

# Estimació del suport latent als partits polítics: Eleccions al Congrés dels Diputats 2011

XAVIER FERNÁNDEZ-I-MARÍN

RESUM. El suport als partits polítics es pot considerar com una variable latent l'estat del qual només es coneix amb certesa quan hi ha eleccions, i s'estima amb un nivell raonable d'incertesa en cada enquesta d'intenció de vot.

L'objectiu d'aquest treball és estimar el suport latent a cadascun dels partits que es presenten a les eleccions al Congrés dels Diputats de novembre de 2011. Els resultats d'aquesta estimació, així com el 95% de la incertesa associada, s'obtenen a partir de les diferents mesures del suport a cada partit que es desprenen de les enquestes electorals. Cada enquesta la realitza una casa (empresa, diari, institució pública) i es pot estimar també el biaix que cadascuna d'aquestes cases tenen respecte a les prediccions. El resultat final és una estimació del suport per a cada partit, una estimació del biaix de cada casa d'enquesta per a cada partit, i una estimació de la variació diària del vot de cada partit.

Aquest document explica els detalls del mètode d'estimació, així com les interpretacions de cadascun dels paràmetres que s'obtenen.

## 1. INTRODUCCIÓ

La predicció dels resultats electorals és l'objectiu de les enquestes electorals. Però les enquestes només són una fotografia fixa d'un moment concret de la campanya, basades en una estimació feta a través d'extreure la informació d'una mostra més o menys representativa de la població. Si fos possible combinar els resultats de les diferents enquestes en una única enquesta seria possible arribar a estimar amb més precisió el suport latent a cada partit. De tota manera, les enquestes es realitzen en dies diferents i, per això, és difícil combinar-les per tal d'aconseguir un únic resultat.

De tota manera, és possible combinar les eleccions (enquestar les enquestes — *polling the polls*) per arribar a una predicció del suport electoral si es considera que són una mesura més o menys imperfecta del suport latent que cada dia té un partit polític concret. Aquest document explica els detalls d'un mètode de predicció del suport electoral basat en “enquestar les enquestes”, mirant de combinar-les per aconseguir informació sobre el suport a cada partit.

En moments electorals, la ciència política es pot beneficiar d'aquest mètode d'estimació, ja que també s'obtenen, com a subproductes de l'estimació, altres paràmetres que poden ser rellevants per a respondre a altres preguntes. Per exemple, quin és l'efecte d'una campanya electoral en la variació dels suports a cada partit polític. Val la pena realitzar una campanya electoral? Varien realment els suports que cada partit obté a les eleccions en funció del que passa els dies de

---

*Date:* 14 de novembre de 2011.

Treball en evolució. Si us plau, no ho citeu sense permís exprés de l'autor. Els detalls de contacte apareixen al final.

campanya electoral o bé el suport a un partit és relativament insensible als esdeveniments d'una campanya? Més enllà de ser un moment on els partits expliquen les seves propostes, quina és la capacitat que tenen les campanyes per variar el vot dels ciutadans? Un segon interès paral·lel rau en mesurar la volatilitat del suport a cadascun dels partits. Hi ha partits que tenen suports més estables que altres? Quins són? Finalment, és també objecte d'interès el biaix que cadascuna de les cases d'enquestes tenen respecte al resultat final de la campanya, i en mesurar com són d'acurades les enquestes.

De tota manera, aquestes darreres mesures només poden ser completes un cop el model pugui incorporar també les dades de l'observació real del que es pretén predir, que no deixa de ser el suport electoral que es coneixerà el dia de les eleccions.

El document recull els detalls concrets del mètode de predicció del suport electoral, així com les interpretacions que es poden fer de les diferents mesures que s'extreuen en el model

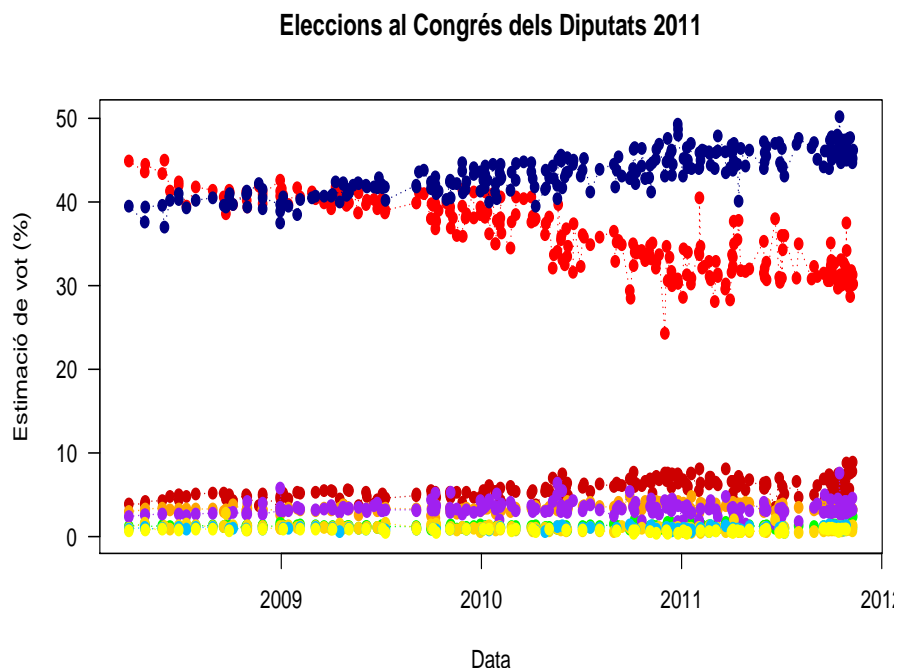
## 2. ESTIMAR EL SUPORT LATENT

El suport latent a un partit polític o a una coalició de govern en base als resultats de les enquestes ha estat analitzat a (Jackman 2005) i desenvolupat posteriorment a (Jackman 2009). El model presentat aquí és una extensió d'aquest treball per a un sistema multipartidista, on no només s'estima el suport a una única opció de vot, sinó a múltiples. La segona característica que estén aquest treball és que Jackman (2005) fa servir el model no tant per estimar el resultat final, sinó per fixar-se en els efectes que té la campanya electoral i els biaixos de cadascuna de les cases d'enquestes. En aquest sentit, en aquest treball l'objectiu final és relativament diferent —estimar el suport latent—, tot i que el procediment emprat sigui fonamentalment el mateix. Per a una referència més general sobre la predicció de suport electoral veure Lock & Gelman (2010).

La idea fonamental és que el suport a un partit polític és un paràmetre que tan sols es pot conèixer amb certesa quan hi ha una cita electoral. En el cas català, això és cada quatre anys, si no hi ha eleccions anticipades. Però entremig d'aquests quatre anys hi ha una manera més o menys precisa de conèixer el suport, i és mesurar-lo a través d'enquestes. Aquestes enquestes són una manera de tenir una estimació del suport latent. És a dir, mesuren el suport, tot i que amb un error associat respecte del valor latent en el moment en què es porta a terme l'enquesta. Per tant, cal distingir entre el suport *real*—el que es pot mesurar amb les eleccions—, el suport *mesurat*—el reportat per una enquesta concreta— i el suport *latent*—el suport diari que una enquesta només pot estimar amb un cert marge d'error—. Distingirem suport mesurat de suport latent.

La Figura 1 presenta l'evolució del suport mesurat a cada partit, d'acord amb les estimacions de cada enquesta.

Si només es té en compte el suport mesurat, l'efecte és que el suport als partits polítics és força variable, especialment quan les enquestes són molt properes en el temps. El suport latent a cada partit no té unes desviacions tant grans d'un dia a un altre, encara que és plausible que el suport latent variï més per als partits amb més percentatge de suport. A més a més, cada “casa” o empresa o organisme que realitza una enquesta té els seus propis mètodes de mostreig i imputació de la no resposta que fan que el suport mesurat sigui susceptible a biaixos. Aquests biaixos



*Figura 1.* Suport als partits polítics. Estimació directa del vot a cada una de les enquestes.

però, són propis de cada casa  $i$ , per tant, sistemàtics i constants en el temps, sempre i quan els mètodes utilitzats siguin els mateixos.

Per tot això, un model d'estimació del suport latent ha de tenir en compte el següent:

- Cada “casa” d'enquestes té el seu propi sistema i mètode, a més de fer les estimacions basant-se en diferents paràmetres únics a cada casa. Es tracta de biaixos sistemàtics de cada empresa d'enquestes cap a un partit o altre, que es poden considerar fixos en el període estudiat.
- El suport latent varia lentament, amb petites variacions diàries.
- Cada partit té una volatilitat que li és pròpia. Això és, cada partit pot variar el seu suport diari d'una manera més o menys acusada que els altres partits.

Cadascuna de les estimacions de vot (suport) que fa cada enquesta  $y$  assumeixo que provenen d'una distribució normal amb mitjana  $\mu$  i desviació típica  $\sigma$ . Cada observació correspon a una enquesta ( $e$ ) i a un partit ( $p$ ). A més a més, cada enquesta està realitzada per una casa ( $c$ ) que s'acostuma a repetir. Cada observació, doncs, es pot entendre com una realització de:

$$(1) \quad y_{e,p} \sim \mathcal{N}(\mu_{e,p}, \sigma_e)$$

L'Equació 1 tan sols trasllada la incertesa associada a l'estimació que qualsevol enquesta dóna, i cada enquesta té associat un nivell d'incertesa determinat especificat per  $\sigma_e$ .

El component sistemàtic del suport mesurat és simplement la suma del suport latent al partit concret en aquell moment del temps ( $t$ ) més el biaix propi que de cada casa per a cada partit, definit per  $\delta_{c,p}$ :

$$(2) \quad mu_{e,p} = \xi_{t,p} + \delta_{c,p}$$

El suport latent  $\xi_{t,p}$  és una sèrie temporal on cada observació és igual a l'observació anterior més una desviació concreta associada a cada partit (*random walk*), que es pot entendre com la volatilitat del suport a un partit determinat ( $\omega_p$ ):

$$(3) \quad \xi_{t,p} \sim \mathcal{N}(\xi_{t-1,p}, \omega_p)$$

El biaix de cada casa per a cada partit s'expressa com una mostra d'una distribució multinormal amb mitjanes 0 i una matriu de variança-covariança ( $W$ ) amb covariances fixades a zero i variança 4. Una variança de 4 correspon a una desviació de cada casa de fins de 4 punts per sobre i 4 punts per sota. Amb un marge tan ample el que es pretén és que les dades siguin les que informin totalment dels efectes de cada casa:

$$(4) \quad \delta_{c,p} \sim \mathcal{MN}(0, W)$$

**2.1. Restriccions.** Una manera de restringir el model és fixar que la suma del suport latent a cadascun dels partits és igual a 100, bo i tenint en compte el suport a la opció "altres":

$$(5) \quad \sum_{p=1}^P \xi_{t,p} = 100$$

**2.2. Priors.** Els *a priori* sobre la incertesa de les estimacions són poc informatius.  $\omega_p \sim \mathcal{U}(0, 1)$  representa la incertesa *a priori* de la variació de vot diària de cada partit, uniforme entre 0 i 1. Es tracta d'un escenari on la variació diària del suport latent és d'un màxim de dos punts percentuals. Una manera d'estendre el model seria poder relaxar aquest paràmetre a què de tant en tant pogués ser més elevat, com a resposta a estímuls sobtats en una campanya. Tot i la limitació, una variació del suport de 2 punts diaris és raonable.

Per la seva banda,  $\sigma_e$ , l'error de cada enquesta, depèn de la mostra de cada enquesta. En un model simplificat s'ha considerat que pot ser entre 0 i 5 ( $\sigma \sim \mathcal{U}(0, 5)$ ).

**2.3. Eleccions.** Les eleccions es tracten com una realització de  $\xi$  sense variació. Es fixen i ajuden a l'estimació de la sèrie del suport latent. Són valors observats de  $\xi$ . De fet, són els únics moments on  $\xi$  s'observa.

### 3. DADES

Les dades provenen de les enquestes d'intenció de vot al Congrés dels Diputats des de les eleccions de març de 2008, i es van actualitzant a mesura que noves medicions del suport van apareixent en forma d'enquesta. La data de cada enquesta he considerat que era la mitjana dels dies del treball de camp.

La Taula 1 presenta el nombre d'enquestes per cada casa i la mitjana de la mostra que cada casa pren com a referència.

	Enquestes	Mediana de la mostra
Celeste-Tel	44	1100
La Razón	36	1000
El País	28	1000
El Mundo	23	1000
Público	21	2044
La Sexta	19	
COPE	18	501
Antena 3, Onda Cero	17	1000
CIS	15	2478
El Periódico	8	1000
SER	8	1000
ABC	7	1076
La Vanguardia	7	1272
Intercampo	3	1724
Fundación Ortega y Gasset	1	800
Fundación Ortega-Marañón	1	1505
GAD3	1	400
Ipsos	1	
La Gaceta	1	1400

*Taula 1.* Nombre d'enquestes realitzades per cada casa i mediana de les seves mostres.

3.1. **Partits.** Els partits considerats són els que actualment tenen representació parlamentària (PSOE, PP, IU, CiU, PNB, UPyD, ERC, BNG i CC).

3.2. **Temps.** La sèrie temporal comprèn des de les eleccions de març de 2008 i les eleccions de 20 de novembre de 2011. Això fa un total de 1352 dies, d'un dels quals se'n coneix el suport i s'estima per als 1351 restants.

#### 4. ESTIMACIÓ

D'acord amb les equacions de la secció 2 els paràmetres a estimar són els següents:

	num	n
$\xi_{t,p}$	T*P	12168
$\delta_{c,p}$	C*P	171
$\omega_p$	P	9
$\sigma_e$	E	259

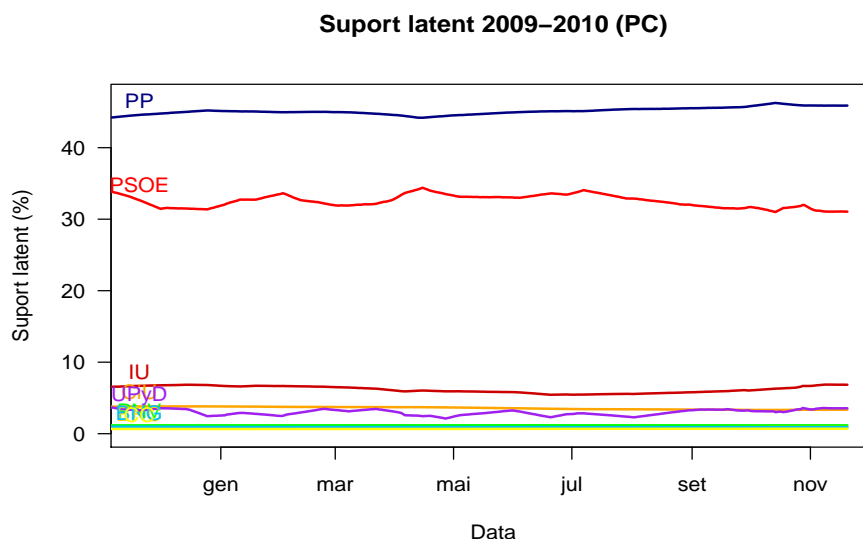
*Taula 2.* Nombre de paràmetres a estimar.

L'estimació es fa fent servir inferència bayesiana. Els resultats s'obtenen simulant les distribucions *a posteriori* de cadascun dels paràmetres d'interès. La simulació es fa per mètodes Monte Carlo. Més concretament, es fa servir el mostrejador Gibbs (*Gibbs sampler*) implementat pel programa JAGS (Plummer 2010). Les cadenes resultants s'analitzen amb R (R Development Core Team 2010) amb les llibreries

coda (Plummer, Best, Cowles & Vines 2010) i *boa* (Smith 2007). Es simulen 200,000 valors (iteracions) de les quals només se'n desa una de cada 10, i amb un període inicial de 100,000 valors que serveixen per a què les cadenes arribin a un estat de convergència. No hi ha evidència de manca de convergència d'acord amb el test Geweke (Geweke 1992).

## 5. RESULTATS

**5.1. Estimació del suport latent als partits polítics.** La Figura 2 presenta l'estimació de suport latent mitjà per als partits polítics en el període entre les eleccions de 2008 i les de 2011. En contrast amb la Figura 1, aquesta figura ja contempla les desviacions de cada casa d'enquestes i, doncs, presenta el suport latent afinant l'estimació i suavitzant les línies del suport estimat.



*Figura 2.* Evolució temporal de l'estimació del suport latent mitjà per a cada partit polític entre les eleccions de 2008 i 2011 ( $\xi$ ).

La Figura 3 presenta per a cada partit polític, el suport latent amb l'interval de credibilitat del 95%. Els punts representen enquestes on es reporta el suport mesurat segons cadascuna de les cases d'enquestes. Altre cop, el suport latent es pot entendre com una mitjana ponderada de cada enquesta havent descomptat el biaix de cada casa d'enquestes i tenint en compte una línia temporal que ve des de les darreres eleccions el març de 2008.

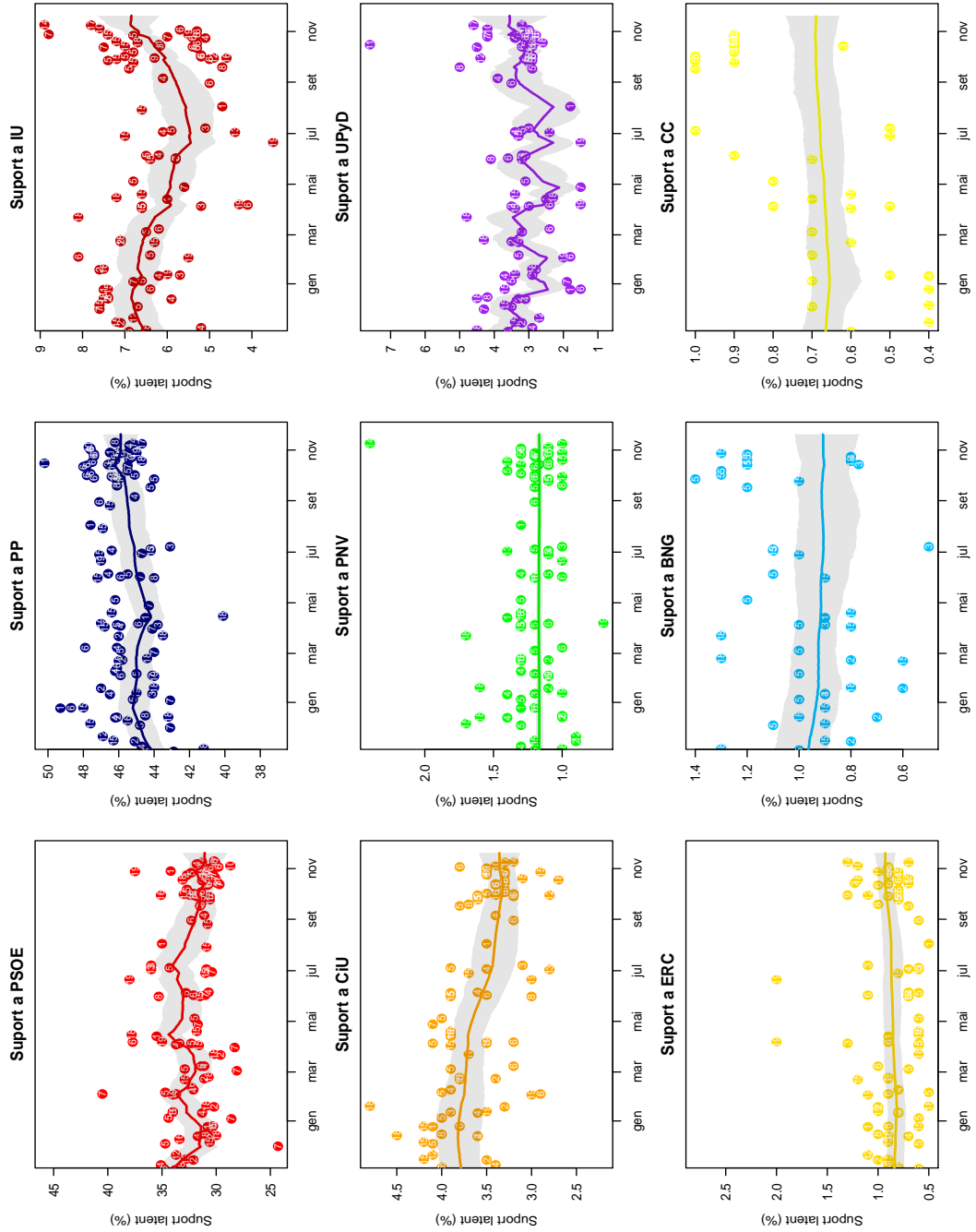


Figura 3. Evolució temporal de l'estimació del suport latent i el 95% de credibilitat per a cada partit polític els darrers mesos abans de les eleccions de 2010 ( $\xi$ ).

5.2. **Biaix de cases.** La Figura 4 presenta les estimacions del biaix que cada casa d'enquestes té respecte a cadascun dels partits. El punt representa el biaix mitjà, mentre que les barres corresponen a la incertesa de vot al 95%. L'escala de la figura és en punts percentuals.

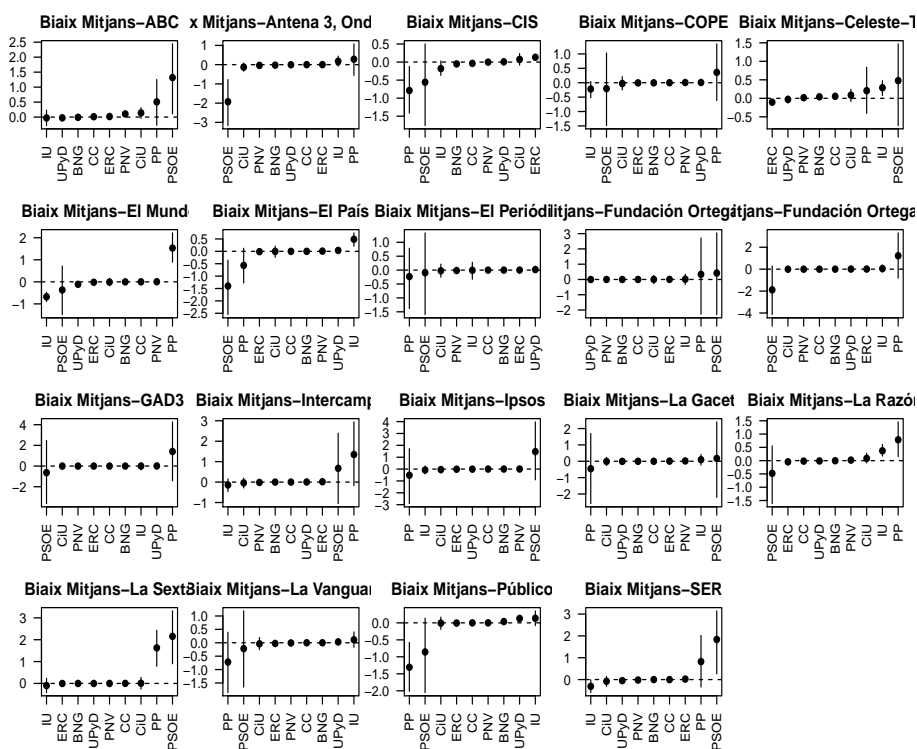


Figura 4. Biaix de cadascuna de les cases d'enquestes respecte a cada partit polític ( $\delta$ ).

La Figura 5 presenta les mateixes dades que la Figura 4, però en aquest cas la comparació es fa entre cases d'enquestes per a cada partit polític.

5.3. **Diferència PSOE-PP.** El model també permet capturar la diferència entre cadascun dels partits polítics. La Figura 6 presenta l'evolució d'ambdós partits i l'estimació de la diferència entre ambdós partits, amb un 95% de credibilitat.

5.4. **Volatilitat del vot.** Els paràmetres  $\omega_p$  fan referència a la volatilitat del suport latent a cada partit. És a dir, de la variació en el suport que rep cada partit a través del temps. Una volatilitat major significa que el suport a un partit concret és més variable de dia a dia. La Figura 7 presenta els valors de la volatilitat estimats per a cada partit.

## 6. CONCLUSIONS

El document ha documentat el model emprat en l'estimació del suport latent de cadascun dels partits polítics que es presenten a les eleccions al Congrés dels

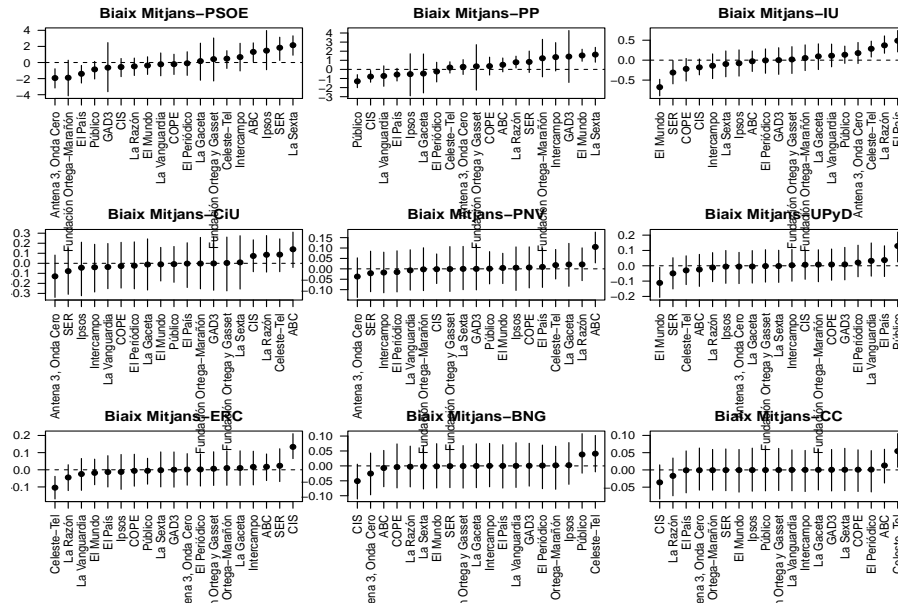


Figura 5. Biaix que cada partit té respecte a cadascuna de les cases d'enquestes ( $\delta$ ).

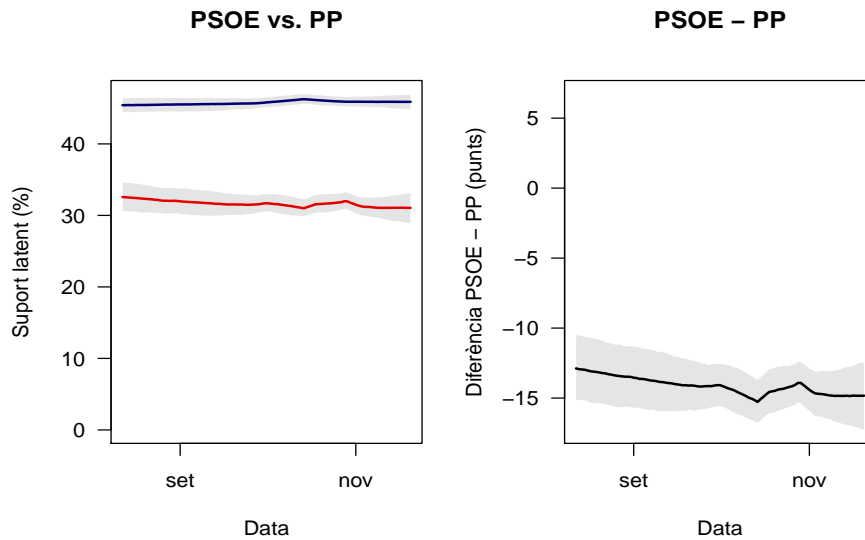
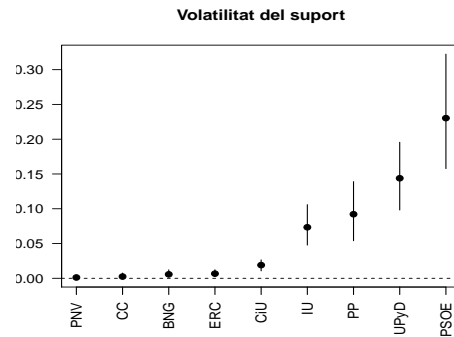


Figura 6. Diferència de suport latent entre PSOE i el PP els darrers dies abans de les eleccions ( $\xi_{psoe} - \xi_{pp}$ ).

Diputats el novembre de 2011. Els resultats es basen en combinar les dades de diferents enquestes per tal d'obtenir una predicció de vot. A més a més, en el procés



*Figura 7.* Volatilitat estimada del suport latent a cadascun dels partits polítics ( $\omega$ ).

d'estimació es produeixen també altres paràmetres que poden resultar d'interès per a estudiar la dinàmica electoral en un context de partits múltiples. Aquests paràmetres d'interès inclouen el biaix de cada casa que realitza enquestes, l'efecte que les campanyes electorals tenen en la variació del suport electoral als partits, i la volatilitat del vot a cada partit polític.

## REFERÈNCIES

- Geweke, J. F. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in J. Berger, J. Bernardo, A. Dawid & A. Smith (eds), *Bayesian Statistics 4*, Clarendon Press, Oxford, UK.
- Jackman, S. (2005). Polling the polls over an election campaign, *Australian Journal of Political Science* **40**(4): 499–517.
- Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*, John Wiley & Sons, New Jersey.
- Lock, K. & Gelman, A. (2010). Bayesian combination of state polls and election forecasts, *Political Analysis* **18**(3): 337–348.
- Plummer, M. (2010). *JAGS: Just Another Gibbs Sampler*. Version 2.1.0. <http://www-fis.iarc.fr/~martyn/software/jags/>.
- Plummer, M., Best, N., Cowles, K. & Vines, K. (2010). *coda: Output analysis and diagnostics for MCMC*. R package version 0.13-5.  
**URL:** <http://CRAN.R-project.org/package=coda>
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org>
- Smith, B. J. (2007). boa: An r package for mcmc output convergence assessment and posterior inference, *Journal of Statistical Software* **21**(11): 1–37.

INSTITUT BARCELONA D'ESTUDIS INTERNACIONALS (IBEI)

*E-mail address:* [xfernandez@ibei.org](mailto:xfernandez@ibei.org)

*URL:* <http://xavier-fim.net>