Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Survey Sliders and Visual Analog Scales

Robert Kubinec

New York University Abu Dhabi

March 1st, 2020

Abstract

I propose a new model, ordered beta regression, for data collected from human subjects using slider scales/visual analog scales with lower and upper bounds. This model employs the cutpoint technique popularized by ordered logit to simultaneously estimate the probability that the outcome is at the upper bound, lower bound, or any continuous number in between. This model is contrasted with existing approaches, including ordinary least squares (OLS) regression and the zero-one-inflated beta regression (ZOIB) model. Simulation evidence shows that the proposed model, relative to existing approaches, estimates effects with more accuracy while capturing the full uncertainty in the distribution. Furthermore, an analysis of data on U.S. public opinion towards college professors reveals that the proposed model is better able to combine variation across continuous and degenerate responses. The model can be fit with the R package `brms`.[1]

---

[1]For a reproducible version of this paper, please see the Github repository https://github.com/saudiwin/ordbetareg. To see how to fit the model with R package `brms`, which includes a full range of regression model options including multilevel/hierarchical models, see the vignette available at https://htmlpreview.github.io/?https://github.com/saudiwin/ordbetareg/blob/master/estimate_with_brms.html.

1 Introduction

The increasing sophistication of data collection in the social, medical and behavioral sciences in the last twenty years has led to adoption of slider and visual analog scales as a way to capture nuanced information from human respondents (Cooper and Kagel 2016; Roster, Lucianetti, and Albaum 2015). A visual analog scale (VAS) is a continuous scale with potentially an infinite number of points along which a respondent may make a selection (though in practice, as many as 100 points are often used) (Wewers and Lowe 1990; Myles et al. 2017). As such, it can record much more variation than a Likert scale, which rarely has more than ten categories.

However, despite the increasing popularity of these scales, the statistical methodology used to analyze them is still very much in development. While practitioners often fall back on ordinary least squares (OLS) as a convenient and easily interpretable way to analyze the data, in reality the Normal distribution cannot capture important characteristics of the sample. Respondents often locate responses at either end of the scale, leading to bi- or even tri-modal distributions. As the Normal distribution is by nature symmetric and unimodal, OLS cannot capture the dynamics of the distribution very well, leading to distorted confidence/uncertainty intervals and coefficient estimates, as I show in this paper.

The most natural alternative, and one that has been used in the literature since Ferrari and Cribari-Neto (2004), is the beta regression model as it is a flexible continuous distribution with natural upper and lower bounds. While this clever parameterization permits powerful inference on these bounded, potentially bi-modal distributions, the beta regression's main flaw is that it cannot model observations that are degenerate, i.e., equal to the lower bound of 0 or the upper bound of 1. Various solutions have been proposed, such as transforming the outcome so that all observations are strictly less than 0 or 1 (Smithson and Verkuilen 2006), and more recently modeling 0s and 1s through separate processes via zero-one inflated beta regression (ZOIB) (Ospina and Ferrari 2012).

A new approach is given in this paper that seeks to capture the best points of the existing approaches while simplifying the model so that it is maximally efficient in its use of information in the data. I employ ordered cutpoints, similar in spirit to an ordered logit model, to estimate the joint probability of 0s (the lower bound), continuous proportions, and 1s (the upper bound) in VAS/slider scale data. As only one predictive model is used for all of the outcomes, the effect of covariates is identified with greater precision than existing approaches. The use of cutpoints permits the model to fit even very degenerate distributions with many observations on the bounds.

To compare this new model, simulations are employed to examine how each of the previous approaches compares along an array of criteria, including root mean squared error (RMSE), predictive validity, and the

uncertainty and bias in the estimation of marginal effects defined over the whole distribution. The results show that while OLS is able to compete in the RMSE category, its inability to capture higher-order moments of the distribution, particularly kurtosis, make its inferences strongly influenced by the number of degenerate responses. The ordered beta regression model, on the other hand, avoids this vulnerability but is competitive with OLS in terms of estimating precise and accurate marginal effects.

To apply the model, I use data from the Pew Research Forum American Trends Panel survey of U.S. residents to examine the drivers of affect towards college professors. As the survey employs feeling thermometers, a type of VAS scale that is bounded at 0 and 100, the issue of responses at either end of the scale is a pronounced problem. The analysis shows that although the ordered beta regression model and the ZOIB model are similar in many respects, the estimated effects of predictors diverge when the number of degenerate responses varies with the values of the predictor. Unhelpfully, the ZOIB model conflates the effect of a given covariate with the number of degenerate responses at each level of the covariate. The ordered beta regression model, on the other hand, produces a much more straightforward interpretation in which the effect of the covariate is equal to an average effect on the feeling thermometer combining both the degenerate and continuous values.

2 Background

As online and digital data collection possibilities increase, so has the employment of slider scales capable of capturing finely nuanced datasets along a dimension of interest. While too broad to list all the possible applications here, slider/VAS scales have been employed prominently in medical pain research (Roster, Lucianetti, and Albaum 2015; Myles et al. 2017), psychological research (Monk 1989; Lee, Hicks, and Nino-Murcia 1991), and political science and economics (Cooper and Kagel 2016; Nelson 2008; M. Liu and Wang 2015). As the possibilities of dynamic interactive instruments increases, these scales are likely to become even more sophisticated, such as permitting respondents to rank options along the continuous scale. As such, it is an important empirical domain for applied statistical analysis, and one that has only been recently examined in the literature.

The standard approach for modeling this type of variable is OLS regression as the variable is at least in part continuous. OLS, as the maximum entropy estimator for any continuous distribution with finite variance (Jaynes 2003), is likely to capture at least some of the relevant features of the distribution. The more "Normal" the distribution, the more likely this approximation will give interpretable answers. However, Smithson and Verkuilen (2006) raise important questions about this application of OLS to upper and lower-bounded dependent variables. They argue that OLS' failure to capture higher-order moments of the distribution represents a serious shortcoming because these moments, such as skewness and variance, may well affect

what can be learned from the model and even the theoretical questions one can ask. As I describe in the empirical analysis, being able to model the full distribution of the outcome allows the analyst to pose more detailed questions about clustering and heterogeneity about respondents' opinions.

As a result, beta regression has become an increasingly popular technique, though it likely is still a minority preference when handling this particular kind of outcome. The main drawback of beta regression, as mentioned earlier, is that it cannot handle 0s and 1s (or whatever the bounds of the scale may happen to be). The reason for this limitation is that the beta distribution can be understood as the prior conjugate distribution for a binomial probability, which is never strictly equal to 0 or 1. However, while probability theory discounts such extreme events from occurring, human beings are much less sanguine. Particularly when there is a reason to expect extreme stimuli in the population of interest, humans have no qualms about expressing complete certainty in either a positive or negative direction.

There are two straightforward ways to handle zeros and ones. The first is to simply drop these responses and model the remaining data. If the count of 0s and/or 1s in the data is small, this strategy would seem reasonable. However, this naturally represents an unfortunate loss of data that will only get worse as the number of 1s and 0s increases. A more sophisticated strategy is to normalize the outcome within a set of bounds that are close to, but never equal to, one. The formula from Smithson and Verkuilen (2006) that has received considerable attention from researchers is as follows. For a given outcome $y_i \in [0,1]$, $i \in \{1, 2, ...N\}$, define a normalized outcome y_j:

$$y_j = \frac{y_i(N-1) + 0.5}{N}$$

The distribution of y_i is nudged so that the values can never reach 0 or 1. This transformation permits beta regression modeling without any need for further modeling choices. As such, it is a computationally simple and straightforward solution. However, it is not immediately clear what ramifications this transformation has on inferences as the transformation is non-linear. Ultimately, and as I show in this paper, this transformation tends to deflate the effects of predictors on the outcome and has worrying distortions in how it affects the estimation of sample characteristics.

The shortfall in existing strategies is what led to the development of the zero-one inflated beta regression (ZOIB) model, an approach that this paper builds upon. While Ospina and Ferrari (2012) proposed a zero or one-inflated beta regression, where a discrete process could be used to model either category separately, F. Liu and Kong (2015) show how to model both 0s and 1s in a zero-one inflated beta regression model with two distinct processes for degenerate responses. Their approach is worth considering in detail as it offers a convenient jumping off position for the model I intend to explicate. Given a slider/VAS response y_i

with observations at the bounds of the scale, they estimate three separate probabilities which I label as α for $Pr(y_i = 0)$, γ for $Pr(y_i = 1)$, and δ for $Pr(y_i > 0 \cap y_i < 1)$. Given these probabilities, we can define a conditional distribution over y_i that depends on the realization of y_i in these three mutually exclusive outcomes, along with parameters μ and ϕ to model the continuous outcomes via the Beta distribution (defined below):

$$
f(y_i | \alpha, \gamma, \delta, \mu, \phi) = \begin{cases} \alpha & \text{if } y_i = 0 \\ (1 - \alpha)\gamma & \text{if } y_i = 1 \\ (1 - \alpha)(1 - \gamma)\text{Beta}(\mu, \phi) & \text{if } y_i \in (0, 1) \end{cases}
\tag{1}
$$

where the Beta distribution is defined in as follows:

$$
f(y_i \in (0, 1); \omega, \tau) = \frac{\Gamma(\omega + \tau)}{\Gamma(\omega)\Gamma(\tau)} y_i^{\omega - 1}(1 - y_i)^{\tau - 1}
\tag{2}
$$

To directly estimate the mean of the beta distribution, we can substitute the parameters μ and ϕ (dispersion) for the shape parameters ω and τ:

$$
\mu = \frac{\omega}{\omega + \tau}
\tag{3}
$$

$$
\phi = \omega + \tau
\tag{4}
$$

Returning to the ZOIB model in (1), we can see that the Beta distribution is being deflated by the probabilities α and γ such that the contribution of the Beta distribution cannot exceed $(1 - \alpha)(1 - \gamma)$, i.e., δ. To parameterize the model, we can include regressors $\alpha = g(X'\beta_\alpha)$, $\gamma = g(X'\beta_\gamma)$, and $\mu = g(X'\beta_\delta)$ for a given matrix of covariates X. These linear models are rescaled with the inverse logit function $g(\cdot)$ to map on to $(0, 1)$. While the covariates X for each of the sub-models could be different or shared, the parameters β_α, β_γ, and β_δ need to be distinct for each sub-model as the three processes are functionally independent. These categories are not ordered, and as such the outcomes of y_i are exchangeable (can be re-ordered) for any given value of α, γ and δ.

While this point is rather subtle, it is very important for the modeling exercise that follows. It is possible in this model for $Pr(y_i = 0)$ and $Pr(y_i = 1)$ to both increase independently of $Pr(y_i > 0 \cap y_i < 1)$. This independence can isolate heterogeneity in either end of the slider scale such that the decision to choose a 1 or a 0 are distinct choices with no necessary connection to each other.

This exchangeability, however, comes at a cost. As the number of covariates X increases, the number of parameters necessarily triples (assuming that all covariates are used to predict all parts of the model). The

independence between sub models means that X could positively predict $Pr(y_i = 0)$ and negatively predict $Pr(y_i = 1)$ in the same model without contradiction. While this formulation is potentially quite powerful, and it solves the problems of fit and ad-hoc transformations mentioned earlier, it results in a model that does not have a single effect of X and has an abundance of parameters.

As such, the model seems to go beyond what many practitioners want from the model, which is the ability to evaluate the ability of covariates of interest to affect the combined distribution of y_i. The functional independence of outcomes, while a potentially powerful way to examine in depth the processes that lead to degenerate outcomes, is not necessarily the object of interest to scholars who collect this data. There should be a way to more parsimoniously handle the inflation of 0s and 1s without needing to have exchangeable outcomes. The probabilities of the three categories do not need to be independent since the same set of respondents in a survey or treatment group are all employing the same scale to produce the outcome in question.

3   Model

To resolve this problem, I present *ordered* beta regression. The main difference between this approach and the ZOIB model is to induce dependence between the three probabilities α, γ and δ. To do so, I borrow ideas from the literature on the ordered (cumulative) logit model (McCullagh 1980). We can think of the distribution of y_i as being a realization of a latent cumulative distribution function $F(y_i^*)$ defined by a set of ordered cutpoints $k \in \{1, 2\}$ where $k_1 < k_2$. As we only have one latent variable y_i^*, we can also have a single set of regressors $X'\beta$ that predict this latent variable. At very low values of y_i^*, we will have a discrete outcome $y_i = 0$, and at very high values of y_i^* we will have a discrete outcome $y = 1$. For intermediate values of y_i^*, we will observe the Beta-distributed outcome $y_i \in (0, 1)$.

Using y_i^* and cutpoints k, we can now re-defined the probabilities α, γ and δ where:

$$
\left\{
\begin{array}{l}
\alpha = 1 - g(X'\beta - k_1) \\
\delta = [g(X'\beta - k_1) - g(X'\beta - k_2)] \operatorname{Beta}(g(X'\beta), \phi) \\
\gamma = g(X'\beta - k_2)
\end{array}
\right\}
\tag{5}
$$

We can see that we can still obtain the necessary probabilities α, γ and δ to combine the 0s, 1s, and proportions into a single distribution. However, unlike in (1), the probabilities are no longer exchangeable, but rather ordered due to the cutpoints. The position of the cutpoints k_1 and k_2 will affect each outcome by either decreasing or increasing the probability of that outcome occurring. As k_1 increases, $Pr(y_i = 0)$ must

increase and $Pr(y_i > 0 \cap y_i < 1)$ must decrease. Similarly, an increase in k_2 will necessarily make $Pr(y_i = 1)$ decrease and increase $Pr(y_i > 0 \cap y_i < 1)$.

The use of ordered cutpoints further implies that separate intercepts are no longer necessary for the linear model as in (1). The linear model is defined implicitly relative to the cutpoints on y_i^* so that no further parameters are needed. This implication of using ordered cutpoints means that only two parameters, the cutpoints themselves, are required in addition to the auxiliary parameter ϕ. Thus only two parameters more than OLS are required to fit this model, improving efficiency and information retrieval.

I can express the model as a log-likelihood for a given distribution of y_i:

$$ll(y_i|K,\beta,\phi) = \sum_{i=1}^{N} \begin{cases} \log\ [1 - g(X'\beta - k_1)] & \text{if } y_i = 0 \\ \log\ [g(X'\beta - k_1) - g(X'\beta - k_2)] + \log \text{Beta}(g(X'\beta),\phi) & \text{if } y_i \in (0,1) \\ \log\ g(X'\beta - k_2) & \text{if } y_i = 1 \end{cases} \quad (6)$$

To consider the model from a Bayesian perspective, I assign weakly informative priors to the parameters. We define these as:

$$\beta \sim N(0,5) \tag{7}$$

$$\phi \sim E(.1) \tag{8}$$

$$k_2 - k_1 \sim N(0,5) \tag{9}$$

I define a difference prior over the cutpoints K because it is difficult to know a priori the location of k_1 and k_2. The weakly informative prior on the difference implies that we expect the difference to lie somewhere between $[-5,+5]$ on the logit scale, which is quite wide. Similarly, the prior on β is weakly informative on the logit scale. The exponential prior on ϕ similarly puts prior mass on a wide range of values due to its long tail.

It is possible to further parameterize ϕ to model higher moments in the distribution. A higher value of ϕ for a given value of μ is associated with extreme values, either 0, 1 or both depending on the value of μ. This kind of information can be useful to analyze in a context where understanding which respondents/subjects tend to choose middle versus extreme values is a research question of interest. To do so we simply replace ϕ in (6) with a set of regressors β_ϕ and a covariate matrix X (the covariates could be shared or different from those used to predict the mean of the distribution). Fitting covariates to ϕ can also dramatically improve the fit of the model as is shown in the empirical example.

I can now define a joint log posterior distribution over y_i conditional on the log-likelihood function and set of parameters:

$$\log p(K, \beta, \phi | y_i) \propto \sum_{i=1}^{N} \log p(K) + \log p(\beta) + \log p(\phi) + ll(y_i | K, \beta, \phi) \tag{10}$$

where $\propto$ indicates that the posterior is calculated proportional to the normalizing constant.

4   Estimation

Estimation of the model is done using Hamiltonian Markov Chain Monte Carlo with the software Stan (Carpenter et al. 2017). The model converges fairly rapidly with less than 1,000 iterations on simulated data. In addition to sampling the model above, I also draw from the posterior-predictive distribution of y_i, denoted $\int_\Theta p(\tilde{y}_i | \theta) p(\theta | y_i) \mathrm{d}\theta$, conditional on the posterior estimate of the model parameters (denoted θ) for a given number of MCMC draws S. To do so I first sample a categorical outcome $y_{repO} \in \{1, 2, 3\}$ based on an ordered categorical tuple of the probabilities α, γ and δ:

$$y^s_{repO} \sim \mathrm{Cat}(\{1, 2, 3\}, \{\alpha^s, \delta^s, \gamma^s\}) \tag{11}$$

If y^s_{repO} is equal to 1 or 3, then assign 0 and 1 respectively to y^s_{rep}:

$$y^s_{rep} = 0 \text{ if } y_{repO} = 1 \tag{12}$$

$$y^s_{rep} = 1 \text{ if } y_{repO} = 3 \tag{13}$$

I then draw from the Beta distribution if $y_{repO} = 2$.

$$y^s_{rep} \sim \mathrm{Beta}(\mu_s, \phi_s) \tag{14}$$

In addition to this predictive distribution, I also examine measures of model fit. I use an estimate of leave-one-out (LOO) predictive density in which the posterior predictive distribution is evaluated by dropping a data point y_i, estimating the model, and predicting the held out y_i. Given that this measure is computationally challenging with Bayesian inference, I employ an approximation from Vehtari, Gelman, and Gabry (2016), the Pareto-stabilized important sampling (PSIS)-LOO predictive density:

$$el\hat{p}d_{psis_{loo}} = \sum_{i=1}^{N} \log \left(\frac{\sum_{s=1}^{S} \omega_i^s p(y_i|\theta^s)}{\sum_{s=1}^{S} \omega_i^s} \right) \tag{15}$$

The ω_i are weights derived from importance sampling of the joint posterior for each data point y_i and smoothed by the Pareto distribution to account for outliers. The resulting quantity can be interpreted as the log density of a future dataset $\tilde{y}_i$ from the "true" data-generating process. Importantly, this quantity can be evaluated on any of the models discussed so long as the same data are used to fit the model.

Finally, I also estimate sample average marginal effects for each parameter c in β on the expected value of y_i. I evaluate these marginal effects through numerical differentiation of $\frac{\partial E(y_i|\beta_{-c},K)}{\partial \beta_c}$, iterating over all elements c in β. I suppress ϕ in the notation because it does not by definition factor into the calculation of the expected value.

5   Simulations

To compare the models, I simulate data in a manner consistent with the distribution by using the formula described above. Because the results of simulations can be sensitive to the particular values of parameters, I draw from a broad range of possible values to simulate the ordered beta regression model. For a given covariate effect β_x, scalar dispersion parameter ϕ, and cutpoints $c \in \{1, 2\}$, these ranges are defined as:

$$\beta_x \sim U(-2, 2) \tag{16}$$

$$\phi \sim U(0.5, 4) \tag{17}$$

$$c_1 \sim U(-5, -1) \tag{18}$$

$$c_2 \sim c_1 + U(0.5, 5) \tag{19}$$

The total number of observations N is also sampled from a uniform distribution between 100 and 2000. For this simulation, 10,000 independent variates were drawn from the distributions above. Because the broad bounds on the parameters permit relatively sparse distributions, such as no zeros or ones (or no continuous values), any set of parameter values that could not produce at least five observations from the zeroes, ones and continuous parts of the distribution was discarded. This procedure ensures that a range of models can all be compared to each other without encountering estimation problems.

In addition to the ordered beta regression model specified above, four other models were fit to the data. Two kinds of beta regression models were fit, one using the transformation specified in Smithson and Verkuilen

(2006) and the second only using the continuous values of the distribution (discarding degenerate outcomes). The two other models include the ZOIB model specified above and a Bayesian version of OLS.

All of the estimated values came from Hamiltonian Markov Chain Monte Carlo chains using Stan (Carpenter et al. 2017). All models are run with two chains of 1,000 iterations each with 500 samples discarded as warmup in each chain. The use of the same sampler ensures that the results are comparable across models. The beta regression models and the OLS regression were fit with `rstanarm` (Goodrich et al. 2020).

For each draw of the simulation I calculate the marginal effect of β_x on the outcome to allow for easier comparability across models, in particular for OLS, which does not produce any covariates on a scale other than the data. Marginal effects have the additional advantage of being very easy to interpret as they represent the partial second derivative of the parameter β_x on the expected value of the outcome, which in this case is a number bounded between 0 and 1.

I compare the simulations based on a range of criteria, which are reported in Figure 1. The top left corner of the figure shows the coverage rate of 5%-95% high posterior density interval for the estimated marginal effects, i.e., whether these quantiles of the posterior happened to include the true marginal effect. Theoretically, this value should be equal to 90% in expectation, which is represented by the vertical dotted line in the figure. As can be seen, the ordered beta regression is almost exactly equal to the theoretically correct value. The ZOIB model and OLS have marginally lower coverage rates, while the two types of beta regressions have quite low coverage rates, below 50% of the simulation draws.

The upper right panel of 1 explores M-errors (Gelman and Carlin 2014), which are defined as the ratio of the estimated marginal effect to the true marginal effect. This statistic can capture bias across the distributions, as a completely un-biased estimator would equal exactly 1. As can be seen, in finite samples this is not so, although the ordered beta regression model approaches 1 (the dotted line). While the ZOIB model and OLS both perform well on this statistic, the beta regression model on continuous values is wildly inflated, meaning that it is likely to over-estimate the effect on average. The beta regression on transformed values, however, is very likely to under-estimate the marginal effect.

Immediately below the M-error panel is the S-error panel, which is defined as the proportion of estimated marginal effects that have a different sign than the true marginal effect (Gelman and Carlin 2014). While ordered beta regression performs the best, all of the models are relatively close, except for beta regression on continuous values, which has nearly double the incorrect sign rate.

The bottom right panel shows root mean square error (RMSE), defined as the square root of the squared residuals in the model. In this metric, ordered beta regression marginally out-performs competitors, but OLS and ZOIB are both very close. The beta regression on continuous values is artificially lower on this

metric because extreme values were excluded, and the transformed beta regression shows poorer performance because RMSE was evaluated on the original (untransformed) outcome.

The middle and lower-left panels in the figure show results for $\hat{elpd}_{PSIS-LOO}$ (the statistic cannot be calculated for the beta regressions because the outcome must be the same across models). In a departure from the other metrics, ordered beta regression does not receive the highest absolute $\hat{elpd}_{PSIS-LOO}$ scores (middle panel) nor does it rank the highest on average (bottom left panel). This discrepancy suggests that in terms of predictive validity, or the likelihood of the model being true given the observed data, OLS in fact provides a better fit to the data. However, the average ranks suggest that OLS does not always fit better, as its average rank is only slightly better than ordered beta regression. This finding suggests that OLS' performance is dependent on certain characteristics of the distribution which are worth considering further.

As the RMSE metric shows in Figure 1, OLS does an excellent job capturing the average value of the outcome. However, because OLS is symmetric, unimodal, and defined over all real numbers, it does not perform so well at capturing other sample moments. Given the peculiar nature of the distribution defined above, which can be effectively tri-modal, it is worth considering kurtosis as an important sample characteristic. Kurtosis measures the heaviness of the tails of a distribution, such that more extreme values increase kurtosis. The Normal (Gaussian) distribution is defined as having a fixed kurtosis value of 3, which means that it cannot by definition capture any higher or lower kurtosis values.

To show kurtosis may be affecting the quality of OLS estimates, I plot the sample kurtosis from each simulation draw against $\hat{elpd}_{PSIS-LOO}$ scores in Figure 2. It is important to note that higher $\hat{elpd}_{PSIS-LOO}$ are associated with greater predictive validity. As can be seen, at lower values of kurtosis, OLS tends to have worse predictive validity than ordered beta regression. However, once kurtosis values reach a threshold of approximately 9, OLS starts to perform better than ordered beta regression.

The reason for this differing performance has to do with what kurtosis measures: the extent to which the density of the probability distribution (i.e., the amount of variance) is concentrated in the tails of the distribution. For slider scale/VAS data, higher kurtosis is a sign of a unimodal, primarily continuous distribution. Because the Normal distribution is uni-modal and symmetric, it does a much better job capturing the moments of this distribution. However, at lower values of kurtosis where there is a pronounced U-shape, with fewer continuous values and large numbers of degenerate responses, OLS performance suffers dramatically.

As a result, even without looking at empirical examples, it is difficult to recommend OLS as a general approach for analyzing this type of data. The fact that model performance is so strongly determined by a fairly common feature of the data-generating process means that inferences will tend to be obscured by whether sample kurtosis is high or low. Furthermore, while in this simulation the use of a single continuous
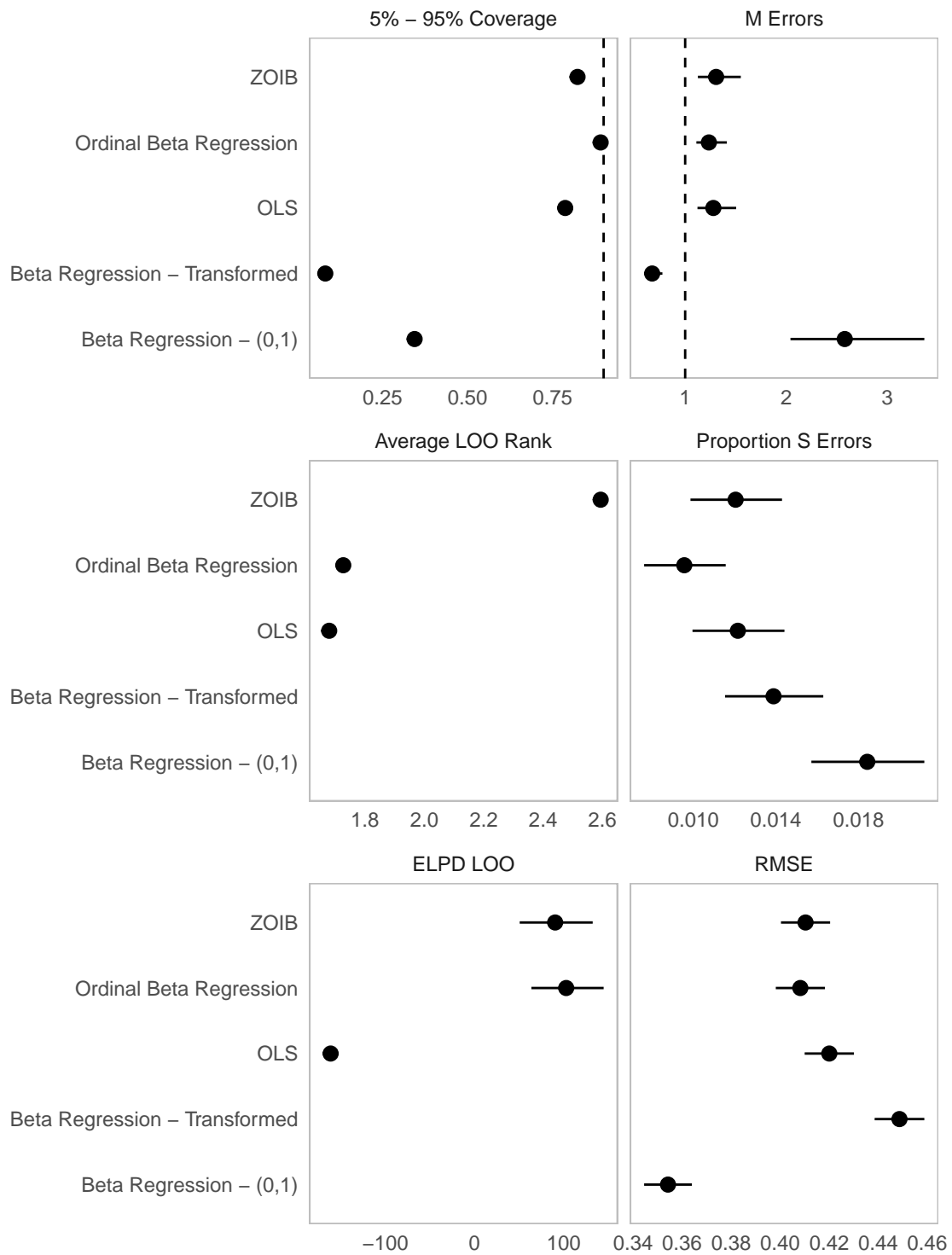
Figure 1: Comparison of Simulation Performance Results

predictor means that we can evaluate kurtosis on the entire sample, in realistic situations, it is *conditional* kurtosis which matters, or the distribution of the outcome conditional on covariate values. While overall sample kurtosis could be quite high, conditional kurtosis could still be quite low for a subset of the respondents who tend to report extreme values, affecting OLS' estimation of that value of the covariate. As will be shown in the empirical example to follow, it is fairly easy to find this kind of conditional heterogeneity between groups of human subjects, and it is difficult to diagnose these issues through simple visual inspection of histograms.
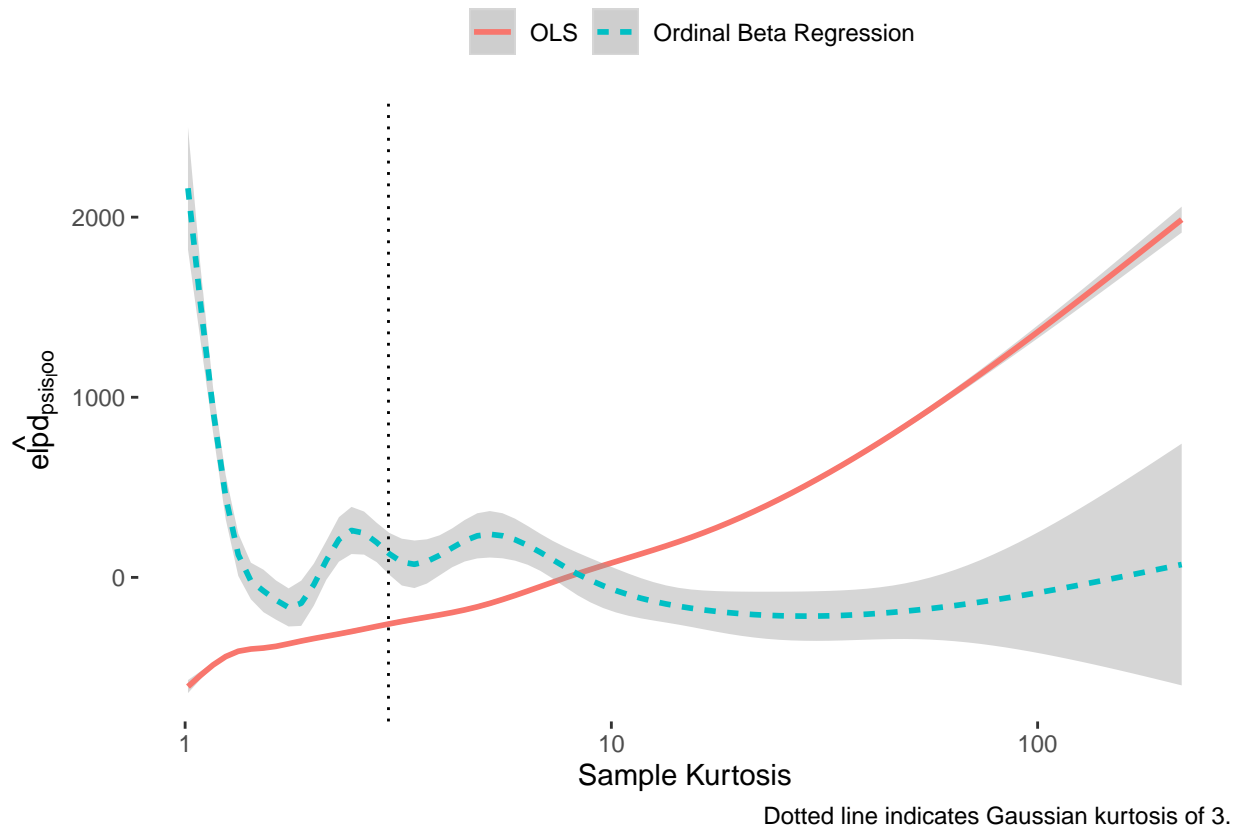

Figure 2: Comparison of $e\hat{l}pd_{psis_{loo}}$ by Sample Kurtosis Values

6 Empirical Examples

To analyze the model in an applied setting I use data from the Pew Forum American Trends Panel. This survey tracks a sample of Americans over time on a variety of political and social opinions. In this section I study the August 2017 wave of the panel, which consisted of 4,971 U.S. adults via an online sampling frame.

As I am interested in studying the effect of certain variables on an outcome in the survey rather than making population inference via sample proportions, I ignore the sampling design in the analysis that follows.[2]

The main dependent variable of interest is a feeling thermometer question in which respondents were asked to rate a number of different social actors based on a 0 to 100 scale, where higher values indicate warmer feelings. The thermometer I will analyze is the respondents' affect towards college professors. As this thermometer was only shown to half of the panel, the final dataset amounts to 2,538 complete observations. A histogram of the dependent variable is shown in Figure 3.
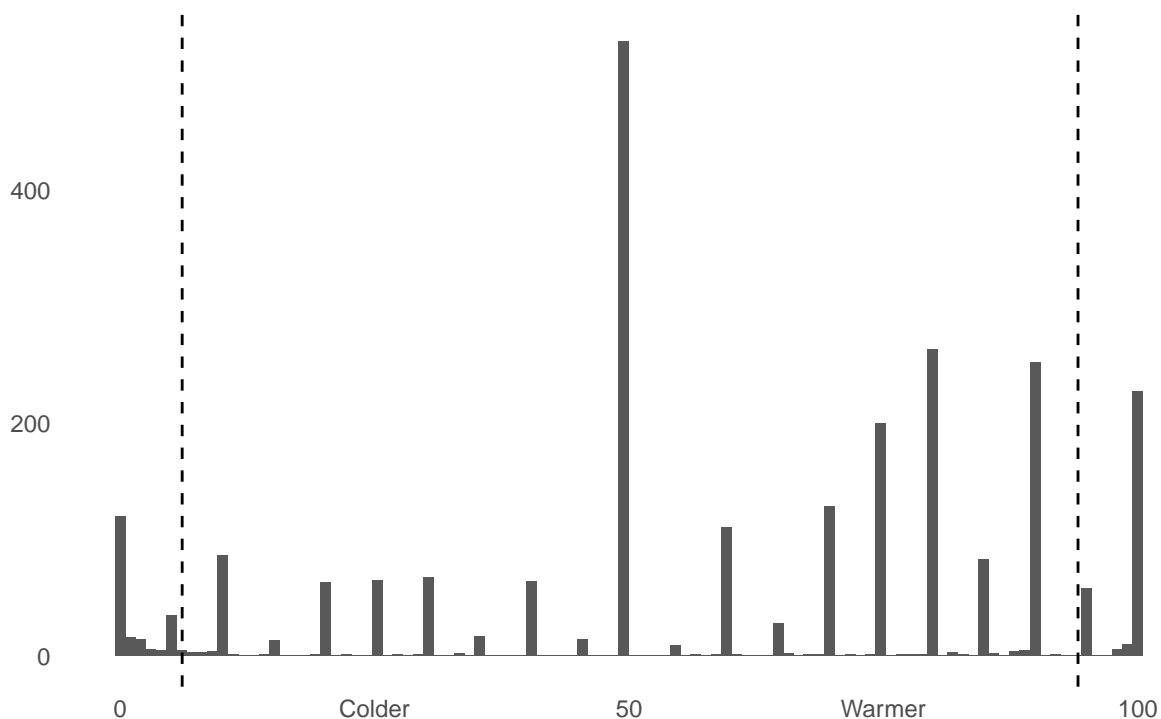
Figure shows the distribution of 2538 non–missing survey responses.
The two vertical lines indicate the estimated cut points from an ordered beta regression model
where responses have a high probability of being considered degenerate (i.e. close to 0 or 1).

Figure 3: Feeling Thermometer Scores for College Professors from the August 2017 wave of the Pew Forum American Trends Panel

Figure 3 reveals some of the very common issues with these sorts of scales with human subjects. The distribution is effectively tri-modal, with a large number of respondents with presumably neutral (score of 50) feelings. A sizable minority have very strong feelings in favor of college professors (perhaps a relief to

_____

[2]As variables that affect sample inclusion, such as education and political affiliation, enter into the model, sample inclusion probabilities are implicitly controlled for when calculating the effects of interest [@gelman2007].

many of the readers of this article), with a second mode at or near 100. Finally, there is a thankfully smaller yet still noticeable mode at 0. I will consider each of the strategies attempted in the simulation to model this dependent variable effectively given a set of covariates relevant to socio-political opinion, including age, sex, race, party identification, political ideology, income, approval of President Donald Trump and (of course) education. With the additional covariates, the total number of non-missing observations comes to 1,475. As such, the dataset is of a size typical in social-scientific analyses, particularly of survey data. The two vertical lines in the figure show the estimated cut point locations from a ordered beta regression model, indicating how the model tends to group responses together at the low and high ends of the scale that conforms to intuitions about which values are considered to be extreme.

I fit each of the models in the simulation to the empirical data. Unfortunately, as I do not know the "true" marginal effects, I cannot calculate M-error or S-error rates, nor can I directly compare the variance of the estimates to each other. However, I do calculate RMSE and kurtosis for each model, which is shown in Figure 4. Similar to the simulation, OLS tends to over-estimate kurtosis while the beta regression on transformed values over-estimates the moment. Interestingly, the beta regression with only continuous values comes closest to the sample kurtosis, though both the ZOIB model and ordered beta regression are quite close. Similar to the simulation, OLS does quite well at minimizing squared errors, out-performing both the ZOIB and ordered beta regression, albeit only slightly.

While not included in the coefficient plot, OLS did have superior performance over other models in terms of $\hat{elpd}_{psis_{loo}}$. As documented in the simulation section, this is likely because of the pronounced mode at 0.5, which helps OLS to fit the range of data values. However, this form of predictive validity obscures much of importance for actual analysis. Figure 5 shows the posterior predictive distributions (draws of the y_i from the posterior values of the parameters, denoted y_{rep}) for each of the models. In this figure, the empirical cumulative density of the true outcome is denoted by a thick blue line, while the individual draws from the posterior predictive distribution are in thin gray lines, indicating the approximate range of uncertainty over y_i. As can be seen, OLS does well by closely fitting the mode at 0.5, but tends to under-predict at the extremes, especially for zeroes in the distribution.

By comparison, ordered beta and ZOIB fit more of the distribution (the thick blue line is inside the gray shaded region), with the exception of the mode at 0.5. The transformed beta regression shows the worst fit overall, with a large number of values outside of the posterior predictions. As such, the question becomes whether the behavior of the ZOIB and the ordered beta regression is preferable to that of OLS. In this case, the answer is almost certainly yes.

For OLS, which is defined over all real numbers, the value of 0.5 is not intrinsically different than other numbers. However, given that this is slider scale/feeling thermometer data, a value of 0.5 is quite different
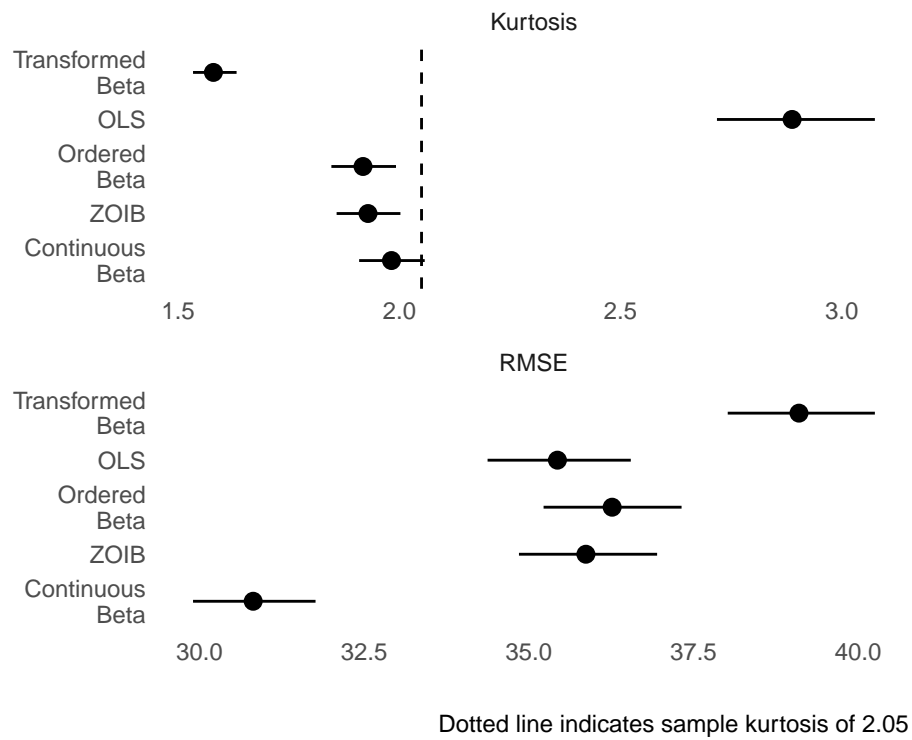
Figure 4: Comparison of Model Diagnostics for Regression of Thermometer Ratings Towards U.S. College Professors

from a value of 0 or a value of 1. Substantively, the midpoint of the scale implies uncertainty: the respondent does not have pronounced views on the subject. For beta-related distributions, this uncertainty is reflected in the shape of the posterior such that it is unlikely that a pronounced mode at 0.5 will re-produce itself. Instead, the beta distribution provides a tighter fit to values that contain more information about respondent views on the $(0, 0.49)$ and $(0.51, 1)$ intervals.

It is possible to capture the pronounced mode at 0.5 using ordered beta regression by adding additional ordered cutpoints. I do not do so here as more precisely modeling this mode is unlikely to affect the substantive results in the model given that a value of 0.5 communicates little information about respondents' views; as such, additional cutpoints would result in additional complexity with little additional information gain. In all, the visual analysis provides yet another reason to reject OLS as a suitable model for this type of data as it will down-weight values at the extremes, which contain a lot of information about respondent preferences, in favor of values in the middle which contain relatively little information about respondent preferences. As can be seen, absent this one mode, both ordered beta regression and ZOIB provide reasonable coverage of the full empirical cumulative density of the posterior relative to the observed data.

Because I do not have access to the true underlying data generative process, I compare estimated marginal effects from each model. These are reported for the different predictors in the model in Figure 6. These results are intriguing both for their similarities and differences. For some covariates, such as age, sex and political ideology, the models return almost identical results (even for the beta regression on continuous values). However, for other variables, the models diverge quite significantly. The most pronounced differences are for education and income where the effects for ordered beta regression and the ZOIB are noticeably far apart, and sometimes even of different signs.

This difference is remarkable given the previous findings showing that ordered beta regression and the ZOIB are similar in many respects, although ordered beta regression tends to be a more efficient estimator. In Figure 6 we can see evidence of the issue raised earlier in the article about the ZOIB modeling 0s, 1s and continuous values as *separate* processes. Even though the ZOIB provides a generally similar fit to the data as ordered beta regression, the interpretation assigned to covariate effects is not generally the same. In this case, the difference appears to arise from how the two models interpret respondent decision-making by re-weighting the importance of degenerate versus continuous responses.

Table 1 shows the separate calculations of average thermometer rankings for continuous and degenerate responses by education and income, along with the proportion of values in each category that contain degenerate responses. Comparing first for average thermometer values, there does appear to be evidence that the unadjusted scores are higher for less educated and poorer respondents. Comparing continuous and degenerate responses in the middle two columns, the difference between categories is much more pronounced
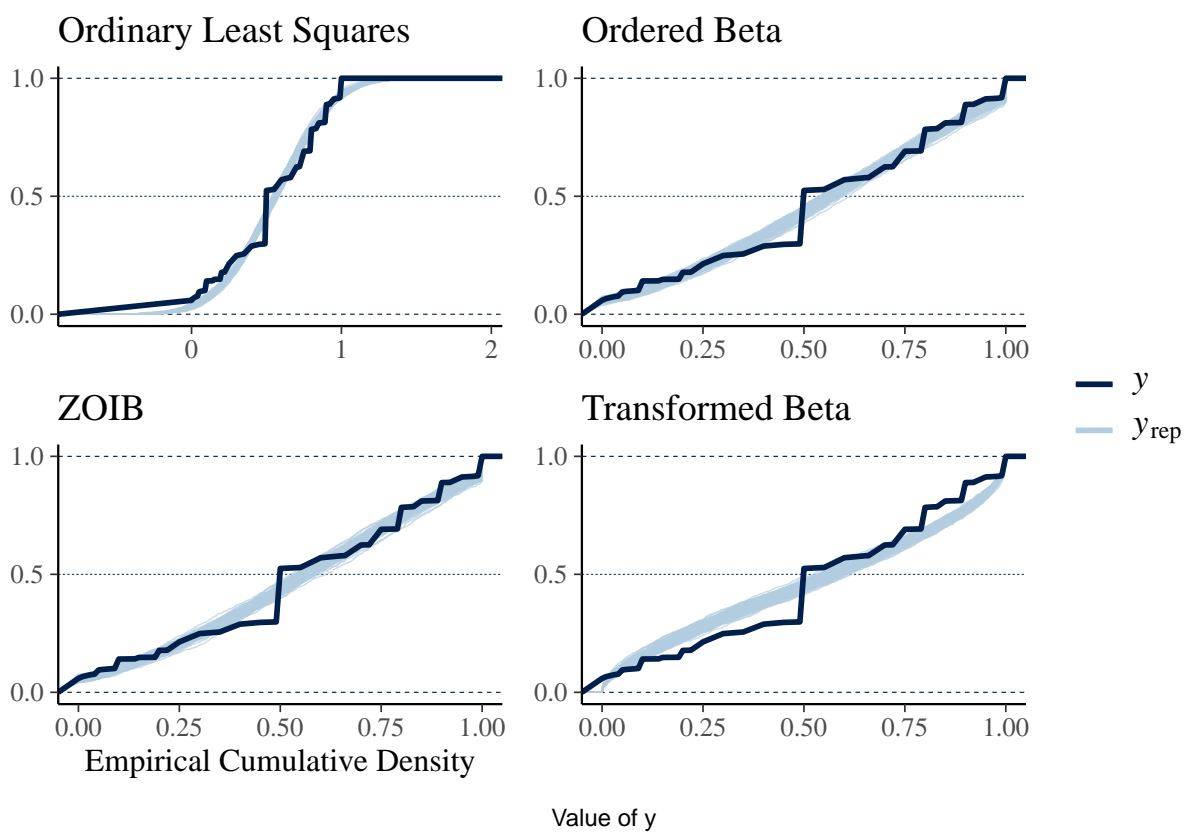
Figure 5: Comparison of Sample to Posterior Empirical Cumulative Distribution for Models
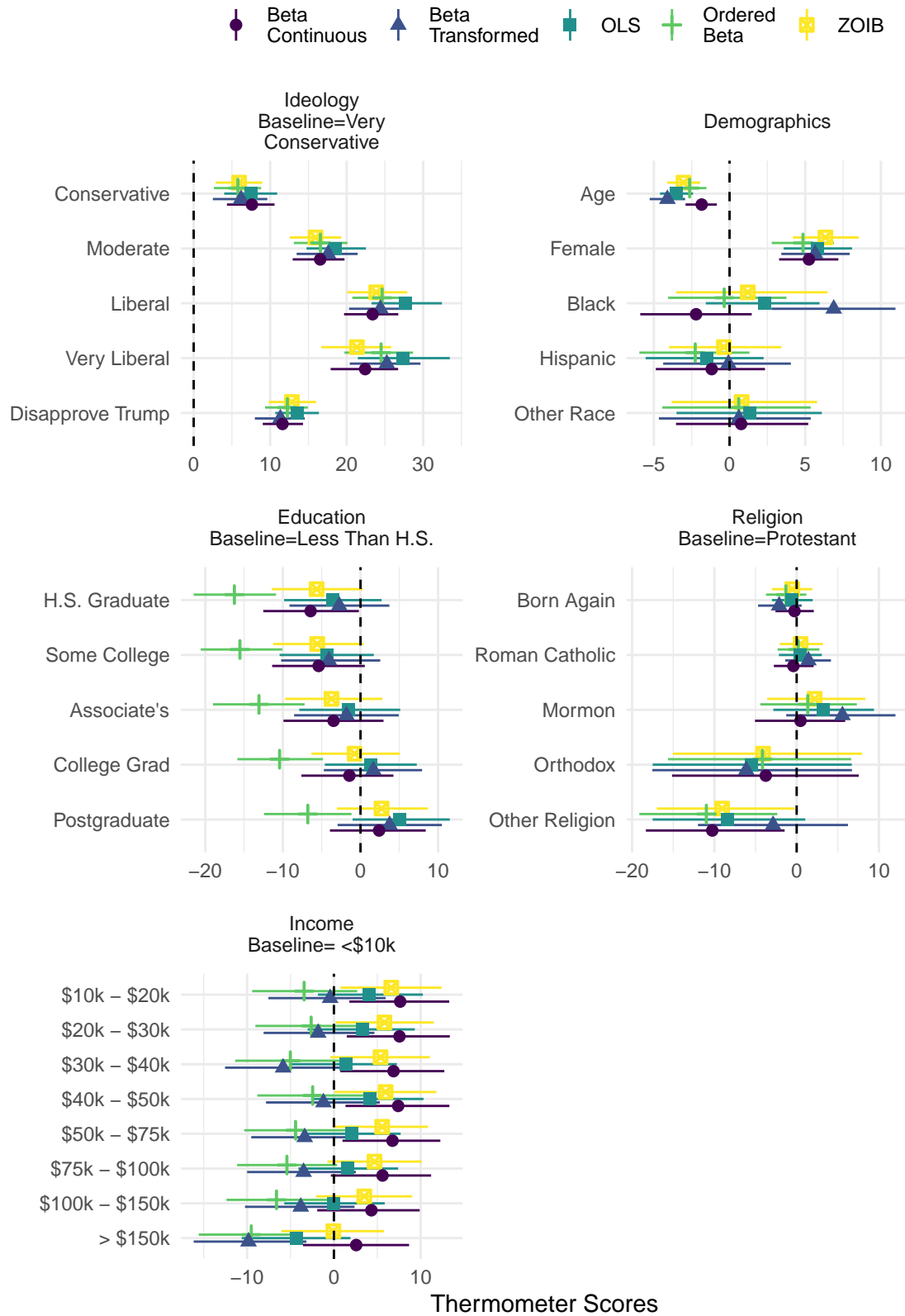
Figure 6: Estimated Marginal Effects from Regression of Feeling Thermometer Ratings Towards College Professors

for degenerate responses than for continuous responses. In particular for income, poorer respondents are much more likely to give higher scores among degenerate values (an average of 0.81) to professors while wealthier respondents give a dispiriting 0.31. For the continuous values, on the other hand, there are small to weak differences between poorer and richer respondents.

These comparisons can help us understand the divergent marginal effects reported in Figure 6. Beta regression with only continuous values shows no differentiation among education categories, while for income categories it shows an *opposite* relationship, with richer respondents holding stronger views for college professors. OLS, which also downplays values at the extremes, is close to the continuous beta distribution in its interpretation of these effects. The transformed beta regression is close to the continuous distributions for education, and quite confusingly, close to the ordered beta regression for income (as has probably become clear by this point, this transformation has potentially severe consequences on correctly modeling the response).

While it is intuitive to understand why OLS and the continuous beta regression would diverge from ordered beta regression, I still have not explained why the ZOIB does so as well. The answer can be seen in the right-most column in Table 1. Not only do poorer and less educated respondents give higher scores among the degenerate responses for college professors, but they are also more likely to give degenerate responses as a proportion of their total responses. As a result, the ZOIB model *down-weights* the degenerate responses for the poorer/less educated respondents while it *up-weights* continuous responses for the richer/more educated respondents.

This confusing behavior can be visualized by examining the model coefficients for each of the outcome's components, $Pr(y_i = 0)$, $Pr(y_i = 1)$ and $Pr(y_i > 0 \cap Pr(y_i < 1))$. The effects of income and education on each of these components for the ZOIB and ordered beta regression are shown in Figure 7. First, the top row shows the coefficient values for the predictors of the unweighted Beta regression component. As can be seen, the two models have coefficients that run in the same directions but are different by a constant value. This difference is due to again to the different weights that the two models assign to continuous as opposed to degenerate responses in y_i.

These weights can be seen in the bottom row of the table for the prediction of $Pr(y_i > 0 \cap Pr(y_i < 1))$. While the ordered beta regression shows no particular differences across categories, the ZOIB shows positive values, indicating that richer and more educated respondents are between 5 and 10 percent more likely to use continuous versus degenerate responses. As a result, the continuous values are more influential for the total marginal effect reported in Figure 6.

However, it would not seem that the re-weighting differences on their own would permit the coefficients to diverge as much as they do in Figure 6. The second significant difference between the two models has to

Table 1: Thermometer Ratings by Income, Education and Response Type (Degenerate or Continuous)

Value	Mean Thermometer	Mean Continuous Thermometer	Mean Degenerate Thermometer	Proportion Degenerate Responses
Education				
Less than high school	59.1	58.8	60.0	0.27
High school graduate	51.4	50.3	56.8	0.18
Some college, no degree	50.0	50.7	45.0	0.13
Associate's degree	53.2	53.1	53.8	0.16
College graduate/some postgrad	56.6	55.6	63.0	0.14
Postgraduate	63.4	62.6	69.4	0.11
Income				
Less than $10,000	60.8	52.5	81.0	0.29
10 to under $20,000	61.0	58.0	80.0	0.14
20 to under $30,000	57.8	56.8	63.2	0.16
30 to under $40,000	55.2	57.3	45.8	0.18
40 to under $50,000	57.6	56.4	63.6	0.17
50 to under $75,000	54.6	55.2	51.2	0.16
75 to under $100,000	54.2	54.1	55.6	0.07
100 to under $150,000 [OR]	55.5	54.9	59.4	0.13
$150,000 or more	51.8	54.0	31.2	0.10

do with the middle row in Figure 7, the marginal changes in probability for $Pr(y_i = 0)$ and $Pr(y_i = 1)$. The ordered beta regression returns relatively sensible results for these two probabilities, suggesting that the effects for education and income are mirror images of each other. High school graduates, for example, tend to hold more extreme negative views of professors, and are less likely to hold extremely positive views of professors.

The ZOIB model, on the other hand, reports that high school graduates are simultaneously less likely to have extremely negative views of professors and extremely positive views of professors. Because these are the two ends of the scale, the diverging effects of being a high school graduate means that the *total impact of the degenerate responses on the marginal effect of education is zero*, and the ZOIB marginal effect collapses to beta regression on continuous values.

This effect decomposition is very puzzling because the subjects are individual respondents. How can the average high school graduate simultaneously hold fewer extreme negative views and fewer extreme positive views on professors separately from the graduate's largely negative views over continuous values of the distribution? If the reader is at this point somewhat perplexed, it would appear to be a result of the over-parameterization of the ZOIB model in this situation. We generally do not think of respondents having divergent preferences at different points of the same scale, and as such the more parsimonious ordered beta regression is easier to interpret: high school graduates are more likely to give extreme negative scores for professors, more likely to give negative continuous scores for professors, and less likely to give extreme positive scores. These effects are not all mirror images of each due to the scaling effect of the cutpoint intercepts.

One final note on the bottom row of Figure 7 are the fixed probabilities for $Pr(y_i > 0 \cap Pr(y_i < 1))$. These fixed values are due to the fact that the cutpoints are fixed for the entire sample. It is possible to allow the cutpoints to vary across groups, or even to parameterize the cutpoints with covariates to permit more subtle gradations. However, as the analysis with the ZOIB model shows, it is important to know theoretically why this flexibility is required, as it will affect the interpretation of the underlying estimates. The employment of cutpoints has the advantage that any additional flexibility can be incorporated as-needed rather than allowing for completely independent processes as with the ZOIB.

Finally, Figure 8 shows the effect of covariates on ϕ, or the dispersion/scale parameter in an ordered beta regression. Higher values of ϕ indicate that respondents tend to cluster together in their views; lower values imply respondents tend to move towards the extremes in opposite directions. As can be seen, the effects of predictors on ϕ are intriguing, as wealthier and more educated respondents are more likely to hold much more heterogeneous views, while poorer, less educated and more conservative respondents tend to cluster together. As this figure shows, a properly parameterized beta regression model can provide fascinating insights about human behavior and opinion. For an OLS model, this type of distinction among respondents is impossible
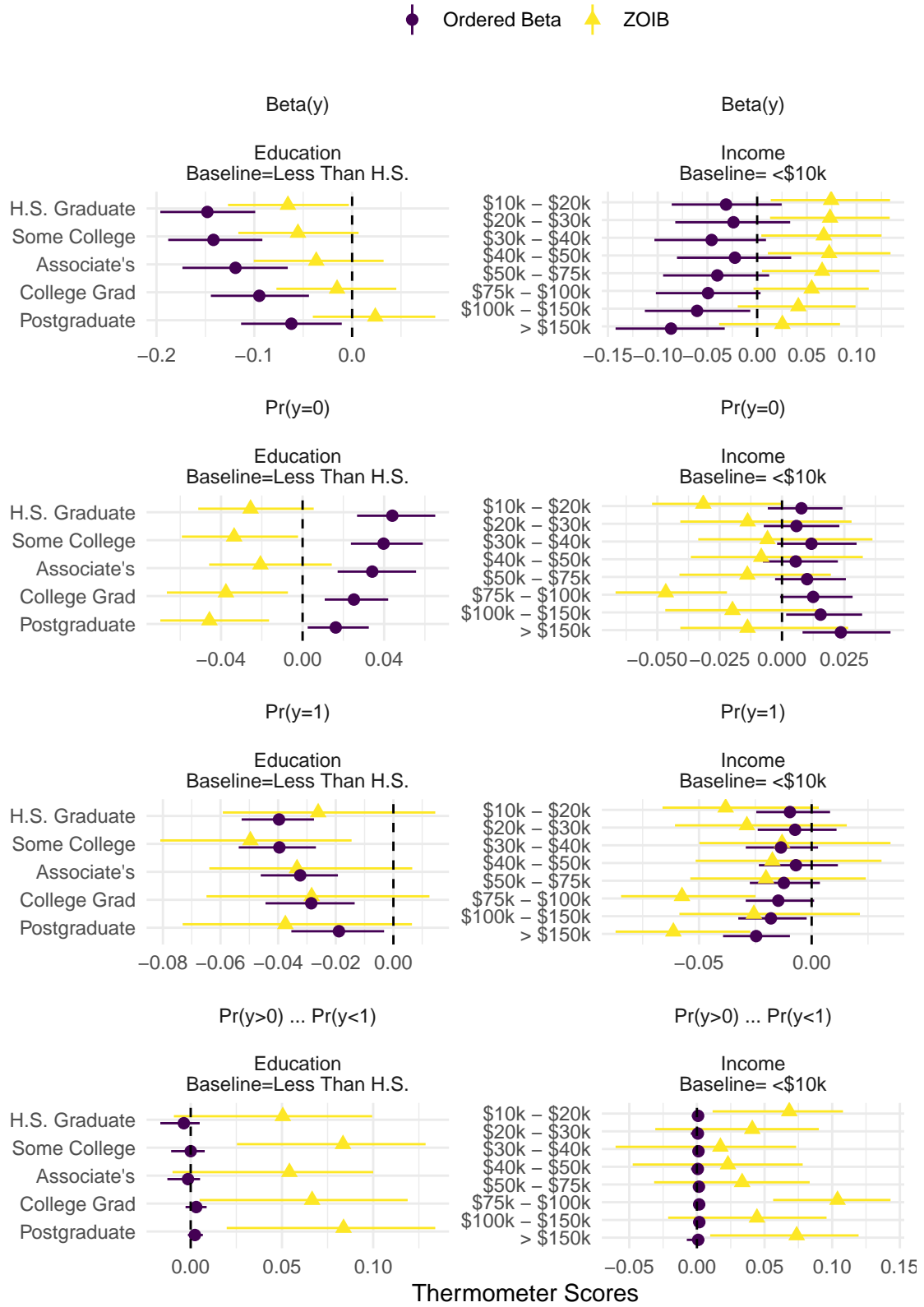
Figure 7: Comparison of Intermediate Probabilities for Ordered Beta Regression and ZOIB

as a U-shape and a cluster (inverted-U shape) can have the same mean and variance.
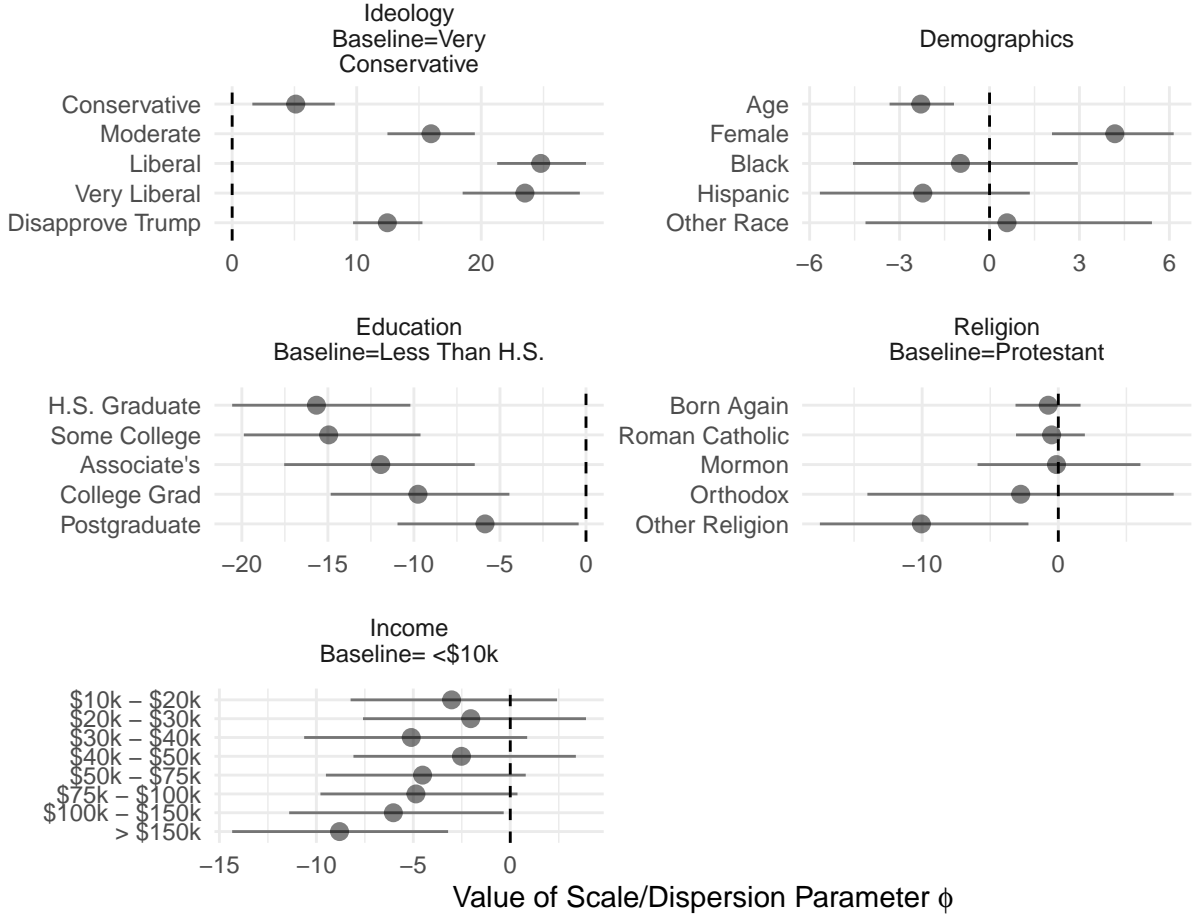

Figure 8: Effects of Covariates on Thermometer Dispersion (Extreme Responses)

7   Discussion

The empirical and simulation-based analysis of the ordered beta regression model show that it is an appealing alternative to existing approaches. The application of the model in this paper is considered primarily in terms of individuals answering survey scales of one kind or another. This aspect is important to evaluating the model, as it suggests that the ordered beta regression's more parsimonious interpretation of the effect of covariates on the outcome is preferable to the more fully parameterized ZOIB. It is assumed that what the analyst desires is the effect of a covariate on the full distribution, which includes both degenerate and continuous responses. There is no particular reason to believe, however, that the effect should vary between

the degenerate and continuous responses, or even between positive and negative degenerate responses.

There are several other points that are worth considering when comparing ordered beta regression to the ZOIB. As Figure 7 reveals, the employment of independent processes for y_i requires much more data and is likely to return large uncertainty intervals with smaller datasets. This difficulty in fitting the model is another reason that the parsimony of the ordered beta regression appears preferable, as it will allow scholars to parse heterogeneity in the different components of the scale with adequate power. The ordered beta regression model's predictions are still well-identified with very few degenerate responses in y_i.

In fact, due to the employment of weakly informative priors, it is perfectly possible to fit the ordered beta regression model to distributions without any degenerate responses. These estimations can still be useful when the analyst believes degenerate responses could have risen (for example, substantial observations very close to the boundary), but did not in the particular sample of data under analysis.

These results also show that predictive validity measures, especially $\hat{elpd}_{psis_{loo}}$ and RMSE, are of limited value when evaluating models with regard to the interpretation of findings. Another assumption in this paper is that the primary objective is to learn about the effect of covariates on the outcome rather than predicting future responses. If that were the aim, it would be worthwhile to consider a broader range of models with considerable predictive power but limited explanatory value, including random forests and neural networks.

8 Conclusion

This paper presented a new model for bounded continuous distributions with considerable observations on the bounds. The principal application of the model is to data collected from slider/visual analog (VAS) scales, in which a respondent has to give a score for a certain item. This paper builds on prior models in the literature incorporating the beta distribution, which has admirable properties for evaluating respondent-level data on this scale. In particular, it allows for multi-modality and considered values near the extremes to be more certain than those values in the middle.

The primary problem this paper solves is to incorporate degenerate responses, or responses at 0 or 1, along with the beta distribution without sacrificing model parsimony. This is made possible by employing cutpoints as scaling intercepts that permit the 0 and 1 values to be jointly estimated with the beta regression. Compared to existing approaches, particularly zero-one-inflated beta regression (ZOIB), ordered beta regression is able to capture a unitary marginal effect of a given covariate on the outcome that is monotonic over the full response range.

For these reasons, I recommend ordered beta regression as a straightforward model to applied researches with

survey scale data. The model can separately predict the probabilities of each of the outcome's components, both continuous and degenerate responses, but will do so in a way that still permits a single interpretation for the effect of the covariate on the outcome. Furthermore, the model's parsimony means that it is able to capture marginal effects at relatively lower cost in terms of data collection than other approaches.

References

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76.

Cooper, David J., and John H. Kagel. 2016. "Other-Regarding Preferences: A Selective Survey of Experimental Results." In *The Handbook of Experimental Economics, Volume Ii*, 217–89. Princeton University Press.

Ferrari, Silvia, and Francisco Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31 (7): 799–815.

Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9 (6): 641–51.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. "Rstanarm: Bayesian Applied Regression Modeling via Stan." *CRAN*.

Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press.

Lee, Kathryn A, Gregory Hicks, and German Nino-Murcia. 1991. "Validity and Reliability of a Scale to Assess Fatigue." *Psychiatry Research* 36 (3). Elsevier: 291–98.

Liu, Fang, and Yunchuan Kong. 2015. "Zoib: An R Package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression." *The R Journal*.

Liu, Mingnan, and Yichen Wang. 2015. "Data Collection Mode Effect on Feeling Thermometer Questions: A Comparison of Face-to-Face and Web Surveys." *Computers in Human Behavior* 48. Elsevier: 212–18. https://doi.org/10.1016/j.chb.2015.01.057.

McCullagh, Peter. 1980. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society* 42 (2): 109–42.

Monk, Timothy H. 1989. "A Visual Analogue Scale Technique to Measure Global Vigor and Affect." *Psychiatry Research* 27: 89–99.

Myles, PS, DB Myles, Wendy Galagher, D Boyd, C Chew, N MacDonald, and A Dennis. 2017. "Measuring Acute Postoperative Pain Using the Visual Analog Scale: The Minimal Clinically Important Difference and Patient Acceptable Symptom State." *BJA: British Journal of Anaesthesia* 118 (3). Oxford University Press: 424–29.

Nelson, Shannon C. 2008. "Feeling Thermometer." *Encyclopedia of Survey Research Methods* 277. Sage Thousand Oaks, CA.

Ospina, Raydonal, and Silvia L. P. Ferrari. 2012. "A General Class of Zero-or-One Inflated Beta Regression Models." *Computational Statistics and Data Analysis* 56 (6): 1609–23.

Roster, Catherine A., Lorenzo Lucianetti, and Gerald Albaum. 2015. "Exploring Slider Vs. Categorical Response Formats in Web-Based Surveys." *Journal of Research Practice* 11 (1).

Smithson, Michael, and Jay Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Independent Variables." *Psychological Methods* 11 (1): 54–71.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2016. "Practical Bayesian Model Evaluation Using Leave-One-Outcross-Validation and Waic." *Statistical Computing*, no. 27: 1413–32.

Wewers, Mary Ellen, and Nancy K. Lowe. 1990. "A Critical Review of Visual Analogue Scales in the Measurement of Clinical Phenomena." *Research in Nursing & Health* 13 (4): 227–36.