"A Moment-Matching Approach to Maximum Likelihood Estimation of the Beta Distribution"

Philip Paolino

University of Texas at Austin

The beta distribution is a flexible distribution that can produce a uniform, unimodal, or bimodal distribution of points that can be either symmetric or skewed. Substantively, the beta distribution has been applied recently to examining individuals' subjective beliefs about the probability of the occurrence of some outcome, such as a candidate's chances of winning an election (e.g. Paolino 1998), and estimating the location of a party's ideal point (e.g. Mebane 1998). In many instances, researchers may be interested not just in the expected value of the function, but also in the variance of a model using this probability distribution. Modelling the variance of a distribution can be used, for example, to estimate factors influencing an individual's certainty about candidate ideal points (e.g. Franklin 1991) or the subjective probability of a candidate's chances of winning an election (e.g. Paolino 1998)

Unlike other distributions, such as the normal or the Poisson, the beta distribution is expressed with shape parameters that do not correspond directly to either the mean or variance of the distribution. Rather, the mean and variance of a (standard), two-parameter beta distribution is a function of the two parameters.

$$f(y|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \tag{1}$$

where $0 < y < 1$, $\alpha, \beta > 0$,

$$E(y) = \frac{\alpha}{\alpha+\beta} \tag{2}$$

and

$$Var(y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{3}$$

Consequently, it is not obvious how researchers using a beta distribution should estimate the substantive and statistical significance of some independent variable upon the mean and variance of some dependent variable.

One approach for researchers modelling some phenomenon according to the beta distribution is to specify the two shape parameters in terms of some set of independent variables, estimate the effects of those variables upon the non-substantive parameters, and then calculate the effects of those variables upon the dependent variable using first differences. Such a means of estimation, however, does not provide a straight-forward way of testing for the statistical significance of an independent variable upon either the mean or the variance of the dependent variable. Using a Wald test, one can obtain some idea about the statistical significance of a variable by determining whether or not some substantively meaningful first difference is statistically significant.[1]

Such a means of estimation, however, still requires the researcher to place all independent variables into the model for both α and β — and, by extension, the mean and variance — even if the researcher has reason to believe that a variable affects only the mean or the variance of the dependent variable. At a minimum, this means a loss of degrees of freedom that could be particularly damaging in small samples. In some cases, this will not pose much of a problem because the lack of independence between the mean and variance of the beta distribution means that a variable that has a significant effect upon the mean may also have an effect upon the variance and vice versa.[2]

Consequently, one might desire a way of estimating the distribution where the mean and variance of a dependent variable could be specified directly as the function of some set of, possibly non-identical, sets of independent variables. One approach would be to create a functional form for the mean and variance and solve those forms in terms of the two beta parameters. Such a "moment-matching" means of estimation requires that the

[1] Thanks to Walter Mebane for this suggestion.

[2] Below, I will present empirical results where this is not the case.

researcher can properly specify functional forms for the mean and variance. For the mean, a reasonable form is one that is bounded by 0 and 1. For the variance, a reasonable form has to be bounded by 0 and .25, the maximum variance for the beta distributed variable. But such a form may be inaccurate if the data generating process is not bimodal, in which case the maximum variance is less than 1/12. Because most problems that we deal with in political science probably concern variables that are distributed unimodally, I will focus upon these cases and ignore bimodally-distributed variables for now.

In this paper, I will first develop the moment-matching approach. Second, I will use Monte Carlo simulations to examine this approach, employing maximum likelihood estimation, to estimating models distributed according to a beta distribution. Third, I will present some empirical results using the moment-matching approach and compare these results with those obtained from the "standard" approach. Finally, I will present conclusions about the desirability of using the moment-matching approach to estimate models generated by a beta distribution.

1 A "Standard" Approach to Beta Estimation

To compare the proposed "moment-matching" approach with a standard approach, I will first examine how the effects of parameters upon the mean and variance are determined by this standard approach.[3] In the standard approach, α and β are specified as functions of a set of independent variables, $\mathbf{X}$. Since α and β must be positive, an obvious specification is (see Brehm and Gates 1993)

$$\alpha = \exp(\mathbf{X}\Gamma) \tag{4}$$

[3] Throughout this paper, I will describe the approach that directly models the shape parameters of the two-parameter beta distribution as the "standard" or "direct" approach, in comparison to the "moment-matching" approach, which as I will show below, specifies the shape parameters through link-functions of the mean and variance.

and

$$\beta = \exp(\mathbf{X}\Phi) \tag{5}$$

where $\mathbf{X}$ is an nxk data matrix of explanatory variables and Γ and Φ are kx1 parameter vectors. Note that the same set of explanatory variables are used in both expressions even if a variable is thought to affect only the mean or variance because both the mean and variance are functions of both α and β. From here, one obtains the log-likelihood function for the beta distribution.

$$\log L = \sum_{i=1}^{N} \ln \Gamma(\alpha + \beta) - [\ln \Gamma(\alpha) + \ln \Gamma(\beta)] + (\alpha - 1)\ln(y) + (\beta - 1)\ln(1 - y) \tag{6}$$

where α and β are specified by equations 4 and 5, respectively.

A drawback to this method is that the parameters cannot be interpreted directly as having a statistically significant effect upon either the mean or the variance. At a minimum, the effects of the independent variables upon the mean and the variance can be examined only with respect to substantive effects using the method of first differences (King 1989). One way of assessing statistical significance, however, is to perform a Wald test on the first differences for some substantively meaningful change in one variable while holding the others at their means.

The first step in using a Wald test is obtaining the change in the mean or variance produced by holding all independent variables at their means, with the exception of the k^{th} independent variable, which is set to two different arbitrary or substantively meaningful values, and calculating the predicted α and β parameters. The Wald statistic,

$$W \;=\; \mathcal{D}\,(\frac{\partial \mathcal{D}}{\partial(\Gamma'\,\Phi')'})'\,\Sigma^{-1}\,(\frac{\partial \mathcal{D}}{\partial(\Gamma'\,\Phi')'})\,\mathcal{D}', \tag{7}$$

is distributed as a χ^2 variable with 1 degree of freedom, where Σ is the variance-covariance matrix of the estimated parameters.[4] One criticism of using the Wald test to determine the

[4]Hypotheses about sets of variables can also be tested, with degrees of freedom equal to the number of variables involved. Specific formulas for $\frac{\partial \mathcal{D}}{\partial(\Gamma'\,\Phi')'}$ are given in the appendix.

statistical significance of a variable is that a given variable can, under some circumstances, be found to be statistically significant or insignificant depending solely upon the researcher's choice of a first difference.

An alternative means of estimation is one that estimates the effect of an independent variable upon the mean and/or the variance of the dependent variable — the "moment-matching" approach — by specifying α and β in terms of the mean and variance, which are themselves specified by link-functions of estimated parameters and the independent variables. In this approach, a functional form is specified for the mean and variance. One specification is:

$$E(Y) = \frac{\exp(\mathbf{X}\Gamma)}{1+\exp(\mathbf{X}\Gamma)} \tag{8}$$

$$Var(Y) = \frac{\exp(\mathbf{Z}\Phi)}{12[1+\exp(\mathbf{Z}\Phi)]} \tag{9}$$

where, $\mathbf{X}$ and $\mathbf{Z}$ are (potentially non-identical) matrices of independent variables and Γ and Φ are parameter vectors for the effects of $\mathbf{X}$ and $\mathbf{Z}$ upon the expected value and variance of $\mathbf{y}$, respectively. The form in equation 8 reflects the constraint that the expected value of Y is bounded by 0 and 1, and the functional form in equation 9 reflects the constraint that the variance of Y, given a unimodal distribution of Y, is bounded by 0 and 1/12. Many problems in political science using a beta distribution are probably unimodally distributed, so the simulations in this paper will be based upon a specification of the variance as in equation 9.[5]

From here, α and β are specified in terms of the mean and variance.

$$\alpha = \frac{[E(Y)]^2[1-E(Y)]}{Var(Y)} - E(Y) \tag{10}$$

$$\beta = \frac{E(Y)[1-E(Y)]^2}{Var(Y)} - [1-E(Y)] \tag{11}$$

Given the specification of the mean and variance of Y expressed in equations 8 and 9, the

[5]In the conclusion, I will comment on estimating a bimodal beta distribution.

above equations can also be written as:

$$\beta = -\frac{-12\exp(\mathbf{X}\Gamma) - 10\exp(\mathbf{X}\Gamma)\exp(\mathbf{Z}\Phi) + \exp(\mathbf{Z}\Phi) + \exp(\mathbf{X}\Gamma)^2\exp(\mathbf{Z}\Phi)}{\exp(\mathbf{Z}\Phi)(\exp(\mathbf{X}\Gamma)^3 + 3\exp(\mathbf{X}\Gamma)^2 + 3\exp(\mathbf{X}\Gamma) + 1)} \tag{12}$$

and $\alpha = \exp(\mathbf{X}\Gamma)\beta$. Written in this form, one can see that this means of estimation places a restriction on the parameters such that, if $\alpha, \beta > 0$ then

$$\begin{aligned} 12\exp(\mathbf{X}\Gamma) + 10\exp(\mathbf{X}\Gamma)\exp(\mathbf{Z}\Phi) - \exp(\mathbf{Z}\Phi) - \exp(\mathbf{X}\Gamma)^2\exp(\mathbf{Z}\Phi) &> 0 \\ -\exp(\mathbf{Z}\Phi)[\exp(\mathbf{X}\Gamma)^2 - 10\exp(\mathbf{X}\Gamma) + 1] &> -12\exp(\mathbf{X}\Gamma) \\ \frac{12\exp(\mathbf{X}\Gamma)}{\exp(\mathbf{X}\Gamma)^2 - 10\exp(\mathbf{X}\Gamma) + 1} &> \exp(\mathbf{Z}\Phi) \\ \ln(12) + \mathbf{X}\Gamma - \ln(\exp(\mathbf{X}\Gamma)^2 - 10\exp(\mathbf{X}\Gamma) + 1) &> \mathbf{Z}\Phi \end{aligned} \tag{13}$$

While not necessarily a problem for the results, this restriction does suggest that estimation using this approach, particularly arriving at suitable starting values, may be more difficult.

In the next section, I will use Monte Carlo simulations to assess both means of estimation under a variety of conditions, including changes in sample sizes and changes in the expected value and distribution of the dependent variable.

2 Monte-Carlo Results

To examine the accuracy of the moment-matching method, a series of Monte Carlo simulations were performed to determine how well the moment-matching method reproduced the results that would be expected from the standard method. In these simulations, a beta-distributed variable was randomly generated using Chen's BA algorithm (see Johnson, Kotz and Balakrishnan 1995, 216), where $(\alpha\beta) = \exp(\mathbf{XB})$, α and β are Nx1 vectors of shape parameters, $\mathbf{X}$ is an Nx4 matrix of a constant and three columns of random standard random normal variables, and $\mathbf{B}$ is a 4x2 matrix of parameters. The approach in these Monte Carlo simulations will be to compare the average value of the first differences for the mean and variance of each variable from the parameters estimated by the direct and moment-matching approaches with the expected values of the first differences, given the known parameters. For all simulations, 1000 trials were performed using the BHHH algorithm, which I have found to be the least sensitive to poor starting values with the moment-matching method.

2.1 Simulation #1

The purpose of the first simulation was to examine how well the moment matching method performed for a dependent variable generated with $\alpha, \beta > 1$ and a mean value around 0.5. For this, the parameter vectors for α and β were (3.5 0.3 0.5 0.7)' and 3.5 0.6 0.4 0.2)', respectively. The expected value for α and β, therefore, is 33.11, the expected mean for the dependent variable is 0.5, and the expected variance is 0.0037.[6] A graph of the distribution is below in Figure 1.

[Figure 1 about here]

The expected first differences for each independent variable were determined by calculating the values of α and β by setting each independent variable at its expected value one standard deviation above and below its mean, i.e. +1 and -1 respectively, while holding the other two variables at their means, i.e. 0, and finally, calculating the expected first differences for the mean and variance of each independent variable. If the moment-matching method generates accurate estimates of the effect of a set of independent variables upon the mean and variance of a dependent variable, then the predicted first differences from these estimates should mirror those expected from the known parameters. For further purposes of comparison, the average predicted first differences obtained from Monte Carlo simulations of a normal model, with expected value $\mathbf{X}$B and variance exp($\mathbf{X}$B).

[Table 1 about here]

[6] High values of the constants in α and β were chosen to keep all α and β values in the simulations greater than 1.

The results in Table 1 show that the moment-matching method reproduces the true mean and variance values very well (and better than the estimates obtained from the normal distribution), but usually does not do as well as the values obtained through direct estimation of α and β. In both cases of large and small samples, the moment-matching approach does produce estimates that are comparable to those produced by the direct approach, which indicates that while the direct method is usually preferable to the moment-matching method because of the ease of estimating the parameters, in small samples, a researcher who wanted to save degrees of freedom by constraining the effect of a variable upon the mean or variance to 0 could obtain reliable results from the moment-matching method.[7]

2.2 Simulation #2

The purpose of the second simulation is to move away from an "easy" specification of the beta distribution to one that is skewed and examine how well the moment-matching method performed for a dependent variable generated with $\alpha < 1, \beta > 1$ and a mean value closer to 0. For this, the parameter vectors for α and β were (-0.65 0.3 0.5 0.7)' and (1 0.6 0.4 0.2)', respectively. The expected value for α and β, therefore, is .52 and 2.71, respectively, the expected mean for the dependent variable is 0.1609, and the expected variance is .0319. The density function for this distribution is shown in Figure 2.

[Figure 2 about here]

The expected first differences for each independent variable were determined, as before, by calculating the values of α and β by setting each independent variable at its expected value one standard deviation above and below its mean, i.e. +1 and -1 respectively, while holding the other two variables at their means, i.e. 0, and finally, calculating the expected

[7] I should note that these results are based upon 1000 successful trials. Failed trials are those where the model did not converge or the variance-covariance matrix of the parameters was not returned. Examination of the descriptive statistics for the dependent and independent variables for successful and failed trials did not indicate any major differences.

first differences for the mean and variance of each independent variable.

[Table 2 about here]

The results from this second simulation indicate that the moment-matching method does a fair job of reproducing the true results. With the exception of the variance for X_1, the estimated first differences from the moment-matching method are within 2 standard deviations of the true values, but the estimates from the moment matching method are, in some cases, actually worse than the estimates provided by the normal maximum likelihood estimates. The data from these simulations indicate that the direct method is clearly preferable to the moment-matching method in cases where one of the shape parameters is less than 1. This suggests that researchers who have a reason for using the moment-matching method should first estimate their model using the direct method to insure that the predicted shape parameters are both greater than 1.

3 Empirical Examples

The Monte Carlo simulations provide some general indications about the performance of the moment-matching approach under some representative conditions, but this section will provide comparisons of the moment-matching and the direct approach using real survey data and representing a variety of situations. The data for these empirical examples come from the 1988 National Election Study's Super Tuesday survey data.

The Super Tuesday study asked respondents how they perceived each of the candidates' chances of winning the nomination on a 0-100 point scale. These subjective evaluations of the candidates' probabilities of winning the nomination are well-suited for estimation using a beta distribution as the likelihood function. First off, these evaluations are bounded by zero and one. Second, as subjective probabilities, it makes theoretical sense to use a distribution where the mean and variance are not independent of one another. With the beta distribution, the maximum value of the variance occurs when $\alpha = \beta$, or when the mean value is 0.5, and minimized as α or β approach 0, and the mean value approaches 0 and 1, respectively.[8] If we interpret the variance as representing an individual's certainty about the subjective probability of a candidate winning the nomination, this relationship between the mean and variance of the beta distribution is relevant because it is hard to imagine an individual being very certain that a candidate had a 50-50 chance of winning the nomination or being very uncertain that a candidate had no chance of winning or losing. At the same time, the variance of two beta-distributed variables can differ even if those variables have the same mean. This allows for individuals with the same subjective assessment of a candidate's viability to hold that belief with different levels of certainty.

With respondents giving evaluations on these scales to 6 Republican and 7 Democratic candidates, it is not clear that we can treat each raw response as equivalent to a probability because the sum of the raw scores often exceeds 100. Furthermore, as Bartels (1988) points out, the question that respondents were asked:

> Now, thinking about these nominating conventions, who do you think is likely to win the Democratic nomination for President. We will be using a scale which runs from 0 to 100, where 0 represents no chance for the nomination, 50 represents an even chance, and 100 represents certain victory. You may use any number between one and one hundred. What do you think ALBERT GORE's chances are?

uses 50 as the point where a candidate is perceived as having as good a chance as any other candidate. To transform these raw scores into probabilities, I follow Bartels's (1988) method for normalizing these scores. First, the raw score is divided by 100. Then, this score is raised to the 2.59 power for Republicans and the 2.81 power for Democrats so that a response of 50 (or an even chance) is transformed into a score of .166 for Republicans and

[8] In all cases, Jarque-Bera tests rejected the null hypothesis of normal errors for these models.

a score of .142 for Democratic candidates. Finally, the scores are summed and each divided by that sum to produce probabilities. Missing data are scored as 50 if the respondent could evaluate the candidate on a feeling thermometer and 0 if the respondent had not heard of the candidate.

The model for the empirical examples simply makes respondents' perceptions of a candidate's viability a function of their preference for the candidate, whether or not they believe that the candidate has the best chance of winning the primary in their state, their level of political information, their exposure to the media, the number of candidates from that candidate's party whose viability that are able to assess, and several other control variables. More specific details about the model are available in Paolino (1998). For both the direct means of estimation α and β and the moment-matching method, equivalent models are estimated, with all independent variables being unconstrained to have an effect upon both the mean and variance of the dependent variable.

The results for the Republican candidates (Table 3) indicate that the moment-matching estimation produces conclusions that are little different from those reached by the standard method of estimation.[9] There is no case where the predicted first differences from the moment-matching method are significantly different from the direct method of estimating α and β. There is one case, however, where the moment-matching method produces a statistically insignificant parameter, t=-1.90, for the effect of the respondent's evaluation of Bush upon the certainty of that evaluation.

[Tables 3 and 4 about here]

The differences between the two means of estimation for the Democratic candidates presented (Table 4) are similar to those for the Republicans, but the differences for Hart

[9]The three Republican candidates presented here were chosen because they represent cases of a symmetric distribution with a mean close to .5, a moderately-skewed distribution with a mean between 0 and .5, and a highly-skewed distribution with a mean close to 0. Results from the models of the other Republican candidates show no significant differences between the direct and moment-matching means of estimation.

and Jackson, the two candidates who with predicted α values less than 1, are not as close as for the other candidates. This seems to corroborate the Monte Carlo results, which indicated that the moment-matching method was not as successful when one of the shape parameters is less than 1.[10]

4 Conclusions

In general, the results from both the Monte Carlo and empirical examples indicate that the direct method of estimating α and β is probably the safer one to use, especially when one is dealing with a distribution of points where one (or both) of the shape parameters is less than one; although, one will generally reach similar conclusions using either method. One case where the moment-matching method may have an advantage over the standard means of estimation is where the researcher does not have any theoretical reason to believe that a variable has an effect upon the mean or variance and the researcher needs to conserve degrees of freedom and estimates indicate that both shape parameters are greater than 1. In the empirical examples, such a case could be made for the variable indicating the respondent's belief that the candidate was most likely to win the primary in the respondent's state of residence always had a positive and significant effect upon the mean assessment of the candidate's viability, but was almost never significantly related to the variance of that assessment.[11] Overall, the results from the analysis in this paper would lead me to recommend use of the standard method in all cases, with a switch to the moment-matching method only in circumstances where a researcher would like to present a more parsimonious empirical model.

[10] As with the Republican candidates, the differences between the direct and moment-matching methods of estimation were trivial for the 4 other Democratic candidates.

[11] And in the one case where the variable was significantly related to the variance, the above results suggest that the standard means of estimation is preferable.

Appendix

Partial Derivatives for First Differences

$$\begin{aligned}\frac{\partial \mathcal{D}_{mean}}{\partial \Gamma} &= \frac{X_h \exp(X_h \Gamma)}{\exp(X_h \Gamma) + \exp(X_h \Phi)} - \frac{\exp(X_h \Gamma)^2 X_h}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^2} \\ &- \frac{X_l \exp(X_l \Gamma)}{\exp(X_l \Gamma) + \exp(X_l \Phi)} + \frac{\exp(X_l \Gamma)^2 X_l}{[\exp(X_l \Gamma) + \exp(X_l \Phi)]^2}\end{aligned} \tag{14}$$

$$\frac{\partial \mathcal{D}_{mean}}{\partial \Phi} = -\frac{X_h \exp(X_h \Gamma) \exp(X_h \Phi)}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^2} + \frac{X_l \exp(X_l \Gamma) \exp(X_l \Phi)}{[\exp(X_l \Gamma) + \exp(X_l \Phi)]^2} \tag{15}$$

$$\begin{aligned}\frac{\partial \mathcal{D}_{variance}}{\partial \Gamma} &= \frac{X_h \exp(X_h \Gamma) \exp(X_h \Phi)}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^2 (\exp(X_h \Gamma) + \exp(X_h \Phi) + 1)} \\ &- \frac{2 \exp(X_h \Gamma)^2 \exp(X_h \Phi) X_h}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^3 (\exp(X_h \Gamma) + \exp(X_h \Phi) + 1)} \\ &- \frac{\exp(X_h \Gamma)^2 \exp(X_h \Phi) X_h}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^2 (\exp(X_h \Gamma) + \exp(X_h \Phi) + 1)^2} \\ &- \frac{X_l \exp(X_l \Gamma) \exp(X_l \Phi)}{[\exp(X_l \Gamma) + \exp(X_l \Phi)]^2 (\exp(X_l \Gamma) + \exp(X_l \Phi) + 1)} \\ &+ \frac{2 \exp(X_l \Gamma)^2 \exp(X_l \Phi) X_l}{[\exp(X_l \Gamma) + \exp(X_l \Phi)]^3 (\exp(X_l \Gamma) + \exp(X_l \Phi) + 1)} \\ &+ \frac{\exp(X_l \Gamma)^2 \exp(X_l \Phi) X_l}{[\exp(X_l \Gamma) + \exp(X_l \Phi)]^2 (\exp(X_l \Gamma) + \exp(X_l \Phi) + 1)^2}\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{\partial \mathcal{D}_{variance}}{\partial \Phi} &= \frac{X_h \exp(X_h \Gamma) \exp(X_h \Phi)}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^2 (\exp(X_h \Gamma) + \exp(X_h \Phi) + 1)} \\ &- \frac{2 \exp(X_h \Gamma) \exp(X_h \Phi)^2 X_h}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^3 (\exp(X_h \Gamma) + \exp(X_h \Phi) + 1)} \\ &- \frac{\exp(X_h \Gamma) \exp(X_h \Phi)^2 X_h}{[\exp(X_h \Gamma) + \exp(X_h \Phi)]^2 (\exp(X_h \Gamma) + \exp(X_h \Phi) + 1)^2} \\ &- \frac{X_l \exp(X_l \Gamma) \exp(X_l \Phi)}{[\exp(X_l \Gamma) + \exp(X_l \Phi)]^2 (\exp(X_l \Gamma) + \exp(X_l \Phi) + 1)}\end{aligned}$$

$$+ \quad \frac{2\ \exp(X_l\,\Gamma)\ \exp(X_l\,\Phi)^2 X_l}{[\exp(X_l\,\Gamma) + \exp(X_l\,\Phi)]^3\ (\exp(X_l\,\Gamma) + \exp(X_l\,\Phi) + 1)}$$

$$+ \quad \frac{\exp(X_l\,\Gamma)\ \exp(X_l\,\Phi)^2 X_l}{[\exp(X_l\,\Gamma) + \exp(X_l\,\Phi)]^2\ (\exp(X_l\,\Gamma) + \exp(X_l\,\Phi) + 1)^2} \qquad (17)$$

X_h and X_l are the vectors of values for the independent variables held at their means, with one variable set to the higher and lower values for the first difference.

References

Bartels, Larry M. 1988. *Presidential Primaries.* Princeton: Princeton University.

Brehm, John and Scott Gates. 1993. "Donut Shops and Speed Traps: Evaluating Models of Supervision on Police Behavior." *American Journal of Political Science* 37(2):555–81.

Franklin, Charles H. 1991. "Eschewing Obsfucation? Campaigns and the Perception of U.S. Senate Incumbents." *American Political Science Review* 85(4):1193–1214.

Johnson, Norman L., Samuel Kotz and N. Balakrishnan. 1995. *Continuous Univariate Distributions.* Vol. 2 2nd ed. New York: John Wiley and Sons, Inc.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge: Cambridge University Press.

Mebane, Walter R. Jr. 1998. "Rational Expectations Coordinating Voting in American Presidential and House Elections." Prepared for delivery at the 1998 Annual Meeting of the Midwest Political Science Assocation.

Paolino, Philip. 1998. "Voters' Perceptions of Candidate Viability: Uncertainty and the Prospects for Momentum.". Prepared for presentation at the 1998 Annual Meeting of the Midwest Political Science Association.

Figure 1: Beta Density Function, $\alpha = 33.11, \beta = 33.11$

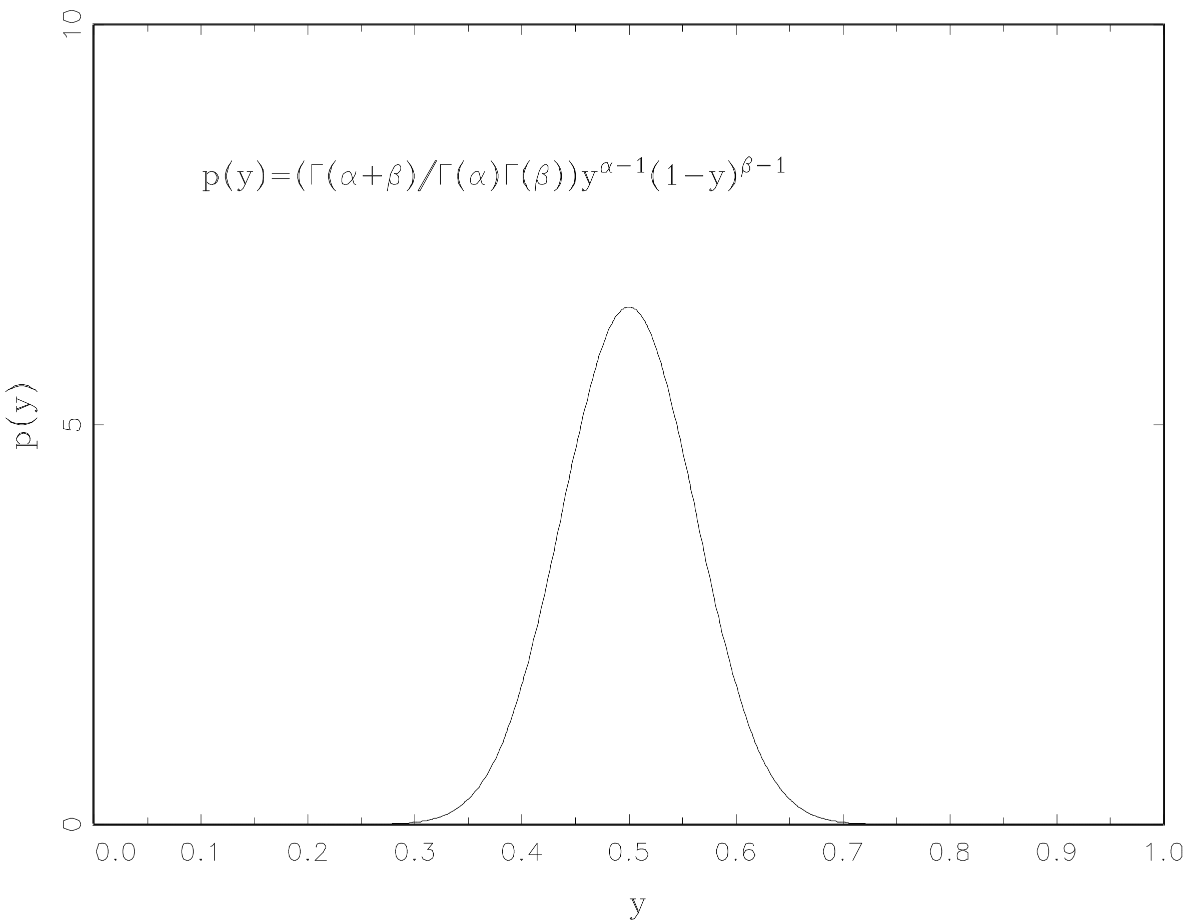

Figure 2: Beta Density Function, $\alpha = .52, \beta = 2.71$

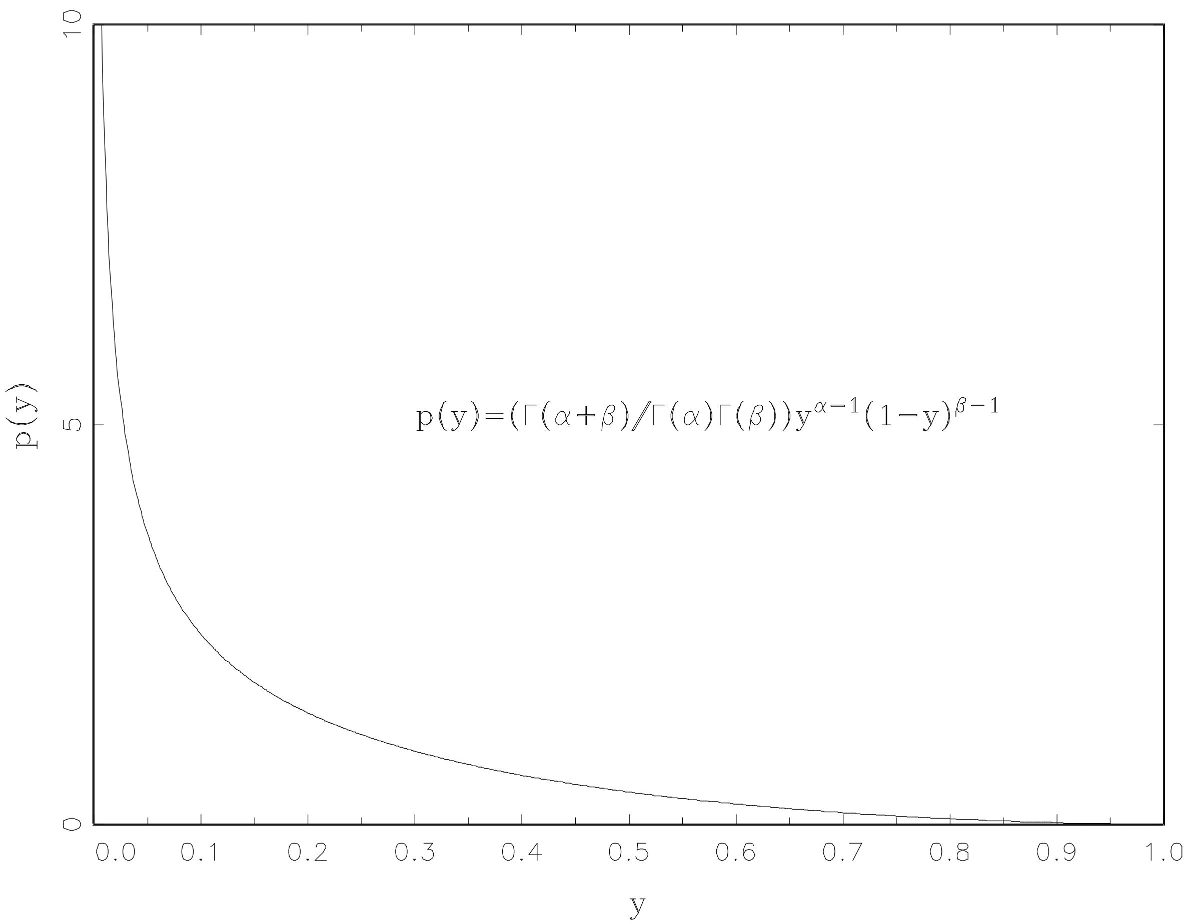

Table 1: Monte Carlo Results, $\alpha, \beta > 1$

	Expected	Average Direct	Average M-M	Average Normal
		First Differences for N=500		
Mean				
x_1	-0.1489	-0.1486	-0.1482	-0.1387
		(0.0064)	(0.0067)	(0.0063)
x_2	0.0500	0.0500	0.0497	0.0466
		(0.0052)	(0.0051)	(0.0051)
x_3	0.2499	0.2448	0.2435	0.2306
		(0.0087)	(0.0086)	(0.0080)
Variance				
x_1	-0.00328	-0.00326	-0.00306	-0.00311
		(0.00055)	(0.00055)	(0.00054)
x_2	-0.00338	-0.00340	-0.00304	-0.00297
		(0.00055)	(0.00052)	(0.00053)
x_3	-0.00310	-0.00311	-0.00304	-0.00284
		(0.00050)	(0.00052)	(0.00055)
		First Differences for N=50		
Mean				
x_1	-0.1489	-0.1496	-0.1472	-0.1379
		(0.0237)	(0.0234)	(0.0217)
x_2	0.0500	0.0494	0.0501	0.0474
		(0.0186)	(0.0177)	(0.0183)
x_3	0.2499	0.2422	0.2412	0.2310
		(0.0296)	(0.0301)	(0.0270)
Variance				
x_1	-0.00328	-0.00325	-0.00322	-0.00309
		(0.00207)	(0.00202)	(0.00208)
x_2	-0.00338	-0.00332	-0.00315	-0.00290
		(0.00212)	(0.00197)	(0.00194)
x_3	-0.00310	-0.00296	-0.00308	-0.00279
		(0.00182)	(0.00198)	(0.00200)

Simulations based upon 1000 trials. Standard deviations in parentheses. BHHH algorithms used for beta estimation for both beta simulations for N=500 (0 failed trials), direct (50 failed trials) and moment-matching for N=50 (332 failed trials). Newton-Raphson algorithms used for all normal models.

Table 2: Monte Carlo Results, $\alpha < 1, \beta > 1$

	Expected	Average Direct	Average M-M	Average Normal
		First Differences for N=500		
Mean				
x_1	-0.0813	-0.0815	-0.0639	-0.0715
		(0.0149)	(0.0138)	(0.0136)
x_2	0.0270	0.0269	0.0315	0.0257
		(0.0135)	(0.0125)	(0.0141)
x_3	0.1362	0.1358	0.1270	0.1215
		(0.0136)	(0.0140)	(0.0125)
Variance				
x_1	-0.04041	-0.04012	-0.03185	-0.03656
		(0.00567)	(0.00431)	(0.00625)
x_2	-0.01577	-0.01566	-0.01213	-0.01360
		(0.00509)	(0.00465)	(0.00590)
x_3	0.00719	0.00694	0.00869	0.00707
		(0.00422)	(0.00478)	(0.00612)
		First Differences for N=50		
Mean				
x_1	-0.0813	-0.0806	-0.0684	-0.0562
		(0.0531)	(0.0486)	(0.0528)
x_2	0.0270	0.0260	0.0315	0.0292
		(0.0477)	(0.0488)	(0.0494)
x_3	0.1362	0.1315	0.1307	0.1189
		(0.0488)	(0.0520)	(0.0462)
Variance				
x_1	-0.04041	-0.03698	-0.03340	-0.03524
		(0.01867)	(0.01654)	(0.02536)
x_2	-0.01577	-0.01589	-0.01374	-0.01061
		(0.01723)	(0.01979)	(0.02297)
x_3	0.00719	0.00348	0.00700	0.01378
		(0.01547)	(0.02072)	(0.02709)

Simulations based upon 1000 trials. Standard deviations in parentheses. BHHH algorithm used for direct beta N=500 (10 failed trials) and N=50 (8 failed trials) and moment-matching beta N=500 (380 failed trials) and N=50 (302 failed trials). Newton-Raphson algorithms used for all normal models.

Table 3: Beta Models for Republican Candidates

	Bush		Dole		Robertson	
Independent Variable	$\mathcal{D}_{dir}$	$\mathcal{D}_{m-m}$	$\mathcal{D}_{dir}$	$\mathcal{D}_{m-m}$	$\mathcal{D}_{dir}$	$\mathcal{D}_{m-m}$
	Estimated First Differences for the Mean					
Candidate Evaluation	.020	.013	.066*	.067*	.110*	.108*
Best in State	.089*	.082*	.088*	.090*	.031*	.032*
Political Information	.030	.030	.023	.026	-.012	-.011
Media Exposure	-.032	-.026	.026	.024	.017	.018
# of Candidates Placed	-.201*	-.221*	-.028	-.017	-.066*	-.062*
	Estimated First Differences for the Variance					
Candidate Evaluation	-.014*	-.015	-.001	-.001	.012*	.013*
Best in State	.004	.005	.005	.007	.003	.003
Political Information	-.011*	-.011*	.006	.006	-.001	-.000
Media Exposure	.001	.000	-.001	-.003	.002	.003
# of Candidates Placed	-.028*	-.022*	-.016*	-.015*	-.013*	-.013*
N	402	402	331	331	315	315
$\hat{\mu}$	.515	.520	.324	.320	.098	.097
$\hat{\sigma}^2$	.036	.034	.023	.024	.007	.007
$\hat{\alpha}$	3.04	3.23	2.76	2.64	1.19	1.19
$\hat{\beta}$	2.85	2.98	5.75	5.61	10.96	11.04

Source: NES Super Tuesday Study. $*p < .05$.

Table 4: Beta Models for Democratic Candidates

	Dukakis		Hart		Jackson	
Independent Variable	$\mathcal{D}_{dir}$	$\mathcal{D}_{m-m}$	$\mathcal{D}_{dir}$	$\mathcal{D}_{m-m}$	$\mathcal{D}_{dir}$	$\mathcal{D}_{m-m}$
	Estimated First Differences for the Mean					
Candidate Evaluation	.102*	.105*	.223*	.214*	.079*	.068
Best in State	.070*	.067*	.096*	.068*	.069*	.062*
Political Information	.010	.018	-.029	-.014	-.005	-.004
Media Exposure	-.001	-.008	-.034	-.050*	-.027	-.039
# of Candidates Placed	-.184*	-.206*	-.277*	-.270*	-.201*	-.206*
	Estimated First Differences for the Variance					
Candidate Evaluation	-.003	.002	.040*	.045*	.007	.002
Best in State	-.011	-.011	.006	.008	.016*	.013
Political Information	.003	.008	-.010	-.010	-.003	-.000
Media Exposure	-.002	.002	-.007	-.013	-.007	-.015
# of Candidates Placed	-.042*	-.042*	-.067*	-.058*	-.059*	-.051*
N	209	209	391	391	387	387
$\hat{\bar{\mu}}$	.332	.341	.241	.234	.204	.207
$\hat{\bar{\sigma}}^2$	.038	.041	.038	.043	.034	.038
$\hat{\bar{\alpha}}$	1.60	1.53	0.94	0.74	0.78	0.69
$\hat{\bar{\beta}}$	3.21	2.95	2.96	2.42	3.06	2.65

Source: NES Super Tuesday Study. $*p < .05$.