Electoral predictions by post-stratification and imputation

A. M. Jaime (amjaime@gmail.com) y M. Escobar (modesto@usal.es)

Spanish Stata Users Group meeting (2012)

12 de septiembre de 2012

Introduction

The general framework of this work is to obtain the best method to predict electoral outcomes using surveys. Our work is relevant for a Stata User Meeting because Stata is well suited to deal easily with three complex operations involved in electoral forecasting:

- First, we need to deal with weights in complex samples by using the module **svy**, which implements sample calibration by using post-strata.
- On the other hand, we need to use imputation procedures, which are implemented by other Stata module updated in version 12: **mi** (multiple imputation).
- Finally, we use Mata, which allows us to use matrices in order to compute a special index for the evaluation of the estimated models: the absolute weighted average error.

Forecasting

"To forecast an election means to declare the outcome before it happens" (Lewis-Beck, 2005). The literature on electoral forecasting has focused almost exclusively on predicting aggregate electoral outcomes using other aggregate magnitudes such as economic growth, unemployment, or popularity rates. Predictions derived from econometric models perform relatively well, but electoral decisions at the individual level become a black-box.

On the other hand, the literature on electoral behavior has grown in recent decades to explain the micro-foundations of electoral choices, but the aim of this line of research is to explain voters' behaviors instead of producing accurate predictions of electoral outcomes.

Forecasting

In this work we use 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 and thus producing more accurate predictions.
There is empirical evidence showing that the electoral behavior of voters who answer to survey questions about voting intentions differs of those who do not say which party they are going to vote for. Moreover, the non-respondents have been more inclined to support different parties in different political periods (Urquizu-Sancho, 2006).

Theoretical framework

Electoral forecasting based upon the data on voters who declare their voting intentions will be misleading and we cannot anticipate the direction and size of the bias.

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 we have to consider to predict voters' decisions.

There are three different approaches to explain electoral behavior: the *party identification approach*, the *rational voter approach*, and the *socio-structural approach*.

Each approach is based on different theoretical assumptions and focuses on different predictors of electoral behavior at the individual level.

Party identification

The theory of *party identification* argues that voters' choices depend on individual allegiances to political parties. These party attachments develop during the early years of childhood (through the socialization process) and become and enduring influence on electoral behavior in adulthood. Harrop and Miller (1987) summarize the main points of this model of electoral behavior:

- Most voters develop a party identification, which is learnt from the family.
- Party identification has not only a direct impact on electoral choices but an indirect effect because party identification also affects how voters evaluate policies and candidates.

Party identification

- The strength of party identification increases with time (positive correlation between party identification and age). Changes in party identification are mostly due to social or geographical mobility.
- Voters may vote eventually against their party identification because of short-term shocks, but this does not change party identification. After the shock is gone voters will vote in line with their party identifications again.

Rational voter

The theory of the *rational voter* is based upon the economic approach to politics. Voters have self-centered motivations and behave like utility maximizers.

The political arena is a market in which parties compete for votes in order to get into power.

On the supply-side, parties propose electoral platforms and each voter chooses the platform expected to produce the best outcome for her/himself.

According to Downs (1957), voters compute the benefits they have got from the party in power and the expected utility from choosing a new government. If the difference is positive they will vote for the incumbent. Otherwise they will vote for the challenger.

Rational voter

The basic device that voters use to compute their utilities is the situation of the economy, since governments are supposed to be responsible for the economic outcomes.

Therefore, voters' evaluations of the economy will be the most relevant variables explaining electoral choices. Those who believe that economy is getting better will vote for the incumbent.

At the aggregate level, changes in electoral outcomes can be explained by changes in the economic situation.

Socio-structural approach

The *socio-structural* theory of voting outlines the relevance of social variables as predictor of electoral choices.
According to this model, electoral behavior is determined by voters' position on the social structure. Therefore, individuals belonging to the same social group will behave in similar ways.
Social groups could be defined by social class, gender, ethnicity, age or any other relevant variable.
Political parties are supposed to be a device to represent interests' groups in the political arena. Hence, their constituency will be group of voters they represent.

Socio-structural approach

The boundaries of these social groups have been defined historically according to the relevant cleavages that exist in each society (i. e. religious conflicts, economic conflicts, ...). These cleavages are the basis for social mobilization that produces political action.
Although cleavages evolve historically, their effects on voting behavior remain stable over time.
Therefore, structural variables (class, gender, age) will be the most relevant variables to predict electoral choices at the individual level.

Approach

From the perspectiv of the academic leterature, the maun noverlty of this research is to put together two different strands of the literature on voting:

- The studies on electoral forecasting
- The studies on voting behavior

We emphasize the contribution to the academic literature, since pollster and research institutes use different procedures to estimate vote distributions, although these procedure are not well-known and rely on non-statistical inferences.

Data

Data come from the Center for Sociological Research (CIS). We use the last two electoral polls:

- The pre-electoral survey was conducted in October (one month before the polls-day): **17.236** interviewed people sampled polietapicly.
- The post-electoral survey, conducted between November the 24th and January the 15th , with **6.062** subjects from a planned sum of 7.547, among those that in the former study didn't mind to be interviewed again.

Design

We want to test and compare different ways of vote estimation through the use of different statistical procedures :
a) Pre-electoral or post-electoral survey
b) Post-estratification or non post-estratification
c) Imputation or non imputation
At the same time, we want to test the different hypothesis about determinants of voting behavior:
a) Previous behavior(remembered vote)
b) Identification (ideology)
c) Rational behavior (govern evaluation , economic situation assessment)
d) Socio-demographic factors (level of education, age, gender)

Stratification

- To stratify a sample consists in making a simple random sample in every relevant division of the population. Obviously, one of the most relevant divisions in electoral studies is constituency.
  - In Spain, there are 52.
- We have to establish a priory the number of elements of every stratum .
  - Generally, this number is proportional to its populational size, but in big size electoral samples, it is frequent to over-sample small constituencies, so small errors may be made.

Weighting

When a sample does not have proportional representation, its member has to be weighted through a coefficient (w_k) whose value must be:

$$w_k = n_k^* / n_k$$

where $n_k^* = N_k / N$; n_k, is the actual size of the sample in the k stratum; N_k, is the populational size of every stratum, and N, the whole size of the population.

- The weight variable has to have a value for every subject; but there will only be k different values, let's say, as many strata as the sample has.

Treatment of non proportional samples

- When there is non proportionality, it is convenient to employ the Stata module **svy**
- The preliminary order of this module is *svyset*
  - Its syntax for stratified samples is the following:

svyset _*n* **[pweight=***peso***]**, **strata(***estrato***)**

where *peso* is the variable that takes account of weight and *estrato* is the variable that identifies each stratum

Posterior treatment of tabulations

Once the structure of weighting is defined, the subsequent analysis must be preceded by the Stata preinstruction **svy**
For example, a univariate distribution can be obtained in this way:

svy: tab variable [, options]

Among specific options in tabulation, the following must be remarked:
cell count obs ci

Outcome of svy: tab (just one variable)


```
. quietly: svyset _n [pweight=peso], strata(strato)

. svy: tab prov if prov>50, coun cell obs per
(running tabulate on estimation sample)

Number of strata   =         11            Number of obs      =        393
Number of PSUs     =        393            Population size    =    55.0697
                                           Design df          =        382

------------------------------------------------
Provincia |      count   percentages        obs
----------+-------------------------------------
    Ceuta |      29.42         53.42        200
  Melilla |      25.65         46.58        193
          |
    Total |      55.07           100        393
------------------------------------------------
  Key:  count     =  weighted counts
        percen~s  =  cell percentages
        obs       =  number of observations
```

Post-stratification

- We call the artificial procedure to repair the representation of a sample with unintended biased results, post-stratification (also calibration).
  - It is different from weighting, because the weight could not be calculated a priory, but a posteriori, once we detect a clear bias in a particular sample.
    - That is the case of polls, due to diverse reasons. In these studies, the most used criterion to calibrate samples is memoirs of vote.

Post-stratified weighting

The weight coefficient to calibrate must be applied after a weight to fix an stratification, according to the following formula:

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

being $\widehat{N}_l = n_l w_k N \,/\, n$, i.e., the estimate size of a populational stratum after stratificational correction and before calibration.

- Note that, if not divided by n, frequencies would be in populational figures, instead of sample ones.

Post-stratification syntax

In order to post-stratify, you have to add two options to the precommand **svyset** : poststrata(post-estrato) and postweight(tamaño)
So, to combine stratification and post-stratification, you can write:

svyset _n **[pweight=**peso**], strata(**estrato**)** ///
postrata(postestrato**) postweight(**tamaño**)**

being **tamaño**, the post-stratum's real size and postestrato the group variable indicating the post-stratum which every subject belongs to.

How to give weights?

- The long way
- The short way
- Vectorial mode (matricial)

The long way

Use if:

generate peso=0
replace peso=0.8 **if** prov==1
replace peso=0.7 **if** prov==2
...

The short way

Use recode:

recode prov **(**1=0.8**)(**2=0.7**)**..., **into(**peso**)**

Matricial way

Through the use of vectors (matrices)

```
matrix Pesos=[0.8\0.7\...]
for numlist 1/52: replace peso=Pesos[X,1] if prov==X
```

Unweighted table


```
. tab vote
```

Vote pre-2011	Freq.	Percent	Cum.
PP	5,379	47.10	47.10
PSOE	3,063	26.82	73.92
IU	674	5.90	79.82
Otro	2,304	20.18	100.00
Total	11,420	100.00	

Weighted table through strata (tabulate)

```
. tab vote [iweight=peso]
```

Vote pre-2011	Freq.	Percent	Cum.
PP	5,252.1245	45.59	45.59
PSOE	3,079.7077	26.74	72.33
IU	775.1442	6.73	79.06
Otro	2,412.381	20.94	100.00
Total	11,519.3574	100.00	

Weighted table through strata (svy: tabulate)


```
. svy: tab vote, count cell obs format(%5.2fc)
(running tabulate on estimation sample)

Number of strata   =        52              Number of obs      =      11420
Number of PSUs     =     11420              Population size    = 11519.357
                                            Design df          =      11368

-------------------------------------------------
Vote      |
pre-2011  |      count  proportions           obs
----------+--------------------------------------
       PP |    5252.12         0.46       5379.00
     PSOE |    3079.71         0.27       3063.00
       IU |     775.14         0.07        674.00
     Otro |    2412.38         0.21       2304.00
          |
    Total |   11519.36         1.00      11420.00
-------------------------------------------------
  Key:  count      =  weighted counts
        propor~s   =  cell proportions
        obs        =  number of observations
```

Weighted table through poststrata (svy: tabulate)


```
. quietly:svyset _n [pweight=peso], strat(prov) poststrata(recuerdo) postweight(po

. svy: tab vote, count cell obs format(%14.2fc)
(running tabulate on estimation sample)

Number of strata    =         52              Number of obs      =       9365
Number of PSUs      =       9365              Population size    =   25734866
N. of poststrata    =          9              Design df          =       9313
```

Vote pre-2011	count	proportions	obs
PP	12,291,034.31	0.48	4,414.00
PSOE	7,110,842.63	0.28	2,661.00
IU	1,528,672.38	0.06	572.00
Otro	4,804,316.68	0.19	1,718.00
Total	25,734,866.00	1.00	9,365.00

Multiple imputation

Multiple imputation, proposed by Rubin (1987), is aimed to build new datasets giving new values to missing cases, assigned by an stochastic function implying other related variables

In contrast to single imputation, which only makes one estimation, MI makes a number m of $\widehat{Q}$ estimations, that gives way to a new estimation $\overline{Q}$ with $\overline{U}$ internal variance and B external variance

Impute methods

There are different imputation methods to obtain $\widehat{Q}$ for missing cases. We are going to use only one instance of each general method:

- Univariate, only imputes one variable (vote in our case)
- Chained, that uses iterative series of imputations for each non-regular variable of our model as a function of the other variables (vote, vote memoirs, ideology, govern evaluation and economic evaluation)

Codes to impute (I)

First step: To declare multiple-imputation data
mi set {flong|wide|mlong|flongsep}
mi svyset **peso**
a) _n [pweight=peso], strat(prov)
b) _n [pweight=peso], strat(prov) poststrata(recuerdo)
postweight(pobl)
Second step: To register and classify variables (imputed, regulars and passives)
mi register {imputed | regular | passive} **varlist**
Third step : To analyze missing patterns
mi misstable {summarize|patterns|tree|nested} **varlist**

Codes to impute (II)

Fourth step: To impute properly

mi impute **method**

a) mlogit voto i.recuerdo i.ideologia estudios i.sexo edad

b) chain (mlogit) voto recuerdo (ologit) gobierno ideologia economica ///

$=$ estudios i.sexo edad

Fifth step: To estimate from imputations

mi estimate: svy: proportion vote

mi estimate, post: svy: regress vote **varlist**

How to measure the accuracy of our estimations?

We need:

- Real data (missing in nearly all research)
- Survey estimates (trough different methods)
- A formula
- To apply the formula to the data

Real data

You can have real data in a dataset and convert then into a Stata matrix:

use "Matriz Electoral Nacional.dta"**, clear**
mkmat PSOE-Otros, **rownames(**Año**) matrix(**E**)**
matrix Real=E["2011",.]

-

Or you can write them directly:

matrix Real=(.446, .288, .069, .197)

Forecasted data

You have to count on the estimation results of svy:tab
The target matrix (vector) is **e(Prop)**.
matrix Pronostico=**e(Prop)**

Formula

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

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

where p_k are the real results in proportions for every political option (k), and $\widehat{p}_p$ are every estimation obtained from the subject's answers.
Obviously, this error measure only can be obtained after the polling day.

Formula application

These three alternatives can be used:

- Stata loop
- Mata function
- Mata call

Stata code


```
local NP=rowsof(Pronostico)
scalar wame=0
forvalues i=1(1)`NP' {
scalar wame=scalar(wame)+abs(Pronostico[`i',1]-Real[`i',1])*
        Real[`i',1]*100
}
```

Wame with a Mata function


```
mata:
function wame(a, b)
{
X=(st_matrix(a))
Y=(st_matrix(b))
R=sum((abs(X-Y)):*Y)
st_numscalar("wame", R:*100)
}
end

mata: wame("Pronostico",Real")
```

Wame with Mata call

It is also possible to calculate wame with just one line of code using mata call:

```
mata: st_numscalar("Wame",
sum((abs(st_matrix("Pronos")-st_matrix(Real"))
:*st_matrix(Real"):*100)))
```

Test structure (20)

The design is ProcedureXCalibrationXRegressionXMethodXSurvey. However, the so called mere estimation does not differ neither with regressions nor methods

		Survey							
		Preelectoral				Postelectoral			
		Method				**Method**			
		Univariate **Regression**		Chained **Regression**		Univariate **Regression**		Chained **Regression**	
Procedure	**Calibration**	Simple	Enhan.	Simple	Enhan.	Simple	Enhan.	Simple	Enhan.
Estimated	Without	1	1	1	1	11	11	11	11
	Calibrated	2	2	2	2	12	12	12	12
Imputed	Without	3	4	5	6	13	14	15	16
	Calibrated	7	8	9	10	17	18	19	20

Missing tree structure (Preelectoral)

Preelectoral, missing vote(*)					
Vote	Ideolog.	Memoir	Govern.	Econom.	%
24.1%	**17.2%**	**14.1%**	**3.2%**	**0.7%**	
4,149	**1,165**	**627**	**84**	**9**	<1
				75	<1
			543	**7**	<1
				536	3
		538	**62**	**11**	<1
				51	<1
			476	**3**	<1
				473	3
	2,984	**933**	**46**	**3**	<1
				43	<1
			887	**5**	<1
				882	5
		2,051	**72**	**5**	<1
				67	<1
			1,979	**10**	<1
				1,969	11

(*)Bold for missing cases

Preelectoral, no missing vote(*)					
Vote	Ideolog.	Memoir	Govern.	Econom.	%
24.1%	**17.2%**	**14.1%**	**3.2%**	**0.7%**	
13,052	**1,793**	**213**	**16**	**1**	<1
				15	<1
			197	**2**	<1
				195	1
		1,580	**116**	**14**	<1
				102	<1
			1,464	**9**	<1
				1,455	8
	11,259	**648**	**23**	**3**	<1
				20	<1
			625	**3**	<1
				622	4
		10,611	**137**	**12**	<1
				125	<1
			10,474	**25**	<1
				10,449	61

Missing tree structure (Postelectoral)

Postelectoral, missing ideologie(*)

Ideolog.	Memoir	Vote	Govern.	Econom.	%
14.4%	**12.5%**	**9.5%**	**2.9%**	**0.5%**	
875	**189**	**97**	**11**	**0**	0
				11	<1
			86	**2**	<1
				84	1
		92	**7**	**0**	0
				7	<1
			85	**0**	0
				85	1
	686	**107**	**10**	**1**	<1
				9	<1
			97	**0**	0
				97	2
		579	**51**	**3**	<1
				48	<1
			528	**1**	<1
				527	9

(*)Bold for missing cases

Postelectoral, no missing ideologie(*)

Ideolog.	Memoir	Vote	Govern.	Econom.	%
14.4%	**12.5%**	**9.5%**	**2.9%**	**0.5%**	
5,181	**570**	**157**	**9**	**0**	0
				9	<1
			148	**1**	<1
				147	2
		413	**15**	**4**	<1
				11	<1
			398	**3**	<1
				395	5
	4,611	**212**	**6**	**1**	<1
				5	<1
			206	**2**	<1
				204	3
		4,399	**69**	**6**	<1
				63	1
			4,330	**7**	<1
				4,323	73

Main results

Preelectoral: univariate models

Preelectoral univariate models									
		Simple model W. calibr.		Simple model Calibrated		Enhanced mod. W.calibr.		Enhanced mod. Calibrated	
Vote	**Real**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**
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

Errors	**Real**	**Est.**	**Real**	**Est.**	**Real**	**Est.**	**Real**	**Est.**
Estimated	1.30		2.00		1.30		2.00	
Imputed	0.70	1.30	2.10	0.20	0.70	1.30	2.10	0.20

Main results
Preelectoral: chained models

		Preelectoral chained models							
		Simple model		Simple model		Enhanced mod.		Enhanced mod.	
		W. calibr.		Calibrated		W.calibr.		Calibrated	
Vote	**Real**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**
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

Errors		**Real**	**Est.**	**Real**	**Est.**	**Real**	**Est.**	**Real**	**Est.**
Estimated		1.30		2.00		1.30		2.00	
Imputed		0.80	1.40	2.10	0.10	0.90	1.10	2.30	0.40

Main results

Postelectoral: univariate models

		Preelectoral univariate models							
		Simple model		Simple model		Enhanced mod.		Enhanced mod.	
		W. calibr.		Calibrated		W.calibr.		Calibrated	
Vote	**Real**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**
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	28.0	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

Errores	Real	Est.	Real	Est.	Real	Est.	Real	Est.
Estimado	0.50		2.00		0.50		2.00	
Imputado	0.50	0.10	2.40	0.40	0.50	0.10	2.40	0.40

Main results
Postelectoral: chained models

		Postelectoral chained models							
		Simple model		Simple model		Enhanced mod.		Enhanced mod.	
		W. calibr.		Calibrated		W.calibr.		Calibrated	
Vote	**Real**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**	**Est.**	**Imp.**
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

Errores	**Real**	**Est.**	**Real**	**Est.**	**Real**	**Est.**	**Real**	**Est.**
Estimado	0.50		2.00		0.50		2.00	
Imputado	0.40	0.20	2.00	0.10	0.40	0.10	2.10	0.10

Main Results

Results obtained from imputation are quite accurate.

- Imputed equations produces more accurate predictions than estimated equations in pre-electoral survey. Therefore, imputation techniques allow us to improve electoral forecasting.
- However, estimated equations perform better when we use strata based on previous vote. This is because we are losing information for those who did not vote in previous election.
- Simple models preform relatively well. Error in chained models is greater than in univariate models.

Main Results

Enhanced models including more variables do not reduce error. Possible explanations:

- Endogeneity. Some authors argue that individual evaluations of the economy are colored by ideology or previous vote.
- Economic perceptions have low variance in this election. Most voters (including government supporters) perceive that the economy was in very bad shape by the time the election took place.

Determinants of vote

	Original equation			Imputed equation		
	PP	PSOE	IU	PP	PSOE	IU
Did not vote	1.930***	3.400***	1.859***	1.939***	3.442***	1.905***
	(0.181)	(0.328)	(0.392)	(0.188)	(0.333)	(0.380)
Voted PSOE	2.023***	4.596***	2.451***	2.023***	4.658***	2.526***
	(0.173)	(0.318)	(0.361)	(0.177)	(0.320)	(0.355)
Voted PP	4.014***	1.868***	1.258**	3.997***	1.931***	1.283**
	(0.197)	(0.416)	(0.574)	(0.197)	(0.416)	(0.566)
Voted IU	0.955***	1.797***	4.269***	0.907***	1.859***	4.361***
	(0.346)	(0.439)	(0.382)	(0.348)	(0.452)	(0.371)
Voted CiU	-0.706**	0.704	-0.546	-0.740**	0.728	-0.720
	(0.318)	(0.463)	(1.075)	(0.307)	(0.474)	(1.066)
Voted PNV	-2.709***	-0.512	-0.713	-2.735***	-0.410	-0.647
	(0.745)	(0.650)	(1.062)	(0.723)	(0.660)	(1.012)
No ideology	-0.164	0.0637	-0.630	-0.149	0.0118	-0.792*
	(0.145)	(0.164)	(0.425)	(0.133)	(0.159)	(0.410)
Left	-2.162***	0.930***	1.728***	-2.150***	0.866***	1.694***
	(0.210)	(0.142)	(0.208)	(0.238)	(0.151)	(0.216)
Center-left.	-1.257***	0.851***	1.007***	-1.262***	0.822***	1.003***
	(0.117)	(0.106)	(0.182)	(0.113)	(0.101)	(0.192)
Center-right.	1.327***	-1.021***	-1.082	1.306***	-1.016***	-1.088
	(0.158)	(0.320)	(0.936)	(0.160)	(0.316)	(0.926)
Right	1.465***	-0.727	-1.252	1.455***	-0.732	-1.172
	(0.295)	(0.622)	(1.055)	(0.296)	(0.596)	(1.081)

Determinants of vote

	Original equation			Imputed equation		
	PP	PSOE	IU	PP	PSOE	IU
Education	-0.284***	-0.297***	-0.0981**	-0.286***	-0.298***	-0.0906**
	(0.0327)	(0.0317)	(0.0470)	(0.0313)	(0.0306)	(0.0434)
Female	-0.137	0.261***	-0.0605	-0.105	0.247***	-0.0683
	(0.0880)	(0.0865)	(0.133)	(0.0878)	(0.0904)	(0.136)
Age	0.00151	0.0172***	-0.00637	0.00150	0.0174***	-0.00577
	(0.00277)	(0.00284)	(0.00486)	(0.00278)	(0.00278)	(0.00474)
Constant	-0.307	-3.661***	-3.388***	-0.283	-3.667***	-3.496***
	(0.249)	(0.367)	(0.529)	(0.257)	(0.368)	(0.501)
N	10,731	10,731	10,731	13,320	13,320	13,320

Standard errors in brackets (*** $p<0.01$, ** $p<0.05$, * $p<0.1$)

Determinants of vote

- Previous voting behavior and ideology have a strong and significant effect on vote choices.
- However, those who voted for PSOE in PSOE have significant chances of voting for other parties.
- The probabilities of voting for PP increase toward the right and the probabilities of voting for PSOE and IU increase toward the left.
- Education has a negative impact on the probabilities of voting PP, PSOE and IU. Well educated voters prefer to vote for other parties.
- Gender and age have a modest impact on vote choices. However, women and the elderly have greater chances of voting for PSOE.

Determinants of vote

- Perceptions of the economy have a barely significant effect on the probability of voting for PSOE. This party would get better results among who believed that the economic situation was good.
- Vote choices were mostly driven by ideological factors such as ideological proximity and party loyalty.

Remarks (Stata)

- Easy to stratify with Stata
- Easy to impute with Stata
- Advantages of working with results and matrices
- Advantages of creating own functions
- Use of Mata inside Stata

Remarks (Forecasting)

- As it was expected, postelectoral-polls are more accurate than pre-electoral surveys.
- Post-stratification has been extensively used in pre-electoral, but it does not always work better.
    - That is because of social desirability.
    - Post-stratification by previous vote is enough
- Imputation seems to work well. Even better than post-stratification.
- However, the use of both at the same time doesn't improve estimation, since they give similar results.