Mixed and Mixture Regression Models for Continuous Bounded Responses Using the Beta Distribution

Jay Verkuilen and Michael Smithson

The online version of this article can be found at:

Published on behalf of

American Educational
Research Association

American Educational Research Association

and
®SAGE

http://www.sagepublications.com

Additional services and information for *Journal of Educational and Behavioral Statistics* can be found at:

Email Alerts: http://jebs.aera.net/alerts

Subscriptions: http://jebs.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions

Mixed and Mixture Regression Models for Continuous Bounded Responses Using the Beta Distribution

Jay Verkuilen
City University of New York

Michael Smithson
The Australian National University

Doubly bounded continuous data are common in the social and behavioral sciences. Examples include judged probabilities, confidence ratings, derived proportions such as percent time on task, and bounded scale scores. Dependent variables of this kind are often difficult to analyze using normal theory models because their distributions may be quite poorly modeled by the normal distribution. The authors extend the beta-distributed generalized linear model (GLM) proposed in Smithson and Verkuilen (2006) to discrete and continuous mixtures of beta distributions, which enables modeling dependent data structures commonly found in real settings. The authors discuss estimation using both deterministic marginal maximum likelihood and stochastic Markov chain Monte Carlo (MCMC) methods. The results are illustrated using three data sets from cognitive psychology experiments.

Keywords: *beta distribution; general linear model; mixed model; mixture model*

1. Introduction

Continuous dependent variables with scales bounded at both ends are fairly common throughout the social and behavioral sciences. Examples from psychology are the percentage of time attending one stimulus versus another as in habituation studies, subjective probability or confidence ratings common in cognitive research, or total scores on a symptom questionnaire applied in a community setting. In behavioral economics examples include most utility scales (usually assumed to be bounded at both ends) and proportional measures such as allocations to investments in a portfolio and leverage (the ratio of debt to assets plus

1

debt). These continuous doubly bounded dependent variables frequently present special problems for modelers. Chief among these are uncorrectable skew and heteroscedasticity, both often arising from the fact that as the mean response moves toward either scale boundary, the variance tends to decrease and skew tends to increase. In short, in a bounded response space, the moments of the distribution are inextricably linked. Furthermore, the closer one is to a boundary the stronger this linkage is.

Several authors (Kieschnick & McCullogh, 2003; Paolino, 2001; Smithson & Verkuilen, 2006) have presented arguments that traditional methods such as regression-linearizing and variance-stabilizing transformations or robust estimators are not always appropriate for modeling this kind of limited dependent variable. They propose an alternative in the form of using the beta distribution to model such variables. The beta distribution is a well-known member of the exponential family of distributions for which the regularity conditions that help ensure maximum likelihood estimates exist and are well defined. Nevertheless, the literature on generalized linear models (GLMs) for beta-distributed dependent variables (i.e., beta regression) is sparse. In the otherwise exhaustive *Handbook of Beta Distributions*, there is no mention of beta regression (Gupta & Nadarajah, 2004), although there is a sizeable parallel literature on beta-binomial regression (an early instance is Crowder, 1978), suited to modeling proportions of counts where the beta distribution is used as a hierarchically specified random effect to address overdispersion.

An early example of beta regression is Brehm and Gates's (1993) model of police compliance with supervision. However, Paolino (2001) was the first to employ the mean-precision parameterization of the beta distribution that greatly simplifies interpretation. Apparently independently, Ferrari and Cribari-Neto (2004) derived a similar beta regression model that recently has been implemented in the SAS GLIMMIX procedure (SAS, 2008), and likewise Kieschnick and McCullogh (2003) compared the performance of a beta regression model for proportions, as employed in economics and finance research, with several alternatives and concluded that it is often the best option. Noël and Dauvier (2007) and Noël (2008) have presented unfolding and dominance item-response models for continuous doubly bounded scale items based on the beta distribution.

Some of the above-mentioned versions of beta regression treat the precision parameter as a nuisance. Recently Smithson and Verkuilen (2006) followed and extended Paolino's model by explicitly modeling precision (and therefore dispersion) as well as the mean response, utilizing the extended GLM framework for joint modeling of means and dispersions described in Chapter 10 of McCullagh and Nelder (1989) and developed by Smyth (1989) for random variables in the exponential family of distributions. There are good reasons for regarding the modeling of dispersion as important in its own right. We concur with Carroll's (2003) observation that ignoring variance structure or treating it as simply a nuisance not only can lead to inefficient estimation but to misleading conclusions.

2

Smithson and Verkuilen (2006) presented real-world examples of this and the present article includes additional illustrations.

Likelihood maximization has been the dominant approach in estimating beta regression models (e.g., Paolino, 2001; Smithson & Verkuilen, 2006), with Vasconcellos and Cribari-Neto (2005) and Ospina, Cribari-Neto, and Vasconcellos (2006) proposing bias-correcting adjustments. Epsinheira, Ferrari, and Cribari-Neto (2008a, 2008b) explore alternative weighted standardized residuals and influence diagnostics for the Ferrari-Cribari-Neto beta GLM. Buckley (2002) implemented the Paolino model in a Bayesian Markov chain Monte Carlo (MCMC) procedure, and Smithson and Verkuilen (2006) also did so with their version.

A number of gaps remain in our knowledge about beta regression models. Chief among them is the absence of methods for handling dependent observations. In this article, we develop and explore generalized linear mixed models (GLMMs) and related dependent-observation models for beta-distributed dependent variables. Justifications for GLMMs as a way of handling dependencies are well known. The dependencies among observations may arise naturally from effects such as autocorrelation over time, within-subject correlations in repeated measures experiments, or clumping due to within-cluster homogeneity. We begin by reprising the independent-observations beta regression model and then developing an overall framework for beta GLMMs, including both mixed and mixture distribution models. A simulation example is presented, comparing maximum likelihood and Bayesian MCMC estimations, along with a discussion of variance partition issues and goodness of fit. Three examples of applications to real data are provided.

2. Regression With the Beta Distribution

2.1. Independent Observations

This section summarizes a longer discussion found in Smithson and Verkuilen (2006). Let $y \sim \text{Beta}(\alpha, \beta) \in (0, 1)$, where $\alpha, \beta > 0$ are shape parameters.[1] The probability density function (PDF) is

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

where $\Gamma(\cdot)$ is the complete gamma function. The two shape parameters pull density toward 1 (α) or 0 (β). Standard results show that

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

and

3

$$\text{var}(y) = \frac{\text{E}(y)(1 - \text{E}(y))}{\alpha + \beta + 1}. \tag{3}$$

The standard parameterization of the beta is most useful when it is used as a prior, where the parameters correspond to degrees of freedom. However, it is inconvenient for a regression model. Let $\mu = \text{E}(y)$ and precision parameter $\phi = \alpha + \beta$, which can be inverted to show $\alpha = \phi\mu$ and $\beta = \phi(1 - \mu)$. Upon reparameterization,

$$f(y, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu)\Gamma(\phi(1 - \mu))} y^{\phi\mu - 1} (1 - y)^{\phi(1 - \mu) - 1}. \tag{4}$$

We use the mean-precision parameterization henceforth, denoting such a distribution as $\text{Beta}(\phi\mu, \phi(1 - \mu))$. Whenever, $\phi > 2$, the distribution has a mode on the interior of $(0, 1)$. Whenever $\phi < 2$, the distribution is bimodal, with μ controlling the relative height of the anti-modes, and $\phi = 2$ implies a J- or L-shaped distribution, depending on whether $\mu < 0.5$ or $\mu > 0.5$, respectively. When $\mu = 0.5$ and $\phi = 2$, the distribution is uniform. Finally, the distribution reflects about 0.5 in that if $y \sim \text{Beta}(\phi\mu, \phi(1 - \mu))$, then $1 - y \sim \text{Beta}(\phi(1 - \mu), \phi\mu)$.

To form a regression model, consider two design matrices $\mathbf{X}, \mathbf{V}$ associated with the location and dispersion, respectively, so that $\mathbf{x}_i, \mathbf{v}_i$ are their ith row vectors. It is assumed that $\mathbf{X}, \mathbf{V}$ are of full rank (typically they have a column vector $\mathbf{1}$ for an intercept). Define two vectors of regression weights, $\boldsymbol{\beta}, \boldsymbol{\delta}$, which are related to μ and ϕ through two link functions as

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \mathbf{x}_i\boldsymbol{\beta}, \tag{5}$$

$$\log(\phi_i) = \mathbf{v}_i\boldsymbol{\delta}. \tag{6}$$

(More generally, $g(\cdot) \in (0, 1)$ and $h(\cdot) > 0$ may be used for location and dispersion link functions.) We use the logit/log pair in this article, which restrict $0 < \mu < 1$ and $\phi > 0$ as required. Other link functions such as the probit or complementary log–log for location or reciprocal for dispersion could be used as well. We believe the logit link is the most broadly useful for the location submodel and the log for the dispersion submodel, but this is a matter for specific consideration in light of theory and data. For instance, theory may support the use of a complementary log–log link for location, as is the case for a discrete time survival model (Fahrmeier & Tutz, 2001).

In summary, beta regression assumes

$$y|\mathbf{X}, \mathbf{V} \sim \text{Beta}(\phi(\mathbf{V})\mu(\mathbf{X}), \phi(\mathbf{V})(1 - \mu(\mathbf{X}))). \tag{7}$$

It should be noted that the *unconditional* distributions can be markedly different from the beta distribution. For instance, a beta distribution cannot be bimodal on

4

the interior of the unit interval, but the distribution of a 50% Bernoulli mixture of Beta$(5, 2)$ and a Beta$(2, 5)$ will be.

Let $\mu_i = \mu(\mathbf{x}_i)$ and $\phi_i = \phi(\mathbf{v}_i)$, where $\mathbf{x}_i, \mathbf{v}_i$ are vectors of regressors for the ith case. The likelihood function is

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\delta}; \mathbf{y}, \mathbf{X}, \mathbf{V}) = \prod_{i=1}^{n} \frac{\Gamma(\phi_i)}{\Gamma(\phi_i\mu_i)\Gamma(\phi_i(1 - \mu_i))} y_i^{\phi_i\mu_i - 1}(1 - y_i)^{\phi_i(1 - \mu_i) - 1}. \tag{8}$$

Taking the natural logarithm gives

$$\begin{aligned} \log\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\delta}; \mathbf{y}, \mathbf{X}, \mathbf{V}) = \sum_{i=1}^{n} &\log\Gamma(\phi_i) - \log\Gamma(\phi_i\mu_i) - \log\Gamma(\phi_i(1 - \mu_i)) \\ &+ (\phi_i\mu_i - 1)\log(y_i) \\ &+ (\phi_i(1 - \mu_i) - 1)\log(1 - y_i). \end{aligned} \tag{9}$$

The score function and Hessian of this likelihood can be derived using the chain rule; they are rather unwieldy expressions and thus relegated to the Appendix. The most important conclusion to draw from them is the fact that μ and ϕ are not separable, unlike, say, the mean and dispersion parameters in the Gaussian case.

2.2. Estimation in the Independence Case

The beta forms a two-parameter exponential family and, for models lacking dispersion regressors that are free of the usual trouble spots for any GLM such as improperly scaled design matrices or collinearity, is generally quite numerically tractable. The estimation theory for the independent case has been examined in great detail by Ospina et al. (2006), who note that beta regression for independent observations satisfies all usual regularity conditions for maximum likelihood estimation (MLE). Estimates of location parameters are very close to unbiased even for modest samples while estimates for the precision parameter tend to be somewhat optimistic, that is, too large, with the bias increasing as the true precision increases. They propose a correction for this bias, which we do not pursue here. In our experience, behavioral science data tend to have relatively lower values of the precision parameter and are less biased as a consequence. With a dispersion submodel, estimation is trickier and better starting values for the parameters are required. We have found the best procedure is to get starting values by fitting a reasonable location-only model before considering dispersion covariates.

MLE proceeds through numerical maximization of the log likelihood using a Newton or quasi-Newton method with standard errors for the parameters coming from the inverse of the final Hessian, that is, observed information, or using Fisher scoring and expected information. In general, Bayesian estimation seems to be a bit more forgiving than MLE in the analyses we have conducted because the priors serve to smooth the likelihood. Unsurprisingly, given that both are

likelihood based, where MLE has problems, in particular, when $\phi < 2$, MCMC usually exhibits poor mixing as well.

2.2.1. Handling Boundary Observations

One important problem with data on a bounded interval is the presence of exact boundary observations, that is, 0 or 1. These observations occur in real data but are impossible given the sample space of $(0, 1)$. For instance, if the response scale is, say, a discrete response scale on $\{0, 1, \ldots, 100\}$ that is being analyzed as a continuous score transformed into a proportion, exact values of 0 or 100 can and do occur. Even in a slider response with extremely high resolution, it is not unlikely that responses will pile up a bit at the boundary, causing a probably unimportant but notable deviation from the model. Points on the interior such as the midpoint may also occur more often than is entirely consistent with the model.

Provided that boundary observations do not represent qualitatively different responses but instead are reasonably assumed to be the result of finite precision of measurement, it is reasonable to proceed with analysis by "cheating" the observations slightly away from the boundary, via either of two methods. The first rescales all observations by a small amount:

$$y_{\text{new}} = \frac{\varepsilon}{2} + \left(\frac{1 - \varepsilon}{2}\right) y_{\text{old}}, \tag{10}$$

for a small $\varepsilon > 0$. Larger values of ε shrink the data more toward 0.5, and thus, ε should be kept as small as possible. Reasonable choices of ε will, of course, depend on the application. For instance, with a discrete response scale on $\{0, 1, \ldots, 100\}$ a reasonable choice might be $\varepsilon = 1/100$. The second method replaces 0 by $\varepsilon > 0$ and 1 by $1 - \varepsilon$, leaving the other observations unchanged. Aitchison (2003) discusses the boundary problem in detail, favoring the rescaling approach due to the fact that it preserves the mean of the data and maintains certain collapsibility conditions that are unimportant in the present context.

We advise use of sensitivity analysis to ensure that estimates and inference are not notably affected by the choice of ε. One selling point for the beta distribution as compared to, say, logit-transforming observations and then assuming normality of the transformed responses is the fact that the log likelihood of the logit-normal near the boundary is particularly sensitive while the log likelihood of the beta is less so. In particular, examining the log likelihood of the logit-normal shows that it contains exponential terms while the beta's derivatives are algebraic, implying that small perturbations in the logit-normal are amplified compared to the beta.

2.3. The Mixed Model

It is somewhat unfortunate that the usual terminology distinguishes between mixed and mixture models as, conceptually, they are essentially the same. In both

6

cases, one or more missing independent variables constant within a cluster are posited to account for dependence among observations in a cluster. Upon conditioning on these variables and observed explanatory variables, the observations are assumed independent. Beta regression in the independent case states that, conditional on the regressors $\mathbf{X}, \mathbf{V}$, the response variable is distributed beta. The mixed model extension asserts that, conditional on the regressors and the additional missing information, the response variable is distributed beta. We develop the two-level model equations here for general location and dispersion link functions $g(\cdot)$ and $h(\cdot)$, respectively, assuming balanced data for simplicity, that is, equal numbers of observations within each cluster. The general principles carry over to higher level structures and unbalanced data.

We follow the basic scheme given in McCulloch, Searle, and Neuhaus (2009, chap. 8), which we take as a basic reference. Pinheiro and Chao (2006) provide a very useful summary of the estimation theory for mixed location models. Much less is known about estimation of dispersion mixtures or location/dispersion mixtures. The discussion found in Johnson (2003), where a mixed location/dispersion model was considered for discrete ordinal variables is helpful. Proactive Monte Carlo simulation of models that are likely to be used in a given study may be useful if one is in doubt about estimation properties (Steiger, 2001). To do this, simply generate replicated data sets using a random number generator and estimate the desired model on these replications.

Let $i = 1, \ldots, I$ index subjects and $j = 1, \ldots, J(i)$ index observations within the ith subject. For simplicity of presentation here we assume that $J(i) = J$; that is, each subject has exactly the same number of responses, so there are $IJ = N$ total observations. Consider four matrices of regressors, $\mathbf{X}, \mathbf{Z}, \mathbf{V}, \mathbf{W}$. $\mathbf{X}$ and $\mathbf{V}$ are as previously. $\mathbf{Z}$ and $\mathbf{W}$ are the regressors for random effects $\mathbf{b}$ and $\mathbf{d}$, respectively. Then

$$g^{-1}(\mu_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}, \tag{11}$$

$$h^{-1}(\phi_{ij}) = \mathbf{v}_{ij}\boldsymbol{\delta} + \mathbf{w}_{ij}\mathbf{d}. \tag{12}$$

We do not assume that $\mathbf{b}$ is independent of $\mathbf{d}$, though this may often be sensible in practice. Denote this joint distribution as $q(\mathbf{b}, \mathbf{d})$. In this article, we will use either multivariate normal or multinomial mixing distributions or a combination of multivariate normal and multinomial. The likelihood function is

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{b}, \mathbf{d}; \mathbf{y}, \mathbf{X}, \mathbf{V}, \mathbf{Z}, \mathbf{W}) = \prod_{i=1}^{I} \prod_{j=1}^{J} f(y_{ij}, \mu_{ij}, \phi_{ij}). \tag{13}$$

Unfortunately, as is well known, it is not possible to estimate $\hat{\mathbf{b}}$ or $\hat{\mathbf{d}}$ directly in a statistically consistent manner. Instead, it is usual to integrate out the random effects by averaging over the random parameters (here, the integral sign should be understood as either a continuous integral or finite sum, as the case may be):

7

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\delta}; \mathbf{y}, \mathbf{X}, \mathbf{V}) = \prod_{i=1}^{I} \int_{B \times D} \prod_{j=1}^{J} f\left(\phi_{ij}\mu_{ij}, \phi_{ij}(1 - \mu_{ij})\right) q(\mathbf{b}, \mathbf{d}) d\mathbf{b} d\mathbf{d}. \qquad (14)$$

Strategies for doing this are much the same for nearly any GLMM, as shown in Pinheiro and Chao (2006). In this article we use Gauss-Hermite quadrature, leading to the marginal maximum likelihood (MML) solution, and fully Bayesian estimation, which uses MCMC through the Gibbs sampler to approximate the appropriate integrals. The GLIMMIX procedure in SAS implements the penalized quasi-likelihood approach for a location-only model (SAS Institute, 2008). This approach approximates the integrals for a multivariate normal mixing distribution by a particular Taylor series expansion. The penalized quasi-likelihood approach has yet to be thoroughly studied. It is relatively fast and can accommodate larger numbers of random effects than Gauss-Hermite quadrature. However, the properties it exhibits for categorical data seem to be present in dealing with beta-distributed variables, namely, that it is biased when the random effects are relatively large. We do not consider it further here. For the data sets we have considered, ranging between a few hundred and a few thousand observations with between one and seven random effects, MML and MCMC seem to perform as expected from other GLMM contexts. MML by quadrature works for between one and three random effects and we have obtained sensible estimates from MCMC estimation for up to seven random effects in a not unreasonable amount of time (i.e., minutes to hours).

An important special case is the finite mixture model, which asserts that, conditional on regressors $\mathbf{X}, \mathbf{V}$, the response variable is distributed as a finite weighted sum of C beta random variables:

$$y|\mathbf{X}, \mathbf{V} \sim \sum_{c=1}^{C} \gamma_c f_c(y), \qquad (15)$$

where $0 \leq \gamma_c \leq 1$, $\sum_c \gamma_c = 1$, and

$$f_c(y) = \text{Beta}(\phi_c(V) \mu_c(X), \phi_c(V)(1 - \mu_c(X))). \qquad (16)$$

The γ_c, in turn, may also have a submodel with its own regressors $\mathbf{Q}$. For $c < C$,

$$\log\left(\frac{\gamma_{ic}}{\gamma_{i_C}}\right) = q_i \eta_c. \qquad (17)$$

Both MLE and Bayesian estimation methods may be used for this model, subject to the caveats mentioned earlier. Smithson, Merkle, and Verkuilen (2009) explore mixture models in more depth and present examples of their application.

8

2.4. Remarks on Estimation

We now offer some remarks on the two estimation strategies used in this article in the context of mixed and mixture models. First, we discuss issues that are common to both methods and then remark on aspects that apply to each individually.

Estimation of a GLMM is markedly more complex than in the independent observations case. In particular, conditions guaranteeing global concavity of the log likelihood do not hold here, and it is crucial to get good starting values. Because numerical optimization algorithms are only locally convergent, having good starting values greatly speeds estimation. Similarly, for MCMC, starting too far from the region of high posterior density will slow convergence. We have found that estimating a sequence of simpler models and feeding previous parameter estimates into later estimations greatly facilitates the fitting process. For instance, a fixed effects model with no random components should provide decent starting values for the fixed effects in a mixed model and a simplified random effects only model often provides a good guess for the mixing components. Even providing the correct signs of the coefficients will help convergence. The other area where estimation can go awry for the beta is in the matter of the log-gamma function and its derivatives near 0. These functions will underflow for argument values of approximately 1×10^{-8}, which can be obtained for very small values of ϕ or whenever $\mu \approx 0$ or $\mu \approx 1$. Users programming their own routines need to be sure to properly scale their independent variables (e.g., by standardizing all of them to have mean 0 and variance 1) and bound predicted values away from such values.

MML estimation involves using a double iteration to fit the models, alternating between a quadrature step (using Gaussian quadrature) to integrate out the random effects and an optimization step to update the fixed effect parameters of the marginal likelihood. This process is computationally intensive. It is incumbent on the user to make a careful choice of the number of quadrature nodes and to have good starting values for model parameters. Adaptive Gaussian quadrature is slower (often much slower) but more accurate and requires fewer quadrature nodes. Ensuring that the solution is not sensitive to the number of quadrature nodes by manually varying that number is highly advisable. In our experience, beta regression seems to require more quadrature nodes than, say, logistic regression, which is perhaps unsurprising because the likelihood function is more complex than for the Bernoulli. The only other major area of trouble is when the maximum of the log-likelihood function is near a degeneracy. In this case, the Hessian will be poorly conditioned and optimization will usually fail (MCMC usually exhibits poor mixing in this case as well). Developing a regularization strategy for MML, for example, by adding a small number of pseudo-observations to shrink the likelihood toward a relevant null model and smooth the log likelihood, would constitute a useful research topic.

MCMC estimation generally performs well, both in successfully estimating complex models and in accuracy and interval coverage. Our conclusions regarding the strengths and weaknesses of this approach are largely in line with Browne and Draper's (2006) comparisons with likelihood and quasi-likelihood approaches. For medium to large random effects and moderate to large samples, Bayesian MCMC estimates and interval coverage seem to be reasonably well behaved. In particular, MCMC interval coverage fidelity appears superior to that provided by asymptotic Gaussian confidence intervals. MCMC estimates also are less biased than the quasi-likelihood ones (especially regarding variance components). Furthermore, in our experience MCMC often can successfully estimate complex models where likelihood-based approaches fail or are intractable, particularly when the number of random effects is larger than 3.

However, MCMC interval coverage fidelity is not well studied for small-to-medium samples in non-Gaussian random effects or hierarchical models. Browne and Draper (2006) found that undercoverage in logistic random-effects models could be corrected for by using appropriate posterior summary measures, but as they point out this is not a satisfactory solution. Likewise, MCMC may not perform as well for models in which random effects are small, but in general neither does any other known method. Finally, large data sets, models with many parameters, or models that yield full conditional distributions that are not log-concave can cause MCMC to be time-consuming, although this is likely to become less of an issue as computing power increases.

It has been widely noted that the choice of priors for variance parameters in random effects models can be problematic. We recommend avoiding the use of improper priors and evaluating the effects of choosing alternative "noninformative" priors even when they are proper, for reasons summarized by Congdon (2003, p. 21). Of course it may be reasonable to use an informative prior for substantive and/or technical reasons. See Chen, Ibrahim, Shao, and Weiss (2003) for recent investigations regarding informative priors for GLMMs. In the analyses reported here, we adopted two commonly recommended priors for variance parameters: A uniform prior for σ^2 on $(0, 1/\varepsilon)$ (e.g., Gelman & Rubin, 1992), and a Gamma$(\varepsilon, \varepsilon)$ prior for $1/\sigma^2$, where ε is a small positive number. We found very little difference between results using either of these priors in the examples presented here, but both Browne and Draper (2006) and Gelman (2006) provide useful criticisms and alternatives regarding the effects of using these priors for small sample sizes. At present the only reliable check is sensitivity analysis.

2.5. Goodness of Fit and Model Checking

A full explication of model evaluation, model checking, and model comparison for beta GLMMs is beyond the scope of this article. These topics raise some open questions in the GLMM literature generally. We briefly review the

10

available tools for evaluating and comparing beta GLMs and indicate unresolved issues. We also address some of these issues in the examples.

Global model comparison may be achieved using well-known measures such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance Information Criterion (DIC). These measures are all essentially attempts to estimate relative Kullback-Liebler divergence based on the optimized log likelihood and a penalty term representing model complexity. For MML, comparative measures of fit based on the likelihood (e.g., AIC and BIC) and the log-likelihood chi-square test for comparing nested models all are applicable. For Bayesian modelers, the posterior log likelihoods and DIC may be used. These measures are all subject to the usual caveats that apply in other contexts and we emphasize that they are best used in a comparative fashion. One particular warning that applies to users of beta regression is that all integration constants need to be included in the computation of these measures when they are used to compare across distributions (Pawitan, 2001).

Model evaluation and checking involve two related issues: How well the model "predicts" the data and whether there are unduly influential observations (e.g., outliers). The first issue may be dealt with via simulations from the posterior predictive density. Typically the match between the replicated and actual data is evaluated by comparing their cumulative relative frequencies. A similar approach can be used in MML as each point predicts a density, conditional on the regressors and parameter estimates. The main idea is generating replicated data by predicting μ_{ij} and ϕ_{ij} and then examining the resulting beta density.

Lack of a deviance or an appropriate residual is an undesirable feature of the beta GLMM. Clearly, the usual notion of a standardized residual is not adequate even for beta GLMs, because these models typically are heteroscedastic. Ferrari and Cribari-Neto (2004) proposed a deviance residual for the beta GLM, but Epsinheira et al. (2008a) found this residual to be inaccurate, particularly when the location submodel estimates are close to 0 or 1. Epsinheira et al. propose and evaluate two residuals based on the Fisher scoring algorithm. Epsinheira et al. (2008b) develop a Cook-like distance measure of the influence of deleted observations on parameter estimates. The Bayesian tools for influence diagnostics include the conditional predictive ordinate (CPO; Weiss & Cho, 1998), which is readily applicable to beta GLMMs and probably is the most well-established option for these models. This method is essentially an approximation to the jackknife and, more broadly, deletion of suspect observations (or blocks of them) provides a useful means to diagnose models. Longford (2001) proposes a similar approach in a likelihood context, using the parametric bootstrap to simulate the distribution of residuals at various levels in a multilevel model. An algorithm such as Atkinson's forward search seems promising in this context (Atkinson, 1994). The principal disadvantage of approaches such as the CPO, parametric bootstrap, or forward search is, of course, computational intensity. Ntzoufras (2009) provides CPO code in WinBUGS.

11

TABLE 1
Results of the 4 × 1 Simulation Study

Parameter	Inv. Link			Link		
	True	MML	Bayes	True	MML	Bayes
β_1	−1.39	−1.300 (0.150)	−1.288 (0.151)	0.2	0.210 (0.025)	0.217 (0.026)
β_2	0	0.010 (0.100)	0.009 (0.105)	0.5	0.500 (0.026)	0.502 (0.026)
β_3	0	−0.010 (0.110)	−0.007 (0.113)	0.5	0.500 (0.028)	0.498 (0.028)
β_4	1.39	1.370 (0.130)	1.359 (0.130)	0.8	0.800 (0.021)	0.795 (0.021)
δ_1	0	0.160 (0.160)	0.153 (0.156)	1	1.180 (0.190)	1.180 (0.188)
δ_2	0.69	0.680 (0.130)	0.669 (0.130)	2	1.980 (0.250)	1.968 (0.255)
δ_3	0.69	0.680 (0.120)	0.675 (0.117)	2	1.990 (0.240)	1.978 (0.233)
δ_4	1.1	1.100 (0.130)	1.094 (0.129)	3	3.030 (0.380)	3.010 (0.382)
σ	0.5	0.430 (0.100)	0.412 (0.111)			

Note: MML = marginal maximum likelihood.

3. A Simulation Study

To demonstrate the basic properties of mixed beta regression, we have performed parameter recovery simulation studies, one of whose results we show here. The simulated data and code are available from the authors. The study in question had 100 replications of $N = 100$ subjects and $J = 4$ items per subject. The items had means and precisions given in Table 1 and were mixed by subject with a Gaussian random intercept in the location submodel. It should be noted that these items have markedly non-Gaussian marginal densities: Given the configuration of the parameters, the first item is left skewed and exhibits mild bimodality, the middle two items are uniform, and the third item is right skewed. The resulting model equations are

$$\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_j + b_i, \tag{18}$$

$$\log(\phi_{ij}) = \delta_j, \tag{19}$$

$$b_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \tag{20}$$

The true values of the parameters are given in Table 1.

This model is quite similar to several commonly employed in the behavioral sciences, including unidimensional item response theory models or those for repeated measures experiments. We fit the model in many different ways but report the results from two runs, one using MML estimation and the other using Bayesian MCMC estimation. Given good starting values, MML convergence is

12

rapid, each replication taking under 30 seconds to run on a 2.4 GHz Intel Core 2 Duo with 2 GB of RAM running Windows XP Pro using SAS 9.1.3 NLMIXED (SAS Institute, 2006). Simulation studies with MCMC are slower, but each run took no more than 5 minutes to converge running winBUGS 1.4.3 (Spiegelhalter, Thomas, Best, & Lunn, 2004) on a similar machine.

As can be seen in the table, the parameters are nearly unbiased even for this relatively modest sample, though of course inside the link they are more biased (Pawitan, 2001). Further examination shows that all parameters both inside the link and as linear predictors have reasonably symmetric distributions; location parameters are better approximated by a Gaussian than dispersion parameters but the skew exhibited for the latter is quite mild. The only parameter that exhibits any sort of downward bias and nonnormality is σ, although its true value is still within the standard error of estimate. With more observations at Level 2, this bias is alleviated. The random effects are recovered adequately—$\text{cor}(b_i, \hat{b}_i) = 0.63$— at least as well as one would expect from only four observations per subject. As one would expect, a substantial amount of shrinkage is observed, as can be seen from $\hat{\sigma}$. This shrinkage is alleviated by a larger sample at Level 2, that is, by more items, exactly as expected. In short, mixed beta regression has statistical properties fairly similar to other generalized linear mixed models that have been extensively studied over the last 10 years.

3.1. Partitioning the Variance

One common issue in GLMMs involves partitioning the variance into that due to different levels in the model. For instance, in a two-level model with a random interecept, what proportion of the variation in the dependent variable is attributable to fixed effects, unexplained Level 1 variability and the Level 2 variability accounted for by the random intercept? This was discussed quite thoroughly by Goldstein and colleagues (Browne, Subramanian, Jones, & Goldstein, 2005; Goldstein, Browne, & Rasbash, 2002) in the context of mixed binary regression. In particular, in a linear mixed model, the question is much more straightforward to answer than in a nonlinear model such as mixed beta regression. Mixed beta regression suffers the same problems as mixed binary regression regarding the impact of nonlinearity. In particular, because the variance is a function of the mean, it is not possible to compute straightforward quantities such as the intraclass correlation frequently used in a mixed model context to measure the proportion of dependence at the cluster level.

We have adapted the methods they proposed, which we illustrate here. The approximation method, as explained in Goldstein et al. (2002), uses a first-order Taylor expansion of the link function at the mean of the distribution of the appropriate random effects. Browne et al. (2005) extend their approach to mixed models with more than two levels. We adapt their approach for the logistic link

13

by incorporating the dispersion effect parameter ϕ. We then illustrate both methods using one of the random samples from the 1×4 simulated data.

We start with an estimated location submodel

$$\mu_{ij} = \frac{\exp\left(\sum_k \beta_k x_{ijk} + b_i\right)}{1 + \exp\left(\sum_k \beta_k x_{ijk} + b_i\right)} \tag{21}$$

with $b_i \sim N(0, \sigma_u^2)$ and dispersion submodel $\phi = \exp(-\delta_0)$. The Level 1 variance is

$$\sigma_{\mu_j}^2 = \frac{\left(1 - \mu_{(j)}\right)\mu_{(j)}}{\phi + 1}, \tag{22}$$

where $\mu_{(j)}$ denotes the predicted value for a combination of x_j values.

In a two-level model, following Goldstein et al.'s (2002) first-order Taylor expansion of the location submodel, our adaptation of their formulation yields the following approximate estimate of the Level 2 variance,

$$\sigma_{u_{(j)}}^2 = \exp\left(-2\sum_{j=1}^{J} x_j \beta_j\right)\mu_{(j)}^2 \sigma_u^2 \tag{23}$$

for any given combination of x_j values. Finally, the approximate variance partition coefficient (VPC, the proportion of variation accounted for at Level 2) is

$$\tau_j \approx \frac{\sigma_{b_{(j)}}^2}{\sigma_{b_{(j)}}^2 + \sigma_{\mu_j}^2}. \tag{24}$$

The simulation approach includes the following steps:

1. We start with an estimated location submodel.
2. Generate a large number (say, $M = 5{,}000$) of random Level 2 residual values $b_m \sim N(0, \sigma_b^2)$, for m from 1 to M.
3. Select one or more appropriate combinations of values for the covariates x_j. For each such combination, use sample estimates of the relevant parameters to compute M predicted values

$$\mu_{m(j)} = \frac{\exp\left(\sum_k \beta_k x_{jk} + b_m\right)}{1 + \exp\left(\sum_k \beta_k x_{jk} + b_m\right)}. \tag{25}$$

In our 1×4 example, we do this for each of the four combinations of 0–1 values for the dummy x_{jk} variables.

14

TABLE 2
Approximation and Simulation Method VPC Results

	$\sigma^2_{\mu_1}$	$\sigma^2_{\mu_2}$	$\sigma^2_{\mu_3}$	$\sigma^2_{\mu_4}$
	0.0607	0.0827	0.0902	0.0427
	$\sigma^2_{\mu_{(1)}}$	$\sigma^2_{\mu_{(2)}}$	$\sigma^2_{\mu_{(3)}}$	$\sigma^2_{\mu_{(4)}}$
Approximation	0.0067	0.0191	0.0192	0.0080
Simulation	0.0072	0.0165	0.0165	0.0081
	τ_1	τ_2	τ_3	τ_4
Approximation	0.0997	0.1879	0.1751	0.1580
Simulation	0.1060	0.1666	0.1550	0.1604

4. Compute the variance $\sigma^2_{b(j)}$ of the $\mu_{m(j)}$.
5. Compute $\sigma^2_{u_j}$.
6. The approximate VPC is τ_j.

The approximation method is illustrated here; a similar demonstration of the simulation method is available from the authors. In our sample of 100 observations from our repeated-measures simulation, the sample means are $\mu_{(1)} = .1807$, $\mu_{(2)} = .5136$, $\mu_{(3)} = .5106$, and $\mu_{(4)} = .7974$. The sample $\phi_{(j)}$ are $\phi_{(1)} = 1.438$, $\phi_{(2)} = 2.019$, $\phi_{(3)} = 1.770$, and $\phi_{(4)} = 2.787$. Finally, the sample $\sigma^2_u = 0.3067$. The Level 1 variances are shown in the upper row of Table 2. For example, $\sigma^2_{\mu_1} = \mu_{(1)}(1 - \mu_{(1)})/(\phi_{(1)} + 1) = 0.0607$. The Level 2 variances for the approximation method are listed in the row beneath the Level 1 variances. For instance,

$$\sigma^2_{u_{(1)}} = \frac{\mu^2_{(1)}\sigma^2_u}{\left(\left(1 - \mu_{(1)}\right)\mu_{(1)} + 1\right)^2} = .0067.$$

The VPC results are shown in the lower part of Table 2, for both the approximation and the simulation (from a typical simulation run) methods. They are reasonably close, with the simulation method returning a somewhat lower VPC for the second and third categories than the approximation method does. In either case, the VPC magnitudes reflect moderate clustering due to Level 2 effects, as can be seen from the τ_j magnitudes which range from about .10 to .19.

4. Applications

We provide three real examples of mixed or mixture beta regressions using data from several cognitive experiments. The first example shows a mixture model applied to judged probabilities elicited in an experiment by Gurr

15

(2009). It demonstrates the performance of a finite mixture of betas in a relatively simple setting. The second example is a more complicated reanalysis of judged probability data from Fox and Rottensreich (2006). This example illustrates the fact that modeling the dispersion structure often leads to different conclusions than modeling location alone, more in line with theoretical expectations. The third example considers numerical confidence ratings taken from Roy and Liersch (2009). This model includes both location and dispersion submodel random effects.

4.1. Partition Effects on Probability Judgments by a Finite Mixture Model

On grounds of insufficient reason, a probability of $1/K$ is assigned to K mutually exclusive possible events when nothing is known about the likelihood of those events. Fox and Rottenstreich (2003) found evidence that subjective probability judgments are typically biased toward this distribution. Smithson and Segale (2009) extended the study of partition effects to judgments of lower and upper probabilities and introduced a beta regression mixture modeling approach to analyzing this influence.

Gurr (2009) investigated partition priming effects by setting up an ambiguous sample space and introducing an experimental manipulation to influence perceptions of this space. One hundred fifty-five undergraduate students at The Australian National University (108 females, 43 males, and 4 unspecified) were recruited for his study. Their ages ranged from 17 to 43 years ($M = 21.41$, $SD = 4.46$). The participants completed two probability judgment tasks. The first task was a "no-information" condition and participants were told that "In this block you will be asked to make decisions with very little information. A man, Geoff, works as an engineer in a big hydraulic power production firm. Geoff has three acquaintances, Pat, James and Chris." They were then asked to assign probabilities to the three acquaintances being hydraulic power engineers in Geoffs firm. The second "information" condition repeated the task but provided information about each acquaintances engineering qualifications.

The resultant sample space is ambiguous because the number of positions available in Geoff's firm and the number of potential applicants to those positions are not specified. The hypothesis associated with this part of the study was that participants would be more likely to return "ignorance" priors (i.e., assigning equal probabilities to all three targets) under the no-information condition. In particular, it was expected that the no-information condition would elicit more ignorance priors comprising probabilities of 1/2 for each target, by virtue of participants thinking in terms of a two-alternative sample space for each target.

Participants also were randomly assigned to one of two priming conditions, one asking them which applicant was most likely to get a position in Geoff's firm (the comparison prime) before they were asked for their probability assignments and the other not asking this question. The major hypothesis for this aspect of the

16

experiment was that the comparison prime would make participants more likely to make their probabilities additive (i.e., summing to 1 across the three targets) because it would cue them to think that there might be only one position available. The provision of three targets enables a distinction between judges expressing ignorance by assigning probabilities of 1/2 to all three targets and who express ignorance by assigning 1/3 to all targets, whereas using only two targets would conflate additivity with ignorance prior assignments.

Initially, probability assignments in the two tasks were analysed separately using mixture beta-regression models via MLE methods, using SAS 9.2 and IBM SPSS Statistics 18 (IBM SPSS results are reported here). No-effects models were determined by comparing model fit for a single-distribution, two-component, and three-component mixture models. All two-component mixture models fixed one component distribution's mean either at 1/2 or 1/3 while freeing the other mean to be estimated, and the three-component model fixed two component means at 1/2 and 1/3 while freeing the third component mean to be estimated. The fixed-means component distributions were modeled by $f_c(y) =$ Uniform $(\mu_c - \varepsilon, \mu_c + \varepsilon)$, with ε given values of .001, .01, and .05. We report findings for $\varepsilon = .01$, as the other values yielded similar results.

Beginning with the no-information condition task, both two-component mixture models (one with a component mean fixed at 1/2 and the other fixed at 1/3) outperform the single-distribution model, difference in log-likelihood chi-squares: $\chi^2(1) = 108.08$, $p < .0005$, and $\chi^2(1) = 50.48$, $p < .0005$, respectively, and a three-component model (with two-component means fixed at 1/2 and 1/3) outperforms the best of the two-component models, $\chi^2(1) = 135.66$, $p < .0005$. Turning to the information condition, both two-component mixture models (one with a component mean fixed at 1/2 and the other fixed at 1/3) outperform the single-distribution model, difference in log-likelihood chi-squares: $\chi^2(1) = 29.75$, $p < .0005$, and $\chi^2(1) = 102.09$, $p < .0005$, respectively and a three-component model (with two component means fixed at 1/2 and 1/3) outperforms the best of the two-component models, $\chi^2(1) = 26.62$, $p < .0005$. The three-component model therefore is adopted for both the no-information and information tasks. The distributions for both tasks under the two experimental conditions are shown in Figure 1.

A mixed three-component mixture model was estimated in WinBUGs 1.4.3 to assess the joint effects of the prime and the no-information versus information conditions. This model is limited to a random intercept in the location submodel, due to there being only two data points per participant. The location submodel is

$$\mu_{1j} = 1/2,$$

$$\mu_{2j} = 1/3,$$

$$g(\mu_{3ji}) = \beta_{0j} + b_i, \qquad (26)$$
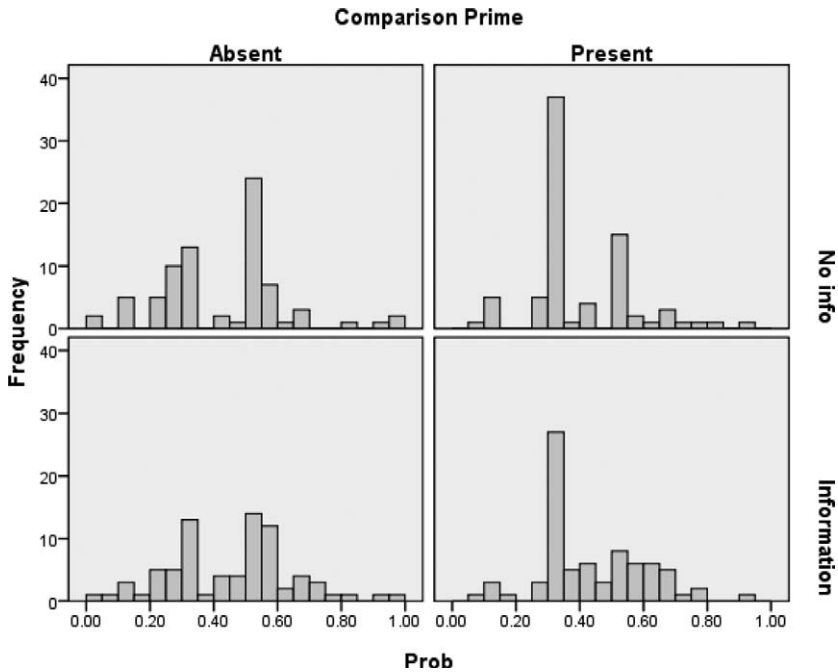
17

Comparison Prime

FIGURE 1. *Component distributions for prime by information conditions.*

where $b_i \sim N(0, \sigma_b^2)$. The dispersion submodel is

$$h(\phi_{cj}) = \delta_{cj}, \tag{27}$$

$$\delta_{cj} \sim N\left(\mu_{\delta cj}, \sigma_{\delta cj}^2\right).$$

For $j = 1$ and 2, we used informative priors to simulate compressed component distributions centered at $1/2$ and $1/3$, whereas for $j = 3$, we used a noninformative prior. Finally, the composition submodel is

$$\log(\gamma_{1ji}/\gamma_{3ji}) = \eta_{10j} + \eta_{11j}q_i$$

$$\log(\gamma_{2ji}/\gamma_{3ji}) = \eta_{20j} + \eta_{21j}q_i \tag{28}$$

$$\gamma_{3ji} = 1 - \gamma_{1ji} - \gamma_{2ji},$$

where $q_i = 0$ for the no-prime condition and 1 for the prime.

We used a two-chain model with a 10,000 iteration burn-in and estimates based on the subsequent 10,000 iterations. The parameter estimates, standard errors, and 95% credible intervals are shown in Table 3. The priming effect on

18

TABLE 3
Beta-Regression Mixture Models for Gurr (2009) Study

Parameter	M	SD	Credible Lower	Interval Upper
Location				
β_{01}	−0.330	0.129	−0.570	−0.065
β_{02}	−0.109	0.107	−0.307	0.100
σ_b	1.002	0.087	0.840	1.180
Dispersion				
δ_{11}	8.737	0.352	8.037	9.402
δ_{12}	6.980	0.522	6.120	8.180
δ_{21}	16.800	0.318	16.170	17.410
δ_{22}	15.900	0.316	15.270	16.510
δ_{31}	2.550	0.303	1.972	3.165
δ_{32}	3.092	0.280	2.568	3.643
Compos.				
η_{101}	−1.409	0.381	−2.233	−0.731
η_{102}	−1.782	0.488	−2.864	−0.973
η_{201}	−0.640	0.261	−1.169	−0.143
η_{202}	−1.567	0.327	−2.238	−0.968
η_{111}	1.645	0.455	0.803	2.579
η_{112}	1.329	0.571	0.304	2.557
η_{211}	−0.176	0.425	−1.005	0.647
η_{212}	−0.505	0.565	−1.674	0.588
$\eta_{102} - \eta_{101}$	−0.373	0.615	−1.642	0.788
$\eta_{202} - \eta_{201}$	−0.927	0.417	−1.756	−0.132
$\eta_{112} - \eta_{111}$	−0.315	0.731	−1.702	1.169
$\eta_{212} - \eta_{211}$	−0.328	0.710	−1.749	1.072

relative composition is similar in both conditions, as can be seen from the credible intervals for $\eta_{112} - \eta_{111}$ and $\eta_{212} - \eta_{211}$. The large positive η_{111} and η_{112} values suggest that the presence of the comparison prime increases the proportion of additive responses (i.e., those summing to 1). However, the small η_{211} and η_{212} values indicate that it has no discernible impact on the proportion of responses centered on 1/2. On the other hand, the large negative $\eta_{202} - \eta_{201}$ difference suggests that the no-information condition increases the proportion of responses summing to 1/2, whereas the weak $\eta_{102} - \eta_{101}$ difference shows that it has little or no effect on the proportion of additive responses. Thus, both of the aforementioned hypotheses received some support.

The estimated and observed proportions are shown in Table 4. There is a slight tendency to underestimate the proportions of additive responses in the no-prime

19

TABLE 4
Recovery of Mixture Composition Structure

Condition	Component Mean	Estimated No Prime	Observed No Prime	Estimated Prime	Observed Prime
No-info	1/3	0.138	0.156	0.468	0.470
	1/2	0.298	0.299	0.163	0.167
	Free	0.564	0.545	0.369	0.359
Information	1/3	0.122	0.143	0.361	0.346
	1/2	0.152	0.156	0.072	0.077
	Free	0.726	0.701	0.568	0.577

condition and to overestimate them in the prime condition for the information-condition task, but otherwise the composition structure is recovered well. Of the 310 mean component classification scores produced by the MCMC run, 288 (92.9%) of these were within 0.1 of their correct classification, so the model correctly assigned most cases to a component distribution.

4.2. Dispersion Modeling to Address Response Style in Judged Probabilities

See, Fox, and Rottenstreich (2006) presented a series of studies of the influence that the number of possibilities (the state-space partition) has on judged probabilities. In their Study 1, participants observed a set of randomly ordered restaurant receipts for various meals on various days by a person named "Joe." The receipts contained two pieces of information: day of the week (Sunday, Monday, ..., Saturday) and meal category (breakfast, lunch, and dinner). Participants were told they would be asked to judge the likelihood of the events they had observed. See et al.'s primary hypothesis was that subjective probability estimates would be biased toward the ignorance prior suggested by the target attribute's partition: meal of the day (three alternatives, therefore an ignorance prior of 1/3) or day of the week (1/7). Following the learning phase, 267 participants were asked to estimate the likelihoods of all 10 attributes (meal categories and days of the week). One peculiarity of their data is the substantial number of exact boundary observations, particularly concentrated in certain participants. These values needed to be adjusted slightly. In our analysis here, we replaced exact 0 values (12 in all) with 0.001 and exact 1 (5 in all) with 0.999. The exact 1 responses are particularly important because these are highly unlikely responses given the stochastic specification of the model.

To assess the relative influence of the partition and the information provided during the learning phase, See et al. conducted separate regression analyses for each participant. Their model was derived from Fox and Rottenstreich's (2003) ignorance prior model, which can be written as follows:

20

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{0i} + \beta_{1i}\log\left(\frac{n_{fj}}{n_{aj}}\right) + \beta_{2i}\log\left(\frac{F_{fj}}{F_{aj}}\right), \tag{29}$$

where p_{ij} is the probability of the jth target judged by the ith subject, n_{fj} and n_{aj} are the number of distinct states that the jth focal target and its alternatives, respectively, can take on (i.e., 1 and 2 for meals and 1 and 6 for days of the week), and F_{fj} and F_{aj} are the frequencies with which the jth focal target and its alternatives appeared during the learning phase. They reported the median intercept (-0.13), regression coefficients ($\hat{\beta}_1 = .36$ and $\hat{\beta}_2 = .46$), and multiple R^2 ($.61$) for these regressions.

One obvious weakness with this analysis is the fact that each participant only provides 10 observations, which means each individual-level regression only has six degrees of freedom. We reanalyzed their data using a beta GLMM approach. All models shown here were fit using WinBUGs 1.4.3. We ran two-chain models with a burn-in of 5,000 iterations and an additional 10,000 iterations for estimation purposes. (MML was also attempted. The final Hessian was singular but coefficient estimates were very close to those from MCMC.) In these models, i indexes subjects, j indexes stimuli, and k indexes dependent variables including the intercept. In line with convention, when $k = 0$, $x_{ij0} = 1$ and β_0 = intercept. We began with a model incorporating a random intercept in the location submodel (Model 0):

$$p_{ij} \sim \text{Beta}(\mu_{ij}\phi_{ij}, (1 - \mu_{ij})\phi_{ij}), \tag{30}$$

$$\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \sum_{k} \beta_k x_{ijk} + b_i, \tag{31}$$

$$\log(\phi_{ij}) = \delta_0, \tag{32}$$

where $\beta_k \sim N(\mu_{\beta_k}, \sigma_{\beta_k}^2)$, $b_i \sim N(0, \sigma^2)$, and $\delta_0 \sim N(\delta, \sigma_\delta^2)$. We then estimated a model with random b_k and independent random-effects terms (Model 1), so that the right-hand side of Equation 33 is replaced with $\sum_{k}(\beta_k + b_{ik})x_{ijk}$, where $b_{ik} \sim N(0, \sigma_k^2)$, and a model with correlated random-effects (Model 2), in which $b_{ik} \sim N(0, \sigma_k^2)$ is replaced with $\mathbf{b_i} \sim N(\mathbf{0}, \mathbf{\Sigma})$.

Table 4 shows the posterior mean $-2\log\mathcal{L}$ figures for the three models, indicating clear gains in model fit for each of them. That said, the location and dispersion fixed-effects submodels were nearly unchanged in these three models. The Model 2 estimates and 95% credibility intervals are $\hat{\beta}_0 = -0.193$, $[-0.259, -0.130]$; $\hat{\beta}_1 = 0.233, [0.161, 0.303]$; $\hat{\beta}_2 = 0.502$, $[0.449, 0.555]$; and $\hat{\delta}_0 = -2.886, [-2.944, -2.827]$. These results show a somewhat different picture from See et al.'s median regression coefficients, suggesting a greater influence from the learning component and a lesser one from the partition.

Mixed and Mixture Regression Models

Other models are possible. One obvious alternative is a model that incorporates a random intercept in the dispersion submodel instead of one in the location submodel. Including random intercepts in both submodels resulted in models that failed to converge in WinBUGS. However, a model with a random intercept in the dispersion submodel alone converged well. Moreover, it considerably outperformed all of the models presented thus far (posterior mean $-2\log \mathcal{L} = -5783.86$). More importantly, this model yields a somewhat different location submodel from the other approach. The model estimates and 95% credibility intervals are $\hat{\beta}_0 = -0.149, [-0.198, -0.101]$; $\hat{\beta}_1 = 0.395, [0.339, 0.444]$; $\hat{\beta}_2 = 0.403, [0.366, 0.438]$; and $\hat{\delta}_0 = -3.012, [-3.130, -2.896]$. These results suggest that the partition effect is nearly equal to the effect of the information provided in the learning phase, more in line with See et al.'s findings.

What, then, does the dispersion random effect mean here? If one examines the raw data as well as $\hat{d}$, it appears that it essentially addresses systematic differences in response style away from the primed anchors. A location random effect here would mean that participants systematically anchored off the primes. There seems to be no evidence for this in these data, which is unsurprising given the theory.

See et al. (2006) also collected confidence judgements and knowledge judgments to measure how knowledgable their subjects were. The motivation behind this was that more knowledgeable subjects might be less susceptible to priming effects. A subject-level knowledgeability covariate could be incorporated into a more elaborate model, with a cross-level knowledgeability-by-partition interaction term to ascertain whether greater knowledgeability lessens the partition effect.

As a comparison and more in line with the original analysis, we also attempted to fit logit-normal mixed models by transforming the dependent variable and assuming it is normal. These models exhibited substantial sensitivity to the choice of ε, the shrinkage parameter in Equation 11. In particular, the variance of the random effect was pulled substantially by choice of ε. This is, no doubt, because the usage of boundary values varies drastically within participants.

4.3. Location and Scale Mixing: Numerical Confidence Ratings

The data used for this example come from Roy and Liersch (2009), who considered the "better than average" effect, also known as the Lake Wobegon Effect, the well-known finding that more than 50% of subjects are willing to rate themselves as being better than average, impossible in a symmetric distribution (e.g., Alicke & Govorun, 2005). The participants are 35 female undergraduates at a large Midwestern university (an additional 20 male participants were dropped from the analysis here to avoid dealing with a substantial gender by skill interaction). They provided numerical confidence ratings (on a scale of 0–100) of their ability to perform ten skills: riding a bike (bke), dancing (dnc), driving a car
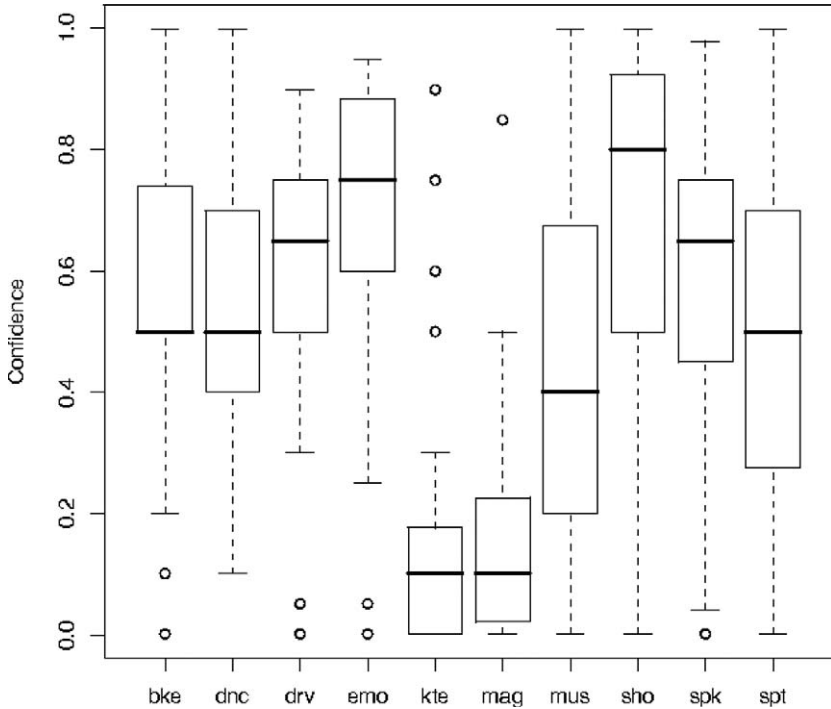
22

FIGURE 2. *Roy and Liersch (2009) confidence ratings for 35 female participants.*

(drv), sensing emotion (emo), karate (kte), performing magic tricks (mag), musical performance (mus), speaking in public (spk), playing "ball" sports (spt), and tying a shoe (sho).

These skills are expected to have substantially varying difficulties in the study population—for instance, karate and magic are both expected to be difficult while tying shoes is expected to be easy. The experimenters hoped to see if participants would rate themselves as being worse than average on difficult skills and better on easy skills, which would imply substantially skewed responses, both left and right skewed, as difficulty varied. Of course, it would not be surprising for there to be substantial differences in response style in these data, in location, dispersion, or both, much like what was observed for the judged probabilities in the previous example. A location difference indicates a systematic anchor point of "generalized perceived self-competency" for a given participant while a dispersion difference indicates how spread a given participant's responses are around that. Box plots of the ratings are shown in Figure 2.

We rescaled the response to lie inside the unit interval by letting

$$y = .0025 + .9975 * (y_{\text{raw}}/100).$$

23

As can be seen, the distributions are substantially skewed (particularly for the harder skills) and are overdispersed relative to the beta conditional on skill alone. There are some notable outliers, particularly for the difficult skills such as karate and magic, about which more below. The full specification will be given; simpler versions are devised by restricting one or more parameters to 0. Let $i = 1, \ldots, 55$ index subjects and $j = 1, \ldots, 10$ index skills. We assume that observed response y_{ij} is beta-distributed conditional on skill and random effects; that is,

$$y_{ij}|\text{skill}_j, b_i, d_i \sim \text{Beta}(\phi_{ij}\mu_{ij}, \phi_{ij}(1 - \mu_{ij})), \tag{33}$$

where

$$g(\mu_{ij}) = \beta_j * \text{skill}_j + b_i, \tag{34}$$

$$h(\phi_{ij}) = \delta_j * \text{skill}_j + d_i. \tag{35}$$

The random effects are multivariate normal; that is,

$$\begin{pmatrix} b \\ d \end{pmatrix} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}). \tag{36}$$

For the unrestricted model, $\mathbf{\Sigma}$ is unstructured. Imposing $\text{cov}(b, d) = 0$ would force the random effects to be independent and of course setting a variance term to 0 in effect drops the term from the model. We use the logit-exponential links and in this case no column of 1s is used for the design matrices. In sum, the models fit here posit separate beta distributions for each skill, possibly mixed in both location and dispersion.

We consider four different models for these data. The first model fits only fixed location and dispersion effects for skills with no random effects. The second model considers a location random effect only. The third fits a dispersion random effect only. The fourth considers random effects for both location and dispersion, allowing these to be correlated. The parameter estimates shown here are from the MML estimation but MCMC estimates are very similar. As there are many model parameters, we consider only the random effects and fit statistics for the moment. Examination of the reproduced means for Model 4 shows that it almost exactly recovers the raw data means for each skill (as indeed do the others). However, as can be seen from the BIC statistics, Models 3 and 4 are quite soundly preferred to the other two, again suggesting substantial differences among subjects in scale usage. However, Model 4 has some modest support over Model 3 (and indeed would have more if ρ_{bd} were constrained to 0), which suggests some location anchoring on the part of subjects. There are 20 fixed effects parameters, which we do not show to conserve space. Model 4 has generally more extreme parameter estimates, as is common in random effects models.

24

TABLE 5
Posterior Mean −2log L for Three Models

	−2log $\mathcal{L}$	Diff.
Model 2	−5193.08	
Model 3	−5337.02	143.94
Model 4	−5474.13	137.11

TABLE 6
Random Effects Parameter Estimates and Fit Statistics for the Confidence Rating Data

Parameter	Model 1	Model 2	Model 3	Model 4
$\hat{\sigma}_b$	—	0.37	—	0.34
$\hat{\sigma}_b$	—	—	0.68	0.70
$\hat{\rho}_{bd}$	—	—	—	0.16
−2log $\mathcal{L}$	−193.6	−205.2	−256.0	−266.2
df	20	21	21	23
BIC	−122.5	−130.5	−181.3	−184.5

Note: BIC = Bayesian Information Criterion.

All is not well with Model 4, however. To evaluate it in more detail, we computed the Pearson correlation between $\hat{\mu}_{ij}$ and y_{ij} by skill. These correlations range between 0.4 and 0.7 for most skills. However, two skills, karate and music, have correlations that are effectively 0. As is evident in the boxplot, the confidence ratings for karate are extraordinarily dispersed. Many participants rate their confidence at 0 and the distribution is L-shaped. However, a few participants rate their karate ability as being very high (perhaps accurately?). Music performance is not as dispersed but behaves similarly. We speculate that this effect might be due to the fact that both skills are ones in which some participants have much more objective information about their ability compared to others, such as dancing or driving. Switching to a different random effects distribution might accommodate this overdispersion better (Table 6).

To assess the effect of these outliers on the random effects estimates, we refit Model 4 omitting karate and music. In this case, the fit statistics are no longer comparable, but parameter estimates should be similar. The fixed effects are very similar but given the specification of the model this is to be expected. Examining $\hat{\sigma}_b = 0.45$, $\hat{\sigma}_d = 0.81$, and $\hat{\rho}_{bd} = 0.22$. This is comforting because the random effects parameters would be more likely to be distorted by the outliers in other parts of the model, but they are not. By contrast, the logit-normal model parallel of Model 4 exhibits very poor convergence, even when started from the same point as the beta model.

5. Discussion and Conclusion

The principal focus of this article has been a GLMM framework for dependent random variables with conditional beta distributions. In particular, we have shown that for subjective rating data, a random effects dispersion model is highly useful in modeling individual heterogeneity in response style. Response style is a pervasive problem with self-report data such as numerical confidence ratings, judged probabilities, sliders, and so on, all of which are used quite commonly by psychologists and other behavioral scientists. While in many cases it is simply a nuisance behavior, in other cases, response style heterogeneity might be an important study outcome in its own right, for instance, when group differences are expected as in attitude polarization (Dolnicar & Grun, 2007). We have seen in two examples how partition-priming effects on probability judgments are more accurately described by response-style heterogeneity than by a shift in the mean response.

This framework may be extended and developed in several ways. We briefly discuss the following starting points for such developments: Alternative estimation techniques, model diagnostics, special mixture models, alternative link functions, and alternative distributions.

Beginning with estimation methods, at least three major approaches remain to be explored. This article dealt with MML and Bayesian MCMC techniques but omitted restricted maximum likelihood (REML), penalized quasi-likelihood (PQL), and generalized least squares (GLS) approaches. These techniques were neglected primarily for lack of software tools with which to investigate them. As mentioned earlier, there is a limited implementation of PQL estimation in the SAS GLIMMIX procedure. Investigations of REML and GLS await software developments.

We now turn to extensions of the GLMM itself. Some applications require ''x_0-inflated'' models, where x_0 is a specific value on the unit interval. The most well-known situation is where $x_0 = 0$ (see Kieschnick & McCullogh, 2003), and a natural model here is a mixture of a beta random variable and one whose PDF concentrates its entire mass at 0. However, situations can arise such as the judged probability example (Smithson & Segale, 2009) in which values near x_0 should be included in the ''inflation'' component distribution. This component distribution therefore can no longer concentrate all of its mass at x_0, although it will have a small variance. Mixture models with a component distribution of this kind are notoriously difficult to estimate, and alternative estimation methods have yet to be fully investigated. One promising avenue is a Bayesian approach utilizing an informative prior for the inflation component distribution. Smithson et al. (2009) and Merkle, Verkuilen, and Smithson (2009) investigate and demonstrate the utility of these models in analyzing such phenomena as scale response style, ''nice number'' bias, and nonadditivity in probability judgments.

26

More general extensions of the GLMM may be obtained by considering alternative link functions and distributions. As indicated earlier, the logit is not the only twice-differentiable link function that transforms the unit interval to the real line. Little is known about how these alternative link functions perform in comparison with the logit.

Likewise, several alternatives to the beta distribution have been proposed. Kumaraswamy (1980) provided a two-parameter (α, β) distribution closely related to the beta. A Kumaraswamy-distributed random variable is the αth root of a Beta$(1, \beta)$ random variable. This distribution's main advantage over the beta distribution is the closed form of its density and cumulative density functions, but the expressions for its moments do not have closed forms whereas those for the beta distribution do.

A more popular alternative is Johnson's S_B distribution, which begins with a linear scaling to the (0,1) interval, followed by a logit transformation and finally a linear scaling to the standard normal distribution. The logit-logistic (LL) distribution (Tadikamalla & Johnson, 1982) replaces the normal with the standard logistic distribution. The chief advantage of the LL over the S_B distribution is that its CDF has a simple closed invertible form. The shape domain of the LL distribution includes that of the beta distribution. However, its moments do not have closed-form expressions. A third alternative is the simplex distribution (Barndorff-Nielsen & Jorgensen, 1991), which is constructed from the inverse Gaussian distribution in a manner similar to the way the beta is contsructed from the gamma distribution. Like the beta parameterization used in GLMs, the simplex distribution is parameterized in terms of a mean and standard deviation. Unlike the beta model, the simplex distribution is naturally a deviance-based model. Kieschnick and McCullogh (2003) compare it with their beta regression model on two data sets. Qiu, Song, and Tan (2007) develop a simplex distribution mixed model.

Fourth, the logit-normal and logit-logistic distributions are members of the otherwise unexplored symmetric family that utilizes an invertible transformation $H : [-\infty, \infty] \to [0, 1]$ such that $x = H(z)$, and a standardized PDF, $g(z)$, whose support is the real line. The resultant PDF of X is

$$f(x) = g(H^{-1}(x)) \frac{\partial H^{-1}(x)}{\partial x}.$$

The H function can be any CDF whose quantile function has a closed form expression. For some of these distributions, as with the logit-logistic, the cumulative density and quantile functions can be explicitly written, and they have natural location and dispersion submodels. However, very little is known about this subfamily.

Finally, Cox (1996) and Papke and Wooldridge (1996) apply a quasi-likelihood approach to modeling a variable on the unit interval. The quasi-likelihood approach specifies the first and second moments of the conditional

27

distribution as functions of the mean but does not specify the full distribution; in this sense it is a second-order analog of maximum likelihood. This approach is useful when the relationship between the mean and variance is not captured by a standard distribution. Kieschnick and McCullogh (2003) compare a special case of Papke and Wooldridge's model with their beta regression model on two data sets and find that both work reasonably well and give similar results. A strategy using a semiparametric model based on the Dirichlet process might also be a useful strategy to consider (McAuliffe, Blei, & Jordan, 2006).

Summing up, this article has extended beta regression to deal with dependent observations in a general way, along the lines of the multilevel modeling literature. The resulting GLMM has been shown to be effective in a simulation study and three real-world applications and, as indicated earlier, it has applicability in a wide range of disciplines where doubly bounded constructs are commonplace. We also have demonstrated the practicality of MLL and MCMC estimation of this GLMM under a variety of conditions. Code for the examples in this article and additional resources are available at http://dl.dropbox.com/u/1857674/betareg/betareg.html. As should be apparent from this section, there is considerable potential for extensions and developments in modeling doubly bounded dependent variables, but the state of the art there is rapidly approaching a richness and sophistication comparable to the models available for other kinds of variables.

Appendix
Score and Information Matrix

To simplify the presentation of the score and Hessian, we only consider the log-density of one observation. These terms are all sums and can be pieced together afterward as needed. Let

$$\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x), \tag{A1}$$

that is, the digamma function. The mean-precision parameterized beta density is

$$f(y, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu)\Gamma(\phi(1-\mu))} y^{\phi\mu-1} (1-y)^{\phi(1-\mu)-1}. \tag{A2}$$

Taking the natural logarithm, expanding fully and rearranging the terms slightly gives

$$\begin{aligned} \log f(y, \mu, \phi) = {}& \log \Gamma(\phi) - \log \Gamma(\phi\mu) - \log \Gamma(\phi(1-\mu)) \\ & + \phi\mu \log y + \phi(1-\mu)\log y - \log y - \log(1-y). \end{aligned} \tag{A3}$$

Components of the score function are given by the vector of first partials and the Hessian by the matrix of second partials. These are

28

$$\frac{\partial \log f}{\partial \mu} = -\phi(2 \, \text{arctanh}(1 - 2y) + \psi(\phi\mu) + \psi(\phi(1 - \mu)));$$

$$\frac{\partial \log f}{\partial \phi} = -2\mu \, \text{arctanh}(1 - 2y) + \psi(\phi) - \mu\psi(\phi\mu) - (1 - \mu)\psi(\phi(1 - \mu)),$$
(A4)

and

$$\frac{\partial^2 \log f}{\partial \mu^2} = -\phi^2 \left(-\psi'(\phi\mu) + \psi'(\phi(1 - \mu)) \right);$$
(A5)

$$\frac{\partial^2 \log f}{\partial \phi^2} = \psi'(\phi) - \mu^2 \psi'(\phi\mu) - (1 - \mu)^2 \psi'(\phi(1 - \mu));$$

$$\frac{\partial^2 \log f}{\partial \mu \partial \phi} = -2 \, \text{arctanh}(1 - 2y) - \psi(\phi\mu) - \psi(\phi(1 - \mu))$$
$$- \phi\mu\psi'(\phi\mu) + \phi(\mu - 1)\psi'(\phi(1 - \mu)),$$

respectively. Depending on the algorithm and software used, these quantities may or may not be necessary. Modern software such as SAS uses symbolic differentiation to compute the needed derivatives and only require the likelihood. Other algorithms use numerical approximations to the derivatives. However, they are of theoretical importance. In particular, the fact that μ and ϕ appear in the cross-partial of the Hessian shows that the location and scale parameters are not separable, unlike in models such as the Gaussian.

To consider a regression model for link pair $\mu(\mathbf{X}) = g(\mathbf{X}\beta)$ and $\phi(\mathbf{W}) = h(\mathbf{W}\delta)$, it is necessary to use the chain rule, considering the various partial derivatives. See Ospina et al. (2006).

Note

1. The *four-parameter beta distribution* allows for the boundaries to be different than (0, 1) by including two additional parameters. Assuming these boundaries are known a priori the beta distribution on interval (l, *u*) is a linear transformation of one on (0, 1) and thus, without loss of generality, the boundaries can be ignored. If the boundaries are unknown, estimation becomes highly irregular. We believe it is rarely the case that a doubly bounded distribution such as the beta would be a plausible model when reasonable bounds are unknown a priori.

References

Aitchison, J. (2003). *The statistical analysis of compositional data.* London, England: The Blackburn Press.

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment.* New York, NY: Psychology Press, 85–108.

Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association*, *89*, 1329–1339.

Barndorff-Nielsen, O. E., & Jorgensen, B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis*, *39*, 106–116.

Brehm, J., & Gates, S. (1993). Donut shops and speed traps: Evaluating models of supervision on police behavior. *American Journal of Political Science*, *37*, 555–581.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*, 473–514.

Browne, W. J., Subramanian, S. V., Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit over-dispersion. *Journal of the Royal Statistical Society, Series A*, *168*, 599–613.

Buckley, J. (2002). Estimation of models with beta-distributed dependent variables: A replication and extension of Paolino (2001). *Political Analysis*, *11*, 1–12.

Carroll, R. J. (2003). Variances are not always nuisance parameters. *Biometrics*, *59*, 211–220.

Chen, M.-H., Ibrahim, J. G., Shao, Q.-M., & Weiss, R. E. (2003). Prior elicitation for model selection and estimation in generalized linear mixed models. *Journal of Statistical Planning and Inference*, *111*, 57–76.

Congdon, P. (2003). *Applied Bayesian modelling.* Chichester, England: Wiley.

Cox, C. (1996). Nonlinear quasi-likelihood models: Applications to continuous proportions. *Computational Statistics and Data Analysis*, *21*, 449–461.

Crowder, M. J. (1978). Beta-binomial ANOVA for proportions. *Applied Statistics*, *27*, 34–37.

Dolnicar, S., & Grun, B. (2007). Cross-cultural differences in survey response patterns. *International Marketing Review*, *24*, 127–143.

Epsinheira, P. L., Ferrari, S. L. P., & Cribari-Neto, F. (2008a). On beta regression residuals. *Journal of Applied Statistics*, *35*, 407–419.

Epsinheira, P. L., Ferrari, S. L. P., & Cribari-Neto, F. (2008b). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis*, *52*, 4417–4431.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models.* New York, NY: Springer.

Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, *31*, 799–815.

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, *14*, 195–200.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 516–534.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, *7*, 457–511.

Goldstein, H., Browne, W. J., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics*, *1*, 223–231.

Gupta, A. K. & Nadarajah, S. (Eds.). (2004). *Handbook of beta distribution and its applications.* New York, NY: Marcel Dekker.

Gurr, M. (2009). *Partition dependence: Investigating the principle of insufficient reason, uncertainty and dispositional predictors*. Unpublished Honours thesis, The Australian National University, Canberra, Australia.

Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, *68*, 563–583.

Kieschnick, R., & McCullogh, B. D. (2003). Regression analysis of variates observed on (0,1): Percentages, proportions, and fractions. *Statistical Modelling*, *3*, 193–213.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, *46*, 79–88.

Longford, N. T. (2001). Simulation-based diagnostics in random-coefficient models. *Journal of the Royal Statistical Society, Series A*, *164*, 259–273.

McAuliffe, J. D., Blei, D. M., & Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, *16*, 5–14.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, England: Chapman and Hall.

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models* (2nd ed.). New York, NY: Wiley.

Merkle, E., Smithson, M., & Verkuilen, J. (2009). *Hierarchical Bayesian beta models for subjective probabilities*. Paper presented at the Joint Annual Convention for the Society of Mathematical Psychology and the European Mathematical Psychology Group, Amsterdam, August 1–4.

Noël, Y. (2008 August). *When extreme responses are substantial: A generalized beta response model of behavior change*. Paper presented at the Joint Annual Convention for the Society of Mathematical Psychology and the European Mathematical Psychology Group, Amsterdam, the Netherlands.

Noël, Y., & Dauvier, B. (2007). A beta item response model for continuous bounded responses. *Applied Psychological Measurement*, *31*, 47–73.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS.* Hoboken, NJ: Wiley.

Ospina, R., Cribari-Neto, F., & Vasconcellos, K. L. P. (2006). Improved point and interval estimates for a beta regression model. *Computational Statistics and Data Analysis*, *51*, 960–981.

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, *9*, 325–346.

Papke, L., & Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics*, *11*, 619–632.

Pawitan, Y. (2001). *In all likelihood: Statistical modeling and inference using likelihood.* Oxford, UK: Clarendon Press.

Pinheiro, J. C., & Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear models. *Journal of Computational and Graphical Statistics*, *15*, 58–81.

Qiu, Z., Song, P. X. K., & Tan, M. (2008). Simplex mixed-effects models for longitudinal proportional data. *Scandinavian Journal of Statistics*, *35*, 577–596.

Roy, M., & Liersch, M. (2009). *Most people are above average, sometimes*. Unpublished manuscript, Elizabethtown College, Elizabethtown, PA.

SAS 9.1.3. (2006). Cary, NC, U.S.A.: SAS Institute, Inc.

SAS 9.2 (2008). Cary, NC, U.S.A.: SAS Institute, Inc.

See, K. E., Fox, C. R., & Rottenstreich, Y. S. (2006). Between ignorance and truth: Partition dependence and learning in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*, 1385–1402.

Smithson, M. J., Merkle, E., & Verkuilen, J. (2009 August). *Beta regression finite mixture models of polarization and priming*. Paper presented at the Joint Annual Convention for the Society of Mathematical Psychology and the European Mathematical Psychology Group, Amsterdam, the Netherlands.

Smithson, M. J., & Segale, C. (2009). Partition priming in judgments of imprecise probabilities. *Journal of Statistical Theory and Practice*, *3*, 169–182.

Smithson, M. J., & Verkuilen, J. (2006). A better lemon squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*, 54–71.

Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society, Series B*, *51*, 47–60.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2004). WinBUGS Version 1.4.3. Computer software Cambridge, England: Medical Research Council, Biostatistics Unit.

Steiger, J. H. (2001). Driving fast in reverse: The relationship between software development, theory and education in structural equation modeling. *Journal of the American Statistical Association*, *96*, 331–338.

Tadikamalla, P. R., & Johnson, N. L. (1982). Systems of frequency curves generated by transformation of logistic variables. *Biometrika*, *69*, 461–465.

Vasconcellos, K. L. P., & Cribari-Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics*, *19*, 13–31.

Weiss, R. E. & Cho, M. (1998). Bayesian marginal influence assessment. *Journal of Statistical Planning and Inference*, *71*, 163–177.

Author

JAY VERKUILEN is Assistant Professor of Educational Psychology in the Graduate Center, City University of New York, 365 Fifth Avenue New York, NY 10016 U.S.A.; jverkuilen@gc.cuny.edu. His research interests include nonlinear and generalized linear models, item response theory and paired comparison models.

MICHAEL SMITHSON is Professor in the Psychology Department at The Australian National University, Canberra A.C.T. 0200, Australia; Michael.Smithson@anu.edu.au. His research interests include judgment and decision making under uncertainty, non-linear and generalized linear models, and fuzzy logic methods for the human sciences.