# Using Beta Regression to Model Normative Aspects of Prejudice and Political Attitudes: With Applications in R

Mark H. White II[1]

[1] University of Kansas

Author Note

Author note will go here.

Correspondence concerning this article should be addressed to Mark H. White II. E-mail: markhwhiteii@gmail.com

# Abstract

Abstract will go here.

*Keywords:* beta regression, hurdle models, gamlss, norms, social attitudes

Using Beta Regression to Model Normative Aspects of Prejudice and Political Attitudes: With Applications in R

## Statistical Background

### The Beta Distribution

The beta distribution can be used in a generalized linear model when the values of the dependent variable are bounded $0 < y < 1$ (Coxe, West, & Aiken, 2013). The probability density function (pdf) of the beta distribution is determined by two parameters, $\alpha$ and $\beta$, that are called "shape" parameters:

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

where $\Gamma(.)$ is the gamma function. One of the benefits of the beta distribution is that it is flexible and can take a number of distributional shapes (Figure 1).

These parameters are not inherently meaningful to researchers, however. R. Rigby, Stasinopoulos, Heller, and De Bastiani (2017) "reparameterized" the beta distribution so that the two parameters determining the shape of the distribution would be more useful in a regression framework (but see Ferrari & Cribari-Neto, 2004 for a different parameterization). Instead of predicting $\alpha$ and $\beta$, they parameterize (i.e., algebraically rearrange parameters) so that beta regression predicts two different parameters: $\mu$ (called the "location" parameter) and $\sigma$ (called the "scale" parameter), where $\mu = \frac{\alpha}{(\alpha+\beta)}$ and $\sigma = \sqrt{\frac{1}{(\alpha+\beta+1)}}$. $\mu$ is equivalent to the mean, and $\sigma$ is related positively to the variance (note that $\sigma$ is *not* the standard deviation, even though $\sigma$ commonly refers to the standard deviation). The variance is equivalent to $\sigma^2\mu(1-\mu)$, and there are two things to note from this equation: First, the greater the $\sigma$, the greater the variance; second, the variance depends on the mean, which means that beta regression will be naturally heteroskedastic (covered shortly). Using this distribution in a regression framework cannot handle the dependent variable taking on values 0 and 1—how can we model dependent variables in the range $0 \leq y \leq 1$?

## Zero-One Inflated Beta Distribution

R. Rigby et al. (2017) show that the beta distribution can be "inflated" at 0 and 1, meaning the dependent variable contains 0s and 1s (i.e., $0 \leq y \leq 1$), the pdf for this beta inflated distribution, beinf, is:

$$\text{beinf}(y; \mu, \sigma, \nu, \tau) = \begin{cases} p_0 & \text{if } y = 0 \\ (1 - p_0 - p_1) f(y; \mu, \sigma) & \text{if } 0 < y < 1 \\ p_1 & \text{if } y = 1 \end{cases}$$

for $0 \leq y \leq 1$, where $f(y; \mu, \sigma)$ is the beta pdf with $\mu$ and $\sigma$ bounded *between* zero and one. The two additional parameters, $\nu$ and $\tau$, are related to $p_0$ and $p_1$, respectively. $p_0$ is the probability that a case equals zero, $p_1$ is the probability that a case equals one, and $p_2$ (i.e., $1 - p_0 - p_1$) is the probability that the case comes from the beta distribution. In terms of these two additional parameters, $p_0 = \frac{\nu}{(1+\nu+\tau)}$ and $p_1 = \frac{\tau}{(1+\nu+\tau)}$. Rearranging these algebraically, $\nu$ is the odds that a case is 0 to being $0 < y < 1$, $\nu = p_0/p_2$, and $\tau$ is the odds that a case is a 1 to being $0 < y < 1$, $\tau = p_1/p_2$.

## Beta Regression Models

The goal for a researcher now is to predict these four parameters, $\mu, \sigma, \nu, \tau$, from any number of predictor variables of theoretical interest. All four parameters can be predicted by an identical set of predictors, none of the same predictors, or any mixture. Both $\mu$ and $\sigma$ have to be between zero and one (since the beta distribution is bounded between 0 and 1), so one can use the logistic link function to fit predicted values in this range. Both $\nu$ and $\tau$ have to be greater than zero (since they are odds), so one can use the log link function to fit predicted values in this range. Imagine a researcher has one predictor of interest, $x$. They could include this variable as a predictor of all four variables with the equations:

$$\log(\tfrac{\mu}{1-\mu}) = \beta_{10} + \beta_{11}X$$
$$\log(\tfrac{\sigma}{1-\sigma}) = \beta_{20} + \beta_{21}X$$
$$\log(\nu) = \beta_{30} + \beta_{31}X$$
$$\log(\tau) = \beta_{40} + \beta_{41}X$$

Or equivalently:

$$\mu = \tfrac{1}{1+e^{-(\beta_{10}+\beta_{11}X)}}$$
$$\sigma = \tfrac{1}{1+e^{-(\beta_{20}+\beta_{21}X)}}$$
$$\nu = e^{\beta_{30}+\beta_{31}X}$$
$$\tau = e^{\beta_{40}+\beta_{41}X}$$

where $e^x$ is the natural exponential function. A different inflated beta distribution can also be used when the dependent variable contains 0s but not 1s (e.g., $0 \leq y < 1$) or when it contains 1s but not 0s (e.g., $0 < y \leq 1$). Let $c$ be the value—0 or 1—that is included (i.e., inflated). The pdf is:

$$\text{beinf}_c(y; \mu, \sigma, \nu) = \begin{cases} p_c & \text{if } y = c \\ (1-p_c)f(y; \mu, \sigma) & \text{if } 0 < y < 1 \end{cases}$$

where $\nu = p_c/(1-p_c)$ and the remaining notation is the same as in the zero-and-one inflated beta distribution. The same link functions are used, but the fourth parameter, $\tau$, is not included, since the distribution is only inflated at $c$. Lastly, if no 0s or 1s are observed, the beta distribution alone can be used as the pdf. This results in a beta regression where the researcher is only predicting the location, $\mu$, and shape, $\sigma$. Since there is no inflation at 0 or 1, the latter two parameters, $\nu$ and $\tau$, are not included.

## Modeling Bounded Variables Beyond the Zero Through One Range

These beta regression models are commonly used when rates and proportions are dependent variables, given that these are naturally bounded $0 \leq y \leq 1$ (Buntaine, 2011;

Eskelson, Madsen, Hagar, & Temesgen, 2011; Gallardo, Bovea, Colomer, & Prades, 2012; Hubben et al., 2008; Peplonska et al., 2012). More generally, however, the beta distribution is a doubly bounded continuous distribution: Although it is on a continuous scale, values cannot be exceed the upper bound, $u$, or be less than the lower bound, $l$. In the case of the zero-and-one inflated beta distribution, $l = 0$ and $u = 1$.

Most scales used to measure prejudice and political attitudes are doubly bounded and continuous. Although researchers generally model variables on Likert scales and sliding scales (e.g., feeling thermometers) as being conditionally normally distributed (i.e., using ordinary least squares regression), these variables are not strictly normal. Observations from a normal distribution can take on any value on the real number line, while observations from a standard 7-point Likert scale can only take on values 1 through 7—and in studying controversial and polarizing issues like prejudice and politics, many participants score at the lower or upper bounds, causing floor or ceiling effects. In these cases, the assumptions of normality and homoskedasticity are likely to be violated. Many times researchers create, update, or seek out measures to so that the distributional form of the dependent variable takes on a normal shape—statistical assumptions are directing the phenomenon researchers choose to observe and the theoretical questions that they ask. Instead, one can explicitly take into account that the response is doubly bounded and heteroskedastic by using the beta regression models described above, using a simple linear rescaling of the data.

If one observes a dependent variable limited between two bounds, there is a straightforward way to rescale the variable to the $0 \leq y \leq 1$ range:

$$y'_i = (y_i - l)/(u - l)$$

where $y$ is the variable on the original scale, $y'$ is the rescaled variable, $u$ is the upper bound (i.e., the largest possible value on the scale), $l$ is the lower bound (i.e., the smallest possible value on the scale), and the $i$ subscript denotes an individual's score. On a standard 7-point Likert scale, $l = 1$ and $u = 7$. This rescaling allows a researcher to explicitly model conditional variance, floor effects, and ceiling effects using beta regression.

**Conditional variance.** Observing a dependent variable that is doubly bounded and continuous can produce heteroskedasticity. One of the assumptions of an OLS regression is homoskedasticity—that the variance of the errors are unrelated to any predictor, any linear combination of predictors, and the predicted values. A regression equation with one predictor variable $x$ is often written as $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $\epsilon_i$ is how far an osbserved value $y_i$ is from the predicted value (i.e., the residual). Let $\hat{y}_i$ be the predicted value for observation $i$, where $\hat{y}_i = \beta_0 + \beta_1 x_i$. We can simplify the equation to $y_i = \hat{y}_i + \epsilon_i$, where the residuals $\epsilon$ are normally distributed with a mean of 0 and some variance—that is, $\epsilon \sim N(0, \sigma_\epsilon^2)$. We can further simplify this to: $y_i | x_i \sim N(\hat{y}_i, \sigma_\epsilon^2)$, which means that each observation of the dependent variable we make, given the predictors we use, is normally distributed with a mean of that observation's predicted score and some variance. The assumption of homoskedasticity is found in that $\sigma_\epsilon^2$ does *not* have a subscript $i$. This means that there is one *common variance* of the residuals.

Imagine one runs a regression and observes $\hat{y}_i = 2.5 + 1.5 \times x_i$ and $\sigma_\epsilon^2 = 9$. Say that the first individual's score on $x$ is 0, $x_1 = 0$, and the second individual's score is $x_2 = 5$. The model then assumes that the first person comes from a normal distribution with a mean of 4 (i.e., $\hat{y}_1$) and a variance of 9 (i.e., $\sigma_\epsilon^2$), while the second person comes from a normal distribution with a mean of 10 and that same variance of 9. The violation of this assumption can lead to inflated Type I errors or diminished statistical power, depending on the type and source of the heteroskedasticity (Hayes & Cai, 2007; Long & Ervin, 2000; Rosopa, Schaffer, & Schroeder, 2013).

While common remedies for heteroskedasticity (e.g., robust standard errors, transforming the dependent variable, weighted least squares) treat it as a nuisance to be corrected for when calculating $p$-values (Hayes & Cai, 2007; Long & Ervin, 2000; Rosopa et al., 2013), it could be that explicitly modeling this conditional variance is an interesting phenomenon per se. Heteroskedasticity can arise when the error variance is predicted by one of the observed predictor variables, and since beta regression models predict a shape parameter, $\sigma$, a researcher can explicitly model this relationship. This allows researchers to ask questions about conditional variance: "Does the variance in $y$ increase as $x$ also increases?" Since the focus of OLS regression is the *location* parameter (e.g., conditional mean), researchers might be ignoring interesting, potentially theoretically-relevant parts of their data. After discussing floor and ceiling effects, I will argue that normative influences are one of these aspects.

**Floor and ceiling effects.** Floor and ceiling effects occur when there are many observations of the dependent variable at the lower and upper bound of the scale, respectively (Everitt & Skrondal, 2010). Variables displaying these effects are likely to violate numerous assumptions of ordinary least squares regression, such as conditional normality and homoskedasticity of variance, and can be modeled as censored or truncated. Censoring occurs when all scores beyond some threshold are recorded at that threshold. For example, if a researcher is measuring reaction times, censoring occurs if they decide that any times above 5 seconds are scored as 5. If many people took longer than 5 seconds, this censoring can cause a ceiling effect. Truncating occurs when people who score beyond some treshold are excluded from the data. If the researcher simply does not record trials that take longer than 5 seconds, then the data are truncated. A common regression technique for censored or truncated variables are Tobit models (McBee, 2010; Smithson & Merkle, 2013). However, standard Tobit models assume that the underlying latent construct is normally distributed.

In the realm of prejudice and politics, I argue that it is *not* useful to think of prejudice as coming from a latent normal distribution. I find it unlikely that if, instead of a Likert scales going from 1 to 7, it extended from -7 to +7, one would observe a normal distribution for the Attitude Towards Blacks scale, for example (Brigham (1993)). I find it likely that participants would opt for the new floor—the negative 7. This can be thought of as a decision-making process. Using prejudice as an example, people decide whether or not they are going to admit to feeling any prejudice. If they decide not to, they simply circle or click all 1s (i.e., the floor). If they decide to admit prejudice, they respond to the items. As another example: When offering one's attitudes toward a polarizing political figure, a participant decides whether or not to respond entirely in the negative (i.e., at the floor), entirely in the positive (i.e., at the ceiling), or somewhere in between. This could result in a bimodal distribution, with both ceiling and floor effects.

This decision making process can be modeled with the inflated beta regression models described above. Inflations at zero, one, or zero-and-one allow researchers to measuring ceiling effects, floor effects, and both simultaneously. While methodologists in beta regression refer to these models as "inflated" models, a more general label is "two-part" models (Coxe et al., 2013). When one assumes a decision-making process behind the two parts, these models are often referred to as "hurdle" models in the econometrics literature (Cameron & Trivedi, 2005; Wooldridge, 2010).

Cragg (1971) demonstrated that hurdle models are a generalization of the Type I Tobit model (but note that, unlike the standard Tobit model, the latent dependent construct is assumed to be conditionally *beta* distributed, which can take on many distributional forms; see Figure 1). In economics, hurdle models are often used when dependent variables are counts with excess 0s (see Carlevaro, Croissant, & Hoareau, 2017 for a review of applications). The log likelihood function for these models can be separated in two parts: First, a logistic (or probit) regression modeling the probability of observing $y = 0$ versus $y > 0$; second, a truncated negative binomial model using just the cases where $y > 0$.

The inflated beta regressions can be seen as hurdle models, as the log likelihood for the zero-and-one inflated beta regression can also be separated in two parts: First, a multinomial logistic regression modeling the probabilities of observing a $y = 0$, $y = 1$, or $0 < y < 1$; second, a beta regression model using just the cases where $0 < y < 1$. Similarly, the inflated beta regression at *either* zero *or* one has a log likelihood function that can be separated in two parts: First, a logistic regression modeling the probability of observing $y = c$ or $0 < y < 1$; second, a beta regression modeling just the cases where $0 < y < 1$ (R. Rigby et al., 2017).

Since their likelihoods can be separated, this implies a critical assumption: That the two processes are independent, given the observed predictors. In other words, the error terms of each submodel are assumed to be independent of one another (where submodels refer to models predicting each parameter, $\mu$, $\sigma$, $\nu$, or $\tau$, separately). Dependencies between them can be modeled (referred to as "selection" models; Carlevaro et al., 2017; Wooldridge, 2010), but I will not discuss these models here, as they are not extended to beta regression.

## Norms Produce Invariance, Ceiling and Floor Effects

Louis, Mavor, and Terry (2003) argue that norms produce invariance. In this section, I demonstrate the relationship between normative strength and invariance of attitudes empirically. I replicated the methodology used by Crandall, Eshleman, and O'brien (2002). In their Study 1, the authors presented participants with 105 social groups (e.g., Black people, librarians, deaf people, ugly people, Nazis, drug dealers). Participants were assigned to one of two groups. Half of the participants indicated how negatively they felt toward the group (i.e., prejudice), while the other half indicated how socially acceptable they perceived it was to feel and express negativity toward that group (i.e., acceptability). Crandall and colleagues averaged the prejudice and acceptability scores at the group level, resulting in an $N = 105$ and two variables for each group: prejudice and acceptability. They found that expressed prejudice and acceptability of prejudice correlated at $r = .96$; people report prejudices that are socially acceptable (see Cary and Page-Gould (2014) for a replication).

I recruited 405 people from Amazon's Mechanical Turk website and randomly assigned them to either a *prejudice* or *acceptability* condition. All participants answered questions about 30 groups, randomly sampled from a pool of 120 groups. In the prejudice condition, participants were asked "How do you feel toward each of these groups?" and to indicate how they feel on a scale from "very negatively (0) to very positively (100)." I reverse-scored these items so that higher scores indicated more prejudice. In the acceptability condition, participants were asked "How OK is it in society for people to feel and say negative things about each of these groups?" and to indicate their perception on a scale from "totally unacceptable (0) to completely acceptable (100) for people to feel and say negative things about these groups."

To demonstrate the relationship between norms and invariance, I calculated the median social acceptability score and the variance of the self-reported prejudice scores for each of the 120 groups. There was no linear relationship between social acceptability and variance of reported prejudice, $b = 0.31, SE = 0.84, t(118) = 0.37, p = .714$; however, there was a quadratic relationship regressing variance of reported prejudice on median social acceptability, $\Delta R^2 = .28, b = -1250.07, SE = 187.44, t(117) = -6.67, p < .001$ (Figure 2). The closer a prejudice was perceived to be by some participants as totally unacceptable (0) or compeltely acceptable (100), the lower was the variance of self-reported prejudice by other participants.

I also investigated normative influences on floor effects. Crandall (1994) proposed a measure for anti-fat prejudice. In Study 5 of this paper, Crandall makes the argument that anti-fat attitudes are more socially acceptable (and less subject to social desirability biases) than anti-Black prejudice. He does this by demonstrating the decision-making process described above in the hurdle model: How often do people *always* select the least prejudiced option? That is, how often do people respond *always* at the lower bound of the distribution? He calls the proportion of people at the floor of a scale the "politically correct (PC) index," as this shows how often people respond in the most socially acceptable (i.e., politically correct) way possible. He finds that anti-fat prejudice is less subject to these floor effects: 10% of participants always responded at the lower bound for anti-Black prejudice; this was only 3% for Crandall's anti-fat scale.

To show that norms produce floor effects, I calculated Crandall's PC index for each of the 120 groups by counting how many people responded with a 0 on the prejudice thermometer (i.e., "totally positive" feelings). Using a negative binomial regression (Venables & Ripley, 2002) for these count data, I regressed the PC index on median social acceptability of the prejudice. Perceived social acceptability of the prejudice by one group of participants negatively predicted the PC index, calculated using a separate group of participants, $b = -0.03, SE = 0.002, Z = -10.97, p < .001$ (Figure 3). The more socially acceptable a prejudice, the less people opt for the "politically correct" floor of the scale.

This evidence shows that norms do positively predict invariance; predicting the scale parameter, $\sigma$, in beta regression could be used to investigate how much people adhere to norms regarding the expression of certain attitudes. These data also show that normative influence can lead to floor effects (or ceiling effects); predicting the shape parameters, $\nu$ and $\tau$, in beta regression could also be used to examine how much people adhere to norms. I now turn to demonstrating how these beta regression models can be used in R (R Core Team, 2017) via the `gamlss` package (R. A. Rigby & Stasinopoulos, 2005).

## Implementation in GAMLSS

The `gamlss` name stands for generalized additive models for location (i.e., $\mu$), scale (i.e., $\sigma$), and shape (i.e., $\nu$ and $\tau$). This package gives researchers the ability to use a wide variety of models, including beta regression models. A number of algorithms to estimate the coefficients can be used, described in detail by M. D. Stasinopoulos, Rigby, Heller, Voudouris, and De Bastiani (2017). To first demonstrate that the package works according to the parameterization discussed in the Statistical Background section, I will generate a zero-and-one inflated beta distribution and estimate the parameters with an intercepts-only regression using the `gamlss` function. Then, I will demonstrate how to model prejudice as a function of ideology using real data.

### Intercepts-Only Regression

We can set a number of parameters for this zero-one-inflated regression using the following code:

```r
n <- 5000
mu <- 0.40
sigma <- 0.60
p0 <- 0.13
p1 <- 0.17
p2 <- 1 - p0 - p1
a <- mu * (1 - sigma ^ 2) / (sigma ^ 2)
b <- a * (1 - mu) / mu
```

The sample size `n` $= 5000$, the mean `mu` $= 0.40$, the shape parameter `sigma` $= 0.60$, the probability of being 0 is `p0` $= 0.13$, the probability of being 1 is `p1` $= 0.17$, and the probability of being between 0 and 1 is `p2` $= 1 - p_0 - p_1 = .70$. Using the equations in the Statistical Background section, we can convert `mu` and `sigma` back to the original shape parameters `a` ($\alpha$) and `b` ($\beta$). The dependent variable `y` can now be generated using the following code:

```r
set.seed(1839)
y <- vector("numeric", n)
for (i in 1:n) {
  rand <- runif(1)
  if (rand <= p0) {
    y[i] <- 0
  } else if ((p0 < rand) & (rand <= p0 + p1)) {
    y[i] <- 1
  } else {
    y[i] <- rbeta(1, a, b)
  }
}
```

I generate a random number, `rand`, between 0 to 1. If `rand` $\leq$ `p0`, then the case is 0; if `rand` $>$ `p0` and `rand` $\leq$ `p0 + p1`, then the case is 1; otherwise, the case comes from a beta distribution with shape parameters of `a` and `b`. The `gamlss` function has four arguments for formulas—one for each of the parameters. It also has an argument, `family`, that determines what distribution will be used in fitting the model. For a zero-and-one inflated beta regression, the family is `BEINF()`. One can fit the intercepts-only model using the following code:

```
fit <- gamlss(
  formula = x ~ 1, # formula for mu
  formula.sigma = ~ 1, # formula for sigma
  formula.nu = ~ 1, # formula for nu
  formula.tau = ~ 1, # formula for tau
  family = BEINF() # distribution for model
)
```

We can use the inverse link functions to transform the coefficients back into the original scale. Then we can compare it to the parameters we set above. The estimated parameters are estimated using the code, according to the equations and link functions described in the Statistical Background section:

```
# define the inverse of the logit link function
# the exp function is the inverse of the log link
inv_logit <- function(x) exp(x) / (1 + exp(x))
fit_mu <- inv_logit(fit$mu.coefficients)
fit_sigma <- inv_logit(fit$sigma.coefficients)
fit_nu <- exp(fit$nu.coefficients)
fit_tau <- exp(fit$tau.coefficients)
fit_p0 <- fit_nu / (1 + fit_nu + fit_tau)
```

```
fit_p1 <- fit_tau / (1 + fit_nu + fit_tau)
```

The estimates for $\mu$, $\sigma$, $p_0$, and $p_1$ are .41, .60, .13, and .17, respectively. These are close estimates to the population parameters set above. I now turn to a model using real data.

## Modeling Prejudice from Political Ideology

## References

Brigham, J. C. (1993). College students? Racial attitudes. *Journal of Applied Social Psychology*, *23*(23), 1933–1967.

Buntaine, M. T. (2011). Does the Asian Development Bank respond to past environmental performance when allocating environmentally risky financing? *World Development*, *39*(3), 336–350.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York, NY: Cambridge university press.

Carlevaro, F., Croissant, Y., & Hoareau, S. (2017). *Multiple hurdle Tobit models in R: The mhurdle package*. Retrieved from https://cran.r-project.org/web/packages/mhurdle/vignettes/mhurdle.pdf

Cary, L. A., & Page-Gould, E. (2014). The prevalence and impact of sexism, racism and homophobia in online gaming environments. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Austin, TX.

Coxe, S., West, S. G., & Aiken, L. S. (2013). Generalized linear models. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods, volume 2* (pp. 26–51). New York, NY: Oxford University Press.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829–844.

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology*, *66*(5), 882.

Crandall, C. S., Eshleman, A., & O'brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359.

Eskelson, B. N., Madsen, L., Hagar, J. C., & Temesgen, H. (2011). Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, *57*(3), 212–221.

Everitt, B., & Skrondal, A. (2010). *The cambridge dictionary of statistics*. New York, NY: Cambridge University Press.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815.

Gallardo, A., Bovea, M. D., Colomer, F. J., & Prades, M. (2012). Analysis of collection systems for sorted household waste in Spain. *Waste Management*, *32*(9), 1623–1633.

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, *39*(4), 709–722.

Hubben, G. A. A., Bishai, D., Pechlivanoglou, P., Cattelan, A. M., Grisetti, R., Facchin, C., … Tramarin, A. (2008). The societal burden of HIV/AIDS in Northern Italy: An analysis of costs and quality of life. *AIDS Care*, *20*(4), 449–455.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*(3), 217–224.

Louis, W. R., Mavor, K. I., & Terry, D. J. (2003). Reflections on the statistical analysis of personality and norms in war, peace, and prejudice: Are deviant minorities the problem? *Analyses of Social Issues and Public Policy*, *3*(1), 189–198.

McBee, M. (2010). Modeling outcomes with floor or ceiling effects: An introduction to the tobit model. *Gifted Child Quarterly*, *54*(4), 314–320.

Peplonska, B., Bukowska, A., Sobala, W., Reszka, E., Gromadzińska, J., Wasowicz, W., … Ursin, G. (2012). Rotating night shift work and mammographic density. *Cancer, Epidemiology, Biomarkers, & Prevention*, *21*(7), 1028–1037.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, *54.3*, 507–554.

Rigby, R., Stasinopoulos, M., Heller, G., & De Bastiani, F. (2017). *Distributions for modelling location, scale, and shape: Using GAMLSS in R. Unpublished draft.* Retrieved from gamlss.com/books-articles

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods*, *18*(3), 335.

Smithson, M., & Merkle, E. C. (2013). *Generalized linear models for categorical and continuous limited dependent variables.* Boca Raton, FL: CRC Press.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible regression and smoothing: Using GAMLSS in R.* Boca Raton, FL: CRC Press.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s.* New York, NY: Springer.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT press.

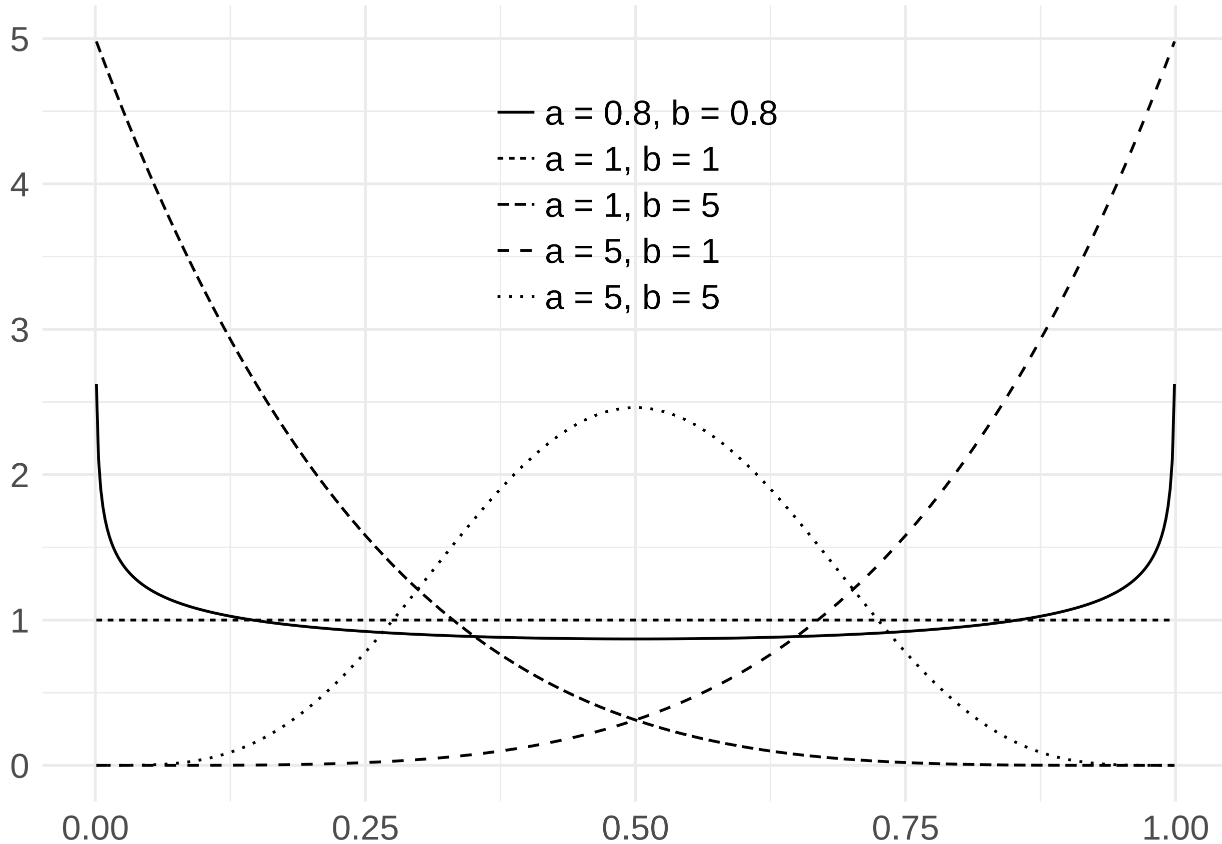


*Figure 1*. Beta probability density functions with various combinations of shape parameters.

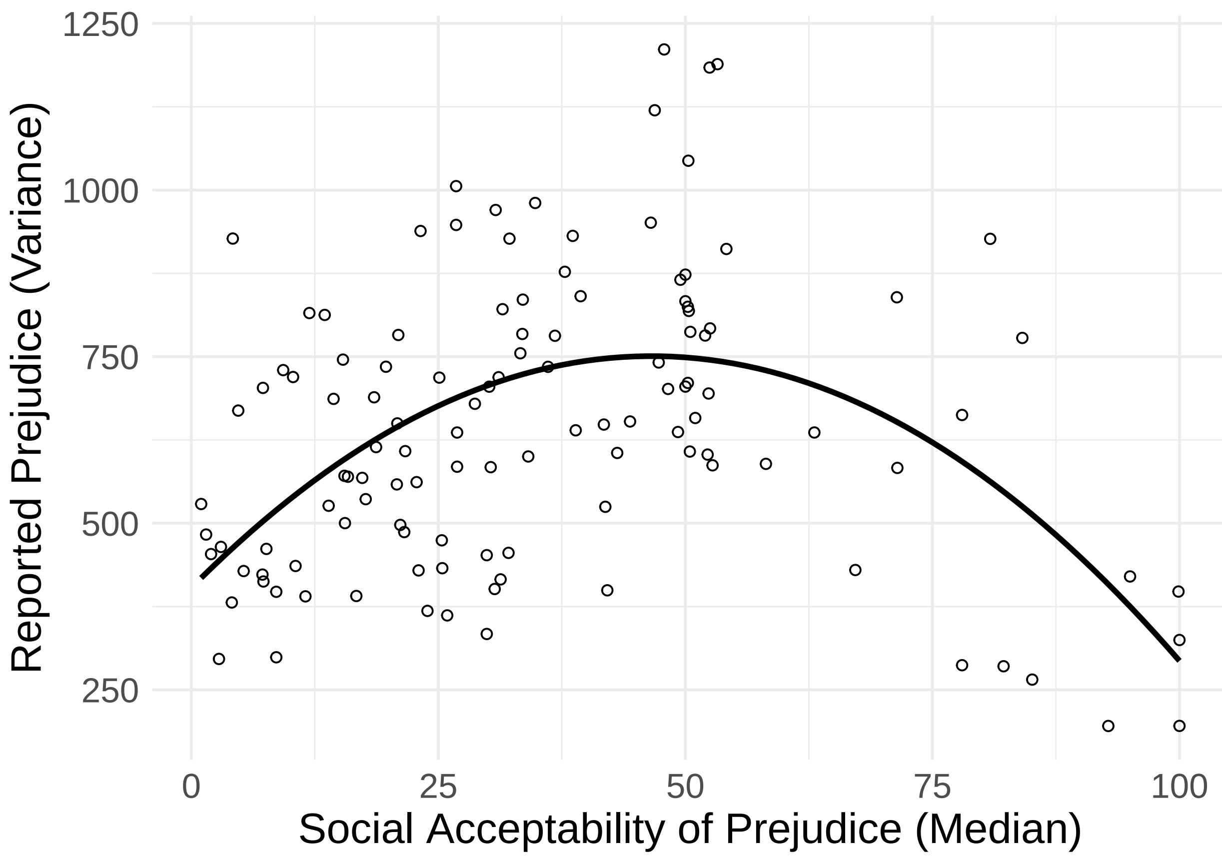


*Figure 2.* The stronger the normative influence on prejudice, the smaller the variance of reported prejudice.

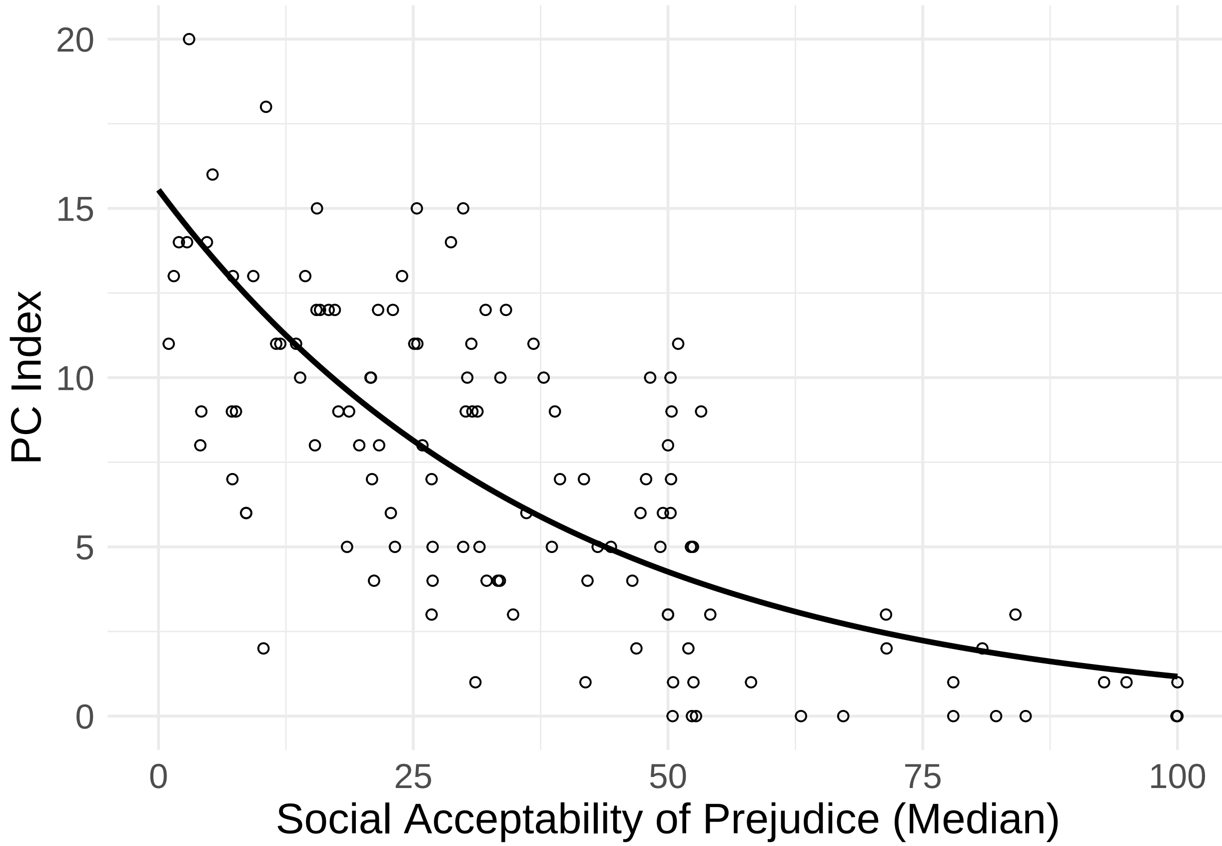


*Figure 3*. The more socially acceptable the prejudice, the less people opt for the politically correct response.