

# Notas de Investigación

## Evaluar el desempeño de las encuestas

*Nydia Iglesias/ Carlo Varela*

### RESUMEN EJECUTIVO

---

- El objetivo de esta nota es evaluar las encuestas pre-electorales publicadas en México desde 1994 hasta el 2016, tanto estatales como federales. Con ello pudimos valorar la veracidad de la idea de que las encuestas de las elecciones estatales del 2016 han sido “las peores de la historia”.
- Si bien las encuestas no son los únicos elementos para pronosticar resultados electorales, son los únicos públicos que existen en México. Nosotros estudiamos en particular los datos de encuestas más cercanos al evento electoral que estimaron.
- Revisamos los métodos disponibles para evaluar el desempeño de encuestas en términos de su capacidad predictiva del resultado observado de las elecciones.
- También revisamos las únicas tres evaluaciones que se realizaron sobre las encuestas del 2016 e identificamos varias limitaciones de las mismas.
- Elegimos dos criterios de evaluación: el error de predicción de cada encuesta respecto a cada uno de los candidatos principales, y si aciertan o no al ganador. Con ello proponemos identificar cuáles fueron los mejores/peores años de las encuestas y cuáles han sido las mejores/peores casas encuestadoras.
- Concluimos que, en efecto, el 2016 fue el peor año de las encuestas ya que sólo 39% de las mismas acertó al ganador aunque, en cuanto a su precisión estimativa no figura entre los peores años (lugar 11 de los 17 años evaluados). No obstante, al considerar simultáneamente ambos criterios con una ponderación equivalente, el 2016 quedó en el lugar 15.
- Los mejores años en esa serie de encuestas fueron, en orden descendente, 2007, 1994 y 2011 y los peores, 2016, 2015 y 2010.
- Para evaluar el desempeño de las distintas casas encuestadoras no tomamos en cuenta si habían hecho pocas o muchas encuestas a lo largo de estas dos décadas, ni si las habían realizado al inicio del periodo o si sólo habían participado recientemente.
- Las mejores encuestadoras, de las 25 consideradas, resultaron ser Berumen, Consultores y Marketing político, y Alduncin y Asociados. Y en los lugares 23 a 25 se ubicaron El Financiero/AM, MEBA y Demotecnia/RG . El consenso de las encuestas no resultó ser el mejor pronóstico, quedó en el lugar 10.

El 5 de junio del 2016, el PRI perdió el mandato en seis de los nueve estados que gobernaba. Las encuestas preelectorales<sup>1</sup> no nos prepararon para este resultado. Lejos de ello, al cierre de las campañas, encuestadores y analistas anticiparon que el PRI ganaría al menos nueve de las doce gubernaturas en juego, que el PAN podría obtener dos; y ya fuese Morena o el PRD, una. Finalmente, el PAN ganó siete -cuatro por sí mismo y tres en coalición con el PRD- y el PRI se quedó con cinco. Lo sorprendente de los resultados rehabilitó el debate sobre el desempeño de las encuestas preelectorales y sobre la evolución de esa industria en México.

En las semanas que siguieron a las elecciones, el juicio público contra las encuestas fue contundente: habían fallado en sus estimaciones y pronósticos. Se publicaron tres evaluaciones sobre las encuestas preelectorales, que emplearon los mismos métodos pero distintos criterios y arrojaron diferentes resultados. Ello evidenció que no existe aún consenso respecto al propósito último de las encuestas ni una metodología estándar para evaluarlas. Derivado de esa premisa, en esta nota: a) revisamos los métodos disponibles para evaluar el desempeño de encuestas, b) empleamos la metodología mejor conocida para evaluar el desempeño anual de la industria de encuestas en México desde 1994, c) revisamos el desempeño por casa encuestadora de las que han publicado a lo largo de ese período de tiempo y, finalmente d), a partir de esa evaluación, proponemos un índice para medir no sólo el desempeño

estadístico sino también el predictivo de las encuestas mexicanas desde 1994.

### ¿Qué información proporcionan las encuestas?

Dado que una elección es un evento de distribución de un total de votos entre los partidos o candidatos participantes en cada evento, una manera de entenderlas es, justamente, como estimaciones de distribución del voto. Se trata de estimaciones estadísticas que sólo pueden aproximarse a un fenómeno. Siempre presentarán un margen de error porque del estudio de una muestra derivan conclusiones para el universo general. Parte de ese margen es metodológicamente inevitable y aumenta o disminuye con la robustez de la muestra empleada, pero no desaparece del todo. Otra parte del error se origina en factores no muestrales, que van desde la organización y eficiencia del trabajo de campo, restricciones presupuestales, cambios en la estructura de competencia, falta de confianza social para responder a una encuesta, ocultamiento de la intención de voto, el momento en que se levanta la encuesta, eventos imponderables que modifican el sentimiento del electorado, hasta la decisión voluntaria de publicar datos alterados con fines propagandísticos. En estricto rigor estadístico, la evaluación de una encuesta debiera basarse en qué tanto se desvió el estimador del resultado oficial, para cada uno de los participantes y concluir que, a menor error, mejor desempeño de una encuesta. ¿Cómo medir ese error?

### Cómo medir la precisión de las encuestas

La literatura fundacional sobre la materia proviene del estudio

patrocinado en 1949 por el Social Science Research Council (SRC) de Estados Unidos a Frederick Mosteller, Herbert Hyman, Philip J. McCarthy, Eli S. Marks y David B. Truman para evaluar la precisión de las encuestas electorales levantadas durante la campaña presidencial de 1948. Estos investigadores propusieron ocho métodos alternativos dos de los cuales siguen siendo, a la fecha, los más empleados para medir el error de las encuestas y generar comparaciones a lo largo del tiempo. Cada uno de esos métodos, conocidos como Mosteller 1 a 8<sup>2</sup>, evalúa aspectos distintos de la estimación ofrecida por las encuestas preelectorales. Aunque no se ha homogeneizado el uso de una sola de las métricas propuestas por Mosteller et.al., las dos más comúnmente usadas son la 3 y la 5, a saber:

Mosteller 3 o error promedio. Es la media aritmética de los errores en las estimaciones de una encuesta para cada partido. Es decir, la

---

2 Además de las dos métricas de Mosteller referidas en el cuerpo del texto, existen otras seis que se detallan a continuación: Mosteller 1. La diferencia en puntos porcentuales entre el voto efectivo total obtenido por el candidato o partido puntero en la encuesta y el obtenido en la jornada electoral;

Mosteller 2. La diferencia en puntos porcentuales entre el voto obtenido por cada candidato de los partidos más grandes en la encuesta y en la jornada electoral;

Mosteller 4. La diferencia promedio entre el número uno y la razón/proporción de cada candidato, donde la razón/proporción de cada candidato se define como la estimación de voto hecha por la encuesta para un candidato dividida entre el voto que éste obtiene en la jornada electoral.

Mosteller 6. La máxima diferencia entre el porcentaje de voto estimado para un partido y lo efectivamente obtenido.

Mosteller 7. La prueba de significancia para evaluar la congruencia de las distribuciones de voto estimado y real.

Mosteller 8. La diferencia entre el voto electoral pronosticado y el registrado, de acuerdo al sistema de colegio electoral empleado en los Estados Unidos.

---

1 / Las encuestas "preelectorales" son aquéllas que se realizan antes de la jornada electoral y se distinguen de las "de salida", que se realizan entre electores que ya emitieron su voto.

sumatoria de la diferencia absoluta del porcentaje del partido  $i$  estimado en la encuesta menos el porcentaje obtenido por el partido  $i$  en el resultado oficial y se lee como:  $\sum$  absoluta (% voto partido  $i$  encuesta - % voto partido  $i$  oficial... $n$ ) /  $n$ .

Mosteller 5 o error en la estimación de la ventaja. Es el valor absoluto de la diferencia entre el porcentaje de voto obtenido por el ganador y el segundo lugar en el resultado oficial menos la diferencia entre el porcentaje de voto estimado por la encuesta para ese ganador oficial y el segundo lugar oficial.

La mayoría de los estudios sobre encuestas en el mundo usa el error promedio (Mosteller 3) para calificar su precisión. Optar por este método alinea al evaluador con la idea de que lo que debe calificarse es qué tan bien estimó una encuesta la distribución del voto entre los partidos o candidatos participantes. Las evaluaciones más comunes en México calculan el error promedio y, si acaso, muestran en columna aparte la información de acierto, como dos datos desconectados.

2016 ofrece un buen ejemplo, no es la primera ocasión en que los resultados electorales distan de lo previsto por las encuestas, en el pasado hay debates registrados sobre el pobre desempeño de la industria demoscópica mexicana y las razones de ése. Desde 1992 la revista Nexos, y desde 1997 la

revista Voz y Voto, han convocado a diferentes encuestadores a debatir las razones por las que fallan las encuestas y, desde el año 2000, el Instituto Federal Electoral ha articulado seminarios de reflexión sobre la calidad de las encuestas mexicanas. Es interesante por ejemplo que, en esos foros y de manera gradualmente más abierta, se ha discutido el tema del uso propagandístico de las encuestas. Por supuesto que no es un fenómeno privativo de México, ocurre en todas partes pero lo cierto es que los esfuerzos por transparentar a las encuestas han sido infructuosos aquí, lo mismo para detectar elementos de manipulación propagandística de los datos que para medir su desempeño. Los encuestadores mexicanos no enfrentan el juicio de un evaluador autónomo y con autoridad para establecer una metodología estándar de evaluación. Los evaluadores de encuestas en México han sido siempre los encuestadores mismos y, en consecuencia, los debates sobre sus resultados han sido desordenados y escasamente rigurosos. Con frecuencia se asume que son los propios encuestadores quienes eluden el establecimiento de una metodología estándar de evaluación. Aquí consideramos relevante advertir al lector que uno de los dos autores de esta nota, Carlo Varela, es también un encuestador y que ha estado activo en la industria mexicana desde

1993. La base utilizada en esta nota incluye datos suyos publicados bajo la etiqueta de Varela y Asociados. Son todos los que cumplieron con los criterios definidos para integrar esta base.

A partir de 1994, algunos evaluadores comenzaron también a calcular el error en la estimación de la ventaja (Mosteller 5), para afinar la evaluación ya que, cuando todas las encuestas aciertan a ganador, su desempeño puede diferenciarse por la precisión con la que cada uno calculó la distancia entre el primer y el segundo lugar. Sin embargo, creemos que ambos métodos ofrecen visiones parciales del error de una encuesta. Veamos el siguiente ejemplo en el que se simulan datos de tres encuestas sobre una elección hipotética (cuadro 1). En la primera tabla se presentan primero los datos estimados para los tres partidos que, en el resultado oficial, obtuvieron 10% o más del voto efectivo (A, B y C). En la columna Otros se agruparon los datos agregados de los otros partidos participantes. En el último renglón se muestran los resultados de la elección. En la segunda tabla se presenta el cálculo de error promedio (Mosteller 3) y de error en la estimación de la ventaja (Mosteller 5), así como la información respecto a si cada encuesta acertó o no a ganador.

**Cuadro 1. Ejemplo hipotético sobre Mostelle 3, Mosteller 5 y acierto a ganador**

TABLA 1

	Partido A	Partido B	Partido C	Otros	Diferencia	Acierto
					primero - segundo	
Encuesta 1	39	36	10	15	-3	No
Encuesta 2	35	45	8	12	10	Sí
Encuesta 3	40	37	20	3	-3	No
Resultado oficial	33	36	14	17	3	

TABLA 2

	Error	Error	Error	Error	Error	Error estimación	Acertó
	Partido A	Partido B	Partido C	Otros	promedio (Mosteller 3)	de la ventaja (Mosteller 5)	
Encuesta 1	6	0	4	2	3	6	0
Encuesta 2	2	9	6	5	5.5	7	1
Encuesta 3	7	1	6	14	7	6	0

Como puede observarse, si nos quedáramos con el cálculo de Mosteller 3, deberíamos asumir que la encuesta 1 estuvo dentro de los márgenes teóricos de error aceptables y que fue la mejor. Sin embargo, se trata de una encuesta que equivocó el ganador. Esto es, por ejemplo, lo que pasó en las elecciones presidenciales de 2000 y 2006, cuando la mayoría de las encuestas equivocó ganador. Los encuestadores se autocalificaron bien porque sus márgenes de error no variaron tanto del resultado oficial pero el que no hubieran logrado predecir el ganador les generó descrédito entre la opinión pública.

Ahora bien, si sólo atendiéramos al dato de acierto, la mejor encuesta sería la 2, pero esa evaluación pasaría por alto que esta encuesta tuvo un alto error en la estimación de la distribución general de voto. Algo así sucedió en la elección presidencial de 2012, cuando todas las encuestas acertaron ganador

pero tuvieron uno de los peores récords de precisión.

Finalmente, si nos quedáramos únicamente con Mosteller 5, concluiríamos que las tres encuestas tuvieron un desempeño muy similar, lo cual es falso.

### ¿Cómo les fue a las encuestas preelectorales del 2016?

Los únicos dos analistas que publicaron evaluaciones de las encuestas preelectorales de 2016 fueron Alejandro Moreno y Lorena Becerra. Ambos midieron la magnitud del error de cada encuesta según Mosteller 3 y Mosteller 5 pero obtuvieron conclusiones distintas. Esto sucedió así debido a que optaron por definiciones metodológicas distintas para integrar sus bases de datos.

Moreno publicó dos evaluaciones, una primera en El Financiero<sup>3</sup> y una

segunda en la revista Campaigns & Elections<sup>4</sup>. En la de El Financiero sólo evaluó las 17 encuestas que él mismo publicó ahí a lo largo de los dos meses que duraron las campañas, para 11 de las 12 elecciones de gobernador que se llevaron a cabo en junio. En nueve casos publicó más de un dato porque ofreció también su filtro de votantes probables. Estado por estado, Moreno calculó el error promedio (Mosteller 3) de sus datos directos y de los que obtuvo después de filtrarlos con su modelo de votantes probables. Evaluó por separado a los partidos políticos que obtuvieron más de 10% de la votación en el correspondiente Programa de Resultados Electorales Preliminares (PREP) y al resto lo agrupó en Otros. Las conclusiones de esa evaluación son ilustradas en tablas por estado pero, en síntesis, Moreno reconoce que, entre sus encuestas preelectorales de 2016, "...hubo algunas con un

3/ Moreno, Alejandro:

<http://www.elfinanciero.com.mx/opinion/aciertos-y-fallas-de-las-encuestas.html>

4/ Moreno, Alejandro:

<http://www.campaignsandelections.com/single-post/2016/07/02/Encuestas-2016>

alto grado de precisión y otras fuera de los márgenes tolerables. Siete de las 17 encuestas (41%) tuvieron un error promedio menor o igual a tres puntos, pero cinco del total (24%) tuvieron un error promedio de más de seis puntos.” En este texto Moreno afirma que “de un total de 26 estimaciones publicadas (incluyendo las de votantes probables de las encuestas finales), los ejercicios de El Financiero acertaron ganador o previeron un escenario cerrado con error promedio menor o igual a tres puntos en 17 casos (65% del total), mientras que en el resto (35%) no se proyectó correctamente al ganador”. Sin embargo, de los 17 datos publicados como pronóstico en los últimos días de las campañas, sólo acertó a ganador en 16.7%, y su porcentaje de desacierto fue 83.3%.

En Campaigns & Elections, el mismo Moreno comparó sus datos con los de todas las encuestas publicadas que logró rastrear, y utilizó sólo los últimos números publicados por cada casa encuestadora, sin importar la fecha de levantamiento. En esta segunda evaluación, el encuestador no fue transparente respecto a los criterios con los que integró su base, por ejemplo, en lugar de utilizar el umbral mínimo de 10% del voto para medir a un partido por separado, se refirió a “las tres principales fuerzas de cada estado” y agregó a los demás en Otros. No explicitó cuál fue su umbral para determinar cuándo un partido es o no “fuerza principal” en cada estado. En esta segunda evaluación, Moreno ordenó primero a las encuestas estado por estado según su margen de error promedio y su error en la estimación de la ventaja y luego, en tabla aparte, reportó si las encuestadoras evaluadas acertaron o no a ganador. Como ya se expuso, Mosteller 3 y Mosteller 5 constituyen mediciones distintas, por lo cual el ordenamiento de las casas encuestadoras es distinto en

cada caso. En ese texto, Moreno afirma que hasta antes de 2016, el 2010 había sido el año de peor desempeño de las encuestas y, luego, presenta el cálculo de Mosteller 3 para 2016, por casa encuestadora para cada estado. Sin haber ofrecido cifras generales, Moreno concluye que 2016 ha pasado a ser el peor año de las encuestas preelectorales en México -aunque a las suyas no les fue tan mal, al menos en términos de error distributivo.

Lorena Becerra publicó en Reforma<sup>5</sup> una evaluación de encuestas circunscritas a los estados en los que Reforma hizo levantamientos en 2016. Ella se compara con las que, a su juicio, fueron “las otras principales casas encuestadoras”, pero no argumenta los criterios de ese juicio. Notamos que los datos incluidos en su base no corresponden todos al mismo período, para algunos estados incluye información publicada en abril y mayo, para otros sólo menciona datos de mayo. Esta investigadora también usa Mosteller 3 y Mosteller 5 para determinar qué encuestas incurrieron en mayor error distributivo en 2016. Y reporta luego, también por separado, quiénes acertaron a ganador y quiénes no. Becerra no ofrece mayor análisis de los cálculos de error que obtiene. En cualquier caso, esta publicación no abona análisis para comprender el desempeño de la industria general de encuestas en 2016.

Ni Moreno ni Becerra realizaron un balance general de las encuestas publicadas en 2016, dos de los tres estudios son parciales ya que solo analizan sus propias encuestas y no ejercen un criterio claro en cuanto a fechas de levantamiento de las encuestas evaluadas, por lo que no

5/ Becerra, Lorena: “Encuestas: Las lecciones del 2016” en Periódico Reforma, Revista R, pp 8-10. 12 de junio del 2016.

se sabe si el error medido es causado por la distancia en el tiempo o por la mala calidad de la encuesta. Tampoco fueron claros respecto a cómo seleccionaron los partidos que evaluaron por separado, pero la diferencia en los resultados evidencia que cada evaluación se hizo con una selección diferente. En suma, no es posible comparar las únicas tres evaluaciones de encuestas publicadas en 2016.

### Evaluación del desempeño histórico de las encuestas en México

Para discernir sobre la evolución del desempeño de encuestas en el país, es indispensable primero establecer qué se calificará de los datos de una encuesta: que se aproxime a las porciones del voto obtenidas por los participantes, que logre establecer la ventaja con que el ganador se impondrá sobre el segundo lugar, que acierte al ganador o todas las anteriores.

Luego debieran acordarse al menos dos criterios metodológicos: 1. qué umbral de voto debe alcanzar un partido o candidato para ser evaluado de forma independiente y 2. cuál es el período de levantamiento que debe observarse. Para que las evaluaciones resulten comparables en el tiempo y sean replicables, debiera emplearse un mismo método.

A partir de nuestra revisión de los métodos existentes a continuación ofrecemos una evaluación comparativa, por año, de las preelectorales publicadas en México desde 1994. La intención es determinar cómo les fue a las encuestas del 2016 en relación con la trayectoria histórica de la industria. Construimos una base de datos (Anexo 1) que reúne 240 encuestas preelectorales para elecciones de Presidente, gobernador, diputados federales y asambleístas de la Ciudad de



México debido a que sólo esas metodologías comparables<sup>6</sup>. Únicamente incluimos encuestas levantadas máximo 16 días antes de la elección con la intención de capturar la mayor aproximación posible al evento. Pudimos haber usado 'el último dato publicado' por casa encuestadora para atenernos a un principio de parsimonia pero, dado que algunas casas publican un único dato temprano durante las campañas, ello nos impedía controlar por el momento del levantamiento, por lo que no sabríamos si el margen de error es atribuible a problemas metodológicos de la encuesta o a que ésta se levantó mucho antes del evento<sup>7</sup>. Es posible que haya datos de encuestadoras locales que

no estén ahí debido a que construimos la base únicamente con la información disponible en medios nacionales y en la internet. No todos los datos incluidos corresponden a encuestadoras conocidas y, de hecho, hay registros de origen desconocido que probablemente tuvieron un propósito propagandístico. En cualquier caso, decidimos incluir todos los registros que reunieron los criterios señalados.

Para calcular el tamaño del error según Mosteller 3 y 5 revisamos por separado únicamente los datos de los partidos/candidatos que obtuvieron al menos 10% del voto efectivo en la elección. Esto último con la finalidad de no diluir el error

de estimación para los partidos grandes con el de los pequeños.

Generalmente, el porcentaje de voto que obtienen las organizaciones políticas pequeñas presentan una menor variación en sus estimaciones, en consecuencia, si se los incorpora individualmente al promedio tienden a reducir el cálculo del error.

El cuadro 2 muestra el número de encuestas que reunieron las características mencionadas, por año, así como los cálculos de error promedio (Mosteller 3) y error en la estimación de la ventaja (Mosteller 5), por año para toda la industria.

**Cuadro 2. Error promedio y error de ventaja de encuestas preelectorales 1994 a 2016**

Año	Encuestas publicadas	Error Promedio	Error Ventaja	Porcentaje Acierto	Mediana	Rango	Desvest
1994	2	2.5	5.0	100	2.5	1.4	1.0
1997	24	3.3	6.7	92	2.8	8.8	2.1
1998	5	2.7	5.2	80	2.3	3.9	1.5
1999	8	2.9	5.8	75	2.6	5.4	1.8
2000	14	2.9	7.5	50	3.0	3.5	1.1
2001	10	6.0	5.6	90	5.0	12.8	4.4
2003	8	2.6	4.7	88	2.4	4.0	1.5
2004	3	4.5	8.6	100	4.8	3.4	1.7
2005	12	3.3	7.9	92	2.7	7.2	2.0
2006	15	2.5	3.8	53	1.7	6.4	1.9
2007	2	1.4	2.7	100	1.4	1.4	1.0
2009	14	3.5	8.6	86	3.4	5.4	1.6
2010	29	5.3	11.7	62	5.1	9.2	2.3
2011	12	2.7	6.9	92	2.9	2.2	0.9
2012	24	3.2	9.3	92	2.9	4.2	2.0
2015	30	4.7	10.6	60	3.9	13.6	3.3
2016	28	3.5	6.9	39	3.0	8.8	1.8
<b>Promedio</b>	<b>14</b>	<b>3.7</b>	<b>7.9</b>	<b>72</b>	<b>3.1</b>	<b>6.0</b>	<b>1.9</b>

Fuente: Banamex con datos de su archivo interno

6/ Fueron levantadas en vivienda, cara a cara, con un tamaño de muestra de entre 1000 y 1500 cuestionarios y publicaron datos para todos los partidos participantes

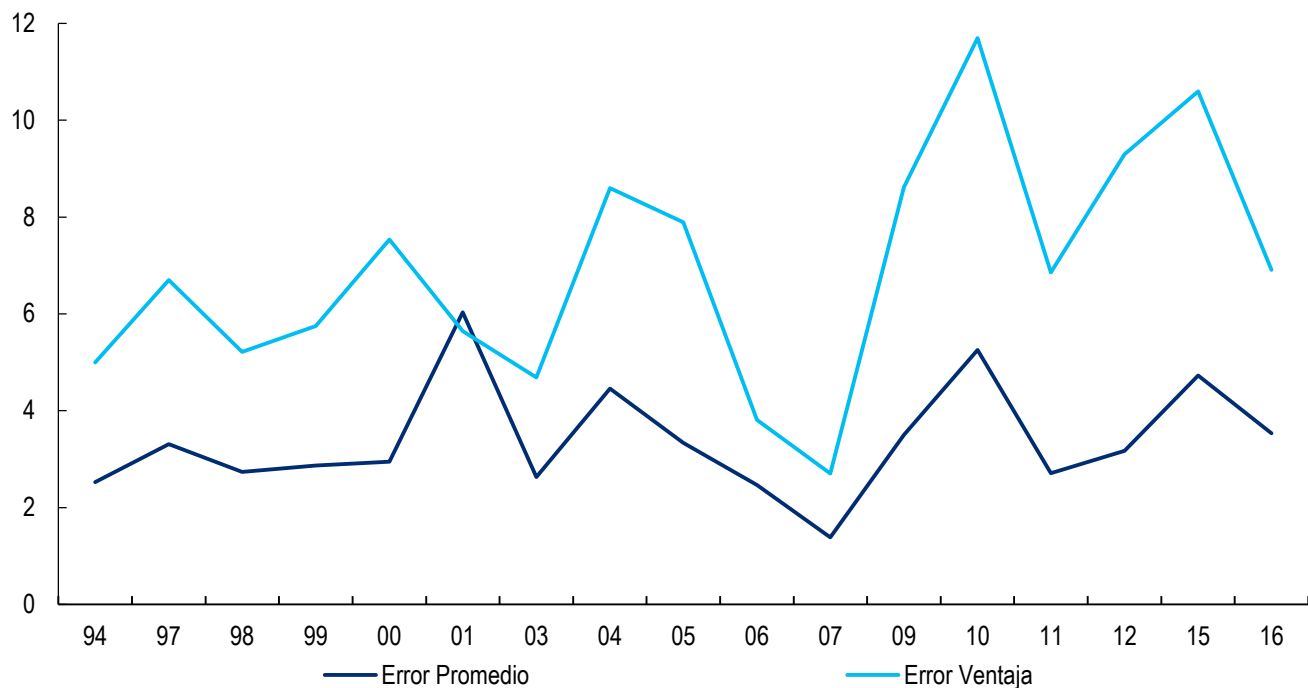
7/ Varela, Carlo: Factores muestrales y no muestrales que alteran el margen de error de una encuesta electoral en su pronóstico: una primera aproximación vía empírica, tesis de licenciatura, ITAM, México, D.F., 2010.

Antes del año 2009 el número de encuestas preelectorales era bajo e irregular. La excepción fue 1997 porque la primera elección para Jefe de Gobierno del entonces aún Distrito Federal marcó un hito y fue abundantemente medida por las encuestas. Después, sólo los años de elecciones presidenciales registran más interés de las encuestas y, a partir del 2010, hay

un incremento sustantivo. La calendarización de las elecciones no tiene una distribución proporcionada a lo largo del sexenio, cada estado sigue una dinámica propia, por lo que hay años sin elecciones para gobernador (1996, 2002, 2008), otros con pocas y algunos con muchas. También existen años en los que concurren elecciones

presidenciales con legislativas federales y múltiples locales (1994, 2000, 2006 y 2012) y otros en los que la concurrencia sólo involucra diputados federales con algunas locales (1997, 2003, 2009 y 2015).

**Gráfica 1. Encuestas 1994 a 2016: error promedio y error de estimación de la ventaja**



Fuente: Banamex con datos de su archivo interno

De la gráfica 1 destacan los siguientes resultados, que se refieren de manera agregada a toda la industria:

- a) Por margen de error (Mosteller 3), el peor año fue 2001 con 6.0%, en segundo lugar 2010 con 5.3% y en tercero 2015 con 4.7%. El mejor año fue 2007 con 1.4% y en segundo lugar 1994 y 2006 con 2.5% ambos. De acuerdo a este indicador, 2016 fue el treceavo peor año de los 17 estudiados.
- b) Por error en la estimación de la ventaja (Mosteller 5), el mejor año fue nuevamente 2007 con 2.7%, luego 2006 con 3.8% y 2003 con 4.7%. El peor año fue 2010 con 11.7%, el segundo peor año fue 2015 con 10.6% y en tercer lugar 2004 y 2009 con 8.6% ambos. Según Mosteller 5, 2016 fue el noveno peor año con 6.9%.
- c) Sólo ocho años (menos de la mitad de las observaciones) han estado por debajo de 3% de error promedio, es decir, por lo general, la precisión de las encuestas es mayor al margen de error teórico que presentan en sus metodologías.
- d) En lo que los análisis no se han equivocado es en señalar que 2016 ha sido el peor año en cuanto a acierto a ganador, solo 39% de las encuestas publicadas predijo correctamente al ganador en los últimos 16 días anteriores a la jornada electoral.
- e) En el cuadro II podemos observar otras variables de

tendencia central como la mediana, el rango y la desviación estándar, que nos permiten identificar la existencia de datos atípicos en algunos años. Hay tres datos que presentan una mediana muy por encima de la media, 2001, 2010 y 2011. Y hay dos con un rango considerablemente más amplio que el del resto, 2001 y 2015. Evidentemente, el 2001 fue un año atípico, lo cual probablemente se deba a que sólo hubo levantamientos en dos estados, Michoacán y Tabasco.

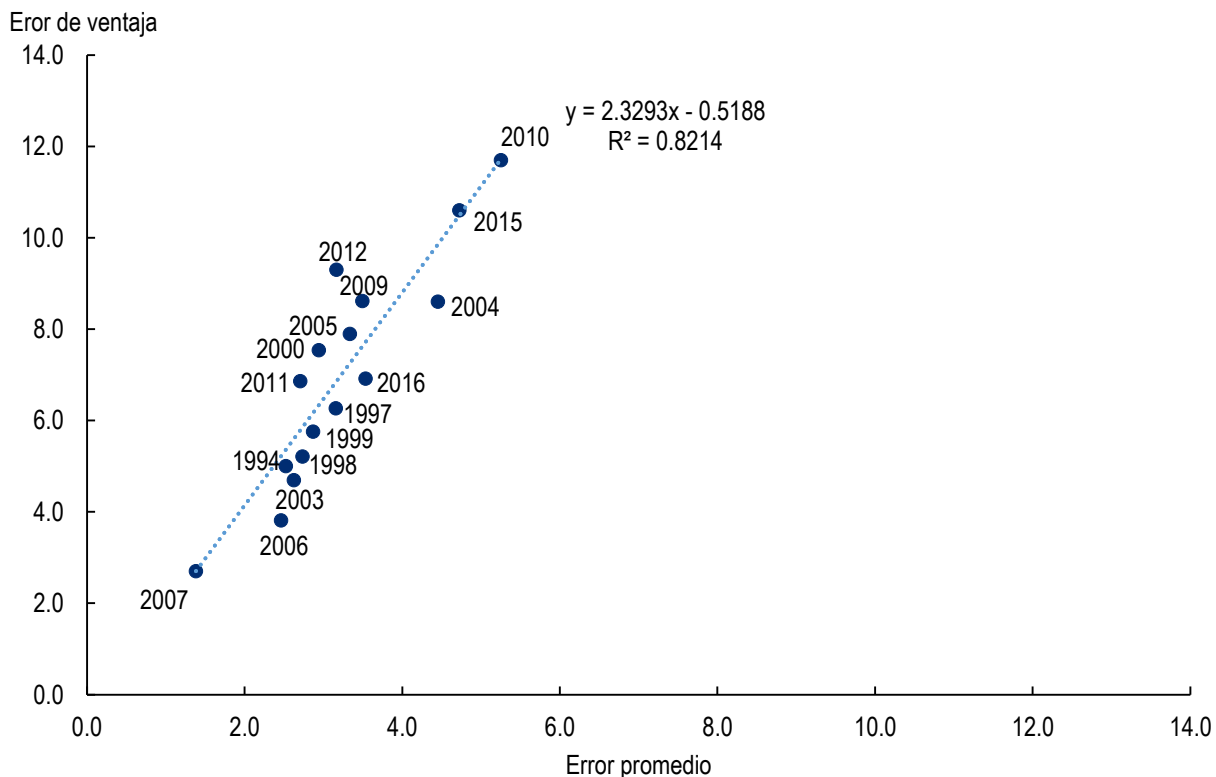
2016, en contraste, no aparece como un mal año para la industria de las encuestas según estas otras variables.

Ahora, al graficar Mosteller 3 y Mosteller 5 (gráfica 2) resulta difícil optar por uno de ellos como un mejor indicador del desempeño general de las encuestas. El rango de Mosteller 3 es menor que el de Mosteller 5. Si el dato está por encima de la línea de regresión, significa que las encuestas de ese año tuvieron un peor desempeño

según Mosteller 5 que según Mosteller 3 si, por el contrario, el dato está por debajo de la línea de regresión, significa que Mosteller 3 fue mayor a Mosteller 5.

El año 2001 es atípico, se trata del mayor Mosteller 3 de toda la serie, aunque su Mosteller 5 sea bastante aceptable. 2010 y 2015 son, observablemente, los peores años de la industria. Y 2016 se ubica "a media tabla".

**Gráfica 2. Encuestas 1994 a 2016: error promedio y error de estimación de la ventaja**



Fuente: Banamex con datos de su archivo interno

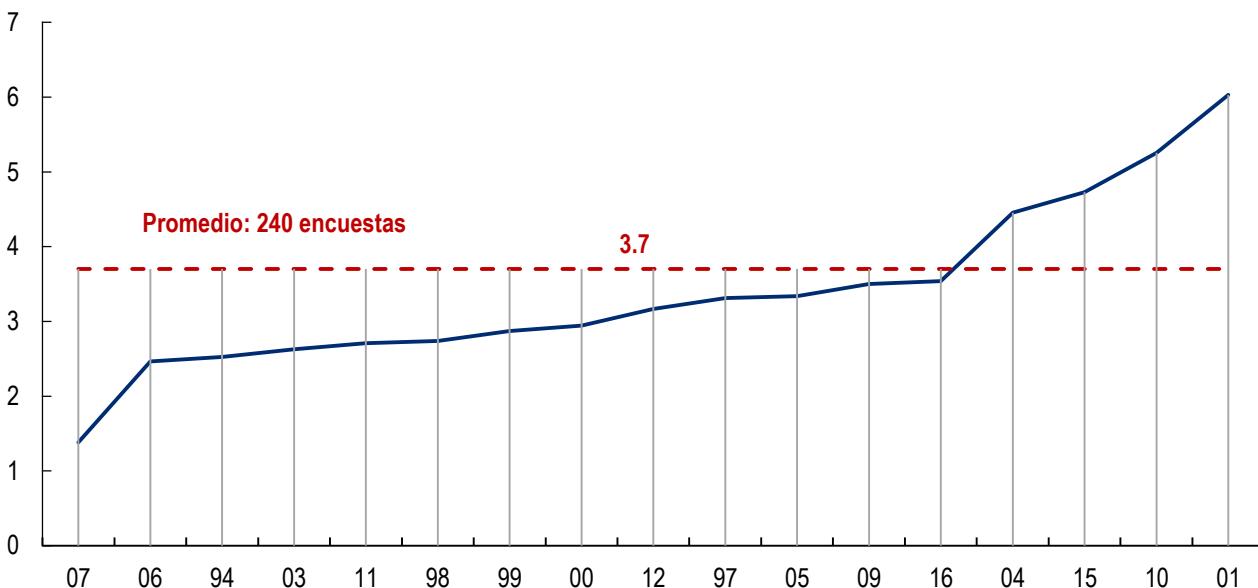
Si graficamos sólo Mosteller 3 (Gráfica 3) la serie se divide en tres grupos: en el primero está sólo el año 2007; en el segundo grupo hay

12 años cuyo promedio de error es 3% (las mediciones van de 2.5 a 3.5) y en el tercer grupo hay cuatro

años con un error promedio de más de 5% (2004, 2015, 2010 y 2011).



Gráfica 3. Error promedio por año



Fuente: Banamex con datos de su archivo interno.

### Desempeño por casa encuestadora

Para obtener información desagregada por casa encuestadora agrupamos sacamos el promedio de error por casa encuestadora (Cuadro III). Hay 24 empresas que han publicado por lo menos tres encuestas preelectorales que cumplen con los requisitos que definimos para integrar la base. Adicionalmente, incluimos al consenso o mediana de las encuestas, por tipo de elección, donde por lo menos tres encuestadoras hubieran publicado datos, esto es en 34 diferentes tipos

de elección. Es relevante aclarar que, a lo largo del período que aquí estudiamos, el periódico Reforma ha tenido tres investigadores responsables de sus encuestas: Rafael Gímenez (RG) de 1994 a 1999; Alejandro Moreno (AM) de 1999 a 2015 y Lorena Becerra (LB) desde el 2016. AM es ahora, desde el 2016, el encuestador de El Financiero. En El Universal, Alduncin y Asociados se encargó de todas las encuestas publicadas antes del 2000, desde ese año y hasta el 2015 Carlos Ordoñez (CO) fue el investigador encargado del departamento de opinión pública y, a partir del 2016, no se ha revelado

quién es el responsable de las encuestas que el periódico publica, por ello únicamente denominamos a este período como Nueva Época (NE). Beltrán, Juárez y Asociados es la empresa encargada de los levantamientos del Excelsior/BGC. Demotecnia/MH con María de las Heras fue la empresa encargada de publicar encuestas en Milenio Diario del 2000 al 2011; después de la muerte de De las Heras, su hijo Rodrigo Galván se hizo cargo de la empresa Demotecnia/RG y durante 2016 publicó encuestas para el Sol de México.

**Cuadro 3. Error promedio y error de ventaja de encuestas preelectorales 1994 a 2016**

Empresa	Encuestas publicadas	Error Promedio	Error Ventaja	Porcentaje Acierto
Consultores y Marketing Político	4	1.8	3.9	75
Berumen	4	2.4	6.2	100
El Universal NE	4	2.5	5.0	50
Reforma RG	12	2.6	5.7	83
Excelsior/BCG	5	2.8	4.2	60
El Universal CO	17	3.0	6.8	82
Reforma LB	3	3.0	7.3	67
Arcop	5	3.0	6.4	80
Consenso	34 *	3.1	6.6	71
Reforma AM	29	3.1	6.6	66
Alduncin y Asociados	6	3.1	3.4	100
Indermerc Louis Harris	3	3.3	9.1	100
Parametría	11	3.3	8.3	64
El Financiero AM	6	3.4	8.4	17
Espacio Muestral	6	3.4	8.5	50
CEO	5	3.6	10.1	60
Demotecnia/MH	18	3.7	6.3	89
Buendía	3	3.7	10.8	100
Covarrubias	4	3.9	10.2	75
Varela y Asociados	6	3.9	6.2	50
Consulta	15	4.2	9.0	80
GEA-ISA	8	4.5	11.0	75
GCE	14	5.0	9.7	86
MEBA	5	5.4	9.0	60
Demotecnia/RG	5	5.4	10.6	20

Fuente: Banamex con datos de su archivo interno

Reforma AM es la empresa que más encuestas ha publicado seguida de Milenio con Demotecnia (18) y El Universal CO (17). Las casas encuestadoras con el Mosteller 3 más bajo fueron: Consultores y Marketing Político, Berumen y El Universal NE, todos ellos han estado en promedio por debajo del margen de error teórico.

### Un nuevo índice para evaluar encuestas

La enorme mayoría de los encuestadores sostiene aún que el criterio para evaluar sus datos debiera ser qué tanto se acercan a la distribución final de voto producida en una elección, sin importar si aciertan o no al ganador.

Afirman que, en la medida que una elección sea más competida, se incrementará la probabilidad de que sus datos fallen el ganador, sin que ello deba descalificarlos. Es un argumento que escuchamos hace un año en Gran Bretaña, hace algunas semanas en el referéndum sobre Brexit, en otras elecciones europeas recientes y, con mucha frecuencia, en México. No obstante, el prestigio de cada casa encuestadora se relaciona con su desempeño predictivo y no sólo con su trayectoria de estimaciones distributivas. La manera en que los medios publican datos de encuestas es ilustrativa de esto<sup>8</sup>. En principio,

un bajo desempeño predictivo debería retirarle prestigio y clientela a una casa encuestadora. Por supuesto, el problema de las encuestas propagandísticas y la lógica detrás de éstas explica que el principio anterior se modifique a veces. Como fuere, las encuestas son consumidas por el público en general, y por el elector en particular, como un “atajo informativo” para saber quién ganará. Y, en la medida que esa expectativa se cumple, las encuestas se convierten en constructoras de confianza democrática. Cuando las encuestas fallan por márgenes estrechos, este

<sup>8</sup> En el 2016, los medios que publicaron datos de encuestas optaron por cabecearlos con el nombre del

partido o candidato “ganador” y no con la distribución de voto estimada.

mecanismo no se altera pero, cuando fallan notoriamente, se convierten en raíz de desconfianza social. Basados en lo anterior decidimos insistir en que la capacidad de una encuesta para acertar al ganador sí importa y que debe constituir parte de la medición de desempeño, tanto como la precisión.

Buena parte de las encuestas que se publican incorporan a sus datos ponderaciones y/o consideraciones derivadas de la experiencia histórica que el encuestador tiene, su conocimiento específico del electorado y los candidatos en juego, de los elementos institucionales y/o políticos que pudieran generar sesgos en las respuestas de los encuestados, en fin. Lo hacen con la intención de afinar la precisión de sus datos y, como resultado, aplican un “tratamiento” a los datos estadísticos que busca extraer de ellos algo de un pronóstico. Por supuesto, no afirmamos aquí que las encuestas deban considerarse pronósticos en sí mismas. Entendemos que los modelos predictivos son construcciones más complejas que una encuesta, que buscan determinar y medir los principales factores de cambio electoral, para emplearlos como base de una predicción explícita, previa a la elección.

Existen distintos modelos básicos para generar pronósticos electorales: 1. los basados en mercados de apuestas sobre el resultado electoral; 2. los de modelos indizados, basados en las características de los candidatos y los temas de campaña; 3. los generados por expertos basados en su experiencia e intuición; 4. los de expectativa ciudadana, basados en encuestas sobre quién cree la gente que ganará, no por quién votarán; y 5. los basados directamente en encuestas de intención de voto. En otros países, además de estos modelos básicos, se han construido

otros que combinan enfoques en un intento por incrementar su confianza en el pronóstico. La mayor parte de los modelos combinados incorpora información de encuestas. Se trata, en efecto, de su insumo más dinámico de información<sup>9</sup>. Quienes construyen modelos predictivos con formalidad generalmente explicitan su metodología, parámetros y ponderaciones. La elección presidencial actualmente en curso en Estados Unidos es un buen ejemplo de esto. Cada día se produce una enorme cantidad de información (tanto de encuestas como otra) que alimenta diversos modelos predictivos, publicados y actualizados en tiempo real<sup>10</sup>. La cantidad de información que estos modelos sintetizan y el hecho de que sean publicaciones de acceso no restringido permiten al electorado norteamericano y a la población en general, formarse una opinión respecto al proceso electoral. Y, sobre todo, les permite generar una expectativa de cuál es el resultado más probable.

A falta de modelos predictivos semejantes, en México tenemos encuestas. Y dado que no existe regulación de aplicación general, la cantidad y calidad de la información metodológica que los encuestadores revelan tiende a ser escasa. Por supuesto, es posible que algunas de las encuestas preelectorales que se dan a conocer publiquen estrictamente los datos de distribución que obtienen pero también sabemos que hay algunas otras que buscan generar pronósticos. Con ese propósito, las últimas revisan y ponderan sus estimaciones a la luz de la experiencia del encuestador y del análisis de factores alternos como los que emplean algunos modelos de predicción (condiciones económicas, popularidad del político

en funciones, temas de la campaña y otros acontecimientos de impacto idiosincrático). En consecuencia, si bien no puede afirmarse que todas las encuestas pretendan obtener pronósticos, parece razonable asumir que algunas lo hacen y, sobre todo, aquellas que se publican ya en el último tramo de las campañas. Como Romero y Varela<sup>11</sup> afirman, “sólo si la coincidencia de encuestas y resultados electorales fuese aleatoria, si el comportamiento de las encuestas fuese sistemáticamente independiente del resultado electoral, podríamos concluir que las encuestas preelectorales no se emplean para construir pronósticos; sin embargo, la evidencia estadística sugiere que las últimas encuestas preelectorales publicadas tienden a acertar el resultado”. En consecuencia, las encuestas pueden estimar una distribución y, además, ser insumo central de un pronóstico. En ambos propósitos pueden tener un desempeño satisfactorio o equívoco. Dado que no existe un método para evaluar el error de una encuesta simultáneamente en más de una dimensión, en la última parte de esta nota proponemos uno.

### Índice General de Error (IGE)

En nuestro ejercicio hipotético, mientras las encuestas 1 y 2 (Cuadro I) fueron más precisas que la 3, sólo la encuesta 2 acertó a ganador. ¿Es factible resumir esa conclusión en un solo dato? Para intentar lo anterior utilizamos dos variables: a) el error promedio y b) el porcentaje de desacierto de las encuestas por año. El error promedio califica la precisión del escenario de distribución del voto; el desacierto es una calificación de la predicción de ganador ofrecida por las encuestas. A partir de la

9 Mann, Thomas: “What do the models say about who will win in November?” en FixGov, septiembre 15, 2016, Brookings Institution, sitio electrónico.

10 Idem.

11 Romero, Vidal y Carlo Varela: “La precisión de las encuestas preelectorales” en Revista de Estudios Jurídico Electorales, México, D.F. pp 30-37 Año2, No. 0 Mayo-Julio 2011.

normalización de esas variables construimos un índice.

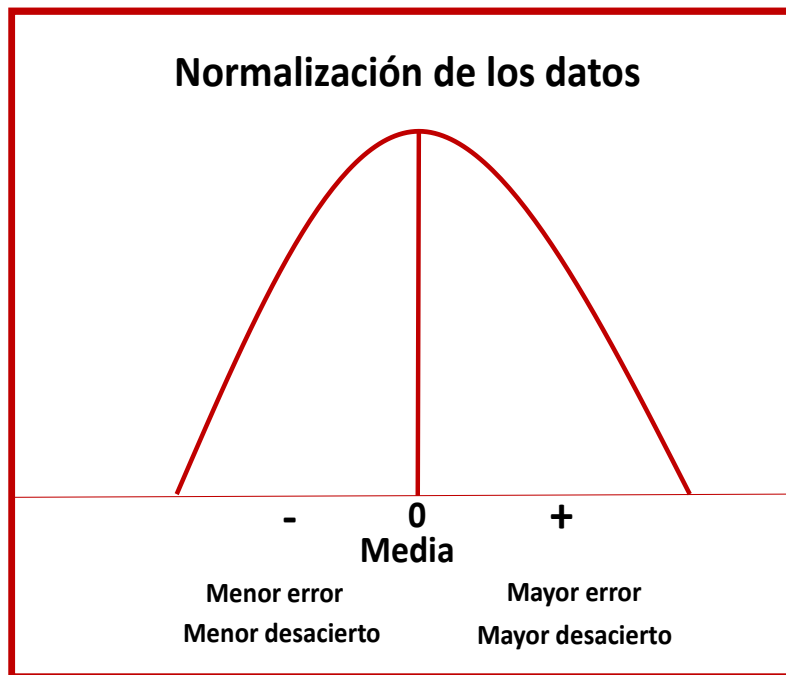
Somos conscientes de que, en términos estadísticos, la precisión y el acierto son dimensiones distintas; para cada encuesta, la primera se expresa en valores nominales que tenderán a exhibir una varianza discreta mientras la segunda dimensión se expresa en valores binarios de 0 y 1 que, si acaso, pueden convertirse en porcentajes y, al hacerlo, se ubicarán en los extremos de la escala de 0 a 100%

para un año determinado o para cada una de las casas encuestadoras a evaluar. Esa diferencia de proporcionalidad supondría, en rigor estadístico, que se trata de dimensiones no comparables. Sin embargo, con el propósito de obtener una evaluación que sintetice en un solo dato la calidad de la encuesta en tanto ejercicio estadístico y pronóstico de un evento, nos permitimos la que asumimos como una licencia metodológica para intervenir los valores de ambas dimensiones de

modo tal que se logra acercarlos y volverlos comparables.

Para volver comparables los dos componentes: error promedio y porcentaje de desacierto proponemos normalizar o estandarizar las dos variables. Una vez estandarizados, los valores obtenidos se ubican dentro de una dimensión nominal que permite promediarlos (Diagrama 1).

Diagrama 1. Estandarización de las variables de error promedio y desacierto



Al normalizar las variables los datos van de negativo, si están por debajo del promedio, a positivo si están por arriba del promedio. Mientras más negativo sea el valor significa que el error de las encuestas es menor y, cuanto más positivo, el error será mayor. Lo mismo para el desacierto,

mientras más negativo sea el dato significa que el grado de desacierto es menor y cuanto más positivo sea el dato, mayor será el grado de desacierto (Diagrama 1). Por último, asignamos el mismo peso relativo al error distributivo que al predictivo.

La ventaja del indicador que se obtiene es que incorpora, en un solo dato, el error promedio y, por primera vez, el acierto (o más bien Desacierto a ganador dado que, por congruencia conceptual, era necesario convertir esa información en un formato de error).

**Cuadro 4. Índice general de error por componentes y año**

Año	Encuestas publicadas	Error Promedio	Porcentaje Desacierto	Estandarización		Índice General
				Promedio	Desacierto	
1994	2	2.5	0	-0.8	-1.1	-0.9
1997	24	3.3	8	-0.1	-0.6	-0.4
1998	5	2.7	20	-0.6	0.0	-0.3
1999	8	2.9	25	-0.5	0.2	-0.1
2000	14	2.9	50	-0.4	1.5	0.6
2001	10	6.0	10	2.3	-0.6	0.9
2003	8	2.6	13	-0.7	-0.4	-0.5
2004	3	4.5	0	0.9	-1.1	-0.1
2005	12	3.3	8	0.0	-0.6	-0.3
2006	15	2.5	47	-0.8	1.4	0.3
2007	2	1.4	0	-1.7	-1.1	-1.4
2009	14	3.5	14	0.1	-0.3	-0.1
2010	29	5.3	38	1.6	0.9	1.3
2011	12	2.7	8	-0.6	-0.6	-0.6
2012	24	3.2	8	-0.2	-0.6	-0.4
2015	30	4.7	40	1.2	1.0	1.1
2016	28	3.5	61	0.1	2.1	1.1

Fuente: Banamex con datos de su archivo interno

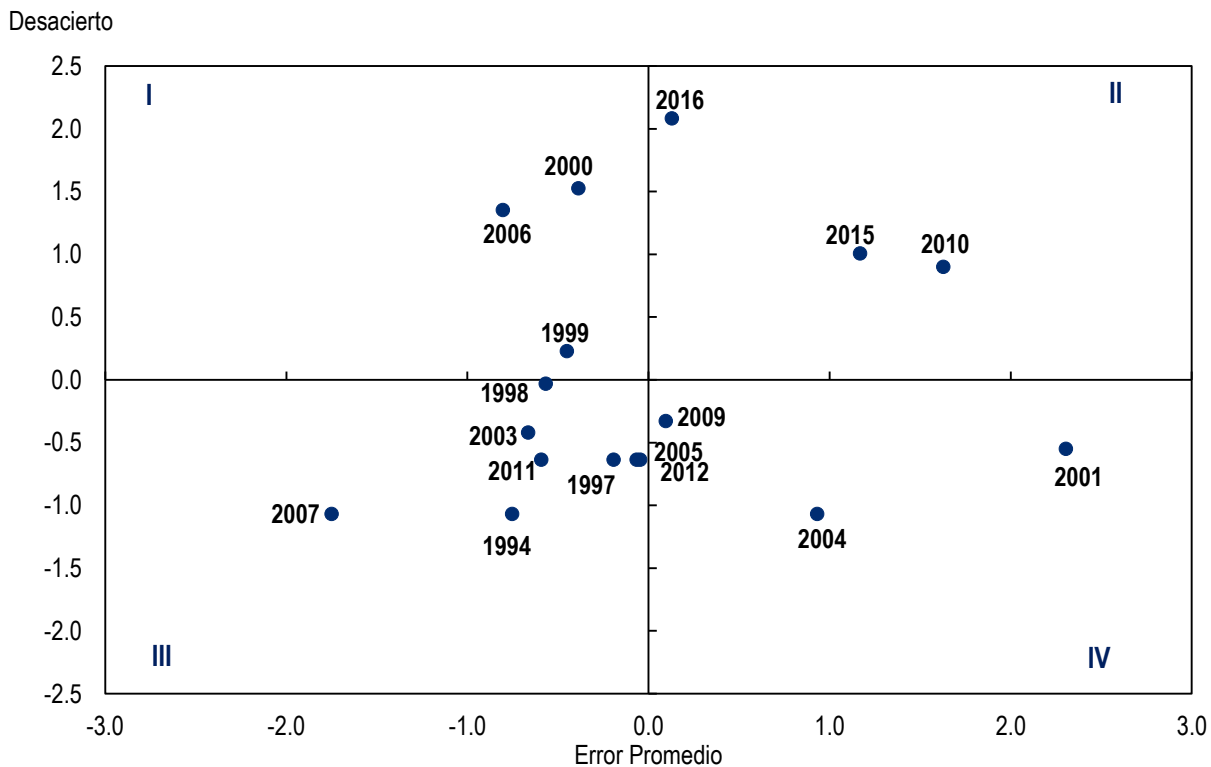
Como puede observarse en cuadro 4, si juzgamos por el porcentaje de desacierto, 2016 sí ha sido el peor año, por amplio margen, sólo 39% de las encuestas publicadas predijo correctamente al ganador. En el pasado, los peores años habían sido 2006 con 53% y 2000 con 50%, en ambos casos ese error se concentró en las elecciones

presidenciales. En 1994, 2004 y 2007 100% de las encuestas consideradas en nuestra base acertaron a ganador. Al graficar en dos dimensiones los datos estandarizados de error promedio y porcentaje de desacierto, agregados por año (Gráfica 4), los datos del cuadrante III son los más precisos y certeros, son 1994, 1997,

2003, 2007 y 2011; los años con el peor desempeño son los ubicados en el cuadrante II, 2010, 2015 y 2016. En el cuadrante I están los años más precisos que certeros, 1998, 1999, 2000 y 2006. Y en el cuadrante IV se ubican los años más certeros que precisos, 2001, 2004, 2005, 2009 y 2012.



Gráfica 4. Estandarización del error compuesto y del índice general de error



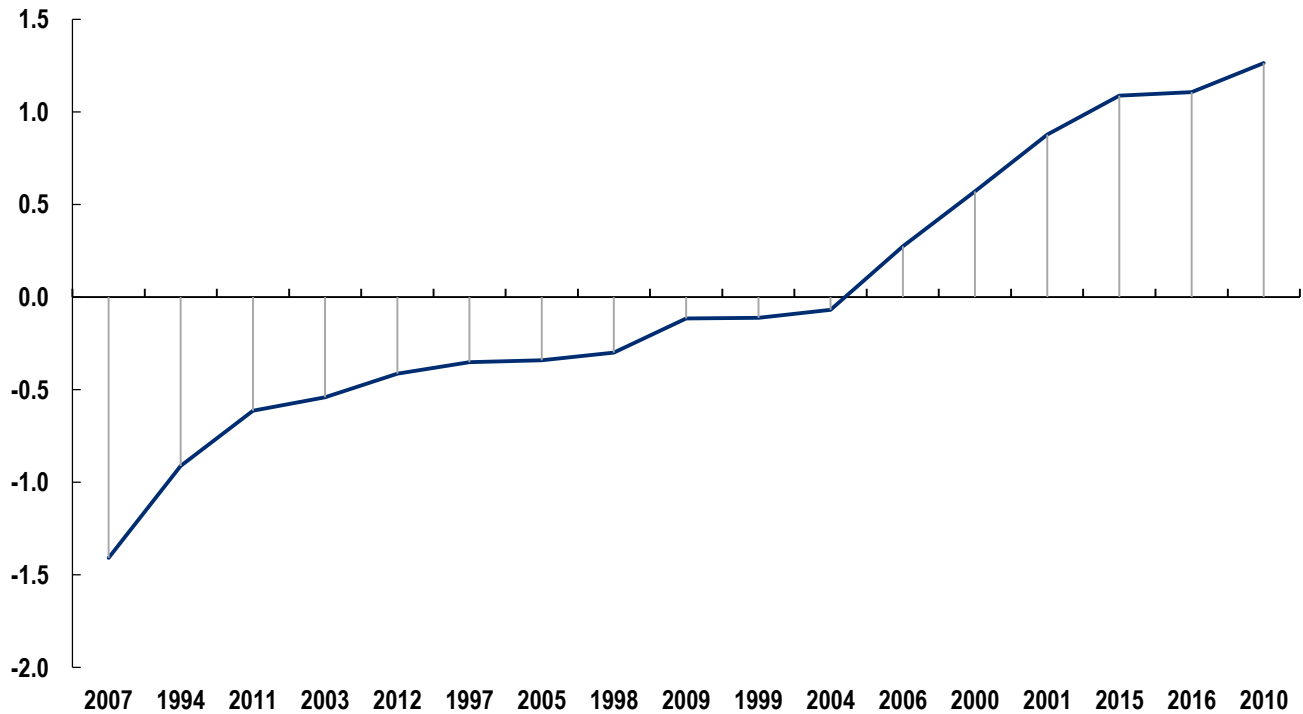
Fuente: Banamex con datos de su archivo interno

En términos generales, los primeros años de esta serie de encuestas tuvieron un mejor desempeño que los últimos: 1994 y 1997 se ubican en el cuadrante 1 mientras que las encuestas de los últimos años noventa y los primeros del siglo XXI han sido mejores para estimar la distribución del voto que para acertar a ganador: 1998, 1999, 2000 y 2006. Resulta interesante que las encuestas de los años en los que se eligieron ejecutivos federales de alternancia, 2000 y 2006, fueron muy precisas en su estimación distributiva pero fallaron en su predicción de ganador, la mayoría de ellas sugirió un triunfo de Labastida, no de Fox y pronosticaron que ganaría AMLO,

no Calderón. En contraste, a mediados del decenio pasado varias encuestas parecen haber comenzado a tener problemas de precisión aunque su record de acierto haya sido bueno, 2001, 2004, 2005, 2009 y 2012; destacan los últimos dos porque la mayoría de las encuestas acertó a ganador, pero su error promedio se disparó por arriba del promedio. Por último, es claro que 2010, 2015 y 2016 han sido años de mal desempeño general, pese a que se han levantado muchas encuestas (particularmente en 2010). Ello explica que la percepción sobre el desempeño de las encuestas se haya deteriorado sustantivamente en años recientes.

La Gráfica 5 ilustra con elocuencia la simplificación implícita en el Índice General de Error, que nos permite pasar de una evaluación multidimensional difícil de sintetizar a otra unidimensional, que facilita conclusiones ordinales. Los mejores años fueron 2007 y 1994, cuyas encuestas tuvieron un desempeño muy por encima del de las otras de la serie, aunque sólo estudiamos dos encuestas en cada uno de esos años. En contraste, los peores años para la industria han sido 2016, 2015 y 2010 no han sido ni precisas ni certeras. La combinación de estos factores en el índice permite tener una mayor claridad en la evaluación.

Gráfica 5. Índice general de error por año



Fuente: Banamex con datos de su archivo interno

**Desempeño por casa encuestadora**

Finalmente, a partir de la nueva medición de error aquí propuesta, en la Gráfica 6 presentamos el cálculo por casa encuestadora, con registros que van de 1997 al 2016 para las 22 casas encuestadoras y el consenso que publicaron más de tres encuestas<sup>12/</sup>. En general puede observarse que hay una concentración de datos en las tres casas encuestadoras que publicaban en periódicos, dos de las cuales eran los

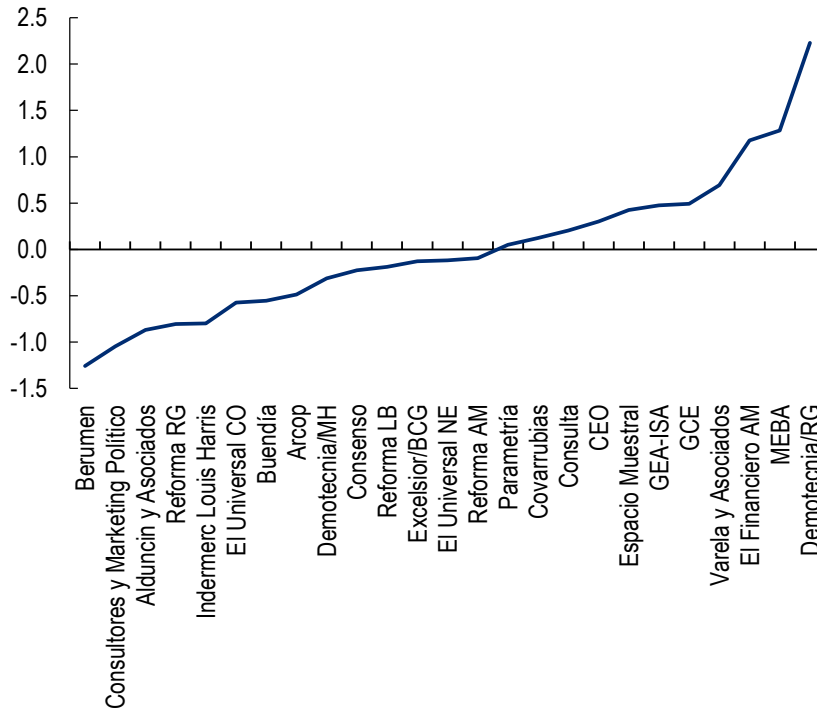
departamentos de investigación de los propios medios: Reforma, bajo la dirección de Alejandro Moreno, El Universal con Carlos Ordoñez y Milenio con Demotecnia, bajo la dirección de María de las Heras. Entre ellos tres concentraron 25% de las encuestas publicadas.

A lo largo de la serie, Berumen resulta la encuestadora que más ha combinado precisión con acierto; es seguida por Consultores y Marketing Político y en tercer lugar Alduncin y Asociados. En el otro extremo las encuestadoras con el peor

desempeño general han sido Demotecnia bajo la dirección de Rodrigo Galván, Mendoza Blanco y Asociados que ha sido impreciso aunque medianamente certero y El Financiero con Alejandro Moreno que, aunque publicó datos bastante precisos este 2016, tuvo de los peores records de acierto.

12/ Hubo 38 encuestadoras que publicaron una o dos encuestas. Si agrupamos a las empresas que publicaron menos de dos ocasiones y las distinguimos de las que publicaron en más de 3 ocasiones, resulta que el segundo grupo tiene una mejor precisión y certeza de sus resultados. Las encuestadoras que sólo siguen uno o dos eventos son por lo general más pequeñas y menos experimentadas que las otras. Resulta interesante que, entre las empresas que menos datos publican, 20% son universidades.

Gráfica 6. Índice General de Error, por casa encuestadora



Fuente: Banamex con datos de su archivo interno

### Conclusiones generales

2016 no fue el peor año para la industria de encuestas electorales, 2010 y 2015 tuvieron un peor desempeño. En los tres casos se trata de años en los que se publicó una gran cantidad de encuestas a las que los medios vuelven de consumo generalizado, más allá de la comunidad de análisis político y estrategia electoral, por lo que su alta tasa

generalizada de desacierto produce una crisis de confiabilidad hacia la industria de encuestas en pleno. Indudablemente, el incremento en esos errores no sólo obedece a problemas metodológicos o muestrales sino también a otros que aquí no han sido discutidos. Para pasar a esa discusión, sin embargo, nos parecía fundamental primero establecer cuál es el valor de las distintas dimensiones de información que

proporciona una encuesta y, en consecuencia, cómo puede evaluarse de forma estandarizada cuando una encuesta produce datos erróneos. Dejamos para un documento posterior el análisis de qué variables pueden explicar la ocurrencia y el tamaño de esos errores.

## BIBLIOGRAFÍA

Abundis, Francisco; Alcocer, Jorge; Becerra, Lorena; Campos, Roy y Moreno, Alejandro, (2016) “La vida privada de las encuestas”, No. 282, septiembre Revista Voz y Voto

Becerra, Lorena, (2016), “Encuestas: Las lecciones del 2016” en Periódico Reforma, Revista R, pp 8-10. 12 de junio del 2016.

Estévez, Federico (2010), “¿Las encuestas en decadencia?”, Este País, 233, septiembre, pp. 40-42.

Mann, Thomas: “What do the models say about who will win in November?” en FixGov, Septiembre 15, 2016, Brookings Institution, sitio electrónico.

Moreno, Alejandro (2010), “Las fallas de las encuestas”, Reforma, sección Enfoque, pp. 10-12, 25 de julio.

Moreno, Alejandro (2016), “Aciertos y Fallas de las encuestas” El Financiero, 6 de agosto 2016.

Moreno, Alejandro (2016), “Encuestas 2016, Lo bueno, lo malo y lo peor”, Num. 62, julio, Revista Campaigns&Elections.

Mosteller, Frederick et al. (1949), The Pre-Election Polls of 1948, Nueva York, Social Science Research Council.

Romero, Vidal (2012) “Notas para la Evaluación de las encuestas preelectorales, Las elecciones para gobernador de 2010 en México”, Volumen XIX, Número 1, 1 Semestre 2012 pp101-124 Política y Gobierno.

Romero, Vidal y Varela Carlo (2011) “Las precisiones de las encuestas preelectorales” Año 2, No. 0, Mayo-Julio 2011, Revista de Estudios Jurídicos Electorales. pp 30-37

Varela, Carlo (2010), “Factores muestrales y no muestrales que alteran el margen de error de una encuesta electoral en su pronóstico: Una primera aproximación vía empírica”, ensayo de titulación, México, Instituto Tecnológico Autónomo de México.