

Proceso Electoral: Modelo de agregación de encuestas

Resultados después de la elección

Fernando Antonio Zepeda Herrera
10 de julio del 2018

 **Numérica**[®]

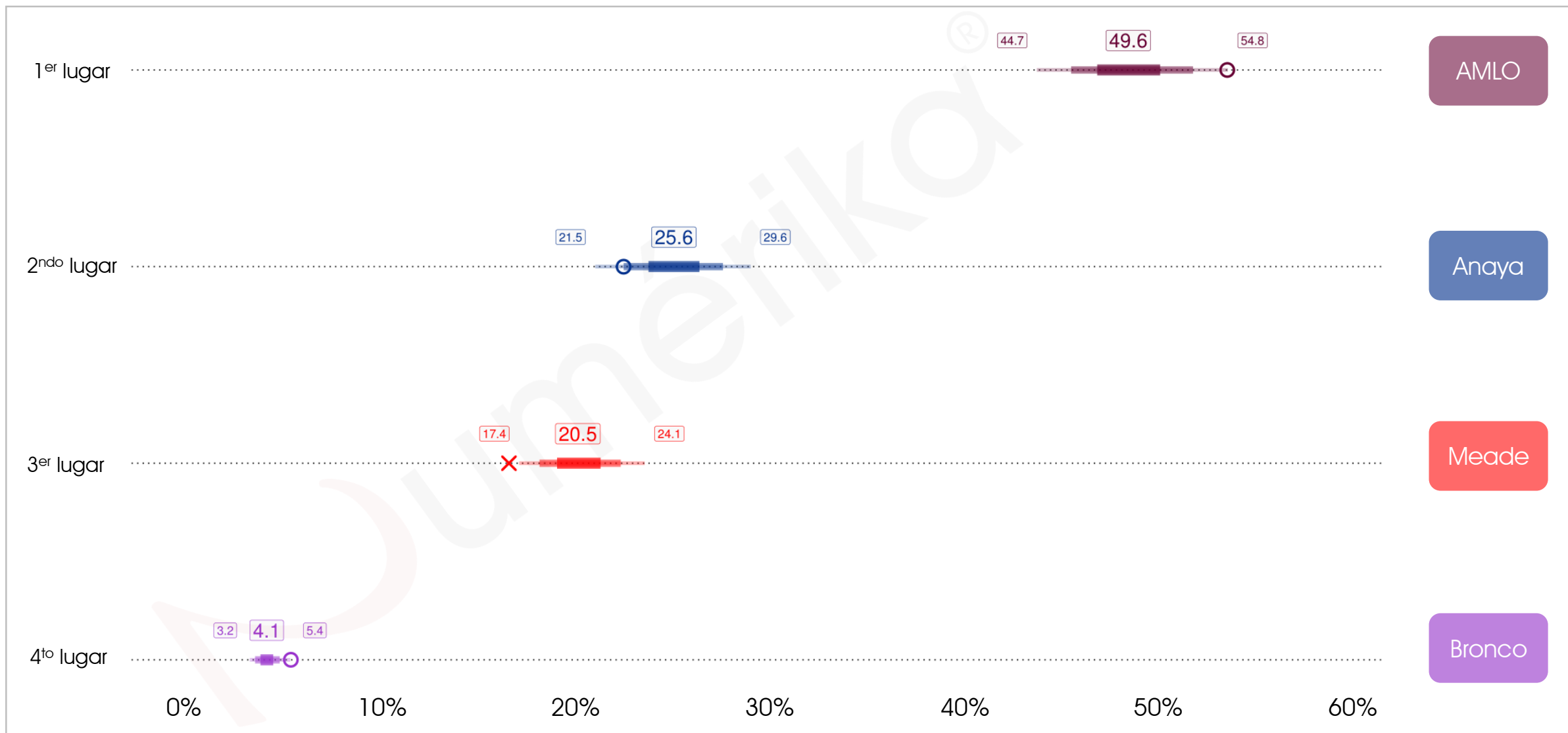
- A lo largo de un proceso electoral, las diferentes casas encuestadoras y campañas de los candidatos realizan encuestas con el objetivo de obtener información sobre el estado de las preferencias de los votantes.
- Debido a la volatilidad natural de las encuestas, en los últimos años han cobrado relevancia los **modelos de agregación de encuestas**, conocidos como *Poll of Polls*.
- Uno de los principales atractivos de estos modelos es que, al considerar varias estimaciones individuales, utilizan una mayor cantidad de información. Asimismo, **permiten realizar estimaciones más estables** pues tratan de separar el “ruido” estadístico de la tendencia “real”.
- De manera particular, **el modelo aquí presentado buscó reflejar la incertidumbre existente** sobre el resultado de la votación presidencial en cada semana previa a la jornada electoral, **agregando los resultados de las encuestas publicadas** por distintos medios electrónicos y escritos entre enero y junio del 2018.

Los **principales supuestos** en los que se basó el modelo son los siguientes:

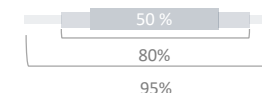
- En cada punto en el tiempo existen **preferencias subyacentes entre los votantes** que determinarían el resultado de la elección, *si esta se realizara en ese momento*.
- Cada encuesta buscó estimar dichas preferencias pero, debido a errores muestrales y no muestrales, el dato obtenido fue una desviación de las mismas.
- Dicha desviación dependió, primordialmente, de la casa encuestadora que realizó el ejercicio demoscópico a través de dos fuentes independientes:
 - Las posibles propensiones, constantes en el tiempo, a sobre o subestimar a cada candidato. **Es decir, pueden existir *efectos constantes de casa***.
 - Asimismo, cada casa encuestadora tiene una **volatilidad común entre sus encuestas**, proveniente de todos los errores muestrales y no muestrales que no pueden clasificarse como parte de los sesgos constantes.

En este sentido, el modelo de agregación de encuestas fue un **modelo de variables latentes, dinámico y jerárquico**.

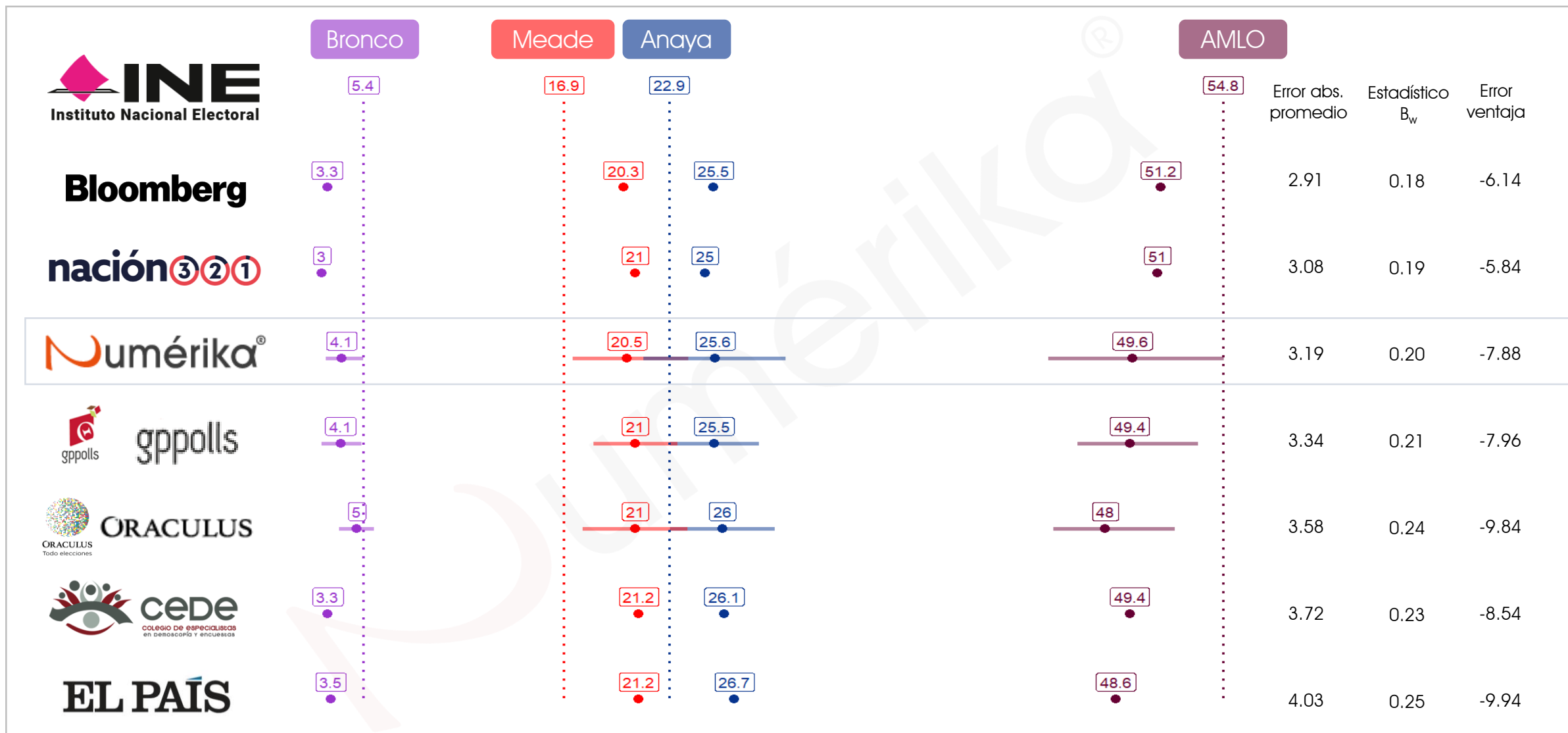
Estimación modelada de las preferencias y resultados oficiales



Estimación puntual y límites de los intervalos de probabilidad al 95% para las preferencias efectivas entre los 4 candidatos de la elección presidencial. Por cuestiones de redondeo, la suma no da 100%. Los segmentos de diferente grosor representan intervalos centrales de probabilidad al 50%, 80% y 95% como se muestra en el diagrama a la derecha. Un círculo representa que el intervalo contuvo al verdadero valor obtenido por el candidato en la jornada electoral, mientras que un tache significa que la votación obtenida se encontró fuera del intervalo.



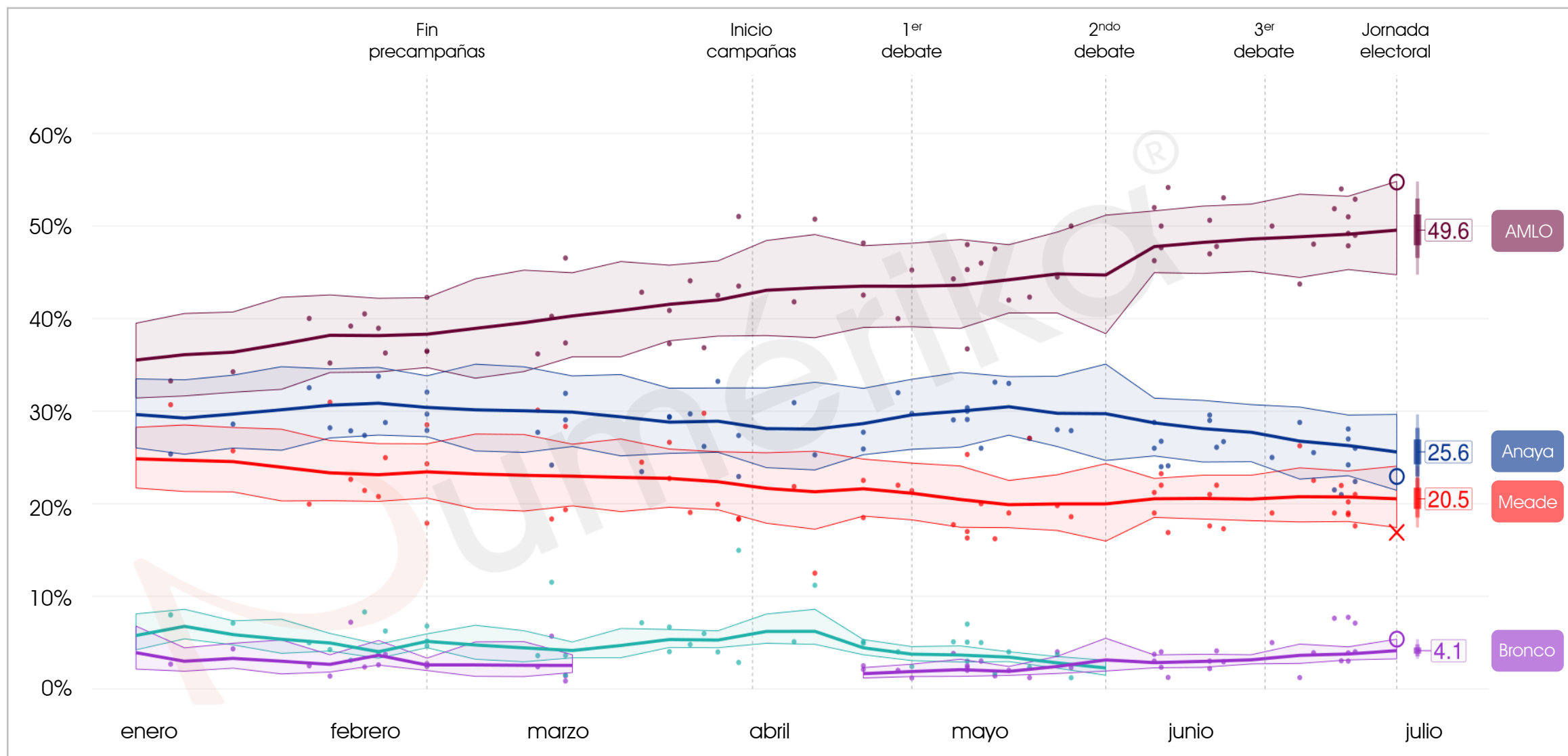
Comparativo de agregadores con los cómputos distritales



Estimación puntual y límites de los intervalos de probabilidad al 95% para las preferencias efectivas entre los 4 candidatos de la elección presidencial. Por cuestiones de redondeo, las sumas pueden no dar 100%.

La línea punteada indica el porcentaje efectivo de los cómputos distritales del INE. El error promedio considera las desviaciones absolutas para los 4 candidatos, el estadístico B_w fue propuesto por Arzheimer y Evans en 2014 para sistemas multipartidistas, mientras que el error ventaja es la diferencia en la estimación del margen entre los dos punteros. Las tres medidas son mejores entre más cercanas sean al cero.

Evolución de las preferencias modeladas y resultados oficiales

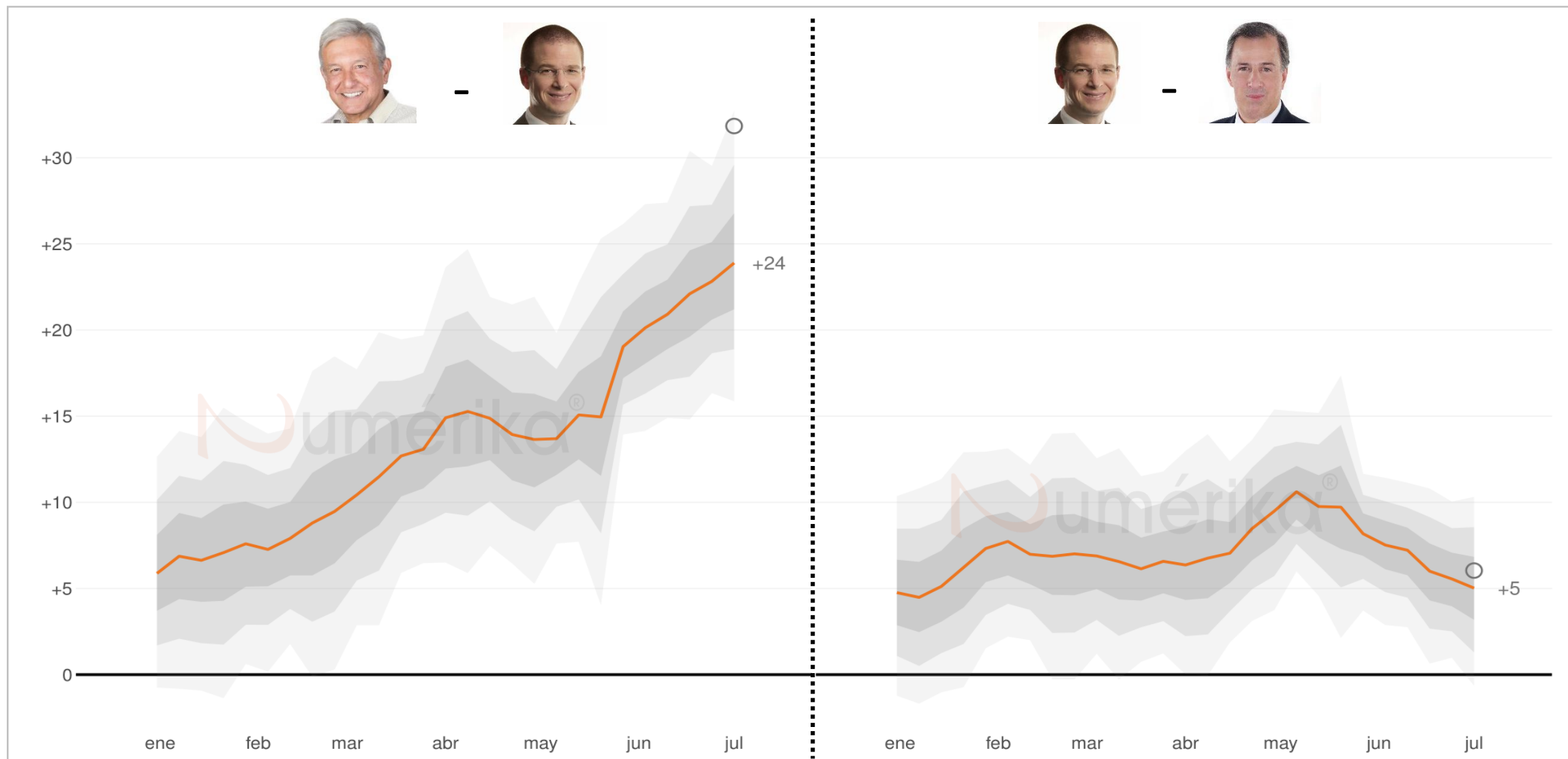


Estimaciones individuales de encuestas públicas para la proporción efectiva de votos entre los 4 candidatos de la elección presidencial. Evolución semanal de las preferencias modeladas.

Los últimos datos se presentan como estimaciones puntuales e intervalos centrales de probabilidad al 50%, 80% y 95%, como se muestra en el diagrama del costado derecho.

Un círculo representa que el intervalo contuvo al verdadero valor obtenido por el candidato en la jornada electoral, mientras que un tache significa que la votación obtenida se encontró fuera del intervalo.

Estimación del margen entre candidatos y márgenes finales



Última estimación puntual de la diferencia en puntos porcentuales de las preferencias efectivas entre cada par de candidatos (margen C1 – C2).

Evolución de la estimación del margen, la línea central indica la estimación puntual y las bandas son los intervalos centrales de probabilidad al 50%, 80% y 95%. El círculo representa que el intervalo contuvo al verdadero margen entre los candidatos.

Los **principales resultados** del modelo fueron los siguientes:

- Los últimos intervalos estimados **lograron capturar a 3 de los 4** resultados efectivos de los candidatos, solamente José Antonio Meade obtuvo un resultado fuera de su intervalo.
- El **error promedio** en las estimaciones puntuales fue de 3.19 puntos porcentuales. El de los diferentes modelos de agregación fue pequeño, variando entre 2.91 y 4.03.
- Esto **refleja la precisión** que tuvieron las *encuestas publicadas*, pues las más precisas tuvieron errores de aproximadamente 1 punto porcentual. Al mismo tiempo, podemos decir que **los modelos lograron su objetivo** de reducir los sesgos de las encuestas que más se alejaron de los resultados oficiales, mismas que llegaron a tener errores promedio de más de 7 puntos.
- Algo que tuvieron en común todas las agregaciones fue la **subestimación del margen de victoria** de Andrés Manuel López Obrador. En particular, nuestro modelo subestimó en 7.88 puntos porcentuales su diferencia con Ricardo Anaya. A pesar de lo anterior, *los dos intervalos de márgenes* entre candidatos que presentamos **sí capturaron los resultados finales**. Más aún, la estimación del margen entre Anaya y Meade fue precisa, con tan solo un error de 1 punto porcentual.

Puede consultar la metodología completa en <https://www.numerika.mx/spanish/elecciones-2018/poll-of-polls/>

