

# Cocina electoral

## Encuestas en tiempos de incertidumbre

Pablo Cabrera Álvarez |  @pablocalv

Universidad de Salamanca

28 feb. 2019

# Problema

# Reino Unido, 2015

B

B

C

Sign in

News

Sport

Weather

Shop

Reel

Travel

M

NEWS

Home

Video

World

UK

Business

Tech

Science

Stories

Entertainment & Arts

Politics

Parliaments

Brexit

# Election 2015: How the opinion polls got it wrong

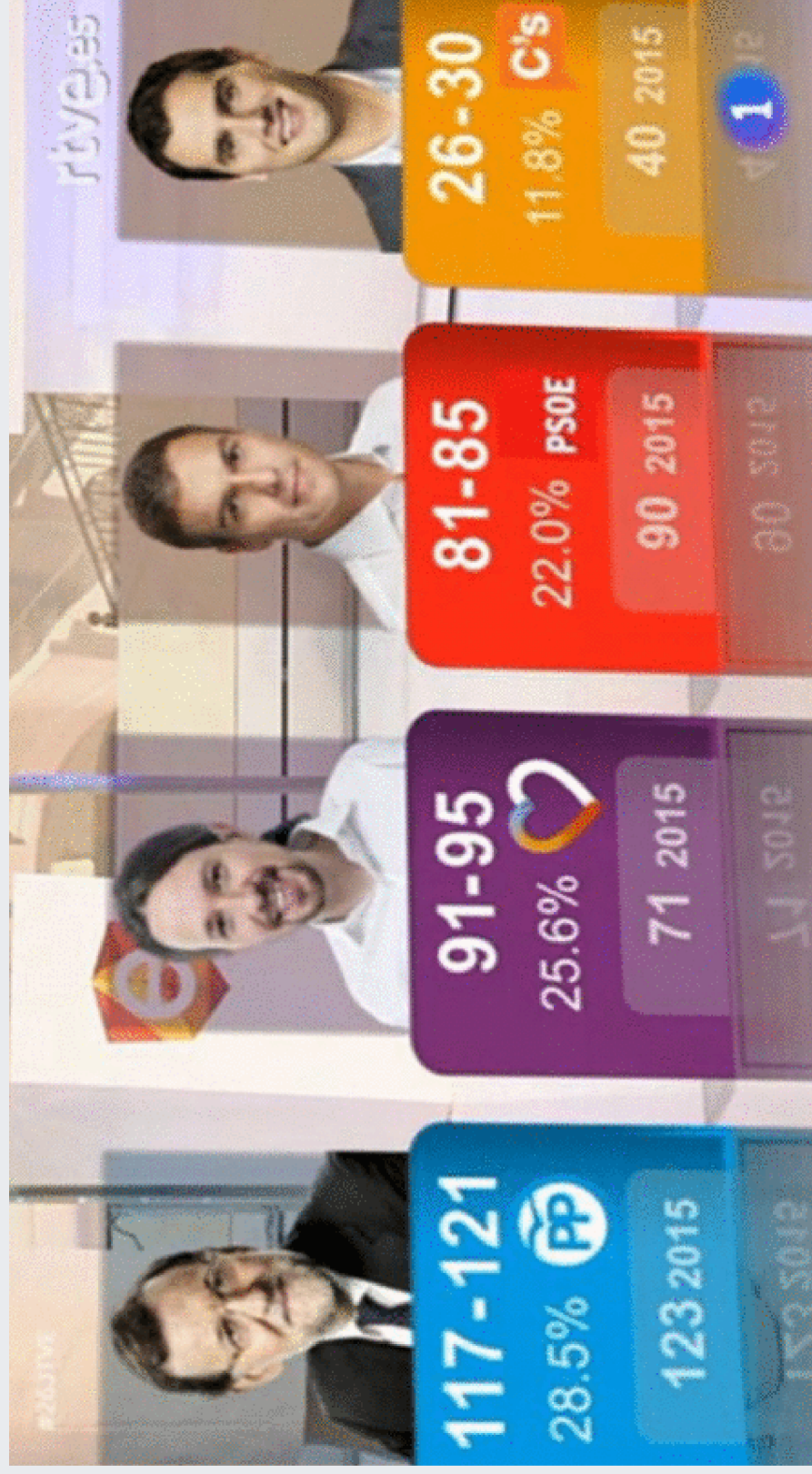
By David Cowling  
Editor, BBC Political Research Unit

© 17 May 2015

f

Share

# España, 2016



# Y no son casos únicos...

- Estados Unidos 1948
- Francia 2002

# Y no son casos únicos...

- Estados Unidos 1948
- Estados Unidos 1992
- Virginia 1989
- Reino Unido 1970
- Reino Unido 1992
- Reino Unido 2015
- Nicaragua 1990
- Hungría 2002
- Francia 2002
- Francia 2007
- Francia 1997
- Quebec 1998
- España 1993
- España 1996
- España 2016
- Italia 2006

# ¿Qué puede fallar en una estimación de voto?

- Problemas de **medición**
  - Reporte erróneo deliberado (Bodor, 2012)
  - Cambio de opinión de los votantes (Durand *et al.*, 2002; MRS, 1994)
  - Modo de administración (Abramson, 2007; Callegaro y Gasperoni, 2008)
- Problemas de **representatividad**
  - Voto por correo y votantes en el extranjero
  - Estimación de la participación (Abrams, 1970; Callegaro y Gasperoni, 2008)
  - No respuesta y sesgo de autoselección (Jowell *et al.*, 1993; Durand *et al.*, 2002; Callegaro y Gasperoni, 2008; Sturgis *et al.*, 2016)
- Efecto **rebaño**

# La vida de un *pollster* en la cocina





# Preguntas

# Población

Población de cinco  
**hombres** y cinco **mujeres**  
(censo)

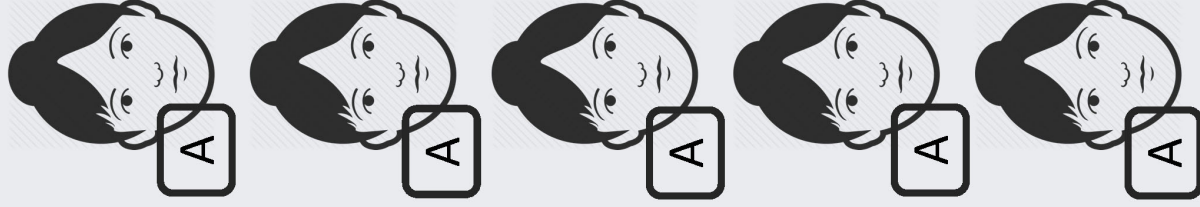


Investigar su **intención de  
voto**

# Población

Población (censo):

5 (50%) **Mujeres**  
5 (50%) **Hombres**



Población (modo Dios):

5 (50%) **Partido A**  
5 (50%) **Partido B**

# No respuesta total

Muestra:

2 (29%) **Mujeres**  
5 (71%) **Hombres**



Muestra:

2 (29%) **Partido A**  
5 (71%) **Partido B**

# Ponderación de no respuesta

Ponderación como:

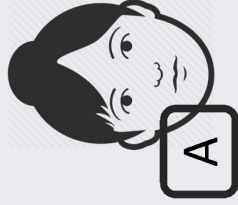
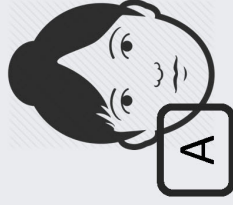
$$w_k = \frac{N_k}{n_k}$$

Para las **mujeres**:

$$w_m = \frac{5}{2} = 2.5$$

Para los **hombres**:

$$w_h = \frac{5}{5} = 1.0$$



Estimar voto con  
**ponderaciones:**

$$\text{M - A: } 2 * 2.5 = 5 \text{ (50\%)}$$

$$\text{H - B: } 5 * 1.0 = 5 \text{ (50\%)}$$

# Ponderación de no respuesta

- **Ajuste estadístico** que permite compensar las diferencias en las probabilidades de respuesta de diferentes subgrupos (Särndal, 2007)
- **Variables auxiliares** disponibles para los que responden y los que no responden, o al menos los totales poblacionales (*por ej.* sexo, edad, voto en anteriores elecciones...)
- **Variables auxiliares** que estén correlacionadas con la **probabilidad de responder** y la **variable de interés** (*por ej.* intención de voto)
- La **distribución de la variable de interés** dentro de los subgrupos de las variables auxiliares debe ser similar a la distribución en la población (Sturgis *et al.*, 2016)
- En estudios para estimar voto uso recurrente de la variable **recuerdo de voto**. Debate sobre su uso (Crespi, 1988; Voss *et al.*, 1995; Durand *et al.*, 2015; Crewe, 2001; Escobar *et al.*, 2014)

# Problemas del recuerdo de voto

Recuerdo de voto barómetros CIS (1994-2004)  
En porcentaje sobre censo



# No respuesta parcial

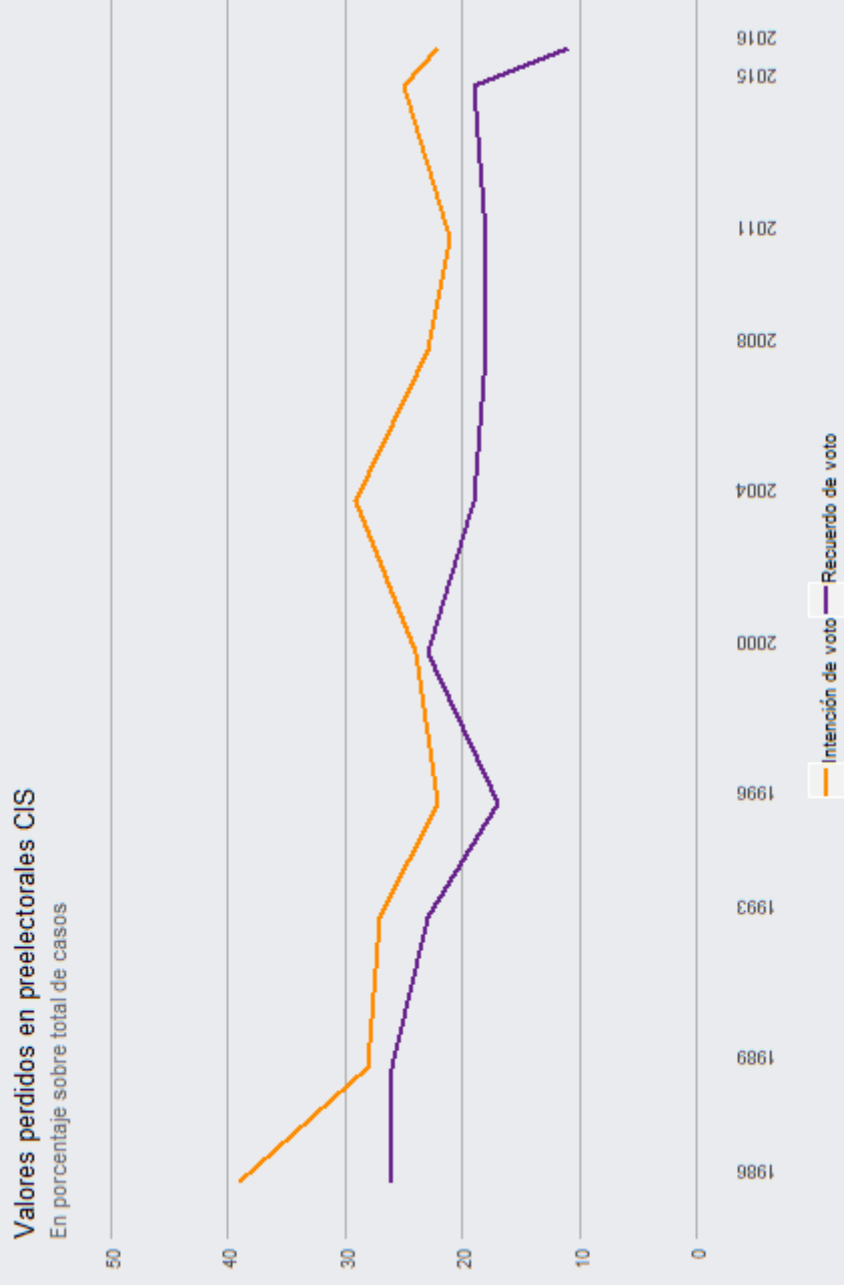
Dos entrevistados  
rechazaron revelar  
su **intención de voto**



Usar otra información  
de la encuesta para  
**imputar**



# No respuesta parcial



# Imputación

- Diversidad de **métodos de imputación**: imputación de la media/moda, *hot-deck*, imputación condicional, imputación múltiple
- Posibilidad de utilizar otras variables del cuestionario para **informar la imputación** (*por ej.* imputación condicional, *hot-deck*)
- La **imputación múltiple** consiste en generar  $m$  bases de datos completas con lo que se añade a la varianza un componente derivado de la incertidumbre debida a los datos perdidos (Rivero, 2011)
- La imputación de la **intención de voto** llevada a cabo con técnicas *ad hoc* según la casa de encuestas (Crespi, 1988; Sturgis *et al.*, 2016)
- Algunos intentos de utilizar imputación múltiple en el ámbito electoral desde la academia (*por ej.* Bernhagen y Marsh, 2007; Escobar y Jaime, 2013; Liu, 2014)

# Hipótesis

- H1: La **ponderación** de la muestra por **recuerdo de voto** contribuye a incrementar la precisión de la estimación de voto
- H2: El uso de la **ponderación** que combina las variables **sociodemográficas** y el **recuerdo de voto** será el más efectivo para reducir el nivel de error de la estimación de voto
- H3: El uso de técnicas de **imputación múltiple** para tratar la variable **recuerdo de voto** reduce el nivel de sesgo presente en esta variable y, en consecuencia, su uso en la ponderación incrementa la precisión de la estimación de voto
- H4: El uso de técnicas de **imputación múltiple** para asignar una **intención de voto** a aquellos que no saben o no contestan incrementa la precisión de la estimación de voto
- H5: El uso de la variable **recuerdo de voto** para ponderar tiene un efecto positivo en la precisión de las estimaciones en las **elecciones en las que hay continuidad política**, también después de 2015

# Datos y métodos

- Datos de encuestas **preelectorales y postelectorales del CIS** para las elecciones 1982-2016
- **Ponderaciones** generadas usando **calibración logística**
- **Imputación múltiple** utilizando un sistema **ecuaciones encadenadas y univariada**
- Estimaciones evaluadas utilizando el **error medio absoluto ponderado (EMAP)** con respecto a los resultados electorales:

$$EMAP = \sum_{k=1}^k |\hat{p}_k - p_k| p_k$$

donde  $\hat{p}_k$  es la estimación de voto para el partido  $k$  y  $p_k$  su resultado en las elecciones

- EMAP de cada estimación y recuerdo de voto imputado modelado usando regresión lineal (MCO)

# Métodos: ponderación

Nombre	Variables de ponderación	RV imputado	Ámbito del RV
EB	Provincia		
SD	Provincia y sociodemográficas (SD)		
RV0	Provincia y recuerdo de voto (RV)	No	NR, NV y NVJ excluidos
RV1	Provincia y RV	No	NR, NV y NVJ incluidos
RV2	Provincia y RV	Sí (NR imputado)	NR, NV y NVJ incluidos
RV3	Provincia y RV	Sí (NR y NVJ imputado)	NR, NV y NVJ incluidos
SD+RV	Provincia, SD y RV	Sí (NR y NVJ imputado)	NR, NV y NVJ incluidos

# Métodos: imputación múltiple

- Recuerdo de voto:

Nombre	Variable objetivo (modelo)	Predictores
RV1 sin imputar		
RV2 imputado (NR)	Recuerdo de voto (multinomial)	Conjunto mixto
RV3 imputado (NR y NVJ)	Recuerdo de voto (multinomial)	Conjunto mixto

**Conjunto mixto:** Sexo, edad, tamaño de hábitat, comunidad autónoma, intención de voto e ideología

# Métodos: imputación múltiple

- Intención de voto:

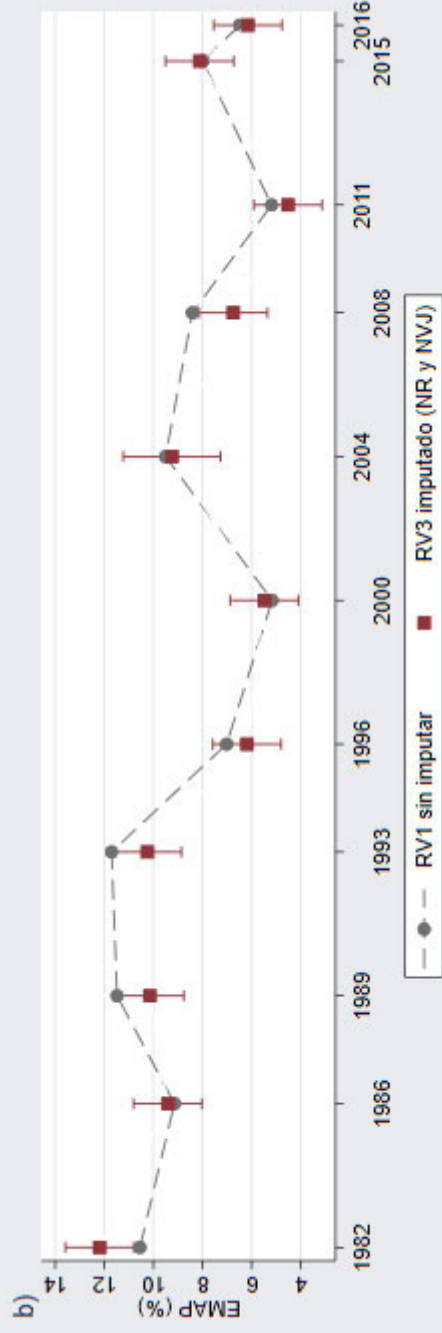
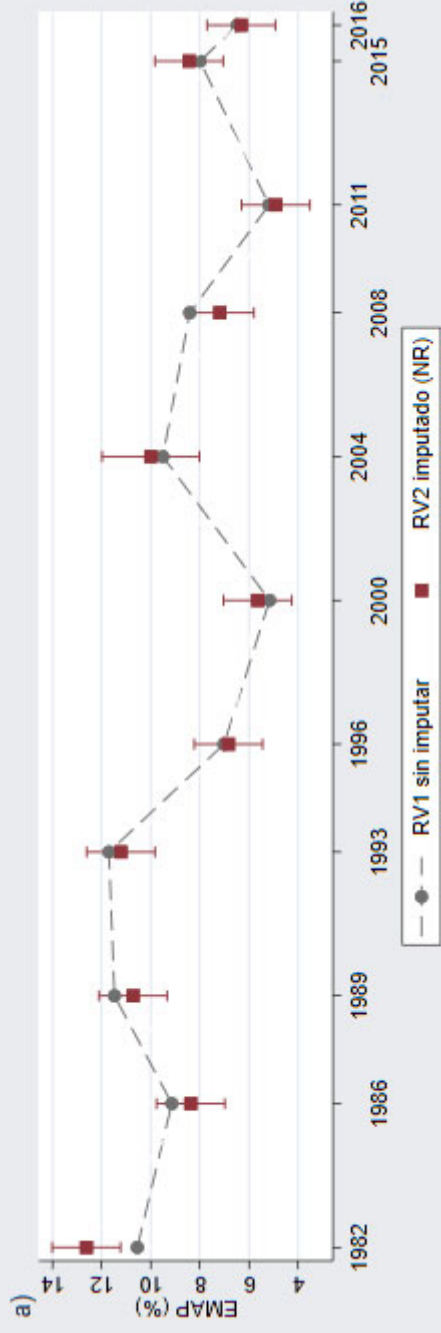
Nombre	Variable objetivo (modelo)	Predictores
No imputada (EB)		
Univariada básica	Intención de voto (multinomial)	Conjunto básico
Univariada ampliada	Intención de voto (multinomial)	Conjunto ampliado
Encadenada básica	Intención de voto (multinomial) Recuerdo de voto (multinomial) Ideología (ordinal)	Conjunto básico
Encadenada ampliada	Intención de voto (multinomial) Recuerdo de voto (multinomial) Ideología (ordinal)	Conjunto ampliado

**Conjunto básico:** Sexo, edad, nivel educativo, recuerdo de voto e ideología **Conjunto ampliado:** Sexo, edad, nivel educativo, recuerdo de voto, ideología, evaluación de la situación económica y evaluación de la situación política

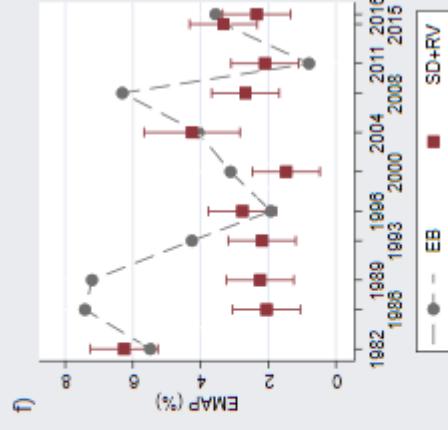
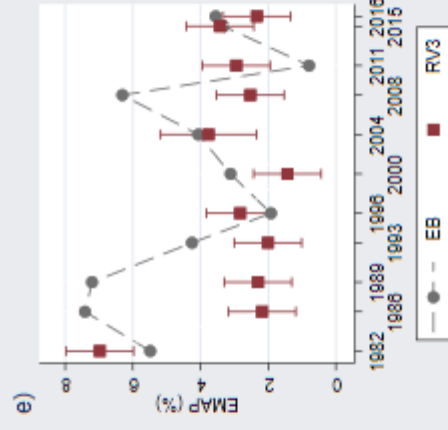
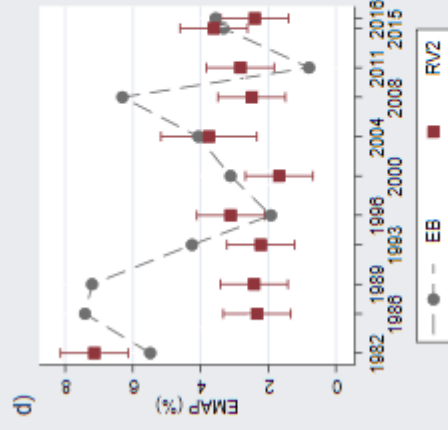
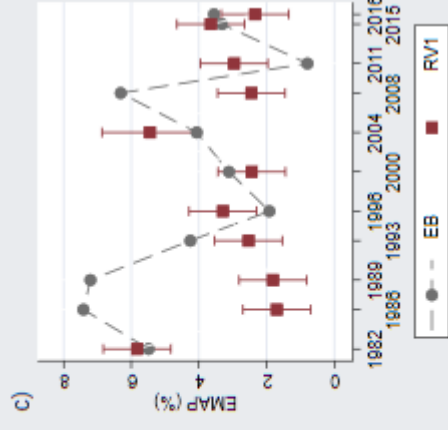
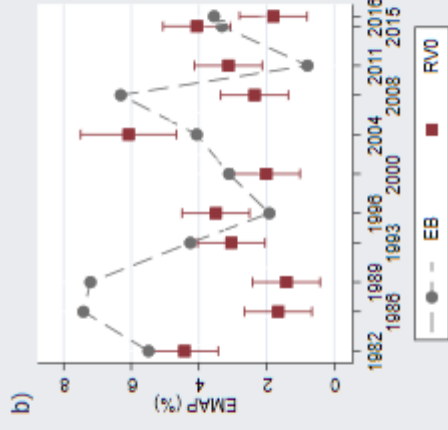
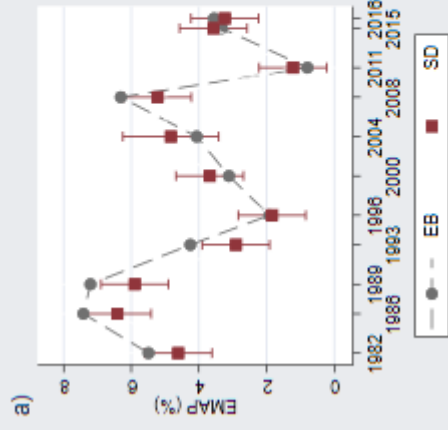
# Indicios



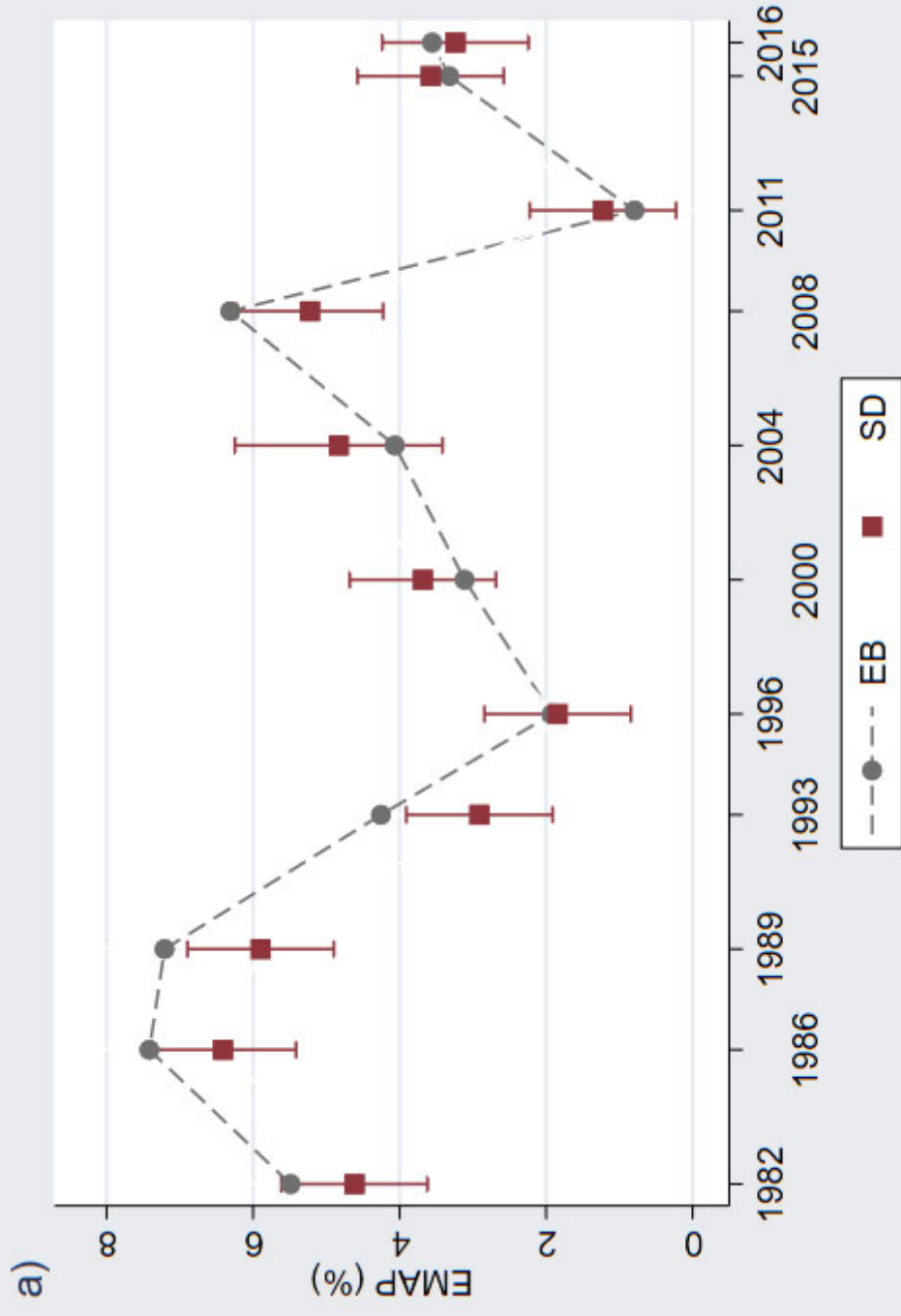
# ¿Imputar el recuerdo de voto?



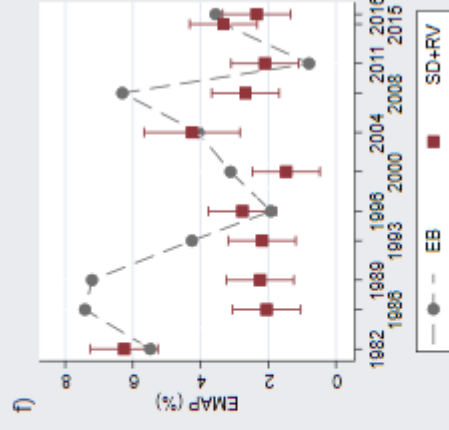
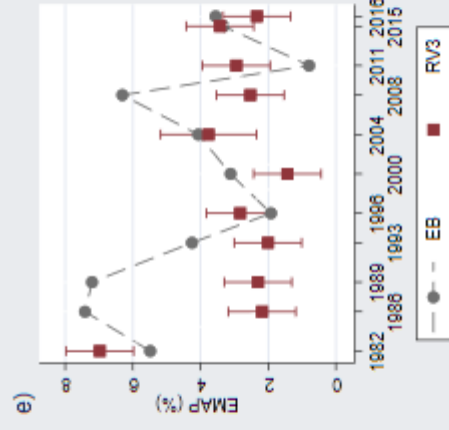
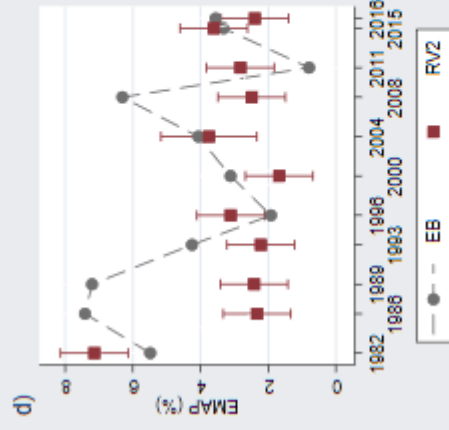
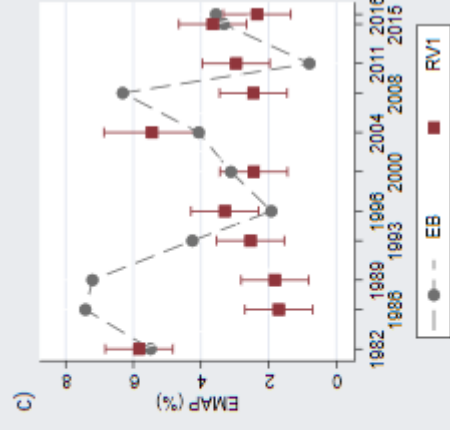
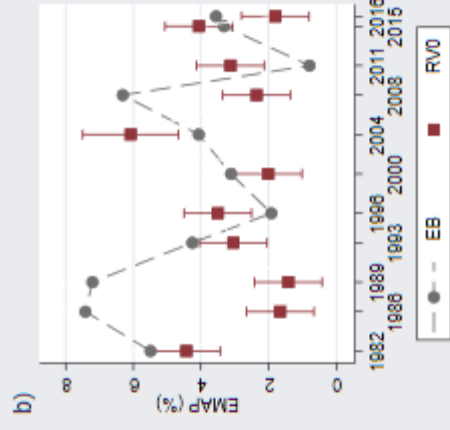
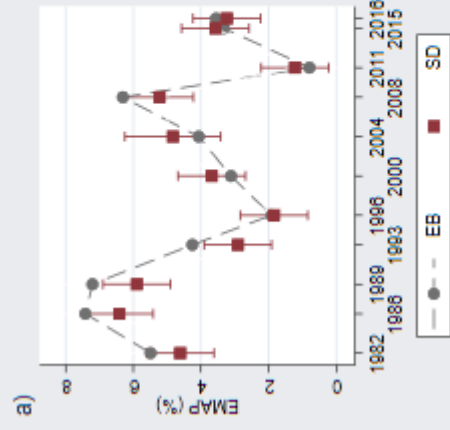
# ¿Ponderar por el recuerdo de voto?



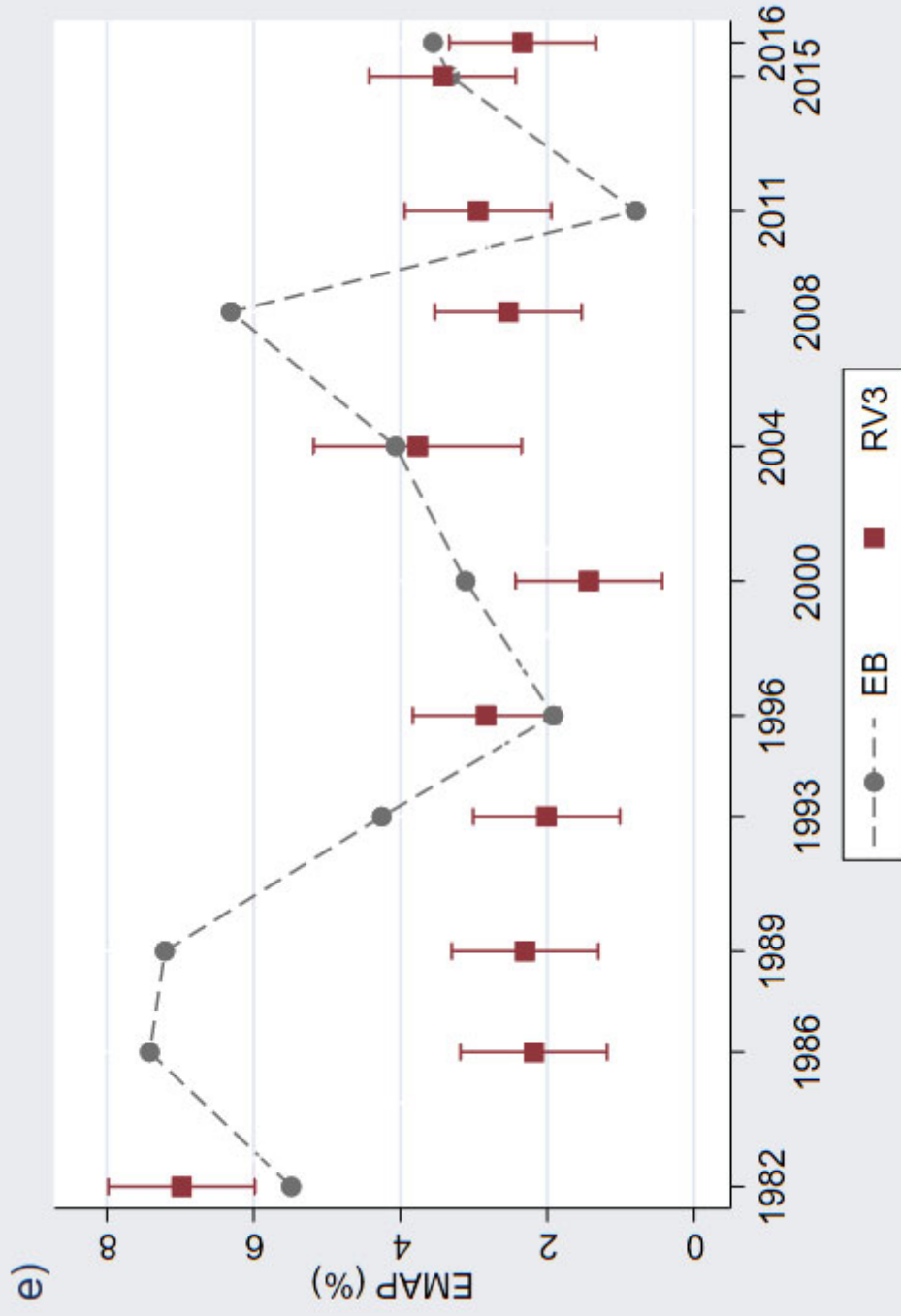
# ¿Ponderar por el recuerdo de voto? (SD)



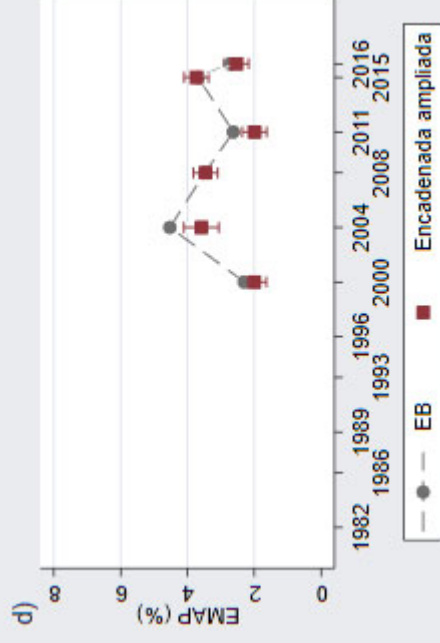
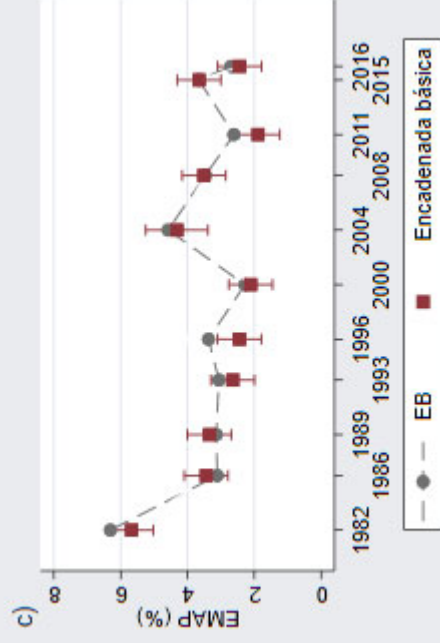
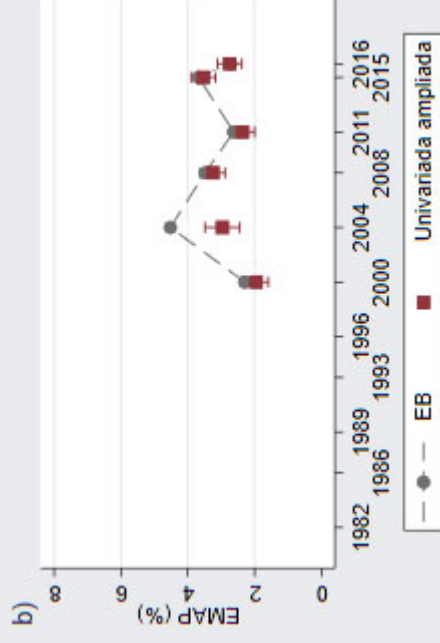
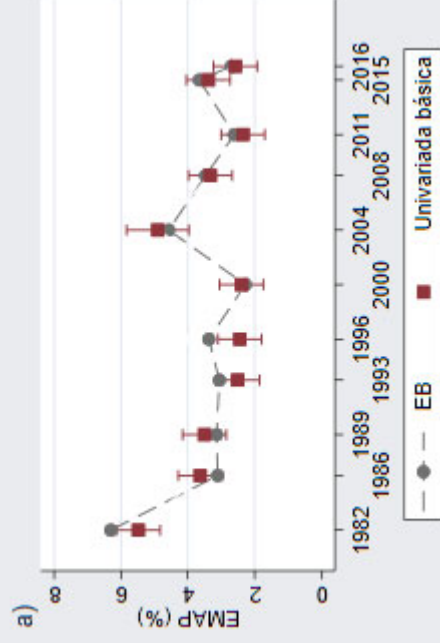
# ¿Ponderar por el recuerdo de voto?



# ¿Ponderar por el recuerdo de voto? (RV3)



# ¿Imputar la intención de voto?



Final

# Conclusiones

- Uso de la ponderación por recuerdo de voto (**H1**) mejora la precisión de las estimaciones aunque añadir las variables sociodemográficas (**H2**) no aporta un incremento en la precisión
- Usar la imputación múltiple para corregir la variable recuerdo de voto (**H3**) no repercute en el nivel de ajuste de la propia variable y tampoco en la precisión de la estimación
- El uso de la imputación múltiple para tratar los casos perdidos de la variable de intención de voto (**H4**) apenas afecta a la precisión de las estimaciones
- El momento del ciclo político, según se trate de elecciones de continuidad o de cambio (**H5**), sigue estando relacionado con la capacidad de la ponderación de mejorar la estimación de voto tras los cambios en el sistema de partidos



# Gracias

[pablocal@usal.es](mailto:pablocal@usal.es) | [@pablocalv](https://twitter.com/pablocalv)

Cabrera-Álvarez, Pablo y Escobar, Modesto (2019). «El efecto de la ponderación y la imputación en el sesgo de los estudios electorales en España». Revista Española de Investigaciones Sociológicas, 165: 45-64.

Esta investigación cuenta con el apoyo del programa de becas de la Fundación "La Caixa"  
LCF/BQ/ES16/11570005



# Recuerdo voto CIS

Pregunta 35aa  
 RECUERDO DE VOTO EN ELECCIONES GENERALES DE 2016.

PP	17,3
PSOE	23,1
Unidos Podemos	8,2
C's	9,1
En Comú Podem	1,9
Compromís-Podemos-EUPV	1,8
ERC	3,0
CDC	1,1
En Marea	0,7
EAJ-PNV	0,8
EH Bildu	0,7
CC-PNC	0,1
Otros	1,6
En blanco	1,8
Voto nulo	0,3
No tenía edad	2,9
No votó	14,5
No tenía derecho a voto	0,4
No recuerda	4,4
N.C.	6,2
(N)	(2.989)