Political Analysis, 9:4

Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables

Philip Paolino
Department of Government, The University of Texas at Austin,
536 Burdine Hall, Austin, TX 78712-1087
e-mail: ppaolino@mail.la.utexas.edu
http://www.la.utexas.edu/∼ppaolino

Research in political science is often concerned with modeling dependent variables that are proportions. Proportions are relevant in a wide variety of substantive areas, including elections, the bureaucracy, and interest groups. Yet because most researchers rely upon an approach, OLS, that does not recognize key aspects of proportions, the conclusions we reach from normal models may not provide the best understanding of phenomena of interest in these areas. In this paper, I use Monte Carlo simulations to show that maximum likelihood estimation of these data using the beta distribution may provide more accurate and more precise results. I then present empirical analyses illustrating some of these differences.

1 Introduction

RESEARCH IN POLITICAL science is often concerned with modeling dependent variables that are proportions. Researchers may be interested in things like people's political knowledge, which is often measured as the proportion of correct responses to a series of factual questions (e.g., Delli Carpini and Keeter 1996), or whether or not welfare recipients relocate to receive higher benefits, modeled in terms of the proportion of poor single mothers with children migrating from one state to another (e.g., Schram et al. 1998). Much of this work relies upon the robustness and computational ease of "normal" modeling of data using ordinary least squares (OLS). We know, however, that many of the OLS assumptions, such as normality and homoskedasticity, are often violated when the dependent variable is a proportion.

Others (e.g., Palmquist 1999) have recognized these problems and proposed alternatives to OLS when modeling proportions. When these data arise from the generation of grouped binary data (as with political knowledge), where the researcher can calculate the ratio of

Author's note: I would like to thank Chris Achen, John Brehm, Mel Hinich, Gary King, Tse-min Lin, Walter Mebane, Jonathan Nagler, Kevin Quinn, and two anonymous reviewers for helpful suggestions. I would also like to thank the Center for Basic Research in the Social Sciences for providing support and intellectual stimulation crucial to this paper's development. Finally, I would like to thank all the authors who graciously provided their replication data to me for reanalysis in this paper. Replication data sets and code are available at the Political Analysis web site. All references to results not present are also available at that web site.

Copyright 2001 by the Society for Political Methodology

successes to the total number of trials, Palmquist argues that the beta-binomial distribution provides a more appropriate statistical model. There are, however, other proportions of interest in political science research that do not arise from grouping binary processes or where the data from which the proportions arise are not available to use the beta-binomial distribution. For example, Brehm and Gates (1993) examine police officers' shirking behavior. In that paper, estimates of the percentage of time an officer shirks are based upon researchers' own observations. These data do not arise from a generating process where an officer is presented with a series of discrete opportunities to shirk or work. The measure is truly a continuous one. In cases such as these, I argue that the beta distribution should be considered as an appropriate distribution.

The beta distribution is a flexible distribution that can produce a unimodal, uniform, or bimodal distribution of points that can be either symmetrical or skewed. Substantively, the beta distribution has been applied recently to examining police officers' shirking (Brehm and Gates 1993), the location of a party's ideal point (Mebane 2000), and individuals' subjective beliefs about the probability of a candidate's chances of winning an election (Paolino 1998).

One particularly attractive feature of the beta distribution is its recognition of a relationship between the mean and the variance that may occur with proportions. A normally distributed variable can have any variance. But a beta-distributed variable with a mean close to either 0 or 1 generally has a smaller variance than a proportion with a mean of 0.5.[1] Any covariate in a quantitative model that has a large effect upon the mean is also likely to imply a heterogeneous variance. While it is possible to model heterogeneous variance in normal models, it is not readily apparent how one models the variance so that it is also a function of the mean. In many instances, researchers may also be interested in not just how a covariate influences the expected value of the function, but also the variance. Modeling the variance of a distribution can be used, for example, to estimate factors influencing individuals' certainty about their own ideal points (e.g., Alvarez and Brehm 1995; Alvarez and Brehm 1997; Franklin 1991) or heterogeneity across voters' beliefs about a candidate's chances of being elected (e.g., Paolino 1998). I have argued, for instance, that heterogeneity in voters' beliefs about candidates' probability of winning the presidential nomination can provide a basis for predicting which candidates are most likely to generate momentum in the primaries. When a researcher is dealing with a proportion and the theory calls for heterogeneity, one should definitely consider the use of the beta distribution instead of the normal distribution. Even in cases where the researcher does not have any theoretical concern with the variance function, the potential for heteroskedasticity should raise concerns with normal-least squares estimates and lead one to consider using a beta distribution to model the heteroskedasticity.

In this paper, I show that the use of a normal least-squares estimation approach can lead us to inaccurately evaluate the effects of covariates in our models. In Section 2, I briefly discuss the beta distribution, lay out a standard approach to maximum likelihood estimation using the beta distribution, and provide Monte Carlo results comparing the beta estimation with normal models. In Section 3, I introduce an alternative method of using the beta distribution in maximum likelihood estimation more in line with the common practice of direct specification of the central moments of interest. In Section 4, I present empirical examples showing how estimation using the beta distribution can lead researchers to conclusions

[1] It is, however, easy to think of examples where this does not hold. For instance, a beta random variable with a unimodal distribution will have a smaller variance than a beta variable with a bimodal distribution, regardless of the mean.

different than the ones they reach using traditional normal-linear models. Finally, in Section 5, I summarize how the results of these findings can be used by researchers faced with the estimation of proportion data.

2 Estimation Using the Beta Distribution

Unlike many distributions familiar to political scientists, such as the normal and the Poisson, the beta distribution is defined 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 are functions of the two parameters, α and β:

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

where $0 \leq y \leq 1$, $\alpha, \beta > 0$, $\Gamma()$ denotes the gamma function,

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

and

$$\begin{aligned} \text{Var}(Y) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= E(Y)[1-E(Y)][1/(\alpha+\beta+1)] \end{aligned} \tag{3}$$

Note that the variance is a function of the mean and a dispersion parameter. The form of the variance in the last line of Eq. (3) is relevant below when I introduce an alternative approach to estimating the beta distribution.

Brehm and Gates (1993) provide one example in the political science literature that uses what I refer to as a "standard" approach to beta maximum likelihood estimation (BMLE) of the effects of several covariates upon police officers' defection (shirking) and compliance (working) rates. In this approach, y is an $N \times 1$ vector of observations of the dependent variable, α and β [see Eq. (1)] are specified as functions of $\mathbf{X}$, an $N \times K$ matrix of covariates, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$, respectively, are $K \times 1$ vectors of parameters representing the effects of the covariates upon compliance and defection. Because α and β must be positive, obvious specifications are

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

and

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

From here, one estimates the parameters by maximizing the log-likelihood function,

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

In the standard approach, the same set of covariates is used in both expressions. Because the mean and variance are functions of both α and β, estimating both shape parameters as functions of the full set of covariates should provide better estimates of the data. With the estimates, researchers can calculate the estimated effects of any quantity of interest, such as the mean, the variance, or the dispersion of the distribution. Brehm and Gates (Figs. 1–3)

illustrate how estimation of the beta distribution provides a better representation of the data than OLS. Of greater relevance, they also note several violations of the classical regression assumptions and show that OLS estimates of these data produce estimates with signs and inferential statistics different from those obtained from BMLE.

2.1 *Normal Alternatives for Estimating Proportions*

Given the relative inconvenience of estimating BMLE—none of the major packages have canned routines for beta estimation—one could easily decide to avoid using the beta distribution and instead estimate the model assuming normally distributed errors. There are several reasons why this might not be wise. First, the normal distribution ignores the constraints upon the expected value of a beta-distributed variable and, as a result, produces predicted values outside of the logically possible domain, such as negative percentages or probabilities greater than 1. More importantly, because the variance of the beta distribution is a function of the expected value, it would not be surprising for beta-distributed variables to violate the assumption of homoskedasticity—meaning that the standard errors generated by a normal-linear model would be incorrect, as would the statistical inferences that a researcher would draw. Finally, using a normal distribution, where the mean and variance are independent, may ignore an important aspect of the data generating process and produce a poorer fit of the data. While researchers can estimate a heteroskedastic OLS model, standard practice often leads researchers to ignore this possibility, causing them to overlook theoretically interesting aspects of the data generating process that could broaden our understanding of the phenomena under investigation. For this reason, researchers using OLS should at least consider modeling the variance as a function of some of the independent variables.

Many researchers often try to surmount some of these other problems by transforming the dependent variable. One common approach uses the normal distribution for estimation after performing a logistic transformation on the dependent variable,

$$y = \ln\left(\frac{\pi}{1-\pi}\right) \tag{7}$$

where π represents the original proportion. This approach obviously avoids the possibility of producing expected values outside of permissible bounds, as it is easy to transform the expectation back into the original scale:

$$\pi = \frac{\exp(y)}{1+\exp(y)}. \tag{8}$$

While this transformation removes the problem of using the normal distribution for a bounded variable, I show below that this approach can produce even greater problems.[2]

Essentially, the tradeoff between using BMLE and normal approaches concerns largely accuracy versus increased programming effort. I hope to convince both less and more sophisticated users that the additional effort can be worthwhile.[3] It is not my purpose to say that BMLE is always called for when dealing with proportions. Instead, I hope to provide

[2]One can also argue that "covering up" the boundary problem in this way may lead a researcher to overlook specification problems that would be noticed with an untransformed dependent variable (Achen and Shively 1995).

[3]To reduce the effort even more, my Gauss code is available at the Political Analysis web site.

some intuition as to when BMLE is most likely to matter and show the kinds of benefits that researchers can get from putting in the additional effort. In the next subsection, I use Monte Carlo simulations to assess these 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.

2.2 *The Monte Carlo Procedure*

To compare the accuracy of BMLE with these normal approaches, I performed 12 Monte Carlo simulations. In these simulations, I generated y, an $N \times 1$ vector of beta-distributed random variables with shape parameters $\alpha = \exp(\mathbf{X\Psi})$ and $\beta = \exp(\mathbf{X\Phi})$ using Cheng's BA algorithm (see Johnson et al.1995, p. 216) when $\max(\alpha,\beta) > 1$ or Jöhnk's algorithm (see Gentle 1998, p. 97) when $\alpha,\beta < 1$. $\mathbf{X}$ is an $N \times 2$ matrix of a constant and one column of a random standard normal variable. Four sets of parameters were set so that y was distributed symmetrically, $\Psi = (2.0 \;\; 0.3)'$ and $\Phi = (2.0 \;\; -0.3)'$; asymmetrically, $\Psi = (2.5 \;\; 0.3)'$, and $\Phi = (1.5 \;\; -0.3)$; one-sided at the upper bound, $\Psi = (1.5 \;\; 0.3)'$ and $\Phi = (-0.4 \;\; -0.3)$; and bimodally, $\Psi = (-0.8 \;\; 0.3)'$ and $\Phi = (-0.4 \;\; -0.3)'$. Each of these four distributions was run with N equal to 500, 100, and 30 "observations." Figure 1 shows the resulting distributions. For all simulations, 1000 trials were performed and the Newton–Raphson algorithm was used.

Generally, one compares the parameter values from the simulations with the "true" values. Because the likelihood functions are different, it makes little sense to compare parameters with one another because we do not know the "true" values of the parameters for the normal models. The first differences across all approaches, however, can be compared. Consequently, I compare the average values of the first differences for the mean from the parameters estimated by the beta and normal approaches with the "true" values of the first differences, given the data and known parameters. The first differences are calculated by varying the single covariate from 1 standard deviation above its mean value to 1 standard deviation below its mean. Comparisons are also made for the efficiency and "overconfidence" (described below) of each estimation approach.

Because one of the reasons that BMLE may be useful when modeling proportions is the potential relationship between the mean and the variance, results from the Monte Carlo simulations are compared with respect to both the mean and the variance functions. The relevant comparisons involving the mean function are between BMLE and homoskedastic OLS. When the results concern the effect of $\mathbf{X}$ upon the variance function, however, the relevant comparison is between BMLE and a heteroskedastic normal model. This is done for several reasons. First, researchers concerned with only the mean function might not consider using a heteroskedastic model. As it is, the results from the heteroskedastic normal model do not differ greatly from those produced by OLS (see the Web appendix). In contrast, the heteroskedastic linear model is used when comparing the effect of $\mathbf{X}$ upon the variance function because presumably it is a model that a researcher with a theoretical concern with modeling the variance would use.

2.2.1 Results of the Simulations

The first concern is how well the estimation procedures reproduce the data. To determine this, I examine the mean squared prediction error and the difference for the effect of $\mathbf{X}$ evaluated 1 standard deviation above and below its mean. Across all simulations, BMLE

estimates the true value of *y* slightly better than the normal approaches, but the differences are so slight that I do not report the average mean squared errors here. It is worth mentioning, however, that the OLS estimates using the logit transformation produce a noticeably worse prediction error than either BMLE or OLS using the untransformed variables (see the web appendix). Instead, I focus on the predicted first differences from the different estimation approaches (Table 1).

With respect to the estimation of effects, BMLE generally dominates linear-normal approaches, with the latter often off by at least 5–10%. There are two noteworthy points about these results. First, BMLE's advantage over OLS increases with sample size. In small samples, it seems likely that using twice as many parameters imposes some costs upon BMLE. Second, normal estimation using a logit-transformed variable produces the greatest discrepancies between the true and the estimated first differences, especially in

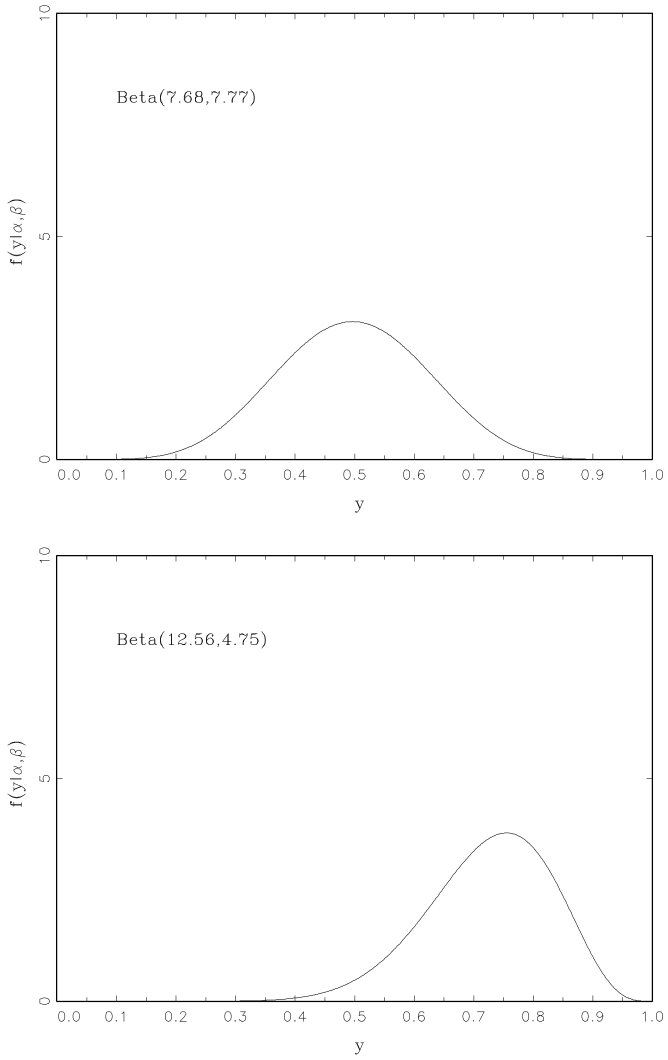


Fig. 1 Illustrative beta density functions.

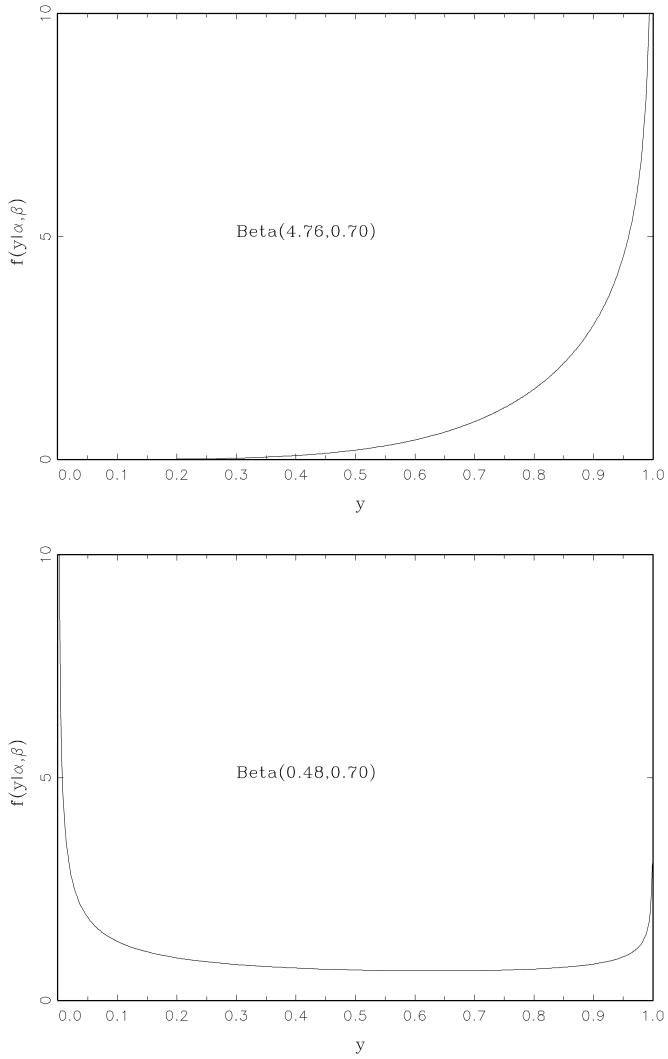


Fig. 1 *Continued.*

the bimodal case. These results and the others that follow indicate that researchers who want to use normal models are better off leaving the dependent variable untransformed. Overall, these results do not indicate, however, that estimates derived from normal models with the untransformed variable are grossly inaccurate. Below, I present empirical examples supporting these results, where OLS does both a reasonable job and a clearly inferior job of reproducing the actual data.

Beyond the average accuracy of predicted effects, we are also concerned with the efficiency of those predictions. In most Monte Carlo simulations, we would compare the efficiency of various estimators by examining the variance around a known parameter estimate. When we are comparing various normal-linear estimators, this is simply the predicted effect of a covariate upon the dependent variable. As I noted above, the effect that we can use for comparison is a first difference, not the specific parameters. As such, the efficiency

Table 1 Model fit of Beta MLE vs. alternatives

A. Expected value function

Model	*N*	$\mathcal{D}$	*Beta*	*OLS*	OLS_{lt}	$Beta_{alt}$
Symmetrical	500	.291	.291	.276	.311	.289
	100	.283	.282	.270	.301	.281
	30	.243	.244	.232	.259	.241
Asymmetrical	500	.234	.235	.227	.246	.234
	100	.239	.238	.229	.248	.238
	30	.248	.248	.250	.261	.247
One-sided	500	.132	.131	.137	.113	.131
	100	.133	.133	.140	.117	.131
	30	.140	.135	.143	.122	.133
Bimodal	500	.288	.287	.272	.556	.273
	100	.310	.312	.294	.568	.293
	30	.261	.257	.247	.468	.241

B. Variance function

Model	*N*	$\mathcal{D}$	*Beta*	*Normal*	$Normal_{lt}$	$Beta_{alt}$
Symmetrical	500	.0001	.0001	−.0002	.0002	−.0001
	100	.0003	.0004	.0014	.0001	.0007
	30	.0008	.0011	.0012	.0013	.0011
Asymmetrical	500	−.0081	−.0081	−.0074	−.0101	−.0080
	100	−.0088	−.0085	−.0089	−.0103	−.0087
	30	−.0075	−.0067	−.0067	−.0087	−.0066
One-sided	500	−.0219	−.0219	−.0224	−.0284	−.0215
	100	−.0221	−.0212	−.0231	−.0287	−.0207
	30	−.0232	−.0198	−.0255	−.0260	−.0193
Bimodal	500	.0291	.0288	.0278	.3951	.0274
	100	.0363	.0347	.0354	.3770	.0328
	30	.0248	.0226	.0314	.2306	.0214

Values are the average first differences from the Monte Carlo simulations, calculated at the value of the covariate one standard deviation above and below its mean value. OLS models in section A are homoskedastic, while normal-linear models in section B are heteroskedastic. The "lt" subscript signifies that the proportion was transformed using a logit transformation, and Beta_{alt} refers to the alternate approach to BMLE, as described in Eqs. (11)–(15).

is computed:

$$\text{Efficiency} = 100\frac{\sqrt{\sum_{i=1}^{1000}(\hat{\mathcal{D}}_i - \mathcal{D})^2}}{\sqrt{\sum_{i=1}^{1000}(\tilde{\mathcal{D}}_i - \mathcal{D})^2}} \tag{9}$$

the ratio of the sum of the squared errors in the predicted first differences of the normal-linear estimators, $\hat{\mathcal{D}}$, compared with the beta estimator, $\tilde{\mathcal{D}}$. Efficiency scores greater than 100 indicate that BMLE is superior to the approach being compared.

The efficiency comparisons in Table 2 indicate that BMLE is generally more efficient than normal-linear approaches, particularly with large samples and especially when the dependent variable has been transformed. The exceptions, again, appear to be for effects

Table 2 Efficiency of Beta MLE vs. alternatives

		Mean			*Variance*		
Model	*N*	*OLS*	OLS_{lt}	$Beta_{alt}$	*Normal*	$Normal_{lt}$	$Beta_{alt}$
Symmetrical	500	161.0	205.0	100.4	111.5	130.7	100.3
	100	100.8	129.2	99.8	114.6	130.1	100.7
	30	80.0	99.8	98.4	121.5	138.5	99.2
Asymmetrical	500	113.5	156.7	99.2	119.3	194.1	99.9
	100	103.1	115.0	99.4	115.4	133.6	100.0
	30	92.2	105.6	99.5	114.0	130.4	99.9
One-sided	500	117.6	178.3	98.0	118.9	244.9	97.7
	100	114.5	121.1	97.4	124.1	172.9	97.9
	30	105.0	100.1	98.4	138.5	139.8	98.4
Bimodal	500	111.8	1002.0	104.8	154.1	4678.7	99.7
	100	96.9	456.4	93.7	148.6	2059.8	100.4
	30	98.1	245.9	92.2	171.6	1155.9	98.7

Measures of efficiency were determined from the Monte Carlo trials and defined as $\text{Efficiency} = 100\frac{\sqrt{\sum_{i=1}^{1000}(\hat{\mathcal{D}}_i - \mathcal{D})^2}}{\sqrt{\sum_{i=1}^{1000}(\tilde{\mathcal{D}}_i - \mathcal{D})^2}}$. Values over 100 indicate the relative superiority of BMLE. The "lt" subscript signifies that the proportion was transformed using a logit transformation, and Beta_{alt} refers to the alternate approach to BMLE, as described in Eqs. (11)–(15).

upon the expected value with smaller samples. In these cases, OLS on untransformed variables is more efficient than the results from BMLE. The big difference in efficiency, however, comes with estimation of the covariates' effect upon the variance function, where BMLE is at least 10% more efficient than heteroskedastic normal approaches—much more so as the data move away from a unimodal, symmetrical distribution. Of course, these differences represent the worst case scenario, when we know the true data generating process to be the beta distribution. Nonetheless, differences of this magnitude can greatly affect our interpretation of the data.

Bias and efficiency are only two bases of comparison between BMLE and OLS. We are also generally concerned about the statistical inferences we make. As noted above, there is reason to believe that OLS will not produce the correct standard errors when the dependent variable is a proportion. For one thing, the data generation process linking the mean and variance suggests that the classical linear assumption of homoskedasticity may often be violated. Consequently, using OLS with proportion data may result in type I or type II errors. As one way of assessing how well the various estimation approaches reproduce the correct standard errors, I present measures of overconfidence, a measure that Beck and Katz (1995) use to examine the accuracy of standard errors.

$$\text{Overconfidence} = 100\frac{\sqrt{\sum_{i=1}^{1000}(\hat{\Psi}_i - \bar{\hat{\Psi}})^2}}{\sqrt{\sum_{i=1}^{1000}(s.e.(\hat{\Psi}_i))^2}} \tag{10}$$

Overconfidence scores greater than 100 indicate that the estimated standard errors underestimate the true variability of the parameters. Such overconfidence, of course, can lead researchers to commit type I errors, the rejection of a true null hypothesis. Likewise, overconfidence scores less than 100 indicate that the estimated standard errors overestimate the

Table 3 Overconfidence of Beta MLE vs. alternatives

Mean function

Model	*N*	*Beta*	*OLS*	OLS_{lt}	$Beta_{alt}$
Symmetrical	500	103.2	91.0	106.9	99.2
	100	101.3	90.3	103.4	98.8
	30	106.5	85.1	100.1	102.7
Asymmetrical	500	108.1	100.8	113.1	104.1
	100	103.2	96.1	104.5	102.1
	30	104.1	95.0	100.3	103.1
One-sided	500	101.4	108.9	114.8	99.3
	100	95.4	103.8	106.7	93.1
	30	102.6	106.6	101.3	107.5
Bimodal	500	100.6	90.6	116.9	94.1
	100	102.7	90.0	122.3	97.0
	30	102.7	92.8	106.0	97.3

Variance function

Model	*N*	*Beta*	*Normal*	$Normal_{lt}$	$Beta_{alt}$
Symmetrical	500	103.7	93.7	118.4	100.4
	100	99.3	95.0	111.6	99.7
	30	97.2	104.1	116.6	107.2
Asymmetrical	500	98.4	95.2	116.4	97.7
	100	99.7	101.4	115.8	101.0
	30	96.0	101.7	115.4	105.6
One-sided	500	101.7	146.6	152.3	97.7
	100	94.9	138.3	144.1	98.9
	30	100.8	159.8	135.8	109.1
Bimodal	500	100.5	85.6	184.1	96.7
	100	98.3	85.2	167.0	97.5
	30	100.3	96.1	146.2	102.1

Measures of overconfidence were determined from the Monte Carlo trials and defined as Overconfidence $= 100\frac{\sqrt{\sum_{i=1}^{1000}(\hat{\Psi}_i-\bar{\hat{\Psi}})^2}}{\sqrt{\sum_{i=1}^{1000}(s.e.(\hat{\Psi}_i))^2}}$. Values greater than 100 indicate that the approach underestimates the standard errors, while numbers under 100 indicate an overestimate of the sampling variability. The "lt" subscript signifies that the proportion was transformed using a logit transformation, and $Beta_{alt}$ refers to the alternate approach to BMLE, as described in Eqs. (11)–(15).

true sampling variability of the parameters, leading researchers to commit type II errors, the acceptance of a false null hypothesis.

The overconfidence scores from the Monte Carlo simulations are presented in Table 3. These results indicate that BMLE generally outperforms other estimation approaches across the different distributions and sample sizes. While BMLE provides reasonably accurate estimates of the standard errors for different distributions of the data and different sample sizes, normal methods on untransformed variables tend to overstate the size of the standard errors, while normal methods on logistically transformed variables tend to understate the standard errors. Across all three means of evaluating the methods, BMLE appears to be a reliable way of estimating proportions data across a variety of sample sizes and distributions of the data and generally dominates normal estimation approaches, especially when we are concerned with the variance function.

3 An Alternative Approach to Beta Estimation

The Monte Carlo results indicate that maximum likelihood estimation using the beta distribution can be a very attractive approach for proportions data, but interpreting results from the distribution is not as straightforward as for normal-linear models. As researchers, we are rarely interested in estimating a covariate's relationship to a shape parameter. Rather, we want to know things like whether or not bureaucrats who are members of minority social groups are, on average, more or less likely to determine that other minority members are eligible for government assistance. With this goal, it is not obvious how researchers using BMLE should estimate the substantive and statistical significance of the effects of some covariates upon the mean and variance of a dependent variable because there is no one-to-one correspondence of an explanatory variable with the mean or the variance from which to derive inferential statistics. Perhaps for this reason, among others, many researchers fall back on the normal distribution when examining such variables.

Estimating two parameters for a single covariate to determine its relationship to the expected value of the dependent variable can lead to some interpretational difficulties. Brehm and Gates (1993, p. 569) argue that using the beta distribution is preferable to OLS because

> the two-parameter feature of the beta distribution adds yet another desirable property for the analysis of compliance: we can model factors leading toward compliance separate from those that lead toward defection. In OLS we are confined to two parameters: one describes the mean; the other the variance.

Yet a difficulty arises in their analysis when interpreting the individual coefficients in both the α and the β parameters for the dummy variable for Chicago police officers because both coefficients are positive and statistically significant. Does this mean that officers in Chicago are statistically more likely both to shirk and to work than police officers in other cities? The difficulty, here, is that one coefficient indicates a "push" away from compliance, while the other one indicates a "push" away from defection. Fundamentally, we are interested in knowing whether or not Chicago officers are more or less likely to do their job. Brehm and Gates (1993, p. 574) answer this, using the method of first differences, estimating that "an officer in Chicago in 1966 is likely to shirk 7% more than officers in other cities, controlling for other variables."

This relationship, however, is not immediately obvious from the estimates. In this case, it seems reasonable that a covariate having a greater effect upon α than upon β, holding everything else constant, would produce a higher mean value (i.e., more shirking) because as α increases relative to β, the expected value of y increases. Yet this intuition does not work for "functional dislike" because the first difference predicts greater shirking even though the effect of "dislike" upon β is greater than that upon α.

A second difficulty in the standard approach comes in determining the statistical significance of any variable. The variance–covariance matrix provides only the standard errors for the parameters' relationship to the shape parameters. We are generally interested, however, in the effect of a variable upon the expected value of the dependent variable. While Brehm and Gates (1993, p. 570) point out that the mean of the distribution might not be an interesting point, they provide very little discussion of the shape of the distribution implied by their estimates and instead focus more on the expected amount of shirking. The standard approach to BMLE, however, does not provide a ready means of statistical inference about a variable's effect upon the expected value of the dependent variable. For instance, Brehm and Gates's results indicate that "satisfaction with supervisor" has a negative and statistically

significant effect upon both the α (compliance) and the β (defection) shape parameters but *increases* the amount of shirking by 3%. If the .03 effect of increased satisfaction upon shirking is statistically significant, that would be a counterintuitive finding, but Brehm and Gates do not provide that information in their table. In contrast, we know that the OLS result is negative and statistically significant from the estimated coefficient and standard error. It seems likely, therefore, that the effect of satisfaction does not have a statistically significant effect even though both parameters have a significant effect upon their respective shape parameters.

Two alternatives exist for assessing the statistical significance of variables upon substantively interesting quantities: a simulation method (King et al. 2000) and the Wald test. Both options involve the researcher choosing a substantively relevant first difference and calculating a test statistic to determine with what probability that difference is statistically different from some other substantively important value, usually 0. A presentational difficulty with this approach is that substantively interesting differences may not be the same for each variable. One might take the first differences set 1 standard deviation above and below their means for some variables, but then reasonably determine the first difference between the minimum and the maximum values for dummy variables—although they do not say this explicitly, this is presumably what Brehm and Gates do with their Chicago police officer example. In the alternate approach, significance levels are easily determined from the estimated coefficients and standard errors.

A further criticism of the standard approach is that it 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 a mean or the dispersion of the dependent variable. Even in cases where a variable affects both the mean and the variance, it might do so only through the mean, not the dispersion. If so, estimating unnecessary parameters means a loss of degrees of freedom that may be particularly damaging in small samples. While we may have some intuition or interest about whether or not a variable has an effect upon the dispersion, it is more difficult to think of the intuition that would lead someone to argue that a variable affects α but not β.

An alternative approach to BMLE is one that estimates the effect of an independent variable upon the mean and the dispersion of the dependent variable by specifying α and β in terms of the mean and dispersion, 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 dispersion. One specification is

$$E(Y) = \frac{\exp(\mathbf{X\Psi})}{1 + \exp(\mathbf{X\Psi})} \tag{11}$$

$$\phi_y = \exp(\mathbf{Z\Phi}) \tag{12}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are (potentially distinct) matrices of independent variables and $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are parameter vectors for the effects of $\mathbf{X}$ and $\mathbf{Z}$ upon the expected value and dispersion of y, respectively. The form of Eq. (11) reflects the constraint that the expected value of Y is bounded by 0 and 1. Similarly, ϕ_y must be positive because it is the sum of two positive quantities, α and β. Estimating the dispersion, rather than the variance, reflects the relationship between the mean and the variance. Recalling Eq. (3),

$$\text{Var}(Y) = E(Y)[1 - E(Y)]/(\phi_y + 1) \tag{13}$$

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

$$\alpha = \frac{[E(Y)]^2[1 - E(Y)]}{\text{Var}(Y)} - E(Y) \tag{14}$$

$$\beta = \frac{E(Y)[1 - E(Y)]^2}{\text{Var}(Y)} - [1 - E(Y)] \tag{15}$$

and these values are plugged into the likelihood function [Eq. (6)].

This approach has several desirable qualities. First, it is easy to interpret the covariates' effect upon the expected proportion and make statistical inferences about those effects. The same goes for the effect of covariates upon the dispersion. If one is interested in the effect upon the variance, one uses a simulation or Wald approach to determine statistical significance, but that makes the researcher no worse off than under the standard approach. Second, if a researcher does not have any reason to believe that a variable has an effect upon the dispersion of the distribution, the effect of that variable upon the dispersion can be constrained to 0, making the variable's effect upon the variance entirely a function of its effect upon the mean and saving degrees of freedom. Finally, a researcher might be interested theoretically if a variable contributes to the variance beyond its effect upon the mean. This alternative approach allows us to separate out those effects and determine their significance.

To insure that the alternative approach to estimating the beta likelihood is as good as the standard approach, I also ran the Monte Carlo simulations described above with this alternative way. Even though the alternate approach allows for constant dispersion, the simulations for the alternative approach specify both the mean and the dispersion as a function of a constant and the sole covariate for purposes of comparability with the standard approach. The results in the last columns in Tables 1–3 show that this alternative approach to BMLE produces results that are similar to those produced by the standard approach and remain generally superior to those produced by linear-normal approaches. The Monte Carlo results, therefore, provide support for considering BMLE, in both its standard and its alternate forms, but a real question is the difference it makes with actual data.

4 Empirical Examples

There are many instances where political science researchers use proportions data in their analysis. In this section, I compare the results from several published articles where OLS is used and compare these results with those from BMLE. The purpose of this section is not to show that BMLE must always be used when estimating proportions. In fact, I cannot offer any hard and fast guidelines to indicate when BMLE will be better than OLS. Rather, I use these examples to help support the conclusions of the Monte Carlo results and provide readers with some examples to follow when trying to determine whether or not BMLE is appropriate for their particular analysis. The results in this section focus upon the comparison of the alternate approach with linear approaches using both untransformed and transformed proportions. When using the standard approach to BMLE produces different results from the alternate approach, I, however, point out these differences.

In one article that uses proportions, Selden et al. (1998) examine the factors that influence the determination of eligibility for the Rural Housing Loans program of the Farmers Home Administration. In particular, the researchers wanted to know whether or not bureaucrats socialized to see themselves in a "minority representative role" were more likely to grant minority applicants eligibility for this program. Their analysis provides strong support for

Table 4 OLS vs. Beta estimation

	Percent FmHA approvals			
Variable	*OLS*	$\mathcal{D}$	*Beta*	$\mathcal{D}$
Expected value				
Constant	−.386**		−4.945	
	(−3.28)		(−7.52)	
Bureaucratic role	−0.097	−.005	2.145	.028
	(−0.17)		(0.73)	
Minority rep. role	0.027**	.323	0.138**	.380
	(10.32)		(9.29)	
Minority hardship index	0.483**	.090	2.668**	.120
	(2.96)		(3.12)	
Minority bureaucrat	0.011	.010	0.084	.017
	(0.30)		(0.41)	
Dispersion				
Constant			1.026**	
			(10.66)	
N	182		182	
Mean squared error	.037		.037	

Source: Selden et al. 1998.
T-statistics in parentheses. First differences calculated for each variable one standard deviation above and below its mean value while holding all other variables at their means. ** $p < .05$.

this hypothesis. For the current paper, the relevant question, however, is whether or not their conclusions were influenced by using OLS instead of BMLE.

Before jumping right into BMLE, the first thing a researcher might want to consider is whether or not the classic OLS assumptions are violated. Two considerations are whether or not the errors are normally distributed and whether or not the errors are homoskedastic. A Cook–Weisberg test cannot reject the null hypothesis of homoskedasticity, but a Wilks–Shapiro test allows one to reject the hypothesis of normally distributed errors. This indicates that Selden et al.'s hypothesis tests are misleading because they use homoskedastic OLS. In this case, it is reasonable to consider BMLE. Compared with the results from the BMLE, one would not greatly change the substantive conclusions that the authors reached from their analysis (Table 4).[4] There are, however, several noteworthy points. First, there is very little difference between the mean squared errors of the estimates whether using OLS or BMLE. Second, the estimated effects of a minority representative role and minority hardship are 18 and 30% greater, respectively, when estimated under BMLE than under OLS. Third, consistent with the Monte Carlo results, BMLE produces a smaller confidence interval. Finally, the beta model's estimated shape parameters suggest that the beta distribution is more likely to be the data generating process, as the Kolmogorov–Smirnov test cannot reject the null hypothesis that the data are distributed Beta(1.62, 1.22).

In spite of the evidence that indicates that BMLE provides better estimates than OLS, it is clear that the authors of the original study did not reach any dramatically different

[4]The lead author graciously provided me with the data and code for replication. For whatever reason, the code did not replicate the data reported in the article. I was able, however, to come close to the published results, missing only two observations from the number reported in the article. For my purposes here, these minor differences are unimportant.

conclusions than they would have had they used BMLE. While the data are skewed toward the higher end of the distribution, fewer than 10% of the observations grant eligibility in 90% or more of the reviewed cases, and none of the predicted values from the OLS estimates fall outside of the 0–1 domain. I should also note that OLS on logistically transformed proportions would not change the interpretation. Fundamentally, the advantage of BMLE in this case is that it produces more accurate and more precise estimates.

Keiser and Soss (1998) use proportion data to examine the factors that influence when bureaucrats grant to welfare claimants "good-cause" exemptions from helping locate missing parents who owe child support. In their analysis, Keiser and Soss consider the characteristics of the bureaucracy, the task environment, and the political environment as influencing the proportion of good-cause exemptions granted. They find that characteristics of the bureaucracy, such as the bureaucratic mission to recapture revenue, are significantly related to granting exemptions. They also find that the political environment is related, as states with unified Democratic control are more willing to grant these exemptions. Contrary to their expectations, they do not find that aspects of the task environment, such as a state's need for financial resources, have a significant effect upon bureaucrats' granting of good-cause exemptions.

The data of Keiser and Soss have some different characteristics than those analyzed by Selden et al. One cannot, for instance, reject the null hypothesis that the errors from the OLS model are normal. On the other hand, the modal point in these data occurs at the upper boundary, and one can reject the null hypothesis that the dependent variable comes from a normal distribution. A reanalysis of the data using BMLE also shows that some of Keiser and Soss's conclusions are affected by how the model is estimated. A state's economic need for resources *is* statistically significantly related to the bureaucracy's willingness to grant good-cause exemptions, while the hypothesis that Democratic governments are significantly more likely to grant such exemptions is no longer supported (Table 5).

The BMLE results provide a different conclusion about the role of economic capacity in the bureaucracy's willingness to grant good-cause exemptions, as well as the influence of Democratic governments. There is some reason to doubt the BMLE estimates with constant dispersion because the predicted shape parameters describe a one-sided distribution, Beta(2.72, 0.96) that is rejected as describing the data by a Kolmogorov–Smirnov test. The BMLE estimates, however, indicate that bureaucratic values, the number of applications, and the presence of unified Democratic government significantly affect the variance. While Keiser and Soss assume homoskedasticity, there is evidence that the data should be modeled as heteroskedastic.

Further analysis shows that modeling the dispersion as a function of both bureaucratic values and the number of applications provides a significant improvement in the fit of the model, $\chi^2_2 = 26.15$. This less constrained model indicates that both of the environmental factors of economic capacity and unified Democratic government influence decisions to grant good-cause exemptions. Modeling the dispersion function also improves the model fit, as measured by the mean squared error, making it only slightly larger than the homoskedastic OLS estimates. Finally, a Kolmogorov–Smirnov test cannot reject at $p < .01$ the null hypothesis that the data come from a beta distribution with the estimated shape parameters, $\alpha = 3.11$, $\beta = 1.16$. These parameters are also quite close to the estimates of α and β one would derive from plugging the sample mean and variance into Eqs. (14) and (15), $\alpha = 3.80$, $\beta = 1.24$. A Kolmogorov–Smirnov test cannot reject at $p < .05$ the null hypothesis that the data were produced by a beta distribution with this shape.

By taking heterogeneity across observations into account, BMLE also provides some insight beyond what the homoskedastic model reveals. Both the standard and alternate

Table 5 OLS vs. Beta estimation

	Percent good cause exemptions					
Variable	*OLS*	$\mathcal{D}$	*BMLE*	$\mathcal{D}$	*BMLE*	$\mathcal{D}$
Expected value						
Constant	.761**		1.16		1.79	
	(3.43)		(0.86)		(1.38)	
Bureaucratic resources	−.003	−.014	−.010	−.009	.044	.040
	(−0.40)		(−.023)		(1.28)	
Bureaucratic values	−.025**	−.066	−.199**	−.104	−.125**	−.067
	(−2.22)		(−3.05)		(−1.98)	
Number of applications	−.012**	−.085	−.097**	−.133	−.072**	−.102
	(−2.77)		(−4.08)		(−5.24)	
Demand	−.002	−.002	−.015	−.003	−.270	−.051
	(−0.05)		(−0.07)		(−1.23)	
Economic capacity	.132	.051	1.406**	.105	1.310**	.101
	(1.42)		(2.43)		(2.53)	
Democratic government	.075**	.073	.245	.046	.312**	.060
	(2.28)		(1.23)		(1.74)	
1991	.007	.006	.009	.002	.010	.018
	(0.19)		(0.05)		(0.59)	
1992	−.024	−.022	.007	.001	.117	.022
	(−0.66)		(0.04)		(0.70)	
Dispersion						
Constant			1.306**		−0.407	
			(11.47)		(−0.61)	
Bureaucratic values					0.215**	
					(1.97)	
Number of applications					0.183**	
					(5.25)	
N	140		140		140	
Mean squared error	.027		.029		.028	

Source: Keiser and Soss 1998.
T-statistics in parenthesis. First differences calculated for each variable one standard deviation above and below its mean value while holding all other variables at their means. ** $p < .05$ (one-tailed test).

approaches to BMLE suggest that the variance of good-cause exemptions granted decreases as more applications are made. This might indicate that bureaucracies that are faced with a greater number of applications have to adopt or adhere to more highly standardized procedures for determining merit, thereby reducing bureaucratic discretion. That this effect is not simply a function of the mean, but works through the dispersion parameter, bolsters this interpretation.[5] In contrast, while the variance is also smaller in states with unified Democratic government, this effect works mostly through the mean, as including "Democratic government" in the dispersion function did not significantly improve the model fit (see the web appendix). Together with the positive effect of Democratic government on the expected value of exemptions granted, this seems to indicate the adoption of operating procedures

[5]Remember, the coefficients in the dispersion parameter are inversely related to the variance.

that are the result of political influence. Bureaucrats may come to believe, or be instructed, that political appointees want exemptions granted and so adopt a routinized compliance with superiors' preferences.

Given these different results, it is worth considering which estimation procedure is more accurate. Obviously, we do not know which, if either, distribution produced the observed data. The normal results have a slightly smaller mean squared error and do not produce any predicted values outside of the permissible range. On the other hand, the predicted values from the beta estimations appear to do a better job of replicating the true underlying distribution. The OLS results' advantage in reducing the mean squared error is seen with the concentration of predicted values around the mean of the actual values. Furthermore, the greater precision in the BMLE estimates is consistent with the Monte Carlo results showing that OLS tends to overestimate the standard errors in samples of this size as well as the results showing greater efficiency for BMLE in estimating covariates' effects upon the variance. Finally, the results from the BMLE analysis yield some plausible and compelling insights that are overlooked by the normal analysis.

Haider-Markel and Meier (1996) examine the factors that affect the proportion of a state's population covered by laws dealing with gay and lesbian rights. They investigate the hypotheses that the amount of coverage can be explained by interest-group factors and "morality" factors. They conclude that interest-group factors, including interest-group membership and wealth, the presence of gay and lesbian businesses, and elite support, explain the presence of antidiscrimination laws. There is reason to believe, however, that their analysis overstates the degree to which some indicators of interest-group activity affects the presence of antidiscrimination laws in certain policy domains.

Haider-Markel and Meier are aware that they are using heavily skewed data for their analysis, and for this reason, they transform their variables by taking the natural log of their dependent measures and many of the indicators of interest-group activity, such as interest-group membership and wealth, the presence of gay and lesbian businesses, and elite support. Analysis of the regression residuals shows that one cannot reject the hypothesis of normal errors, while the same is clearly not true for the untransformed data. In this sense, the transformation obviates a violation of the classic OLS assumptions. The Monte Carlo results above indicate, however, that transforming the data may be problematic because the standard errors underestimate the sampling variability. The conclusions from those results also hold for a simple log transformation (results in the web appendix).

Haider-Markel and Meier analyze a variety of antidiscrimination laws, but I provide a fairly detailed analysis of one area, laws against discrimination in private employment. This issue is interesting because the distribution of coverage of these laws is clearly nonnormal, with observations concentrated heavily at 0 and 1, albeit more heavily 0 (Fig. 2(a)). One can also reject the assumption of normality for the residuals for both the models using untransformed and transformed data. So unlike the Selden et al. and the Keiser and Soss examples, Haider-Markel and Meier provide an opportunity to see the benefits of BMLE when the data are clearly nonnormal.

There are a few others things to note about these data. First, the null hypothesis of homoskedasticity is rejected for both the transformed and the untransformed models, $p < .05$. Figures 2(c) and 2(d) also shows that, unlike the Selden et al. and the Keiser and Soss data, 28 and 14% of predicted values from both the untransformed and the transformed models, respectively, fall outside of the permissible 0–1 domain. The predicted values from the untransformed data look nothing like the actual distribution, and the transformed data only approach the actual data by overpredicting the values at the upper bound. All of this suggests that a normal-linear approach is probably faulty.

Both BMLE and OLS produce substantively comparable results for most of the covariates' effect upon the mean (Table 6). Unlike the results for the OLS model using a log-transformed dependent variable, neither of the variables for task force membership or wealth is significant. This is also consistent with the Monte Carlo results for bimodally distributed data that indicate that OLS with transformed variables often underestimates the standard errors. Figure 2(d) also shows that while the log-transformed model produces predicted values that accurately estimate observations around the lower mode, this model misses the values at the upper bound.

Between the BMLE and the OLS on the untransformed proportion, choosing the better model is reasonably easy. Although BMLE has a 14% greater mean squared error prediction when the BMLE is estimated with constant dispersion, these differences are significantly

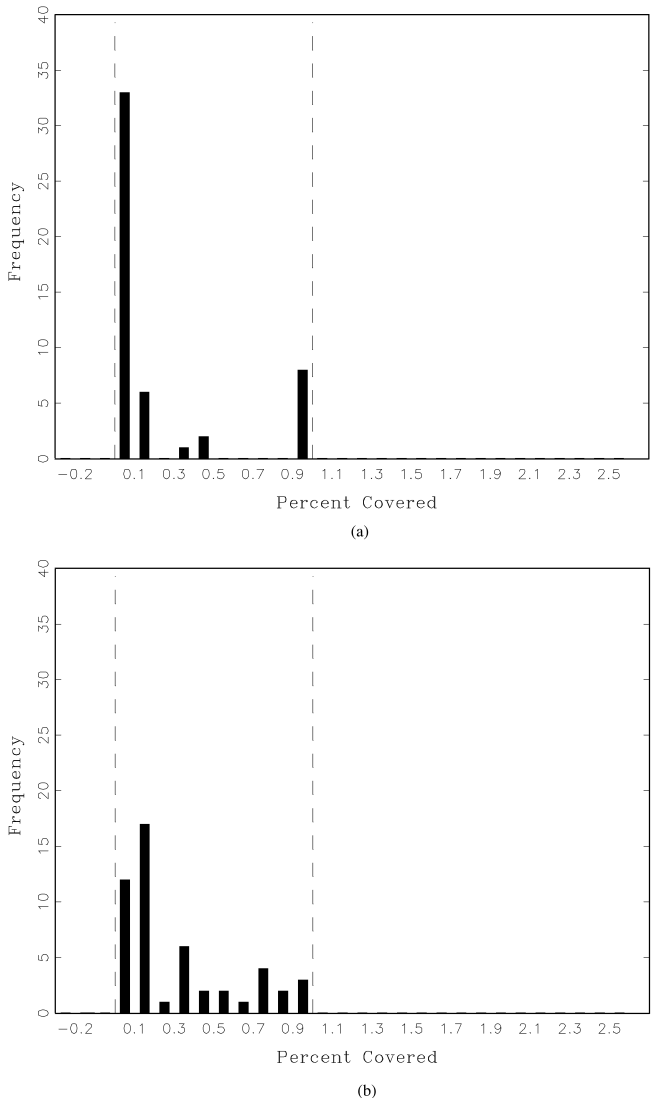
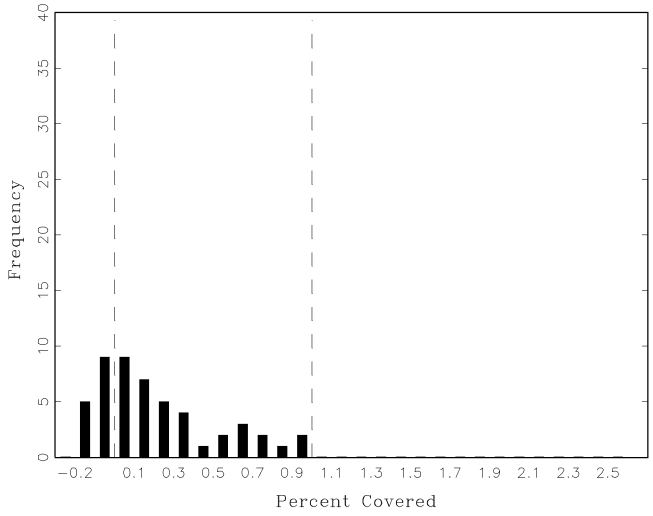
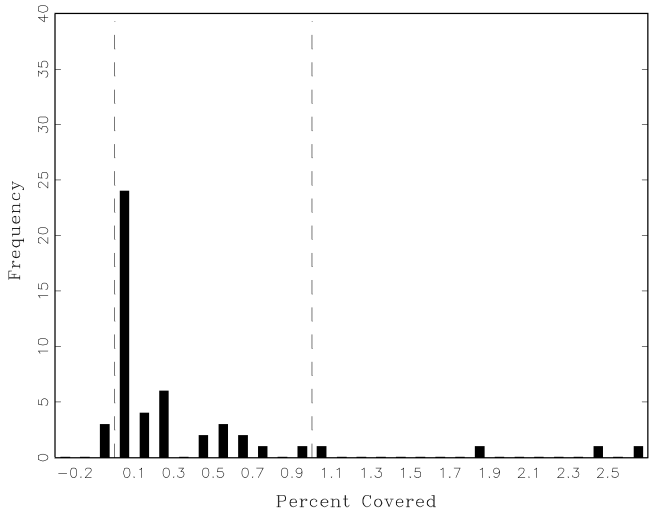


Fig. 2 Haider–Markel and Meier private employment laws data: (a) actual values, (b) predicted values – Beta, (c) predicted values – OLS, and (d) predicted values – OLS (logged).



(c)



(d)

Fig. 2 *Continued.*

reduced if the dispersion is modeled as a function of elite support and other civil rights laws. At this point, the mean squared error predictions for BMLE are 25% smaller, and the BMLE model may be said to fit the data better because of the high proportion of predictions that fall outside of the permissible bounds. The BMLE model with heterogeneous dispersion also captures the distribution of actual values better (Fig. 3). While the constant-dispersion BMLE model predicted the modal range of points between 0.1 and 0.2, BMLE with heterogeneous dispersion correctly predicts that most of the observations are between 0 and 0.1. Finally, because the null hypothesis of normality of the OLS residuals in the private employment rights model can be rejected, we might place greater trust in the inferential statistics from the BMLE model.

As with the Keiser and Soss data, estimating the gay rights data using BMLE produces some insights missed by linear models. Without estimating dispersion coefficients for early

Table 6 OLS vs. Beta estimation—Haider-Markel and Meier

	Percent covered by anti-discrimination laws					
Variable	*OLS*	$\mathcal{D}$	*BMLE*	$\mathcal{D}$	*BMLE*	$\mathcal{D}$
Expected value						
Constant	−0.983**		−6.67**		−11.16**	
	(−4.23)		(−4.81)		(−8.02)	
Task force	0.014	.138	0.103	.192	0.209**	.130
	(1.50)		(1.63)		(3.61)	
Members civil rights	0.021	.063	0.177	.103	0.464**	.086
	(0.77)		(1.29)		(3.34)	
Laws elite support	0.976**	.340	4.93**	.327	10.717**	.292
	(4.03)		(3.83)		(7.78)	
Salience	−.045	−.097	−0.179	−.075	−.046	−.006
	(−1.22)		(−1.01)		(−0.32)	
College graduates	0.035**	.205	0.134	.154	0.029	.010
	(2.34)		(1.72)		(0.40)	
Wisconsin	0.830**	.234	3.332**	.182	6.062**	.109
	(3.60)		(2.83)		(5.29)	
Louisiana	0.168	.048	1.624	.089	3.672**	.063
	(0.74)		(1.29)		(6.18)	
Delaware	−0.218	−.062	−1.265	−.070	−1.057	−.018
	(−0.96)		(−1.13)		(−1.02)	
Rhode island	−0.536**	−.152	−3.046**	−.167	−2.269	−.038
	(−2.23)		(−2.50)		(−1.58)	
Dispersion						
Constant			0.105		8.231**	
			(0.42)		(7.27)	
Civil rights					−0.386**	
					(−2.77)	
Laws elite support					−9.771**	
					(−6.24)	
N	50		50		50	
Mean squared error	.040		.045		.032	

Source: Haider-Markel and Meier1996.

T-statistics in parentheses. First differences calculated for each variable one standard deviation above and below its mean value while holding all other variables at their means. $^{**}p < .05$ (two-tailed test).

civil rights laws and elite support, only elite support has a statistically significant effect upon the variance. But when the dispersion is modeled as a function of both, civil rights laws also influence the variance in the coverage of antidiscriminatory laws. One might interpret these results as implying that while the absence of early civil rights laws reduces the variance through lower expected legislative support for laws against gay discrimination, the effect of the absence of civil rights laws reducing the variance through the dispersion may suggest something related to the absence of both early civil rights laws and gay rights laws as reflecting a traditional conservatism or reluctance to innovate in areas of progressive public policy beyond civil rights. Also note that the effect of task force membership is again significantly related to the coverage by civil rights laws, as one might expect and as Heider-Markel and Meier found in their analysis—albeit through an inferior approach. Finally, this

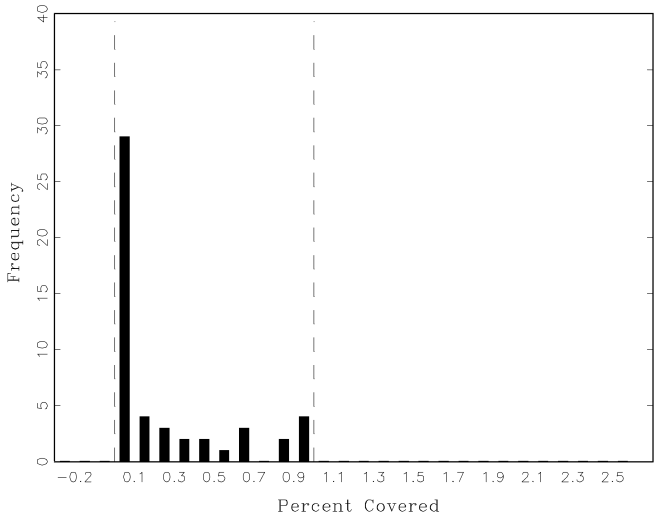


Fig. 3 Haider – Markel and Meier data: predicted values – heterogenous beta.

example illustrates a value of the alternate approach to BMLE. Given that Haider-Markel and Meier's data contain only 50 observations, estimating a full set of parameters with the standard approaches taxes the data quite heavily. The alternate approach, in contrast, saves a significant number of degrees of freedom and allows for estimation of the variance even for variables not included in the dispersion function.

5 Conclusions

The results from both the Monte Carlo and the empirical examples indicate that a BMLE approach to the analysis of proportions provides some significant advantages over normal estimation in some cases. The major points can be summed up as follows.

- BMLE does not always improve the model fit, but when the underlying data generating process produces variables according to the beta distribution, BMLE can produce significantly more accurate estimates of the effects of covariates upon the expected value.
- When the proportions data are generated according to the beta distribution, it may be unreasonable to assume homoskedasticity. And because even heteroskedastic normal models are inefficient, modeling the heteroskedasticity using BMLE may yield insights that would be missed or less subtly understood otherwise.
- Estimating BMLE with the alternate approach that I propose in this paper can provide many of the desirable features of OLS (such as relatively easily interpretable significance of effects) and the standard approach (better description of the data generating process).

For these reasons, researchers have much to gain by utilizing the beta distribution when dealing with proportions data.

This is not to say that I believe that all analysis of political science proportions data must be done using BMLE. I analyzed other data where BMLE made absolutely no difference in the interpretation of the results. These data, which were used by Atkeson (1998) and Lowery and Gray (1998), both had no values at either boundary and were fairly symmetrically

distributed. In these circumstances, the analysis presented in this paper does not suggest that BMLE will necessarily provide much insight beyond what we get from normal methods. The more important point of this paper is that, as researchers, we never truly know the data generating process. If we blithely assume that all of our models with proportions are generated by a normal distribution, we will get reasonably accurate estimates in some cases, even when our assumption is incorrect. We can, however, also greatly misunderstand the phenomenon we are studying because we have not looked carefully enough at the data, including the results we get from using BMLE.

References

Achen, Christopher H., and W. Phillips Shively. 1995. *Cross-Level Inference*. Chicago: University of Chicago Press.

Alvarez, R. Michael, and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science* 39(4):1026–1055.

Alvarez, R. Michael, and John Brehm. 1997. "Are Americans Ambivalent Toward Racial Policies?" *American Journal of Political Science* 41(2):345–374.

Atkeson, Lonna Rae. 1998. "Divisive Primaries and General Election Outcomes: Another Look at Presidential Campaigns." *American Journal of Political Science* 42(1):256–271.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series-Cross-Section Data in Comparative Politics." *American Political Science Review* 89:634–647.

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):555–581.

Delli, Carpini, Michael X., and Scott Keeter. 1996. *What American Know About Politics and Why It Matters*. New Haven, CT: Yale University Press.

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

Gentle, James E. 1998. *Random Number Generation and Monte Carlo Methods*. New York: Springer-Verlag.

Haider-Markel, Donald P., and Kenneth J. Meier. 1996. "The Politics of Gay and Lesbian Rights: Expanding the Scope of the Conflict." *Journal of Politics* 58(2):332–349.

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

Keiser, Lael R., and Joe Soss. 1998. "With Good Cause: Bureaucratic Discretion and the Politics of Child Support Enforcement." *American Journal of Political Science* 42(4):1133–1156.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):347–361.

Lowery, David, and Virginia Gray. 1998. "The Dominance of Institutions in Interest Representation: A Test of Seven Explanations." *American Journal of Political Science* 42(1):231–255.

Mebane, Walter R., Jr. 2000. "Coordination, Moderation, and Institutional Balancing in American Presidential and House Elections." *American Political Science Review* 94(1):37–58.

Palmquist, Bradley. 1999. "Analysis of Proportions Data." Paper presented at the 1999 Annual Meeting of the Political Methodology Society, College Station, TX.

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.

Schram, Stanford, Lawrence Nitz, and Gary Krueger. 1998. "Without Cause or Effect: Reconsidering Welfare Migration as a Policy Problem." *American Journal of Political Science* 42(1):210–231.

Selden, Sally Coleman, Jeffrey L. Brudney, and J. Edward Kellough 1998. "Bureaucracy as a Representative Institution: Toward a Reconciliation of Bureaucratic Government and Democratic Theory." *American Journal of Political Science* 42(3):717–744.