National Electoral Thresholds and Disproportionality

Online Appendix (not intended for publication)

Tasos Kalandrakis* Miguel R. Rueda†

March 12, 2020

*Department of Political Science and Department of Economics, University of Rochester. E-mail: kalandrakis@rochester.edu.

†Department of Political Science, Emory University.

A Identification: Proof of Theorem 1

We first argue that β_0 is identified. By (A_2), F_v has strictly positive mass on

$$\{(v,s) \mid \text{There exist } i_1, i_2, i_3 \text{ with } v_{i_1} > v_{i_2} > v_{i_3} > 0 \text{ and } s_{i_3} > 0\}.$$

For all (v,s) in that set, the conditional distribution of seats for the two top vote-getting parties (without loss of generality parties 1 and 2) satisfies

$$(s_1, s_2) \sim \mathit{Binomial}\left(\frac{v_1^{\beta_0}}{v_1^{\beta_0} + v_2^{\beta_0}}, \frac{v_2^{\beta_0}}{v_1^{\beta_0} + v_2^{\beta_0}}; s_1 + s_2\right).$$

This is because within that subset of data, the threshold is known to be below $v_{i_3} < v_i, i = 1, 2$, so that the conditional distribution of seats for parties 1 and 2 given the remaining allocation follows from the preservation of multinomial sampling under conditioning. Now β_0 is identified from these data as the success probability of the binomial is identified and is a one-to-one function of β_0. So it remains to argue that the threshold parameters θ_0, σ_0 are identified. By (A_2), we observe (v,s) with $v \in \mathcal{V}_3$. For such (v,s), let $\underline{i} = \arg\min_i \{v_i \mid v_i > 0\}$ and $\underline{\underline{i}} = \arg\min_i \{v_i \mid v_i > v_{\underline{i}}\}$. We have (for general parameters β, θ, σ):

$$\begin{aligned}
\mathbb{P}\left(s_{\underline{i}} = 0 \mid v\right) &= \left(1 - \Phi\left(\frac{v_{\underline{i}} - \theta}{\sigma}\right)\right) + \Phi\left(\frac{v_{\underline{i}} - \theta}{\sigma}\right)\left(1 - q_{\underline{i}}(v, \beta, 0)\right)^{\sum_{i=1}^N s_i} \\
\mathbb{P}\left(s_{\underline{\underline{i}}} = 0 \mid v\right) &= \left(1 - \Phi\left(\frac{v_{\underline{\underline{i}}} - \theta}{\sigma}\right)\right) + \left(\Phi\left(\frac{v_{\underline{\underline{i}}} - \theta}{\sigma}\right) - \Phi\left(\frac{v_{\underline{i}} - \theta}{\sigma}\right)\right)\left(1 - q_{\underline{\underline{i}}}(v, \beta, v_{\underline{i}})\right)^{\sum_{i=1}^N s_i} \\
&\quad + \Phi\left(\frac{v_{\underline{i}} - \theta}{\sigma}\right)\left(1 - q_{\underline{\underline{i}}}(v, \beta, 0)\right)^{\sum_{i=1}^N s_i}.
\end{aligned}$$

Solving for $\Phi\left(\frac{v_{\underline{i}}-\theta}{\sigma}\right), \Phi\left(\frac{v_{\underline{\underline{i}}}-\theta}{\sigma}\right)$ we obtain

$$
\begin{aligned}
\Phi\left(\frac{v_{\underline{i}}-\theta}{\sigma}\right) =& \mathbb{S}_{\underline{i}}(\beta \mid v) \\
\Phi\left(\frac{v_{\underline{\underline{i}}}-\theta}{\sigma}\right) =& \mathbb{S}_{\underline{\underline{i}}}(\beta \mid v),
\end{aligned}
$$

where

$$
\mathbb{S}_{\underline{i}}(\beta \mid v) = \frac{1-\mathbb{P}\left(s_{\underline{i}}=0 \mid v\right)}{1-\left(1-q_{\underline{i}}(v,\beta,0)\right)^{\sum_{i=1}^{N} s_i}}
$$

$$
\mathbb{S}_{\underline{\underline{i}}}(\beta \mid v) = \frac{1-\mathbb{P}\left(s_{\underline{\underline{i}}}=0 \mid v\right) + \frac{1-\mathbb{P}\left(s_{\underline{i}}=0 \mid v\right)}{1-\left(1-q_{\underline{i}}(v,\beta,0)\right)^{\sum_{i=1}^{N} s_i}}\left(\left(1-q_{\underline{\underline{i}}}(v,\beta,v_{\underline{i}})\right)^{\sum_{i=1}^{N} s_i}-\left(1-q_{\underline{\underline{i}}}(v,\beta,0)\right)^{\sum_{i=1}^{N} s_i}\right)}{1-\left(1-q_{\underline{\underline{i}}}(v,\beta,v_{\underline{i}})\right)^{\sum_{i=1}^{N} s_i}}.
$$

With β_0 identified, both $\mathbb{S}_{\underline{i}}(\beta_0 \mid v)$ and $\mathbb{S}_{\underline{\underline{i}}}(\beta_0 \mid v)$ are identified from the data. It follows that θ_0, σ_0 are identified as two quantiles suffice to identify the parameters of the normal distribution, specifically

$$
\begin{aligned}
\theta_0 =& \frac{v_{\underline{i}}\Phi^{-1}\left(\mathbb{S}_{\underline{\underline{i}}}(\beta_0 \mid v)\right) - v_{\underline{\underline{i}}}\Phi^{-1}\left(\mathbb{S}_{\underline{i}}(\beta_0 \mid v)\right)}{\Phi^{-1}\left(\mathbb{S}_{\underline{\underline{i}}}(\beta_0 \mid v)\right) - \Phi^{-1}\left(\mathbb{S}_{\underline{i}}(\beta_0 \mid v)\right)}, \\
\sigma_0 =& \frac{v_{\underline{\underline{i}}} - v_{\underline{i}}}{\Phi^{-1}\left(\mathbb{S}_{\underline{\underline{i}}}(\beta_0 \mid v)\right) - \Phi^{-1}\left(\mathbb{S}_{\underline{i}}(\beta_0 \mid v)\right)}.
\end{aligned}
$$

B MAP-EM Estimator

With complete data including the unobserved $Z = \{z_t\}_{t=1}^T$ and $\Theta^* = \{\theta_t^*\}_{t=1}^T$, the data-augmented log-likelihood is given by (8) of the main text. The $m+1$-th MAP-EM estimator steps consist of:

1. Expectation (E-step):

$$Q(\theta, \sigma, \beta; \theta_m, \sigma_m, \beta_m) = \mathbb{E}_{Z,\Theta^*}\left[L(\theta, \sigma, \beta \mid X, Z, \Theta^*) \mid X, \theta_m, \sigma_m, \beta_m\right].$$

2. Maximization (M-step):

$$(\theta_{m+1}, \sigma_{m+1}, \beta_{m+1}) = \arg\max_{(\theta,\sigma,\beta)} \{Q(\theta, \sigma, \beta; \theta_m, \sigma_m, \beta_m) + p(\theta, \sigma, \beta)\}.$$

Recall that the prior $p(\theta, \sigma, \beta)$ is given by (7) in the main text. To execute the Expectation step, note that the probability of z_t conditional on $X, \theta_m, \sigma_m, \beta_m$ takes the form

$$\mathbb{P}(z_t \mid X, \theta_m, \sigma_m, \beta_m) = \frac{p(z_t \mid v_t, \theta_m, \sigma_m) p(s_t \mid v_t, \beta_m, u_{t,z_t})}{\sum_{z_t'=1}^{n_t} p(z_t' \mid v_t, \theta_m, \sigma_m) p(s_t \mid v_t, \beta_m, u_{t,z_t'})},$$

by an application of Bayes' rule. Furthermore,

$$\mathbb{P}(\theta_t^* \mid X, Z, \theta_m, \sigma_m, \beta_m) = \frac{\mathbb{I}_{(\ell_{t,z_t}, u_{t,z_t}]}(\theta_t^*) f(\theta_t^* \mid \theta, \sigma)}{p(z_t \mid v_t, \theta, \sigma)}.$$

We can then compute $Q(\theta, \sigma, \beta; \theta_m, \sigma_m, \beta_m)$ as follows:

$$
\begin{aligned}
Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m) &= \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\int_{\ell_{t,z_t}}^{u_{t,z_t}}\log(p(s_t \mid v_t,\beta,u_{t,z_t}))f(\theta_t^* \mid \theta_m,\sigma_m)d\theta_t^* \\
&\quad +\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\int_{\ell_{t,z_t}}^{u_{t,z_t}}\log(f(\theta_t^* \mid \theta,\sigma))f(\theta_t^* \mid \theta_m,\sigma_m)d\theta_t^* \\
&= \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\mathbb{E}_{z_t}[1 \mid \theta_m,\sigma_m]\sum_{i=z_t}^{N}s_{t,i}\log\left(\frac{v_{t,i}^{\beta}}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}\right) \\
&\quad +\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\frac{\theta}{\sigma^2}\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m,\sigma_m] \\
&\quad -\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\frac{1}{2\sigma^2}\mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m,\sigma_m] \\
&\quad -T\left(\log(\sqrt{2\pi}\sigma)+\frac{\theta^2}{2\sigma^2}\right),
\end{aligned}
$$

where $h(X_t \mid \theta_m,\sigma_m,\beta_m)=\sum_{z'_t=1}^{n_t}p(z'_t \mid v_t,\theta_m,\sigma_m)p(s_t \mid v_t,\beta_m,u_{t,z'_t})$, $X_t=(s_t,v_t)$, and with the expectation terms taking the form

$$
\mathbb{E}_{z_t}[\theta^{*p}_t \mid \theta_m,\sigma_m] := \int_{\ell_{t,z_t}}^{u_{t,z_t}}\theta^{*p}_t f(\theta_t^* \mid \theta_m,\sigma_m)d\theta_t^*, p=0,1,2. \tag{1}
$$

These are available in closed form (see online Appendix F).

To execute the Maximization step, we compute the first order conditions for a maximum, taking first partial derivatives:

$$
\frac{\partial Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m)+p(\theta,\sigma,\beta)}{\partial\theta} = \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\frac{1}{\sigma^2}\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m,\sigma_m]-\frac{\kappa(\theta-\mu)+T\theta)}{\sigma^2},
$$

$$\frac{\partial Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m)+p(\theta,\sigma,\beta)}{\partial\sigma} = -\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\frac{2\theta}{\sigma^3}\mathbb{E}_{z_t}[\theta_t^*\mid\theta_m,\sigma_m]$$
$$+\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\frac{1}{\sigma^3}\mathbb{E}_{z_t}[\theta_t^{*2}\mid\theta_m,\sigma_m]$$
$$-\frac{T+\nu+3}{\sigma}+\frac{T\theta^2+\kappa(\theta-\mu)^2+s^2}{\sigma^3},$$

and

$$\frac{\partial Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m)+p(\theta,\sigma,\beta)}{\partial\beta} = \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\times$$
$$\sum_{i=z_t}^{N}s_{t,i}\left(\log(v_{t,i})-\frac{\sum_{j=z_t}^{N}v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}\right). \tag{2}$$

Therefore, the updated iterates θ_{m+1} and σ_{m+1} are obtained as in equations (9) and (10) of the main text by solving the corresponding first order conditions. It is not possible to solve analytically for β_{m+1}, and we obtain it numerically by solving a single non-linear equation setting (2) to zero. We use Newton's method for that purpose[1] and the cross partial second derivative used is given by:

$$\frac{\partial^2 Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m)}{\partial\beta\partial\beta} = -\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\sum_{i=z_t}^{N}s_{t,i}\times$$
$$\left[\frac{\sum_{j=z_t}^{N}v_{t,j}^{\beta}\log(v_{t,j})^2}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}-\left(\frac{\sum_{j=z_t}^{N}v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}\right)^2\right].$$

[1]We use a MATLAB implementation of Newton's method in the *CompEcon Toolbox* by Miranda and Fackler (2002).

With regard to the priors, we opt for a vague prior for σ by setting $\nu = s^2 = 2$. This amounts to an $InverseGamma(1,1)$ distribution and a prior mode at $\frac{1}{2}$ for σ^2. This choice bounds our estimates of the nuisance parameter away from zero. We also set $\mu = 0$ and $\kappa = \frac{1}{100}$, the latter being the value chosen by Fraley and Raftery (2007) in their implementation. In combination, these two choices mildly bias our estimates towards zero expected thresholds.

We monitor convergence by measuring the distance between successive iterates and terminate the algorithm when this distance is less than 10^{-9}. To safeguard against isolating a local maximizer, we initiate the algorithm from different starting values. Specifically, we take all possible combinations of high, intermediate, and large values of the three parameters (27 possible combinations) as follows:

- Initial $\theta \in \{0, \sum_{t=1} \frac{\bar{v}_t}{2T}, \sum_{t=1} \frac{\bar{v}_t}{T}\}$,
- initial $\sigma \in \{\min\{0.1, \frac{se(\bar{v})}{4}\}, \frac{se(\bar{v})}{2}, se(\bar{v})\}$,
- and initial $\beta \in \{.9, 1, 2\}$.

Here, $se(\bar{v}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\bar{v}_t^2}{12} + (\frac{\bar{v}_t}{2})^2\right) - \left(\sum_{t=1}^{T}\frac{\bar{v}_t}{2T}\right)^2})$ is the standard deviation of a random variable that is drawn from the uniform in $[0, \bar{v}_t]$ with probability $\frac{1}{T}$ for each $t = 1, \ldots, T$.

C Standard Errors

The variance covariance matrix is calculated using the Hessian of the posterior evaluated at the point estimates, which is denoted by $H(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})$. We compute this matrix using a result from Dempster, Laird and Rubin (1977, p.10) (see also Jamshidian and Jennrich (2000)) for the EM context and applied in the MAP-EM case

$$H(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) = (\ddot{Q}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}; \widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) + \ddot{p}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})) \left(I_3 - \dot{M}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})\right).$$

Here, $\ddot{Q}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}; \widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) + \ddot{p}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})$ is a 3×3 matrix of second derivatives of $Q + p$ with respect to the components of its first three arguments. Also,

$$M(\theta_m, \sigma_m, \beta_m) := \arg\max_{\theta,\sigma,\beta} \{Q(\theta, \sigma, \beta; \theta_m, \sigma_m, \beta_m) + p(\theta, \sigma, \beta)\},$$

and $\dot{M}(\theta_m, \sigma_m, \beta_m)$ is the Jacobian with respect to θ_m, σ_m, and β_m. The first coordinate of $M(\theta_m, \sigma_m, \beta_m)$ is given by (9) and the second by (10) in the main text. Since the third coordinate is not available in closed form and is obtained by solving (2) numerically, its derivatives with respect to $x \in \{\theta_m, \sigma_m, \beta_m\}$ are computed using the Implicit Function theorem as follows:

$$-\frac{\frac{\partial^2 Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m)+p(\theta,\sigma,\beta)}{\partial x \partial \beta}}{\frac{\partial^2 Q(\theta,\sigma,\beta;\theta_m,\sigma_m,\beta_m)+p(\theta,\sigma,\beta)}{\partial \beta \partial \beta}}.$$

Thus, all the necessary partial derivatives needed to calculate the Jacobian of $M(\theta_m, \sigma_m, \beta_m)$ (and the Hessian $H(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})$) are obtained analytically (see online Appendix G for detailed derivations).

D Estimation of Restricted Model

Without parameters θ, σ and since we have assumed a flat improper prior for β, a MAP-EM estimator of the no-threshold model is equivalent to an EM-estimator which in turn is equivalent to a classic ML estimator. The log-likelihood of the restricted model is

$$L(\beta \mid X) = \sum_{t=1}^{T} \log\left(p(s_t \mid v_t, \beta, 0)\right). \tag{3}$$

The first order condition for the maximization of the log-likelihood is

$$\sum_{t=1}^{T}\sum_{i=1}^{N} s_{t,i}\left(\log(v_{t,i}) - \frac{\sum_{j=1}^{N} v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=1}^{N} v_{t,j}^{\beta}}\right) = 0.$$

We thus obtain the proportionality parameter using Newton's algorithm exploitng the fact that

$$\frac{\partial^2 L(\beta \mid X)}{\partial\beta\partial\beta} = -\sum_{t=1}^{T}\sum_{i=1}^{N} s_{t,i}\left[\frac{\sum_{j=1}^{N} v_{t,j}^{\beta}\log(v_{t,j})^2}{\sum_{j=1}^{N} v_{t,j}^{\beta}} - \left(\frac{\sum_{j=1}^{N} v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=1}^{N} v_{t,j}^{\beta}}\right)^2\right].$$

E Expected Threshold and its Variance

An analytical expression for the mean of the realized (instead of latent) threshold is:

$$\begin{aligned}\bar{\theta}(\theta,\sigma) := \mathbb{E}[\theta_t|\theta,\sigma] &= P\left(\theta_t^* \geq 0 \mid \theta,\sigma\right)\mathbb{E}[\theta_t^*|\theta_t^* \geq 0,\theta,\sigma] \\ &= \left(1-\Phi\left(-\tfrac{\theta}{\sigma}\right)\right)\theta + \sigma\phi\left(-\tfrac{\theta}{\sigma}\right).\end{aligned}$$

To derive an expression for its standard deviation, we first compute the expectation of the squared threshold

$$\begin{aligned}\mathbb{E}[\theta_t^2|\theta,\sigma] &= P\left(\theta_t^* \geq 0 \mid \theta,\sigma\right)\mathbb{E}[\theta_t^{*^2}|\theta_t^* \geq 0,\theta,\sigma] \\ &= \left(1-\Phi\left(-\tfrac{\theta}{\sigma}\right)\right)(\sigma^2+\theta^2) + \theta\sigma\phi(-\tfrac{\theta}{\sigma}),\end{aligned}$$

to obtain

$$\begin{aligned}\bar{\sigma}(\theta,\sigma) &= \sqrt{\left(1-\Phi\left(-\tfrac{\theta}{\sigma}\right)\right)(\sigma^2+\theta^2) + \theta\sigma\phi(-\tfrac{\theta}{\sigma}) - \left(\left(1-\Phi\left(-\tfrac{\theta}{\sigma}\right)\right)\theta + \sigma\phi\left(-\tfrac{\theta}{\sigma}\right)\right)^2} \\ &= \sqrt{\left(1-\Phi\left(-\tfrac{\theta}{\sigma}\right)\right)(\sigma^2+\theta^2) + \theta\sigma\phi(-\tfrac{\theta}{\sigma})\left(2\Phi(-\tfrac{\theta}{\sigma})-1\right) - \left(1-\Phi\left(-\tfrac{\theta}{\sigma}\right)\right)^2\theta^2 - \sigma^2\phi(-\tfrac{\theta}{\sigma})^2}.\end{aligned}$$

Both $\bar{\theta}(\theta,\sigma)$ and $\bar{\sigma}(\theta,\sigma)$ are continuously differentiable functions of (θ,σ,β). Taking their distribution to be (asymptotically) normal with variance $-\mathbb{E}[H(\theta,\sigma,\beta)]^{-1}$, the Delta method yieds estimates of the asymptotic variances:

$$\widehat{\bar{\theta}(\theta,\sigma)} \overset{a}{\sim} N\left(\bar{\theta}(\theta,\sigma), J_{\bar{\theta}(\theta,\sigma)}V_{\theta,\sigma}J'_{\bar{\theta}(\theta,\sigma)}\right), \tag{4}$$

where $J_{\bar{\theta}(\theta,\sigma)} = \left(\frac{\partial\bar{\theta}(\theta,\sigma)}{\partial\theta}, \frac{\partial\bar{\theta}(\theta,\sigma)}{\partial\sigma}\right)$ and $V_{\theta,\sigma}$ denotes the matrix formed by the first two columns and rows of $-\mathbb{E}[H(\theta,\sigma,\beta)]^{-1}$.

Similarly,

$$\widehat{\bar{\sigma}(\theta,\sigma)} \overset{a}{\sim} N\left(\bar{\sigma}(\theta,\sigma), J_{\bar{\sigma}(\theta,\sigma)} V_{\theta,\sigma} J'_{\bar{\sigma}(\theta,\sigma)}\right), \tag{5}$$

where $J_{\bar{\sigma}(\theta,\sigma)} = \left(\frac{\partial\bar{\sigma}(\theta,\sigma)}{\partial\theta}, \frac{\partial\bar{\sigma}(\theta,\sigma)}{\partial\sigma}\right)$.

The partial derivatives involved in the above Jacobians are:

$$\frac{\partial\bar{\theta}(\theta,\sigma)}{\partial\theta} = 1 - \Phi(-\theta/\sigma),$$

$$\frac{\partial\bar{\theta}(\theta,\sigma)}{\partial\sigma} = \phi(-\theta/\sigma),$$

$$\frac{\partial\bar{\sigma}(\theta,\sigma)}{\partial\theta} = \frac{1}{2\bar{\sigma}(\theta,\sigma)} 2\Phi(-\theta/\sigma)\left(\theta(1-\Phi(-\theta/\sigma)) + \sigma\phi(-\theta/\sigma)\right) + 2\theta\phi(-\theta/\sigma)^2(\theta-1), \text{ and}$$

$$\frac{\partial\bar{\sigma}(\theta,\sigma)}{\partial\sigma} = \frac{1}{2\bar{\sigma}(\theta,\sigma)} - 2\theta\phi(-\theta/\sigma) + 2\sigma\left(1-\Phi(-\theta/\sigma)\right) + 2\theta\Phi(-\theta/\sigma)\phi(-\theta/\sigma) - 2\sigma\phi(-\theta/\sigma)^2.$$

F First and Second Moments of Truncated Normal

To write closed form expressions for (1), let

$$\Phi_{z_t}(\theta_m, \sigma_m) := \int_{\ell_{t,z_t}}^{u_{t,z_t}} f(\theta_t \mid \theta_m, \sigma_m) d\theta_t.$$

With this notation we now write

$$\mathbb{E}[\theta_t \mid z_t, \theta_m, \sigma_m] \quad = \quad \Phi_{z_t}(\theta_m, \sigma_m)\theta_m + \left(\phi\left(\frac{\ell_{t,z_t} - \theta_m}{\sigma_m}\right) - \phi\left(\frac{u_{t,z_t} - \theta_m}{\sigma_m}\right)\right)\sigma_m,$$

and

$$\begin{aligned}\mathbb{E}[\theta_t^2 \mid z_t, \theta_m, \sigma_m] \quad &= \quad \sigma_m^2\left(\frac{\ell_{t,z_t} - \theta_m}{\sigma_m}\phi\left(\frac{\ell_{t,z_t} - \theta_m}{\sigma_m}\right) - \frac{u_{t,z_t} - \theta_m}{\sigma_m}\phi\left(\frac{u_{t,z_t} - \theta_m}{\sigma_m}\right) + \Phi_{z_t}(\theta_m, \sigma_m)\right) \\ &\quad +2\mathbb{E}[\theta_t \mid z_t, \theta_m, \sigma_m]\theta_m - \Phi_{z_t}(\theta_m, \sigma_m)\theta_m^2,\end{aligned}$$

where $\phi(.)$ is the standard normal probability density function.

G Derivative Calculations for Standard Errors

All the necessary partial derivatives needed to calculate the Jacobian of $M(\theta_m, \sigma_m, \beta_m)$ (and the Hessian $H(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})$) are obtained analytically and are given by the following expressions:

$$\ddot{Q}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}; \widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) + \ddot{p}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) = \begin{pmatrix} -\frac{T+\kappa}{\widehat{\sigma}^2} & 0 & 0 \\ 0 & -\frac{2(T+\nu+3)}{\widehat{\sigma}^2} & 0 \\ 0 & 0 & \frac{\partial^2 Q(\widehat{\theta},\widehat{\sigma},\widehat{\beta};\widehat{\theta},\widehat{\sigma},\widehat{\beta})}{\partial\beta\partial\beta} \end{pmatrix}.$$

In the above expression, use is made of the fact that $\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}$ are the maximizing values, which implies that the first order necessary conditions are satisfied.

As for the derivatives of the Jacobian of $M(\theta_m, \sigma_m, \beta_m)$, we have:

$$\begin{aligned}
\frac{\partial\theta_{m+1}}{\partial\theta_m} &= \frac{1}{T+\kappa}\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t, \beta_m, u_{t,z_t})}{h(X_t \mid \theta_m, \sigma_m, \beta_m)}\Bigg[\frac{\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]}{\sigma_m^2}\Bigg(\sum_{z_t'=1}^{n_t}\frac{p(s_t \mid v_t, \beta_m, u_{t,z_t'})}{h(X_t \mid \theta_m, \sigma_m, \beta_m)}\Big(\mathbb{E}_{z_t'}[1 \mid \theta_m, \sigma_m]\theta_m \\
&\quad -\mathbb{E}_{z_t'}[\theta_t^* \mid \theta_m, \sigma_m]\Big)\Bigg) + \frac{\partial\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]}{\partial\theta_m}\Bigg]. \\
\frac{\partial\theta_{m+1}}{\partial\sigma_m} &= \frac{1}{T+\kappa}\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t, \beta_m, u_{t,z_t})}{h(X_t \mid \theta_m, \sigma_m, \beta_m)}\Bigg[\frac{\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]}{\sigma_m}\Bigg(\sum_{z_t'=1}^{n_t}\frac{p(s_t \mid v_t, \beta_m, u_{t,z_t'})}{h(X_t \mid \theta_m, \sigma_m, \beta_m)}\Big(\mathbb{E}_{z_t'}[1 \mid \theta_m, \sigma_m] \\
&\quad -\frac{\mathbb{E}_{z_t'}[\theta_t^{*2} \mid \theta_m, \sigma_m] - 2\mathbb{E}_{z_t'}[\theta_t^* \mid \theta_m, \sigma_m]\theta_m + \mathbb{E}_{z_t'}[1 \mid \theta_m, \sigma_m]\theta_m^2}{\sigma_m^2}\Big)\Bigg) + \frac{\partial\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]}{\partial\sigma_m}\Bigg]. \\
\frac{\partial\theta_{m+1}}{\partial\beta_m} &= \frac{1}{T+\kappa}\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t, \beta_m, u_{t,z_t})}{h(X_t \mid \theta_m, \sigma_m, \beta_m)}\Bigg[\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]\Bigg(\frac{\partial\log(p(s_t \mid v_t, \beta_m, u_{t,z_t}))}{\partial\beta_m} \\
&\quad -\sum_{z_t'=1}^{n_t}\frac{p(s_t \mid v_t, \beta_m, u_{t,z_t'})}{h(X_t \mid \theta_m, \sigma_m, \beta_m)}\frac{\partial\log(p(s_t \mid v_t, \beta_m, u_{t,z_t'}))}{\partial\beta_m}\mathbb{E}_{z_t'}[1 \mid \theta_m, \sigma_m]\Bigg)\Bigg].
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \sigma_{m+1}}{\partial \theta_m} &= \frac{1}{2\sigma_{m+1}(T+\nu+3)}\left[\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\left[\frac{\partial \mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m,\sigma_m]}{\partial(\theta_m)} - \frac{\mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m,\sigma_m]}{\sigma_m^2}\right.\right.\\
&\left.\left.\left(\sum_{z'_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z'_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\left(\mathbb{E}_{z'_t}[\theta_t^{*} \mid \theta_m,\sigma_m] - \mathbb{E}_{z'_t}(1 \mid \theta_m,\sigma_m)\theta_m\right)\right)\right] - 2(T+\kappa)\theta_{m+1}\frac{\partial \theta_{m+1}}{\partial \theta_m}\right],\\
\frac{\partial \sigma_{m+1}}{\partial \sigma_m} &= \frac{1}{2\sigma_{m+1}(T+\nu+3)}\left[\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\left[\frac{\partial \mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m,\sigma_m]}{\partial \sigma_m} - \frac{\mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m,\sigma_m]}{\sigma_m}\sum_{z'_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z'_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\right.\right.\\
&\left.\left(\frac{\mathbb{E}_{z'_t}[\theta_t^{2} \mid \theta_m,\sigma_m] - 2\mathbb{E}_{z'_t}[\theta_t^{*} \mid \theta_m,\sigma_m](\theta_m) + \mathbb{E}_{z'_t}(1 \mid \theta_m,\sigma_m)(\theta_m)^2}{\sigma_m^2} - \mathbb{E}_{z'_t}(1 \mid \theta_m,\sigma_m)\right)\right]\\
&\left.-2(T+\kappa)\theta_{m+1}\frac{\partial \theta_{m+1}}{\partial \sigma_m}\right],\\
\frac{\partial \sigma_{m+1}}{\partial \beta_m} &= \frac{1}{2\sigma_{m+1}(T+\nu+3)}\left[\sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)}\left[\mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m,\sigma_m]\left(\frac{\partial \log(g(s_t \mid p_m,v_{tk}))}{\partial \beta_m} - \sum_{z'_t=1}^{n_t}\frac{p(s_t \mid v_t,\beta_m,u_{t,z'_t})}{h(X_t \mid \theta_m,\sigma_m,\beta_m)},\right.\right.\right.\\
&\left.\left.\left.\frac{\partial \log(p(s_t \mid v_t,\beta_m,u_{t,z'_t}))}{\partial \beta_m}\mathbb{E}_{z'_t}(1 \mid \theta_m,\sigma_m)\right)\right] - 2(T+\kappa)\theta_{m+1}\frac{\partial \theta_{m+1}}{\partial \beta_m}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]}{\partial \theta_m} &= \left[\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right]\frac{\theta_m}{\sigma_m} + \mathbb{E}_{z_t}[1 \mid \theta_m, \sigma_m] - \\
&\quad \left[\phi'\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \phi'\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right], \\
\frac{\partial \mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]}{\partial \sigma_m} &= \left[\frac{l_{t,z_t}-\theta_m}{\sigma_m^2}\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \frac{u_{t,z_t}-\theta_m}{\sigma_m^2}\phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right]\theta_m + \\
&\quad \left[\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right)^2\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)^2\phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right] + \\
&\quad \left[\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right], \\
\frac{\partial \mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m, \sigma_m]}{\partial \theta_m} &= 3\left[\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]\left(1 - \frac{\theta_m^2}{\sigma_m^2}\right) - \mathbb{E}_{z_t}[1 \mid \theta_m, \sigma_m]\theta_m\right] + \frac{\theta_m}{\sigma_m^2} \\
&\quad \left(2\mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m, \sigma_m] + \mathbb{E}_{z_t}[1 \mid \theta_m, \sigma_m]\theta_m^2\right) - \sigma_m\left[\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right)\right. \\
&\quad \left.\left(1 - \left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right)^2\right) - \phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\left(1 - \left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)^2\right)\right], \\
\frac{\partial \mathbb{E}_{z_t}[\theta_t^{*2} \mid \theta_m, \sigma_m]}{\partial \sigma_m} &= \left[\frac{l_{t,z_t}-\theta_m}{\sigma_m}\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \frac{u_{t,z_t}-\theta_m}{\sigma_m}\phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right]\left(2\sigma_m + \frac{\theta_m^2}{\sigma_m}\right) + \\
&\quad 2\mathbb{E}_{z_t}[1 \mid \theta_m, \sigma_m]\left(\sigma_m - \frac{\theta_m^2}{\sigma_m}\right) + 2\theta_m\left[\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right)^2\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right)\right. \\
&\quad \left. - \left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)^2\phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right] + 2\mathbb{E}_{z_t}[\theta_t^* \mid \theta_m, \sigma_m]\frac{\theta_m}{\sigma_m} \\
&\quad +\sigma_m\left[\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right)^3\phi\left(\frac{l_{t,z_t}-\theta_m}{\sigma_m}\right) - \left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)^3\phi\left(\frac{u_{t,z_t}-\theta_m}{\sigma_m}\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 Q(\gamma;\gamma_m)}{\partial\theta_m\partial\beta} &= \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\Bigg[\frac{\mathbb{E}_{z_t}[\theta_t^*\mid\theta_m,\sigma_m]-\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\theta_m}{\sigma_m^2}-\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m] \\
&\quad \sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\frac{\mathbb{E}_{z_t}[\theta_t^*\mid\theta_m,\sigma_m]-\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\theta_m}{\sigma_m^2}\Bigg] \\
&\quad \times\Bigg[\sum_{i=z_t}^{N}s_{ti}\Bigg(\log(v_{t,i})-\frac{\sum_{j=z_t}^{N}v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}\Bigg)\Bigg] \\
\frac{\partial^2 Q(\gamma;\gamma_m)}{\partial\sigma_m\partial\beta} &= \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\Bigg[\Bigg(\frac{\mathbb{E}_{z_t}[\theta_t^{*2}\mid\theta_m,\sigma_m]-2\mathbb{E}_{z_t}[\theta_t^*\mid\theta_m,\sigma_m]\theta_m+\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\theta_m^2}{\sigma_m^2}- \\
&\quad \mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\Bigg)\frac{1}{\sigma_m}-\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\frac{1}{\sigma_m} \\
&\quad \Bigg(\frac{\mathbb{E}_{z_t}[\theta_t^{*2}\mid\theta_m,\sigma_m]-2\mathbb{E}_{z_t}[\theta_t^*\mid\theta_m,\sigma_m]\theta_m+\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\theta_m^2}{\sigma_m^2}-\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\Bigg)\Bigg] \\
&\quad \times\Bigg[\sum_{i=z_t}^{N}s_{ti}\Bigg(\log(v_{t,i})-\frac{\sum_{j=z_t}^{N}v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}\Bigg)\Bigg], \\
\frac{\partial^2 Q(\gamma;\gamma_m)}{\partial\beta_m\partial\beta} &= \sum_{t=1}^{T}\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)}\Bigg[\frac{\partial\log(p(s_t\mid v_t,\beta_m,u_{t,z_t}))}{\partial\beta_m}-\sum_{z_t=1}^{n_t}\frac{p(s_t\mid v_t,\beta_m,u_{t,z_t})}{h(X_t\mid\theta_m,\sigma_m,\beta_m)} \\
&\quad \frac{\partial\log(p(s_t\mid v_t,\beta_m,u_{t,z_t}))}{\partial\beta_m}\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\Bigg]\mathbb{E}_{z_t}[1\mid\theta_m,\sigma_m]\times \\
&\quad \Bigg[\sum_{i=z_t}^{N}s_{ti}\Bigg(\log(v_{t,i})-\frac{\sum_{j=z_t}^{N}v_{t,j}^{\beta}\log(v_{t,j})}{\sum_{j=z_t}^{N}v_{t,j}^{\beta}}\Bigg)\Bigg].
\end{aligned}
$$

H Auxiliary Tests

All tests are based on Pearson's goodness-of-fit test statistic. We use simulation under the null hypothesis to compute exact p-values instead of relying on asymptotic arguments. Exact procedures are described below for each test.

H.1 Fit

We assess model fit using the test statistic

$$P((s_t)_{t=1}^{T}, \widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{(s_{t,i} - \overline{s}_{t,i}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}))^2}{\overline{s}_{t,i}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})}. \tag{6}$$

The null hypothesis is that the observed seat allocations are drawn from the specified model multinomial distribution with means as specified by the model parameters, specifically, $\overline{s}_{t,i}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta})$ is the expected number of seats of party i in election t under the null:

$$\overline{s}_{t,i}(\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}) = S_t \sum_{z_t=1}^{N} q_{t,i}(v_t, \widehat{\beta}, u_{t,z_t}) p(z_t \mid v_t, \widehat{\theta}, \widehat{\sigma}),$$

where $p(N \mid v_t, \widehat{\theta}, \widehat{\sigma}) = 1 - \Phi\left(\frac{u_{t,N-1} - \widehat{\theta}}{\widehat{\sigma}}\right)$. We simulate the distribution of the test statistic (and calculate the p-value) by calculating (6) using simulated seat allocations generated for the observed vote shares under the null. Specifically, to obtain one of $10,000$ realizations from the distribution of this statistic:

1. For each t, we draw a latent threshold $\widetilde{\theta}_t^*$ from the normal distribution with mean $\widehat{\theta}$ and variance $\widehat{\sigma}^2$, which determines a non-negative threshold, $\widetilde{\theta}_t$ ($\widetilde{\theta}_t = \widetilde{\theta}_t^*$ if $\widetilde{\theta}_t^* \geq 0$ and $\widetilde{\theta}_t = 0$ otherwise).

2. For each t, we draw a vector $\tilde{s}_t$ allocating S_t total seats from the Multinomial distribution with probabilities $q_{t,i}(v_t, \widehat{\beta}, \widetilde{\theta}_t)$ as defined in (2).

3. We compute the statistic in (6) using the simulated seats $(\tilde{s}_t)_{t=1}^T$ instead of the actual seats $(s_t)_{t=1}^T$, $P((\tilde{s}_t)_{t=1}^T, \widehat{\theta}, \widehat{\sigma}, \widehat{\beta})$.

We repeat this process 10,000 times and compute the p-value as the fraction of the Pearson's statistics calculated with these simulated seats that is greater than or equal to P.

Similarly, the fit test of the restricted model is based on the statistic

$$P^o((s_t)_{t=1}^T, \widehat{\beta}^o) = \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{(s_{i,t} - \overline{s}^o_{i,t}(\widehat{\beta}^o))^2}{\overline{s}^o_{i,t}(\widehat{\beta}^o)}. \tag{7}$$

Party i's expected seats according to the restricted model are given by

$$\overline{s}^o_{i,t}(\widehat{\beta}^o) = S_t q_{i,t}(v_t, \widehat{\beta}^o, 0),$$

where $\widehat{\beta}^o$ denotes the estimated disproportionality parameter from the model with no threshold. The only difference in the computation of the p-value is that we now skip step one and set the threshold to zero instead. Step 2 is now executed with multinomial probabilities $q_{i,t}(v_t, \widehat{\beta}^o, 0)$ and the remaining steps are analogous.

H.2 MAP-EM vs Restricted Zero-Threshold Model

The restricted model has a threshold of zero (both in expectation and realized threshold) and proportionality parameter $\widehat{\beta}^o$. We can perform a simple comparison test between the estimated restricted model and the MAP-EM fully estimated model. The null hypothesis is that the the seats are drawn (conditional on the observed votes) from the restricted model with zero threshold and proportionality $\widehat{\beta}^o$, whereas the alternative is that they are drawn from the full model with parameters $\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}$. We execute this test using the following statistic:

$$D((s_t)_{t=1}^T, \widehat{\theta}, \widehat{\sigma}, \widehat{\beta}, \widehat{\beta}^o) = P^o((s_t)_{t=1}^T, \widehat{\beta}^o) - P((s_t)_{t=1}^T, \widehat{\theta}, \widehat{\sigma}, \widehat{\beta}), \tag{8}$$

where P^o and P are computed as in (7) and (6), respectively. Large values of the statistic provide evidence against the estimated restricted model and in favor of the alternative MAP-EM estimated model with thresholds. We compute the p-value by drawing seat allocations under the null (from the Multinomial distribution with probabilities $q_{t,i}(v_t, \widehat{\beta}^o, 0)$, as in the case of the fit statistic P^o), evaluating (8) using the simulated seats instead of the actual seats, and computing the fraction of simulated statistics that exceed D to obtain the p-value.

H.3 Electoral System Change

Consider a set of elections $t = 1, \ldots, T$. Under the null hypothesis one electoral system governs all T elections with the conditional distribution of seats determined by the estimated parameters denoted by $(\widehat{\theta}^c, \widehat{\sigma}^c, \widehat{\beta}^c)$. The alternative hypothesis is that these T elections are partitioned into two or more (fixed) subsets $T_f \subset \{1, \ldots, T\}$, a different conditional distribution of seats governs seat allocation in each subset, f, and this distribution is determined by the estimated parameters $(\widehat{\theta}^f, \widehat{\sigma}^f, \widehat{\beta}^f)$ obtained by the MAP-EM estimates when the subset of elections T_f is used for their estimation. The comparison test computes the following statistic for elections that are part of the coarser system

$$D^{systems}((s_t)_{t=1}^T, \widehat{\theta}^c, \widehat{\sigma}^c, \widehat{\beta}^c, (\widehat{\theta}^f, \widehat{\sigma}^f, \widehat{\beta}^f)_f) = \sum_t P_t^c - \sum_f \sum_{t \in T_f} P_t^f, \tag{9}$$

where

$$P_t^k = \sum_{i=1}^N \frac{(s_{t,i} - \overline{s}_{t,i}(\widehat{\theta}^k, \widehat{\sigma}^k, \widehat{\beta}^k))^2}{\overline{s}_{t,i}(\widehat{\theta}^k, \widehat{\sigma}^k, \widehat{\beta}^k)}, k \in \{c, f\}.$$

We simulate the distribution of this statistic under the null. We obtain one realization from that distribution as follows:

1. For each t, we draw a latent threshold $\widetilde{\theta}_t^*$ from the normal distribution with mean $\widehat{\theta}^c$ and variance $(\widehat{\sigma}^c)^2$, which determines a non-negative threshold, $\widetilde{\theta}_t$ ($\widetilde{\theta}_t = \widetilde{\theta}_t^*$ if $\widetilde{\theta}_t^* \geq 0$ or $\widetilde{\theta}_t = 0$ otherwise).

2. For each t, we draw a vector $\tilde{s}_t$ allocating S_t total seats from the Multinomial distribution with probabilities $q_{t,i}(v_t, \widehat{\beta}^c, \widetilde{\theta}_t)$ as defined in (2).

3. We compute the statistic in (9) using the simulated seats $(\tilde{s}_t)_{t=1}^T$ instead of the actual seats $(s_t)_{t=1}^T$.

We repeat this process 10,000 times and compute the p-value as the fraction of these 10,000 Pearson's statistics that is greater than or equal to $D^{systems}$.

H.4 Bayesian Tests

We can perform similar tests using information from the full posterior. The tests of fit can be performed using the same test statistics but using the posterior predictive distribution as detailed in Gelman et al. (2004). Specifically, using (6), prior to the first step described in subsection H.1, we would obtain a sample of size one of the model parameters from the posterior distribution, say θ', σ', β'. We would then execute Steps 1-3 to obtain a statistic, drawing threshold realizations $\tilde{\theta}_t$, simulated seats $\tilde{s}_t$, and computing the statistic $P((\tilde{s}_t)_{t=1}^T, \theta', \sigma', \beta')$. This approach would yield a p-value as the fraction of times this statistic exceeds the statistic computed with the actual seats $P((s_t)_{t=1}^T, \theta', \sigma', \beta')$. We can perform the model comparison tests for electoral system change either using a version of the Bayesian Information Criterion (BIC) as suggested in Fraley and Raftery (2007), or the Deviance Information Criterion Spiegelhalter et al. (2002) which has the advantage that it more naturally compensates for model complexity. The latter test is easy to compute once a sample of parameters from the posterior distribution is available.

I Full Results

Table 1: MAP-EM Estimates and Comparison with Restricted Model

				Full			Restricted	Tests (p-values)		
Country	System	T	Dates	$\bar{\theta}(\hat{\theta},\hat{\sigma})$	$\bar{\sigma}(\hat{\theta},\hat{\sigma})$	$\hat{\beta}$	$\hat{\beta}^o$	Fit	Fit^o	Difference
Austria	AUT1	8	1945-1970	3.938	0.585	1.283*	1.371*	1	0.805	0.018$^+$
				(0.471)	(0.154)	(0.064)	(0.061)			
	AUT2	6	1971-1990	3.247	0.628	0.981	1.156*	0.995	0.506	0^+
				(1.064)	(0.18)	(0.057)	(0.05)			
	AUT3	6	1994-2008	3.841	0.603	1.02	1.231*	1	0.589	0^+
				(0.497)	(0.169)	(0.059)	(0.05)			
Belgium	BEL1	16	1946-1991	0.518	0.443	1.202*	1.22*	0.995	0.95	0.203
				(0.22)	(0.098)	(0.028)	(0.027)			
	BEL2	2	1995-1999	0.696	0.539	1.3*	1.326*	1	0.998	0.244
				(0.835)	(0.186)	(0.139)	(0.132)			
	BEL3	3	2003-2010	0.53	0.501	1.322*	1.326*	0.991	0.938	0.287
				(1.225)	(0.326)	(0.103)	(0.101)			
Bulgaria	BGR1	5	1991-2005	4.332	0.564	1.015	1.377*	1	0.353	0^+
				(0.386)	(0.146)	(0.044)	(0.034)			
	BGR2^b	1	2009	3.652	0.605	0.994	1.185*	0.985	0.051	0^+
				(0.656)	(0.175)	(0.1)	(0.09)			
Cyprus	CYP1	1	1981	4.039	0.65	0.68	1.254	0.419	0.232	0.016$^+$
				(1.836)	(0.207)	(0.366)	(0.281)			
	CYP2	2	1985-1991	3.6	0.646	0.969	1.173	1	0.952	0.073
				(1.944)	(0.205)	(0.265)	(0.223)			
	CYP3	3	1996-2006	1.586	0.584	1.067	1.158	1	1	0.123
				(0.5)	(0.166)	(0.118)	(0.105)			
Czech	CZE1	2	1996-1998	4.966	0.652	1.041	1.406*	1	0.15	0^+
Republic				(1.178)	(0.195)	(0.088)	(0.071)			
	CZE2	3	2002-2010	5.231	0.647	1.223*	1.685*	0.999	0.329	0^+
				(0.929)	(0.192)	(0.105)	(0.083)			
Germany	DEU1	1	1949	0.012	0.073	1.139*	1.139*	0.986	0.865	0.374
				(0.116)	(1.107)	(0.058)	(0.058)			
	DEU2	1	1953	0.287	0.385	1.22*	1.225*	0.796	0.007	0.1
				(0.584)	(0.319)	(0.054)	(0.053)			
	DEU3	8	1957-1983	3.866	0.744	1.04	1.182*	0.22	0.138	0^+
				(0.537)	(0.219)	(0.027)	(0.024)			

	DEU4	1	1987	2.019	0.634	0.998	1.103⋆	1	0.795	0.001^{+}
				(2.258)	(0.2)	(0.073)	(0.062)			
	DEU5	1	1990	2.255	0.586	1.109⋆	1.274⋆	0.48	0.065	0^{+}
				(0.59)	(0.169)	(0.055)	(0.049)			
	DEU6	2	1994-1998	3.118	0.627	1.018	1.181⋆	1	0.284	0^{+}
				(1.075)	(0.184)	(0.036)	(0.03)			
	DEU7	2	2002-2005	2.696	0.627	1.092⋆	1.226⋆	0.265	0.26	0^{+}
				(1.174)	(0.187)	(0.044)	(0.037)			
	DEU8	1	2009	2.812	0.639	1.046	1.21⋆	0.931	0.1	0^{+}
				(2.03)	(0.203)	(0.091)	(0.078)			
Denmark	DNK1	2	1945-1947	1.303	0.586	1.047	1.083	0.992	0.978	0.178
				(0.617)	(0.157)	(0.084)	(0.081)			
	DNK2	2	1950	2.134	0.635	0.999	1.052	1	0.999	0.079
			4/1953	(2.206)	(0.2)	(0.088)	(0.081)			
	DNK3	3	9/1953	0.003	0.036	1.085	1.085	0.771	0.664	0.718
			1957-1960	(0.034)	(0.495)	(0.055)	(0.055)			
	DNK4	3	1964-1968	1.835	0.536	0.998	1.169⋆	1	0.765	0.003^{+}
				(0.357)	(0.135)	(0.068)	(0.059)			
	DNK5	14	1971-2005	1.874	0.544	1.006	1.102⋆	0.958	0.958	0.052
				(0.172)	(0.104)	(0.027)	(0.025)			
	DNK6	1	2007	1.426	0.598	0.997	1.051	1	0.993	0.115
				(0.794)	(0.167)	(0.122)	(0.115)			
Spain	ESP1	10	1977-2008	0.187	0.269	1.255⋆	1.273⋆	0.973	0.628	0.054
				(0.105)	(0.074)	(0.02)	(0.018)			
Estonia	EST1	1	1992	1.664	0.608	1.524⋆	1.585⋆	0.992	0.775	0.091
				(0.918)	(0.178)	(0.183)	(0.17)			
	EST2	2	1995-1999	4.231	0.621	1.153	1.453⋆	1	0.822	0.001^{+}
				(0.761)	(0.179)	(0.129)	(0.107)			
	EST3	2	2003-2007	3.503	0.644	1.186	1.381⋆	1	0.984	0.037^{+}
				(1.897)	(0.205)	(0.156)	(0.132)			
Finland	FIN1	18	1945-2007	0.359	0.451	1.195⋆	1.207⋆	0.991	0.946	0.476
				(0.284)	(0.162)	(0.031)	(0.029)			
France	FRA1	3	1945-1946	0.031	0.123	1.347⋆	1.347⋆	1	0.773	0.788
				(0.307)	(1.801)	(0.075)	(0.075)			
	FRA2	2	1951-1956	0.15	0.286	0.779	0.779	0.028	0.001	0.052
				(1.513)	(2.076)	(0.078)	(0.078)			
	FRA3	2	1958-1962	1.713	0.606	1.29⋆	1.359⋆	0.014	0	0.063
				(0.581)	(0.181)	(0.068)	(0.064)			
	FRA4	3	1967-1973	1.03	0.563	1.258⋆	1.268⋆	0	0	0.324

				(0.576)	(0.136)	(0.045)	(0.044)			
	FRA5	2	1978-1981	1.492	0.558	1.307⋆	1.426⋆	0.148	0	0.554
				(0.408)	(0.146)	(0.073)	(0.066)			
	FRA6	1	1986	0.77	0.54	1.199⋆	1.222⋆	0.033	0	0.947
				(0.632)	(0.152)	(0.07)	(0.067)			
	FRA7	5	1988-2007	0.008	0.059	1.694⋆	1.694⋆	0.881	0	1
				(0.086)	(0.947)	(0.038)	(0.038)			
United	GBR1	18	1945-2010	0	0.01	1.486⋆	1.486⋆	1	0	0.976
Kingdom				(0.003)	(0.084)	(0.021)	(0.021)			
Greece	GRC1	1	1974	2.013	0.635	1.722⋆	1.732⋆	1	0.04	0.217
				(2.669)	(0.201)	(0.107)	(0.105)			
	GRC2	2	1977-1981	0.285	0.39	1.742⋆	1.745⋆	1	0.195	0.806
				(1.219)	(0.701)	(0.089)	(0.088)			
	GRC3	1	1985	0.586	0.519	1.673⋆	1.683⋆	1	0.973	0.439
				(1.542)	(0.377)	(0.171)	(0.167)			
	GRC4	3	1989-1990	0.006	0.05	1.248⋆	1.249⋆	1	0.744	0.496
				(0.055)	(0.677)	(0.055)	(0.055)			
	GRC5	4	1993-2004	3.171	0.556	1.254⋆	1.411⋆	0.952	0.362	0.003^{+}
				(0.331)	(0.143)	(0.053)	(0.049)			
	GRC6	2	2007-2009	3.102	0.61	1.105⋆	1.209⋆	0.697	0.309	0.002^{+}
				(0.691)	(0.177)	(0.059)	(0.054)			
Croatia	HRV1[a]	1	1992	0.005	0.048	2.201⋆	2.201⋆	0.167	0	0.998
				(0.059)	(0.741)	(0.225)	(0.225)			
	HRV2[b]	1	1992	2.893	0.595	1.048	1.228⋆	0.998	0.902	0.021^{+}
				(0.629)	(0.173)	(0.145)	(0.127)			
	HRV3[a]	1	1995	0.252	0.369	7.829⋆⋆	7.829⋆	1	1	0.609
				(3.163)	(2.164)	(3.109)	(3.109)			
	HRV4[b]	1	1995	4.032	0.628	1.05	1.294⋆	1	0.524	0.002^{+}
				(0.891)	(0.184)	(0.156)	(0.133)			
	HRV5	3	2000-2007	1.168	0.652	1.305⋆	1.386⋆	0.966	0.703	0.131
				(0.48)	(0.181)	(0.074)	(0.066)			
Hungary	HUN1	1	1990	0.003	0.035	2.332⋆	2.332⋆	0.867	0.001	0.998
				(0.035)	(0.512)	(0.184)	(0.184)			
	HUN2	5	1994-2010	0.002	0.026	2.149⋆	2.149⋆	0.968	0.002	0.995
				(0.019)	(0.328)	(0.081)	(0.081)			
Ireland	IRL1	6	1948-1965	0.209	0.331	1.124⋆	1.128⋆	0.99	0.929	0.382
				(0.411)	(0.338)	(0.05)	(0.049)			
	IRL2	3	1969-1977	2.175	0.596	1.177	1.288⋆	0.975	0.808	0.083
				(0.637)	(0.169)	(0.123)	(0.107)			

	IRL3	9	1981-2007	0.13	0.263	1.196*	1.2*	1	0.922	0.499
				(0.294)	(0.467)	(0.038)	(0.036)			
Iceland	ISL1	5	1946-1956,	5.614	0.669	0.94	1.084	0.282	0.144	0.017$^+$
			6/1959	(1.533)	(0.207)	(0.151)	(0.136)			
	ISL2	8	10/1959,	1.036	0.591	1.138	1.18*	0.979	0.986	0.28
			1963-1983	(0.77)	(0.187)	(0.093)	(0.084)			
	ISL3	4	1987-1999	1.193	0.619	1.13	1.203*	1	1	0.224
				(0.695)	(0.187)	(0.116)	(0.1)			
	ISL4	3	2003-2009	4.442	0.655	1.114	1.258*	1	0.993	0.08
				(1.82)	(0.208)	(0.157)	(0.141)			
Italy	ITA1	1	1946	0.007	0.053	1.04	1.04	0.982	0.978	0.366
				(0.071)	(0.841)	(0.049)	(0.049)			
	ITA2	1	1948	0.005	0.046	1.142*	1.142*	0.999	0.908	0.429
				(0.056)	(0.712)	(0.046)	(0.046)			
	ITA3	1	1953	0.021	0.101	1.14*	1.14*	0.998	0.944	0.648
				(0.189)	(1.374)	(0.05)	(0.05)			
	ITA4	9	1958-1992	0.001	0.014	1.125*	1.125*	1	0.967	0.517
				(0.006)	(0.138)	(0.015)	(0.015)			
	ITA5	2	1996-2001	0.002	0.024	1.221*	1.221*	0.94	0.167	0.488
				(0.017)	(0.298)	(0.041)	(0.041)			
	ITA6[a]	2	2006-2008	0.002	0.025	1.17*	1.17*	0.934	0.148	0.167
				(0.018)	(0.32)	(0.028)	(0.028)			
Lithuania	LTU1[b]	1	1992	1.702	0.597	1.237	1.347*	0.827	0.591	0.041$^+$
				(0.74)	(0.172)	(0.154)	(0.134)			
	LTU2[b]	4	1996-2008	5.13	0.597	1.008	1.656*	1	0.433	0$^+$
				(0.512)	(0.161)	(0.124)	(0.09)			
Luxembourg	LUX1	3	1945	0.258	0.373	1.186	1.187	1	1	0.453
			1954-1959	(2.095)	(1.385)	(0.146)	(0.145)			
	LUX2	2	1948-1951	0.252	0.369	0.617	0.617	0.178	0.134	0.907
				(3.162)	(2.163)	(0.235)	(0.235)			
	LUX3	4	1964-1979	1	0.59	1.23*	1.243*	1	1	0.404
				(1.216)	(0.177)	(0.131)	(0.128)			
	LUX4	6	1984-2009	2.447	0.574	1.162*	1.245*	0.573	0.615	0.47
				(0.428)	(0.154)	(0.098)	(0.09)			
Latvia	LVA1	1	1993	3.584	0.631	0.986	1.279*	1	0.869	0$^+$
				(1.059)	(0.187)	(0.142)	(0.11)			
	LVA2	5	1995-2010	5.056	0.573	1.01	1.538*	1	0.647	0$^+$
				(0.416)	(0.149)	(0.111)	(0.084)			
North	MKD1	1	1994	0.007	0.054	1.373*	1.373*	0.096	0	0.973

Macedonia				(0.073)	(0.858)	(0.125)	(0.125)			
	MKD2[a]	1	1998	0.252	0.369	1.416⋆	1.416⋆	0.042	0	0.955
				(3.163)	(2.164)	(0.213)	(0.213)			
	MKD3[b]	1	1998	5.594	0.662	1.158	1.419	0.999	0.924	0.123
				(1.225)	(0.199)	(0.417)	(0.38)			
	MKD4	3	2002-2008	0.81	0.56	1.215⋆	1.307⋆	0.998	0.978	0.039^{+}
				(0.379)	(0.137)	(0.072)	(0.06)			
Malta	MLT1	2	1966-1971	0.257	0.372	2.84*	2.84	0.999	0.895	0.424
				(3.247)	(2.166)	(1.357)	(1.356)			
	MLT2	2	1976-1981	0.252	0.369	0.439*	0.439	0.721	0.711	0.558
				(3.163)	(2.164)	(3.462)	(3.462)			
	MLT3	2	1987-1992	2.898	0.64	1.015*	1.854	1	0.901	0.342
				(2.633)	(0.203)	(2.764)	(1.02)			
	MLT4	4	1996-2008	2.274	0.637	1.652*	2.099	0.99	0.966	0.404
				(3.466)	(0.202)	(1.807)	(0.937)			
Moldova	MOL1	3	1994-1998	5.076	0.661	1.004	1.496⋆	1	0.447	0^{+}
			2010	(1.631)	(0.21)	(0.111)	(0.083)			
	MOL2	3	2001-2005	6.814	0.679	1.021	1.489⋆	1	0.2	0^{+}
			4/2009	(1.14)	(0.201)	(0.09)	(0.07)			
	MOL3	1	7/2009	3.211	0.642	1.04	1.207	1	0.774	0.015^{+}
				(1.957)	(0.204)	(0.157)	(0.133)			
Montenegro	MON1	2	2006-2009	3.295	0.615	1.007	1.295⋆	1	0.764	0^{+}
				(0.64)	(0.185)	(0.109)	(0.086)			
Netherlands	NLD1	3	1946-1952	0.557	0.506	1.06	1.067	1	1	0.355
				(0.869)	(0.226)	(0.086)	(0.085)			
	NLD2	17	1956-2010	0.535	0.402	1.049⋆	1.07⋆	1	1	0.126
				(0.162)	(0.066)	(0.025)	(0.024)			
Norway	NOR1	2	1945-1949	0.526	0.495	1.272⋆	1.278⋆	0.956	0.499	0.459
				(0.922)	(0.252)	(0.076)	(0.075)			
	NOR2	9	1953-1985	0.412	0.49	1.155⋆	1.17⋆	0.946	0.661	0.474
				(0.345)	(0.22)	(0.034)	(0.031)			
	NOR3	4	1989-2001	0.241	0.367	1.238⋆	1.253⋆	0.959	0.749	0.513
				(0.397)	(0.313)	(0.061)	(0.055)			
	NOR4	2	2005-2009	2.401	0.63	1.076	1.168⋆	0.975	0.905	0.007^{+}
				(1.428)	(0.191)	(0.083)	(0.074)			
Poland	POL1	1	1991	0.002	0.028	1.269⋆	1.269⋆	0.989	0.217	0.701
				(0.023)	(0.378)	(0.054)	(0.054)			
	POL2	2	1993-1997	5.124	0.578	1.638⋆	2.056⋆	1	0.03	0^{+}
				(0.432)	(0.155)	(0.072)	(0.064)			

	POL3	1	2001	6.556	0.673	1.052	1.265⋆	0.995	0.023	0^+
				(1.1)	(0.2)	(0.065)	(0.057)			
	POL4	2	2005-2007	5.074	0.661	1.238⋆	1.499⋆	1	0.139	0^+
				(1.568)	(0.205)	(0.068)	(0.057)			
Portugal	PRT1	2	1975-1976	0.413	0.482	1.327⋆	1.379⋆	0.071	0.01	0.989
				(0.398)	(0.194)	(0.079)	(0.071)			
	PRT2	5	1979-1987	1.538	0.565	1.198⋆	1.334⋆	0.719	0.537	0.02^+
				(0.366)	(0.148)	(0.052)	(0.044)			
	PRT3	6	1991-2009	1.565	0.55	1.276⋆	1.349⋆	1	0.978	0.027^+
				(0.409)	(0.14)	(0.046)	(0.042)			
Romania	ROM1	1	1990	0.364	0.412	0.98	1.047	1	1	0^+
				(0.382)	(0.172)	(0.033)	(0.03)			
	ROM2	2	1992-1996	3.069	0.576	1.017	1.453⋆	0.877	0.4	0^+
				(0.497)	(0.159)	(0.06)	(0.045)			
	ROM3	2	2000-2004	5.491	0.635	0.996	1.441⋆	0.999	0.49	0^+
				(0.664)	(0.184)	(0.064)	(0.047)			
	ROM4	1	2008	4.36	0.651	0.99	1.441⋆	0.997	0.451	0^+
				(1.574)	(0.202)	(0.124)	(0.095)			
Slovenia	SVN1	2	1992-1996	2.965	0.555	0.862	1.352⋆	0.999	0.666	0^+
				(0.415)	(0.149)	(0.128)	(0.096)			
	SVN2	3	2000-2008	3.654	0.585	1.015	1.243⋆	1	0.928	0^+
				(0.465)	(0.161)	(0.088)	(0.073)			
Servia	SER1	2	2007-2008	4.176	0.627	1.013	1.322⋆	1	0.311	0^+
				(0.838)	(0.18)	(0.074)	(0.061)			
Slovakia	SVK1	1	1994	4.312	0.635	1.013	1.324⋆	1	0.22	0^+
				(0.966)	(0.187)	(0.118)	(0.093)			
	SVK2	4	1998-2010	4.829	0.599	1.007	1.424⋆	1	0.268	0^+
				(0.539)	(0.164)	(0.081)	(0.064)			
Sweden	SWE1	1	1948	0.252	0.369	1.16	1.16	0.795	0.567	0.51
				(3.163)	(2.164)	(0.105)	(0.105)			
	SWE2	6	1952-1968	0.006	0.049	1.105⋆	1.105⋆	0.91	0.703	0.955
				(0.061)	(0.756)	(0.038)	(0.038)			
	SWE3	13	1970-2010	3.644	0.552	1.004	1.088⋆	1	0.397	0^+
				(0.36)	(0.135)	(0.021)	(0.02)			
Switzerland	SWZ1	16	1947-2007	0.004	0.04	1.109⋆	1.109⋆	1	1	0.442
				(0.035)	(0.497)	(0.023)	(0.023)			
Ukraine	UKR1	1	1994	0.117	0.249	1.23⋆	1.239⋆	0.911	0.461	0.374
				(0.355)	(0.637)	(0.075)	(0.071)			
	UKR2[b]	2	1998-2002	4.257	0.593	0.993	1.518⋆	0.981	0.097	0^+

			(0.537)	(0.168)	(0.076)	(0.056)			
UKR3[a]	2	1998-2002	0.002	0.028	1.337⋆	1.337⋆	0.991	0.176	0.666
			(0.022)	(0.365)	(0.072)	(0.072)			
UKR4	2	2006-2007	3.422	0.572	1.003	1.356⋆	0.963	0	0^{+}
			(0.454)	(0.154)	(0.054)	(0.046)			

Estimates of expected national electoral threshold, $\bar{\theta}(\hat{\theta}, \hat{\sigma})$, threshold standard deviation, $\bar{\sigma}(\hat{\theta}, \hat{\sigma})$, and disproportionality, $\hat{\beta}$. Estimates of disproportionality for models with no threshold, $\hat{\beta}^{o}$. Standard errors in parentheses.

a SMD partition of mixed system.

b PR partition of mixed system.

* Not statistically different than three (3) at the 5% level of significance (two-tailed test).

⋆ Larger than one at the 5% level of significance (one-tailed test).

\+ Comparison test favors model with threshold at 5% level of significance.

J Tests of Change in Electoral Systems

Table 2: Statistically Significant Electoral System Changes

Panel A: Coarse vs. 5% cutoff definition			*Panel B*: 5% cutoff vs. finest definition		
System	Years	p-values	System	Years	p-values
CYP1-2	1981-1991	0.275	BEL1	1946-1991	0.25
DEU4-5	1987-1990	0.268	DEU3	1957-1983	0.094
DEU6-7	1994-2005	0.306	FRA1	1945-1946	0.283
EST1-2	1992-1999	0.024	FRA4	1967-1973	0.055
FRA3-5,7	1958-1981,1988-2007	0	FRA7	1988-2007	0.058
HRV1,3[a]	1992,-1995	0.006	GBR1	1945-2010	0
HRV2,4[b]	1992,-1995	0.171	IRL1	1948-1965	0.245
HUN1-2	1990-2010	0.542	IRL2	1969-1977	0.321
IRL1-3	1948-2007	0.032	IRL3	1981-2007	0.19
ISL1-2	1946-1983	0.025	ITA4	1958-1992	0.373
LTU1-2	1992-2008	0	LUX1	1945,1954-1959	0.237
LUX1-4	1945-2009	0.088	LUX3	1964-1979	0.407
LVA1-2	1993-2010	0.003	LUX4	1984-2009	0.106
MLT1-2	1966-1981	0.003	MLT1	1966-1971	0.29
MOL1-3	1994-2010	0.428	MOL1	1994-2010	0.245
NLD1-2	1946-2010	0.595	NOR1	1945-1949	0.395
NOR3-4	1989-2009	0.106	NOR2	1953-1985	0.055
PRT1-3	1975-2009	0	PRT1	1975-1976	0.018
ROM2-4	1992-2008	0.04	ROM2	1992-1996	0.389
UKR2,4[c]	1998,2002,2006-2007	0.175	ROM3	2000-2004	0.319
			SVK2	1998-2010	0.424
			SWE2	1952-1968	0.368
			SWE3	1970-2010	0.539
			SWZ1	1947-2007	0.063

Reports tests of change in electoral systems comparing two pairs of alternative *ad hoc* definitions. Panel A presents results of the comparison between our default 5% cutoff definition and a coarser definition. Panel B presents results of the comparison between our default 5% cutoff definition and a finer definition of electoral systems. In each case, the null hypothesis is that the coarser definition is correct (no change within the corresponding period) and small p-values indicate support for the finer definition. 'System' reports systems involved in the test according to our default definition. 'Years' reports the years covered by the test.
a SMD partition of mixed system.
b PR partition of mixed system.
c years 1998 and 2002 PR partition.

K Sensitivity to MAP-EM priors & Comparison with EM

To probe the sensitivity in estimated parameters and standard errors to the prior specification, we run the MAP-EM algorithm with four (4) sets of different prior specifications:

1. The data-independent prior used in the main body of the paper:

   - $\nu = 2$,
   - $s^2 = 2$,
   - $\mu = 0$,
   - $\kappa = \frac{1}{100}$.

   This sets the prior mean of the mean of the latent threshold, θ, at zero and applies an improper inverse-gamma prior $IG(1, 1)$ on σ^2 that has a mode at $\frac{1}{2}$.

2. A data-driven prior specified in the spirit of the prior used by Fraley and Raftery (2007):

   - $\nu = 3$,
   - $s^2 = 2var(\bar{v}_t)$,
   - $\mu = \frac{\sum_{t=1}^{T} \bar{v}_t}{2T}$,
   - $\kappa = \frac{1}{100}$.

   This prior uses the information that the realized threshold in election t is in $[0, \bar{v}_t]$ to set $var(\bar{v}) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\bar{v}_t^2}{4} + (\frac{\bar{v}_t}{2})^2 \right) - \left(\sum_{t=1}^{T} \frac{\bar{v}_t}{2T} \right)^2$, and the mean at the mean of the midpoints of the logical interval of the realized threshold. The calculated mean of the prior on σ^2 is s^2, and it is equal to the variance of a random variable that is drawn from the 'non-informative' $Beta(\frac{1}{2}, \frac{1}{2})$ in $[0, \bar{v}_t]$ with probability $\frac{1}{T}$ for each $t = 1, \ldots, T$.

3. A data-independent prior that leaves the prior on σ^2 the same $IG(1, 1)$ but shifts the prior mean of the latent threshold mean θ from $\mu = 0$ to $\mu = 5$:

    - $\nu = 2$,
    - $s^2 = 2$,
    - $\mu = 5$,
    - $\kappa = \frac{1}{100}$.

4. A data-independent prior that leaves the prior on θ the same as in our default choice, but assumes a proper inverse gamma $IG(1.5, 0.05)$ for σ^2. This prior is much more tightly concentrated near zero with a mode at 0.02 and a mean at 0.1:

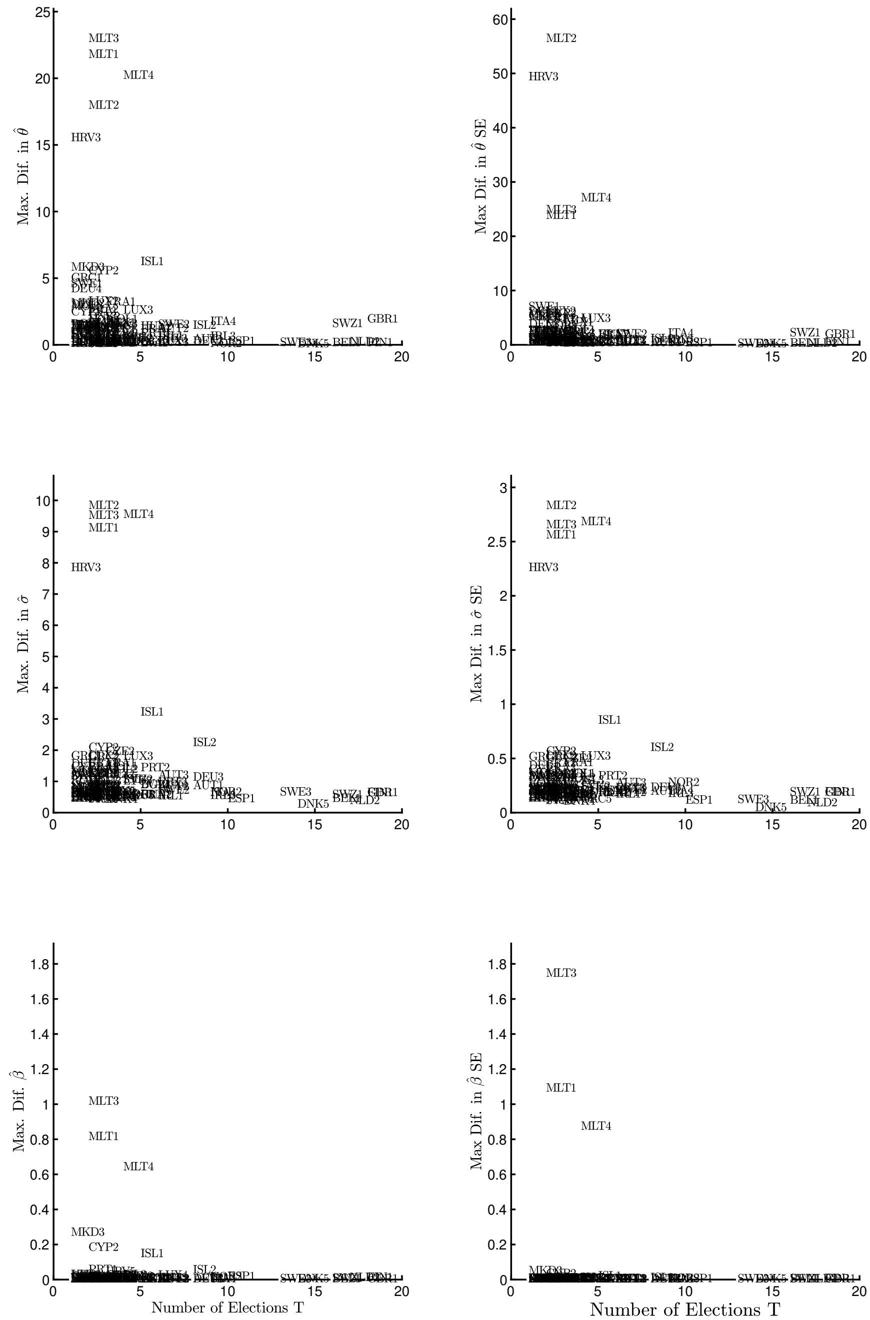
    - $\nu = 3$,
    - $s^2 = 0.1$,
    - $\mu = 0$,
    - $\kappa = \frac{1}{100}$.

For each electoral system, we run MAP-EM for each of these four prior specifications and then compute standard errors using the procedure of Online Appendices C and E as applicable. We then compute the maximum absolute difference in these quantities (parameters and standard errors). This information is then plotted against two measures of the amount of data we have available for each system:

- The number of elections, with the results displayed in Figures 1 and 2;
- The logarithm of the number of parties receiving positive vote shares across the elections in the system $\left(\log\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{I}(v_{i,t} > 0)\right)\right)$ in Figures 3 and 4.

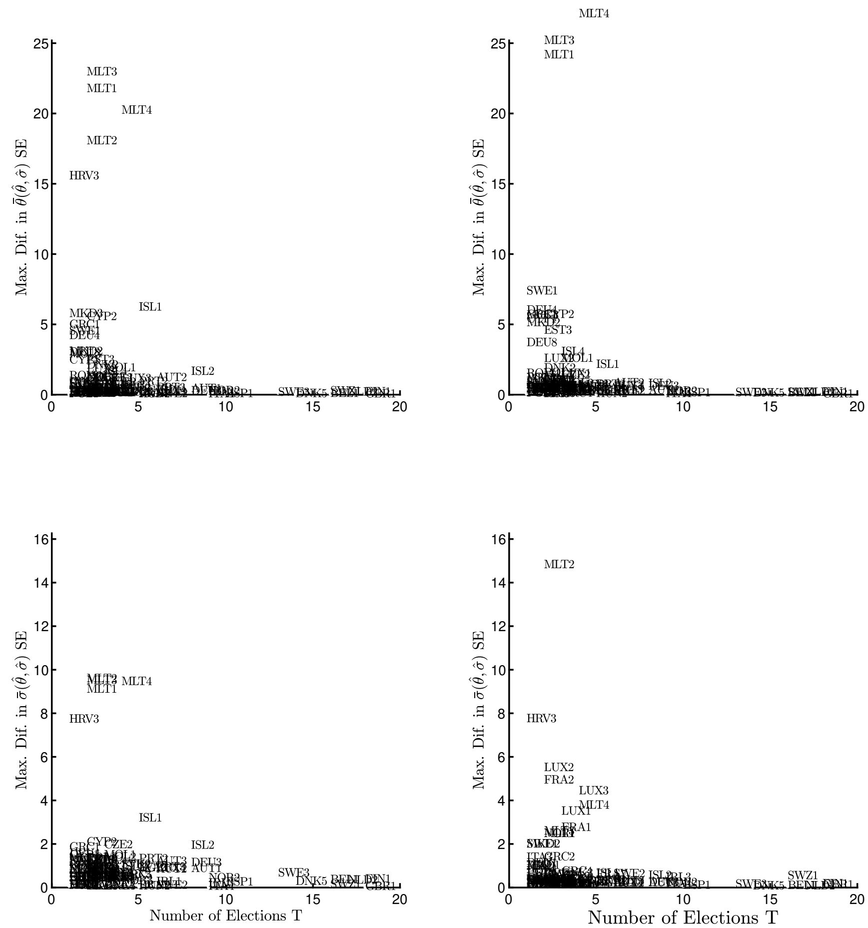
These figures exhibit the behavior we would expect: The estimated parameters and standard errors are quite sensitive to the prior for electoral systems for which the amount of data is small (using either measure for the amount of data). In all cases, the proportionality parameters are the most stable, while the estimated standard deviation of the latent threshold $\hat{\sigma}$ shows the least robustness, with differences in point estimates that are away from zero (but decreasing) with the amount of data. But despite these deviations, the expected threshold, $\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$, is more robust, in part because the differences reflect several cases with negative estimated $\hat{\theta}$, for which the bulk of the mass of the distribution of the latent threshold is below zero. Systems that show particular volatility (for example, the Maltese systems MLT1-4, the Croatian system HRV3) have both a small number of elections and a small number of parties receiving positive votes as can be gleaned by the combined information in Figures 1-4. But also note that many of these systems that exhibit large volatility in their estimated parameters also have large estimated standard errors under our default MAP-EM specification as reported in Online Appendix I, so that the MAP-EM estimator provides a natural warning to place less confidence on these estimates.

Figure 1: MAP-EM Sensitivity to Prior (Parameters & Standard Errors)



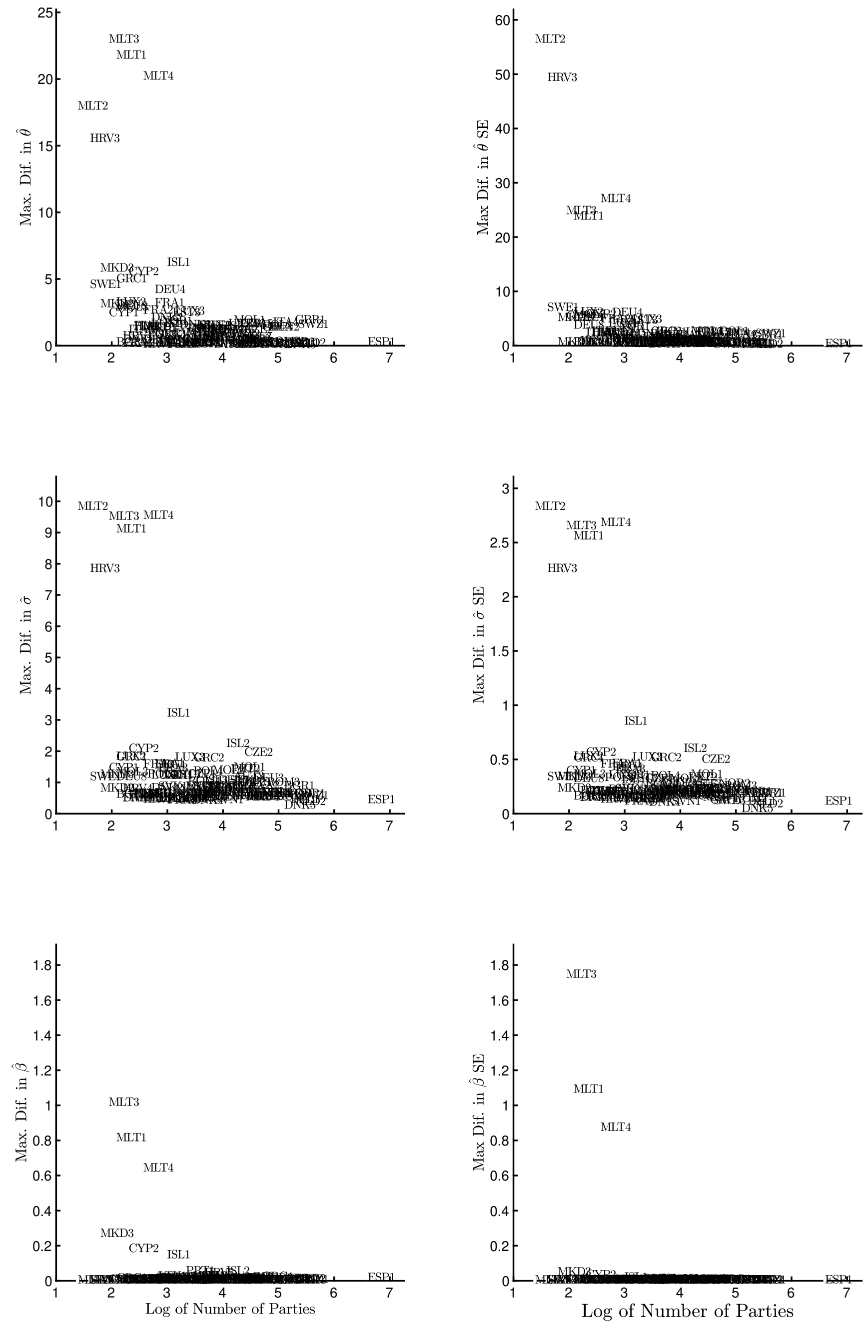
Scatter plots in the left present the difference between the maximum estimate and the minimum estimate obtained with MAP-EM with the four alternative priors (maximum difference) as a function of number of elections in each system being estimated. Scatter plots in the right present the difference between the maximum estimate and the minimum estimate of the standard error of each parameter obtained by applying the procedure in Online Appendix C on the MAP-EM estimates for each of the four alternative priors (maximum difference) as a function of number of elections in each system being estimated.

Figure 2: MAP-EM Sensitivity to Prior ($\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$ and $\bar{\sigma}(\widehat{\theta}, \widehat{\sigma})$ & their Standard Errors)



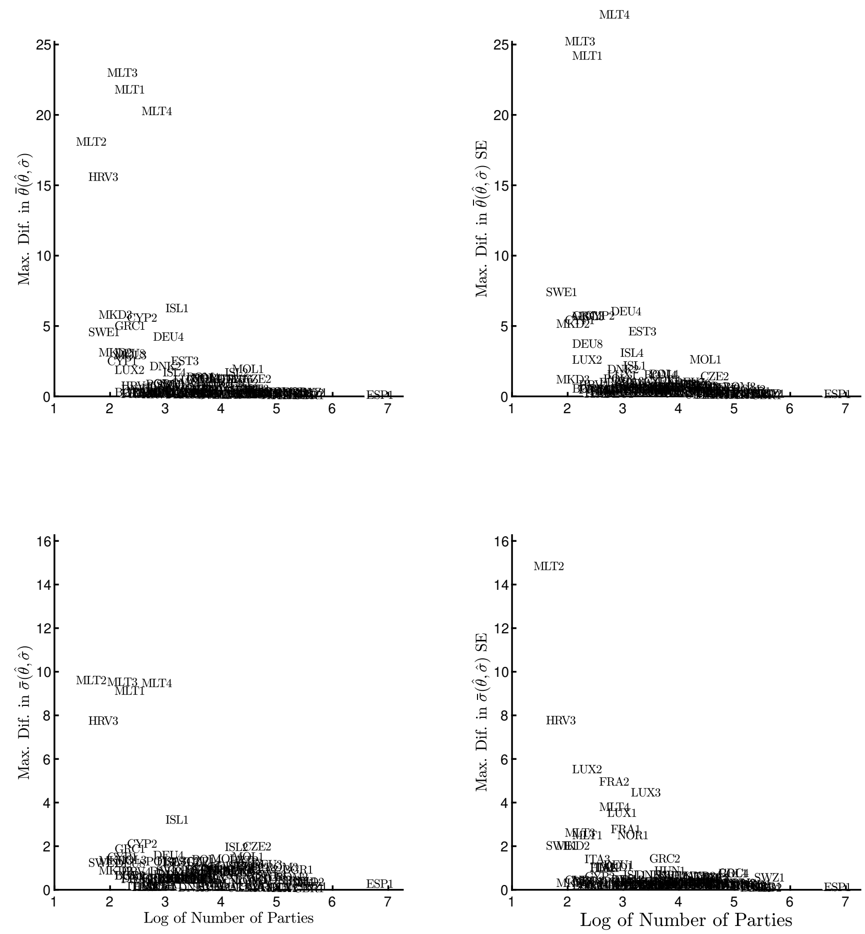
Scatter plots in the left present the difference between the maximum estimate and the minimum estimate obtained with MAP-EM with the four alternative priors (maximum difference) as a function of number of elections in each system being estimated. Scatter plots in the right present the difference between the maximum estimate and the minimum estimate of the standard error of each parameter obtained by applying the procedure in Online Appendix E on the MAP-EM estimates for each of the four alternative priors (maximum difference) as a function of number of elections in each system being estimated.

Figure 3: MAP-EM Sensitivity to Prior (Parameters & Standard Errors)



Displays the same information as in Figure 1 as a function of the logarithm of the sum of the number of parties receiving positive vote shares in each election for each system being estimated.

Figure 4: MAP-EM Sensitivity to Prior ($\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$ and $\bar{\sigma}(\widehat{\theta}, \widehat{\sigma})$ & their Standard Errors)



Displays the same information as in Figure 2 as a function of the logarithm of the sum of the number of parties receiving positive vote shares in each election for each system being estimated.

We now explore the differences between the MAP-EM estimates with the default prior specification used in the main body of the paper ($\nu = s^2 = 2, \kappa = \frac{1}{100}, \mu = 0$) and estimates obtained from application of the EM algorithm without a prior. As we have already discussed in motivating the MAP-EM approach, the pure EM-algorithm is plagued by numerical instability and slow convergence and to obtain the EM estimates for this analysis we had to reduce the convergence criterion, now requiring that successive parameter iterates are within 10^{-7} instead of 10^{-9} of each other. With that relaxed convergence criterion we obtain convergence for all 116 systems (though not for all 27 initial values we initiate the algorithm for each system).[2] We encountered many instances in which our analytical standard errors failed to compute despite convergence. Probing of the eigenvalues of the resulting Hessian from our approximation reveals many cases when the maximum eigenvalue is nearly zero and even positive (it ought to be negative for a negative definite Hessian). Even when the eigenvalue is negative, many standard errors appear implausible. We therefore focus the comparison on the point estimates $\widehat{\theta}, \widehat{\sigma}, \widehat{\beta}$, and $\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$ and $\bar{\sigma}(\widehat{\theta}, \widehat{\sigma})$.

In Figure 5 we plot the MAP-EM estimates of these quantities against the EM point estimates. For most of the systems considered, the estimates from these two approaches are very close when it comes to the proportionality parameter $\widehat{\beta}$, but also for expected latent and realized thresholds $\widehat{\theta}$ and $\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$. Not surprisingly, the biggest differences appear in the case of the nuisance parameters $\widehat{\sigma}$ and $\bar{\sigma}(\widehat{\theta}, \widehat{\sigma})$. This is not surprising, as we have already seen that these parameters are the most sensitive the prior. Furthermore, it is also evident that there is a cluster of systems for which the differences in these estimates are noticeable. These are largely the same set of systems for which we observed noticeable sensitivity to the prior due to a small number of elections and fewer parties.

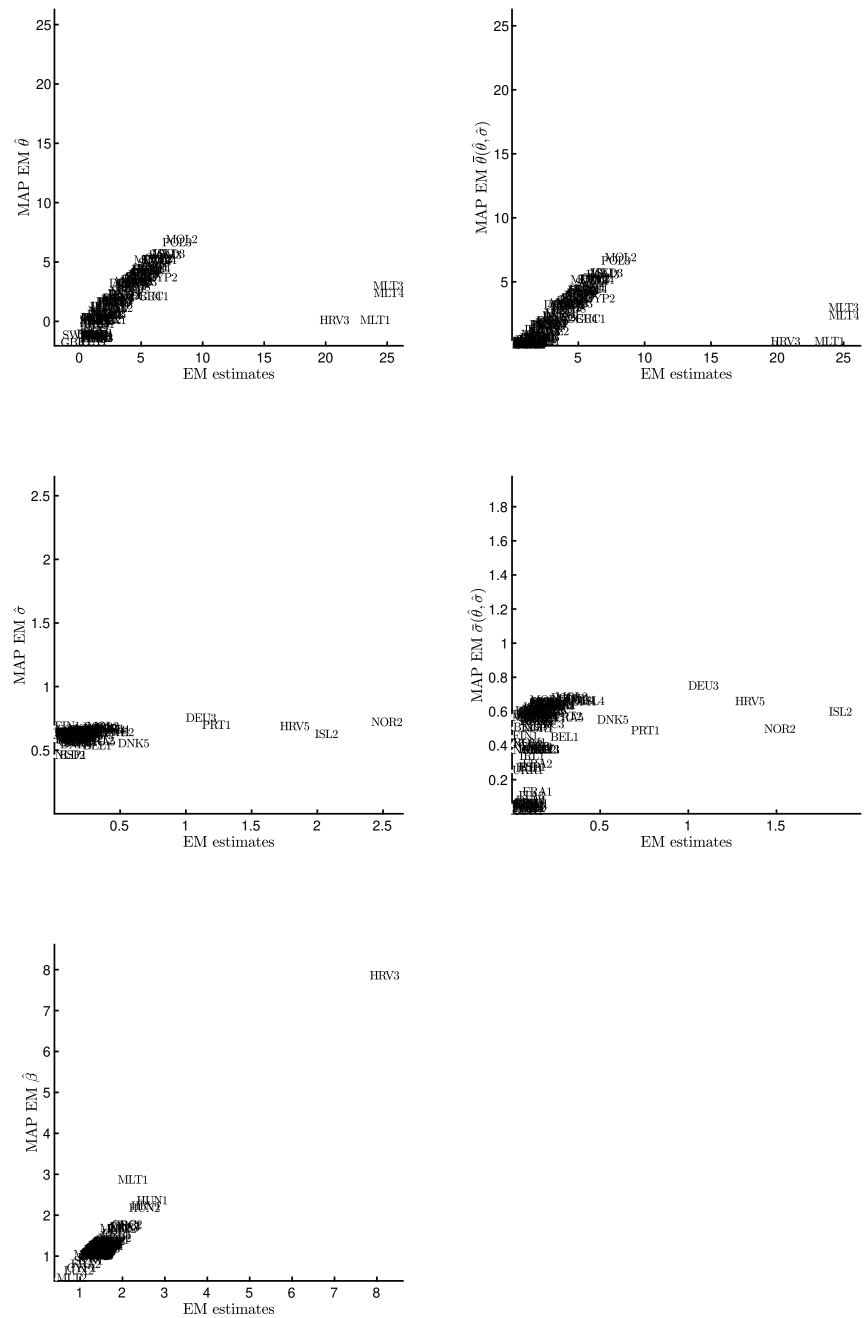
Figure 6 presents the same comparison for systems for which the Hessian is negative

[2]The initial values and all other aspects of the algorithms are the same except for the prior and the convergence criterion.

definite at the estimated parameters (76 systems). These are cases in which we can place bigger confidence that the EM algorithm has converged to the right mode of the likelihood.[3] The relationships show that the MAP-EM and the pure EM estimates of expected thresholds $\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$ and disproportionately, $\widehat{\beta}$, which are substantively the focus of our study, follow each other much more closely in this subset. In the case of $\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$ this is the case despite the differences in $\widehat{\sigma}$. As can be gleaned from the first panel on the left, $\widehat{\theta}$ takes smaller values below zero under MAP-EM to produce the same near zero expected threshold $\bar{\theta}(\widehat{\theta}, \widehat{\sigma})$ as EM when the prior forces the value of $\widehat{\sigma}$ to larger values than at EM.

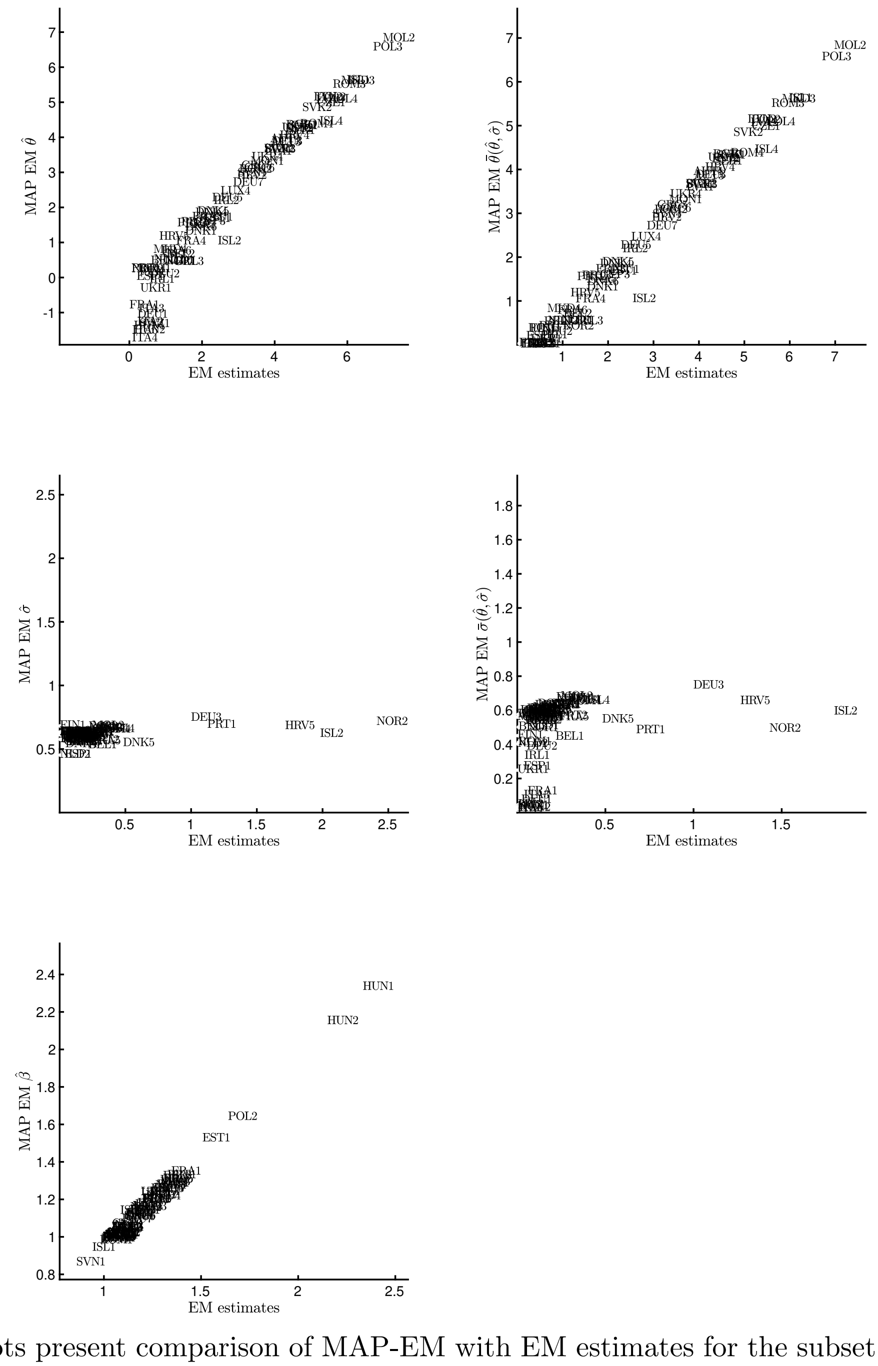
[3]Other cases may have too. The sign of the maximum eigenvalue is numerically sensitive in the analytical Hessian computations when the likelihood is very flat.

Figure 5: MAP-EM vs. EM Estimates



Scatter plots present comparison of MAP-EM with EM estimates.

Figure 6: MAP-EM vs. EM Estimates (EM Hessian Negative Definite)



Scatter plots present comparison of MAP-EM with EM estimates for the subset of systems for which the computed Hessian at the EM convergence point is negative definite.

L Alternative cutoff rules for electoral system determination

In the following table we summarize the number of electoral systems by country that result from application of different rules for determining changes in electoral institutions. The column labeled "Coarse" shows the number of systems that ignore numerical changes in the number of allocated seats, districts, or legal thresholds as long as the allocation formula (including upper tier allocation rules) are stable. "Finest" records a change whenever there is a change in any of the observed quantities, allocation formulas, number of seat, districts, etc. The intermediate columns require either a change in allocation formulas (including upper tier rules) or apply a percentage cutoff rule in the numerical dimensions (number of seats, number of districts, legal thresholds), as described in Appendix M. In 32 out of 36 countries, the number of identified electoral institutions does not change when we move from a 5% to a 7.5% cutoff rule. Similarly, the number of electoral institutions does not change in 27 out of 36 countries when we switch from 5% to a 2.5% cutoff rule. The level of the cutoff makes no difference in 24 out of 36 countries.

Table 3: Number of electoral systems

Country	Coarse	7.5% Rule	5% Rule	2.5% Rule	Finest
AUT	3	3	3	3	3
BEL	3	3	3	3	4
BGR	2	2	2	2	2
CYP	2	3	3	3	3
CZE	2	2	2	2	2
DEU	6	7	8	8	9
DNK	6	6	6	6	6
ESP	1	1	1	1	1
EST	2	3	3	3	3
FIN	1	1	1	1	1
FRA	4	7	7	8	10
GBR	1	1	1	2	8
GRC	6	6	6	6	6
HRV	3	5	5	5	5
HUN	1	2	2	2	2
IRL	1	2	3	5	7
ISL	3	4	4	4	4
ITA	6	6	6	7	7

LTU	1	2	2	2	2
LUX	1	3	4	6	7
LVA	1	2	2	2	2
MKD	4	4	4	4	4
MLT	3	4	4	5	5
MOL	1	3	3	3	5
MON	1	1	1	1	1
NLD	1	2	2	2	2
NOR	3	4	4	5	7
POL	4	4	4	4	4
PRT	1	2	3	4	4
ROM	2	4	4	4	6
SVN	2	2	2	2	2
SER	1	1	1	1	1
SVK	2	2	2	2	3
SWE	3	3	3	3	9
SWZ	1	1	1	2	4
UKR	3	4	4	4	4

.

M Electoral Results and Institutions Data Codebook

The data include information on votes and seats of all parties or coalitions of parties contesting elections, as well as the institutional characteristics that govern the allocation of seats for a given election. Each country data file contains up to five worksheets: *Parties*, *Coalitions*, *Votes*, *Seats*, and *Institutions*.

M.1 Electoral Data

These data are recorded in worksheets: *Parties*, *Coalitions*, *Votes*, *Seats*. The worksheet *Parties* includes the following variables:

PARTY or COALITION NUMBER: Number that identifies a party or coalition in the data set. Includes parties that did not contest elections on their own but in an electoral coalition including other parties. Coalitions are listed at the bottom of the worksheet.

ENGLISH NAME: Name of the party in English. If not available from original sources, name corresponds to the literal translation of the native name.

NATIVE NAME: Native name.

The *Coalitions* worksheets detail the composition of electoral coalitions (if present).

The *Votes* and *Seats* worksheets contain votes and seats (respectively) of parties or coalitions (corresponding to rows) per election (corresponding to columns).

M.2 Primary Institutional Variables

The *Institutions* worksheet records electoral system information per election (corresponding to columns as in the *Seats* and *Votes* worksheets). Each row corresponds to a variable, starting with election date information and continuing with institutional variables (either primary or derivative) which we detail below.

SYSTEM: Electoral systems per country are enumerated starting with 1 for the first system encountered in the data. A system persists if

- There are no changes in the allocation formula.[4]
- No other primary institutional variable (number of districts, number of seats, or numerical thresholds) changes by more than 5% of the system's average (that is if $I_1, \dots, I_t$, and I_{t+1} reflect consecutive values, a change occurs at $t+1$ whenever $\frac{\max\{\frac{\sum_{t'=1}^{t+1} I_{t'}}{t+1}, I_{t+1}\}}{\min\{\frac{\sum_{t'=1}^{t+1} I_{t'}}{t+1}, I_{t+1}\}} > 1.05$).

COARSE SYSTEM: Alternative (coarser) definition of electoral system enumerated starting with 1 for first system encountered in the data. A system persists if

- There are no changes in the allocation rules, even if changes occur in other coded variables (numerical thresholds, number of districts, or number of seats).

FINEST SYSTEM: Alternative (finer) definition of electoral system enumerated starting with 1 for first system encountered in the data. A system persists if

- There are no changes in the allocation formula.
- There are no changes in numerical thresholds, number of districts, or number of seats.

[4] There is one exception in the data: A nominal change in allocation formula for the highest tier in the third Swedish system from Modified Saint Lague to Saint Lague is ignored as the divisor sequence that is actually in use at this tier is identical between the two versions of these two allocation rules.

TOTAL SEATS: Total number of seats to be allocated, absent provisions for seats to be added (to ensure proportionality or legislative majority) or that remain unallocated (due to failure of turnout or threshold provisions).

MAJORITY/PLURALITY BONUS: Binary variable. Takes the value 1 if national tier allocation reserves extra seats for the party with a majority or plurality of votes.

M.2.1 First tier variables

NSEATS: Maximum number of seats out of TOTAL SEATS potentially allocated in the first tier.

NDISTRICTS: Number of districts in first tier.

NTHRES: National vote share that must be exceeded to take part in first tier allocation in any district.

DTHRES: District vote share that must be exceeded to take part in first tier district allocation.

ALLOCATION FORMULA: Formula used for first tier allocation. Possible values are

- One of **Plurality, Majority runoff, Majority plurality**: Possible Majoritarian allocation formulas.
- **STV**: Single transferable vote method.
- One of **Hare, Droop, Hagenbach-Bischoff, Imperiali, $\frac{1}{n+3}$ quota**: Respective quota rule with remainders reserved for allocation in an upper tier.

- One of **Hare, Droop, Hagenbach-Bischoff, Imperiali, $\frac{1}{n+3}$ LR**: Respective quota rule with largest remainders method applied to allocate seats that cannot be allocated on the basis of the quota.

- One of **D'Hondt, Saint Lague, Modified Saint Lague, Modified D'Hondt**: Highest divisors methods with different divisor sequences.

- Hybrids: Combinations of the above formulas, e.g., **Hare LR/Majority**.

M.2.2 Tier k=$2, 3, 4$

Up to 3 additional tiers of allocation are coded. Higher tiers are assigned a value $k = 2, 3, 4$, with higher values reserved for coarser geographic partitions of the country (e.g., in a country with three increasingly coarser tiers of allocation, the first tier may involve allocation at the level of the district, tier $k = 2$ at the level of the state or region, and tier $k = 3$ at the level of the country). If allocation occurs at the same level of aggregation but with two different modes or rules, these tiers are assigned consecutive levels in arbitrary fashion. For example, the first Greek electoral system (GRC1) involves district allocation (tier 1), regional allocation of remainders from tier 1 (tier 2), national allocation of remainders from tier 2 (tier 3), and national allocation of reserved seats in an "at-large" national district (tier 4), so that tiers 3 and 4 apply to the exact same geographic unit.

For each of the possible tiers $k = 2, 3, 4$, the following variables are coded:

TIERk: Binary variable; takes the value 1 if there exists a k-th tier, 0 otherwise.

NDISTRICTS: Number of districts in tier k.

ALLOCATION FORMULA: Formula used for tier allocation. Takes the same possible values as in the first tier formula.

FIXED SEATS: Binary variable. Takes the value 1 if a predetermined number of seats are reserved for allocation in this tier, 0 otherwise.

NFSEATS: Predetermined number of seats to be allocated in this tier (if FIXED SEATS = 1).

NATIONAL ACCESS THRESHOLD: National share of vote that is sufficient to take part in this tier's allocation.

TIER 1 DISTRICT ACCESS CONDITION: Lowest tier district vote share condition that is sufficient for access to this tier's allocation.

OTHER ACCESS CONDITION: Other condition that is sufficient for access to this tier's allocation.

N Electoral Results Sources

We compile electoral returns data using information from online official sources or databases of electoral results. We also relied on printed sources, notably Nohlen and Stöver (2010) and Mackie and Rose (1991), for older elections. If there were discrepancies between official country sources and the data from these books, we use the information from official sources. The following is the list of online sources.

N.1 Several Countries

Election Resources on the Internet by Manuel Alvarez Rivera. http://electionresources.org/

European Election Database. Norgewian Social Science Data Services. http://www.nsd.uib.no/european_election_database/

PARLINE Database on National Parliaments. http://www.ipu.org/parline-e/parlinesearch.asp

Psephos Adam Carr's Election Archive. http://psephos.adam-carr.net/

N.2 Country-specific

Belgium. Federal Portal

2007 Election Results. http://polling2007.belgium.be/en/

2010 Election Results. http://polling2010.belgium.be/en/

Bulgaria. Electoral Commission of Bulgaria. http://rezultati.cik2009.bg/results/proportional/rik_00.html

Czech Republic. Czech Statistical Office. http://www.volby.cz/pls/ps2010/ps2?xjazyk=EN

Croatia. State Electoral Commission of the Republic of Croatia. http://www.izbori.hr/izbori/ip.nsf/wpds/A51BF17BD1E3B7D5C125742000368C5D?open&1

Cyprus. Parliament of Cyprus. http://http://www.parliament.cy/

France. Ministry of Interior. http://www.interieur.gouv.fr/Elections/Les-resultats/Legislatives/

Germany. The Federal Returning Officer. http://www.bundeswahlleiter.de/en/bundestagswahlen/

Italy. Ministry of Interior. http://elezionistorico.interno.it/index.php

Macedonia. State Election Commission http://http://217.16.84.11/Default.aspx

Moldova. E-Democracy. http://www.e-democracy.md/elections/parliamentary/

Portugal. National Electoral Commission. http://eleicoes.cne.pt/index.html

Romania. Romanian Electoral Commission. http://www.becparlamentare2008.ro/

Spain. Ministry of Interior. http://www.infoelectoral.interior.es/min/

O Data Issues – Country Notes

In cases when countries have adopted mixed electoral systems with two distinct partitions of the electorate into districts, two separate ballots cast in each election, one for each partition, and a separate (not necessarily independent) allocation of seats within each partition, we treat the mixed system as one electoral system using the PR vote as an input in the German and Hungarian cases in which allocations across the two partitions are not independent.[5] In all other cases, we estimate separate electoral system parameters for each partition, provided that within that partition the allocation of seats takes the ballot in that partition as the sole input, independent of the vote outcome in the other partition.

When parties or candidates lumped in the 'others' category earn seats, we amend the definition of the upper bound on the realized threshold for election t in (3) to the more conservative

$$\bar{v}_t := \min_{i \in \{0,1,\ldots,N\}} \{v_{t,i} \mid s_{t,i} > 0\}, \tag{10}$$

where $v_{t,0} := 1 - \sum_{i=1}^{n} v_{t,i}$ and $s_{t,0} := S_t - \sum_{i=1}^{n} s_{t,i}$, the vote share and seat number, respectively, of parties or candidates not separately reported in electoral returns. A number of electoral systems make special provisions to ensure the representation of minorities, or special overseas districts (e.g., Finland, Italy, Croatia, Romania, Slovenia). In these cases, we faced the choice whether or not to include these seats and corresponding votes (if separately reported) as part of the estimation. Because these seats are typically allocated using

[5] A third case of mixed system with a similar dependency structure between the two partitions is the Italian 'scorporo' system. We only estimate the majoritarian partition of that system (ITA5) primarily due to data difficulties identifying parties competing across the two partitions.

special provisions, their inclusion in the analysis can lead to erroneous conclusions about the nature of competition induced by the electoral system. This is especially the case when very small minorities are guaranteed representation, thus forcing an inference that the electoral threshold in effect for all parties is small. We have made efforts to exclude electoral returns data from special or singular districts reserved for minority or overseas seats for which special electoral provisions apply. However, ambiguities regarding these exclusions remain in some cases, primarily due to unavailability of disaggregated results by district or inconsistencies between different sources.

The following list states exclusions regarding special districts, the treatment of mixed electoral systems, and any residual ambiguities.

Bulgaria Data include only proportional representation seats in 2009.

Croatia Data exclude minorities and diaspora results. Independent components of mixed systems in 1992 and 1995 are treated as separate electoral systems.

Germany Data exclude representatives from Berlin before unification. Data include proportional representation votes and total seats (i.e., including SMD seats).

Denmark Data exclude representatives from Faroe islands and Greenland.

Finland Data include district of Aland results.

France Results for Metropolitan France up to 1988. Results including overseas territories starting with the 1993 election.

Hungary Data include proportional representation votes and total seats (i.e., including SMD seats).

Italy Data includes Valle D'Aosta results. For the years 1996 and 2001 in which a mixed electoral system is used, the data include only the single member district results. Overseas deputies are excluded.

Lithuania Single member district component of mixed systems is excluded in all elections.

Macedonia Overseas representatives are excluded. The mixed system in place in 1998 is treated as two separate electoral systems.

Montenegro Ethnic minorities' representatives are excluded.

Poland Minorities' representatives are excluded from seat allocation.

Portugal Representatives from Europe and the rest of the World are included.

Romania Overseas results are included for the fourth Romanian system. Minorities' seats are excluded in systems ROM2-ROM4.

Serbia Minority seats are excluded.

Slovenia Minority results are excluded.

Ukraine Mixed system components are treated as separate systems.

References

Dempster, A. P., N. M. Laird and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1):1–38.

Fraley, Chris and Adrian Raftery. 2007. "Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering." *Journal of Classification* 24:155–181.

Gelman, Andrew, John B. Carlin, Hal. S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2 ed. Boca Raton, Florida: Chapman & Hall/CRC.

Jamshidian, Mortaza and Robert I. Jennrich. 2000. "Standard Errors for EM Estimation." *Journal of Royal Statistical Society* 62(2):257–260.

Mackie, Thomas and Richard Rose. 1991. *The International Almanac of Electoral History*. London: Macmillan.

Miranda, Mario J. and Paul L. Fackler. 2002. *Applied Computational Economics and Finance*. Cambridge: MIT Press.

Nohlen, Dieter and Philip Stöver. 2010. *Elections in Europe: A Data Handbook*. Berlin: Nomos Verlagsgesellschaft.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin and Angelika van der Linde. 2002. "Bayesian measures of model complexity and fit (with discussion)." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 64(4):583–639.