

Electoral forecasting with Stata

Four years later

Modesto Escobar & Pablo Cabrera

University of Salamanca (Spain)

2016 Spanish Stata Users Group meeting

Barcelona, 20th October, 2016



Contents

- 1 Introduction
- 2 Theory
- 3 Data
- 4 Design
 - Predictors
 - D. variable
- 5 Results
 - Main results
 - Graphs
- 6 Conclusions

Introduction

Multiple imputation

- In this work, we use post-stratification and multiple imputation techniques to produce accurate predictions of electoral outcomes at the aggregate level from individual data on electoral behavior.
- Imputation allows us to predict the electoral choice of non-respondent interviewees in electoral surveys thus producing more accurate predictions.
- There is empirical evidence showing that the electoral behavior of voters who answer survey questions about voting intentions differs from those who do not state which party they are going to vote for.
- We evaluate 60 different ways of predicting electoral results through the last twelve Spanish general elections using preelectoral surveys conducted by the CIS (The Spanish Center for Sociological Research).



Contents

- 1 Introduction
- 2 Theory
- 3 Data
- 4 Design
 - Predictors
 - D. variable
- 5 Results
 - Main results
 - Graphs
- 6 Conclusions

Theory

Putting together

- From an academic perspective, what is original about this research is that it unites two different strands of the literature on voting:
 - Studies on electoral forecasting
 - Studies on voting behavior
- We emphasize our contribution, since pollsters and research institutes use different procedures to estimate vote distributions, although these procedures are not well-known and rely on non-statistical inferences.

Theory

Approaches to explain electoral behavior

- Electoral forecasting based on the data of voters who declare their voting intentions may be misleading, and the direction and the size of the bias cannot be anticipated.
- In order to impute electoral choices to individual voters, we need to rely on a theoretical model of electoral behavior to decide which relevant variables have to be considered to predict voter decisions.
- There are three different approaches to the explanation of electoral behavior: the party identification approach, the rational voter approach, and the socio-structural approach.

Contents

- 1 Introduction
- 2 Theory
- 3 Data**
- 4 Design
 - Predictors
 - D. variable
- 5 Results
 - Main results
 - Graphs
- 6 Conclusions

Data

- The source of our data is the Center for Sociological Research (CIS). We use 12 pre-electoral polls after the Constitution of 1978 (approved 3 years after the death of Franco):
- The samples are randomly stratified by constituencies (52) in three stages (localities-households-individuals), and were conducted at home one month before polling-day.
- Around 230,000 people were interviewed between 1979 and 2016. The distribution of sample sizes was as follows:

Year and sizes of samples					
1st cycle		2nd cycle		3rd cycle	
Year	n	Year	n	Year	n
1979	25,516	1996	6,642	2015	17,452
1982	24,832	2000	24,040	2016	17,488
1986	25,667	2004	24,109		
1989	27,421	2008	18,221		
1993	2,503	2011	17,236		
	105,939		90,248		34,940

Contents

- 1 Introduction
- 2 Theory
- 3 Data
- 4 Design**
 - Predictors
 - D. variable
- 5 Results
 - Main results
 - Graphs
- 6 Conclusions

Design

Predictors

- We wished to test and compare different methods of vote estimation through the use of different statistical procedures :
 - Questions related: *vote intention*, *vote plus sympathy* or *vote plus sympathy plus memory*.
 - Sample considered: *Complete* or limited to those who already had a *fixed voting intention*.
 - *Post-stratification* by vote memory or *non* post-stratification.
 - Imputation (*univariate* or *chained, simple* or *enhanced*) or *non* imputation
 - Simple imputation by sex, age, level of studies and ideology.
 - Enhanced imputation by previous variables plus evaluation of government and economy.

Test structure (60)

The design is

Question(3)XSample(2)XImputation(5)XPost-stratification(2)

Questions	Sample correction									
	Without post-stratification					Post-stratified				
	Models									
	Univariate		Chained			Univariate		Chained		
	Estim.	Simple	Enhan.	Simple	Enhan.	Estim.	Simple	Enhan.	Simple	Enhan.
All the sample										
Vote	1	7	13	19	25	31	37	43	49	55
Vote & sympathy	2	8	14	20	26	32	38	44	50	56
Vote & sym. & memory	3	9	15	21	27	33	39	45	51	57
Only voters										
Vote	4	10	16	22	28	34	40	46	52	58
Vote & sympathy	5	11	17	23	29	35	41	47	53	59
Vote & sym. & memory	6	12	18	24	30	36	42	48	54	60

WAME

Weighted absolute mean error

- For a multiparty system, a convenient indicator to assess a forecast is the weighted absolute mean error WAME:

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

- where p_k are the real results in proportions to every political party or coalition (k), and \hat{p}_k is every estimation or imputation.
- According to this design, 720 WAMES are possible: 60 predictions for each election out of 12.

Contents

- 1 Introduction
- 2 Theory
- 3 Data
- 4 Design
 - Predictors
 - D. variable
- 5 Results**
 - Main results
 - Graphs
- 6 Conclusions

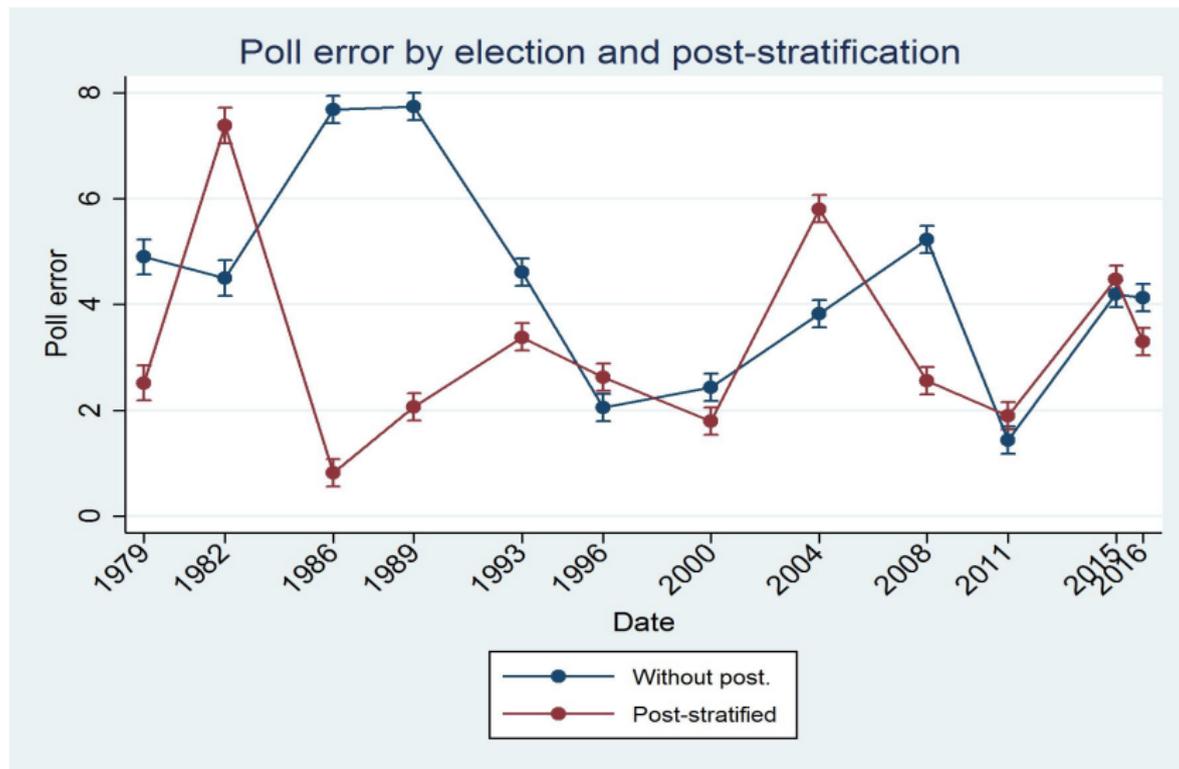
Estimations by post-stratification and imputation

2nd part: Post-estratification and election

		Estimated											
		Vote				Vote + Sympathy				Vote+Sympathy+Memory			
		Complete		Voters		Complete		Voters		Complete		Voters	
Actual		No P.	Post.	No P.	Post.	No P.	Post.	No P.	Post.	No P.	Post.	No P.	Post.
PSOE	36.0	38.1	33.7	38.6	33.9	36.4	32.9	37.1	33.3	37.6	34.1	38.2	34.2
PP	32.2	28.1	31.9	28.8	32.1	26.2	31.2	27.3	31.5	26.3	31.5	27.4	31.7
IU/Pod.	9.3	10.4	9.6	10.7	9.7	10.2	9.7	10.3	9.7	10.0	9.7	10.2	9.7
Otros	22.5	23.3	24.7	22.0	24.2	27.3	26.1	25.2	25.5	26.1	24.8	24.2	24.4
WAME:		2.3	1.5	2.3	1.2	3.2	2.3	2.7	1.9	3.4	1.5	2.8	1.3
		Imputed											
		Vote				Vote + Sympathy				Vote+Sympathy+Memory			
		Complete		Voters		Complete		Voters		Complete		Voters	
Actual		No P.	Post.	No P.	Post.	No P.	Post.	No P.	Post.	No P.	Post.	No P.	Post.
PSOE	36.0	38.0	33.1	38.4	33.3	36.3	32.6	37.0	32.9	37.6	33.8	38.1	33.9
PP	32.2	29.1	33.6	29.8	33.9	27.6	32.9	28.8	33.2	27.7	33.2	28.9	33.5
IU/Pod.	9.3	10.1	9.6	10.4	9.6	10.4	9.8	10.5	9.8	10.3	9.7	10.5	9.8
Otros	22.5	22.8	23.7	21.3	23.1	25.7	24.7	23.6	24.1	24.4	23.2	22.4	22.8
WAME:		1.8	1.8	2.0	1.7	2.4	2.0	1.8	1.9	2.5	1.3	1.9	1.3

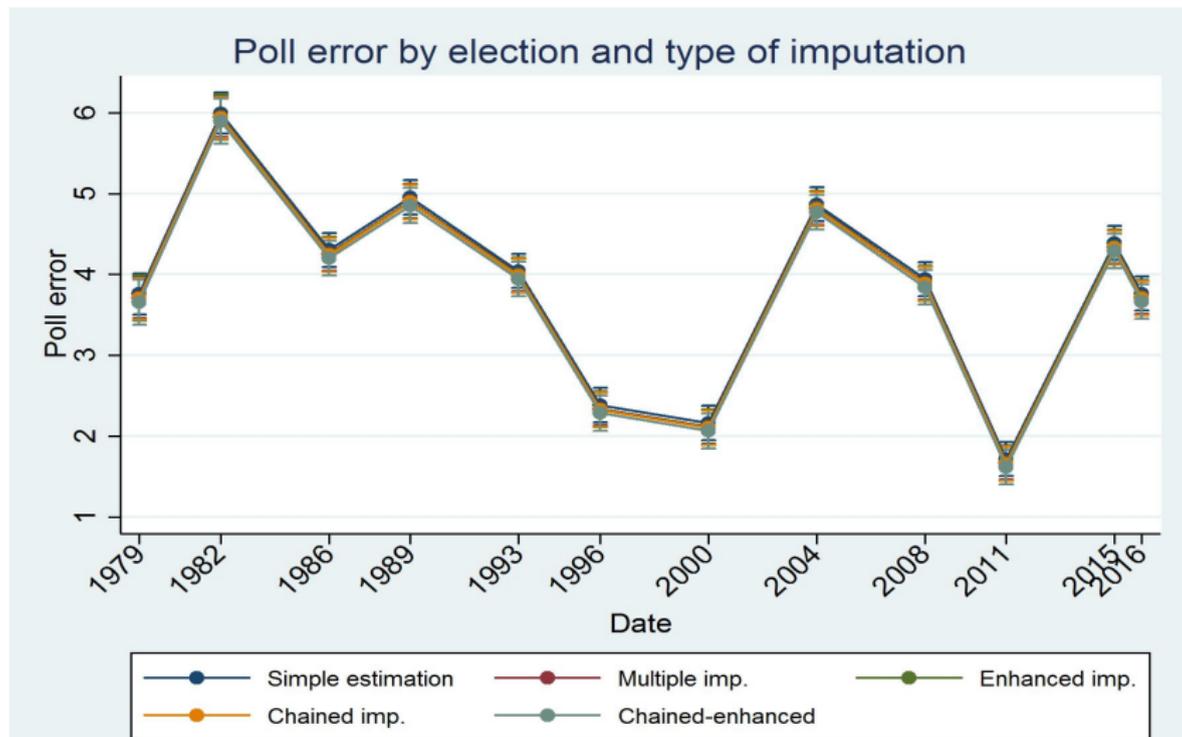
Poll errors

All estimations and imputations (by post-stratification)



Poll errors

By methods: estimation and 4 imputations



Contents

- 1 Introduction
- 2 Theory
- 3 Data
- 4 Design
 - Predictors
 - D. variable
- 5 Results
 - Main results
 - Graphs
- 6 Conclusions

Conclusions

- In predicting electoral results, mix voting intention with party sympathy and select voters with a fixed voting intention.
- Post-stratification has been extensively used in pre-electoral surveys, but it does not always produce the optimal result.
- Post-stratification works better when the incumbent remains in power.
 - This can be attributed to social desirability or hidden voting intentions.
- Imputation seems to work well. But it has less impact than post-stratification.
- Nonetheless, the simultaneous use of both doesn't necessarily improve estimation, since similar results are produced.