

UNA APROXIMACIÓN EMPÍRICA AL ERROR DE DISEÑO MUESTRAL EN LAS ENCUESTAS ELECTORALES DEL CIS¹

Jose M. Pavía

Departamento de Economía Aplicada, Universitat de València

Belén García-Cárceles

Departamento de Análisis Económico, Universitat de València

¹ Los autores desean agradecer la financiación del MICINN a través del proyecto CSO2009-11246, «Estructura Social, Encuestas y Predicciones Electorales», a Valentín Martínez sus respuestas a nuestras preguntas metodológicas y a Lluís Orriols y Vidal Díaz de Rada el habernos facilitado algunas de las referencias bibliográficas. Obviamente cualquier inexactitud es responsabilidad única de los autores.

Una Aproximación Empírica al Error de Diseño Muestral en las Encuestas Electorales del CIS

RESUMEN: Las encuestas electorales elaboradas por el Centro de Investigaciones Sociológicas (CIS) son, con justicia, consideradas un referente en el sector demoscópico español. El CIS emplea un procedimiento de muestreo polietápico, estratificado por conglomerados, con selección de unidades de muestreo (secciones censales) aleatoria con probabilidades proporcionales a la población residente. Los tamaños de las poblaciones residente y electoral, sin embargo, no son coincidentes y sus diferencias presentan un reparto heterogéneo en el espacio. Este trabajo cuantifica (para una muestra de elecciones) el impacto de esta discrepancia y propone emplear como factor de ponderación de cada sección el producto entre el número de españoles residentes mayores de 18 años y la tasa de participación registrada en la sección en las elecciones previas. Los análisis muestran que esta estrategia reduciría significativamente los sesgos introducidos durante el diseño muestral.

PALABRAS CLAVE: Muestreo por conglomerados, Muestreo estratificado, Muestra autoponderada, Factor de elevación, Elecciones españolas.

An empirical approximation to sample design error in CIS election polls

ABSTRACT: Electoral polls conducted by the Centro de Investigaciones Sociológicas (CIS) are fairly thought of as a benchmark in the Spanish polling industry. The CIS uses a multistage stratified cluster sampling procedure, with selection of sampling units (census sections) at random with probabilities of selection proportional to the resident population. The sizes of resident and electoral populations, however, are unequal and their differences are heterogeneously distributed in the electoral territory. This paper quantifies (for a sample of elections) the impact of this discrepancy and suggests using as weighting factor for each census section the product of the number of Spanish residents aged 18 years or more and the section turnout rate in the previous elections. The analyses show that this strategy would significantly reduce the bias introduced in the sampling design.

KEYWORDS: Cluster sampling, Stratified sampling, Auto-weighted sample, Elevation fraction, Spanish elections.

Recibido: 5 de junio de 2012.

Primera decisión: 27 de agosto de 2012.

Aceptado: 25 de septiembre de 2012.

1. Introducción

El muestreo probabilístico es una herramienta muy poderosa que ha permitido a los investigadores encontrar respuestas a cuestiones que, de otro modo, serían inabordables desde una perspectiva práctica. Los enormes costes económicos y/o temporales o, incluso, la imposibilidad real que implica realizar un censo en muchas situaciones habrían impedido disponer siquiera de soluciones aproximadas para multitud de preguntas. Como es sobradamente conocido, sin embargo, las soluciones que se obtienen siguiendo esta metodología son intrínsecamente inciertas, por lo que gran cantidad de técnicas han sido propuestas para mejorar sus prestaciones, tratando de minimizar costes, reducir errores y cuantificar las desviaciones (ver, e.g., Pérez López, 2009). Cuando la investigación implica a personas y se utiliza la encuesta como método de recogida de datos, las potenciales fuentes de error crecen y el investigador debe enfrentarse, además de a la existencia de errores muestrales, a la posibilidad de que se presenten otro tipo de errores, extramuestrales, que incluso pueden superar en magnitud a los derivados de la intervención del azar.

Una encuesta incluye un elevado número de operaciones y puede ser contemplada como un conjunto de procesos interconectados —entre los que se incluyen definir objetivos, seleccionar el marco muestral, planificar el trabajo de campo, diseñar el cuestionario, recolectar y procesar los datos, analizar y difundir los resultados y documentar la investigación (Statistics Canada, 2010)— que, para poder responder con éxito a los objetivos planteados, deben ajustarse a unos principios y metodologías establecidos (Rodríguez Osuma, 1991). Muchos y variados elementos distorsionadores pueden aparecer en cada una de las fases, por lo que no es de extrañar la enorme cantidad de literatura dedicada a intentar identificar/cuantificar su impacto y a tratar de proponer estrategias para atenuar sus potenciales consecuencias.

Además del inevitable error muestral, el sesgo en la estimación y el error en los resultados también puede surgir por otros factores más prosaicos como imperfecciones en el directorio o problemas de cobertura (e.g., Díaz de Rada, 2001; Pasadas del Amo et al., 2006, 2011), el efecto del entrevistador (e.g., Verge Mestre, 2009; Blomm et al., 2011) o del diseño del cuestionario (e.g., Abella Vázquez et al., 2010; Israel, 2010; Fitzgerald et al., 2011), por errores de medición (e.g., Bavdaž, 2010; Kreuter et al., 2010;) o por sesgos debidos a la no respuesta o la autoselección (Bethlehem et al., 2011; Pavía et al., 2011; Pavía y Larraz, 2012), por el modo/canal de recogida de los datos (e.g., Chang y Krosnick, 2009; Sakshaug et al., 2010; Díaz de Rada, 2011), o incluso por errores en la codificación o imputación de los datos faltantes (e.g., Särndal y Lundström, 2001; de Waal, 2011).

Las encuestas del CIS (Centro de Investigaciones Sociológicas) y, en particular, sus encuestas electorales se ajustan a los estándares de calidad exigidos en la industria de estudios de mercado y gozan de un gran prestigio en el sector demoscópico español. Coincidiendo con cada consulta electoral, el CIS investiga el comportamiento de los electores españoles mediante la realización de encuestas pre y post electorales y, con periodicidad mensual, el CIS busca medir el estado de la opinión pública española del momento a través de sus barómetros mensuales. La

profesionalización de la red de encuestadores que utiliza el CIS, el tamaño de sus muestras y los procedimientos de distribución espacial que emplea convierten a los datos recolectados por el CIS en una muy valiosa fuente de información; la cual, afortunadamente, está además disponible para el conjunto de estudiosos del fenómeno electoral español gracias al acceso gratuito que ofrece la institución a los microdatos de todas sus encuestas a través de su página web.²

Las encuestas electorales del CIS, sin embargo, adolecen de un pequeño error de diseño, fácilmente subsanable, que desvirtúa ligeramente los resultados que se derivan de las mismas. En particular, el CIS emplea un procedimiento de muestreo polietápico, estratificado por conglomerados, con selección de unidades de muestreo (secciones censales) de forma aleatoria proporcional a la población residente. Hecho que provoca que determinadas unidades estén ligeramente sobre-representadas, en detrimento de otras que están sub-representadas, debido a la circunstancia de la no coincidencia de los tamaños de las poblaciones residente y electoral y a la heterogénea distribución espacial de sus diferencias.

El presente trabajo tiene por objeto (i) cuantificar (desde una perspectiva empírica y para una muestra de elecciones) el impacto que esta ligera inexactitud tiene en las estimaciones y (ii) proponer una sencilla alternativa en la que, para seleccionar y distribuir la muestra, se sugiere tener en cuenta, además de a los tamaños disponibles de los cuerpos electorales, las tasas de participación previas registradas.

En concreto, el resto del documento está estructurado como sigue. En el apartado segundo se presenta el procedimiento de selección muestral utilizado en el CIS. En el apartado tercero se expone una propuesta de modificación, se detalla la estrategia de evaluación seguida y se presentan los resultados que se habrían obtenido (en condiciones ideales) en tres elecciones de distinta naturaleza, comparándose con los que se logran, en las mismas condiciones, con la estrategia actual. La última sección concluye el documento.

2. El diseño muestral del CIS

De acuerdo con las notas metodológicas que acompañan a las encuestas electorales realizadas por el CIS, el mecanismo seguido en la institución para la selección de unidades muestrales se define como un procedimiento polietápico estratificado por conglomerados, “con selección de las unidades primarias de muestreo (municipios) y de las unidades secundarias (secciones) de forma aleatoria proporcional, y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad.”

En concreto y siguiendo a (Rodríguez Osuna, 1991, 2005), en la primera etapa, se divide el espacio muestral en estratos (utilizando como variable de estratificación el tamaño del hábitat del municipio, medido a través de la población total residente) agrupando a los municipios en 7 categorías/estratos:—(1) menor o igual a 2.000 habitantes; (2) de 2.001 a 10.000; (3) de 10.001 a 50.000; (4) de 50.001 a 100.000; (5) de 100.001 a 400.000; (6) de 400.001 a 1.000.000; y, (7) más de 1.000.000 de

² http://www.cis.es/cis/opencms/ES/1_encuestas/catalogoencuestas.html

habitantes—, para, una vez estratificado el espacio muestral, asignar a cada estrato un número de entrevistas equivalente a la parte proporcional del tamaño muestral que le correspondería tras utilizar como factor de proporcionalidad la población residente del estrato.

Repartida la muestra entre los diferentes estratos, comienza la segunda etapa, de selección de secciones censales, mediante el procedimiento siguiente:

a) Se calcula, a partir del número de entrevistas que le corresponden a cada estrato, el número de secciones (puntos de muestreo) a elegir en cada estrato, procurando que el número de entrevistas en cada sección se sitúe alrededor de 10 (entre 8-12)³.

b) Una vez determinado el número de secciones a muestrear en cada estrato, la selección de las secciones en cada estrato se realiza sin reposición con probabilidad proporcional al total de la población residente en la sección. La manera de proceder en la selección, no obstante, depende del número de municipios en el estrato:

(b1) Si el número de municipios en el estrato es mayor o igual al de secciones a seleccionar en el estrato: las secciones son seleccionadas aleatoriamente con probabilidades de selección proporcionales a su tamaño (número de residentes), con la condición de no seleccionar más de una sección del mismo municipio. Por tanto, en cada municipio se selecciona a lo sumo una única sección.

(b2) Si el número de municipios en el estrato (nM) es mayor que la unidad pero menor que el de secciones a muestrear en el estrato (nS): (i) se selecciona de forma aleatoria proporcional una sección por municipio para, a continuación, (ii) seleccionar, con probabilidades de selección proporcionales a su tamaño, las restantes $nS - nM$ secciones entre el conjunto de todas las secciones no seleccionadas en (i).

(b3) Si el estrato lo compone un único municipio: las secciones son seleccionadas, con probabilidades proporcionales a su tamaño (número de residentes), de forma aleatoria entre todas las secciones del municipio.

Las restricciones (b1) y (b2) persiguen incrementar la dispersión espacial de las observaciones con el objetivo de lograr una muestra más representativa. El coste de tales restricciones, sin embargo, es romper la *supuesta* autoponderación⁴ de la muestra, aunque suele aceptarse que ésta se mantiene de forma aproximada⁵. Empero lo anterior, la autoponderación de la muestra no es tal, ya que la población residente (utilizada para determinar las ponderaciones) no coincide con la población objetivo o universo, que son los electores (en general, y dependiendo de la elección, los españoles mayores de 18 años con derecho a voto).

³ De acuerdo con Díaz de Rada (2008, Tabla 3), sin embargo, en cerca de un 30% de las secciones muestreadas para los barómetros del año 2004 el número de entrevistas realizadas fue inferior a ocho.

⁴ Todas las unidades de la población tienen la misma probabilidad de ser seleccionadas.

⁵ Aunque como señala Díaz de Rada (2008, p. 227) la aplicación durante el trabajo de campo de las normas de selección de los individuos del CIS “supone una gran alteración de la equiprobabilidad”.

A fin de ilustrar cómo funcionan los criterios de selección de secciones utilizados en el CIS, explicaremos a través de un ejemplo con datos reales el proceso de selección. Imaginemos que hubiéramos deseado realizar una encuesta de tamaño 1000 con motivo de las elecciones a la Asamblea de Madrid de 2007. En primer lugar, y de acuerdo con el padrón de habitantes de Enero de 2006 (el más actualizado⁶ en la fecha de celebración de las elecciones autonómicas de Mayo de 2007), obtendríamos el total de población correspondiente a cada una de las siete categorías/estratos en que se dividiría la población (segunda columna de la Tabla 1), para a continuación, determinar el número de encuestas a realizar en cada estrato repartiéndolo proporcionalmente el tamaño muestral (cuarta columna de la Tabla 1): 9 encuestas en el primer estrato, 52 para el segundo estrato, ..., y 521 en el último.

Tabla 1.
Número de encuestas a realizar por estrato para las elecciones a la Asamblea de Madrid 2007

Categoría	Población	Tanto por mil	Encuestas x estrato
≤ 2.000 habitantes	53.133	8.84	9
De 2.001 a 10.000	312.835	52.07	52
De 10.001 a 50.000	623.780	103.82	104
De 50.001 a 100.000	568.783	94.67	94
De 100.001 a 400.000	1.321.052	219.88	220
De 400.001 a 1.000.000	0	0	0
> 1.000.000 de habitantes	3.128.600	520.72	521
Total	6.008.183	1.000	1.000

Fuente: Elaboración propia a partir de datos de población INE (www.ine.es).

Una vez repartida la muestra entre los diferentes estratos, y de acuerdo con el punto a) de la estrategia de muestreo CIS, se seleccionarían (ver Tabla 2)⁷: una única sección en el primer estrato —dónde se realizarán 9 entrevistas (E1: 1→9); cinco secciones en el segundo estrato, debiendo realizarse 10 entrevistas en tres de las secciones y 11 en las otras dos (E2: 3→10, 2→11); E3: 6→9, 5→10; E4: 6→9, 4→10; E5: 22→10; y, E7: 51→10; 1→11.

Tras determinar el número de secciones a muestrear en cada estrato (que aparece en la Tabla 2) hay que proceder a su selección cumpliendo las restricciones de selección establecidos en las encuestas del CIS. En concreto:

- Del estrato 7 (Madrid ciudad) habría que extraer (siempre sin reposición) 52 secciones (siempre al azar y con probabilidad de selección proporcional a la población total de la sección).

⁶ Disponible de modo agregado en el ámbito municipal desde el 29 diciembre de 2006 y por sección censal desde el 13 de enero de 2007.

⁷ Obsérvese, que la solución de reparto no es única. Por ejemplo, en el estrato 3 en lugar de 11 secciones se podría haber optado por seleccionar 10 secciones, con el objetivo de realizar 10 entrevistas en 6 de las secciones y 11 en las 4 restantes.

- Del estrato 5 (8 municipios) habría que extraer una sección de cada municipio y 14 secciones adicionales de entre todas las restantes secciones del estrato.
- Del estrato 4 (8 municipios) habría que extraer una sección de cada municipio y 2 secciones adicionales de entre todas las restantes secciones del estrato.
- Del estrato 3 (25 municipios) habría que extraer un total de 11 secciones de otros tantos municipios diferentes.
- Del estrato 2 (62 municipios) habría que extraer un total de 5 secciones de otros tantos municipios diferentes.
- Del estrato 1 (75 municipios) habría que extraer una única sección.

Seleccionadas las secciones que componen las muestras se aborda la última etapa, la selección de los individuos; la cual se lleva a cabo sobre el terreno directamente por los entrevistadores aplicando un sistema combinado de rutas aleatorias y cuotas de edad y sexo.

Una vez elegido el punto de partida de la ruta, el encuestador va seleccionando a los encuestados siguiendo una serie de normas tanto en lo que se refiere a las viviendas que pueden ser seleccionadas y a los criterios de sustitución, como en lo que respecta a los individuos que pueden ser elegidos dentro de cada vivienda.

Tabla 2.

Número de municipios y secciones a seleccionar por estratos para una encuesta de tamaño 1000 en las elecciones a la Asamblea de Madrid 2007.

Estrato	Municipios en el estrato	Secciones a seleccionar	Encuestas x sección
1	75	1	9
2	62	5	10 en 3, 11 en 2
3	25	11	9 en 6, 10 en 5
4	8	10	9 en 6, 10 en 4
5	8	22	10
6	0	0	0
7	1	52	10 en 50, 11 en 1

Fuente: Elaboración propia.

En concreto, siguiendo a Díaz de Rada (2008), los encuestadores del CIS reciben instrucciones precisas sobre los puntos de partida de las rutas dentro de cada sección censal y sobre los criterios de selección de portales (uno de cada tres en general y uno de cada cinco en viviendas unifamiliares), de forma que “[d]entro de un portal hay que realizar una entrevista cada 12 viviendas o fracción, limitando a tres el número máximo de entrevistas en un portal” (p. 214), no pudiendo realizarse más de una entrevista en un mismo piso y de manera que “cuando una entrevista no se consiga en el primer contacto se puede seguir intentando en la puerta contigua” (p. 216). Seleccionada una vivienda, la elección de la persona a entrevistar se realizará

exclusivamente entre las personas que cumplan las características fijadas en las cuotas todavía no cubiertas, “eligiendo a la persona más joven cuando varios miembros de la vivienda cumplen tales características” (p. 216). Cuando en la vivienda no exista ninguna persona cumpliendo las características el contacto se descarta, procediéndose a la selección de una nueva vivienda de acuerdo con las normas fijadas de antemano.

3. Cuantificando el Error de Diseño Muestral

Como ha sido puesto de manifiesto en el apartado anterior, la selección de secciones censales en las encuestas del CIS se realiza probabilísticamente utilizando como ponderación la población total de la sección. Los tamaños de las poblaciones residente y electoral (que, sin pérdida de generalidad, supondremos está compuesta por españoles de 18 y más años)⁸, sin embargo, no son coincidentes, lo que provoca que aquellos individuos que viven en secciones con un mayor porcentaje de extranjeros y/o menores de edad tengan mayores probabilidades de ser seleccionados.

En este apartado cuantificamos la reducción que se lograría en el error de estimación si para seleccionar las unidades muestrales se empleasen como ponderaciones valores más próximos a los verdaderos tamaños de los electorados de cada sección. En concreto, si se utilizase el número de españoles con 18 o más años (**Pob.Esp**) o, todavía mejor, si se emplease la variable anterior multiplicada por la tasa de participación registrada en la sección en las elecciones previas.

En efecto, consideramos que la segunda de las estrategias es pertinente y debería mejorar la calidad de las estimaciones dado que, por una parte, la variable relevante en predicción electoral es el porcentaje de votos que obtendrá cada candidatura y que, por otra parte, (i) a pequeña escala se observa una evidente y persistente relación entre las tasas de participación de elecciones sucesivas (e.g., Pavía et al., 2008; Pavía-Miralles y Larraz-Iribas, 2008) y que, además, (ii) los abstencionistas suelen presentar una menor propensión a participar en las encuestas electorales (e.g., Brehm, 1993; Weisberg, 2005; De la Calle et al., 2010; Pavía y Larraz, 2012)⁹.

Para evaluar la pertinencia de emplear las variables anteriores como factores de ponderación, mediremos, en ‘condiciones ideales’ y para tres elecciones diferentes (una local, una autonómica y una nacional), la proximidad a los verdaderos valores de las predicciones que se obtienen con cada una de las variables de ponderación

⁸ Hay que notar que en las elecciones municipales también pueden solicitar su inscripción en el censo electoral “aquellas personas que, sin haber adquirido la nacionalidad española, sean ciudadanos de países con los que España haya establecido un Acuerdo reconociendo el derecho a votar en las elecciones municipales a los nacionales miembros de los referidos Estados en España y a los españoles en dichos Estados.” (BOE, 2010, p. 74534) y que, desde 2011, se une “... indefectiblemente el ejercicio de sufragio en elecciones municipales, locales y forales a la condición de vecino” por lo que los españoles ausentes se excluyen del censo electoral para elecciones locales (BOE, 2011, p. 9505), si bien todavía continúan formando parte del mismo para el resto de elecciones.

⁹ Por ejemplo, de acuerdo con Brehm (1993, p.65) “[t]he coefficient on political participation is strongly positive (0.201) and statistically significant at $p < 0.01$. In other words, the more often the respondent participates in politics, the more times that person has participated in an interview”.

consideradas (población residente, población española con derecho a voto y la variable anterior modulada con la tasa de participación registrada en las elecciones previas).

Las ‘condiciones ideales’ han sido simuladas asumiendo que se dispone de una encuesta en la que se han muestreado todas las secciones censales y que, a nivel de sección censal, las estimaciones obtenidas de los porcentajes de votos coinciden con los verdaderos valores registrados en las elecciones. De esta manera, las diferencias agregadas que se observen entre estimaciones y verdaderos valores podrán ser achacadas exclusivamente a los diferentes factores de elevación utilizados; pudiendo ser empleadas, por tanto, las diferencias relativas en el ámbito de la circunscripción como un indicador del error adicional que introduce en las estimaciones el uso de unas ponderaciones que no se aproximan suficientemente a los tamaños de las poblaciones objetivo de cada unidad de muestreo.

En particular, denotando por π_{ij} al porcentaje de votos registrado por el partido i -ésimo en la j -ésima unidad de muestreo, y por $w. cis_j$, $w. esp_j$ y $w. tp_j$, al peso de la sección j -ésima en la circunscripción utilizando, respectivamente, como variables de ponderación la población total, la población española de edad mayor o igual a 18 años y $w. tp_j = w. esp_j \cdot tp_j$ (siendo tp_j la tasa de participación registrada en las elecciones previas en la j -ésima unidad de muestreo), se tiene que las estimaciones de los porcentajes de votos para cada partido, utilizando las distintas ponderaciones, vendrán dadas, respectivamente, por:

$$\begin{aligned}\pi. cis_i &= \sum_j \pi_{ij} w. cis_j \\ \pi. esp_i &= \sum_j \pi_{ij} w. esp_j \\ \pi. tp_i &= \sum_j \pi_{ij} w. tp_j\end{aligned}$$

De donde, inmediatamente es fácil obtener un indicador del grado de ajuste (error de estimación). En concreto, hemos utilizado la suma de las diferencias en valor absoluto entre los verdaderos valores, $\pi. real_i$, y las diferentes estimaciones, $\pi. est_i$.

$$\text{Error} = \sum_i |\pi. real_i - \pi. est_i|$$

A continuación, en los siguientes tres subapartados, se muestran los resultados obtenidos tras aplicar los estadísticos anteriores a los datos de las elecciones locales de Barcelona 2007, las elecciones autonómicas de la Comunidad Valenciana de 2011 y las elecciones Generales de 2011.

3.1 Elecciones Locales de Barcelona 2007

Previo a las elecciones locales de Barcelona los últimos datos de población publicados por el INE (Instituto Nacional de Estadística) correspondían a la explotación estadística del padrón de habitantes de 2006, por lo que tales han sido las cifras de población manejadas en esta aplicación¹⁰. Así mismo, como cifras electorales se han utilizado los datos correspondientes a las elecciones municipales al Ayuntamiento de Barcelona de 2003 y 2007.

Antes de poder realizar las comparaciones, sin embargo, es necesario tratar los datos. Dos son las dificultades que surgen. Por una parte, el nivel de desagregación con el que el INE ofrece la información y, por otro lado, las modificaciones que experimenta la división del espacio en secciones censales a lo largo del tiempo. La segunda de las dificultades ha sido resuelta de manera sencilla comparando los códigos de las secciones y aplicando las reglas de asignación detalladas en Pavía-Miralles (2005, pp. 1117-8). Respecto a las cifras de población, éstas están disponibles de forma agregada para el total de españoles y extranjeros y, sin distinguir la nacionalidad, por grupos quinquenales de edad. La cifra de población que necesitamos no puede ser, por tanto, derivada directamente. Es por ello que la población española en edad de votar ha debido de ser estimada; lo cual se ha conseguido asumiendo hipótesis de distribución uniforme de españoles y extranjeros en cada grupo de edad. En concreto, denotando por P_j^{0-4} , P_j^{5-9} , P_j^{10-14} , P_j^{15-19} , E_j , X_j , respectivamente, a los tamaños de población de 0 a 4 años, de 5 a 9 años, de 10 a 14 años, de 15 a 19 años, española total y extranjera total correspondientes la j -ésima sección, se tiene que la población española de 18 o más años de la sección, **Pob. Esp_j**, ha sido aproximada a través de:

$$\text{Pob. Esp}_j = E_j - \frac{(P_j^{0-4} + P_j^{5-9} + P_j^{10-14} + \frac{3}{5} P_j^{15-19})}{E_j + X_j} E_j$$

Los resultados obtenidos para estimaciones y errores se ofrecen en la Tabla 3. Como se observa claramente, para estas elecciones el uso de la población residente como elemento de ponderación introduce un sesgo agregado importante y significativo en las predicciones, destacando la sobreestimación que se observa para PSC y la subestimación de CiU. La utilización como factor de ponderación del número de residentes españoles, sin embargo, no parece mejorar destacablemente las estimaciones: la reducción del error se concreta en apenas un 14%. El error, no obstante, sí se reduce espectacularmente (disminuye en casi tres cuartas partes) cuando la tasa de participación registrada en las elecciones previas es también utilizada para definir los pesos.

¹⁰ Dependiendo de la antelación con respecto al día de la elección con que se lleven a término las encuestas pre-electorales podrían incluso ser utilizados los censos electorales, ya que el censo electoral de unas elecciones suele estar disponible alrededor de un mes antes de la fecha de celebración.

3.2. Elecciones a Corts Valencianes 2011

Antes de las elecciones autonómicas de Mayo de 2011, las cifras oficiales de población más actuales, disponibles a nivel de sección censal y municipal, habían sido publicadas por el INE en Enero de 2011; las cuales correspondían a la explotación del padrón de habitantes de 2010. Por lo que tales estadísticas, junto con los resultados registrados en las elecciones a Corts Valencianes de 2007 y 2011 han sido los datos empleados en la aplicación que se ofrece en este subapartado. Al igual que en el ejemplo anterior, la información fue tratada a fin establecer la correspondencia entre las secciones censales de los diferentes momentos temporales y de disponer para cada sección censal de estimaciones del número de españoles con 18 o más años.

Las predicciones de los porcentajes de votos que, con cada una de las estrategias, habrían obtenido las principales formaciones que concurrieron a las elecciones, así como los errores asociados se ofrecen (de forma separada para cada una de las tres circunscripciones en que se divide la Comunidad Valenciana) en la Tabla 4. Comparado con el ejemplo anterior, destaca que los errores son, en todos los casos, sensiblemente menores, aunque igualmente se mantiene la relación entre alternativas encontrada en el caso anterior.

Tabla 3.
Estimaciones y Errores. Elecciones locales Barcelona 2007.

	Partidos						Error
	CiU	ERC	ICV	PP	PSC	Resto	
Real	26,54	9,23	9,75	16,27	31,24	6,97	-
CIS	25,56	9,22	9,94	16,09	32,17	7,01	2,36
Pob.Esp	25,65	9,21	9,84	16,15	32,16	6,99	2,09
Propuesta	26,29	9,26	9,78	16,19	31,52	6,96	0,68

Fuente: Elaboración propia a partir de datos electorales facilitados por el Area de Processos Electorals del Departament de Governació i Administracions Públiques de la Generalitat de Catalunya y datos de población obtenidos de la web del Instituto Nacional de Estadística (www.ine.es).

A nivel de circunscripción destaca el caso de Castellón donde el error con los pesos del CIS es 8 veces y media superior al error que se cometería si se emplease la estrategia propuesta en esta investigación. En el caso de Alicante y Valencia la ganancia también es palpable, el error con pesos CIS es en torno a 2.5 veces superior al que se lograría utilizando la población con derecho a voto modulada con la tasa participación previa.

3.3. Elecciones a Cortes Generales 2011

En este subapartado se presentan los resultados logrados para las elecciones al Congreso de los Diputados de 2011. En esta ocasión, y debido a los enormes costes de tratamiento de datos que hubiese supuesto operar con los resultados a nivel de sección censal, se ha considerado oportuno el trabajar con datos municipales. Como contrapartida a la reducción de los costes operacionales, las diferencias que se obtienen se ven reducidas por el efecto igualador que supone la agregación municipal de las secciones, la cual elimina las heterogeneidades intra-municipales.

En la aplicación se han empleado, como datos de población, las cifras más actuales disponibles en Noviembre de 2011 en la web del INE, correspondientes al padrón de habitantes de 2010. Así mismo, en esta ocasión, se han empleado como datos electorales resultados de tres elecciones: las generales de 2011, 2008 y 2004. Se ha experimentado con tasas de participación previas de dos elecciones (2008 y 2004), en lugar de con datos de una única elección, por el efecto distorsionador que en las elecciones de 2008 tuvo sobre la participación la ilegalización de la marca electoral de la izquierda abertzale en las circunscripciones vascas y Navarra.

Tabla 4.
Estimaciones y Errores. Elecciones a Cortes Valencianas 2011.

		Partidos							Error
		Compromís	EUPV	PP	PSOE	UPyD	Verdes	Resto	
Alicante	Real ⁽¹⁾	4,69	5,51	50,97	30,62	3,18	1,92	3,10	-
	CIS	4,50	5,37	51,05	30,59	3,12	2,17	3,20	0,86
	Pob.Esp	4,65	5,64	50,71	30,66	3,26	1,98	3,10	0,60
	Propuesta	4,75	5,52	51,07	30,48	3,17	1,91	3,10	0,33
Castellón	Real ⁽¹⁾	6,93	5,32	49,46	31,90	1,97	1,08	3,34	-
	CIS	7,12	5,35	49,82	30,96	2,10	1,15	3,51	1,87
	Pob.Esp	7,06	5,33	49,86	31,13	2,06	1,12	3,44	1,53
	Propuesta	6,92	5,30	49,57	31,84	1,96	1,08	3,33	0,22
Valencia	Real ⁽¹⁾	9,09	6,53	50,79	26,92	2,28	1,02	3,36	-
	CIS	8,98	6,64	50,52	26,82	2,32	1,05	3,43	0,73
	Pob.Esp	9,00	6,65	50,55	26,80	2,31	1,04	3,41	0,68
	Propuesta	9,11	6,52	50,89	26,80	2,31	1,02	3,35	0,30

Fuente: Elaboración propia a partir de datos electorales disponibles en el archivo histórico electoral de la Comunidad Valenciana, <http://www.pre.gva.es/argos/archivo/index.html>, y datos de población obtenidos de la web del Instituto Nacional de Estadística (www.ine.es).

(1) No incluye el voto de los españoles residentes en el extranjero.

En el ámbito municipal, el INE ofrece información poblacional con un nivel de detalle superior al de sección censal, por lo que, en este caso se ha empleado una

expresión más ajustada para la estimación de la población española con 18 o más años. En concreto, si E_j^{0-15} , P_j^{15-19} , E_j , X_j representan los tamaños de población española de 0 a 15 años, la población total de 15 a 19 años, la española total y la extranjera total correspondientes al j-ésimo municipio, respectivamente, los valores de $Pob. Esp_j$ han sido aproximados esta vez a través de:

$$Pob. Esp_j = E_j - E_j^{0-15} - \frac{2 P_j^{15-19}}{5 E_j + X_j} E_j$$

Así mismo, las estimaciones no han cubierto todos los partidos, sino que de las 62 candidaturas que obtuvieron algún soporte en alguna de las circunscripciones, las estimaciones se han centrado en aquellos partidos o coaliciones que obtuvieron apoyo significativo en alguna de las circunscripciones. En concreto, PP, PSOE, IU-LV, UPyD, CiU, AMAIUR, EAJ-PNV, ESQUERRA, EQUO, BNG, CC-NC-PNC, COMPROMÍS-Q, PACMA, FAC, Eb, PA, PxC, PRC y GBAI, más el resto.

Los resultados de los errores de estimación que se obtendrían utilizando los diferentes factores de elevación se ofrecen, para todas las circunscripciones¹¹, en la Tabla 5. Como se puede observar, aunque mitigado por la mayor agregación que implica trabajar con datos municipales, también en esta ocasión el empleo de unas ponderaciones más próximas a los verdaderos tamaños de las poblaciones objetivo hubiera supuesto un mejor ajuste a los resultados reales. En promedio, el uso de la población española en edad de votar y de la misma variable modulada por la tasa de participación de 2008 reduce el error, en ambos casos, a algo menos de la mitad; mientras que, la reducción es de casi tres cuartas partes cuando se emplea la tasa de participación de 2004. La mayor reducción del error que se observa utilizando las tasas de participación de 2004, se explica por los mayores errores sistemáticos observados en Álava, Guipúzcoa, Navarra y Vizcaya como consecuencia del efecto ya mencionado de la ilegalización de ANV (la formación de la izquierda abertzale para las elecciones de 2008). De hecho, excluyendo Navarra y las circunscripciones vascas, la reducción de error de nuestra propuesta con respecto a la estrategia CIS habría sido de un 85%.

4. Conclusiones

Probablemente, las encuestas electorales más valoradas por el conjunto de los agentes implicados en los estudios electorales en España sean las elaboradas por el CIS. Las encuestas del CIS, que utilizan un procedimiento de selección muestral polietápico, estratificado, por conglomerados, sin embargo, adolecen de un pequeño error de diseño muestral al utilizar el conjunto de la población residente como variable de ponderación. El problema radica en que, como es conocido, los tamaños

¹¹ Ceuta y Melilla han sido omitidas ya que, al estar compuestas por un único municipio, tienen con una ponderación a nivel municipal (cualquiera que este sea) un error de estimación nulo.

de las poblaciones residente y electoral no son coincidentes y en que, además, sus diferencias no se reparten uniformemente por el territorio. Ello provoca que determinadas unidades tengan una probabilidad mayor (menor) de ser seleccionadas de la que les correspondería.

El objetivo de este trabajo ha sido cuantificar (para una muestra de elecciones) el impacto de esta discrepancia y proponer una alternativa de ponderación de las secciones censales más ajustada a los tamaños reales de las poblaciones electorales. El análisis se ha llevado a cabo para tres elecciones: las elecciones locales al Ayuntamiento de Barcelona de 2007, las elecciones a Corts Valencianes de 2011 y las elecciones al Congreso de los Diputados de 2011. A la luz de los resultados, el sesgo de estimación podría ser reducido significativamente si se emplease para determinar los factores de elevación/selección una variable compuesta por el producto del número de españoles mayores de 18 años residentes en la sección y la tasa de participación registrada en esa misma sección en las elecciones previas.

La consideración de la tasa de participación previa, no obstante, introduce (i) la posibilidad de que, como ocurrió en Navarra y las circunscripciones vascas en 2008, puedan aparecer nuevos factores distorsionadores y (ii) la necesidad de establecer una correspondencia entre secciones de momentos temporales diferentes. Respecto a la primera de las cuestiones, obviamente, no hay que hacer una aplicación ciega de la metodología sino que, como ocurre siempre cuando se diseña una encuesta, hay que utilizar el sentido común y tener en cuenta toda la información previa disponible relevante. Por otro lado, respecto al trabajo extra que supone la necesidad de establecer el mapa de relaciones entre secciones en distintos momentos del tiempo, tenemos que, afortunadamente, éste podría ser resuelto muy satisfactoriamente y de forma automática¹² utilizando las propuestas detalladas en Pavía y López-Quilez (2012), supuestos disponibles las divisiones espaciales del espacio electoral en secciones censales. Alternativamente, este establecimiento de relaciones *a priori* podría ser evitado si para determinar las unidades de muestreo se utilizase la división en secciones de la última elección disponible. Esta estrategia además de no introducir mucho ruido en el proceso, podría también conllevar consecuencias muy positivas al permitir, por una parte, diseños muestrales en los que podría ser utilizada toda la información electoral previa disponible para la selección de secciones y, por otro lado, emplear estrategias de predicción electoral basadas en estimadores de pequeñas áreas (Rao, 2003).

¹² Es preciso notar, no obstante, que entre elecciones sucesivas no es habitual que se produzcan grandes modificaciones en la división del espacio electoral. Como excepción encontramos la reestructuración que Barcelona realizó de su división en secciones censales previo a las elecciones europeas de 2009 (Pavía y López-Quilez, 2012).

Tabla 5.
Errores estimación. Elecciones al Congreso de los Diputados 2011.

	CIS	Pob.Esp	Propuesta08	Propuesta04
Alava	0.38	0.09	1.10	0.05
Albacete	0.48	0.18	0.11	0.16
Alicante	0.88	0.27	0.04	0.11
Almería	1.96	0.58	0.06	0.12
Ávila	0.66	0.11	0.08	0.19
Badajoz	1.07	0.56	0.08	0.11
Illes Balears	0.77	0.46	0.05	0.05
Barcelona	0.44	0.19	0.09	0.30
Burgos	0.69	0.28	0.08	0.15
Cáceres	0.87	0.33	0.13	0.21
Cádiz	0.74	0.61	0.09	0.09
Castellón	0.43	0.20	0.07	0.10
Ciudad Real	0.64	0.22	0.01	0.15
Córdoba	1.05	0.99	0.28	0.14
A Coruña	0.56	0.23	0.12	0.17
Cuenca	0.48	0.30	0.06	0.14
Girona	2.35	0.53	0.15	0.24
Granada	0.44	0.26	0.11	0.04
Guadalajara	0.63	0.06	0.11	0.09
Guipúzcoa	0.18	0.23	3.79	0.22
Huelva	1.61	0.73	0.09	0.25
Huesca	0.36	0.19	0.04	0.07
Jaén	0.58	0.45	0.13	0.14
León	0.51	0.20	0.16	0.17
Lleida	0.86	0.21	0.36	0.66
La Rioja	0.29	0.17	0.05	0.04
Lugo	0.68	0.20	0.11	0.32
Madrid	0.07	0.15	0.06	0.08
Málaga	1.54	0.69	0.18	0.07
Murcia	0.17	0.01	0.05	0.09
Navarra	0.23	0.70	2.44	0.62
Ourense	0.58	0.15	0.07	0.06
Asturias	0.12	0.13	0.06	0.08
Palencia	0.69	0.17	0.04	0.11
Las Palmas	1.19	0.40	0.13	0.13
Pontevedra	0.35	0.13	0.07	0.15
Salamanca	0.54	0.14	0.05	0.05
S.C. de Tenerife	0.78	0.49	0.08	0.15
Cantabria	0.70	0.41	0.24	0.34
Segovia	0.42	0.05	0.03	0.11
Sevilla	0.37	0.30	0.05	0.11
Soria	1.29	0.10	0.13	0.26
Tarragona	0.72	0.14	0.43	0.59
Teruel	0.57	0.22	0.09	0.15
Toledo	0.61	0.15	0.05	0.12
Valencia	0.23	0.17	0.04	0.07
Valladolid	0.11	0.08	0.06	0.07
Vizcaya	0.69	0.95	2.57	0.49
Zamora	0.61	0.23	0.16	0.25
Zaragoza	0.13	0.03	0.06	0.07

Fuente: Elaboración propia a partir de datos electorales obtenidos de la web

<http://www.infoelectoral.mir.es/min/areaDescarga.html?method=inicio> y datos de población obtenidos de la web del Instituto Nacional de Estadística (www.ine.es).

Referencias

- Abella Vázquez, C.; Rodríguez Losado, D. y Pérez-Sindín López, X. (2010). Efectos del cambio de orden de variables y categorías de respuesta en los resultados de una encuesta política. *Metodología de Encuestas*, 12, 7-27.
- Bavdaž, M. (2010). Sources of Measurement Errors in Business Surveys. *Journal of Official Statistics*, 26, 25-42.
- Bethlehem, J.; Cobben, F. y Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. Wiley: London.
- Blom, A.G.; de Leeuw, E.D. y Hox, J.J. (2011). Interviewer Effects on Nonresponse in the European Social Survey. *Journal of Official Statistics*, 27, 359-377.
- BOE (2010). Orden EHA/2264/2010, de 20 de julio, por la que se dictan normas e instrucciones técnicas para la formación del censo electoral de residentes en España que sean nacionales de países con Acuerdos para las elecciones municipales. *Boletín Oficial del Estado*, Núm. 208, Viernes 27 de agosto de 2010, Sec. I, 74533-74535.
- BOE (2011). Ley Orgánica 2/2011, de 28 de enero, por la que se modifica la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General. *Boletín Oficial del Estado*, Núm. 25, Sábado 29 de enero de 2011, Sec. I, 9504-9523.
- Brehm, J. (1993). *The Phantom Respondents: Opinion Surveys and Political Representation*. University of Michigan Press, Ann Arbor.
- Chang, L. y Krosnick, J.A. (2009). National Surveys Via RDD Telephone Interviewing Versus the Internet. Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly*, 73, 641-678.
- De la Calle, L.; Martínez, A., y Orriols, L. (2010). Voting without ideology. Evidence from Spain (1979-2008). *Revista Española de Investigaciones Sociológicas*, 129, 107-129.
- de Waal, T.; Pannekoek, J. y Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. John Wiley & Sons, Inc.: New Jersey.
- Díaz de Rada, V. (2001). Problemas de cobertura de las encuestas telefónicas. *Revista Española de Investigaciones Sociológicas*, 93, 133-164.
- Díaz de Rada, V. (2008). La selección de los entrevistados últimos en encuestas presenciales: un análisis de la utilización conjunta del método de rutas y el método de cuotas. *Revista Española de Investigaciones Sociológicas*, 123, 209-247.
- Díaz de Rada, V. (2011). Encuestas con encuestador y autoadministradas por internet. ¿Proporcionan resultados comparables?. *Revista Española de Investigaciones Sociológicas*, 136, 49-90.
- Fitzgerald, R.; Widdop, S.; Gray, M. y Collins, D. (2011). Identifying Sources of Error in Cross-national Questionnaires: Application of an Error Source Typology to Cognitive Interview Data. *Journal of Official Statistics*, 27, 569-599.
- Israel, G.F. (2010). Effects of Answer Space Size on Responses to Open-ended Questions in Mail Surveys. *Journal of Official Statistics*, 26, 271-285.

- Kreuter, F.; Müller, G. y Trappmann, M. (2010). Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data. *Public Opinion Quarterly*, 74, 880-906.
- Pasadas del Amo, S.; Trujillo Carmona, M.; Uribe-Echevarría, M. y Soria Zambrano, M. (2006). El impacto de la telefonía móvil en la cobertura de las encuestas telefónicas. *Metodología de Encuestas*, 8, 137-145.
- Pasadas del Amo, S.; Trujillo Carmona, M.; Sánchez Valenciano, A. y Cañadas Reche, L. (2011). La incorporación de las líneas móviles al marco muestral de las encuestas telefónicas: Pertinencia, métodos y resultados. *Metodología de Encuestas*, 13, 33-54.
- Pavía, J.M.; Larraz, B. y Montero, J.M. (2008). Election Forecasts Using Spatiotemporal Models. *Journal of the American Statistical Association*, 103, 1050-1059.
- Pavía, J.M.; Rausell, P.; Marco-Serrano, F. y Coll Serrano, V. (2011). Las nuevas tecnologías como instrumento de prospectiva electoral. Las elecciones a rector 2010 de la Universitat de València. *Metodología de Encuestas*, 13, 71-96.
- Pavía, J.M. y Larraz, B. (2012). Sesgo de no-respuesta y modelos de superpoblación en encuestas electorales. *Revista Española de Investigaciones Sociológicas*, 137, 121-150.
- Pavía, J.M. y López-Quilez, A. (2012). Spatial Vote Redistribution in Redrawn Polling Units. *Journal of the Royal Statistical Society-A*, in-press. DOI: 10.1111/j.1467-985X.2012.01055.x
- Pavía-Miralles, J.M. (2005). Forecasts from Non-Random Samples: The Election Night Case. *Journal of the American Statistical Association*, 100, 1113-1122.
- Pavía-Miralles, J.M. y Larraz-Iribas, B. (2008). Quick Counts from Non-Selected Polling Stations. *Journal of Applied Statistics*, 35, 383-405.
- Pérez López, C. (2009). *Técnicas de Muestreo Estadístico*. Garceta: Madrid.
- Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley and Sons: Hoboken, New Jersey.
- Rodríguez Osuna, J. (1991). *Métodos de Muestreo* [Colección Cuadernos Metodológicos del CIS, 1]. Centro de Investigaciones Sociológicas, Madrid.
- Rodríguez Osuna, J. (2005). *Métodos de Muestreo. Casos Prácticos* [Colección Cuadernos Metodológicos del CIS, 6]. Centro de Investigaciones Sociológicas, Madrid.
- Sakshaug, J.W.; Yan, T. y Tourangeau, R. (2010). Nonresponse Error, Measurement Error, and Mode of Data Collection: Tradeoffs in a Multi-mode Survey of Sensitive and Non-sensitive Items. *Public Opinion Quarterly*, 74, 907-933
- Särndal, C.-E. y Lundström, S. (2001). *Estimation in the Presence of Nonresponse and Frame Imperfections*. Statistics Sweden, Örebro.
- Statistics Canada (2010). *Survey Methods and Practices*. Minister of Industry: Ottawa.
- Verge Mestre, T. (2009). Deseabilidad social y género: el efecto del sexo del entrevistador(a) sobre las respuestas del encuestado(a). *Metodología de Encuestas*, 11, 27-45.

Weisberg, H.E. (2005). *The Total Survey Error Approach. A Guide to the New Science of Survey Research*. The University of Chicago Press: London.