Modeling Postsecondary Retention Rates in the United States

Emily Mo

12/4/2019

Why Model Retention Rates?

Retention rates are one of the most commonly used measures of an educational institution's success in educating its students and preparing them for successful careers, and a high retention rate is a badge of honor that institutions tend to strive for. However, an institution's retention rate is a product of not only the institution's accessibility to students, but also external factors that impact students—thus, retention rate as a standalone metric may not be very informative about an institution's educational efficacy. I used data from the Integrated Postsecondary Education System Data System (IPEDS) in order to construct multilevel Binomial models of retention rates based on several institutional characteristics and offerings. Because these models only consider predictors on the institutional level and includes no predictors about institutions' student bodies, they are not intended to—and in fact are not able to—paint a complete picture of the factors that impact retention rates. Instead, the models serve to pinpoint how several helpful educational features within institutions may relate to retention rates across the US, but also to demonstrate that it can be misleading to portray institutional success metrics as a direct product of an institution's efforts, due to the inevitable extra-curricular influences that can be confounded with such educational features.

The data used in these models consisted of institutions in IPEDS records that had retention rate and public/private status information (public, private for-profit, private not-for-profit) available from 2017. All variables used were collected in 2017, which was the most recent year for which all variables of interest were available on the IPEDS data portal. I chose to use distance learning options, recreational options, and life experience credit as predictors because they are offerings that are presumably implemented with the purpose of making college a more welcoming place: distance learning makes educational attainment more accessible to those who may be working or may have difficulty physically attending a class, credit for life experiences acknowledges that accreditable learning opportunities do not all occur within the classroom, and recreational options allow students to take a break from traditional academics while still learning new skills. [1] Additionally, a lower student-faculty ratio may make it easier for students to get individualized attention from faculty, which can make the educational experience feel more personalized.

Models

Using this cleaned and narrowed data set, I fit two Binomial regression models with increasingly complex hierarchical structure. The student-faculty ratio variable was centered at its mean, which was 15.268, in order to improve interpretability. Because a Laplacian fitting algorithm produced nearly unidentifiable parameter estimates [2], I used a Bayesian fitting approach, which allowed for more nuanced evaluation of the parameters' convergence, as well as permitted probabilistic statements about the parameters' true values given the data.

---

[1] IPEDS defines "credit for life experiences" as "Credit earned by students for what they have learned through independent study, noncredit adult courses, work experience, portfolio demonstration, previous licensure or certification, or completion of other learning opportunities (military, government, or professional).

[2] Results from the Laplacian approach are detailed in Appendix B.)

Model A: Multilevel model with varying intercepts per state and public/private status

Using a Markov Chain Monte Carlo algorithm, I fit a multilevel binomial model with varying intercepts for state and for public/private status. A multilevel structure was appropriate because it would account for the expected correlation of retention rates between colleges within the same state or within the same public/private status. A binomial regression was appropriate because each institution's retention rate reflects a group of individuals, each of whom were retained or not retained—thus, each institution's retention rate can be viewed as a binomial random variable. The institutional characteristics previously mentioned were used as the model's non-varying predictors. The prior for this model was only specified with respect to the residual variance and the covariance structure of the varying intercepts. For all of those parameters, I supplied a weakly informative Inverse Chi-Square prior with a scale parameter of 1 and 2 degrees of freedom, which can be expressed as an Inverse Gamma distribution with a shape parameter of 1 and a scale parameter of 1. [3] The model was fit with 40,000 iterations, with a burn-in period of 6,000 and a thinning factor of 30. These parameters yielded a roughly stationary trace plot and a roughly symmetric distribution plot for each of the fixed coefficients in the model, which suggested that the parameter estimates were trustworthy. The coefficient estimates and between-group variance estimates are as follows [4]:

```r
set.seed(60)
prior_inf <- list(R = list(V = 1, nu = 2), G = list(G1 = list(V = 1, nu = 2), G2 = list(V = 1,
    nu = 2)))
ipeds_mcmc2 <- MCMCglmm(cbind(retcount, notret) ~ dist17 + lifeex17 + recr_offer17 +
    sfratio_c, ~us(1):state + us(1):pubpriv, data = edu_17, family = "multinomial2",
    prior = prior_inf, nitt = 40000, thin = 30, burnin = 10000, pr = T, verbose = F)
summary(ipeds_mcmc2)
```

```
##
##  Iterations = 10001:39971
##  Thinning interval  = 30
##  Sample size  = 1000
##
##  DIC: 11261844
##
##  G-structure:  ~us(1):state
##
##                                post.mean l-95% CI u-95% CI eff.samp
## (Intercept):(Intercept).state    0.08524  0.05062   0.1213     1000
##
##                  ~us(1):pubpriv
##
##                                post.mean l-95% CI u-95% CI eff.samp
## (Intercept):(Intercept).pubpriv    1.023  0.09918     2.54     1000
##
##  R-structure:  ~units
##
##        post.mean l-95% CI u-95% CI eff.samp
## units     0.8829    0.843   0.9202      815
##
##  Location effects: cbind(retcount, notret) ~ dist17 + lifeex17 + recr_offer17 + sfratio_c
##
##                 post.mean  l-95% CI  u-95% CI eff.samp  pMCMC
```

---

[3] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Analysis, 1(3), 515–534. doi: 10.1214/06-ba117a

[4] full model outputs can be found in Appendix B

2

```
## (Intercept)   1.254295 -0.050227  2.210331   1000.0  0.046 *
## dist17       -0.005937 -0.068010  0.058276    949.5  0.856
## lifeex17     -0.455338 -0.522917 -0.393742   1000.0 <0.001 ***
## recr_offer17 -0.192526 -0.281971 -0.119240   1000.0 <0.001 ***
## sfratio_c    -0.016430 -0.020672 -0.011791   1000.0 <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the model:

- A baseline institution that has no distance learning options, offers no credit for life experiences, offers no recreational or avocational opportunities, and has a typical student-faculty ratio of about 15 is expected to have a log odds retention rate ratio [5] of 1.254, meaning that the estimated retention rate of such an institution would be $logit^{-1}(1.254) = 0.778$.
- An institution that offers distance learning options is expected to have a log odds retention rate ratio that is 0.006 less than that of an institution that does not offer distance learning options, given that credit for life experiences, recreational offerings, and student faculty ratio are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.006 / 4 = -0.001.
- An institution that offers credit for life experiences is expected to have a log odds retention rate ratio that is 0.455 less than that of an institution that does not offer credit for life experiences, given that distance learning offerings, recreational offerings, and student faculty ratio are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.455 / 4 = -0.114.
- An institution that offers recreational or avocational opportunities is expected to have a log odds retention rate ratio that is 0.193 less than that of an institution that does not offer recreational or avocational opportunities, given that distance learning offerings, credit for life experiences, and student faculty ratio are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.193 / 4 = -0.048.
- A unit increase in student-faculty ratio is generally associated with a decrease of 0.016 in the log odds ratio of retention rate, given that distance learning offerings, credit for life experiences, and recreational offerings are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.016 / 4 = -0.004.

The estimate for the variance between states' baseline retention rates, 0.085, was small compared to the estimate for the variance between public/private statuses' baseline retention rates, 1.023. This discrepancy suggested that the state grouping may not provide much additional information about the baseline retention rate after the overall baseline retention rate is already considered. However, there seems to be much more variation in retention rates between public, private not-for-profit, and private for-profit institutions.

The estimated varying intercepts for public, private not-for-profit, and private for-profit were as follows:

```
##                               Mean        SD
## (Intercept).pubpriv.priv_fp  -0.07151329 0.5570228
## (Intercept).pubpriv.priv_nfp  0.14409476 0.5565214
## (Intercept).pubpriv.pub      -0.05844952 0.5575042
```

According to the model, a baseline institution that has no distance learning options, offers no credit for life experiences, offers no recreational or avocational opportunities, and has a typical student-faculty ratio is expected to have a retention rate of:

- $logit^{-1}(1.254 - 0.072) = 0.765$ if it is private for-profit
- $logit^{-1}(1.254 + 0.144) = 0.802$ if it is private not-for-profit
- $logit^{-1}(1.254 - 0.058) = 0.768$ if it is public

---

[5]The log odds retention rate ratio is the natural logarithm of the ratio between retention rate and its complement.

The variation in estimated baseline retention rate between the different public/private statuses remind us that the type of institution alone may be associated with factors that influence a student's likelihood to stay in school, regardless of efforts made by the institution to prevent attrition. However, we must note that the standard deviation of these estimates all suggest 95% credible intervals that contain 0, so there is a great degree of uncertainty of these varying intercepts.

Model B: Multilevel model with varying intercepts per state and public/private status, and varying slope for recreational experiences per public/private status

I next fit a model similar to Model A, but with two major departures: 1.) it eliminated the distance learning predictor since the predictor did not seem very informative in Model A due to collinearity, and 2.) it allowed the slope for recreational offerings to vary across each public/private status. The addition of this random slope reflects the possibility that recreational offerings could have a different relationship with retention rate for public institutions than for private institutions, due to differences in the students that may attend each type of institution.

The coefficient estimates and between-group variance estimates are as follows [6]:

```
set.seed(60)
prior_inf2 <- list(R = list(V = 1, nu = 2), G = list(G1 = list(V = 1, nu = 2), G1 = list(V = diag(2),
    nu = 2)))
ipeds_mcmc3 <- MCMCglmm(cbind(retcount, notret) ~ lifeex17 + recr_offer17 + sfratio_c,
    ~us(1):state + us(1 + recr_offer17):pubpriv, data = edu_17, family = "multinomial2",
    prior = prior_inf2, nitt = 40000, thin = 30, burnin = 10000, pr = T, verbose = F)
summary(ipeds_mcmc3)
```

```
##
##  Iterations = 10001:39971
##  Thinning interval  = 30
##  Sample size  = 1000
##
##  DIC: 11261850
##
##  G-structure:  ~us(1):state
##
##                                post.mean l-95% CI u-95% CI eff.samp
## (Intercept):(Intercept).state   0.08636  0.05186   0.1248     1025
##
##                  ~us(1 + recr_offer17):pubpriv
##
##                                     post.mean l-95% CI u-95% CI eff.samp
## (Intercept):(Intercept).pubpriv       1.89995   0.1387    5.983   1000.0
## recr_offer17:(Intercept).pubpriv      0.04487  -2.4539    3.656    672.2
## (Intercept):recr_offer17.pubpriv      0.04487  -2.4539    3.656    672.2
## recr_offer17:recr_offer17.pubpriv     1.86659   0.1269    6.558   1000.0
##
##  R-structure:  ~units
##
##        post.mean l-95% CI u-95% CI eff.samp
## units    0.8806   0.8423   0.9211    920.6
##
##  Location effects: cbind(retcount, notret) ~ lifeex17 + recr_offer17 + sfratio_c
##
```

---

[6]full model outputs can be found in Appendix B

```
##                 post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)    1.24885 -0.17351  2.76870      1188  0.086 .
## lifeex17      -0.45127 -0.50791 -0.39512      1000 <0.001 ***
## recr_offer17  -0.10823 -1.71475  1.21706      1000  0.888
## sfratio_c     -0.01625 -0.02038 -0.01152      1011 <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the model:

- A baseline institution that offers no credit for life experiences, offers no recreational or avocational opportunities, and has a typical student-faculty ratio of about 15 is expected to have a log odds retention rate ratio of 1.249, meaning that the estimated retention rate of such an institution would be $logit^{-1}$ $(1.249) = 0.777$.
- An institution that offers credit for life experiences is expected to have a log odds retention rate ratio that is 0.451 less than that of an institution that does not offer credit for life experiences, given that recreational offerings and student faculty ratio are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.451 / 4 = -0.11275.
- An institution that offers recreational or avocational opportunities is expected to have a log odds retention rate ratio that is 0.108 less than that of an institution that does not offer recreational or avocational opportunities, given that credit for life experiences and student faculty ratio are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.108 / 4 = -0.027.
- A unit increase in student-faculty ratio is generally associated with a decrease of 0.016 in the log odds ratio of retention rate, given that distance learning offerings, credit for life experiences, and recreational offerings are held constant. While the corresponding change in retention rate itself can vary depending on the original retention rate, an estimated upper bound on the corresponding change in retention rate is -0.016 / 4 = -0.004.

Despite the change in model structure, the predictors' coefficients remain fairly similar—in both direction and magnitude—to their corresponding coefficients in Model A. However, the p-value of recreational offerings, which was close to 0 in Model A, is now very high. Allowing the coefficient of recreational offerings to vary by public/private status may have introduced more uncertainty into the overall coefficient estimate.

The estimate for the variance between states' baseline retention rates, 0.086, was again small compared to the estimate for the variance between public/private statuses' baseline retention rates, 1.9. This again suggested that the state grouping may not provide much additional information about the baseline retention rate after the overall baseline retention rate is already considered. However, there seems to be relatively much more variation in retention rates between public, private not-for-profit, and private for-profit institutions. It is important to note the uncertainty about the variance between public/private statuses' baseline retention rates: the confidence interval about the mean estimate is very wide on the upper limit because the spread of the distribution is so wide that the estimates are truncated at 0 on the lower limit. This is likely because there are only three public/private status groups from which this parameter is estimated. A similar lack of precision is observed with the estimated variance between public/private statuses' slopes for recreational offerings.

The estimated varying intercepts for public, private not-for-profit, and private for-profit were as follows:

```
##                                Mean        SD
## (Intercept).pubpriv.priv_fp  -0.07043963 0.7460763
## (Intercept).pubpriv.priv_nfp  0.12421616 0.7450868
## (Intercept).pubpriv.pub      -0.02949632 0.7461057
```

According to the model, a baseline institution that has no distance learning options, offers no credit for life experiences, offers no recreational or avocational opportunities, and has a typical student-faculty ratio is

expected to have a retention rate of:

- $logit^{-1}$ (1.249 - 0.07) = 0.765 if it is private for-profit
- $logit^{-1}$ (1.249 + 0.124) = 0.798 if it is private not-for-profit
- $logit^{-1}$ (1.249 - 0.029) = 0.772 if it is public

This echoes the pattern seen in Model A's intercepts across public/private statuses: private not-for-profit institutions are expected to have higher baseline retention rates than private for-profit and public institutions. However, similarly to in Model A, all of these estimates' 95% credible intervals contain 0, so we can not be very sure of their precision.

The estimated varying recreational offering slopes for public, private not-for-profit, and private for-profit institutions were as follows:

```
##                               Mean        SD
## recr_offer17.pubpriv.priv_fp  -0.03240173 0.8593610
## recr_offer17.pubpriv.priv_nfp  0.25514642 0.8601168
## recr_offer17.pubpriv.pub      -0.14787813 0.8556300
```

Based on this model's varying slopes, an institution that offers recreational or avocational opportunities is expected to have a log odds retention rate ratio that is:

- -0.108 - 0.032 = 0.141 lower than an institution that does not offer recreational or avocational opportunities, given that credit for life experiences and student faculty ratio are held constant, for for-profit institutions. An estimated upper bound on the corresponding change in retention rate is -0.141 / 4 = -0.035.
- -0.108 + 0.255 = 0.147 higher than an institution that does not offer recreational or avocational opportunities, given that credit for life experiences and student faculty ratio are held constant, for private not-for-profit institutions. An estimated upper bound on the corresponding change in retention rate is 0.147 / 4 = 0.037.
- -0.108 - 0.148 = 0.256 lower than an institution that does not offer recreational or avocational opportunities, given that credit for life experiences and student faculty ratio are held constant, for public institutions. An estimated upper bound on the corresponding change in retention rate is -0.256 / 4 = -0.064.

Interestingly, the varying slopes suggest that the presence of recreational opportunities is associated with higher retention rates in private not-for-profit institutions, yet is associated with lower retention rates in private for-profit and public institutions. Again, this cannot necessarily be interpreted as a causal effect on retention rates—the competition between institutional factors and external student-specific factors may vary between each of these types of institutions. Additionally, the standard deviation of these estimates all suggest 95% credible intervals that contain 0, so there remains much uncertainty about these varying intercepts. With these results in mind, private and public institutions may find it useful to investigate, on the student level, how the quality and variety of recreational programs contribute to the educational experience, or whether participation in these programs are relevant or feasible to students in the first place. For instance, students who work part time to sustain their families, who may be a more prominent subpopulation in public institutions than private not-for-profit institutions, may not have as much time to take advantage of recreational options and may also be more at risk for dropping out.

Model Checking

Model A

In order to check the stability of the MCMC coefficient estimates, I first inspect the trace plots (Figure 1) for stationarity and compare the posterior sample densities (Figure 2) to a Normal distribution. The QQ plots for the posterior sample densities (Figure 3) suggest that aside from the intercept, the coefficient estimates across the fitting iterations are roughly normally distributed. The QQ plot for the intercept, however, suggests that
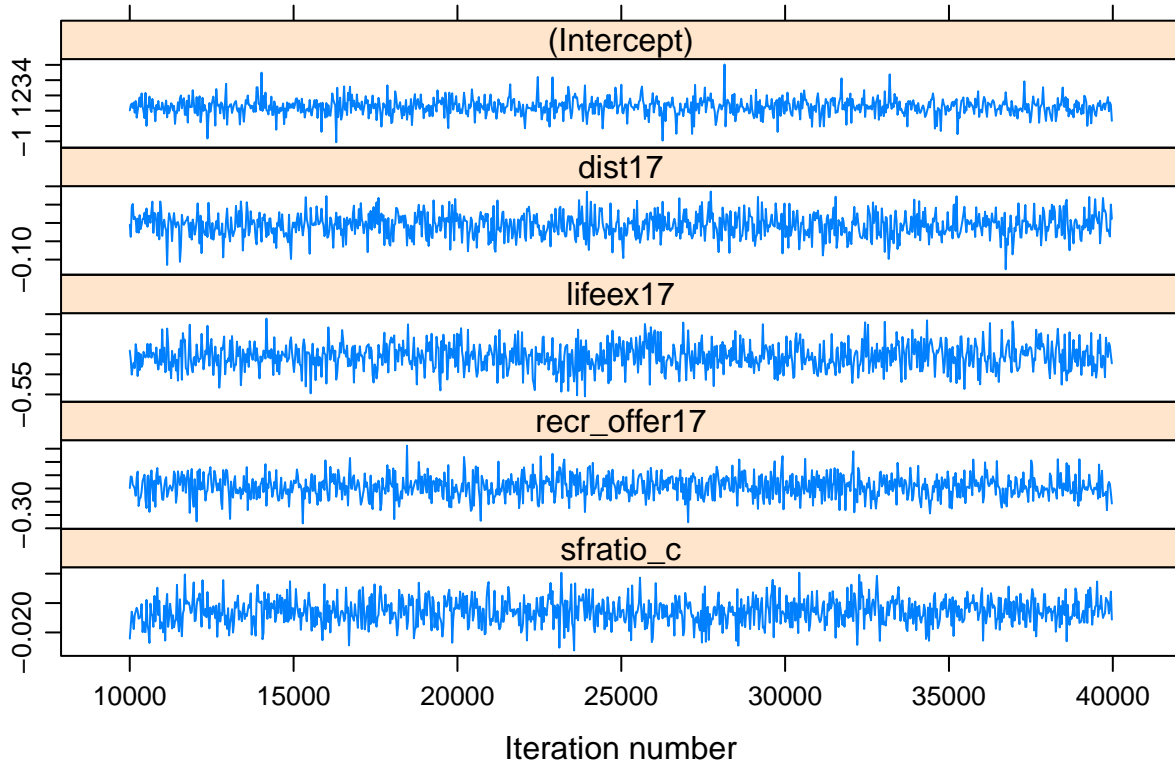
Figure 1: Trace Plots for Model A Non-Varying Coefficients

the distribution has heavier tails than a Normal distribution despite its symmetry, which places more weight on the extreme values of the posterior distribution and makes the posterior mean relatively less probable. The trace plots for all of these coefficient estimates appear stationary aside from occasional outliers, which suggests that the estimated posterior distributions are stable.

Next, I compare Model A's parameter estimates to the unstable estimates from a Laplacian model [7] fitted with the lmer package, which has the same model structure but is nearly unidentifiable.

Laplacian model coefficients:

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unider
##  - Rescale variables?;Model is nearly unidentifiable: large eigenvalue ratio
##  - Rescale variables?

##                 Estimate   Std. Error    z value     Pr(>|z|)
## (Intercept)   1.19100770 0.2504940985   4.754634 1.988065e-06
## dist17        0.10841169 0.0018529786  58.506712 0.000000e+00
## lifeex17     -0.57900105 0.0015413063 -375.656044 0.000000e+00
## recr_offer17 -0.25129337 0.0017213613 -145.985258 0.000000e+00
## sfratio_c    -0.02672116 0.0001463214 -182.619635 0.000000e+00
```

MCMC model coefficients:

```
##                 post.mean     l-95% CI     u-95% CI eff.samp pMCMC
## (Intercept)   1.254295331 -0.05022732   2.21033099 1000.0000 0.046
## dist17       -0.005937468 -0.06801032   0.05827634  949.4929 0.856
## lifeex17     -0.455338310 -0.52291731  -0.39374230 1000.0000 0.001
## recr_offer17 -0.192526000 -0.28197137  -0.11924045 1000.0000 0.001
```

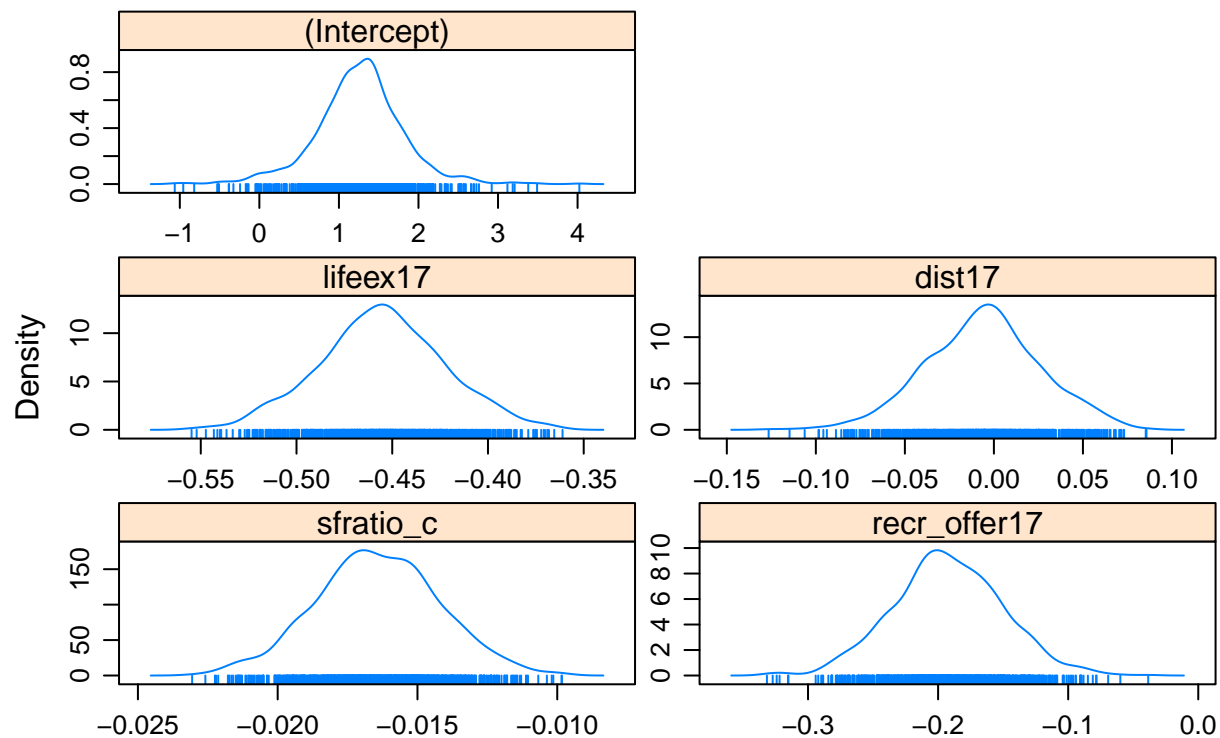[7]More details about the Laplacian model can be found in Appendix B.

Figure 2: Density Plots for Model A Non-Varying Coefficients
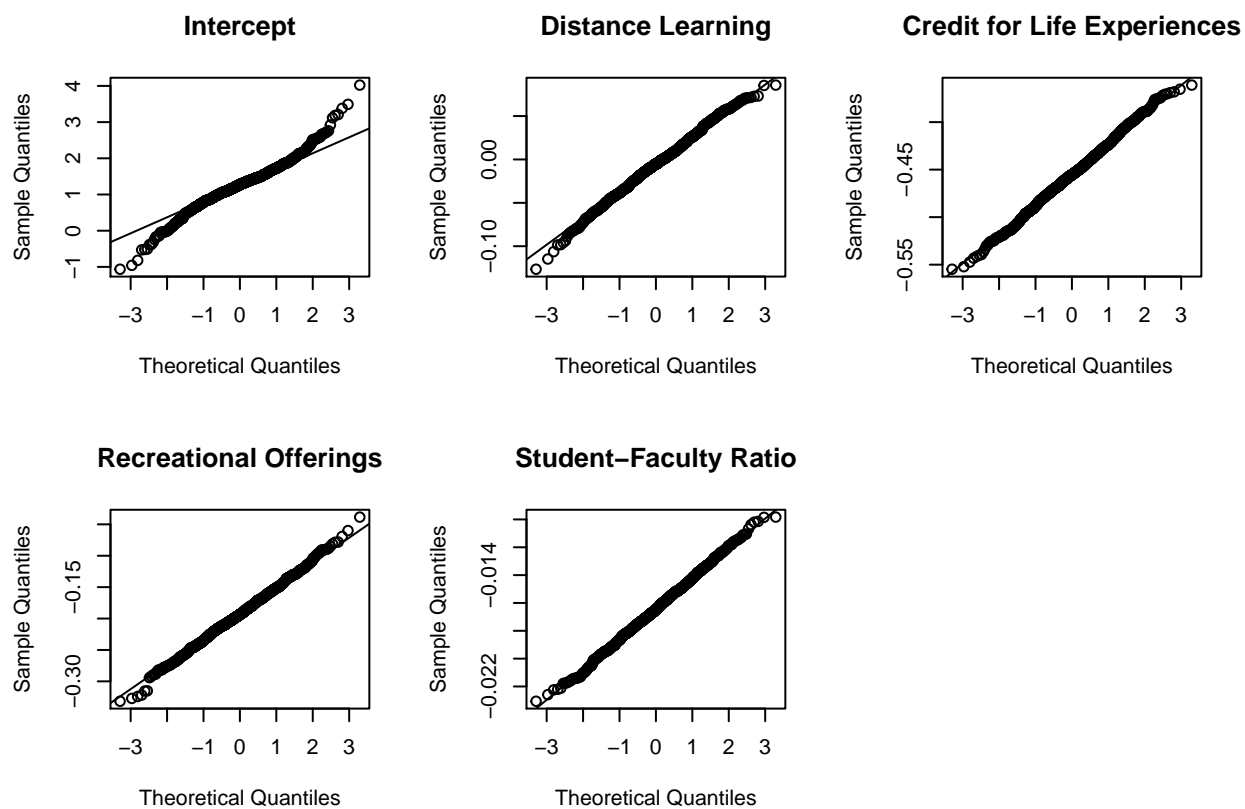

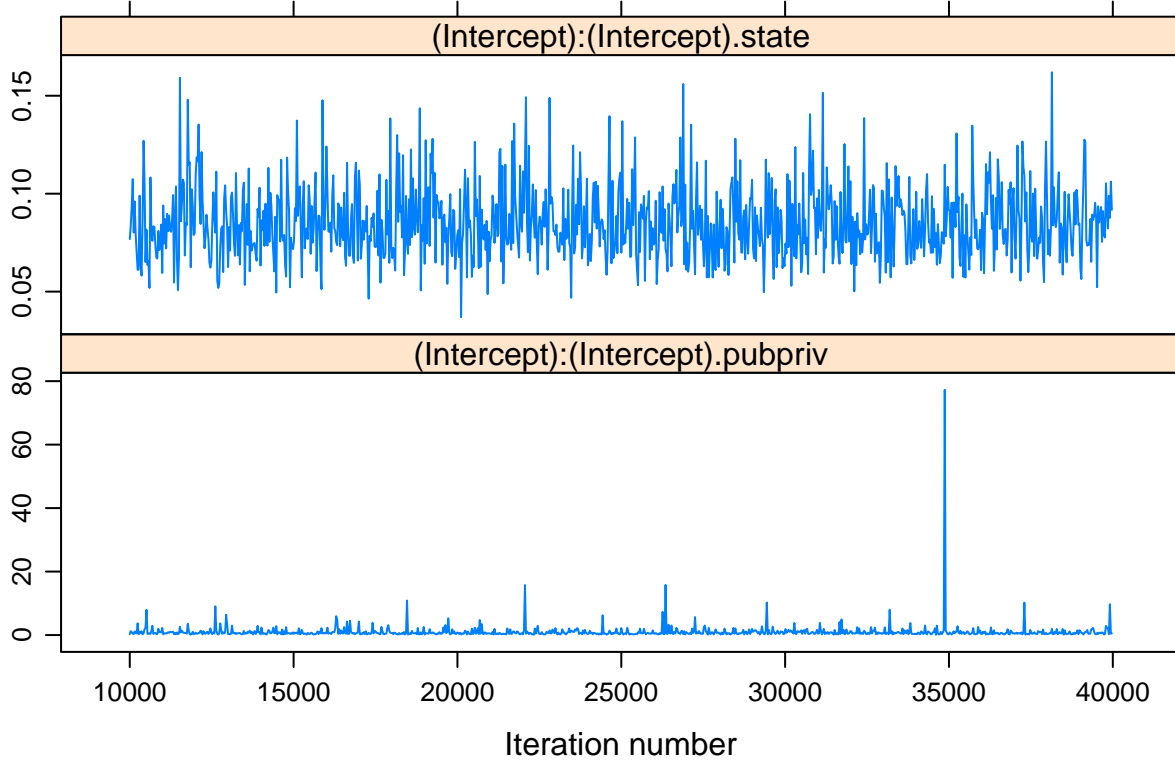Figure 3: QQ Plots for Model A Non-Varying Coefficients

Figure 4: Trace Plots for Model A Varying Intercept Variances

```
## sfratio_c    -0.016429667 -0.02067236 -0.01179105 1000.0000 0.001
```

The most obvious difference is that the coefficient for distance learning offerings has a very small p-value and a much larger positive magnitude in the Laplacian model, whereas the MCMC model has a very large p-value and a negative magnitude closer to 0. The sample density of Model A's iterations suggests that the distance learning coefficient estimate from the analogous Laplacian model, 0.108, is plausible but extremely unlikely. The intercepts between the two models appear to be consistent with one another. The other coefficients, while similar in direction and magnitude, are all of slightly larger magnitude in the Laplacian model, and most of the Laplacian model's coefficients fall outside of the 95% credible intervals produced from the MCMC model. The similarity between most of these parameter estimates affirms Model A's plausibility. Given our knowledge of the collinearity of the three binary predictors, Model A provides a more appropriately conservative suggestion that distance learning may not have an effect on retention rate once the other predictors have been accounted for.

The trace plot and density plot for the variance between state-specific intercepts (Figure 4) suggest that the posterior for the variance estimate is convergent. Based on the narrow spread of the posterior, it is highly credible that the between-state variance is a positive value near 0. However, the variance of intercepts between the public/private statuses (Figure 4) has a spread that is harshly truncated at 0, so while the posterior mean of these intercepts' variance is higher than the posterior mean of the between-state variance, there is much more uncertainty about this estimate, and it is plausible that the true variance between the public/private statuses' intercepts is 0. The trace plots for both of these parameters reveal rough stationarity aside from occasional spikes.
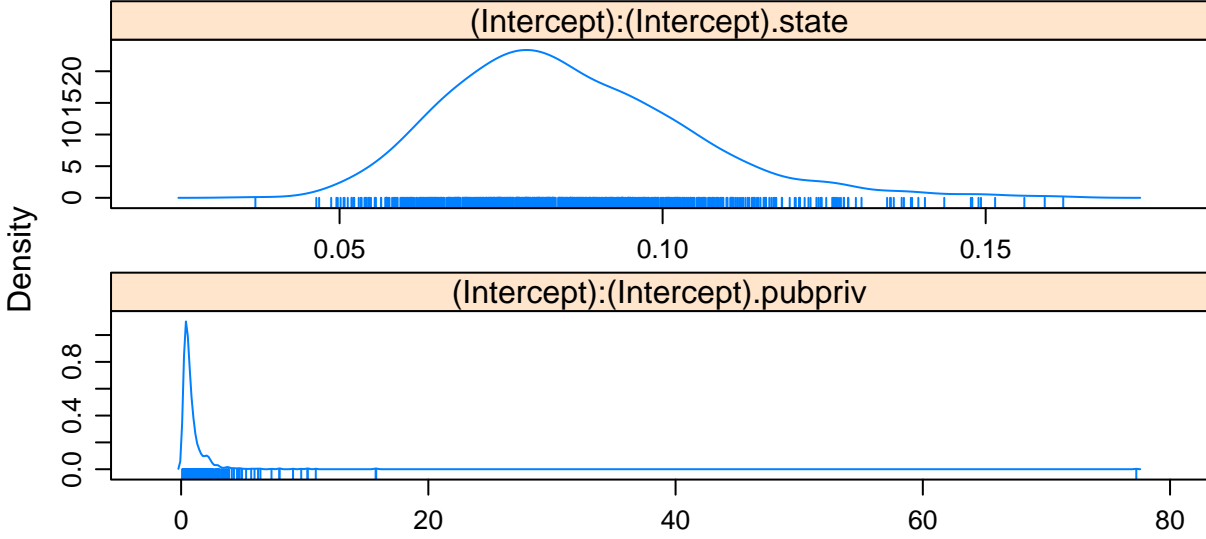
9

Figure 5: Density Plots for Model A Varying Intercept Variances

Model B

The QQ plot for the overall intercept's posterior (Figure 8) again reveals heavy tails, which makes the posterior mean relatively less probable than that of Model A. The QQ plot for the coefficient of recreational offerings (Figure 8) also shows that the tails of the sample posterior are heavier than that of a Normal distribution, which may be a result of the newly added varying slopes for this variable. This kurtosis compounds the uncertainty about the recreational offerings coefficient that was suggested with the drastic increase in its p-value compared to Model A. The remaining QQ plots (Figure 8) for the other coefficients show that the corresponding posterior densities (Figure 7) are roughly Normal, and all trace plots (Figure 6) show stationarity aside from occasional outliers, which suggests that these estimates are stable. The coefficient estimates are roughly consistent with those of Model A, aside from the obviously eliminated term for distance learning.

The density plot of the estimated variation between the state-specific intercepts (Figure 10), with its small spread, again allows for high credibility that the between-state variance is a small, positive, and close to 0. The density plot of the estimated variation for the intercept between different public/private statuses (Figure 10) has a wide spread that get sharply truncated at 0; thus, it is still likely that the true variation of intercepts between public/private statuses is 0, despite the posterior mean being greater than 1. Finally, the density plot of the estimated variation between recreational offerings slopes per public/private status (Figure 10) reveals a very large spread, and the distribution is also sharply truncated at 0, suggesting major uncertainty about this estimate. Despite this uncertainty, the trace plots for all of these parameters (Figure 9) reveal rough stationarity.

Discussion

Fitting the models with the MCMC framework allowed for more probabilistic assessments about the convergence and stability, which cannot be made from a frequentist framework. Therefore, although a Laplacian model fit was nearly unidentifiable, the same model under an MCMC fit could still be deemed valid based on the characteristics of the algorithm's trace over time and the shape of the resulting estimated posterior distribution. Including two non-nested grouping variables in the multilevel model, however, made the model more difficult to fit, and increasingly so when adding a varying intercept on top of the varying slopes. However, a Bayesian framework made it easier to visually pinpoint the credibility of the posterior
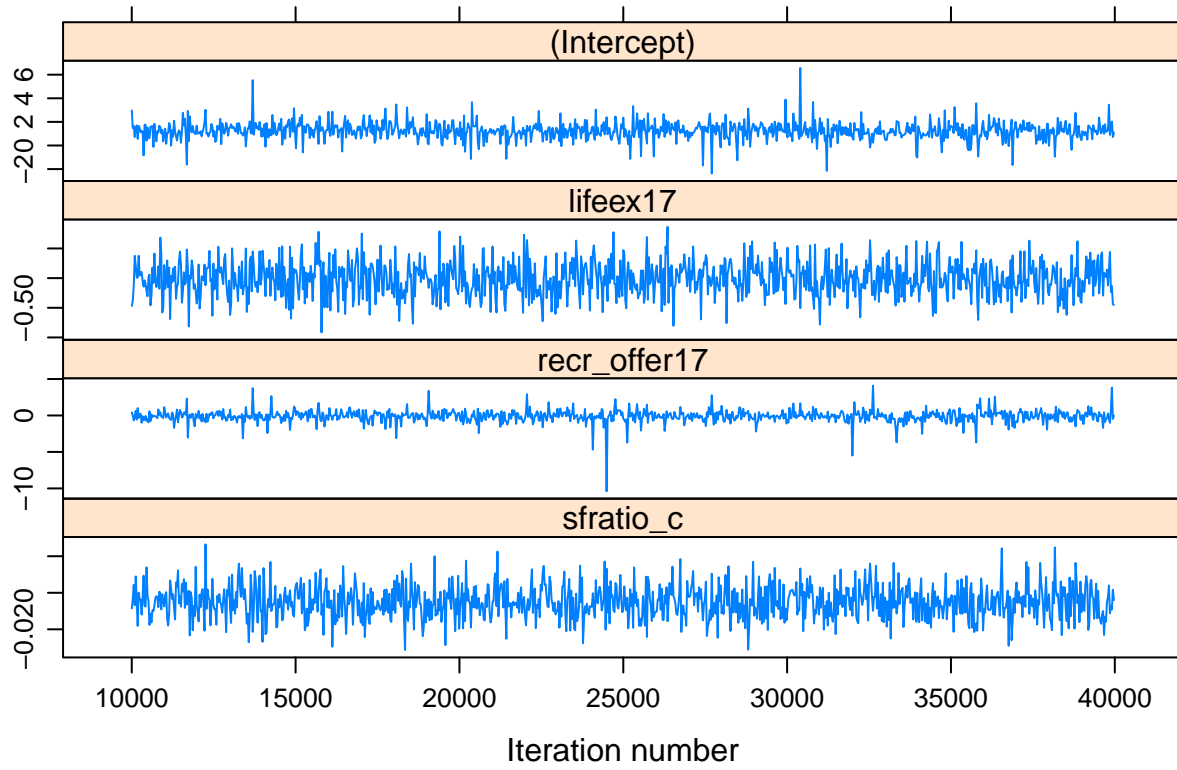
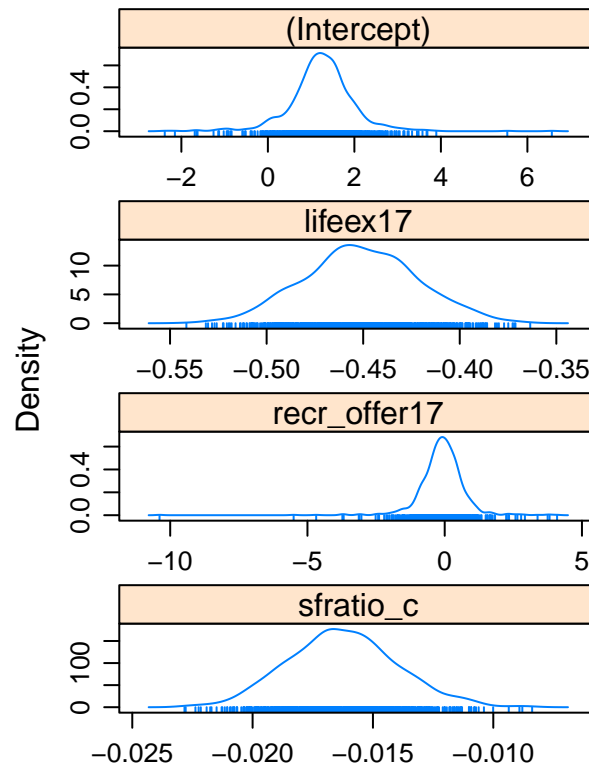Figure 6: Trace Plots for Model B Non-Varying Coefficients

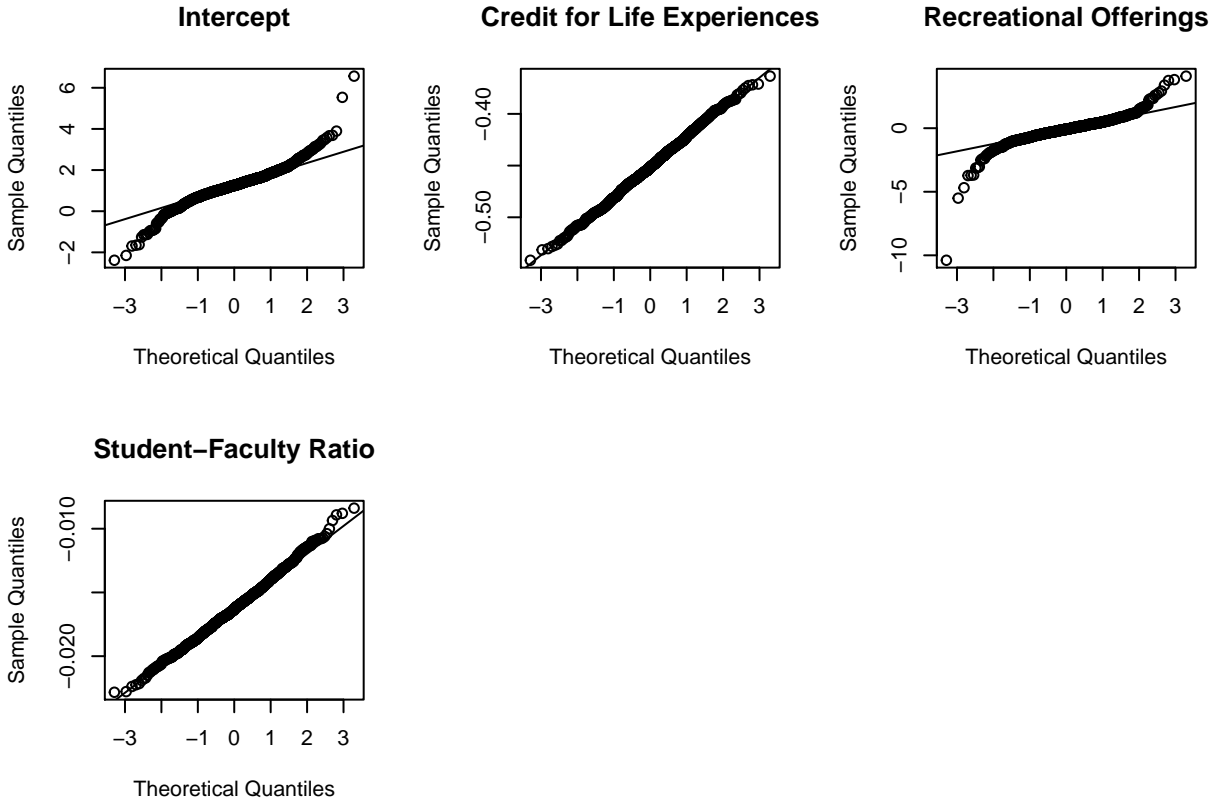Figure 7: Density Plots for Model B Non-Varying Coefficients

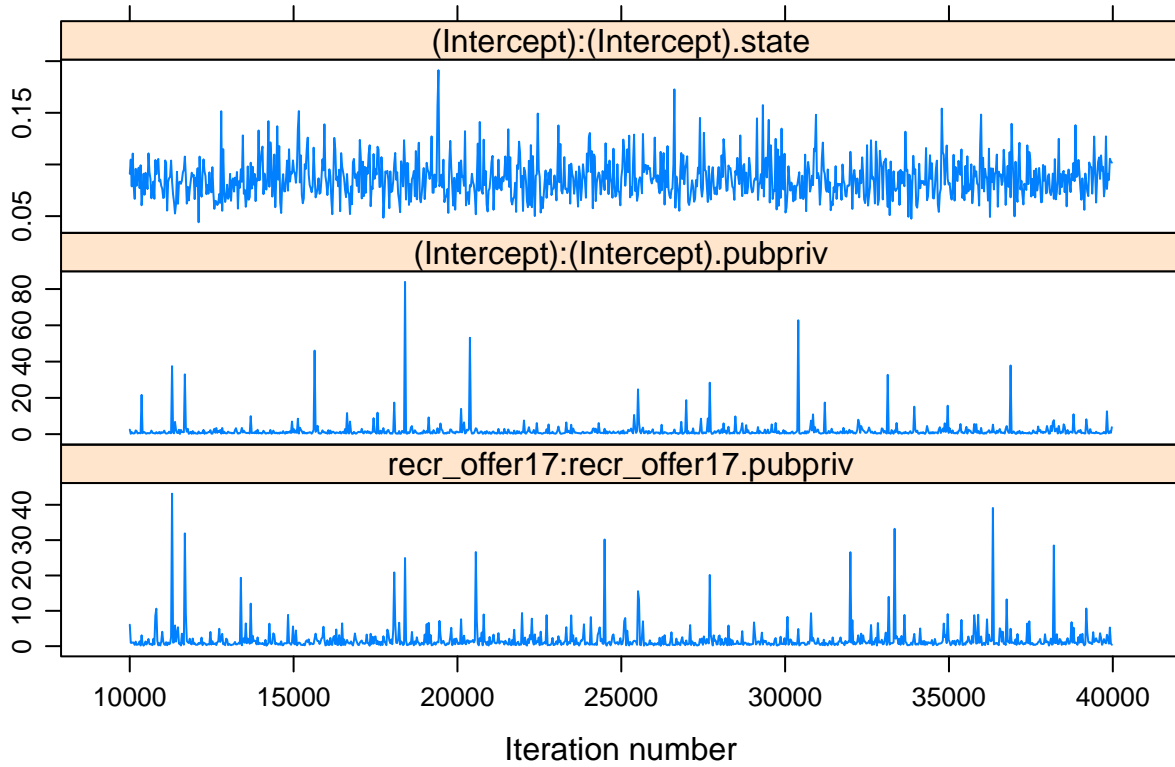Figure 8: QQ Plots for Model B Non-Varying Coefficients

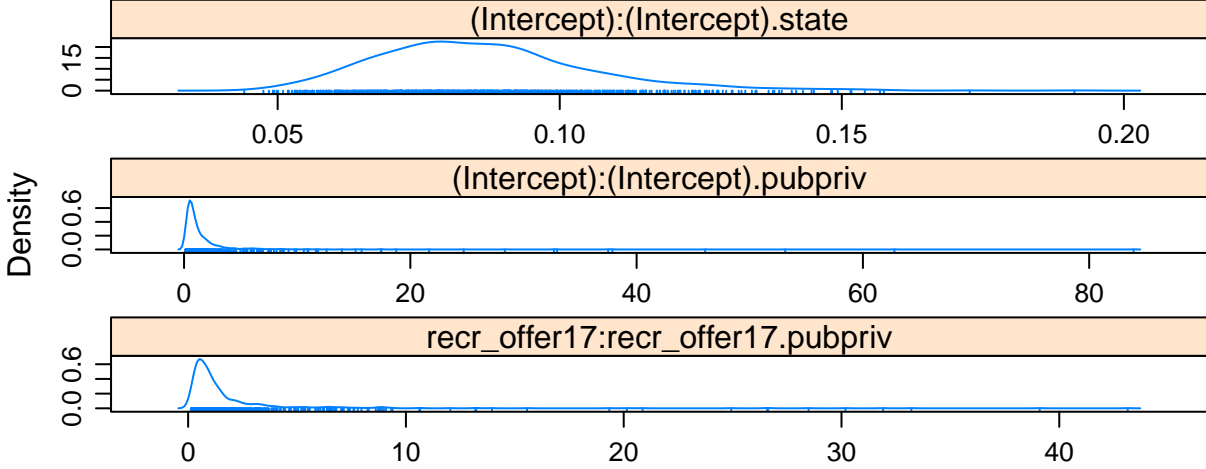Figure 9: Trace Plots for Model B Varying Intercept Variances

Figure 10: Density Plots for Model B Varying Intercept Variances

means for both the coefficient estimates and the between-group variances, as well as to confirm whether the MCMC sampling process was yielding adequately consistent results.

The hierarchical structure of Model A and B assumes a common distribution behind the intercepts corresponding to different states, and a common distribution behind the intercepts corresponding to different public/private statuses. Due to this characteristic of the models, the main coefficients represent trends that fit the entire set of institutions in general, and the coefficients are moderated by further information about an institution's state and public/private status. Both Model A and Model B denote that offering credit for life experiences, offering recreational/avocational options, and having a higher student-faculty ratio are all associated with lower retention rates. Both models identify credit for life experiences and recreational offerings as the predictors that are associated with the largest expected change in retention rate, whereas distance learning and student-faculty ratio contribute more marginally to the retention rate. Student-faculty ratio has a negative association with retention rate, which is consistent with the common intuition that more individualized attention contributes to a more encouraging academic setting. Yet the three binary predictors in the model, all of which represent features that are intended to make the postsecondary education experience more welcoming or accommodating, are all negatively associated with lower retention probabilities, which at first instinct may seem counterintuitive. However, these associations may actually reflect the complex push and pull between external barriers to educational attainment and institutions' efforts to compensate for these barriers. It is possible that an institution with a lower retention rate may actually be doing better in creating an accessible learning environment than its higher-retention-rate counterparts, perhaps because its student body's particular struggles inspired such efforts in the first place. Contrarily, it is also possible that the quality of such efforts genuinely have some sort of negative impact on retention rates.

The varying intercepts for public/private statuses suggest that private not-for-profit institutions are generally expected to have higher retention rates than private for-profit or public institutions. Additionally, the varying slopes reveal that recreational and avocational resources may have completely different roles in the educational experience for private not-for-profit vs. private for-profit and public institutions. When allowing the recreational offerings slope to vary for each public/private status, however, the estimate of the overall recreational offerings coefficient becomes much more uncertain, suggesting that the relationship between recreational offerings and retention rate is dominated by the difference between that variable's three groups. However, the model is unable to provide enough precision on the public/private status groups' varying intercepts or slopes to confidently claim that they differ.

Although these models identify overarching relationships between retention rates and several institutional features believed to improve or facilitate the educational experience, the causality behind these relationships remains unknown. Additionally, the precision of the parameter estimates are limited by the data, which is already a very substantial fraction of the entire population of postsecondary institutions throughout the

nation. The limitations of these models, and the questions that remain after understanding their results, can inspire institutions' future efforts to investigate specifically into the bidirectional causal forces between their own offerings and their retention rates. By starting conversations based on the model results, institutions can evaluate where they individually fit into the picture painted by the models, and use these conversations to improve their programs in accordance with their community's needs. For instance, an institution could obtain much more valuable data for these purposes by administering surveys or hosting student forums about how their resources are addressing student needs. Such measures may include questions about the degree of influence of personal, physical, or social factors on students' ability/desire to take advantage of certain resources, or overall evaluations of how welcome students feel taking advantage of certain resources. These measures would allow for a better understanding of confounding variables that also impact retention rates. Additionally, collecting more standardized information from these causal survey measures across institutions nationwide could help fit an improved model that identifies trends between students' relationships with institutional resources and retention rates, rather than the resources themselves and retention rates. Without considering such factors that exist outside an institution's responsibility or control, however, a retention rate can be a very noisy, uninformative metric of an institution's quality or appeal.