A Summary of the Gelman-King Method Used in JudgeIt

Andrew C. Thomas[*]

August 10, 2008

Abstract

The JudgeIt software package uses a model constructed by Andrew Gelman and Gary King in a series of papers published in the early 1990s to investigate properties of two-party electoral systems, perhaps over a series of elections. Here I summarize the technique and method used in each calculation within the package.

Sources of Error in Electoral Data

In principle, the source of error in an experimental outcome can be divided into systematic error, in which the apparatus for conducting an experiment skews the result in the same way each time it is run, and random error, which is different every time an experiment is conducted. In many cases in political science, the estimated error in an experiment cannot be divided into its two component types.

In the application of two-party elections, however, the "experimental apparatus" to be considered is an electoral system that changes minimally between elections in its construction. This gives us a method for estimating what fraction of an election's variation is caused by properties of the system, and what changes can be considered to be random in nature.

Under this assumption, two different types of elections can be simulated: those in which the existing system is no longer in place, such as the prediction of future elections, and those in which we consider what would happen whether a particular election was run again using the same systematic conditions, two fundamentally different sets of analyses with a common method of generation.

Quantities of interest in the political science literature can often be explained in terms of the outcomes of hypothetical elections or properties of the election system.

---

[*]acthomas@fas.harvard.edu

The Gelman-King Model

JudgeIt is used to model two-party electoral systems; choose one party to identify as "Party 1". (All results for Party 2 are clearly the opposite for those of Party 1.)

In any particular election year, let v_i be the share of the two-party vote received by Party 1 in district i. We model the resulting vote share as

$$v_i = X_i'\beta + \gamma_i + \varepsilon_i$$

where x_i' is a vector of predictor variables with coefficient β, and $\gamma_i \sim N(0, \lambda\sigma^2)$ and $\varepsilon_i \sim N(0, (1-\lambda)\sigma^2)$ are the systematic and random error terms. In this presentation, σ^2 is the total error variance, and λ is the share attributed to the systematic error component. The error terms in each district are independent of each other and of those in each other district in the system.

The standard approach to estimating the unknown quantities is to model each year under the Bayesian framework, with noninformative priors on the β and σ^2 parameters in each year. To estimate σ^2, take the total variance estimate in each year and pool those estimates together; then use the mean of the pooled estimates as the value of σ^2 as the value for each election.

To estimate λ for the electoral system, note that the systematic component of the error is proportional to the votes received in each district in two subsequent elections, yielding [?]

$$v_{i,t+1} = \lambda v_{i,t} + X_i'\beta + \gamma_i + \varepsilon_i,$$

where X_i' may include as many variables as are available from elections t and $t+1$. The value of λ used is the mean of these estimated values. Note that estimates of λ can only be obtained when two subsequent elections use the same electoral map, i.e. no redistricting has taken place.

Generating Hypothetical Elections

Now that the parameters in each election year are estimated, the model can be simulated. For a predictive analysis, the hypothetical vote share in each district is found to be

$$\tilde{v}_i = \tilde{X}_i'\beta + \tilde{\delta} + \tilde{\gamma}_i + \tilde{\varepsilon}_i$$

where $\tilde{X}_i'$ is a (possibly new) vector of predictors corresponding to those terms in β. In this case, the two error terms are unidentifiable and once again combine so that

$$\tilde{\gamma}_i + \tilde{\varepsilon}_i \underset{2}{\sim} N(0, \sigma^2).$$

The $\tilde{\delta}$ term is added under the *general uniform partisan swing* assumption; that is, for small deviations from the observed outcome, a swing in the overall vote share can be represented as the same swing in each district in the system. This allows the user to investigate two scenarios: what would happen if the average vote were to shift by a small amount, or what the electoral map would look like with a particular average vote share (corresponding to a particular shift in the average vote.)

For an evaluation of an election's underlying properties, or to examine what would happen if we re-ran the election under counterfactual circumstances, we note that the systematic error component γ_i can be estimated using the data. Since y_i and γ_i form a bivariate normal distribution

$$\begin{bmatrix} y_i \\ \gamma_i \end{bmatrix} \sim N_2 \left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \lambda\sigma^2 \end{bmatrix} \right)$$

we can obtain the conditional distribution

$$\gamma_i | y_i \sim N(\lambda(y_i - x_i'\beta), \lambda(1-\lambda)\sigma^2).$$

We then use this estimate of γ_i in the simulation equation

$$\tilde{v}_i = \tilde{X}_i'\beta + \tilde{\delta} + \gamma_i + \tilde{\varepsilon}_i.$$

Note that Party 1 wins the election if their share of the two-party vote is greater than one-half. Given β and γ, we then see that the expected seat share is

$$P(v_i > 0.5 | \beta, \gamma, \delta) = 1 - \Phi \left(\frac{0.5 - \tilde{X}_i'\beta - \tilde{\delta} + \gamma_i}{\sqrt{(1-\lambda)\sigma^2}} \right).$$

To generate the probability distribution for this quantity, we then draw values for β and γ given their conditional distributions, and set δ to its required value given the application.

Application: Calculating Partisan Bias

One definition of partisan bias is the extra seat share received by one party when the average vote is split equally between both. We estimate the distribution of partisan bias conditional on β and γ, then use their distributions to determine a probability interval.

To get one draw from the distribution of partisan bias:

- Draw β and γ from their respective distributions. These represent conditions in the electoral system up until the election.

- Calculate the mean and variance of the vote share in each district conditional on the draws of β and γ.

- Calculate the grand mean vote (with whatever weights are being used) and subtract it from 0.5; this is the value of δ we use to adjust the mean vote to 0.5.

- Determine the expected seat share $P(v_i > 0.5 | \beta, \gamma, \delta)$ for each district and take the weighted mean. Twice the (weighted mean minus 0.5) is the partisan bias conditional on β and γ.

Repeating this procedure gives us the distribution of the partisan bias.