

Métodos de imputación múltiple para predecir resultados electorales

Modesto Escobar
Universidad de Salamanca
modesto@usal.es

Antonio M. Jaime
Universidad de Málaga
amjaime@uma.es

1. Introducción

Según la conocida definición de Lewis-Beck (2005), pronosticar una elección significa anticipar el resultado antes de que ocurra. En este sentido, el pronóstico electoral no tiene esencialmente una finalidad de teorizar sobre los comportamientos que se predicen, sino la de establecer un pronóstico lo más ajustado posible a los resultados. Así pues, buena parte de la literatura se ha centrado en pronosticar el resultado de una elección a partir de magnitudes agregadas (*i. e.* Abramowitz 2010; Bafumi et al. 2010; Campbell 2010a,b; Cuzán 2010; Lewis-Beck y Tien 2010; Magalhaes et al. 2011; Whiteley 2008; Whiteley et al. 2011). Si bien tales modelos están informados por los micro-fundamentos del comportamiento electoral (Magalhaes et al. 2011), y su capacidad predictiva es generalmente buena (Lewis-Beck y Tien 1996), los pronósticos funcionan como una *caja negra* sobre la que no es posible hacer inferencias al nivel individual de los electores¹. En paralelo, la literatura sobre el comportamiento electoral ha estado centrada en la explicación de las decisiones de voto a nivel individual (*i. e.* Clarke et al. 2011; Fraile y Lewis-Beck 2010; Sanders et al. 2011), pero sus aplicaciones al pronóstico electoral son potencialmente limitadas, dado que su objetivo es explicar las decisiones electorales a posteriori pero no predecirlas. Estas limitaciones de los enfoques existentes justifican la motivación de este trabajo, cuyo principal objetivo es la utilización de encuestas electorales para el pronóstico de resultados electorales. Para ello usamos técnicas de imputación múltiple que, a diferencia de los enfoques metodológicos tradicionales, nos permiten extender el pronóstico a aquellos individuos para los que no disponemos de información completa, consiguiendo así una importante mejora en la capacidad predictiva de los modelos.

El trabajo ya clásico de Rubin (1987) puede considerarse un punto de partida en el tratamiento de la no respuesta en encuestas a través de técnicas de imputación múltiple. Sin embargo, no ha sido hasta tiempos recientes que Stata y otros paquetes estadísticos han incorporado métodos como el aumento de datos o las ecuaciones encadenadas (Royston 2009), especialmente indicadas para el caso de variables continuas². Por otra parte, el trabajo de King et al. (2001) se ha convertido en un punto de referencia en el ámbito de la Ciencia Política y ha motivado un incremento del interés por este tipo

1. Vid. una excepción temprana en Whiteley (1979) y más recientemente en los trabajos de Rivas et al. (2010) y Varela Mallou et al. (1998) aplicados también al caso español.

2. Una alternativa de software de imputación múltiple bastante popular es la rutina Amelia escrita para R por King et al. (2001).

de técnicas entre los investigadores en Ciencias Sociales. En la literatura en castellano, conviene destacar el cuaderno metodológico de Rivero (2011), que proporciona una guía general para el tratamiento de datos incompletos usando Stata, con un enfoque eminentemente práctico que permite su aplicación a muy diferentes campos de investigación. En este trabajo, utilizamos estos nuevos avances metodológicos, con el fin de obtener predicciones de las decisiones electorales individuales en contextos en los que la no respuesta sesga sistemáticamente los pronósticos. Así pues, nuestro enfoque combina una base teórica en los micro-fundamentos de la decisión electoral con la obtención de pronósticos a nivel agregado.

Con el fin de valorar la capacidad predictiva de este enfoque, lo hemos aplicado al caso concreto de las elecciones generales de 2011 en España. Para el análisis, hemos utilizado los datos de las encuestas pre y post electorales del Centro de Investigaciones Sociológicas que proporcionan información sobre muestras relativamente amplias del electorado español con información de las principales características y motivaciones de los electores. El resto del texto se organiza de la siguiente forma. En la siguiente sección, presentamos los modelos teóricos que justifican la selección de las variables que se utilizan para predecir la decisión electoral. A continuación, describimos los datos que hemos utilizado para llevar a cabo nuestro estudio y la metodología que hemos seguido para la elaboración de los pronósticos electorales. En la sección posterior, discutimos los principales resultados obtenidos. Y en la última sección, proponemos algunas conclusiones tentativas del estudio.

2. Marco teórico

En las últimas décadas se ha producido una extensa literatura sobre los diferentes factores que explican las decisiones electorales en España (Bosch y Riba 2005; De la Calle et al. 2010; Fraile 2005; Fraile y Lewis-Beck 2010; Jaime Castillo y Sáez Lozano 2001, 2007; Lago y Montero 2010; Lago-Peñas y Lago-Peñas 2005; Montero 1994; Montero et al. 2008). No obstante, tales estudios no se han ocupado específicamente del poder predictivo de las encuestas para anticipar los resultados electorales. Existe, igualmente, evidencia empírica de que el comportamiento electoral de quienes no responden a las preguntas de intención de voto no es homogéneo y la inclinación de su voto se ha dirigido mayoritariamente a partidos diferentes a lo largo del tiempo (Urquizu-Sancho 2006). Y esto último implica que cualquier pronóstico realizado a partir de la muestra de votantes con información incompleta es necesariamente sesgado. Ahora bien, podemos usar esta información sobre la relación entre las características de los electores con información completa y sus decisiones de voto para producir predicciones del comportamiento de los electores que no revelan su intención de voto. Para ello hemos de tomar en cuenta los modelos explicativos del comportamiento electoral que nos indican las variables relevantes para establecer predicciones sobre el voto. En términos generales, podemos agrupar tales modelos en tres grandes enfoques (Harrop y Miller 1987; Jaime Castillo y Sáez Lozano 2001). En el modelo de *identificación partidista*, el voto es consecuencia fundamentalmente de la adscripción ideológica aprendida a través del proceso de socialización. En el modelo de *elección racional* los votantes eligen al partido más cercano

a sus propios intereses, valores y prioridades. Y el *enfoque socio-estructural* enfatiza el impacto de la estructura social sobre el comportamiento electoral.

2.1. Teoría de la identificación partidaria

De acuerdo con la tesis de la *identificación partidaria*, el comportamiento de los votantes es explicado con base en su identificación con los diferentes partidos políticos. El origen último de esta identificación partidista a nivel individual está en los procesos de socialización política, a través de los cuales se desarrolla un complejo sistema actitudinal de identificación o aversión hacia los distintos elementos del sistema político, en este caso, los partidos (Abramson 1983; Campbell et al. 1960). Harrop y Miller (1987) resumen las proposiciones del modelo de identificación partidaria en cinco postulados fundamentales: a) La mayoría de los electores sienten una afinidad hacia un partido, la cual es heredada en el seno de la familia; b) La identificación partidista influye directamente en la determinación del voto, pero también indirectamente, a través de la influencia sobre la percepción de las políticas, los candidatos y los vínculos entre partidos y grupos sociales; c) El grado de identificación con un partido se correlaciona con la edad del votante, puesto que aquélla se refuerza con el paso del tiempo. Los cambios en la identificación son personales, ya que reflejan el cambio de grupos de pertenencia, a menudo relacionado con la movilidad social o geográfica; d) Los electores, aunque estén influidos por fuerzas de corto plazo en una campaña determinada a votar contra el partido con el que se identifican, mantienen la identificación con este partido y volverán a votarlo en posteriores elecciones; e) La distribución de las identificaciones partidistas entre un electorado provee la base para calcular el *voto normal*, esto es, el resultado electoral, si no intervinieran factores de corto plazo a favor de ningún partido.

A diferencia del modelo bipartidista norteamericano, en sistemas multipartidistas las identificaciones partidarias pierden peso como variable explicativa del comportamiento político, en favor de la posición ideológica derecha-izquierda. Según Ventura (2001), tanto la identificación partidaria como la posición ideológica derecha-izquierda deberían ser interpretados como heurísticos, a través de los cuales los individuos forman sus opiniones políticas. La diferencia entre uno y otro reside en el papel que cumplen a la hora de organizar el universo político. Mientras que la identificación partidaria es nominal (diferencia un partido del resto), la orientación ideológica es ordinal (crea un ranking de preferencias que ordena los partidos según su posición en el continuum ideológico). De esta forma, en sistemas bipartidistas, la orientación ideológica es un heurístico que tiene poca utilidad, puesto que se hace equivalente a una variable nominal dicotómica, mientras que en sistemas multipartidistas, el heurístico derecha-izquierda puede tener una gran utilidad debido a la complejidad del sistema de partidos, especialmente cuando este está altamente polarizado.

En su versión más simple, los votantes rara vez se dejan llevar por sus preferencias en torno a políticas públicas concretas o temas relevantes (*issues*) a la hora de decidir su voto, sino que votan por el partido con el que se identifican sin importarles demasiado las políticas que éste pretendiese poner en práctica (Campbell et al. 1960). Sin embargo, Key argumentó tempranamente que los *issues* tenían un papel bastante importante en

la determinación de la decisión de voto, esto es, que los electores podían sopesar la idoneidad de las políticas promovidas por los distintos partidos y adoptar una decisión electoral *racional* (Key 1966). En la medida en que las afinidades con los partidos se han hecho más débiles en comparación con las décadas de 1950 o 1960, otros factores como las políticas han ido ganando terreno en la determinación del voto (Nie et al. 1976). Al mismo tiempo, la evidencia para el caso español sugiere que existe un grupo importante de votantes *no ideológicos*, cuyo comportamiento electoral es difícil de predecir a partir de los enfoques tradicionales (De la Calle et al. 2010).

2.2. Teoría del votante racional

Los planteamientos que parten de esta perspectiva aplican el enfoque de la ciencia económica al estudio de la política. Bajo esta perspectiva, el votante se comporta como un maximizador de utilidad con motivaciones egoístas y la política se concibe como un mercado, en el que los partidos políticos compiten por los votos como medio para conseguir las recompensas derivadas de la obtención de representación parlamentaria. El objetivo de los partidos, por tanto, es maximizar su nivel de apoyo electoral, para lo cual utilizan la persuasión política. El votante, como actor racional, efectúa una decisión de voto en función de la información disponible sobre cada partido que se presenta a las elecciones, seleccionando aquel cuyo nivel de desempeño esperado le reporte una mayor utilidad. La versión original de la teoría racional del voto fue formulada por Downs (1957), el cual desarrolló un modelo de competición entre dos partidos, cuyas premisas eran las siguientes: a) Los votantes calculan la utilidad que han recibido del gobierno presente como suma de todos los beneficios obtenidos. b) Los votantes asumen que recibirán en el futuro la misma utilidad que en el pasado si ese gobierno se mantiene en el poder. Es decir, los votantes no juzgan sobre las promesas sino sobre el desempeño real de los gobiernos. c) Los electores también calculan la utilidad que recibirían si el partido en la oposición accediera al gobierno, lo cual es más difícil puesto que no pueden juzgar su desempeño real, pero es necesario para poder comparar a los partidos entre sí. d) Los electores, finalmente, calculan la diferencia entre los dos niveles de utilidad y votan por el partido que maximiza su bienestar.

El planteamiento original de Downs fue revisado posteriormente, entre otros, por Riker y Ordeshook (1968), para quienes los electores evaluaban tanto prospectiva como retrospectivamente las propuestas de los diferentes partidos o candidatos. Más tarde, Fiorina (1977) argumentó que la evaluación que los votantes hacen de los partidos no se limita al desempeño durante la última legislatura, sino que incluye todas las experiencias previas de gobierno, incluidas las que no son inmediatamente anteriores a la elección. Partiendo de estas premisas, la decisión de voto no es el reflejo de predisposiciones actitudinales o la expresión de la pertenencia a un grupo, sino que tiene un valor instrumental, como medio de satisfacción de intereses puramente individuales. De acuerdo, con la *teoría del voto económico*, el criterio básico por el que los votantes se decantan por un partido u otro es el desempeño económico esperado para los distintos partidos (Duch y Stevenson 2008; Fraile 2005; Fraile y Lewis-Beck 2010). En este mismo sentido, cabe esperar que los partidos en el gobierno manipulen las variables de política

económica para presentarse ante el electorado como exitosos y obtener la confianza de los ciudadanos, dando origen así, al llamado *ciclo electoral* de la economía (Williams 1990).

El punto de partida de la teoría del voto económico es la *hipótesis de la responsabilidad*, según la cual los votantes atribuyen los resultados económicos de un determinado periodo a la labor desempeñada por el partido en el gobierno. Por este motivo, recompensarán con su voto a los gobernantes que produzcan buenos resultados económicos y castigarán electoralmente a aquellos bajo cuyo periodo de mandato se produzcan recesiones económicas. Así pues, según este modelo, las variables que mejor predicen el comportamiento electoral son los indicadores económicos (Erikson 1989). Y los cambios de voto pueden explicarse en función de la evolución de determinadas magnitudes de la coyuntura económica (Bosch y Riba 2005), tales como el paro, la inflación, el déficit público o el nivel de imposición fiscal; o alternatively, en función de la percepción que los electores tienen de la coyuntura económica, dado que existe una asimetría de información entre electores y políticos³ (Fraile 2005). No obstante, las decisiones de voto no son un reflejo directo de las condiciones de la economía en un momento dado, en la medida en que los electores no siempre atribuyen la responsabilidad de la situación económica a los partidos en el gobierno (Jaime Castillo y Sáez Lozano 2007; Lago-Peñas y Lago-Peñas 2005). Al mismo tiempo, existen factores institucionales que facilitan o dificultan la claridad en la atribución de la responsabilidad a un partido concreto (Duch y Stevenson 2008), tales como los gobiernos de coalición o la superposición de niveles de gobierno. Adicionalmente, los votantes pueden preocuparse también por otras dimensiones de la gestión política no estrictamente económicas. En el caso español, por ejemplo, existe evidencia de que el nacionalismo es una dimensión relevante del comportamiento electoral que atenúa la aplicación de la regla premio-castigo cuando un partido nacionalista está en el gobierno (Alonso 2008; Aguilar y Sánchez-Cuenca 2007).

2.3. Teoría socio-estructural del voto

Bajo esta denominación se pueden agrupar una serie de enfoques que subrayan la importancia de los factores socio-estructurales como variables explicativas del comportamiento electoral. Según este planteamiento, el voto está determinado por la posición del votante en la estructura social y los grupos con características sociales similares deben tener un comportamiento electoral relativamente homogéneo. La razón última es que los miembros de cada grupo comparten una serie de intereses, cuya defensa en la arena política corresponde al partido que mejor representa esos intereses colectivos (Dunleavy y Husbands 1985; Harrop y Miller 1987). Estos grupos sociales se pueden diferenciar en base a variables como el género, la pertenencia a una clase social o grupo étnico o la edad. Dado que los partidos se conciben como medio de representación de los intereses del grupo, la base social de éstos estará formada, por tanto, por aquel conjunto de electores cuyos intereses expresan. El proceso de formación de grupos sociales se explica históricamente a partir de los conflictos de intereses presentes en cada sociedad

3. Esta asimetría se produce porque los electores tienen menor acceso que los políticos a la información sobre el estado real de la economía.

(*cleavages*), los cuales provocan la unión para la acción política (Lipset y Rokkan 1967). Sin embargo, estas divisorias sociales tienden a perpetuarse en el tiempo, lo que hace que el comportamiento electoral, y el propio sistema de partidos, sean relativamente estables.

No obstante, Flanagan y Dalton (1994) argumentaban que, con el paso del tiempo las escisiones sociales habían perdido importancia en la explicación del *anclaje electoral*, o en otros términos, que las fisuras sociales sobre las que se basaba el sistema de partidos en las sociedades occidentales se habían desdibujado casi por completo en las últimas décadas. Es cierto que tras la Segunda Guerra Mundial la fractura social sobre la que se articulaba la contienda política era la que dividía los partidos de la clase trabajadora de los partidos burgueses. Sin embargo, los cambios socioeconómicos asociados con el post-industrialismo, como el post-fordismo, han hecho que esta fractura social quede cada vez menos definida. El mismo Lipset llegaba décadas más tarde a la conclusión de que la clase social era un factor cada vez menos importante en la explicación del comportamiento electoral (Clark y Lipset 1991). Según apunta la evidencia empírica, en el periodo posterior a la II Guerra Mundial se ha producido una caída sustancial en los niveles de voto de clase en la mayoría de los países occidentales (Nieuwbeerta 1996; Nieuwbeerta y Ultee 1999). Esta erosión del voto de clase ha sido más intensa en donde históricamente había tenido una mayor importancia: en los países escandinavos. El propio concepto de *cleavage* también ha sido objeto de imputaciones críticas. En la versión originaria de Lipset y Rokkan, *cleavage* denota la existencia de un conflicto en la sociedad, cuya articulación política corresponde a los partidos. Sin embargo, con la caída del voto de clase en las sociedades avanzadas, muchos autores han tratado de buscar nuevas escisiones sociales sobre las que vertebrar el comportamiento electoral. Es el caso de Kriesi (1998), quien ha tratado de hacer una relectura de la obra de Lipset y Rokkan en función de lo que él considera una nueva fractura social: el debate entre valores materialistas y post-materialistas.

En cualquier caso, la estructura social sigue siendo una variable clave para la explicación del comportamiento electoral y varios estudios recientes han puesto en duda el declive de su capacidad explicativa del voto (Evans 2000; Manza et al. 1995; Van der Waal et al. 2007). En el caso español, la relación entre clase social ha sido hallada empíricamente en diferentes trabajos por Caínzos (2001) o González (2004), si bien Polavieja (2001) ha argumentado que los mecanismos causales detrás del voto de clase puede deberse a factores no económicos como la socialización política. Del mismo modo, Montero et al. (2008) han hallado que la religión sigue siendo un factor importante en la orientación del comportamiento electoral de los españoles. Por estos motivos, debemos esperar que las variables socio-estructurales tengan un efecto sobre las decisiones de voto individuales, y por tanto, pueden ayudar a predecir el comportamiento de los electores.

3. Datos y metodología

3.1. Datos

Los datos empleados en el presente capítulo proceden del Centro de Investigaciones Sociológicas (CIS), un organismo autónomo dependiente del Ministerio de la Presidencia del Gobierno del Reino de España, cuya función principal es la de contribuir al conocimiento científico de la sociedad española. Este instituto de investigación gubernamental realiza periódicamente una serie de encuestas, entre las que en este contexto destacan los barómetros de opinión, pero sobre todo las encuestas pre-electorales y post-electorales. De hecho, se emplearán las encuestas que se diseñaron para el estudio de las Elecciones Generales que se celebraron el 20 de noviembre de 2011.

Como es bien sabido, España es una democracia parlamentaria con unas Cortes Generales compuesta por dos cámaras renovables cada cuatro años, a menos que el Presidente de Gobierno use su prerrogativa de disolverlas para convocar elecciones. En 2011, José Luis Zapatero, en el epicentro de una crisis económica y política de gran envergadura y en una legislatura con mayoría relativa, que impidió la confección de los presupuestos del 2012, decidió adelantar la celebración de los comicios cuatro meses, a Noviembre de 2011, en lugar de Marzo de 2012.

El estudio pre-electoral de las mencionadas elecciones que llevó a cabo el CIS tuvo lugar entre el 6 y el 29 de octubre, es decir, con un mes aproximado de antelación al día de las elecciones, que fueron celebradas el 20 de noviembre. Utilizó una muestra de 17.397 entrevistas planificadas, de las que se realizaron un total de 17.236. Los puntos de muestreo fueron 1.098 municipios. El muestreo fue polietápico. La primera unidad de selección fueron los municipios estratificados por circunscripción electoral y tamaño de hábitat. Las unidades primarias se seleccionaron mediante rutas aleatorias con cuotas de sexo y edad. El error muestral, tal como consta en la ficha técnica del estudio, fue del $\pm 0.76\%$ para un nivel de confianza del 95.5% y $p=q=50\%$, bajo el supuesto del muestreo aleatorio simple⁴.

La muestra del estudio post-electoral fue diseñada a partir de la pre-electoral, al tratarse de un diseño panel, si bien solo se contó con aquellas personas que no manifestaron expresamente su deseo de no ser entrevistadas en una segunda ocasión. El número de éstas fue de 7.547 y la cantidad final de entrevistas registradas fue de 6.082.

3.2. Variables

La variable imputada principal de este trabajo es la conducta electoral, obtenida principalmente a través de las preguntas sobre intención de voto en la encuesta pre-electoral y sobre la conducta de voto en las mismas elecciones. Las razones de ello son obvias: por un lado, porque se persigue la predicción de los resultados electorales,

4. La información detallada del diseño no está disponible por razones de anonimización de los entrevistados, por lo que en la estimación solamente se ha tenido en cuenta la estratificación por provincia y los pesos correspondientes. No obstante, debe tenerse en cuenta que esto puede subestimar los errores de estimación.

especialmente mediante el estudio pre-electoral y, por otro lado, porque también nos preguntamos por los determinantes de la conducta electoral.

Aunque la pregunta sobre la intención de voto en la encuesta pre-electoral contiene 31 opciones distintas de respuestas por la gran variedad de partidos que se presentan a los comicios, a los efectos de este estudio se ha reducido a cuatro valores, además de los considerados como perdidos por la no contestación o la indecisión del entrevistado. Estos cuatro valores son los tres primeros partidos, a saber, 1) el Partido Popular (PP), partido vencedor (44.3% del voto válido), que representa a la inmensa mayoría de la derecha española, conservadores, liberales y demócrata cristianos incluidos, 2) el Partido Socialista Obrero Español (PSOE) con un 28.8%, cuyo gobierno de centro izquierda moderada fue el que convocó las elecciones y resultó desalojado del poder tras perder las elecciones, 3) Izquierda Unida (IU), coalición que consiguió un 6.9%, aglutinando distintas fuerzas y minúsculos partidos situados a la izquierda de los socialistas en el arco político y 4) el resto de partidos (19.7%), compuesto principalmente por Unión Progreso y Democracia (UPyD), un partido de fundación reciente, cuya principal líder proviene del PSOE, y por los nacionalistas, entre los que sobresalen Convergencia i Unió (CiU) y Esquerra Republicana de Catalunya (ERC) en Cataluña, el Partido Nacionalista Vasco (PNV) en el País Vasco, Coalición Galega (CG) en Galicia y Coalición Canaria (CC) en Canarias. Todos ellos, excepto Esquerra Republicana de Catalunya, tienen orientación ideológica de centro-derecha.

De esta pregunta se ha de descartar a los que no tienen intención de votar, pues a ellos no hace falta imputarles el voto (por tanto serán considerados perdidos no imputables); sólo se ha de hallar aquellos que, aun diciendo que van a votar, no saben o no quieren contestar su opción política. Las variables auxiliares que se han empleado para realizar esta imputación son las siguientes:

a) La ideología, que es un poderoso predictor de la conducta de voto, como se ha visto en la revisión teórica, también ha sido sometida a un proceso de recodificación para evitar categorías excesivamente pequeñas. De este modo, en general, se han reducido las 10 categorías en 5 (juntando dos categorías contiguas). De modo alternativo, cuando la ideología ha sido considerada también variable imputada, se ha considerado como ordinal con sólo tres valores: izquierda (con los tres primeros valores de la escala), centro (con los cuatro centrales, 4-7) y derecha (con los tres últimos).

b) El recuerdo de voto que, aunque tiene una importante similitud con la variable de la post-estratificación, puede considerarse como un magnífico predictor del voto futuro. En la encuesta post-electoral se preguntó por los votos emitidos tanto en las elecciones del 2011, como en las del 2008. El primero ha sido utilizado como variable-respuesta. Sin embargo, no se empleó el recuerdo de voto de 2008 en la encuesta post-electoral, pues se prefirió tomar el dato de la pre-electoral, al poseer menos perturbación por unas elecciones intermedias, haber mayor nivel de respuesta y, por el hecho de que la influencia sobre el voto de su recuerdo es *a priori* y no *a posteriori*.

c) El nivel de estudios, que ha sido considerado como una variable ordinal y que se toma como *proxy* del estatus del sujeto.

d) Tanto el sexo como la edad son variables con muy bajo número de valores perdidos, aunque con tenue influencia sobre las variables imputadas.

Además de estas variables, también se contemplan en los modelos ampliados dos variables adicionales:

a) El gobierno, o valoración de la actividad del gobierno, originalmente medida en cinco categorías, recodificadas en tres: buena más muy buena, regular, y mala más muy mala. Como en la encuesta post-electoral no se incluyó la valoración del gobierno saliente, se ha empleado la información de la pre-electoral, operación posible en la medida en que se trataban de dos estudios con formato de panel.

b) La economía, o valoración de la situación económica del país. También fue originalmente medida en una escala de cinco puntos y transformada en una de tres con valores positivos, neutros y negativos. De modo similar también a la valoración de la actividad del gobierno, se ha empleado el dato de la pre-electoral, en la medida en que en el voto influye la evaluación económica de antes de las elecciones y no la de después.

En el Cuadro 1, aparece la estructura de los valores perdidos de la encuesta pre-electoral y en el Cuadro 2 la estructura de los valores perdidos de la encuesta post-electoral. En la primera fila de cada cuadro se presenta el porcentaje de casos perdidos en cada una de las variables de columna y en las sucesivas filas se presenta el porcentaje de casos perdidos (última columna) para cada uno de los patrones de datos perdidos⁵. En la encuesta pre-electoral se dispone del 61 % de casos con información completa para el conjunto de las cinco variables consideradas, como se observa en la última fila del Cuadro 1, mientras que este porcentaje es del 73 % en la encuesta post-electoral (última fila del Cuadro 2). Adicionalmente, se detecta una diferencia considerable entre ambos conjuntos de datos, como se puede apreciar al comparar la primera fila de los Cuadros 1 y 2. En la pre-electoral, como cabía esperar, la variable con más valores omitidos es la intención de voto, principalmente achacable a los indecisos. En cambio, en la encuesta post-electoral, la variable con menos información es la valoración del gobierno, en la medida en que ha sido preguntada como contrafactual (¿qué hubiera pasado si hubiera sido otro el partido el responsable de la marcha del país antes de las elecciones?), en lugar de ser una opinión directa sobre la labor del ejecutivo. Esto hace que el porcentaje de casos sin información se eleve a casi la cuarta parte de la muestra.

Para realizar las imputaciones electorales y poder realizar comparaciones entre ambas encuestas, se seleccionaron las variables independientes señaladas: sexo, edad y nivel de estudios como datos demográficos; ideología y recuerdo de voto en las últimas elecciones (2008) como información de índole política. Las primeras poseían un bajo nivel de valores perdidos, mientras las segundas presentaban una alta tasa de no respuesta tal como aparecen las frecuencias de datos incompletos en los Cuadros 1 y 2.

5. Cada fila representa un patrón de datos perdidos, el cual se define por una combinación de 1 (valor faltante) y 0 (valor presente). Así pues, la primera fila representa el caso de valores faltantes para todas las variables y la última fila representa el caso en el que no hay valores faltantes en ninguna variable

Cuadro 1: Patrón de casos perdidos. Encuesta pre-electoral

Voto	Ideología	Recuerdo	V.Gobierno	V.Economía	Porcentaje
6.3 %	12.5 %	25.0 %	50.0 %	100.0 %	
1	1	1	1	1	<1
1	1	1	1	0	<1
1	1	1	0	1	<1
1	1	1	0	0	3
1	1	0	1	1	<1
1	1	0	1	0	<1
1	1	0	0	1	<1
1	1	0	0	0	3
1	0	1	1	1	<1
1	0	1	1	0	<1
1	0	1	0	1	<1
1	0	1	0	0	5
1	0	0	1	1	<1
1	0	0	1	0	<1
1	0	0	0	1	<1
1	0	0	0	0	11
0	1	1	1	1	<1
0	1	1	1	0	<1
0	1	1	0	1	<1
0	1	1	0	0	1
0	1	0	1	1	<1
0	1	0	1	0	<1
0	1	0	0	1	<1
0	1	0	0	0	8
0	0	1	1	1	<1
0	0	1	1	0	<1
0	0	1	0	1	<1
0	0	1	0	0	4
0	0	0	1	1	<1
0	0	0	1	0	<1
0	0	0	0	1	<1
0	0	0	0	0	61

(*)El 1 indica un valor perdido en la variable de la columna. El 0 indica la presencia de un valor.
Las cifras de la primera línea son los porcentajes de datos perdidos en la variable de la columna.

Cuadro 2: Patrón de casos perdidos. Encuesta post-electoral

Ideología	Recuerdo	Voto	V.Gobierno	V.Economía	Porcentaje
6.3 %	12.5 %	25.0 %	50.0 %	100.0 %	
1	1	1	1	1	0
1	1	1	1	0	<1
1	1	1	0	1	<1
1	1	1	0	0	1
1	1	0	1	1	0
1	1	0	1	0	<1
1	1	0	0	1	0
1	1	0	0	0	1
1	0	1	1	1	<1
1	0	1	1	0	<1
1	0	1	0	1	0
1	0	1	0	0	2
1	0	0	1	1	<1
1	0	0	1	0	<1
1	0	0	0	1	<1
1	0	0	0	0	9
0	1	1	1	1	0
0	1	1	1	0	<1
0	1	1	0	1	<1
0	1	1	0	0	2
0	1	0	1	1	<1
0	1	0	1	0	<1
0	1	0	0	1	<1
0	1	0	0	0	5
0	0	1	1	1	<1
0	0	1	1	0	<1
0	0	1	0	1	<1
0	0	1	0	0	3
0	0	0	1	1	<1
0	0	0	1	0	1
0	0	0	0	1	<1
0	0	0	0	0	73

(*)El 1 indica un valor perdido en la variable de la columna. El 0 indica la presencia de un valor. Las cifras de la primera línea son los porcentajes de datos perdidos en la variable de la columna.

3.3. Métodos muestrales de estimación

Se han utilizado dos métodos diferentes de estimación⁶. El primero de ellos es esencial al estar trabajando con encuestas obtenidas con muestras polietápicas, ya que éstas

6. El modo de trabajar con estos modelos de encuesta con Stata puede encontrarse en el volumen dedicado a la imputación múltiple (Stata 2011b). También puede consultarse en español un resumen

requieren que los estimadores estén ponderados adecuadamente, cuando el muestreo estratificado se ha realizado con una afijación no proporcional. Estos pesos, que han de ser iguales para todos los entrevistados dentro de una misma circunscripción, se obtienen mediante la fórmula:

$$w_k = C_k/n_k$$

Siendo C_k el número de votantes según el censo electoral y n_k el número de entrevistas realizadas en cada estrato (provincia o circunscripción).

El segundo método de estimación es la llamada post-estratificación de la muestra. Consiste en que, una vez es ésta obtenida, se coteja la distribución de una o varias de sus variables para ver si coinciden con los parámetros del universo y se da unos pesos a cada caso que hacen que los resultados muestrales coincidan con los poblacionales. El coeficiente de post-estratificación se obtendría comparando el tamaño de la muestra, una vez ponderada con el procedimiento anterior, en cada post-estrato con el correspondiente en la población.

Siendo N_l el tamaño real de un determinado post-estrato de la población (l) y \widehat{N}_l el tamaño estimado de la población después de aplicar la estratificación ($n_l w_k N/n$), el nuevo coeficiente de ponderación para cada individuo en el estrato k y el post-estrato l sería:

$$w_{kl} = w_k N_l / \widehat{N}_l$$

Con estos coeficientes de ponderación la estimación de las proporciones ponderadas (\hat{q}) a partir de las iniciales estimaciones \hat{p}_{kl} en el estrato k y el postestrato l adoptaría la siguiente expresión:

$$\hat{q} = \sum_{k=1}^K \sum_{l=1}^L w_{kl} \hat{p}_{kl}$$

En los estudios pre-electorales es común usar el recuerdo de voto como criterio de post-estratificación por los siguientes motivos: En primer lugar, porque hay determinados grupos sociales con menos probabilidad de ser incluidos en la muestra por los filtros sociales que dificultan su accesibilidad, como es el caso de las personas de alto estatus, que a su vez tienden a ser más conservadores. En segundo lugar, porque hay un proceso de auto-selección en las encuestas que se presume es de trascendencia para el estudio de la conducta electoral, pues las personas que no declaran su ideología no votan igual que quienes sí declaran su ideología (De la Calle et al. 2010). Finalmente, existe lo que los psicólogos llaman deseabilidad social, según la cual, las personas tienden a dar respuestas que no sean mal vistas por los otros. Por ejemplo, en España, el voto al PP aparece habitualmente infrarrepresentado en las encuestas por razones históricas, dada la orientación ideológica de derecha que tuvieron regímenes no democráticos precedentes como

en el capítulo 14 de Escobar et al. (2012).

el franquismo. Este sesgo se solventa generalmente mediante la post-estratificación del recuerdo de voto en las elecciones anteriores, operación que se someterá a prueba en este trabajo.

3.4. Modelos de imputación

Se ha empleado el método propuesto por Rubin (1987) para el análisis e imputación de datos incompletos, que consiste en reconstruir nuevos conjuntos de datos, tantos como el investigador establezca, con valores simulados aleatoriamente por otras variables del estudio que contengan mayor información⁷. Frente a la imputación única, que consiste en estimar una sola vez los datos, la imputación múltiple realiza una serie de estimaciones -mediante simulación de un número de conjuntos de datos completos- a partir de los cuales se puede reconstruir una única estimación, complementada con la variación de las diversas estimaciones realizadas. En consecuencia, las varianzas de los parámetros pueden obtenerse de modo más certero que con la imputación única.

De lo que se trata es de obtener m estimadores de datos completos \hat{Q}_i , a partir de los cuales se obtendría un único estimador \bar{Q} (promedio de \hat{Q}_i). De igual modo, en cada estimación se puede obtener una matriz de varianza (varianza interna, \bar{U}) como promedio de las matrices de varianzas estimadas en cada imputación, así como la matriz de varianza de las distintas estimaciones obtenidas (varianza externa, \mathbf{B}). La varianza total \mathbf{T} se obtiene a partir de la suma de la primera y la segunda multiplicada por $(1 + 1/m)$.

Estadísticos derivados de estos estimadores básicos a tener en cuenta son el incremento relativo de la varianza debido a la no respuesta (RVI), la mayor fracción de la información perdida (FMI) y la eficacia relativa (RE). Mientras menores sean los primeros y mayor el último, mejor la calidad de las imputaciones realizadas.

Existen diferentes procedimientos de imputación para completar cada uno de los conjuntos de datos con valores perdidos. Están basados en principios bayesianos y frecuentistas. Básicamente, pueden distinguirse las imputaciones univariadas (una sola variable al mismo tiempo) a partir de la distribución predictiva posterior de los datos perdidos). Dado que la variable de interés para la imputación es el voto (x_i), se empleó un modelo multinomial en el que las probabilidades de las categorías de la variable respondían a la fórmula multinomial:

$$\Pr(x_i = k | \mathbf{z}_i) = \begin{cases} \frac{1}{1 + \sum_{l=2}^K \exp(\mathbf{z}'_i \beta_l)}, & \text{si } k = 1 \\ \frac{\exp(\mathbf{z}'_i \beta_l)}{1 + \sum_{l=2}^K \exp(\mathbf{z}'_i \beta_l)}, & \text{si } k > 1 \end{cases} \quad (1)$$

Donde \mathbf{z}_i es un vector (qx1) de predictores de x_i y β_l son vectores de parámetros desconocidos para $l = 2, \dots, K$.

7. Una introducción clara y básica junto con el modo de obtener estos modelos con Stata puede encontrarse en el volumen dedicado a la imputación múltiple (Stata 2011a). Asimismo, una presentación teórica y aplicada en español se encuentra en el ya citado libro de Rivero (2011).

En primer lugar se obtienen las estimaciones máximo-verosímiles, $\hat{\beta} = (\hat{\beta}'_2, \dots, \hat{\beta}'_K)$, y su matriz de varianzas asintótica \hat{U} , sólo para aquellos casos de la muestra con información completa. A partir de esta primera estimación se simulan nuevos parámetros para cada imputación usando la aproximación normal asintótica, $N(\hat{\beta}, \hat{U})$, a su distribución posterior asumiendo una distribución a priori no informativa de los parámetros.

En segundo lugar, se imputan los valores de los x_i perdidos simulando con el modelo multinomial, tantas veces (m) como se considere conveniente, para lo cual también habrá que simular en el paso anterior tantos vectores de parámetros como de imputaciones se requieran. Todo ello corresponde aproximadamente a extracciones aleatorias de la distribución posterior de los datos perdidos.

De mayor complejidad son las imputaciones multivariadas (en las que se imputan los valores perdidos de más de una variable en el mismo algoritmo). En este caso se opta por el método de ecuaciones encadenadas, que emplea un algoritmo similar al de Gibbs para obtener los valores imputados (Royston 2009). Este procedimiento, que permite especificar las ecuaciones de predicción, realiza un número especificado de iteraciones para obtener un conjunto final de imputaciones de cada variable en la última iteración (con un orden que va desde la más observada a la menos observada) a partir del resto de variables imputadas y de aquellas consideradas regulares (que no poseen casos perdidos) con los parámetros (θ_i), provenientes de distribuciones a priori uniformes específicas para cada densidad de predicción.

$$\begin{aligned}
 & f_i(x_i | x_1, x_2, x_3, \dots, x_{i-1}, \theta_i) \\
 & f_{i-1}(x_{i-1} | x_1, x_2, x_3, \dots, x_{i-2}, x_i, \theta_{i-1}) \\
 & \dots \\
 & f_1(x_1 | x_2, x_3, x_4, \dots, x_i, \theta_1)
 \end{aligned} \tag{2}$$

En este modelo de imputación multivariado, las variables imputadas consideradas han sido, además del voto, el recuerdo de voto y la ideología, ya que ambas suelen presentar un considerable porcentaje de no respuesta. En el caso de la ideología, así como de las evaluaciones del gobierno y la economía, consideradas las tres como variables ordinales, se ha optado por un modelo logístico ordinal en lugar del múltiple seleccionado para las dos variables relacionadas con el voto:

$$\begin{aligned}
 \Pr(x_i = k | \mathbf{z}_i) &= \Pr(\gamma_{k-1} < \mathbf{z}'_i \boldsymbol{\beta} + u \leq \gamma_k) \\
 &= 1 / (1 + \exp(-\gamma_k + \mathbf{z}'_i \boldsymbol{\beta})) - 1 / (1 + \exp(-\gamma_{k-1} + \mathbf{z}'_i \boldsymbol{\beta}))
 \end{aligned} \tag{3}$$

Donde x_i es una variable ordinal con K categorías, \mathbf{z}_i representa los valores de los predictores de \mathbf{x} , $\boldsymbol{\beta}$ es el vector de los coeficientes de regresión y γ_k son los puntos de corte que sirven para separar las K categorías ordinales con valores comprendidos entre $\gamma_0 = -\infty$ y $\gamma_K = \infty$.

3.5. Evaluación

Para evaluar el resultado de estimaciones e imputaciones se hace uso del error absoluto medio ponderado (EAMP), empleado en la literatura de pronóstico en las series temporales (Makridakis et al. 1998; Hyndman y Koehler 2006; Diebold 2007). En este contexto se hará uso de él para analizar el desempeño de las de las diferentes especificaciones de los métodos de imputación sobre las distribuciones de la variable respuesta, es decir, del voto.

Para la evaluación se tomarán como proporciones reales los resultados de las elecciones al Congreso de Diputados, resumidos en cuatro proporciones: la de los tres primeros partidos, es decir PP, PSOE e IU, más una cuarta categoría con el voto al resto de partidos. Por otro lado, para cada encuesta, ponderación, modelo y método, también se dispondrá de cuatro r_k (proporción de votos *obtenidos* por el partido k en una determinada elección), que pueden ser comparadas bien con las respectivas cuatro o_k (proporciones observadas de voto al partido k *sin imputación*⁸), bien con las cuatro c_k (proporciones completadas de voto al partido k *con imputación*).

En consecuencia, son de interés para cada una de las 16 condiciones que se generan (2 encuestas, 2 modos de ponderación, 2 modelos *-simple*, que incluye variables ideológicas y socio-estructurales, y *ampliado*, que añade otras variables relacionadas con el voto racional- y 2 métodos de imputación, a saber: *univariante* y *encadenado*) las tres medidas siguientes:

$$\varepsilon_{or} = \sum_{k=1}^K |o_k - r_k| r_k \quad (4)$$

$$\varepsilon_{cr} = \sum_{k=1}^K |c_k - r_k| r_k \quad (5)$$

$$\varepsilon_{oc} = \sum_{k=1}^K |o_k - c_k| r_k \quad (6)$$

Las dos primeras permiten comparar qué sistema (la estimación sin imputación o cualquier estimación con imputación) se ha aproximado mejor al resultado real, mientras la tercera fórmula indica cuán diferente es la estimación con imputación de la estimación que se hiciera sin ella.

4. Resultados

En esta sección presentamos sumariamente los principales resultados obtenidos en la investigación. En primer lugar, discutiremos la precisión de cada uno de los modelos de imputación aplicados y a continuación el efecto de las variables sobre la predicción del voto. Para organizar la discusión, vamos a centrar el análisis en la comparación de los diferentes procedimientos de estimación que hemos descrito en la sección prece-

8. Las proporciones de voto estimado *sin imputación* han sido calculadas restringiendo el análisis a los casos completos

dente. En primer lugar, hemos de distinguir entre los modelos univariantes (en los que únicamente se imputa la decisión de voto a partir de la información contenida en las variables explicativas) frente a los modelos encadenados (en los que de modo iterativo se atribuyen valores a los datos omitidos tanto de la variable respuesta, como de las variables explicativas). Dentro de cada uno de estos procedimientos de imputación distinguiremos igualmente entre los modelos que utilizan estratificación por provincia y los que utilizan estratificación por provincia y post-estratificación por recuerdo de voto (post-estratificación). Estos últimos devuelven a cada elemento de la muestra su peso correspondiente de acuerdo con la distribución del voto en la elección anterior (celebrada en 2008). Finalmente, dentro de cada uno de estos procedimientos de ponderación utilizamos dos especificaciones de modelos de imputación: uno *simple*, en el que las variables explicativas son las variables socio-estructurales (sexo, edad y nivel de estudios), la ideología y el recuerdo de voto y un modelo *ampliado* en el que además incluimos como predictores la evaluación de la situación económica y la valoración de la gestión del gobierno ⁹. Esto nos da un total de ocho especificaciones de modelos que producen distintos pronósticos sobre el resultado electoral. Para cada una de esas ocho especificaciones, el pronóstico producido por el procedimiento con imputación es comparado con la predicción producida a partir de un modelo estimado sin imputación. Este modelo de voto estimado (sin imputación) se corresponde con el pronóstico electoral que se obtiene de la muestra cuando se utilizan únicamente aquellos casos con información completa. Esta comparación es de gran importancia, ya que nos permite ver en qué medida las técnicas de imputación contribuyen a una mejora de nuestra capacidad predictiva en cada uno de los modelos estimados, lo cual es, en última instancia, el objeto que motiva este trabajo.

Dado que nuestro objetivo central consiste en la obtención de pronósticos electorales, el análisis de los resultados se centrará fundamentalmente en los resultados obtenidos con los datos de la encuesta pre-electoral. No obstante, con el objeto de evaluar la capacidad predictiva teórica de nuestros modelos con imputación, los hemos aplicado igualmente al análisis de la encuesta post-electoral. Si bien en este caso no se puede hablar propiamente de capacidad predictiva, el objetivo de comparar las predicciones que se producen utilizando ambas encuestas es el de averiguar si la capacidad predictiva teórica mejora al utilizar una encuesta que ha de producir una distribución de resultados más cercana al resultado real de la elección, habida cuenta de que incorpora el efecto de la campaña en la estimación del resultado y el porcentaje de datos completos es considerablemente mayor al eliminarse la condición de indeciso. Téngase también presente que en la encuesta pre-electoral se pregunta por la intención de voto mientras que en la encuesta post-electoral se pregunta por el partido al que se votó efectivamente en la elección. Es posible, por tanto, que votantes con intención de votar a un partido deter-

9. En el manual de imputación múltiple de Stata, se sugiere que, cuando se trata de encuestas, se incluyan también en los modelos de imputación las variables estructurales, como los pesos, los estratos y los identificadores de conglomerados (Stata 2011a, p. 8). La experiencia con estos datos y diversas pruebas realizadas, así como consideraciones teóricas sustantivas, nos han aconsejado incluir sólo la variable recuerdo de voto, esto es la variable de post-estratificación. La introducción del peso de los estratos provinciales apenas hacía variar los estadísticos de evaluación en una décima y la introducción del peso de los post-estratos provocaba un empeoramiento sustancial en los estadísticos que comparaban los datos completados con los reales

minado acaben no votándolo, cosa que no debería ocurrir en una encuesta post-electoral en la que la única incógnita desde el punto de vista de la predicción consiste en asignar la opción de voto a quienes no dicen a qué partido votaron.

4.1. Capacidad predictiva de la imputación

Las predicciones de resultados obtenidos en los diferentes modelos aplicados a las encuestas pre y post-electoral se presentan en los Cuadros 3 y 4. La primera columna de cada tabla contiene los datos de voto real a cada partido en la elección. La columna Obs. (datos observados) en cada uno de los modelos hace referencia a la proporción de voto cada partido que se obtiene a partir de la estimación sin imputación, mientras que la columna Comp. (datos completados) refleja las proporciones de voto obtenidas a partir de la estimación con imputación. En cada caso presentan los resultados de los métodos univariantes y encadenados para los modelos simple y ampliado con y sin post-estratificación. Asimismo, a partir de la información contenida en estos Cuadros se ha obtenido el Cuadro 5 en el que se presentan las diferencias obtenidas entre los diferentes métodos de estimación (con y sin imputación) y los datos observados y reales y que ha sido definidas en el último epígrafe de la sección precedente. La medida O-R contiene el error absoluto medio entre datos observados y reales. La medida C-R refleja el error absoluto medio entre datos reales y completados. Y la medida C-O es la desviación absoluta media entre datos observados y completados. En todos los casos, las medidas de error se calculan ponderando por la proporción de voto de cada partido.

Conviene destacar en primer lugar que la predicción de los resultados electorales que se obtiene con imputación es notablemente precisa. El modelo que mejor predice el resultado (el univariante sin post-estratificación) produce un error medio por partido de 0.70% en la encuesta pre-electoral, es decir, menos de un 1% con respecto al resultado real (primera fila del Cuadro 5). Adicionalmente, se observa que en la encuesta pre-electoral el pronóstico con imputación mejora de forma considerable (un 53%) la capacidad predictiva con respecto al pronóstico sin imputación (utilizando solo datos con información completa). Por tanto, cabe concluir que el uso de técnicas de imputación puede ayudar notablemente a la obtención de pronósticos electorales precisos. No obstante, hay una situación en la que el pronóstico basado en la estimación sin imputación es mejor que el pronóstico basado en la estimación con imputación y es aquella en la que se utiliza el recuerdo de voto como criterio de post-estratificación. Por ejemplo, si comparamos el modelo univariante simple de la encuesta pre-electoral sin post-estratificación con el que sí la incorpora vemos que el error medio en la estimación con imputación (la medida C-R) crece desde 0.70% a 2.10%, mientras que el error medio de la estimación sin imputación (la medida O-R) crece desde 1.30% a 2.00%. Cambios similares pueden observarse al comparar cualquier otro modelo con y sin post-estratificación por recuerdo de voto, tanto si nos fijamos en la encuesta pre-electoral como en la encuesta post-electoral.

La explicación a esta aparente paradoja viene del hecho de que al pos-estratificar por recuerdo de voto también estamos perdiendo una información importante: la de aquellos que no votaron en las elecciones anteriores y a los que, por tanto, no se puede

Cuadro 3: Predicción de resultados electorales. Encuesta pre-electoral

Modelos univariantes									
Voto	Real	Modelo simple				Modelo ampliado			
		No post-est.		Post-estrat.		No post-est.		Post-estrat.	
		Obs.	Comp.	Obs.	Comp.	Obs.	Comp.	Obs.	Comp.
PP	44.6	45.6	43.9	47.8	48.0	45.6	43.9	47.8	48.0
PSOE	28.8	26.7	28.4	27.6	27.7	26.7	28.4	27.6	27.7
IU	6.9	6.7	6.6	5.9	5.8	6.7	6.6	5.9	5.8
Otros	19.7	20.9	21.1	18.7	18.5	20.9	21.1	18.7	18.5

Modelos encadenados									
Voto	Real	Modelo simple				Modelo ampliado			
		No post-est.		Post-estrat.		No post-est.		Post-estrat.	
		Obs.	Comp.	Obs.	Comp.	Obs.	Comp.	Obs.	Comp.
PP	44.6	45.6	43.7	47.8	47.9	45.6	44.1	47.8	48.2
PSOE	28.8	26.7	28.5	27.6	27.6	26.7	27.9	27.6	27.2
IU	6.9	6.7	6.4	5.9	5.8	6.7	6.4	5.9	5.8
Otros	19.7	20.9	21.4	18.7	18.7	20.9	21.6	18.7	18.9

asignar un peso por no pertenecer a ninguna de las categorías de post-estratificación. Esto supone una reducción notable de la muestra (alrededor de un 20 % en la mayoría de los casos). Podemos deducir de aquí que, si bien el uso de la post-estratificación por recuerdo de voto puede mejorar la predicción para los votantes habituales (como de hecho lo hace), lleva a mayores desviaciones en el pronóstico global, ya que no tiene en cuenta el comportamiento de los nuevos electores. Y el comportamiento de estos últimos puede ser muy diferente del de los votantes habituales.

Otra nota que es importante destacar es que los modelos relativamente sencillos de imputación tienen un gran poder predictivo. En el caso de la encuesta pre-electoral, la capacidad predictiva de los modelos univariantes es ligeramente mejor que la predicción con modelos encadenados. Concretamente, si nos fijamos en los modelos con imputación sin post-estratificación (primera y tercera filas del Cuadro 5), vemos que el error medio por partido es de 0.70 % en el modelo reducido y de 0.80 % en el ampliado (que incluye la valoración de la situación económica y la gestión del gobierno), mientras que el error en las especificaciones de los modelos encadenados sube a 0.80 % en el modelo reducido y a 0.90 % en el modelo ampliado. Esto indica también que, además de que los modelos encadenados no consiguen mejorar la predicción con los datos de la encuesta pre-electoral, el añadir más variables al modelo tampoco mejora el resultado. En otros términos, el recuerdo de voto y la ideología, controlados por las características socio-demográficas de los votantes, son las variables claves para producir una buena predicción de los resultados.

Existen dos explicaciones potenciales para el hecho de que la evaluación de la situación económica y la valoración de la gestión del gobierno no mejoren la capacidad

Cuadro 4: Predicción de resultados electorales. Encuesta post-electoral

Modelos univariantes									
Voto	Real	Modelo simple				Modelo ampliado			
		No post-est.		Post-estrat.		No post-est.		Post-estrat.	
		Obs.	Comp.	Obs.	Comp.	Obs.	Comp.	Obs.	Comp.
PP	44.6	44.6	44.7	47.6	48.2	44.6	44.7	47.6	48.2
PSOE	28.8	28.1	28.0	28.0	27.7	28.1	28	28.0	27.7
IU	6.9	8.4	8.2	6.9	6.8	8.4	8.2	6.9	6.8
Otros	19.7	19.0	19.0	17.5	17.4	19.0	19.0	17.5	17.4

Modelos encadenados									
Voto	Real	Modelo simple				Modelo ampliado			
		No post-est.		Post-estrat.		No post-est.		Post-estrat.	
		Obs.	Comp.	Obs.	Comp.	Obs.	Comp.	Obs.	Comp.
PP	44.6	44.6	44.4	47.6	47.7	44.6	44.5	47.6	47.8
PSOE	28.8	28.1	28.5	28.0	28.0	28.1	28.3	28.0	28.0
IU	6.9	8.4	8.2	6.9	6.8	8.4	8.2	6.9	6.8
Otros	19.7	19.0	18.9	17.5	17.4	19.0	19.0	17.5	17.4

predictiva de los modelos. En primer lugar, hay un cierto problema de endogeneidad que no puede ser controlado en nuestros modelos, como ha puesto de manifiesto la literatura sobre voto económico (Evans y Pickup 2010). Las evaluaciones de la economía pueden verse influidas por el propio comportamiento electoral previo, por lo que no es posible saber con seguridad si un elector valora bien la economía y por eso vota al gobierno o valora bien la economía porque votó al partido en el gobierno. Y en este último caso estaríamos ante un mecanismo de reducción de la disonancia cognitiva en lugar de un mecanismo de *accountability*. En segundo lugar, hay factores del contexto de estas elecciones que a nuestro juicio pueden resultar más importantes. Como ya hemos mencionado, las elecciones se celebraron en un contexto de crisis económica aguda en el que las encuestas reflejaban una valoración bastante negativa de la gestión del gobierno. Y dado que la mayoría de los electores (tanto votantes del partido en el gobierno como de otros partidos) tenían una percepción bastante negativa de la economía y de los políticos, en especial, de los del gobierno, estas variables muestran una escasa variabilidad, lo cual impide que puedan contribuir de forma significativa a mejorar la capacidad predictiva de los modelos. Queda para futuros trabajos, por tanto, comparar nuestros resultados con elecciones celebradas en contextos diferentes en los que la evaluación de la economía o la gestión del gobierno puedan tener un papel más relevante.

En el Cuadro 4 presentamos los resultados de las estimaciones con la encuesta post-electoral. En términos generales, se observa una mejora importante en la capacidad predictiva con respecto a la encuesta pre-electoral, lo cual es lógico al provenir los datos de una encuesta posterior a la celebración de las elecciones en la que se pregunta por comportamientos y no por intenciones de voto. Según se puede ver, los modelos

Cuadro 5: Evaluación de las imputaciones

Método	Pre-electoral			Post-electoral		
	O-R	C-R	C-O	O-R	C-R	C-O
Univariante simple sin post-estratificación	1.30	0.70	1.30	0.50	0.50	0.10
Univariante simple con post-estratificación	2.00	2.10	0.20	2.00	2.40	0.40
Univariante ampliado sin post-estratificación	-	0.80	1.00	-	0.50	0.10
Univariante ampliado con post-estratificación	-	2.30	0.40	-	2.30	0.40
Encadenado simple sin post-estratificación	-	0.80	1.40	-	0.40	0.20
Encadenado simple con post-estratificación	-	2.10	0.10	-	2.00	0.10
Encadenado ampliado sin post-estratificación	-	0.90	1.10	-	0.40	0.10
Encadenado ampliado con post-estratificación	-	2.30	0.40	-	2.10	0.10

O-R: Error absoluto medio ponderado entre datos observados y reales.

C-R: Error absoluto medio ponderado entre datos completados y reales.

C-O: Desviación absoluta media ponderada entre datos completados y observados.

Las casillas con el signo - tienen, por pares, valores iguales que las superiores.

estimados con esta encuesta reducen notablemente la inflación de voto del PP y otros partidos que se daba en la pre-electoral, al tiempo que corrigen la baja estimación de voto del PSOE. De la misma forma, también se observan pautas similares a los modelos estimados con la encuesta pre-electoral. Los modelos sin post-estratificación funcionan mejor que los modelos con post-estratificación y los modelos reducidos funcionan relativamente bien en comparación con los modelos ampliados que incluyen un mayor número de variables. No obstante, en el caso de la encuesta post-electoral se observa que los modelos encadenados producen mejores predicciones que los modelos univariantes. En este caso, cuando se imputan las variables explicativas, la predicción mejora al reducir el error (medida C-R) de 0.50 % a 0.40 % (filas primera y quinta del Cuadro 5).

Finalmente, también conviene evaluar la imputación a partir de la comparación de la distribución de los datos imputados con los datos observados. Para ello hemos usado el comando `middiagplots` escrito por Eddings y Marchenko (2012) que permite tabular la distribución de las proporciones de voto de cada uno de los partidos a partir de los datos observados, imputados y completados. Para llevar a cabo esta comparación, hemos usado un total de 20 imputaciones obtenidas a partir del modelo univariante simple sin post-estratificación, ya que es el modelo que produce las mejores predicciones de los resultados de la elección. La representación gráfica, que puede consultarse en el apéndice de este capítulo, indica que los datos imputados en la encuesta pre-electoral incrementan de forma notable la proporción de voto del PSOE, lo cual se corresponde con la realidad de lo sucedido en la elección, ya que muchos de los votantes que no declaraban una intención de voto decidida en esta encuesta eran votantes del PSOE. Adicionalmente, las 20 imputaciones muestran una elevada consistencia. Por otra parte, las distribuciones de los datos imputados a partir de la encuesta post-electoral es mucha más cercana a los datos observados. Esto último también es consistente con el hecho

de que en esta encuesta ya no hay votantes indecisos, sino una fracción menor que no revela el partido al que votó en las elecciones.

4.2. Los determinantes del voto

Una vez analizada la capacidad predictiva de los diferentes modelos estimados, comentaremos brevemente cómo influyen las variables explicativas sobre la decisión de voto. Así pues, en esta sección presentamos los resultados de la estimación del modelo univariante reducido ponderado por provincia de la encuesta pre-electoral, aquel en el que únicamente se imputa la decisión de voto a partir de los valores dados de las variables explicativas utilizando la post-estratificación por provincia. El motivo es que éste es el modelo que mejor predice los resultados electorales reales a partir de la encuesta pre-electoral. Asimismo, la dirección y la magnitud de la influencia de las variables son bastante similares en todos los modelos estimados. En todos los modelos discutidos, se ha utilizado un logit multinomial para estimar el efecto de las variables explicativas, cuya variable dependiente es la decisión de voto (PP, PSOE, IU y otros partidos) y se ha tomado como categoría de referencia el grupo de votantes que votan por otros partidos. En el Cuadro 6 presentamos los efectos marginales para cada una de las categorías de respuesta.

Empezando por las variables asociadas a la teoría de la identificación partidaria, en todos los modelos el efecto del recuerdo de voto y la ideología es altamente significativo. En términos generales, el hecho de haber votado por un partido incrementa la probabilidad de volver a votarlo. El efecto más claro se da en el caso del PP, para el cual la probabilidad de votarlo se incrementa más de 0.70 entre quienes lo habían votado en las elecciones anteriores con respecto a quienes habían votado por otros partidos. Y de forma simétrica, la probabilidad de votar por otros partidos se reduce en una magnitud similar entre quienes habían votado al PP en las elecciones anteriores. Se da también el caso de que el hecho de haber votado al PSOE incrementa ligeramente la probabilidad de votar por el PP e IU (siempre en relación con la probabilidad de votar por otros partidos), aunque también reduce drásticamente la probabilidad de votar por otros partidos. Ello viene a ser un reflejo de la pérdida de votos que experimentó el PSOE en estas elecciones y que se dirigieron hacia los partidos situados a su izquierda (IU) y su derecha (PP). Un resultado igualmente interesante es que el hecho de haber votado por partidos nacionalistas (CiU y PNV) en las elecciones de 2008 reduce las probabilidades de votar por el PP en las elecciones de 2011. Con respecto a la ideología, se observa que aquellos que se posicionan en la izquierda o el centro izquierda (por comparación con los que se sitúan en el centro, que representa la categoría de referencia) tienen menos probabilidades de votar al PP, al tiempo que tienen una mayor probabilidad de votar por PSOE o IU. En sentido contrario, la probabilidad de votar por el PP se incrementa entre los electores que se sitúan en el centro derecha y derecha, mientras que en estos colectivos se reducen las probabilidades de voto de PSOE e IU.

En relación a las variables socio-estructurales se observa que un incremento del nivel educativo (medido por la variable de nivel de estudios) tiene un efecto negativo sobre la probabilidad de voto de PP y PSOE, al tiempo que incrementa la probabilidad de

voto a otros partidos. Ello indica que los electores con mayor nivel de estudios están sobrerrepresentados en partidos minoritarios. Finalmente, el género y la edad tienen un efecto limitado sobre las decisiones de voto. El género sólo tiene un efecto significativo sobre las probabilidades de voto del PP y el PSOE, dado que las mujeres tienen más probabilidades de votar por el PSOE y menos probabilidad de votar por el PP. Y la edad solo tiene un efecto significativo sobre las probabilidades de votar al PSOE pero de magnitud muy reducida.

Finalmente, al incluir el efecto de las variables de evaluación de la economía y de la gestión del gobierno en el modelo anterior, los coeficientes no experimentan apenas variación en la significación ni en la magnitud. Se observa asimismo que la evaluación de la situación económica no tiene un efecto significativo sobre las probabilidades de voto de ningún partido si se utiliza el 5% como umbral de significación. El efecto de la valoración de la gestión del gobierno sí que resulta significativo en el caso del PSOE y, en menor medida, en el caso del PP e IU. Las probabilidades de voto por el PSOE se incrementan claramente en aquellos electores que tienen una valoración positiva de la gestión del gobierno, mientras que para este grupo de electores se reduce la probabilidad de voto por el PP.

De estos resultados se desprende que la orientación del voto de los españoles en las elecciones de 2011 estuvo condicionada fundamentalmente por factores ideológicos, como la cercanía ideológica a los partidos o el hecho de haberlos votado previamente. Las variables estructurales tuvieron un peso limitado en la orientación del voto, mientras que la valoración de la gestión del gobierno, y especialmente de la situación económica, tuvieron una influencia muy reducida sobre las decisiones de voto. Como ya hemos mencionado, una posible explicación se encuentra en el hecho de que la percepción mayoritaria del electorado sobre la situación económica y la gestión del gobierno eran bastante negativas al final de la legislatura, tanto entre quienes votaron por partidos de la oposición como entre quienes votaron por el partido en el gobierno. Estos últimos lo hicieron fundamentalmente en función de planteamientos ideológicos, a pesar de que su percepción de la economía era igualmente negativa.

Cuadro 6: Modelo univariante simple ponderado por provincia. Encuesta pre-electoral

	Ecuación con casos completos					Ecuación con datos completados				
	PP	PSOE	IU	Otros	PP	PSOE	IU	Otros	Otros	
No votó	0.289*** (0.032)	0.188*** (0.020)	0.017** (0.008)	-0.494*** (0.030)	0.304*** (0.028)	0.184*** (0.018)	0.018*** (0.007)	-0.507*** (0.026)		
Votó PSOE	0.148*** (0.032)	0.423*** (0.028)	0.022*** (0.008)	-0.593*** (0.027)	0.160*** (0.029)	0.427*** (0.025)	0.024*** (0.007)	-0.612*** (0.024)		
Votó PP	0.725*** (0.026)	-0.009 (0.007)	-0.009* (0.005)	-0.708*** (0.027)	0.740*** (0.023)	-0.009 (0.006)	-0.008* (0.005)	-0.722*** (0.024)		
Votó IU	0.023 (0.057)	0.030* (0.016)	0.388*** (0.065)	-0.441*** (0.056)	0.033 (0.050)	0.031** (0.014)	0.392*** (0.058)	-0.456*** (0.049)		
Votó CiU	-0.089** (0.036)	0.023 (0.015)	-0.005 (0.010)	0.071* (0.039)	-0.066* (0.034)	0.021* (0.013)	-0.006 (0.008)	0.051 (0.036)		
Votó PNV	-0.178*** (0.028)	-0.005 (0.010)	-0.005 (0.009)	0.188*** (0.030)	-0.167*** (0.025)	-0.002 (0.010)	-0.006 (0.008)	0.175*** (0.027)		
Sin ideología	-0.032 (0.029)	0.010 (0.010)	-0.014 (0.009)	0.037 (0.034)	-0.031 (0.023)	0.002 (0.008)	-0.015** (0.006)	0.044* (0.026)		
Izquierda	-0.303*** (0.027)	0.120*** (0.021)	0.160*** (0.033)	0.024 (0.042)	-0.303*** (0.023)	0.119*** (0.020)	0.157*** (0.031)	0.027 (0.038)		
Centro-izq.	-0.237*** (0.022)	0.115*** (0.015)	0.066*** (0.017)	0.056* (0.030)	-0.235*** (0.019)	0.113*** (0.013)	0.063*** (0.015)	0.059** (0.026)		
Centro-der.	0.344*** (0.033)	-0.059*** (0.009)	-0.027*** (0.009)	-0.258*** (0.034)	0.333*** (0.030)	-0.062*** (0.008)	-0.024*** (0.007)	-0.248*** (0.031)		
Derecha	0.370*** (0.057)	-0.056*** (0.012)	-0.028*** (0.009)	-0.286*** (0.058)	0.363*** (0.054)	-0.057*** (0.011)	-0.019** (0.008)	-0.287*** (0.055)		
Estudios	-0.053*** (0.007)	-0.014*** (0.003)	0.001 (0.001)	0.066*** (0.007)	-0.057*** (0.006)	-0.013*** (0.002)	0.002 (0.001)	0.068*** (0.006)		
Mujer	-0.035** (0.018)	0.022*** (0.006)	-0.001 (0.004)	0.014 (0.019)	-0.034** (0.015)	0.016*** (0.005)	-0.004 (0.004)	0.021 (0.017)		
Edad	<-0.001 (0.001)	0.001*** (0.001)	<-0.001 (0.001)	-0.001 (0.001)	<-0.001 (0.001)	0.001*** (0.001)	<-0.001 (0.001)	-0.001 (0.001)		
Observaciones			10,731						13,320	

Las casillas representan el efecto marginal de cada variable en el modelo logit multinomial.

Errores típicos entre paréntesis (***) p<0.01, ** p<0.05, * p<0.1).

5. Conclusiones

Las técnicas de imputación múltiple de datos tienen una historia de varias décadas en Ciencias Sociales, aunque no han sido aplicadas de forma sistemática en la realización de pronósticos electorales, salvo en estudios puntuales como los de Rivas et al. (2010) y Varela Mallou et al. (1998). Por otra parte, la literatura sobre pronósticos electorales ha utilizado generalmente variables agregadas, con lo que los mecanismos de decisión de los votantes quedan en una *caja negra* que no permite hacer inferencias sobre las motivaciones individuales. En este trabajo hemos usado técnicas de imputación múltiple con el fin de producir pronósticos electorales aplicándolo al caso específico de las últimas elecciones generales en España. Para ello hemos utilizado la información de las encuestas pre-electorales pero también hemos comparado esta información con la que proporcionan las encuestas post-electorales. Más concretamente, hemos obtenido una predicción de voto en las elecciones generales de 2010 en España, a partir de la imputación de la variable de intención de voto en la encuesta pre-electoral y de la de recuerdo de voto en la encuesta post-electoral. Para realizar la imputación se ha tenido en cuenta la información proporcionada por variables relacionadas con el comportamiento electoral: variables ideológicas, factores socio-estructurales y valoración de la gestión del gobierno y de la economía.

Los resultados muestran que el uso de las técnicas de imputación mejora de forma consistente la capacidad predictiva de los pronósticos electorales en dicha elección si comparamos la capacidad predictiva de los modelos con datos completados con la capacidad predictiva de los modelos con datos completos. Al mismo tiempo, no obstante, hemos visto que en nuestro caso el pronóstico, usando técnicas de imputación, no mejora con respecto al pronóstico con datos completos cuando se tiene en cuenta el recuerdo de voto como criterio de post-estratificación de los pesos del diseño de la muestra. También hemos podido comprobar que para este conjunto de datos, los dos modelos de imputación simple utilizados producen mejores resultados que los dos modelos de imputación encadenada utilizados. Así pues, los modelos imputación univariantes estimados aquí producen mejores predicciones que los modelos encadenados en los que se imputan tanto la variable dependiente como las explicativas. Del mismo modo, las especificaciones reducidas de modelos (que no incluyen la valoración de la situación económica y la gestión del gobierno) producen predicciones prácticamente igual de acertadas. Parte de la explicación a la falta de poder explicativo de las variables de valoración de la economía y el gobierno puede deberse al contexto específico en el que se celebraron estas elecciones, en el que una mayoría del electorado tenía opiniones negativas sobre ambas variables. Queda para futuros trabajos evaluar la capacidad predictiva de los modelos aquí propuestos en otras elecciones celebradas en circunstancias diferentes, como la de ausencia de crisis económica o la convocatoria adelantada de comicios. De la misma forma, los resultados aquí expuestos podrían ser utilizados para evaluar la calidad de la información proporcionada por las encuestas electorales.

6. Referencias

- Abramowitz, A. I. 2010. How large a wave? Using the generic ballot to forecast the 2010 midterm elections. *PS: Political Science & Politics* 43(04): 631–632.
- Abramson, P. R. 1983. *Political attitudes in America: formation and change*. San Francisco: W.H. Freeman.
- Aguilar, P., y I. Sánchez-Cuenca. 2007. ¿Gestión o representación? Los determinantes del voto en contextos políticos complejos. *Revista Española de Investigaciones Sociológicas* 117: 61–86.
- Alonso, S. 2008. La duración en el poder y el rendimiento electoral de los partidos nacionalistas gobernantes: ¿la ventaja competitiva del nacionalismo? *Revista Española de Investigaciones Sociológicas* 121: 81–103.
- Bafumi, J., R. S. Erikson, y C. Wlezien. 2010. Forecasting House seats from generic congressional polls: the 2010 midterm election. *PS: Political Science & Politics* 43(04): 633–636.
- Bosch, A., y C. Riba. 2005. Coyuntura económica y voto en España. *Papers. Revista de Sociología* 75: 117–140.
- De la Calle, L., I. Martínez, y L. Orriols. 2010. Voting without ideology. Evidence from Spain (1979–2008). *Revista Española de Investigaciones Sociológicas* 129: 107–129.
- Campbell, A., P. E. Converse, W. E. Miller, y D. E. Stokes. 1960. *The American voter*. Chicago: University of Chicago Press.
- Campbell, J. E. 2010a. Forecasts of the 2010 midterm elections. *PS: Political Science & Politics* 43(04): 625–626.
- . 2010b. The seats in trouble forecast of the 2010 elections to the U.S. House. *PS: Political Science & Politics* 43(04): 627–630.
- Caínzos, M. 2001. La evolución del voto clasista en España. *Zona Abierta* 96/97: 91–171.
- Clark, T. N., y S. M. Lipset. 1991. Are social classes dying? *International Sociology* 6: 397–410.
- Clarke, H., D. Sanders, M. Stewart, y P. Whiteley. 2011. Valence politics and electoral choice in Britain, 2010. *Journal of Elections, Public Opinion & Parties* 21(2): 237–253.
- Cuzán, A. G. 2010. Will the republicans retake the House in 2010? *PS: Political Science & Politics* 43(04): 639–641.
- Diebold, F. X. 2007. *Elements of forecasting*. Cincinnati: Thompson-South-Western Publishers.
- Downs, A. 1957. *An economic theory of democracy*. New York,: Harper.

- Duch, R., y R. Stevenson. 2008. *The economic voting: How political and economic institutions condition electoral results*. Cambridge: Cambridge University Press.
- Dunleavy, P., y C. T. Husbands. 1985. *British democracy at the crossroads: voting and party competition in the 1980s*. London ; Boston: Allen & Unwin.
- Eddings, W., y Y. Marchenko. 2012. Diagnostics for multiple imputation in Stata. *Stata Journal* 12(3): 353–367.
- Erikson, R. S. 1989. Economic conditions and the presidential vote. *American Political Science Review* 83(2): 567–573.
- Escobar, M., E. Fernández Macías, y F. Bernardi. 2012. *Análisis de datos con Stata*. 2ª ed. Madrid: CIS.
- Evans, G. 2000. The continued significance of class voting. *Annual Review of Political Science* 3: 401–417.
- Evans, G., y M. Pickup. 2010. Reversing the causal arrow: the political conditioning of economic perceptions in the 2000-2004 U.S. presidential election cycle. *The Journal of Politics* 72(04): 1236–1251.
- Fiorina, M. P. 1977. Outline for a model of party choice. *American Journal of Political Science* 21(3): 601–625.
- Flanagan, S. C., y R. Dalton. 1994. Models of change. In *The West European party system*, ed. P. Mair, 232–246. Oxford: Oxford University Press.
- Fraile, M. 2005. *Cuando la economía entra en las urnas. El voto económico en España (1979-1996)*. Madrid: Centro de Investigaciones Sociológicas.
- Fraile, M., y M. S. Lewis-Beck. 2010. Economic voting in Spain: A 2000 panel test. *Electoral Studies* 29(2): 210–220.
- González, J. J. 2004. Las bases sociales de la política en España. *Revista Española de Sociología* 4: 119–142.
- Harrop, M., y W. L. Miller. 1987. *Elections and voters: a comparative introduction*. Houndmills, Basingstoke, Hampshire: Macmillan Education.
- Hyndman, R. J., y A. B. Koehler. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22(4): 679–688.
- Jaime Castillo, A. M., y J. L. Sáez Lozano. 2001. *El comportamiento electoral en la democracia española*. Madrid: Centro de Estudios Políticos y Constitucionales.
- . 2007. Atribución de la responsabilidad y voto económico: el caso de España. *Trimestre Económico* LXXIV(2): 379–413.
- Key, V. O. 1966. *The responsible electorate. Rationality in presidential voting, 1936-1960*. Cambridge,: Belknap Press of Harvard University Press.

- King, G., J. Honaker, A. Joseph, y K. F. Scheve. 2001. Analyzing incomplete Political Science data: an alternative algorithm for multiple imputation. *American Political Science Review* 95(1): 49–69.
- Kriesi, H. P. 1998. The transformation of cleavage politics: the 1997 Stein Rokkan lecture. *European Journal of Political Research* 33(2): 165–185.
- Lago, I., y J. R. Montero. 2010. Participación y resultados electorales en España. *Revista Española de Investigaciones Sociológicas* 130: 97–116.
- Lago-Peñas, I., y S. Lago-Peñas. 2005. Does the economy matter? An empirical analysis of the causal chain connecting the economy and the vote in Galicia. *Economics and Politics* 17: 215–243.
- Lewis-Beck, M. S. 2005. Election forecasting: principles and practice. *The British Journal of Politics & International Relations* 7(2): 145–164.
- Lewis-Beck, M. S., y C. Tien. 1996. The future in forecasting. *American Politics Research* 24(4): 468–491.
- . 2010. The referendum model: a 2010 congressional forecast. *PS: Political Science & Politics* 43(04): 637–638.
- Lipset, S. M., y S. Rokkan. 1967. *Party systems and voter alignments: cross-national perspectives*. New York: Free Press.
- Magalhaes, P., L. Aguiar-Conraria, y M. S. Lewis-Beck. 2011. Forecasting Spanish elections. WP 17/2011, Nucleo de Investigação en Políticas Económicas.
- Makridakis, S., S. Wheelwright, y R. J. Hyndman. 1998. *Forecasting methods and applications*. New York: John Wiley & Sons.
- Manza, J., M. Hout, y C. Brooks. 1995. Class voting in capitalist democracies since World-War-II. Dealignment, realignment, or trendless fluctuation. *Annual Review of Sociology* 21: 137–162.
- Montero, J., K. Calvo, y Álvaro Martínez. 2008. El voto religioso en España y Portugal. *Revista Internacional de Sociología* 66(51): 19–54.
- Montero, J. R. 1994. Sobre las preferencias electorales en España: fragmentación y polarización (1977-1993). In *Comportamiento Político y Electoral*, ed. P. Del Castillo, 51–124. Madrid: Centro de Investigaciones Sociológicas.
- Nie, N. H., S. Verba, y J. R. Petrocik. 1976. *The changing American voter*. Cambridge, Mass.: Harvard University Press.
- Nieuwbeerta, P. 1996. The democratic class struggle in postwar societies: class voting in twenty countries, 1945-1990. *Acta Sociologica* 39(4): 345–383.

- Nieuwbeerta, P., y W. Ultee. 1999. Class voting in Western industrialized countries, 1945-1990: systematizing and testing explanations. *European Journal of Political Research* 35(1): 123-160.
- Polavieja, J. G. 2001. ¿Qué es el voto de clase? Los mecanismos del voto de clase en España. *Zona Abierta* 96/97: 173-213.
- Riker, W. H., y P. C. Ordeshook. 1968. A theory of the calculus of voting. *American Political Science Review* 62(1): 25-42.
- Rivas, C., M. d. M. Martínez Rosón, y P. Galindo. 2010. La imputación múltiple como alternativa al análisis de la no respuesta en la variable de intención de voto. *Revista Española de Ciencia Política* 22: 99-118.
- Rivero, G. 2011. *Análisis de datos incompletos en Ciencias Sociales*. Madrid: CIS.
- Royston, P. 2009. Multiple imputation of missing values: futher update of ice, with an emphasis on categorical variables. *Stata Journal* 9: 466-477.
- Rubin, D. B. 1987. *Multiple imputation for non-response in surveys*. New York: Wiley.
- Sanders, D., H. D. Clarke, M. C. Stewart, y P. Whiteley. 2011. Downs, Stokes and the dynamics of electoral choice. *British Journal of Political Science* 41: 287-314.
- Stata. 2011a. *Stata multiple imputation reference manual. Release 12*. College Station, TX: Stata Press.
- . 2011b. *Stata survey data reference manual. Release 12*. College Station, TX: Stata Press.
- Urquizu-Sancho, I. 2006. The non-declared vote in the surveys: The Spanish case in the 1980s. *Electoral Studies* 25(1): 103-128.
- Varela Mallou, J., T. Braña Tobío, A. García Carreira, A. Rial Boubeta, y X. G. Vázquez Fernández. 1998. Estimación de la respuesta de los “no sabe/no contesta” en los estudios de intención de voto. *Revista Española de Investigaciones Sociológicas* 83: 269-287.
- Ventura, R. 2001. Family political socialization in multiparty systems. *Comparative Political Studies* 34(6): 666-691.
- Van der Waal, J., P. Achterberg, y D. Houtman. 2007. Class is not dead. It has been buried Alive: class voting and cultural voting in postwar Western societies (1956-1990). *Politics & Society* 35(3): 403-426.
- Whiteley, P. 1979. Electoral forecasting from poll data: the British case. *British Journal of Political Science* 9(02): 219-236.
- . 2008. Evaluating rival forecasting models of the 2005 general election in Britain. An encompassing experiment. *Electoral Studies* 27(4): 581 - 588.

Whiteley, P., D. Sanders, M. Stewart, y H. Clarke. 2011. Aggregate level forecasting of the 2010 general election in Britain: the seats-votes model. *Electoral Studies* 30(2): 278–283.

Williams, J. T. 1990. The political manipulation of macroeconomic policy. *American Political Science Review* 84: 767–795.

7. Apéndice 1: Diagnósticos de la Imputación

En este apéndice presentamos un diagnóstico de la imputación utilizada en el capítulo, comparando las distribuciones de datos observados, imputados y completados. Para realizar este diagnóstico hemos utilizado el comando `midiagplots` escrito por Eddings y Marchenko (2012). En la Figura 1 se presentan las distribuciones para la encuesta pre-electoral y en la Figura 2 se presentan las distribuciones para la encuesta post-electoral. En ambos casos se han obtenido 20 imputaciones usando el modelo univariante sin post-estratificación, que tiene la mejor capacidad predictiva entre todos los estimados.

Figura 1: Distribuciones de datos observados, imputados y completados en la encuesta pre-electoral. Modelo univariante

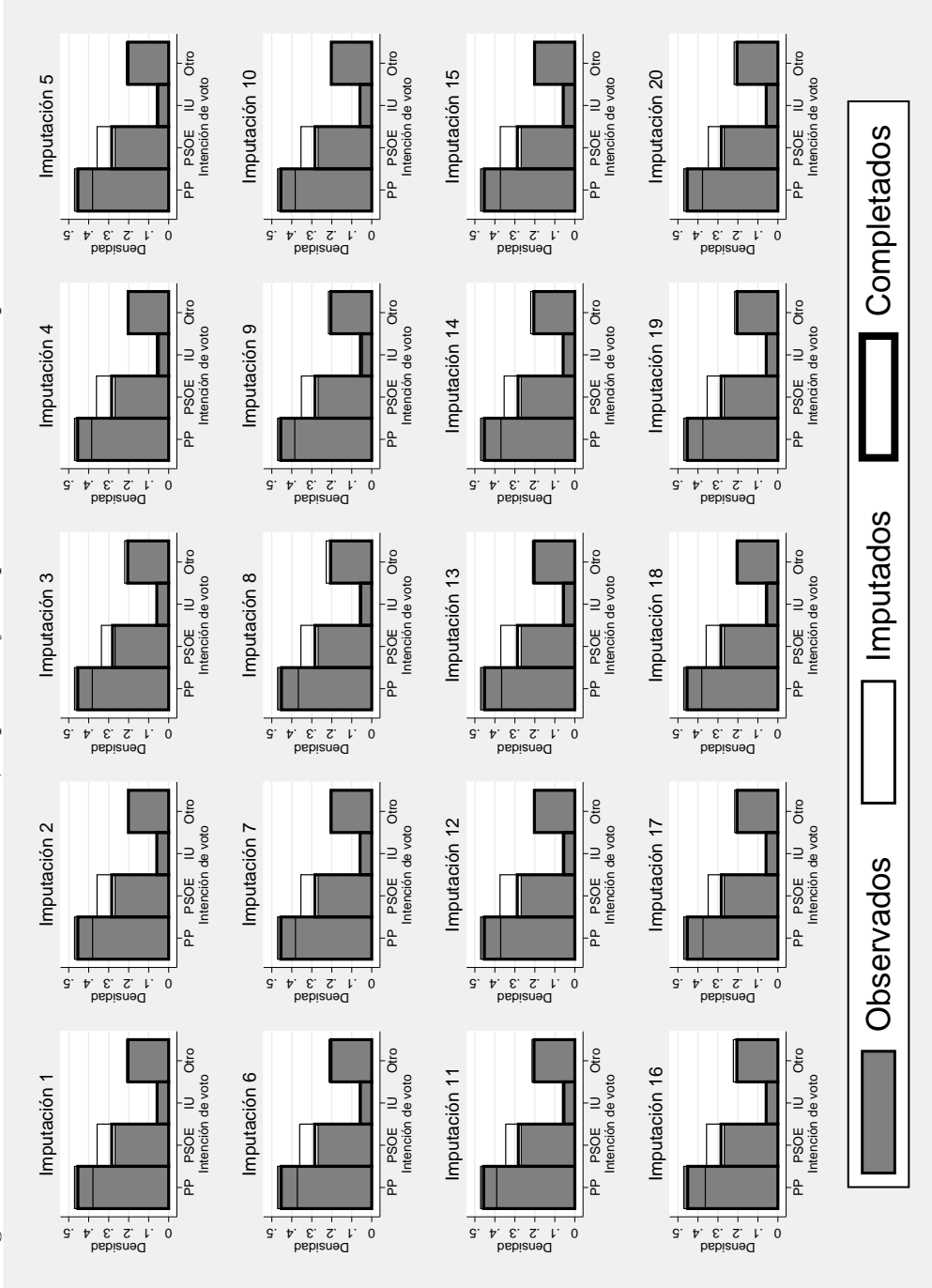


Figura 2: Distribuciones de datos observados, imputados y completados en la encuesta post-electoral. Modelo univariante

