SPANISH EXIT POLLS: SAMPLING ERROR OR NONRESPONSE BIAS?

(ONLINE) SUPPLEMENTARY MATERIAL

Supplementary material of the paper:

Pavía, JM, Badal, E and García-Cárceles, B (2016) "Spanish exit polls: Sampling error or nonresponse bias?", **Revista Internacional de Sociología**, 74 (3), e043. DOI: 10.3989/ris.2016.74.3.043

GRAPHICAL APPENDIX (Figures 1A to 7A)

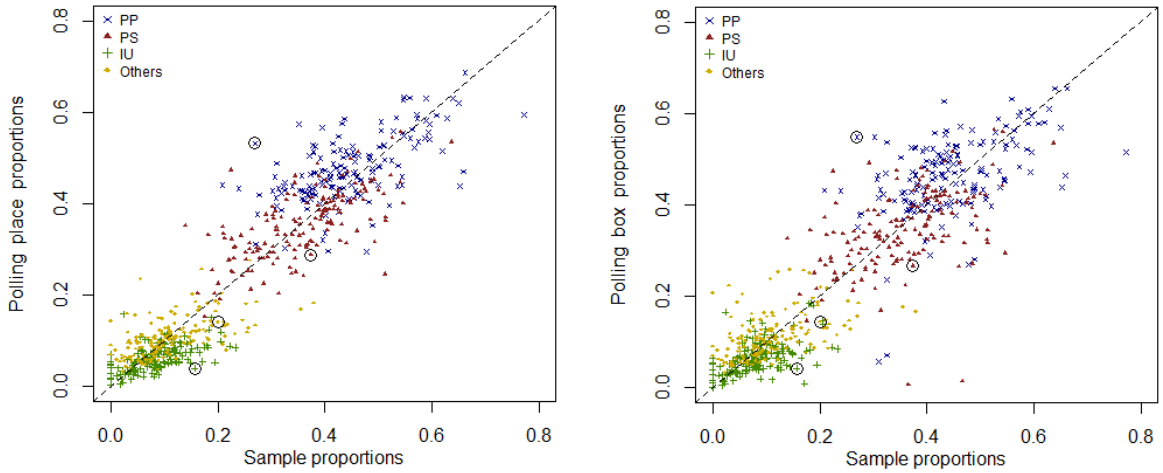

Figure 1A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling box (right scatterplot) levels for 2003 Corts Valencianes election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 640 (160 polling points per four political options). Highlighted, with a circle, the proportions corresponding to the sample where an unusually high amount of false reporting was detected.

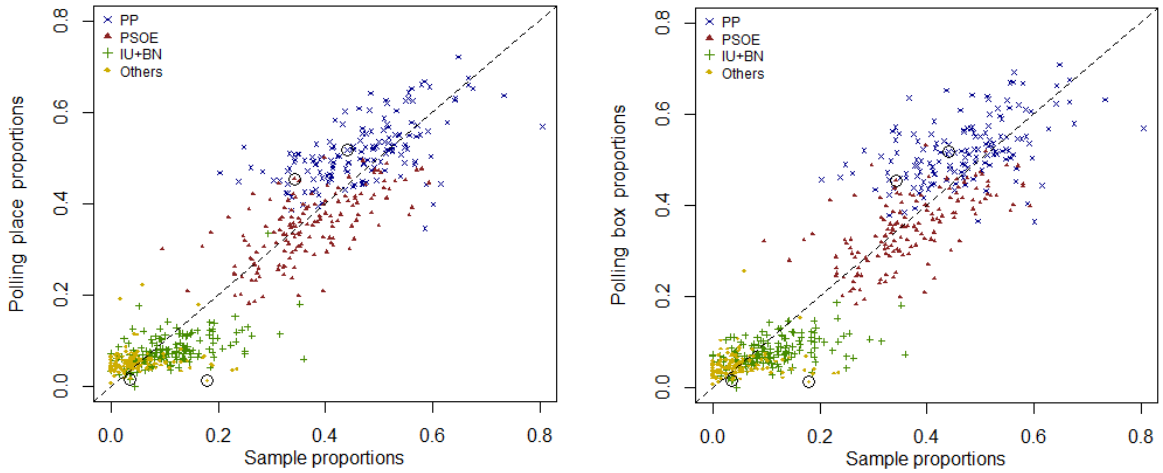

Figure 2A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling box (right scatterplot) levels for 2007 Corts Valencianes election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 604 (151 polling points per four political options). Highlighted, with a circle, the proportions corresponding to the sample where an unusually high amount of false reporting was detected.

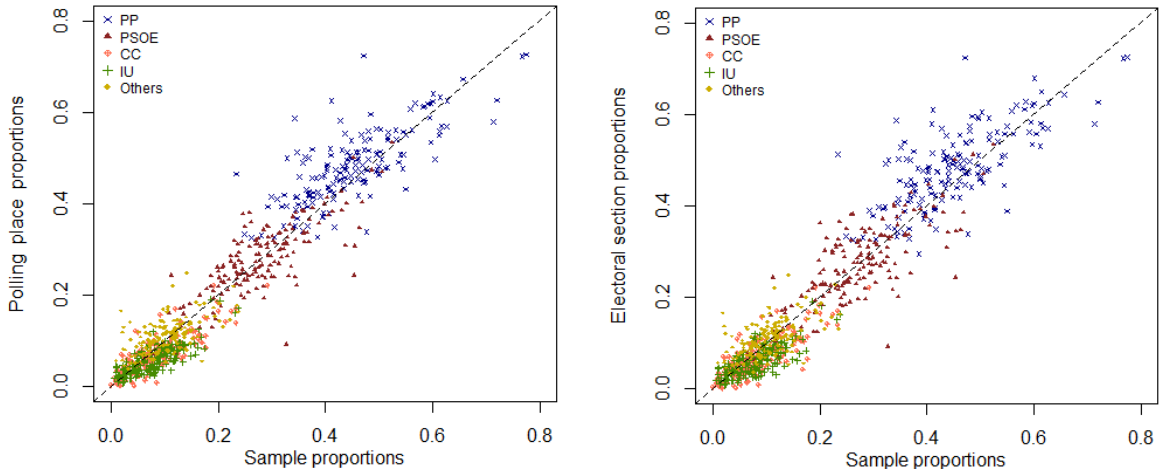

Figure 3A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling section (right scatterplot) levels for 2011 Corts Valencianes election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 790 (158 polling points per five political options).

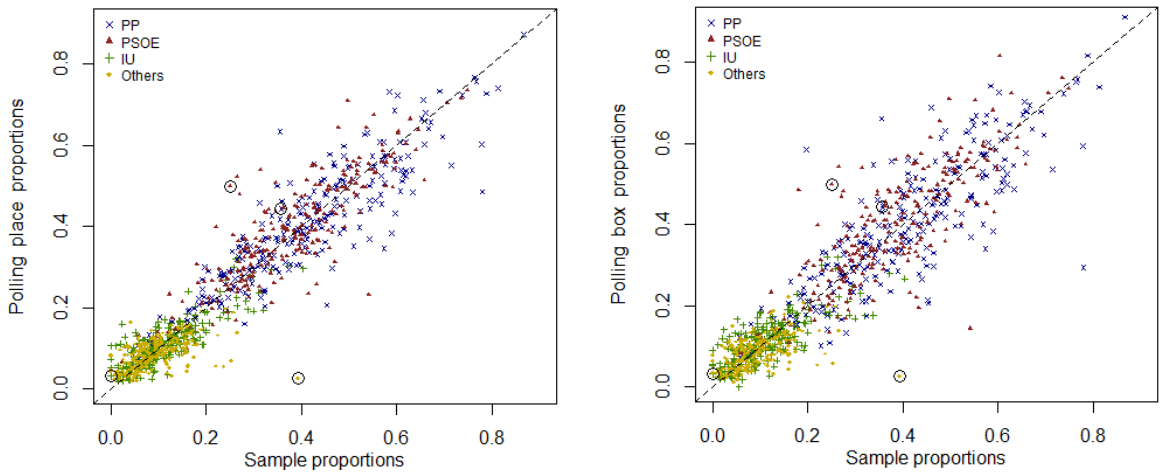

Figure 4A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling box (right scatterplot) levels for 2012 Parlamento de Andalucía election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 960 (240 polling points per four political options). Highlighted, with a circle, the proportions corresponding to the sample where an unusually high amount of false reporting was detected.

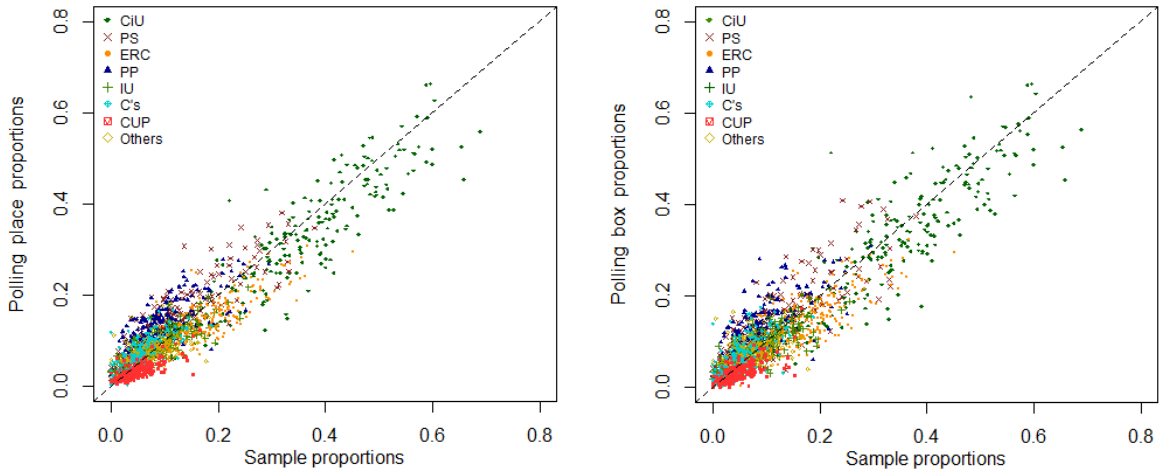

Figure 5A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling box (right scatterplot) levels for 2012 Parlament de Catalunya election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 1,600 (200 polling points per eight political options).

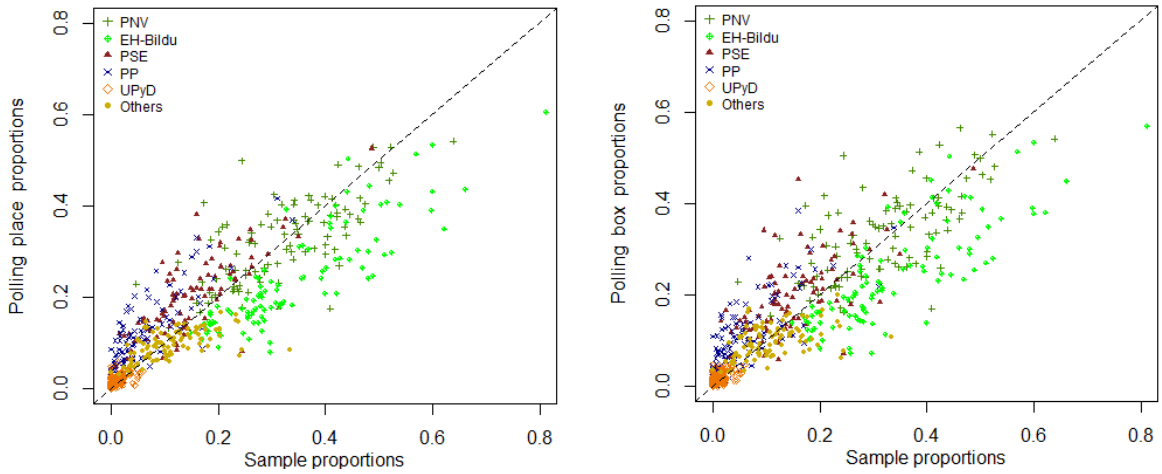

Figure 6A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling box (right scatterplot) levels for 2012 Eusko Legebiltzarra election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 600 (100 polling points per six political options).

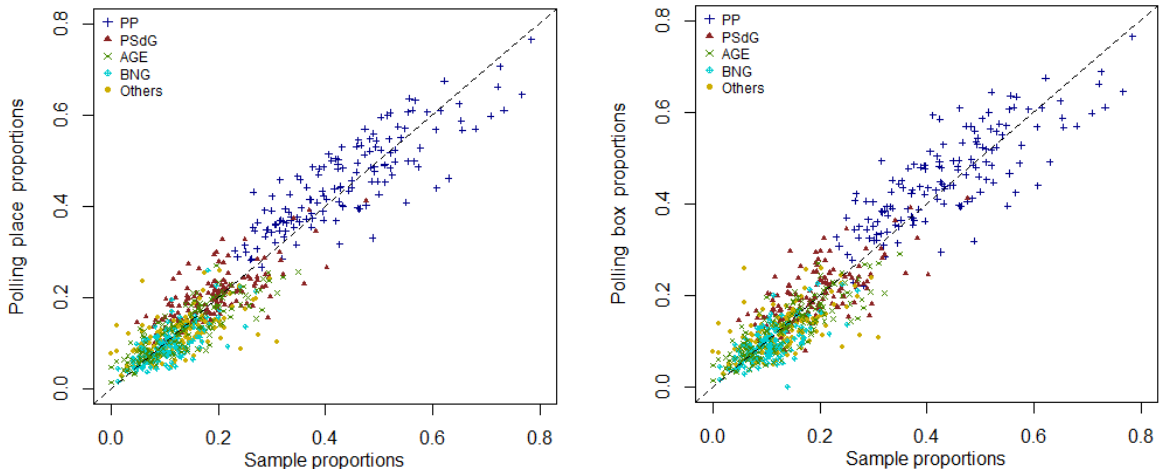

Figure 7A. Comparison between actual and exit poll proportions at polling location (left scatterplot) and polling box (right scatterplot) levels for 2012 Parlamento de Galicia election. The distance from the 45º line indicates how far apart collected data and outcomes are. The number of data points in each scatterplot is 770 (144 polling points per five political options).

SUMMARY STATISTICS (Tables I-A to VII-A)

Table I-A. 2003 Corts Valencianes election actual results and SigmaDos exit poll forecasts.

		Percentages				Seats		
Constituency		PP	PS	IU	Others	PP	PS	IU
Alicante	Results	48.0	36.9	5.9	9.2	16	12	2
	SigmaDos	46.9	37.3	7.6	8.2	15-16	12-13	2
N=6,495	EP raw data	46.7	37.1	8.0	8.1	16	12	2
Castellon	Results	47.7	36.2	4.4	11.7	13	9	1
	SigmaDos	45.0	36.1	4.6	14.3	12-13	9-10	1
N=3,327	EP raw data	41.6	37.1	6.6	14.6	12	10	1
Valencia	Results	46.0	34.9	7.0	12.1	19	14	3
	SigmaDos	44.3	35.8	8.6	11.3	18-19	14-15	3
N=11,382	EP raw data	44.1	34.8	10.0	11.1	18	14	4
Region	Results	46.8	35.7	6.3	11.1	48	35	6
	SigmaDos	45.3	36.4	7.8	10.5	45-48	35-38	6
N=21,204	EP raw data	44.5	35.9	8.9	10.8	46	36	7

Source: Own elaboration from official outcomes and SigmaDos data.

Table II-A. 2007 Corts Valencianes election actual results and SigmaDos exit poll forecasts.

		Percentages				Seats		
Constituency		PP	PS	IU+BN	Others	PP	PS	IU+BN
Alicante	Results	51.4	36.3	6.7	5.6	19	14	2
	SigmaDos	47.9	39.6	7.9	4.6	18	15	2
N=6,751	EP raw data	46.5	39.4	9.0	5.1	17	15	3
Castellon	Results	48.9	37.7	7.7	5.7	12	10	2
	SigmaDos	47.1	39.1	9.6	4.2	12-13	9-10	2
N=3,381	EP raw data	47.2	37.1	11.1	4.6	12	10	2
Valencia	Results	53.3	32.3	8.8	5.6	23	14	3
	SigmaDos	48.1	37.1	10.6	4.2	20-21	15-16	4
N=11,379	EP raw data	47.5	35.1	12.5	4.9	20	15	5
Region	Results	52.2	34.3	8.0	5.6	54	38	7
	SigmaDos	-	-	-	-	50-52	39-41	8
N=21,517	EP raw data	47.1	36.8	11.2	4.9	49	40	10

Source: Own elaboration from official outcomes and SigmaDos data.

Table III-A. 2011 Corts Valencianes election results and GFK-ODEC exit poll forecasts.

		Percentages					Seats			
Constituency		PP	PS	CC	IU	Others	PP	PS	CC	IU
Alicante	Results	48.8	29.4	4.5	5.3	12.1	20	12	1	2
	GFK-ODEC	49.2	29.5	5.3	6.8	9.2	19-20	11-12	1-2	2-3
N=6,700	EP raw data	45.4	29.0	6.3	8.1	11.1	18	12	2	3
Castellon	Results	47.1	30.4	6.6	5.1	10.9	13	9	1	1
	GFK-ODEC	47.3	30.5	6.5	6.0	9.7	13-14	8-9	1-2	1
N=3,911	EP raw data	43.5	30.2	7.8	6.9	11.6	12	8	2	2
Valencia	Results	48.8	25.9	8.7	6.3	10.2	22	12	4	2
	GFK-ODEC	51.9	23.6	8.7	7.2	8.6	23-24	10-11	3-4	3
N=12,596	EP raw data	44.7	25.7	10.6	9.0	10.0	20	12	4	4
Region	Results	48.6	27.6	7.1	5.8	10.9	55	33	6	5
	GFK-ODEC	50.5	26.4	7.3	6.9	8.9	55-58	29-32	5-8	6-7
N=23,207	EP raw data	44.7	27.4	8.9	8.4	10.6	50	32	8	9

Source: Own elaboration from official outcomes and GFK-ODEC data.

Table IV-A. 2012 Parlamento de Andalucía election actual results and Ipsos exit poll forecasts.

		Percentages				Seats		
Constituency		PP	PS	IU	Others	PP	PS	IU
Almeria	Results	51.2	35.4	7.1	6.3	7	4	1
	Ipsos	52.1	33.1	-	-	7-8	4-5	-
N=3,305	EP raw data	50.8	33.4	6.4	9.4	7	5	0
Cádiz	Results	40.5	35.6	12.7	11.3	7	6	2
	Ipsos	41.2	37.4	12.8	8.6	7	6-7	1-2
N=5,442	EP raw data	40.6	36.1	12.9	10.3	7	6	2
Córdoba	Results	39.7	38.9	13.3	8.0	5	5	2
	Ipsos	39.8	37.7	12.8	9.7	5-6	5	1-2
N=3,683	EP raw data	38.9	36.4	15.5	9.2	5	5	2
Granada	Results	43.5	39.5	10.0	7.1	6	6	1
	Ipsos	45.3	37.3	9.8	7.6	6-7	5-6	1
N=3,718	EP raw data	43.4	36.3	10.1	10.2	7	5	1
Huelva	Results	38.6	43.4	10.9	7.1	5	5	1
	Ipsos	38.4	46.5	-	-	4-5	5-6	-
N=3,107	EP raw data	36.9	44.9	10.4	7.8	4	6	1
Jaen	Results	41.1	44.5	8.8	5.6	5	5	1
	Ipsos	43.3	42.4	8.2	6.1	5	5	1
N=3,644	EP raw data	40.7	42.4	10.0	6.9	5	5	1
Malaga	Results	43.7	35.3	12.2	8.9	8	7	2
	Ipsos	46.2	35.8	10.0	8.0	8-9	6-7	2
N=4,993	EP raw data	46.5	33.8	11.3	8.4	9	6	2
Seville	Results	35.3	43.1	12.2	9.4	7	9	2
	Ipsos	36.7	42.4	11.1	9.8	7-8	8-9	2
N=6,729	EP raw data	34.2	42.7	12.9	10.3	7	9	2
Region	Results	40.7	39.6	11.3	8.4	50	47	12
	Ipsos	42.0	39.2	9.9	8.9	52-55	45-48	8-10
N=34,621	EP raw data	41.0	38.3	11.5	9.2	51	47	11

Source: Own elaboration from official outcomes, mass media publications and Ipsos data.

Table V-A. 2012 Parliament de Catalunya election results and Ipsos exit poll forecasts.

		Percentages							Seats						
Constituency		CiU	PSC	ERC	PP	IU	C's	CUP	CiU	PSC	ERC	PP	IU	C's	CUP
Barcelona	Results	28.1	15.4	12.7	13.3	11.1	8.4	3.4	26	14	12	12	10	8	3
	Ipsos	31.2	13.9	13.3	12.3	10.2	6.6	4.6	29-30	12-14	11-13	11-12	9-10	5-6	4
N=16,603	EP raw data	29.1	13.8	16.5	9.2	12.0	6.5	5.2	27	13	15	8	11	6	5
Girona	Results	43.0	10.1	17.8	9.6	5.9	3.6	4.2	9	2	3	2	1	0	0
	Ipsos	-	-	-	-	-	-	-	9-10	1	3-4	1-2	0-1	0	0-1
N=5,190	EP raw data	45.3	6.8	21.4	6.5	6.0	2.5	5.9	9	1	4	1	1	0	1
Lleida	Results	43.1	10.4	17.4	11.3	5.4	3.3	3.0	8	1	3	2	1	0	0
	Ipsos	-	-	-	-	-	-	-	8-9	1	3	1-2	0-1	0	0
N=4,314	EP raw data	43.0	8.9	19.6	9.8	5.7	2.7	4.1	8	1	4	1	1	0	0
Tarragona	Results	31.7	13.6	15.1	15.0	7.3	6.9	3.6	7	3	3	3	1	1	0
	Ipsos	-	-	-	-	-	-	-	7-8	2	2-3	2-3	1	0-1	0-1
N=5,135	EP raw data	36.1	10.3	18.7	11.5	6.6	5.0	4.9	7	2	4	2	1	1	1
Region	Results	30.7	14.4	13.7	13.0	9.9	7.5	3.2	50	20	21	19	13	9	3
	Ipsos	34.0	12.6	14.2	12.0	9.0	5.9	4.7	54-57	16-18	20-23	16-18	10-12	6-7	5-6
N=31,242	EP raw data	34.9	11.4	18.1	9.2	9.2	5.1	5.1	51	17	27	12	14	7	7

Source: Own elaboration from official outcomes, TV3 coverage and Ipsos data.

Table VI-A. 2012 Eusko Legebitzarra election results and Ipsos exit poll forecasts.

		Percentages						Seats				
Constituency		PNV	EH	PS	PP	UPyD	Others	PNV	EH	PS	PP	UPyD
Alava	Results	25.1	21.4	19.0	18.4	3.4	12.6	7	6	6	5	1
	Ipsos	25.5	23.9	18.0	18.8	2.6	11.2	7-8	6-7	5-6	5-6	0-1
N=2,999	EP raw data	25.9	31.8	15.4	11.9	1.5	12.0	8	10	4	3	0
Biscay	Results	37.5	20.9	18.5	11.5	1.8	9.8	11	6	5	3	0
	Ipsos	36.9	25.2	17.1	11.3	1.4	8.1	10-11	6-7	4-5	2-3	0
N=3,222	EP raw data	36.5	27.8	15.9	6.6	2.7	9.5	11	8	4	2	0
Guipuscoa	Results	31.4	31.2	18.8	8.3	1.4	9.0	9	9	5	2	0
	Ipsos	28.3	37.5	16.0	8.3	1.2	8.7	8-9	10-11	4-5	2	0
N=2,766	EP raw data	27.2	44.4	15.0	3.3	0.9	7.9	8	13	4	0	0
Region	Results	33.8	24.3	18.7	11.4	1.9	9.9	27	21	16	10	1
	Ipsos	32.6	28.9	16.9	11.4	1.4	8.8	24-27	23-26	13-15	9-11	0-1
N=8,987	EP raw data	30.1	34.2	15.5	7.3	1.7	11.1	27	31	12	5	0

Source: Own elaboration from official outcomes, EiTB coverage and Ipsos data.

Table VII-A. 2012 Parlamento de Galicia election results and Ipsos exit poll forecasts.

		Percentages					Seats			
Constituency		PP	PS	AGE	BNG	Others	PP	PS	AGE	BNG
La Corunna	Results	45.4	18.8	16.6	9.6	9.6	13	5	4	2
	Ipsos	-	-	-	-	-	12-14	5-6	3-4	2-3
N=7,648	EP raw data	42.7	18.1	17.5	10.9	10.9	12	5	4	3
Lugo	Results	51.5	22.7	10.3	8.7	6.8	9	4	1	1
	Ipsos	-	-	-	-	-	8-9	4	1-2	1
N=3,662	EP raw data	51.0	19.0	10.6	8.7	10.8	9	3	2	1
Orense	Results	49.2	23.7	7.8	8.5	10.9	8	4	1	1
	Ipsos	-	-	-	-	-	8-9	4	0-1	1
N=3,665	EP raw data	46.8	23.6	7.8	8.1	13.6	8	4	1	1
Pontevedra	Results	42.8	20.8	14.4	11.9	10.1	11	5	3	3
	Ipsos	-	-	-	-	-	10-11	5-6	3-4	2-3
N=5,792	EP raw data	38.6	19.0	17.7	12.4	12.3	10	5	4	3
Region	Results	45.8	20.6	13.9	10.1	9.6	41	18	9	7
	Ipsos	44.2	-	-	-	-	39-42	18-20	8-10	7-8
N=20,467	EP raw data	43.7	19.4	14.7	10.5	11.7	39	17	11	8

Source: Own elaboration from official outcomes, TVG coverage and Ipsos data.

multihyper.r

(R code to implement the new statistical tests proposed in the paper)

```
dmultihyper <- function(k,K) {
# Function to calculate the probability of a hypergeometric distribution.
# The hypergeometric distribution applies sampling without replacement
# from a finite population whose elements can be classified into s mutually
# exclusive categories. It can be represented trough a model of an urn
# with marbles of s colors, where there are K_i marbles of color i in the urn
# and a number of marbles are extracted at random without replacement from the urn,
# then the number of marbles of each color in the sample (k_1,k_2,...,k_s) has the multivariate
# hypergeometric distribution.

# INPUT:
#     k: vector of quantiles of order s representing the number of balls drawn of each color
#        from an urn which contains balls of s colors.
#     K: vector of order s with the number of balls of each color in the urn.

# OUTPUT:
#     Probability of observing the vector k in a multi-hypergeometric distribution with
#     parameters K and n=sum(k)

  if (length(K)!=length(k)){
     stop('The order of the vector of balls in the urn and in the sample must be equal')
  }
  if (length(K)<2) stop('At least balls of two different colors are required')
   if (min(K-k)<0){
     output <- 0
  } else {
      output<-exp(sum(lchoose(K,k))-lchoose(sum(K),sum(k)))
  }
  return(output)
}

# ------------------------------------------

LRT.multihyper <- function(K,k0){
# Test based on the log-Likelihood-ratio test to test if a sample k0
# of size n=sum(k0) comes from a Multihypergeometric distribution
# with parameters K and n: MHg(K,n).

# INPUT:
#     K: vector of order s with the number of balls of each color in the urn.
#     k0: vector of order s with the number of balls of each color in the sample.

# OUTPUT:
#     LRT p-value for the test of the null hypothesis that k0 comes from a MHg(K,n)

  if (length(K)!=length(k0)){
     stop('The order of the vector of balls in the urn and in the sample must be equal')
  }
  if (length(K)<2) stop('At least balls of two different colors are required')
  if (min(K-k0)<0){
     p.value <- 0
  } else {
      p.h0<-exp(sum(lchoose(K,k0))-lchoose(sum(K),sum(k0)))
      mle <- (k0/sum(k0))*sum(K)
      p.mle<-exp(sum(lchoose(mle,k0))-lchoose(sum(mle),sum(k0)))
      LRT <- 2*log(p.mle/p.h0)
```

```
      p.value <- 1-pchisq(LRT, length(k0)-1)
  }
  return(p.value)
}

#---------------------------------------
rmultihyper <- function(nn,K,n) {
# Function to generate nn samples of a multi-hypergeometric distribution with
# parameters K and n: MHg(K,n).

# INPUT:
#     nn: the number of samples to be generated.
#     K: vector of order s with the number of balls of each color in the urn.
#     n: number of balls to be drawn without replacement.

# OUTPUT:
#    A matrix of order nn x length(K), where each row contains a sample
#    The element [i,j] of the output accounts for the number of balls of
#    color.j drawn in sample i

  if (length(K)<2) stop('At least balls of two different colors are required')
  if (n<1) stop('At least an extraction is required')
  output <- matrix(0,nn,length(K))
  for (i in 1:nn){
     K0<-K
     prob0 <- K0/sum(K0)
     for (j in 1:n){
        x<-rmultinom(1,1,prob0)
        temp<-which(x==1)
        output[i,temp] <- output[i,temp]+1
        K0[temp] <- K0[temp]-1
        prob0 <- K0/sum(K0)
     }
  }
  rownames(output)<-paste("sample",1:nn,sep="")
  colnames(output)<-paste("color",1:length(K),sep="")
  return(output)
}

#-----------------------------------------------
Dtest.multihyper <- function(K,k0,sim=1000){
# Test based on the Monte Carlo aproximation of the chi-squared distance
# to test if a sample k0 of size n=sum(k0) comes from a multihypergeometric distribution
# with parameters K and n: MHg(K,n).

# INPUT:
#     K: vector of order s with the number of balls of each color in the urn.
#     k0: vector of order s with the number of balls of each color in the sample.
#     sim: number of simulations to calculate the p-value of the test, default 1000.

# OUTPUT:
#     p-value for the test of the null hypothesis that k0 comes from a MHg(K,n)

  if (length(K)!=length(k0)){
     stop('The order of the vector of balls in the urn and in the sample must be equal')
  }
  if (length(K)<2) stop('At least balls of two different colors are required')
  expected <- K/sum(K)*sum(k0)
  chi <- function(ob,es) sum((ob-es)^2/es)
  statistic <- chi(k0,expected)
```

```
  if (min(K-k0)<0){
      p.value <- 0
  } else {
       simul <- rmultihyper(sim,K,sum(k0))
       cal <- apply(simul,1,chi,es=expected)
       p.value <- sum(cal>=statistic)/length(cal)
  }
  return(p.value)
}
```