Scaling Votes

- Nominate, D-Nominate, W-Nominate
  (Poole and Rosenthal, 1991, "Patterns of Congressional Voting," *American Journal of Political Science*; 1996, *A Political-Economic History of Roll Call Voting*, Oxford University Press)
- quadratic utility, logistic errors
  (Ladha, 1991, "A Spatial Model of Legislative Voting with Perceptual Error, *Public Choice*)
- linear factor model
  (Heckman and Snyder, 1997, "Linear Factor Models of Demand for Attributes with an Application to Estimating the Preferences of Legislators," *Rand Journal of Economics*)
- add distributional assumptions on ideal points
  (Lewis, 2001, "Estimating Voter Preference Distributions from Individual-Level Voting Data, unpublished manuscript, UCLA)
- add a model of agenda process
  (Londregan, 2000, "Estimating Legislator's Preferred Points," *Political Analysis*)

Each voter has an ideal point, z_i. There are two alternatives – x_j^Y is the "Yes" alternative and x_j^N is the "No" alternative. A voter votes "Yes" iff

$$\begin{aligned} U(z_i, x_j^Y, \epsilon_{ij}^Y) &> U(z_i, x_j^N, \epsilon_{ij}^N) \\ V(z_i, x_j^Y) + \epsilon_{ij}^Y &> V(z_i, x_j^N) + \epsilon_{ij}^N \\ V(z_i, x_j^Y) - V(z_i, x_j^N) &> -\epsilon_{ij} \end{aligned}$$

where $\epsilon_{ij} = \epsilon_{ij}^Y - \epsilon_{ij}^N$. Letting G be the cdf of ϵ_{ij},

$$Pr(vote_{ij}{=}Yes) = 1 - G(V(z_i, x_j^N) - V(z_i, x_j^Y))$$

Nominate assumes that

$$V(z_i, x) = \theta \exp[-(z_i - x)'(z_i - x)/8]$$

Nominate also assumes that ϵ_{ij} has a logistic distribution. So,

$$Pr(vote_{ij}{=}Yes) = \frac{\exp(\theta \exp[-(z_i - x_j^Y)'(z_i - x_j^Y)/8])}{\exp(\theta \exp[-(z_i - x_j^Y)'(z_i - x_j^Y)/8]) + \exp(\theta \exp[-(z_i - x_j^N)'(z_i - x_j^N)/8])}$$

Other authors assume a simpler quadratic function for the deterministic utility function:

$$V(z_i, x) = -(z_i - x)'(z_i - x)$$

In this case,

$$\begin{aligned} V(z_i, x_j^Y) - V(z_i, x_j^N) &= (x_j^N)'(x_j^N) - (x_j^Y)'(x_j^Y) + 2(x_j^Y - x_j^N)' z_i \\ &= \alpha_j + {\beta_j}' z_i \end{aligned}$$

Ladha assumes in addition that ϵ_{ij} has a logistic distribution, and obtains

$$Pr(vote_{ij} = Yes) = \frac{\exp(a_j + b_j' z_i)}{1 + \exp(a_j + b_j' z_i)}$$

Heckman and Snyder assume in addition that ϵ_{ij} has a uniform distribution on the interval $[-M_j, M_j]$, and obtain

$$Pr(vote_{ij} = Yes) = 1/2 + [\alpha_j + {\beta_j}' z_i]/2M_j = a_j + b_j' z_i$$

This is a linear probability model. The vote equation can then be written as

$$v_{ij} = a_j + b_j' z_i + \eta_{ij}$$

which is a linear factor model.

Note that a large number of parameters must be estimated in any of these models, even if we are only interested in the voters' ideal points. For example, if there were 435 legislators and 1000 roll calls, then a two-dimensional Nominate model would require the estimation of $435 \times 2 + 1000 \times 4 + 1 = 4,871$ parameters. Ladha's model would require estimates of 2,870 parameters. The linear factor model would require the estimation of 870 parameters.

Principal Components and Factor Analysis

There are many different methods for estimating linear factor models, including principal components factor analysis, principal factors, iterated principal factors, maximum likelihood, and minimum-chi-square estimation.

Principal Components

Begin with principal components. (Note, it is fitting that Hotelling, who wrote the seminal paper on spatial competition in 1928, also wrote the seminal paper on principal components analysis in 1933.) Let $y = [y_1, ..., y_m]'$ be a random vector, with $\mathrm{E}(y) = 0$ and $\mathrm{Var}(y) = \mathrm{E}(yy') = \Sigma$. Principal components analysis asks the following questions:

- What linear combination $\theta_1' y$ has maximum variance?
- What linear combination $\theta_2' y$ with $\theta_2'\theta_1 = 0$ has maximum variance?
- What linear combination $\theta_3' y$ with $\theta_3'\theta_1 = 0$ and $\theta_3'\theta_2 = 0$ has maximum variance?
- And so on.

A normalization is necessary, so we stipulate that $\theta_i'\theta_i = 1$ for all i. For any θ, $\mathrm{Var}(\theta' y) = \theta'\Sigma\theta$, so the problem of finding θ_1 then becomes

$$\begin{aligned} &\max_\theta \ \theta'\Sigma\theta \quad \text{s.t.} \quad \theta'\theta = 1 \\ \text{or} \quad &\max_\theta \ \theta'\Sigma\theta - \lambda(\theta'\theta - 1) \end{aligned}$$

Differentiating, the first-order conditions are

$$\Sigma\theta - \lambda\theta = 0$$

This is the equation for finding the eigenvalues and eigenvectors of the matrix Σ. Thus, θ_1 must be an eigenvector of Σ. Which one? (Since Σ is a symmetric, positive-definite matrix, all of its eigenvalues are real and positive.) The quantity to be maximized is $\theta'\Sigma\theta = \theta'\lambda\theta = \lambda\theta'\theta = \lambda$, so θ_1 must be the eigenvector corresponding to the largest eigenvalue.

Now consider θ_2. This problem is

$$\begin{aligned} &\max_\theta \ \theta'\Sigma\theta \quad \text{s.t.} \quad \theta'\theta = 1 \quad \text{and} \quad \theta'\theta_1 = 0 \\ \text{or} \quad &\max_\theta \ \theta'\Sigma\theta - \lambda(\theta'\theta - 1) - \delta(\theta'\theta_1) \end{aligned}$$

The first-order conditions are

$$\Sigma\theta - \lambda\theta - \delta\theta_1 = 0$$

Premultiplying by θ_1 yields

$$\theta_1'\Sigma\theta - \lambda\theta_1'\theta - \delta\theta_1'\theta_1 = 0$$

Now Σ is symmetric, so $(\theta_1'\Sigma)' = \Sigma\theta_1 = \lambda\theta_1$, and thus $\theta_1'\Sigma = \lambda\theta_1'$. Also, $\theta_1'\theta_1 = 1$. So, the equation above reduces to $\lambda\theta_1'\theta - \lambda\theta_1'\theta - \delta = 0$; so, $\delta = 0$. This means the first-order conditions above are

$$\Sigma\theta - \lambda\theta = 0$$

so θ_2 must be an eigenvector of Σ. And, since θ_2 is chosen to maximize, $\theta'\Sigma\theta$, it must be the eigenvector corresponding to the second-largest eigenvalue. Continuing, θ_3, θ_4... are found in an analogous fashion.

In practice, of course, we do not know Σ. Instead, we have a sample of observations Y on y, and use this sample to form an estimate S of Σ. It is common in analyzing votes to first double-center the data – i.e., to subtract the row and column means. We then use the standard formula, $S = Y'Y/(n-1)$, where n is the number of observations in the sample. (Note, the data matrix is

$$Y = [Y_1\ Y_2\ \ldots\ Y_m] = \begin{pmatrix} Y_{11} & Y_{21} & \ldots Y_{m1} \\ \vdots & \vdots & \vdots \\ Y_{1n} & Y_{2n} & \ldots Y_{mn} \end{pmatrix}$$

where m is the number of voters and n is the number roll calls.)

The statistical properties of the estimates are straightforward: S is a consistent estimator of Σ under standard assumptions. And the eigenvalue-eigenvector decomposition is a continuous operation, so by Slutsky's theorem the eigenvectors of S are consistent estimators of the eigenvectors of Σ. Asymptotic normality also follows under standard assumptions.

Factor Analysis

Turn now to the factor-analytic problem. Normally, factor analysis assumes that the underlying variables are standardized, so the diagonal elements of Σ are all 1. The random vector

$y = [y_1, ..., y_m]'$ is assumed to satisfy

$$y = Lf + e$$

where L is an $m \times k$ matrix of factor loadings, f is a $k \times 1$ matrix of common factors, and e is an $m \times 1$ matrix of specific factors (f and e are composed of random variables, but L is not). Assume that $\mathrm{E}(f) = \mathrm{E}(e) = \mathrm{E}(y) = 0$; $\mathrm{E}(ee') = \Psi$ where Ψ is diagonal; $\mathrm{E}(fe') = 0$; and $\mathrm{E}(ff') = I_k$. Then

$$\mathrm{Var}(y) = \mathrm{E}(yy') = L[\mathrm{E}(ff')]L' + \mathrm{E}(ee') = LL' + \Psi$$
$$\text{or} \quad \Sigma = LL' + \Psi$$

We want to estimate L. However, there is an identification problem. Specifically, if (L, Ψ) is a solution to $\Sigma = LL' + \Psi$, then so is (LT, Ψ), where T is any orthogonal matrix. This follows because $(LT)(LT)' = LTT'L' = L(I_k)L' = LL'$. So, additional restrictions on L are necessary. For example, it is typical in maximum-likelihood factor analysis to impose the constraint that the matrix $L'\Psi L$ is diagonal.

If we knew Ψ, then we could recover L (up to a rotation) by applying principal components to $\Sigma - \Psi$. Of course, we know neither Σ nor Ψ. We can construct an estimate S for Σ using the sample Y, as discussed above. Getting Ψ is more difficult. This is usually called the problem of estimating "communalities" – i.e., the diagonal elements of $\Sigma - \Psi$. Let c_i denote the estimate used for the ith communality.

Principal components factor analysis proceeds by choosing $c_i = 1$ for all i (or, equivalently, choosing $\Psi = 0$). The estimated factor loadings are the k eigenvectors of S corresponding to the largest k eigenvectors.

The principal factors method proceeds by choosing one of the following: (i) $c_i =$ the maximum positive correlation between Y_i and the other $(m$–$1)$ Y's; (ii) $c_i =$ the average correlation between Y_i and the other Y's (presumed to be positive); (iii) $c_i =$ the square of the multiple correlation coefficient of Y_i with the other Y's. The estimated factor loadings are the k eigenvectors of S^* corresponding to the largest k eigenvectors, where S^* is the matrix obtained after replacing the diagonal elements of S with the c_i's.

The iterated principal factors method proceeds by starting with some initial c_i to obtain a principal factor solution, then using the sum of squares of the factor loadings from this solution to obtain new values of c_i, and repeating until successive estimates do not differ by much.

Maximum-likelihood estimation relies on the assumption that y has a multivariate normal distribution. Under this assumption, the large-sample properties of the estimates have been characterized in considerable detail.

Most recently, Cragg and Donald (1995, "Factor Analysis Under More General Unknown Form", in *Advances in Econometrics and Quantitative Economics*, Basil Blackwell) have formulated the factor analytic problem as a minimum distance estimation problem and obtained important results.

There are many methods for rotating the set of factors in order to obtain "simple" structures. I don't know what to say about these.

Finally, one crucial issue is determining the appropriate number of factors. There are several possibilities: (i) look for the place where the scree plot (plot of the eigenvalues) levels off; (ii) retain the eigenvectors corresponding to eigenvalues greater than 1; (iii) for maximum-likelihood factor analysis, use a statistic derived by Lawley and improved by Bartlett, which has an approximate chi-square distribution and provides a rigorous test; (iv) use one of the statistics derived in Cragg and Donald (one based on AIC, one based on BIC, and a sequential pretest estimator).

Other Issues

In all of the methods above identification relies crucially on two assumptions: (i) voting is inherently probabilistic – i.e., there is a systematic spatial component and a non-trivial idiosyncratic component; and (ii) we have the correct functional form for the utility function and the distribution of idiosyncratic shocks.

This is easily seen by considering what would happen in a world of "perfect" one-dimensional spatial voting. Clearly, in such a case we can only estimate the *ordering* of voters on a line – or, more precisely, we can only estimate who lies between whom. We cannot recover any information about the distances between voters.

On the plus side: If (i) and (ii) above are true, and we have a large number of roll calls (e.g., 500, or maybe even just 200), then the distribution of "cut-lines" underlying the roll calls does not matter much. It could even be the same cut-line over and over again. We would still obtain good estimates of voters' ideal points. This is because the fraction of votes in favor of the "Left" alternative in a voter's voting record will be monotonically decreasing as the voter's ideal point moves from left to right.

Things To Do

- Distinguish between coalition-formation (partisan or otherwise) and actual preferences.
  - Consider the behavior of those who survive from one congress to the next, but where, say, party control switches. Can we detect new coalitions and agendas?
  - Alternatively, consider the behavior of continuing members pre- and post- Westbury v. Sanders, where court-ordered reapportionment forced the replacement of many rural representatives with suburban and urban representatives. Can we detect new coalitions and agendas?
- Incorporate observables, such as constituency socioeconomic characteristics or partisanship, into the scaling models.
- Distinguish between interest group activities (lobbying, campaign contributions) and actual preferences.
- Distinguish between constituency preferences and personal preferences.
- Compare the issues that map nicely into a one- or two-dimensional spatial model with those that do not. Does constituency partisanship matter more in the former? Do we detect a greater role for interest group activity and/or particular constituency characteristics in the latter?
- Distinguish between changes in preferences and changes in the agenda.
- Use knowledge about the content of bills or the context to help estimate parameters. Consider, for example, a series of roll calls on different amendments to the same bill, each of which loses. In all cases the "No" alternative (the bill) might be treated as the same point in the space.