Assessing the Effects of Supplemental Instruction on Student Success

Ricardo Alexandro Aguilar

2/3/2021

Introduction

Supplemental instruction (SI) is a peer-led academic support model that became available to California State University, Chico students in the Fall 2012 semester. SI provides additional support to students who enroll in courses that have high enrollment, but relatively low pass rates. This institutional level study aims to determine the effect of SI on the likelihood students will receive a D, F, or withdraw from the class (DFW rate) while controlling for academic preparedness and demographic variables.

Data

The data contains 12,730 complete records of undergraduate students' grades, cumulative GPA, high school eligibility index, the course they are enrolled in, the term of course enrollment, gender, whether they attended SI, and whether they are a student of color. From Fall 2012 to Fall 2017, SI was available for 19 courses from various departments including Biological Sciences, Computer Science, History, Sustainable Manufacturing, and Political Science. High school eligibility index (HSEI) is a function of high school GPA, ACT score, and SAT score (continuous values scaled from 0 - 5). The variable for grade was dichotomized to whether the student received a D, F, or withdrew from the class. Originally, there were 18,639 observations, but graduate, post-baccalaureate, and continuing education students were omitted. The original study considered every first attempt at a course as an independent observation. For example, a student can enroll in two new courses and both of those first attempts were considered independent observations. Repeated attempts for a course were omitted from the study. As a result of this approach and to protect the sensitive information of the participating students, the data were de-identified such that multiple observations cannot be traced to an individual student. Additionally, a complete-case analysis was conducted, so the 2,095 observations that were missing a value for any variable included in the model were omitted as well.

Model

A random intercepts, logistic regression model with a Bayesian framework was used to model the data. The model uses the indicator for DFW as the outcome (DFW = 1 and not DFW = 0.) The model includes fixed effect terms for mean-centered cumulative GPA, mean-centered HSEI, gender, student of color status, and SI attendance status. Summary statistics for DFW and the previously mentioned variables can be found in Table 1. The model also included random intercepts for term and course. Using i as the index for term, j as the index for class, and k as the index for the observation within a certain term and class, the mathematical specification of the model can be written as $y_{ijk}|\pi_{ijk} \sim \text{Bernoulli}(\pi_{ijk})$ where

$$\pi_{ijk} = \frac{\exp\left[X'_{ijk}\alpha + \beta_{1,i} + \beta_{2,j}\right]}{1 + \exp\left[X'_{ijk}\alpha + \beta_{1,i} + \beta_{2,j}\right]}.$$

The elements of the vector α are the population regression coefficients and an overall intercept term while β is a vector of term and course specific regression coefficients. The prior densities for α's will be normal with variances set to specified values while the prior densities of the β's will be normal with inverse gamma as the distribution of their variances.

Specifying Priors

All of the fixed effects will have a normal distribution as their prior distribution. Dr. Erik Wasinger, a professor for the chemistry department at California State University, Chico and subject matter expert, said that the DFW rates of courses that implemented SI are approximately 20%, so this was used to determine the center for the prior of the intercept term. The prior for the intercept was centered at -1.38 (ln(0.2/0.8)) with 1 as the variance. The results of a prior study suggest that there is a strong, positive relationship between cumulative GPA and students' grades in a course that provides SI[1]. Therefore the prior distribution for the coefficient of cumulative GPA will be centered at -1.11 with 0.25 as the variance. Evidence has been shown that HSEI is a strong predictor of students' grades in college [2]. The prior for HSEI will be centered at -0.69 with 0.25 as the variance. Gender is not expected to be a significant predictor of DFW [1], so the prior will be centered at 0 with 0.25 as the variance. The odds of passing a class after attending SI at least once is expected to double[3]. In the context of our model, this equates to reducing the odds of DFW by 50%, so the prior will be centered at -0.69 with 0.0625 as the variance. Students of color tend to have slightly lower passing rates than their counterparts[4], so the prior will be centered at 0.095 with 0.0625 as the variance. The priors for the random intercept term are all normal distributions centered at 0 with inverse gamma as the distribution of their variance. All of the inverse gamma distributions are centered at 0.44 with shape and rate equal to 8 and 18, respectively. These values were determined by finding a range of plausible values.

Model Diagnostics

At 5 chains and 50,000 iterations, convergence for the fixed effects and variance of random intercept terms occurs quickly. From Figure 1, the autocorrelation for every fixed effect and variance parameter estimates drops to 0 within 30 lags. The time series plots, for each fixed effect and variance parameter estimate in Figure 2, resemble the fuzzy caterpillar. Thus, we can be confident in the results of our normal model.

Sensitivity Analysis

In the sensitivity analysis, the fixed and random effects were assumed to be from a t-distribution instead of a normal distribution. The centers and taus of the t-distributions are the same as the ones in the normal distributions. Juarez and Steel proposed using Ga(2, 0.1) as a default prior for the degrees of freedom of t-distributions[5]. Thus, priors for the degrees of freedom of all of the fixed and random effects were set to their proposed distribution.

From Figure 5, the posterior distributions of the fixed effects and random effects from the t-distribution priors are very similar to the posterior distributions of the normal priors (with the exception of the intercept term). Figures 3 and 4 reveal that, at 5 chains and 50,000 iterations, convergence occurs quickly for most of the fixed effects and random effects, but the intercept term struggles to converge. Since the estimates from the t-distribution priors in Table 3 are similar to the normal distribution priors in Table 2, there is evidence that suggests the estimates are robust to relaxing the assumption of normal priors.

Additional Analyses

Although the estimates are similar, there are some issues that require more in depth analyses. The first issue is the lack of convergence for the intercept term in the model that uses t-distribution priors. The other issue is the discrepancy between the prior and posterior distributions for the coefficient of cumulative GPA. The posterior distribution for the coefficient of cumulative GPA suggests a massive effect from a one unit increase in cumulative GPA on the odds of a student receiving a D, F, or W. A large effect was expected, but the estimated value of that effect is large enough to require a more thorough examination.

Eigenvalue decomposition was conducted on the $X'X$ matrix shown in Table 4. The results, which can be seen in Table 5, show that $X'X$ is a positive definite matrix since it is symmetric and all of its eigenvalues are positive. To explore the relationship between cumulative GPA and DFW, cumulative GPA was divided into even intervals (of 1) and the DFW rate within each interval was then computed. The results, as seen in Table 6, reveal that the proportion of observations who do not receive a D, F, or W in the lowest interval for cumulative GPA is much greater than the proportion of students who received a D, F, or W in the highest interval. To examine this further, cumulative GPA was divided into smaller, even intervals (of 0.1). Figure 6 shows the nonlinear nature of this relationship. The relationship appears to be linear from values of 0 to 2 of cumulative GPA, but the decrease in DFW rate from 2 to 4 is nonlinear.

To further examine the issues, the model was recreated using relatively vague priors, excluding the random intercept for term, and using different amounts of iterations. The parameter estimates when using relatively vague priors (precision for fixed effects was set to 1), shown in Table 7, are similar to that of Tables 2 and 3 which provides more evidence of the robustness of the parameter estimates in the previous models. Figures 7 and 8 show that similar convergence to the previous models is achieved when using relatively vague priors. Figure 9 shows that, by design, the discrepancy between the prior and posterior distributions is no longer as severe as before. We can see from, Figures 10 and 11, that convergence for the intercept term worsens after removing the random effect for term. However, Table 8 and Figure 12 show that the posterior distributions and parameter estimates are not much different than that of the models that include the random effect for term. From Figure 13, the autocorrelation does not improve even after increasing the amount of iterations to 100,000 (with 5 chains). Figure 14 provides additional evidence that convergence is still somewhat of an issue for the intercept term at larger amounts of iterations. Although lack of convergence for the intercept term is an issue in the models with t-distribution priors, Table 9 suggests that the estimate of the intercept term will eventually converge at a value around -1.73 and -1.71 after 10,000 iterations.

Conclusion

Table 2 contains the regression coefficient estimates for the fixed effects and variance parameters of the random intercept terms when using the normal distribution as the prior. None of the credible intervals for the fixed effects in Table 2 contain zero, so there is evidence of significance for all them. As expected, cumulative GPA has the largest effect on DFW compared to the other covariates. The coefficient for cumulative GPA seems to be estimated at values around -2.9 (log OR) regardless of changes in model assumptions. This could be due to having extremely low DFW rates at the highest values of cumulative GPA while having relatively much higher rates of non-DFW grades at lower cumulative GPA values. Conducting an eigenvalue decomposition did not reveal any issues with the $X'X$ matrix. From Figure 5, we can see that the posterior distributions are much more narrow than the priors with a slight shift from where their respective priors are centered; the most notable difference is the fixed effect for cumulative GPA. Comparing Figures 5 and 9 shows that changing to relatively vague priors, by increasing the variance, produces posterior distributions that are very similar to the ones that used the original priors. After controlling for cumulative GPA, HSEI, gender, and student of color status, undergraduate students who attended at least one SI session have 0.58 times the odds of receiving a D, F, or withdrawing from the class compared to students who did not attend any SI sessions.

Appendix

References

[1] Rabitoy, Eric et al. "Supplemental Instruction: The Effect of Demographic and Academic Preparation Variables on Community College Student Academic Achievement in STEM-Related Fields." Journal of Hispanic Higher Education (2015), Vol. 14(3) 240-255

[2] Geiser, Saul et al. "Validity of High-School Grades in Predicting Student Success Beyond the Freshman Year." Research & Occasional Paper Series: (2007) CSHE.6.07

[3] Guarcello, Maureen et al. "Balancing Student Success: Assessing Supplemental Instruction Through Coarsened Exact Matching." Tech Know Learn (2017), 22:335–352

[4] Rath, Kenneth et al. "Impact of Supplemental Instruction in Entry-Level Chemistry Courses at a Midsized Public University." Journal of Chemical Education (2012), 89, 449-455

[5] Juarez, Miguel et al. "Model-based Clustering of non-Gaussian Panel Data Based on Skew-t Distributions." Journal of Business and Economic Statistics (2010), 28(1):52-66

Table 1: Summary Statistics for DFW and Fixed Effects Variables

	Mean	SD	Range
Cumulative GPA	2.86	0.66	(0.19, 4.0)
HS Eligibility Index	3.62	0.41	(2.04, 4.908)
		N	%
Grade			
	Not DFW	10,590	83.2%
	DFW	2,140	16.8%
Gender			
	Female	8,057	63.3%
	Male	4,673	36.7%
Supplemental Instruction			
	No SI Sessions	6,250	49.1%
	At least 1 Session	6,480	50.9%
Student of Color			
	Not SOC	5806	45.6%
	SOC	6924	54.4%

Table 2: Regression Coefficient Estimates (log OR) for Fixed Effects with Normal Priors and Random Intercept Variance Parameter Estimates

	Estimate	SD	2.5%	97.5%
Intercept	-2.23	0.45	-3.11	-1.34
Cumulative GPA	-2.9	0.07	-3.04	-2.76
HS Eligibility Index	-0.75	0.1	-0.94	-0.56
Gender	-0.23	0.07	-0.37	-0.1
Supplemental Instruction	-0.55	0.07	-0.69	-0.41
Student of Color	0.31	0.06	0.18	0.43
Variance for Term	0.7	0.19	0.37	1.12
Variance for Class	0.53	0.13	0.3	0.82

Table 3: Regression Coefficient Estimates (log OR) for Fixed Effects with t-Distribution Priors and Random Intercept Variance Parameter Estimates

	Estimate	SD	2.5%	97.5%
Intercept	-1.71	0.44	-2.54	-0.8
Cumulative GPA	-2.92	0.07	-3.06	-2.78
HS Eligibility Index	-0.75	0.09	-0.94	-0.57
Gender	-0.24	0.07	-0.37	-0.1
Supplemental Instruction	-0.54	0.07	-0.68	-0.41
Student of Color	0.31	0.07	0.18	0.44
Variance Term	0.7	0.19	0.37	1.12
Variance Class	0.55	0.14	0.31	0.86

Table 4: X'X Matrix

	Intercept	Cumulative GPA	HSEI	Gender	SI	SOC
Intercept	12730	36377	46195	4673	6480	6924
Cumulative GPA	36377	109437	133403	12626	19390	19124
HSEI	46195	133403	169795	16705	23371	24443
Gender	4673	12626	16705	4673	1875	2551
SI	6480	19390	23371	1875	6480	3943
SOC	6924	19124	24443	2551	3943	6924

Table 5: Eigenvalues by Variable for Diagonal Position

	Eigenvalues
Intercept	297437
Cumulative GPA	4090
HSEI	3728
Gender	2504
SI	2156
SOC	123.2

Table 6: DFW by Cumulative GPA Intervals N(%)

	Not DFW	DFW
(0,1]	24(0.14)	143(0.86)
(1,2]	385(0.37)	645(0.63)
(2,3]	4641(0.79)	1249(0.21)
(3,4]	5540(0.98)	103(0.02)

Table 7: Regression Coefficient Estimates (log OR) for Fixed Effects with Vague t-Distribution Priors and Random Intercept Variance Parameter Estimates

	Estimate	SD	2.5%	97.5%
Intercept	-1.76	0.44	-2.62	-0.87
Cumulative GPA	-2.93	0.07	-3.07	-2.79
HS Eligibility Index	-0.75	0.1	-0.94	-0.56
Gender	-0.24	0.07	-0.37	-0.1
Supplemental Instruction	-0.53	0.07	-0.68	-0.39
Student of Color	0.32	0.07	0.19	0.45
Variance Term	0.7	0.19	0.37	1.13
Variance Class	0.55	0.14	0.31	0.86

Table 8: Regression Coefficient Estimates (log OR) for Fixed Effects with t-Distribution Priors and No Random Intercept for Term

	Estimate	SD	2.5%	97.5%
Intercept	-1.65	0.33	-2.32	-1.01
Cumulative GPA	-2.89	0.07	-3.03	-2.75
HS Eligibility Index	-0.76	0.09	-0.94	-0.57
Gender	-0.22	0.07	-0.35	-0.09
Supplemental Instruction	-0.48	0.07	-0.62	-0.35
Student of Color	0.29	0.06	0.16	0.42
Variance Class	0.54	0.14	0.31	0.84

Table 9: Intercept Estimates of T Models at Varying Iterations

	Estimate	SD	2.5%	97.5%
1,000 Iterations	-1.93	0.4	-2.59	-1.07
5,000 Iterations	-1.78	0.45	-2.62	-0.87
10,000 Iterations	-1.73	0.47	-2.66	-0.8
20,000 Iterations	-1.71	0.45	-2.55	-0.79
50,000 Iterations	-1.71	0.44	-2.54	-0.8
100,000 Iterations	-1.73	0.46	-2.61	-0.81

Figure 1: Autocorrelation Plots for Fixed Effects with Normal Priors and Random Intercept Variance Parameter Estimates

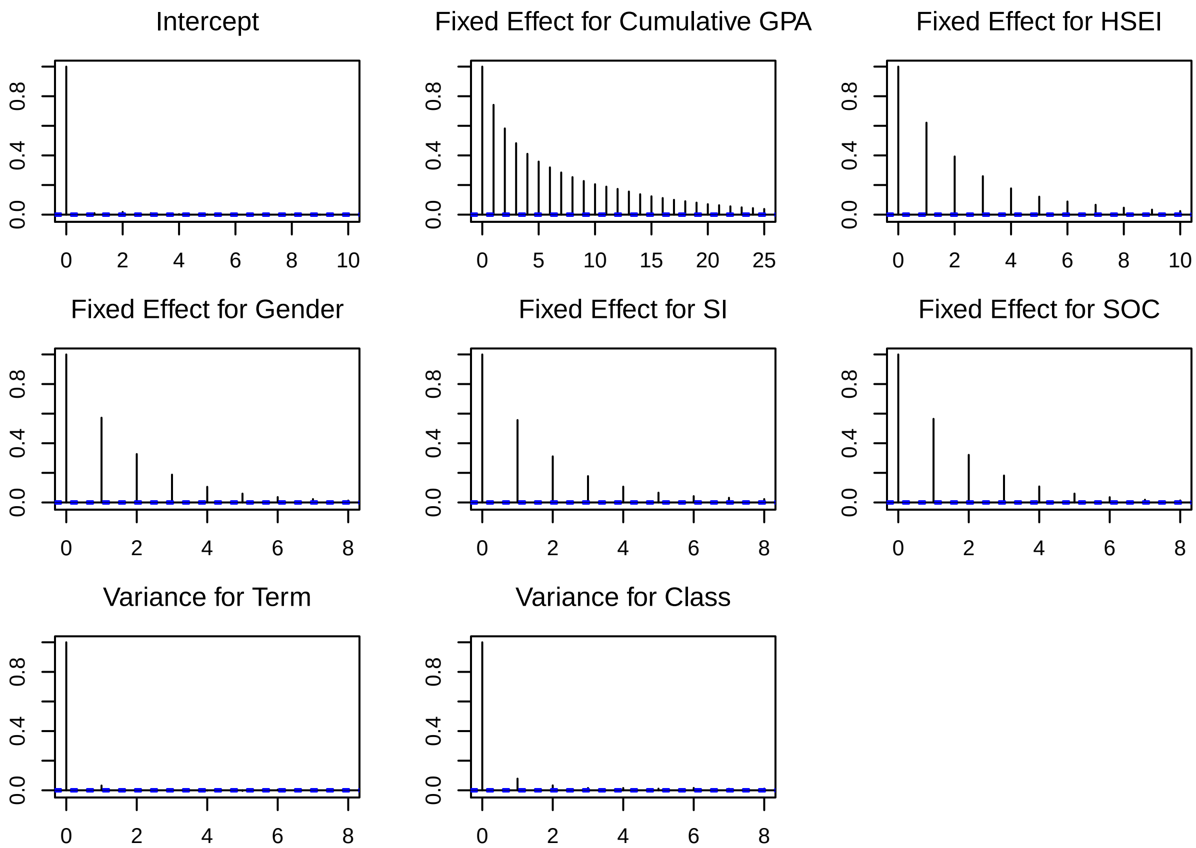

Figure 2: Time Series Plots for Fixed Effects with Normal Priors and Random Intercept Variance Parameter Estimates

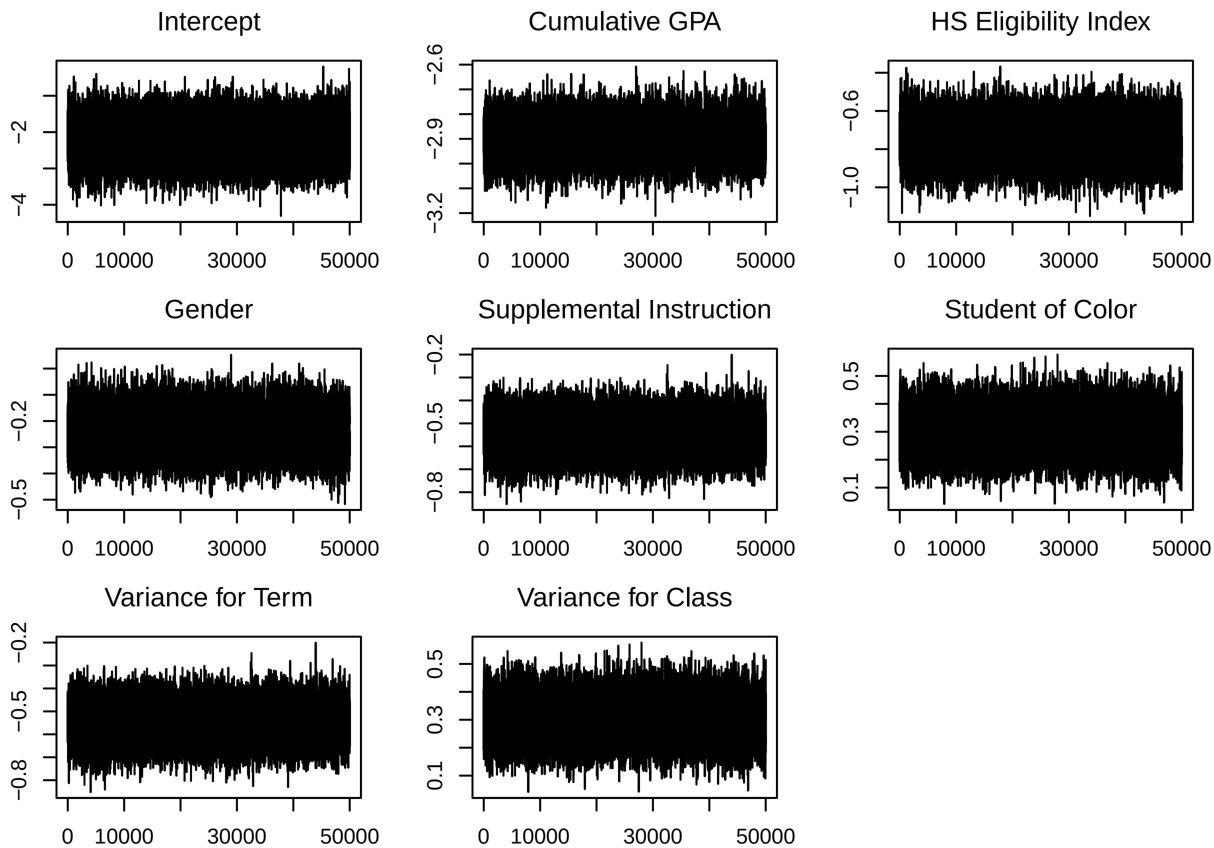

Figure 3: Autocorrelation Plots for Fixed Effects with t-Distribution Priors and Random Intercept Variance Parameter Estimates

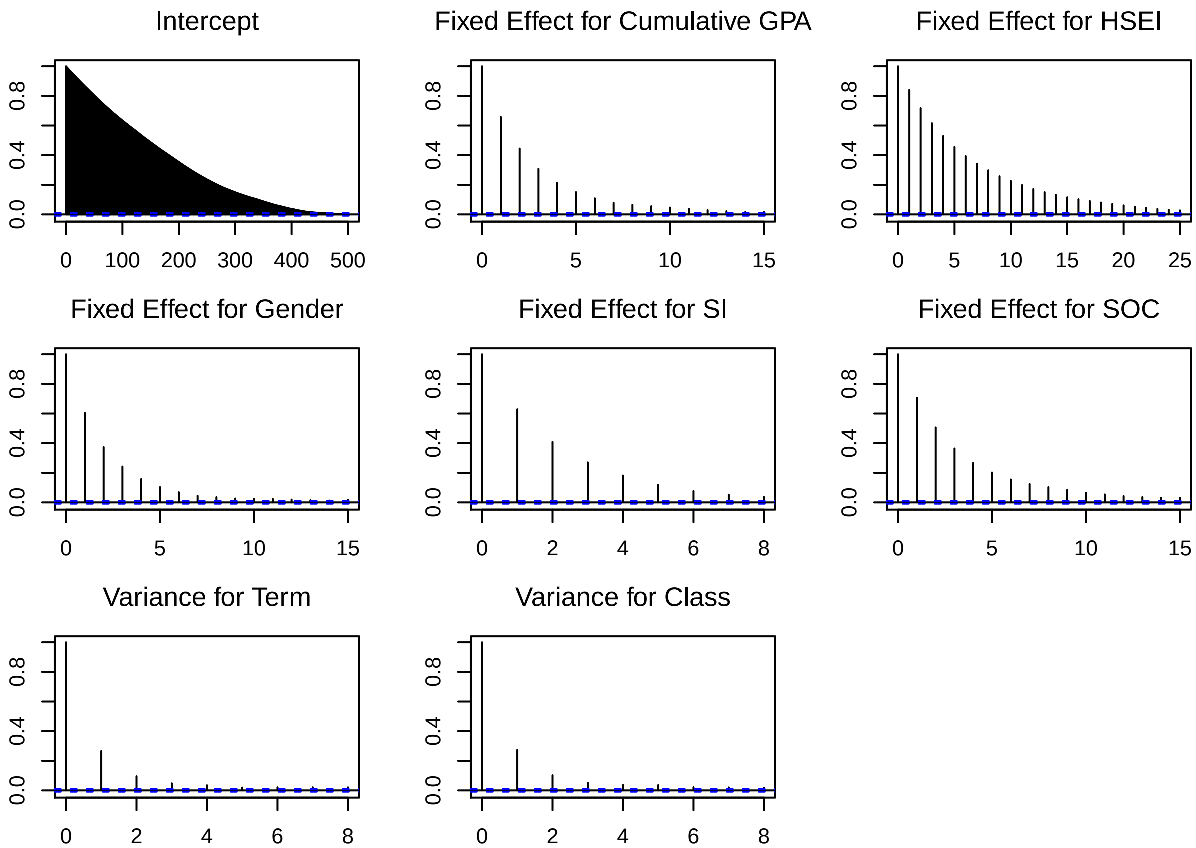

Figure 4: Time Series Plots for Fixed Effects with t-Distribution Priors and Random Intercept Variance Parameter Estimates

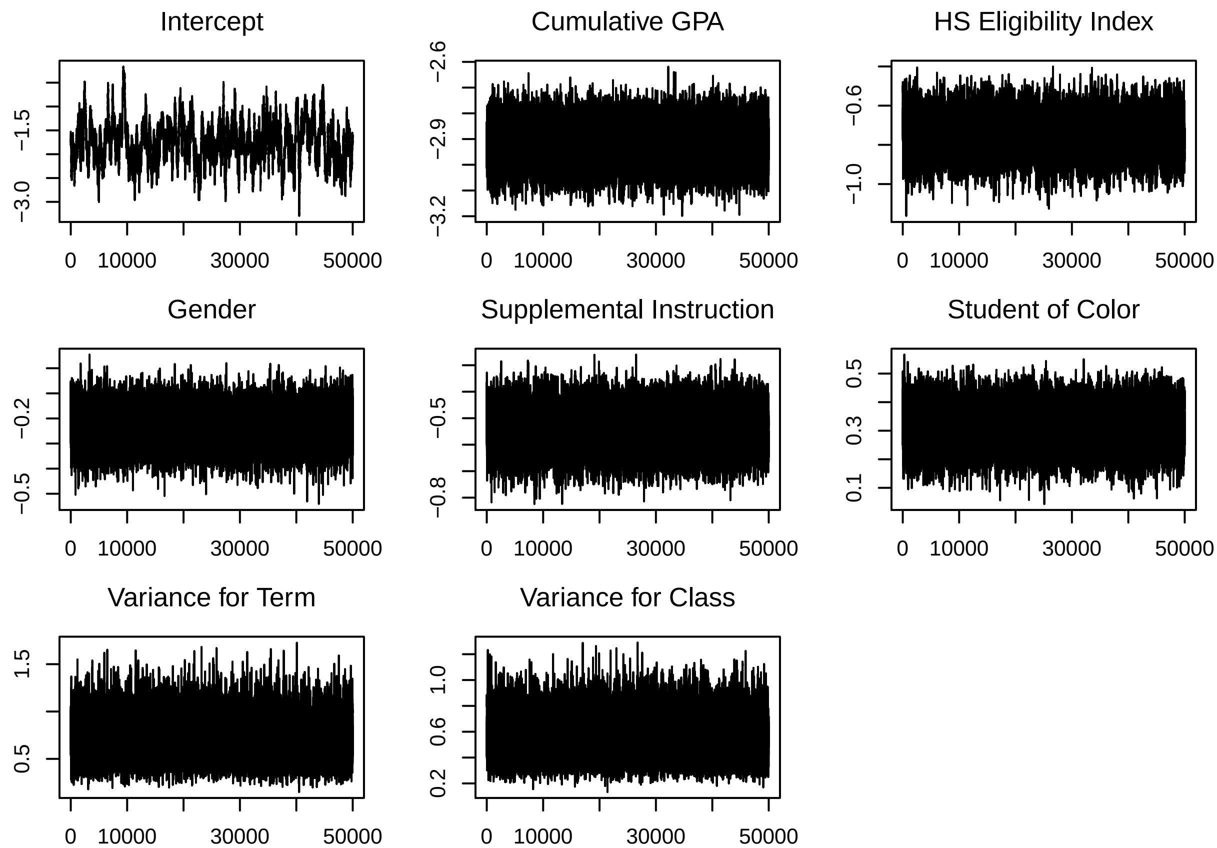

Figure 5: Prior and Posterior Distributions for the Fixed Effects and Random Intercept Variance Parameter Estimates

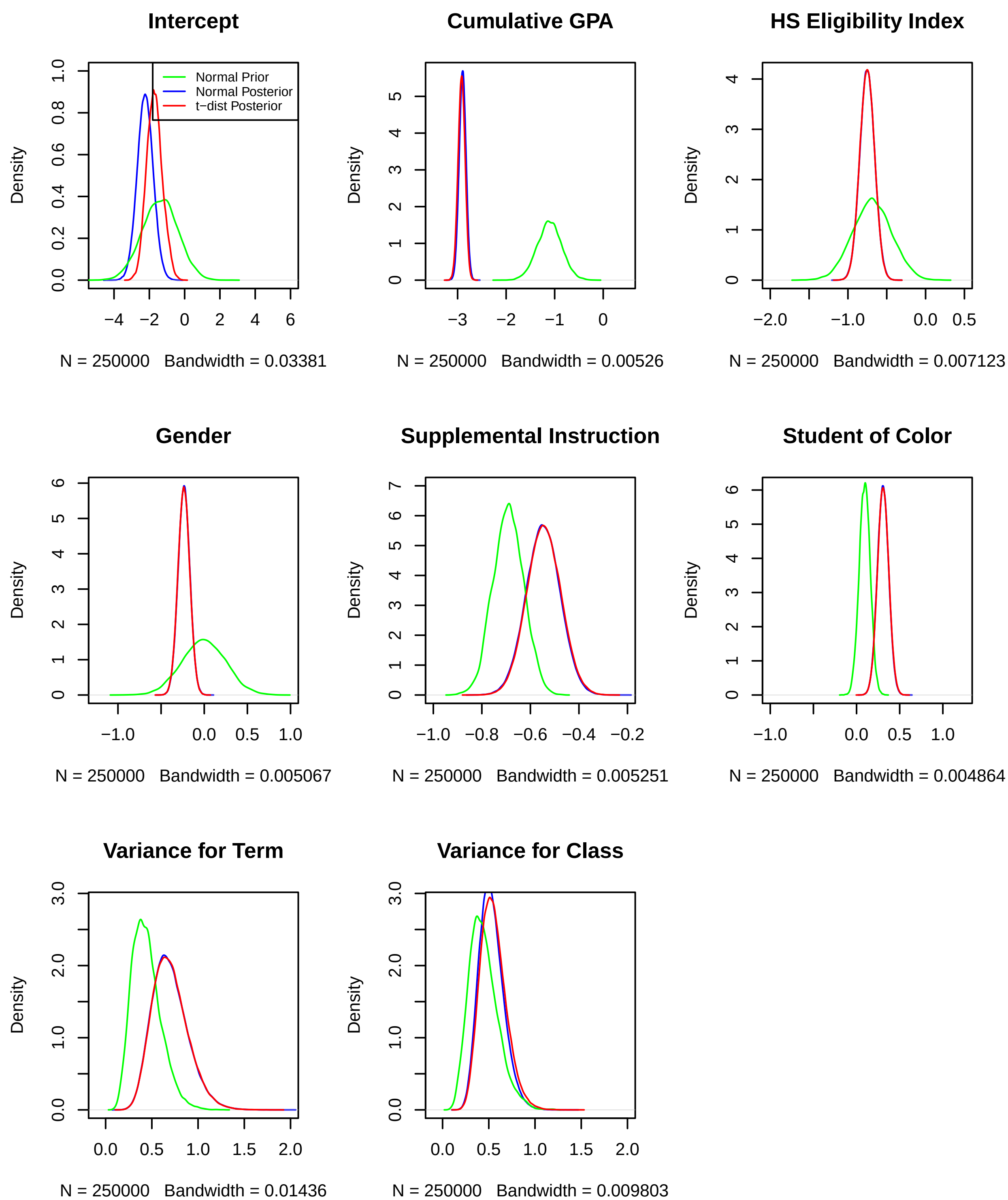

Figure 6: DFW Rate by Cumulative GPA Intervals

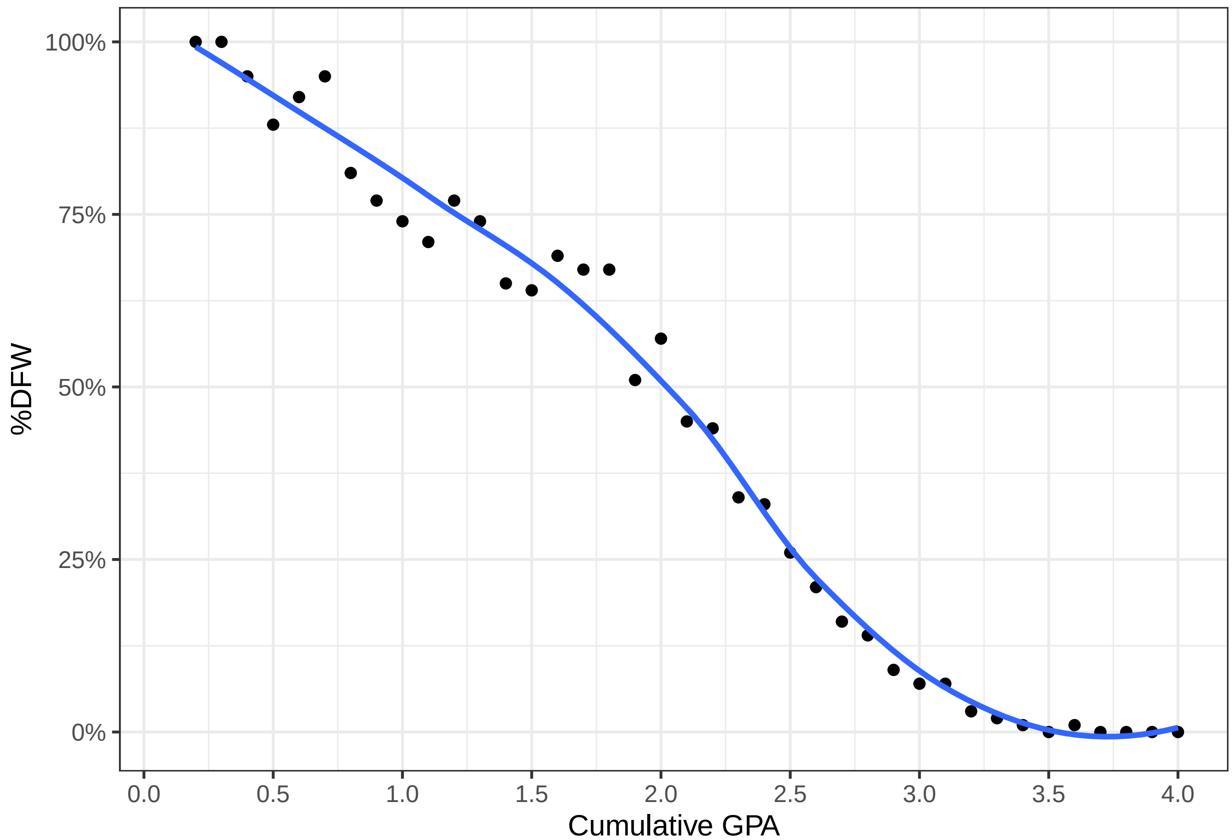

Figure 7: Autocorrelation Plots for Fixed Effects with Vague t-Distribution Priors and Random Intercept Variance Parameter Estimates

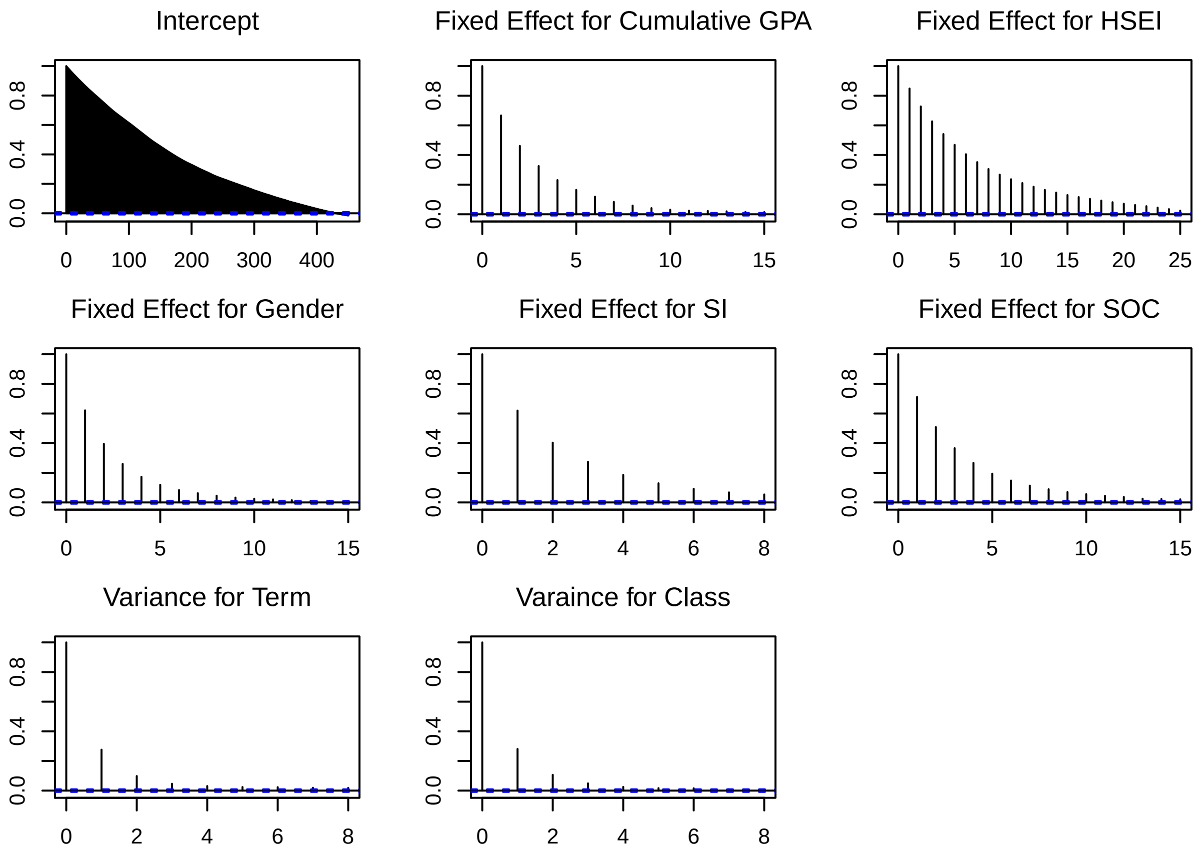

Figure 8: Time Series Plots for Fixed Effects with Vague t-Distribution Priors and Random Intercept Variance Parameter Estimates

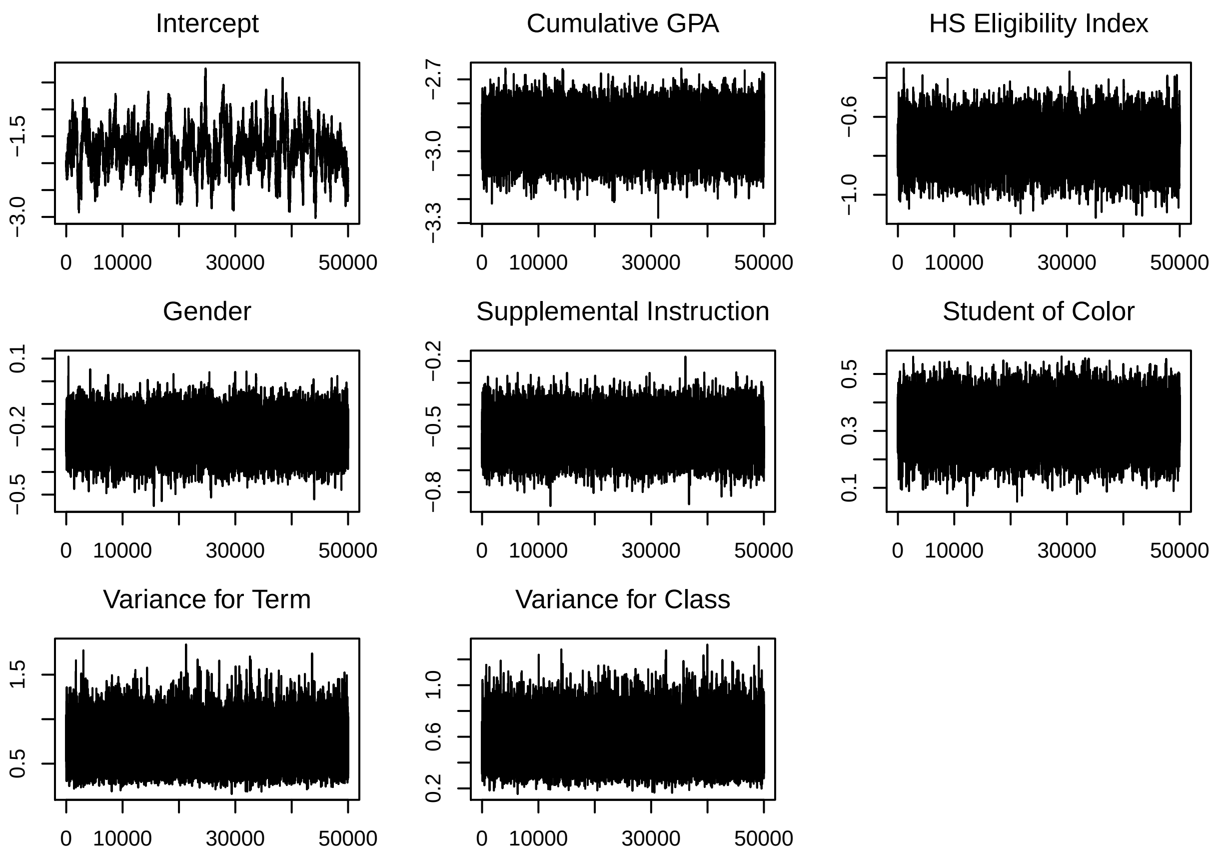

Figure 9: Prior and Posterior Distributions for the Fixed Effects Using Vague Priors and Random Intercept Variance Parameter Estimates

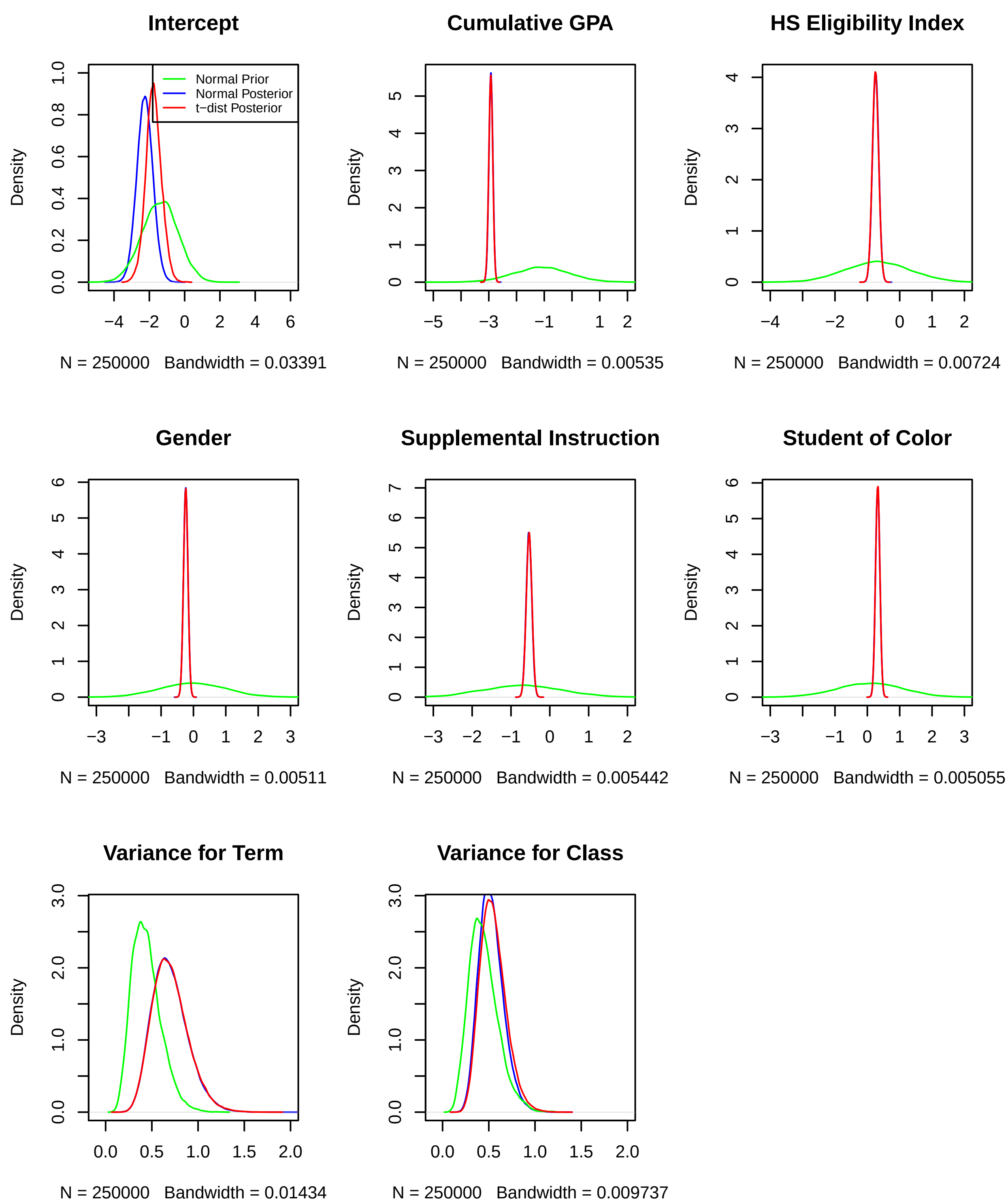

Figure 10: Autocorrelation Plots for Fixed Effects with t-Distribution Priors and No Random Intercept for Term

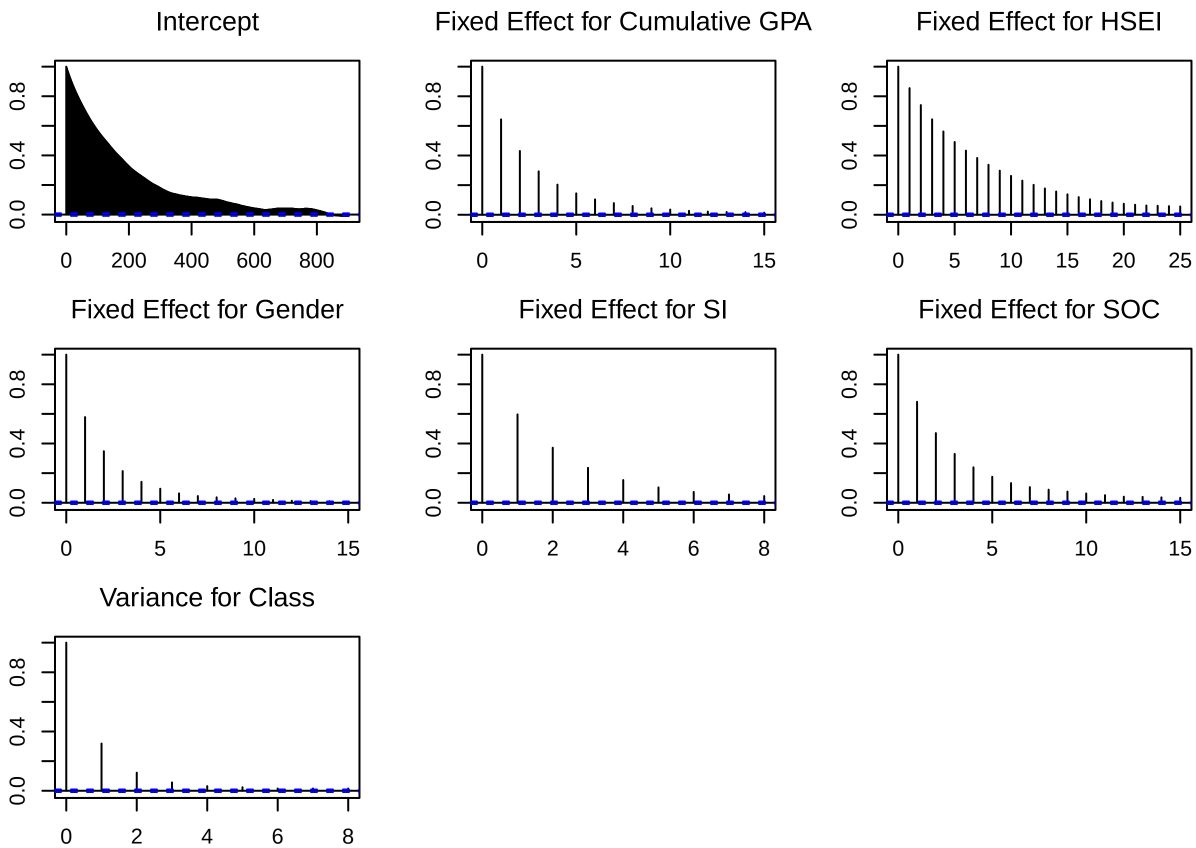

Figure 11: Time Series Plots for Fixed Effects with t-Distribution Priors and No Random Intercept for Term

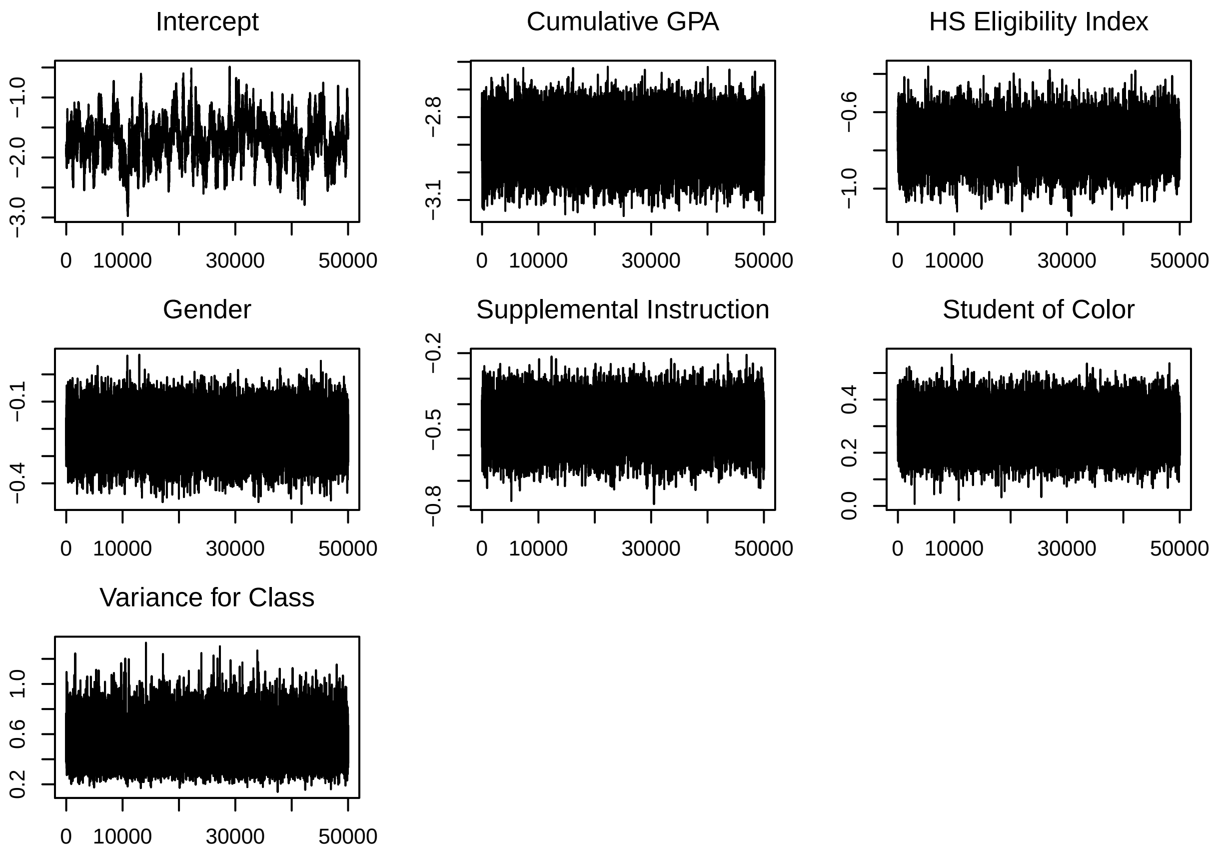

Figure 12: Prior and Posterior Distributions for the Fixed Effects with t-Distribution Priors and No Random Intercept for Term

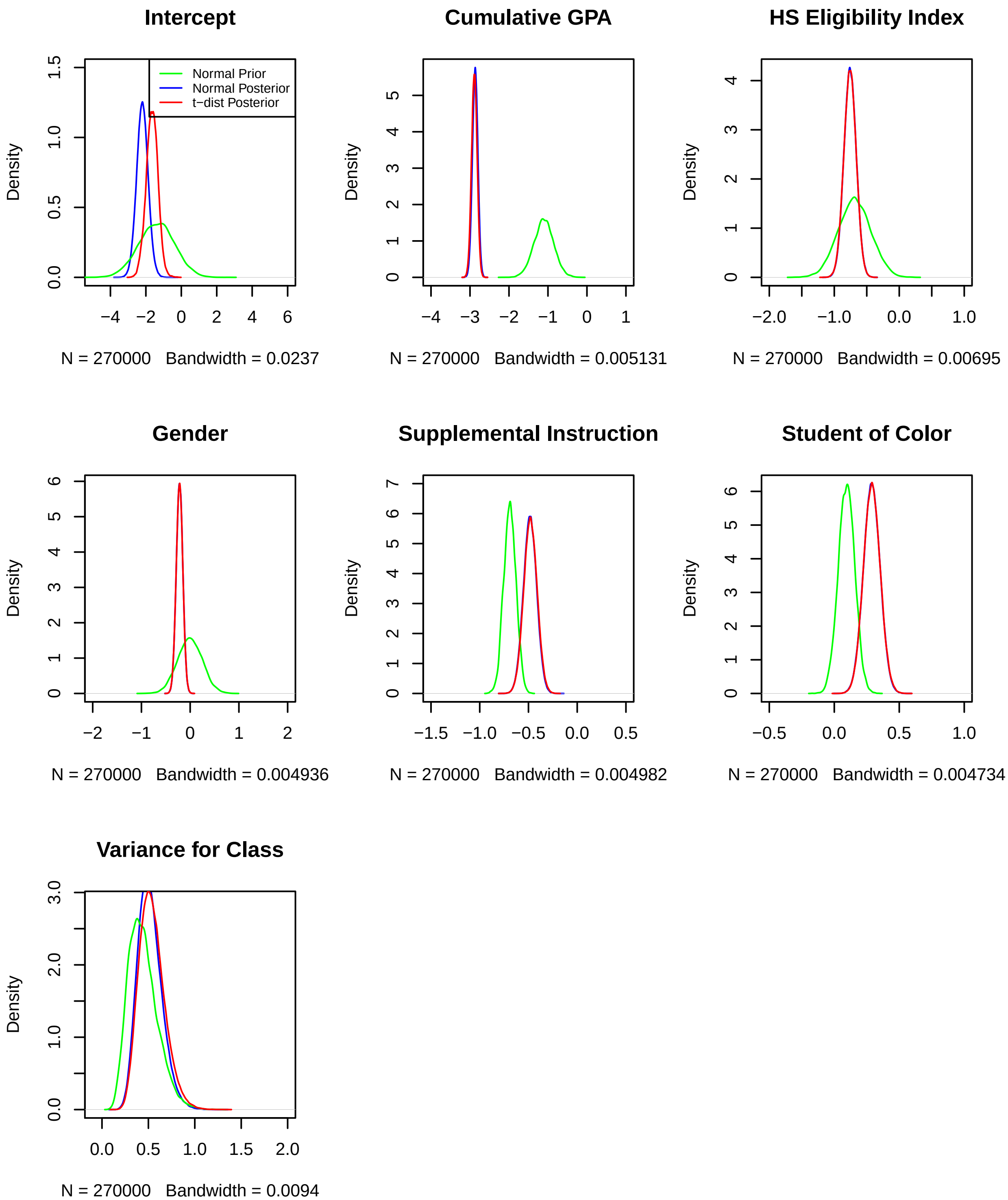

Figure 13: Autocorrelation Plots for Intercept Term of T Models at Varying Amounts of Iterations

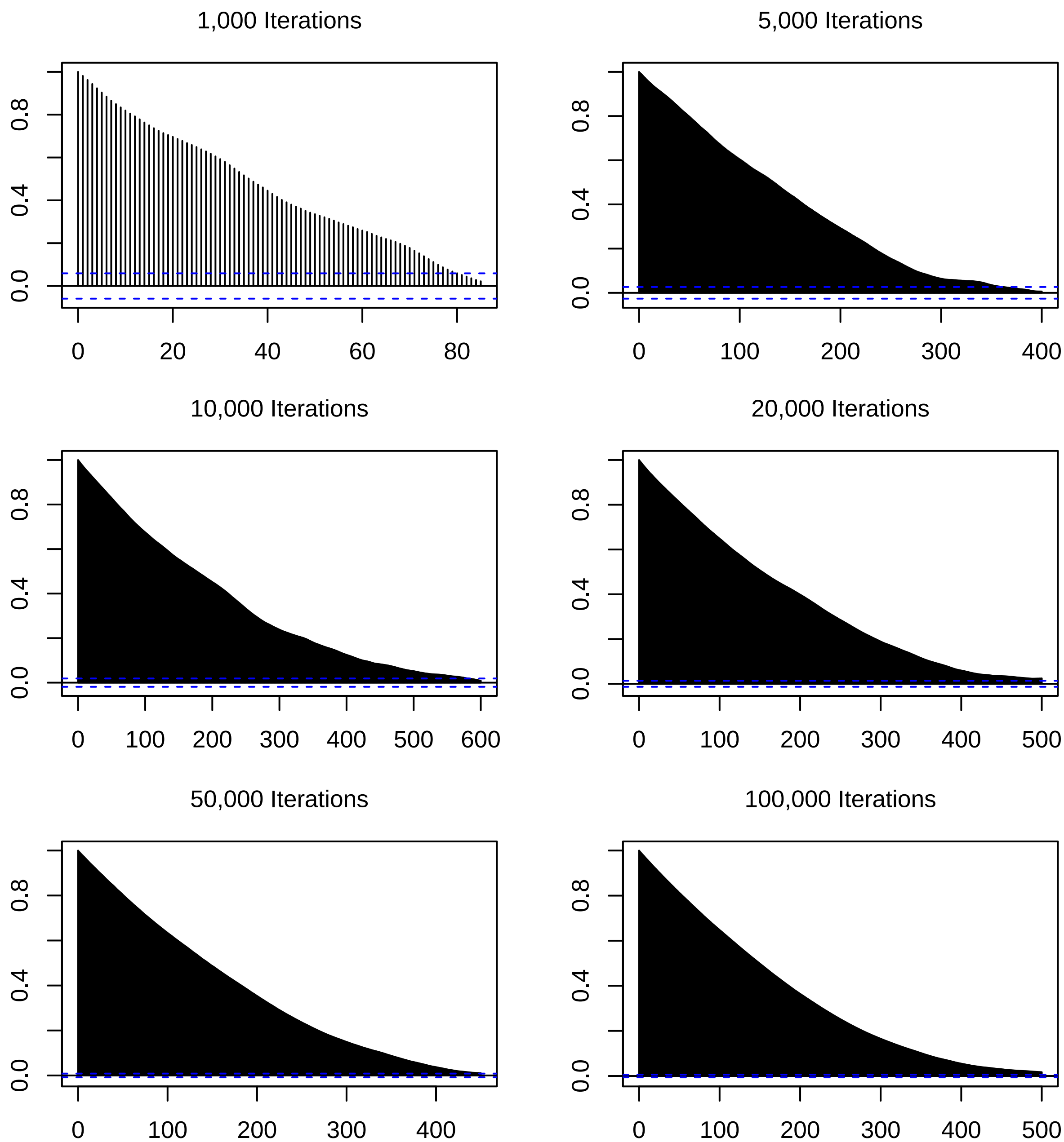

Figure 14: Time Series Plots for Intercept Term of T Models at Varying Amounts of Iterations