia de lo anterior es que, a pesar de un marcado dominio priista en el gobierno estatal, partidos opositores han logrado construir bastiones regionales en diferentes municipios. Los dos casos más importantes son el “corredor azul” de hegemonía panista (formado por los municipios Atizapán de Zaragoza, Naucalpan, Huixquilucan, Tlalnepantla, Tultitlán, Cuautitlán Izcalli y Cuautitlán) y el *corredor amarillo* de dominio perredista con breve alternancia del PRI (lo constituyen municipios como Nezahualcóyotl, Texcoco, Chicoloapan, La Paz, Ixtapaluca, Chalco y Valle de Chalco Solidaridad).⁴ Para tener una referencia del voto en estos bastiones, las tablas 1 y 2 muestran el voto en las últimas elecciones a presidente municipal y gobernador.

4 Valdés Cervantes, Paul. “Desafección política y voto de castigo”. En *La política local importa. La elección a gobernador en el Estado de México*. Coord. Ramiro Medrano Gonzáles y Marco Aurelio Cienfuegos Terrón. 135-156. Toluca de Lerdo: Universidad Autónoma del Estado de México, 2018.

Tabla 1. Votación Morena y PAN 2017, 2018 y 2021 en corredor azul (IEEM)

Corredor Azul	Elección 2017 (Gobernador)		Elección 2018 (Alcaldes)		Elección 2021 (Alcaldes)	
	Morena	PAN	Alianza Morena	Alianza PAN	Alianza Morena	PAN o Alianza
Huixquilucan	20%	30.1%	16.72%	54.57%	21.26%	53.9%
Naucalpan de Juárez	32.8	21%	39%	35.94%	34.6%	52.13%
Atizapán de Zaragoza	33.7%	23.1%	34.985	25.79%	34.57%	55.40%
Cuautitlán	37.5%	12%	41.13%	14.41%	31.23%	36.6%
Cuautitlán Izcalli	40.2%	15.5%	41.80%	33.80%	37.8%	47.8%
Tultitlán	38.3%	9.95%	47.13%	22.80%	45.36%	10.96%
Tlalnepantla de Baz	36.3%	18.4%	44.16%	26.10%	40%	51.8%

Tabla 2. Votación Morena y PRI 2017, 2018 y 2021 en cinturón amarillo (IEEM)

Cinturón Amarillo	Elección 2017 (Gobernador)		Elección 2018 (Alcaldes)		Elección 2021 (Alcaldes)	
	Morena	PRI	Alianza Morena	PRI	Alianza Morena	PRI o Alianza
La Paz	29.8%	35.2%	41.25%	26.72%	25.53%	49.6%
Chicoloapan	32.2%	30.9%	47.82%	28,79%	43.1%	36.5%
Valle de Chalco	31.1%	22.5%	53.68%	12.67%	49%	35%
Texcoco	50.9%	26.4%	53.65%	13.23%	42.01%	18.3%
Ixtapaluca	34.7%	27.9%	37.7%	45.02%	44.97%	38.21%
Nezahualcóyotl	23.9	17.4%	38.9	13.55%	39.40%	21.9%
Chalco	34.0	28.9%	53.21%	19.92%	52.14%	34.8%

Esta regionalización del voto dentro del Edomex es importante considerarla porque, como lo indica la literatura de identidad partidista, el nivel de partidismo está estrechamente relacionado con el número de votos que obtiene un candidato o partido en una región específica.⁵ No obstante, aunque el medio principal para conocer la intención de voto antes de una elección son las encuestas electorales, conocer la opinión del votante en todos los municipios del Edomex implica diferentes retos, tanto metodológicos como económicos o de seguridad. Por estas razones, aplicar un método estadístico que permita estimar la opinión pública abarcando todo el conjunto de áreas dentro de un espacio geográfico es sumamente útil.

5 Bowler, Shaun. "Party identification" en The Routledge Handbook of Elections, Voting Behavior and Public Opinion editado por Justin Fisher, Edward Fieldhouse, Mark N. Franklin, Rachel Gibson, Marta Cantijoch y Christopher Wlezien, 146-157. Abingdon, Oxon ; New York, NY: Routledge, 2018.

Literatura: El modelo multinivel con post estratificación (MRP)

La estimación en áreas pequeñas es un conjunto de técnicas estadísticas que intentan calcular información en espacios donde es imposible tener datos de manera directa. El principal uso de estos métodos es generar inferencia en lugares con poca o nula información (por ejemplo, cuando el diseño de una encuesta asigna pocos casos a un lugar) a partir del uso de datos agregados.⁶

Con datos de las encuestas publicadas a lo largo de esta campaña, aplicamos un modelo multinivel con post estratificación (MRP) para generar estimados de la intención de voto para cada candidata a gobernadora en los 125 municipios de Edomex. Esta técnica ya ha sido usada en diversos estudios a nivel internacional, por ejemplo, para obtener estimados de voto o de opinión sobre temas de política pública (ejemplo la opinión sobre derechos LGBT o el apoyo al Brexit) en áreas geográficas pequeñas.⁷ En México, un ejemplo reciente es el modelo de voto para la Cámara de Diputados de 2021 publicado por Oraculus MX que combinó esta técnica con un modelo de agregación de encuestas para generar estimaciones a nivel distrital.⁸ En un ejercicio similar, pero para las elecciones presidenciales, Cortina, Gelman y Lasala (2008), aplican un modelo MRP a las elecciones de 1994, 2000 y 2006, para entender la influencia del contexto económicos de los estados en los resultados de tales comicios.⁹

Aunque el uso de los modelos MRP para estimar la opinión pública en áreas pequeñas es considerado el “método estándar” en ciencia política, existen otras alternativas. Por ejemplo, una de las primeras soluciones a el número limitado de casos que una encuesta puede tener para cada área geográfica pequeña es la agregación de múltiples encuestas hasta tener un número de observaciones proporcional a cada área y promediar los resultados para tener un estimado de intención de voto.¹⁰ Otra alternativa, que es la utilizada en la mayor parte de estudios de opinión, es ajustar mediante diversos parámetros (los más utilizados son población, recuerdo de voto, escolaridad) los resultados de una encuesta después del levantamiento. La idea es que, mediante ese ajuste o ponderación,

6 Stephen Ansolabehere, Samantha Luks, Brian F. Schaffner, “The perils of cherry picking low frequency events in large sample surveys”. *Electoral Studies*, 40 (2015): 409-410

7 Para el caso del Brexit ver: Introducing the YouGov Referendum Model | YouGov. Y el caso de derechos LGBT: Lax, Jeffrey R., and Justin H. Phillips. “Gay Rights in the States: Public Opinion and Policy Responsiveness.” *American Political Science Review* 2009.103 (3): 367-86

8 Oraculus MX, Modelo Cámara de Diputados 2021. Metodología: <https://oraculus.mx/diputados2021/>

9 Cortina, Jerónimo, Andrew Gelman and Narayani Lasala Blanco. “One vote, many Mexicos: income and vote choice in the 1994, 2000, and 2006 presidential elections” (unpublished).

10 Lax, J.R. and Phillips, J.H. “How Should We Estimate Public Opinion in The States?”. *American Journal of Political Science* 53, 1 (2009): 107-121

lo reportado en la encuesta refleje las características de la población de donde se extrajo la muestra.¹¹

La ventaja de estas alternativas es la relativa simplicidad de la técnica. Para la primera solo basta concentrar un número considerable de encuestas hasta tener una muestra similar en cada subunidad geográfica y después desagregar por cada una. Con la segunda basta incluir información externa, por ejemplo, de censos, y calcular el porcentaje real de personas para cada parámetro, asignando pesos distintos para corregir la desproporción de los resultados (por ejemplo, que haya más hombres que mujeres, o personas de ingresos bajos sobre los de ingreso medio).

Sin embargo, la solución de estas alternativas es bastante limitada. En el caso de la simple agregación, requiere conseguir un número considerable de encuestas hasta tener entrevistas suficientes para cada subunidad, para lo cual no hay un número específico. Una encuesta preelectoral en vivienda tiene en promedio 1,000 a 2,500 casos y la distribución de cada uno depende del tipo de diseño muestral. Y para la ponderación, las limitantes vienen de los datos de la propia encuesta: desde problemas en campo que afectaron la no respuesta, el fraseo de las preguntas de intención de voto, la sobrerrepresentación de alguna categoría, por ejemplo, personas de más ingreso o de más escolaridad pudieran responder con más facilidad una encuesta o incluso, el propio método de levantamiento, dependiendo si es en vivienda, telefónico o en línea.

La propuesta de un modelo MRP, aunque basada en modelos multinivel, incorpora algunas características de las alternativas antes mencionadas. Un modelo MRP consiste en dos etapas principales. Primero, como en la agregación de encuestas, es posible acumular diferentes estudios que contenga una pregunta similar, tanto en fraseo como en categorías, aunque no es paso completamente necesario, incluso con una sola encuesta y cuya muestra incluya pocas subunidades es posible ajustar este modelo. Además de los datos de encuestas es necesario recabar información externa, sobre todo censal, con las características básicas de la población de estudio. Cortina, Gelman y Lasala (2008), utilizaron datos a nivel estatal del PIB per cápita (en dólares), porcentaje de mayores de 18 años en cada grado escolar, por sexo y edad, y el total de población en cada estado. Con estos datos se ajusta un modelo multinivel para calcular la probabilidad de que un entrevistado vote por algún candidato o partido al responder a la pregunta de intención de voto. El modelo multinivel permite asumir que los datos de la encuesta están agrupados dentro de diferentes categorías. Es decir, cada encuestado pertenece a una categoría de edad, de sexo, de escolaridad, y pertenece a una subunidad geográfica. Bajo este supuesto, la respuesta de un entrevistado no solo depende de las variables explicativas, sino que toma en cuenta la variación dentro de cada nivel de edad, sexo, etc. En el caso de datos de encuesta, en estados o municipios donde hay pocas entrevistas, las estimaciones se acercan al promedio dentro del

11 Benjamin E. Lauderdale, Delia Bailey, Jack Blumenau, Douglas Rivers. "Model-based pre-election polling for national and sub-national outcomes in the US and UK", *International Journal of Forecasting*, 36 (2), 2020: 399-413.

grupo, con el fin de sustituir la falta de información. La segunda etapa es similar a la ponderación: con los datos censales se procede a calcular el porcentaje de todas las combinaciones de sexo, edad, escolaridad en cada subunidad (estratos). Con estos porcentajes se pondera cada estimación derivada del modelo multinivel. Este paso permite considerar todos los tipos de posibles votantes que por diseño no entrarían en una encuesta, y les asigna un peso con base en el porcentaje de personas que hay en cada estrato.

Estrategia Empírica: Modelo MRP a las encuestas preelectorales en el Edomex

Para las dos candidatas, Alejandra del Moral y Delfina Gómez, construimos un modelo de regresión logística multinivel que toma como variable principal la respuesta a la pregunta “Si el día de hoy fueran las elecciones al Estado de México, ¿por qué candidata votaría usted?”. Codificamos como 1, la elección por cada candidata y como 0 el resto de las opciones. Es necesario subrayar que solo consideramos encuestas que incluyeran características demográficas y la residencia del encuestado, por lo que dejamos fuera estudios telefónicos, por redes sociales, o encuestas en vivienda sin esta información.¹² La etapa de postestratificación la hicimos calculando el número de hombres y mujeres en cada categoría de edad por municipio de la lista nominal en el Estado de México según el Instituto Nacional Electoral (INE).

Por transparencia, en la tabla 3, reportamos los insumos que utilizamos para la generación del modelo de estimación de preferencias electorales: el número de encuestas incluidas, el número de municipios que fueron incluidos en cada estimación y las observaciones (número de entrevistados). Cabe aclarar que solo modelamos la elección dentro de cada encuesta por Delfina Gómez o Alejandra del Moral, por lo que la no respuesta, el voto nulo u otra opción fueron excluidos.

12 Los estudios incluidos fueron: Parametría 17 de marzo (1,000 encuestados); Covarrubias 21 de marzo, 18 de abril y 16 de mayo (1,200 encuestados cada levantamiento); Reforma 28 de abril y 31 de mayo (1000 encuestados cada levantamiento); Berumen 12 de mayo (1228 encuestados); Áltica 14 de mayo (1000 encuestados); Enkoll 29 de mayo (1233 encuestados).

Tabla 3. Número de municipios y entrevistas dentro de cada estimación

Estimación	Días previos	Encuestas	Municipios	Observaciones
1	79	1	46	1,038
2	75	2	69	1,911
3	49	3	70	2,797
4	37	4	74	3,549
5	20	5	84	4,525
6	17	6	84	5,448
7	4	9	106	10,252

Como variables independientes contextuales incluimos el porcentaje de voto municipal por partido en la elección estatal de 2017, en la elección presidencial y de diputados de 2018 y en la elección federal de 2021 de los partidos en coalición de cada candidata. También añadimos el porcentaje de personas en condición de pobreza extrema y pobreza salarial por municipio (CONEVAL 2020). Además, para controlar la variación por características sociodemográficas incluimos el sexo, el grupo de edad y el municipio del encuestado.¹³ El modelo final para ambas candidatas es el siguiente:

1. Modelo logístico:

$$Pr(Y_i = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \cdot \%VotoGob2017i + \beta_2 \cdot \%VotoPres2018i + \beta_3 \cdot \%VotoDip2018i + \beta_4 \cdot \%VotoDip2021i + \beta_9 \cdot \%POBREZAINGREi + \beta_{10} \cdot \%POBREZAEXT2020i + \beta_{11} \cdot partidismo + u_{SEXO[i]} + u_{EDAD[i]} + u_{ESTADO[i]} + uid_{pri:partidismo}[i])$$

2. Efectos aleatorios

$$u_{SEXO[i]} \sim N(0, \sigma^2 SEXO)$$

$$u_{EDAD[i]} \sim N(0, \sigma^2 EDAD)$$

$$u_{ESTADO[i]} \sim N(0, \sigma^2 ESTADO)$$

$$u_{partidismo:EDAD[i]} \sim N(0, \sigma^2 partidismo:EDAD)$$

¹³ Se utilizaron los porcentajes de personas en pobreza 2020 y el porcentaje de la población con ingreso inferior a la línea de pobreza por ingresos 2020 (debajo de los \$14,240.00 mensuales en un hogar de 4 personas), según el Concentrado de indicadores de pobreza 2020 a nivel municipal del CONEVAL.

Donde $N(0, \sigma^2)$ denota una distribución normal con media 0 y varianza σ^2 .

Cabe mencionar que a lo largo de las campañas hicimos 6 estimaciones antes de las elecciones y una estimación pasados los comicios, esto debido a que la veda electoral prohíbe publicar resultados de encuestas previas a los días de la elección y no posible incorporar esas encuestas en el modelo.¹⁴

Resultados: estimaciones con el modelo MRP

Aunque los estimados estatales dan una idea general de la dinámica de la intención del voto, consideramos que una fotografía más detallada del desarrollo de esta campaña electoral la tenemos al considerar las preferencias electorales por municipio. En los mapas siguientes tenemos las preferencias a gobernadora de nuestra estimación previa a las elecciones. Una de las ventajas de realizar estimaciones electorales es que posible contrastar la efectividad de un modelo con los datos reales, en este caso con el resultado de la elección.

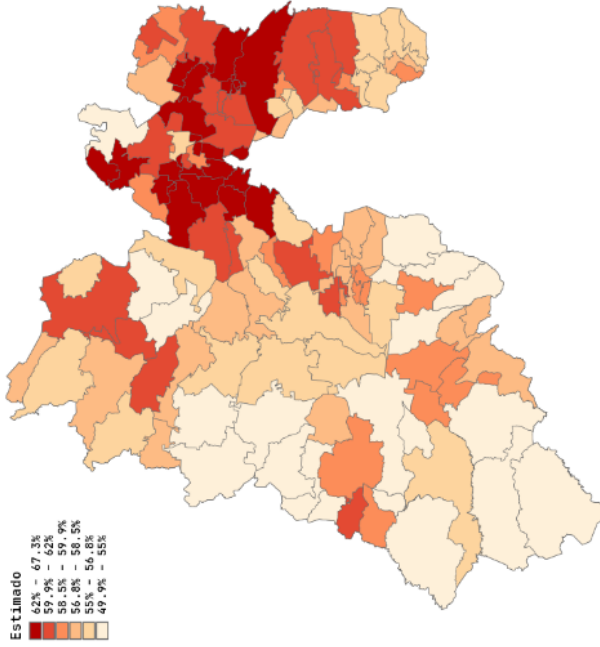
En los mapas la preferencia estimada muestra una concentración de la intención de voto a favor de Delfina Gómez en la zona llamada *corredor azul*. Los municipios de esta zona colindantes con el norte de la Ciudad de México (Atizapán de Zaragoza, Naucalpan, Huixquilucan, Tlalnepantla, Tultitlán, Cuautitlán Izcalli y Cuautitlán) históricamente han tenido gobiernos panistas. Aunque en 2017 y 2018 el voto morenista arrasó en estos municipios los resultados de la elección de 2021 suponían un regreso del PAN a esta región, de la mano de la alianza Va por México, tal y como lo mostramos en la tabla 1.

En la gráfica 1 mostramos las preferencias para las dos candidatas en estos municipios, ordenándolos de menor a mayor margen entre estimaciones. La gráfica muestra que solamente en Huixquilucan (bastión panista) las preferencias estarían cerradas mientras que en el resto del corredor Del Moral es superada por Delfina por más de dos cifras.

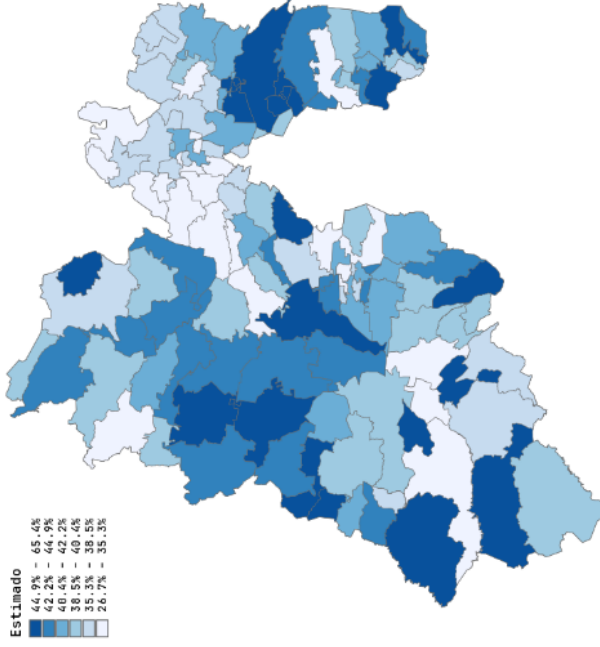
¹⁴ Los modelos finales se ajustaron mediante el software Stan por medio de su interfaz en R y la librería “rstanarm”. Se utilizó la función “sta_glmer()”.

Mapas 1. Preferencia electoral estimada para gobernador por municipios en Edomex
(estimación preelectoral 18 de mayo)

Preferencia electoral - Delfina Gómez

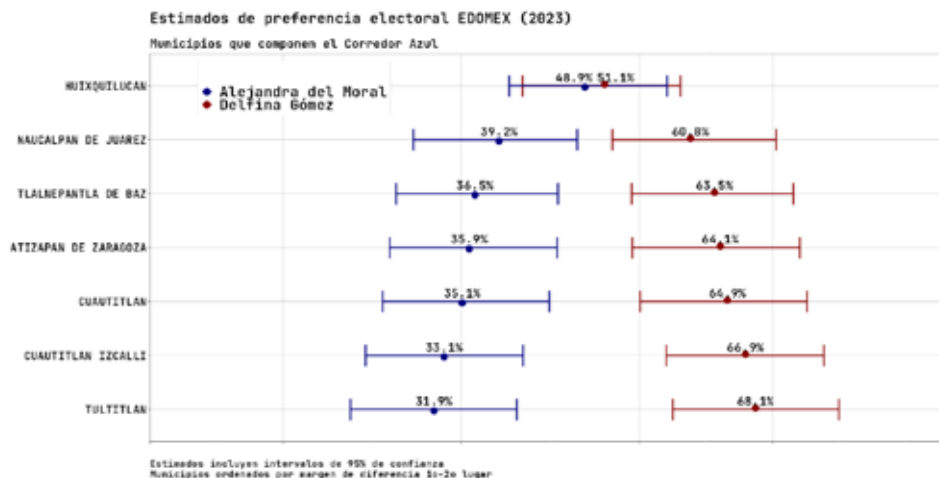


Preferencia electoral - Alejandra del Moral



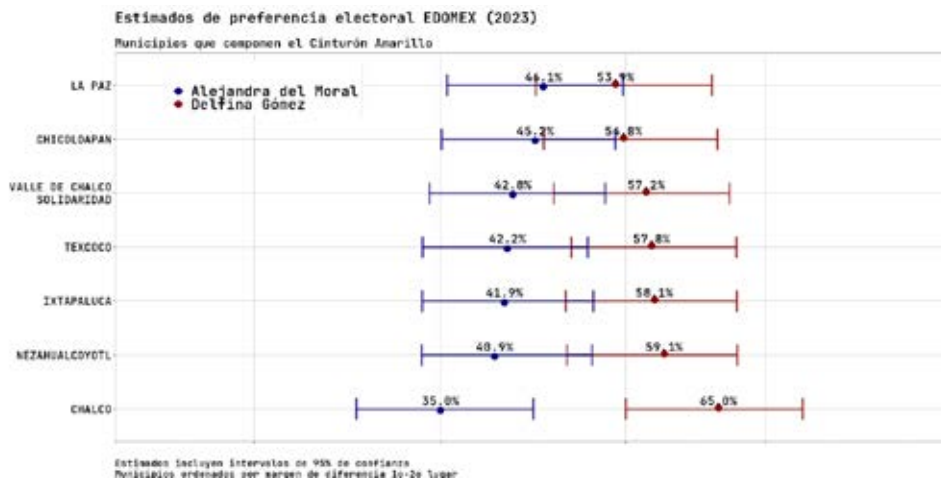
Última estimación realizada

Gráfica 1. Preferencias electorales estimadas en municipios del corredor azul (estimación preelectoral 18 de mayo)



Por su parte, la preferencia estimada para Alejandra del Moral luce dispersa en todo el estado y tiene los porcentajes más altos en Luvianos (65.4%), Chimalhuacán (54.3%), Villa Victoria (52.1%) y Huixquilucan (48.9%); aunque de estos municipios únicamente Chimalhuacán es gobernado por Morena. Incluso en los municipios del también llamado *cinturón amarillo*, la preferencia por Del Moral solo es alta en La Paz y Chicoloapan (46.1% y 45.2% respectivamente). En la gráfica 3 mostramos las estimaciones para los municipios del *cinturón amarillo* ordenadas también de menor a mayor diferencia entre estimaciones. Lo que vemos ahí es que la ventaja de Delfina Gómez sobre Alejandra del Moral, a diferencia del *corredor azul*, es un poco menos cerrada. Únicamente en Chalco la candidata de Morena parece tener una ventaja holgada, mientras que en La Paz y Chicoloapan Del Moral la diferencia es menor a los 10 puntos.

Gráfica 2. Preferencias electorales estimadas en municipios del cinturón amarillo (estimación preelectoral 18 de mayo)



En este *cinturón amarillo* —cuya alternancia entre PRI y PRD le ha dado ese nombre—, su característica principal ha sido una alta competencia partidista.¹⁵

Las cifras aquí presentadas reflejan el carácter local que pueden tener elecciones tan competidas como las de gobernador en el Estado de México. Por un lado, la tendencia general de las preferencias aquí estimadas muestra una ventaja considerable entre la candidata Delfina Gómez sobre su contraparte Alejandra del Moral. Nuestra primera estimación a 79 días de la elección ponía la preferencia por Delfina Gómez (67%) al doble de lo estimado para Alejandra del Moral (32.9%). Con el transcurso de la campaña esa diferencia entre candidatas disminuyó ligeramente, pero sin un cambio notable en la intención de voto. En la última estimación, 17 días previos, la preferencia por Del Moral llegó a 41.1% mientras que para Delfina Gómez apenas bajó a 58.9%.

Por otro lado, las estimaciones a nivel municipal nos dan cuenta de lo heterogénea que puede ser la competencia electoral al interior del estado. Regiones que históricamente han sido ganadas por un partido muestran un cambio significativo hacia un candidato o, por el contrario, regiones con alternancia conservan una marcada competencia. Las preferencias estimadas para *el corredor azul* y *el cinturón amarillo* son ejemplo de estos dos escenarios. En el *corredor azul* los cálculos aquí presentados sugieren un retorno de la fuerza morenista, aunque la disputa por Huixquilucan persiste al darse un margen cerrado para las dos candidatas—incluso en Cuautitlán Izcalli, donde gobernó Del Moral, la distancia con

15 González Anaya, Juan. *Nación 321*. “¿Los municipios gobernados por la oposición en Edomex votan contra el PRI?” 27 de marzo de 2017: <https://www.nacion321.com/elecciones/los-municipios-gobernados-por-la-oposicion-en-edomex-votan-contra-el-pri>

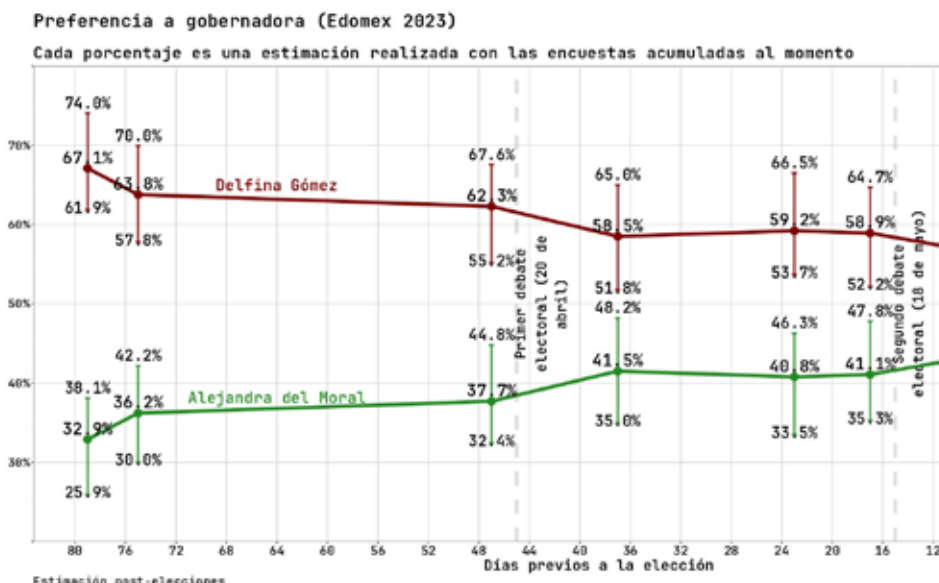
la morenista es amplia. En cambio, la intención electoral en el *cinturón amarillo* aunque favorece a la candidata Delfina Gómez, luce más reñida entre municipios siendo La Paz el más competido de esa zona.

Es importante señalar que el modelo de preferencias electorales que presentamos en este artículo depende de disponibilidad de las bases de datos de las encuestas publicadas—las cuales son depositadas en la página de internet del Instituto Electoral del Estado de México a veces días o incluso semanas después de haberse levantado. A continuación, comparamos los resultados de nuestra estimación de las preferencias electorales —tanto a nivel estatal y municipal— con los resultados de la elección. En resumen, encontramos que nuestra proyección a nivel estatal como municipal reportaron una muy alta precisión respecto de los resultados de la elección.

Resultado: modelo MRP vs resultados de la elección a nivel estatal y municipal

En la gráfica 3 reportamos las preferencias estimadas para Delfina Gómez y Alejandra del Moral basados en nuestro modelo de agregación de encuestas. Como explicamos, nuestro modelo únicamente incluye aquellas encuestas que, además de contar con información demográfica del encuestado, indican el municipio en

Gráfica 3. Preferencia electoral estimada para las candidatas a gobernadora en Edomex

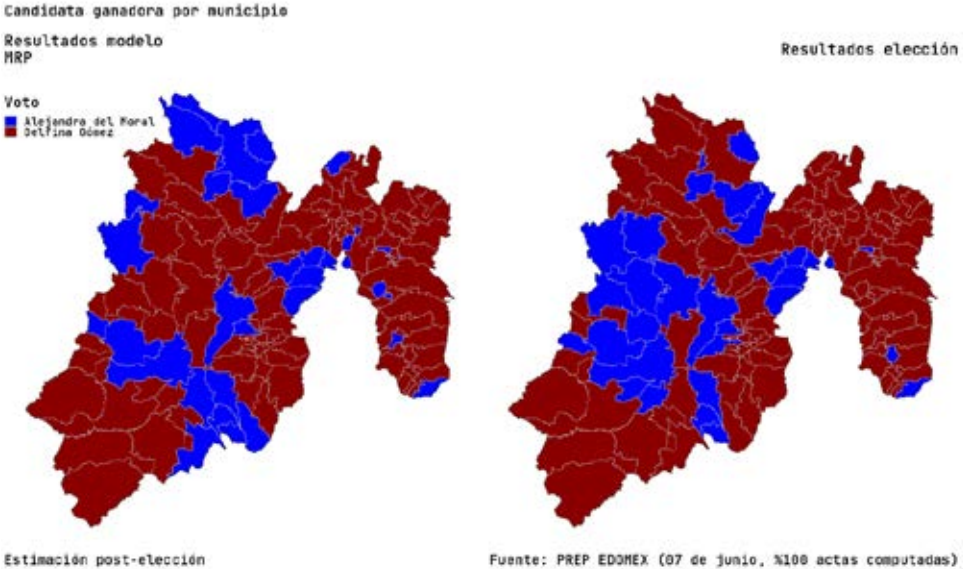


donde viven. También es necesario resaltar que esta última estimación la hicimos pasadas las elecciones ya que, por las restricciones a la publicación de encuestas, aunque estas encuestas se hayan levantado la semana previa a la elección, los datos de estos estudios solo estuvieron disponibles después del 4 de junio en el portal del Instituto Electoral del Estado de México (IEEM). En la gráfica cada porcentaje equivale a una estimación hecha con las encuestas acumuladas al momento.

Las estimaciones que reportamos en la Gráfica 3 para ambas candidaturas quedaron muy cerca de su votación obtenida (Delfina Gomez: estimada 55.3%, elección 54%; Alejandra del Real: estimada 44.7%, elección 46%). Para hacer la comparación entre las preferencias estimadas por nuestro modelo y resultados finales quitamos los votos nulos y de candidatos no registrados. La última diferencia estimada por nuestro modelo para estas candidatas fue de 10.7%, una cifra muy parecida al día de la elección que fue de 8%. También es interesante resaltar que la diferencia entre Delfina Gómez y Alejandra del Moral se redujo 7.1 puntos respecto de la estimación previa al segundo debate. En otras palabras, las campañas sí afectaron la intención del voto, particularmente en la recta final de las mismas.

En los siguientes mapas mostramos el comparativo entre la candidata ganadora por municipio con los resultados del PREP al 100% y la ganadora proyectada por nuestro modelo. De los 125 municipios del Estado de México, nuestro modelo logró acertar a la candidata ganadora en 100, es decir, alcanzó una precisión bastante alta del 80%. Como explicamos al inicio, el diseño muestral de las encuestas que

Mapa 2. Comparativo entre candidata ganadora proyectada y real en los municipios del Estado de México



utilizamos en nuestro modelo fue elegido por las casas encuestadoras para estimar las preferencias electorales a nivel estatal. No obstante, con nuestro modelo logramos aprovechar la información de esas encuestas y modelar las preferencias electorales a nivel municipal con un nivel bastante alto de precisión en la gran mayoría de los municipios del Estado de México. La precisión de nuestro modelo de preferencias electorales fue menor en los municipios del norte y oeste de esta entidad.

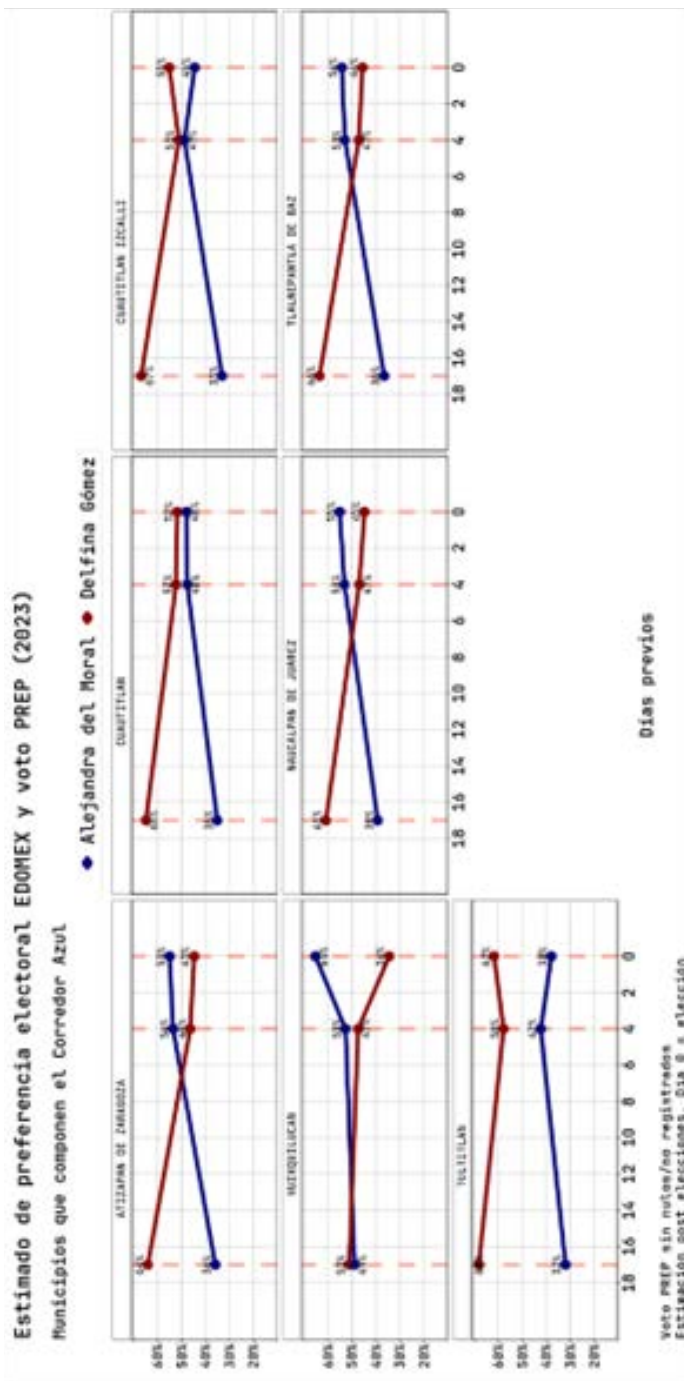
En la gráfica 4 mostramos el cambio entre estimación que realizamos después del segundo debate entre candidatas y la estimación días antes de la elección, en comparación con el voto final recibido en los siete municipios del *corredor azul*. La proyección al día siguiente del segundo debate era que Delfina Gómez ganaría en los siete municipios por una amplia ventaja, excepto en Huixquilucan, situación que cambió en nuestra segunda proyección al actualizar el modelo. De esta forma, nuestro modelo proyectó correctamente a la ganadora de la elección en los siete municipios del *corredor azul*. Únicamente en Huixquilucan y Tultitlán hubo algunas discrepancias en los porcentajes de voto, no obstante, se proyectó correctamente a la ganadora. Al igual que en la gráfica 4, en la gráfica 5 también se puede observar que en las últimas dos semanas de la campaña hubo un cambio notorio en la intención de voto con un repunte significativo en los municipios donde triunfó Alejandra del Moral.

De hecho, en un análisis sobre los bastiones partidistas en el Estado de México utilizando la misma técnica que los modelos de intención de voto, calculamos que la opinión positiva hacia el PRI-PAN en los municipios del *corredor azul*, donde ganó Alejandra del Moral, era bastante más alta respecto del promedio estatal: Huixquilucan, Tlalnepantla de Baz, Naucalpan de Juárez y Atizapán de Zaragoza.¹⁶ Nuestro análisis sugiere que, a pocos días de la elección, la opinión partidista positiva hacia PRI y PAN activó el voto a favor de la candidata de la alianza “Vamos por el Estado de México” en estos municipios.

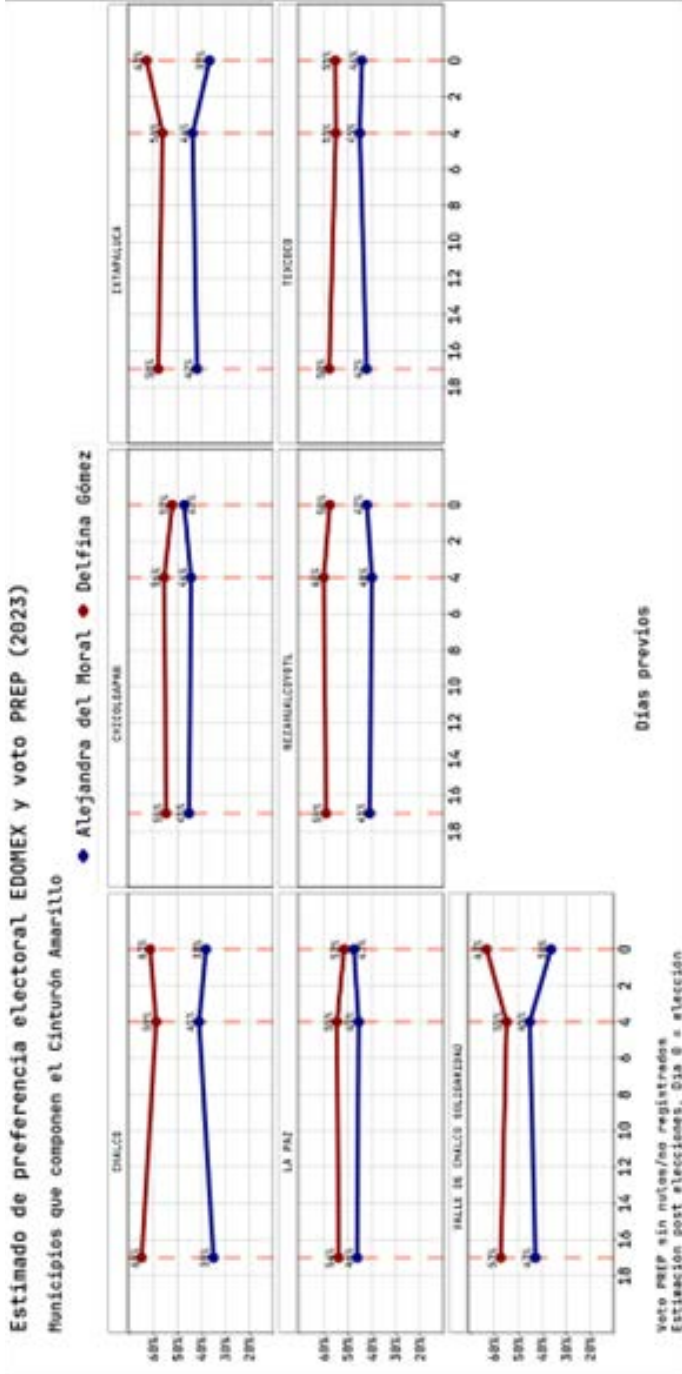
Pasando a los municipios del *cinturón amarillo*, reportamos una gráfica similar incluyendo la estimación al día del segundo debate y la elaborada con datos previos a la veda electoral. Ambos datos los comparamos con el voto PREP. En la gráfica 5, de nueva cuenta, observamos que nuestro modelo proyectó correctamente a la ganadora en todos los municipios del *corredor amarillo*. A excepción de Ixtapaluca y Valle de Chalco Solidaridad, la estimación previa a la elección se acercó bastante a los números del PREP, con una diferencia promedio de 3 puntos. La tendencia de estos municipios mexiquenses tuvo pocos cambios desde el segundo debate hasta el final de las campañas colocando a Delfina Gómez en primer lugar.

16 Pérez García, Daniel y Rodrigo Castro Cornejo. “¿Cuáles son los bastiones partidistas en el EDOMEX? Sobre el apoyo electoral a nivel municipal”. *Nexos, Taller de Datos*, 4 de mayo de 2023. <https://datos.nexos.com.mx/cuales-son-los-bastiones-partidistas-en-el-edomex-analisis-de-la-opinion-publica-a-nivel-municipal/>

Gráfica 5. Preferencia final estimada y voto PREP en municipios del corredor azul



Gráfica 6. Preferencia final estimada y voto PREP en municipios del cinturón amarillo

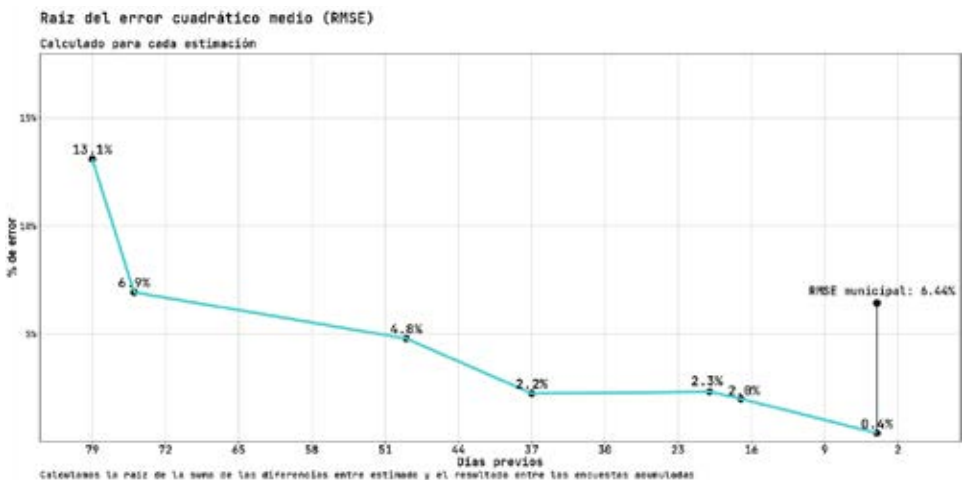


Resultados: evaluación del modelo de preferencias electorales

Las estimaciones realizadas con las encuestas publicadas a lo largo de la campaña mostraron que la diferencia en la intención de voto entre las candidatas mexicanas comenzó a disminuir pasado el segundo debate. Aunque nuestra última estimación fue realizada pasadas las elecciones (dado que las bases de datos de las últimas encuestas en el portal del IEEM solo estuvieron disponibles después de la elección), nuestro modelo proyectó un resultado muy cercano al voto registrado en el PREP. Además, la precisión de nuestro modelo fue aumentando conforme incluíamos más encuestas y más cercana estaba la elección.

En la gráfica 6, hacemos una evaluación de nuestras estimaciones a partir del cálculo de la raíz de las desviaciones del error cuadrático medio (RMSE), medida usada para conocer la precisión de nuestras proyecciones.¹⁷ Por ejemplo, la primera estimación realizada 79 días antes de la jornada electoral tuvo un error promedio de 13.1% de los resultados finales, este error se redujo a solo 0.4%. La gráfica 4 muestra la evolución del error en nuestras estimaciones. Nuestro error bajó considerablemente a partir de la tercera estimación y usando solo tres encuestas (ver tabla 3).

Gráfica 7. Cambio en el error de cada estimación respecto del resultado PREP



17 Houshmand Shirani-Mehr, David Rothschild, Sharad Goel & Andrew Gelman (2018) Disentangling Bias and Variance in Election Polls, *Journal of the American Statistical Association*, 113:522, 607-614

Sin embargo, como indicamos en la misma gráfica, el error final de las estimaciones municipales fue mucho más alto que el estimado estatal. Consideramos al menos dos explicaciones para este error. La primera es la concentración de las muestras de las encuestas en algunos municipios. Generalmente los municipios más poblados tienen mayor probabilidad de entrar en el diseño muestral de una encuesta y esto sesga las estimaciones. Dentro de un modelo multinivel las estimaciones de categorías faltantes, en este caso municipios, son acercadas a la media grupal.

La segunda explicación la relacionamos con la falta de información individual en las encuestas utilizadas. Aunque utilizamos sexo y edad para controlar la variación por grupos demográficos, en la literatura se sugiere incorporar otro tipo de datos como la escolaridad del encuestado, su ingreso, religión o identidad partidista. Con esta información es posible controlar la intención de voto por características más detalladas del votante y también, estimar efectos aleatorios más complejos en el mismo modelo multinivel.¹⁸ Esta información no siempre está disponible en las bases de datos de las encuestas que se suben a los institutos electorales locales durante las elecciones.

Sin embargo, a pesar del tamaño del error estimado a nivel municipal fue posible, como se mencionó al principio de este artículo, acertar a la candidata ganadora en el 80% de los municipios mexiquenses. Incluso como se ve en las gráficas 2 y 3, el porcentaje de voto estimado para Alejandra del Moral y Delfina Gómez tanto en el *corredor azul* como en el *cinturón amarillo* fue bastante cercano al voto final. Esto también muestra que, aun en el supuesto de que las encuestas sobreestimaran o subestimaran la preferencia hacia alguna de las candidatas, el modelo multinivel con post estratificación utilizado permitió corregir los posibles sesgos en la estimación estatal y tener proyección de voto en todos los municipios del Edomex.

Discusión

Como explicamos anteriormente, cuando incluimos en nuestro modelo de preferencias electorales todas las encuestas que se realizaron unos días antes de la elección, el resultado del modelo es muy parecido a los resultados de la elección. De los 125 municipios del Estado de México, nuestro modelo logró acertar a la candidata ganadora en 100, es decir, alcanzó una precisión bastante alta del 80%. Y esto a pesar que las encuestas que se analizaron en este estudio no estaban diseñadas específicamente para estimar las preferencias electorales a nivel mu-

18 Trangucci, Rob, Imad Alid, Andrew Gelman y Doug Rivers. "Voting patterns in 2016: Exploration using multilevel regression and poststratification (MRP) on pre-election polls": <https://arxiv.org/abs/1802.00842>

nicipal. Sin embargo, gracias al modelo MRP, se pudieron generar estimaciones a nivel municipal con un muy alto grado de precisión.

No obstante, es necesario resaltar mucha de esa información no estuvo disponible los días previos a la elección dada que la ley electoral prohíbe publicar encuestas desde el primer minuto del jueves anterior a la elección.¹⁹ Como nuestros modelos sugieren, muchos votantes tomaron su decisión electoral en los últimos días de la campaña. Pero las encuestas y modelos de agregación no siempre pueden reflejar esos últimos cambios debido a la prohibición de difundir encuestas. Lo cual aumenta la percepción de error de las encuestas, cuando en realidad tiene que ver, en parte, por la imposibilidad de levantar y difundir estos estudios en los últimos días de la campaña electoral. Esto impacta en la calidad de información que los votantes y candidatos reciben de parte de las encuestadoras y medios de comunicación. Por lo mismo, podría ser deseable extender el periodo de divulgación de las encuestas y únicamente prohibirlo durante el fin de semana cuando la elección se celebra.

Referencias

- Bisbee, J. (2019). BARP: Improving Mister P Using Bayesian Additive Regression Trees. *American Political Science Review*, 113(4), 1060–1065.
- Bowler, Shaun. “Party identification” en *The Routledge Handbook of Elections, Voting Behavior and Public Opinion* editado por Justin Fisher, Edward Fieldhouse, Mark N. Franklin, Rachel Gibson, Marta Cantijoch y Christopher Wlezien, 146-157. Abingdon, Oxon ; New York, NY: Routledge, 2018.
- Valdés Cervantes, Paul. “Desafección política y voto de castigo”. En *La política local importa. La elección a gobernador en el Estado de México. Coord. Ramiro Medrano Gonzáles y Marco Aurelio Cienfuegos Terrón*. 135-156. Toluca de Lerdo: Universidad Autónoma del Estado de México, 2018.
- Benjamin E. Lauderdale, Delia Bailey, Jack Blumenau, Douglas Rivers. “Model-based pre-election polling for national and sub-national outcomes in the US and UK”, *International Journal of Forecasting*, 36 (2), 2020: 399-413.
- Bowler, Shaun. “Party identification” en *The Routledge Handbook of Elections, Voting Behavior and Public Opinion* editado por Justin Fisher, Edward Fieldhouse, Mark N. Franklin, Rachel Gibson, Marta Cantijoch y Christopher Wlezien, 146-157. Abingdon, Oxon ; New York, NY: Routledge, 2018.
- Stephen Ansolabehere, Samantha Luks, Brian F. Schaffner, “The perils of cherry picking low frequency events in large sample surveys”. *Electoral Studies*, 40 (2015): 409-410

¹⁹ Ley General de Instituciones y Procedimientos Electorales, Capítulo III “De las encuestas y sondeos de opinión” (Instituto Nacional Electoral: México, 2014), art. 213, 251 párr. 5-7 y 253.

Cortina, Jerónimo, Andrew Gelman and Narayani Lasala Blanco. "One vote, many Mexicos: income and vote choice in the 1994, 2000, and 2006 presidential elections" (unpublished).

Lax, J.R. and Phillips, J.H. "How Should We Estimate Public Opinion in The States?". *American Journal of Political Science* 53, 1 (2009): 107-121

Pérez García, Daniel y Rodrigo Castro Cornejo. "¿Cuáles son los bastiones partidistas en el EDOMEX? Sobre el apoyo electoral a nivel municipal". *Nexos, Taller de Datos*, 4 de mayo de 2023. <https://datos.nexos.com.mx/cuales-son-los-bastiones-partidistas-en-el-edo-mex-analisis-de-la-opinion-publica-a-nivel-municipal/>