HW 9

Minci Zhang

2022-04-16

Research Question:

1. Would the pattern of gpa predicting changes in popularity over time differ for Hispanic vs Asian American students?
2. Would the pattern of gpa predicting changes in likability over time differ for Hispanic vs Asian American students?
3. Would the pattern of popularity predicting changes in gpa over time differ for Hispanic vs Asian American students?
4. Would the pattern of likability predicting changes in gpa over time differ for Hispanic vs Asian American students?

Variables:

log transformation has been performed on the original gpa1 popularity1 and likability1 values to fix the skewness of the distributions

- ‘ethnic1final’: ethnicity of the students: 1= Hispanic 2=Asian
- ‘loggpa1’: log transformed gpa at time one
- ‘logpop1’: log transformed value for how popular students are at time one
- ‘loglike1’: log transformed value for how likable students are at time one
- ‘gpa_d’: difference between gpa at time 1 and time 2(one year later)= (gpa2-gpa1)
- ‘pop_d’: difference between popularity at time 1 and 2= pop2- pop1
- ‘like_d’: difference between likability at time1 and 2 = like1-like2

Import Data

```
psyc573 <- read_sav("psyc573.sav")

# Recode `ethnic1final` to a factor variable
psyc573$ethnic1final <- psyc573$ethnic1final -1
psyc573$ethnic1final <- factor(psyc573$ethnic1final,
    levels = c(0, 1),
    labels = c("Hispanic", "Asian"))
```

Variable Summary

```
# descriptive statistics summary by ethnic groups
datasummary( (gpa_d + pop_d + like_d + logpop1 + loglike1 + loggpa1 )  *
                (N + Mean + SD + Min + Max + Histogram) ~
                factor(ethnic1final, labels = c("Hispanic", "Asian")),
            data = psyc573)
```

		Hispanic	Asian
gpa_d	N	117	218
	Mean	0.17	0.11
	SD	0.62	0.43
	Min	−1.80	−1.40
	Max	1.50	1.68
	Histogram		
pop_d	N	117	218
	Mean	−0.04	0.22
	SD	0.69	0.81
	Min	−2.64	−1.74
	Max	1.54	3.95
	Histogram		
like_d	N	117	218

	Mean	−0.10	0.25
	SD	0.88	1.02
	Min	−2.27	−2.63
	Max	1.90	3.85
	Histogram		
logpop1	N	117	218
	Mean	0.02	−0.11
	SD	0.94	0.96
	Min	−1.82	−1.59
	Max	2.03	2.47
	Histogram		
loglike1	N	117	218
	Mean	0.03	−0.01
	SD	0.94	1.00
	Min	−3.53	−2.95
	Max	1.74	1.90
	Histogram		
loggpa1	N	117	218
	Mean	1.18	1.47
	SD	0.26	0.16
	Min	0.34	0.69
	Max	1.61	1.61
	Histogram		

```
# correlation by ethnic groups
hisp<- psyc573 %>% filter(ethnic1final== 'Hispanic')
asian <- psyc573 %>% filter(ethnic1final == 'Asian')

hisp %>%
  select(gpa_d, pop_d, like_d, loggpa1, logpop1, loglike1) %>%
  datasummary_correlation(method="pearson")
```

	gpa_d	pop_d	like_d	loggpa1	logpop1	loglike1
gpa_d	1	.	.	.	.	.
pop_d	−.21	1	.	.	.	.
like_d	−.07	.30	1	.	.	.
loggpa1	−.46	.26	−.02	1	.	.
logpop1	.13	−.45	−.29	−.20	1	.
loglike1	−.08	−.11	−.57	.17	.53	1

```
asian %>%
  select(gpa_d, pop_d, like_d, loggpa1, logpop1, loglike1) %>%
  datasummary_correlation(method="pearson")
```

	gpa_d	pop_d	like_d	loggpa1	logpop1	loglike1
gpa_d	1	.	.	.	.	.
pop_d	−.17	1	.	.	.	.
like_d	−.09	.37	1	.	.	.
loggpa1	−.39	−.04	.05	1	.	.
logpop1	.02	−.14	−.12	−.03	1	.
loglike1	−.06	−.09	−.40	.10	.52	1

Model for gpa time 1 predicting popularity

change score :

Model

Let G = loggpa1, P = pop_d, E = ethnic1final

$$P_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \beta_0 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i \times E_i$$

Prior:

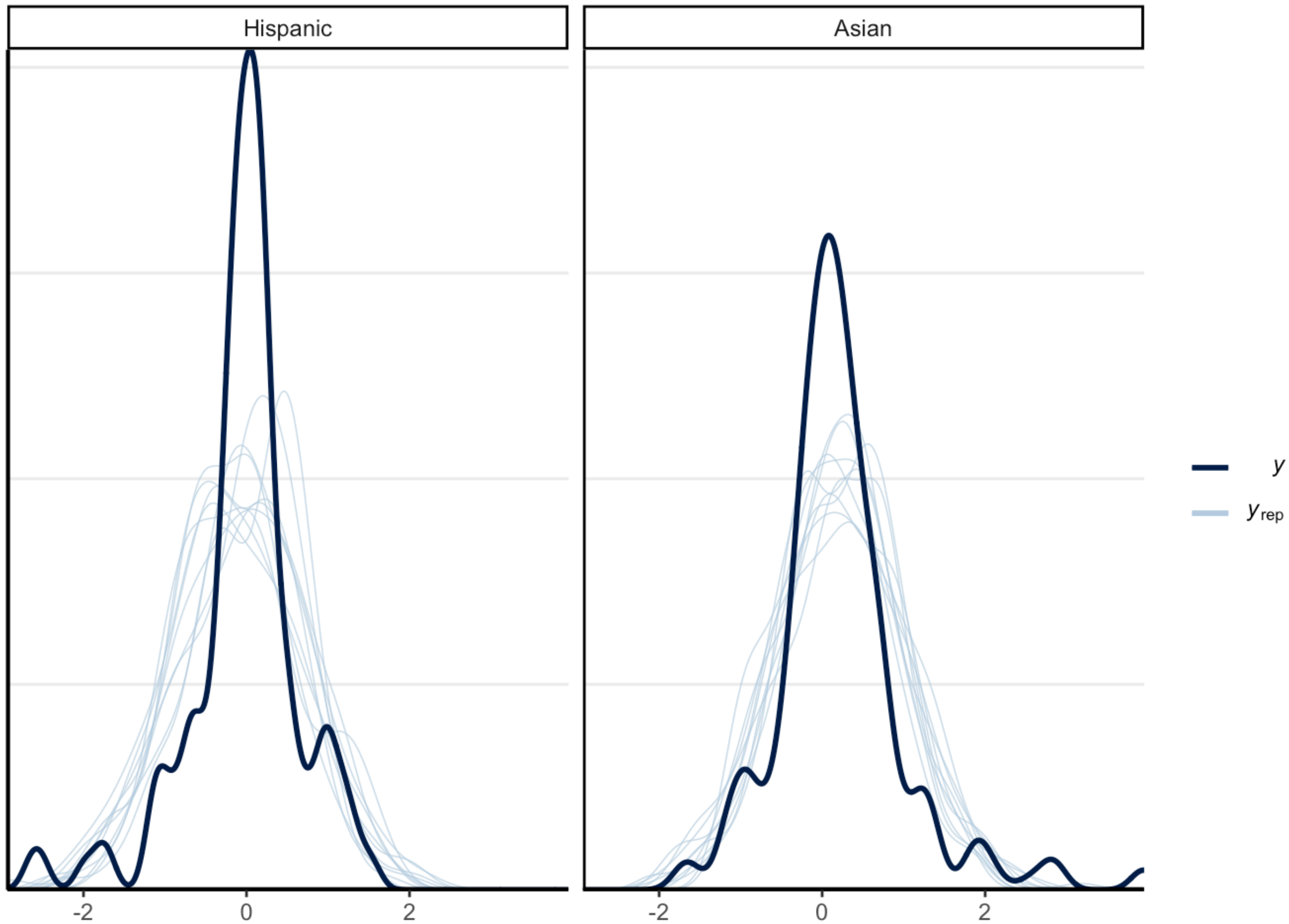
$$\beta_0 \sim N(0, 1)$$
$$\beta_1 \sim N(0.1, 0.05)$$
$$\beta_2 \sim N(0.1, 0.05)$$
$$\beta_3 \sim N(0.1, 0.05)$$
$$\sigma \sim t_3^+(0, 2.5)$$

```
# gpa1 predicts popularity change score
m1 <- brm(
    pop_d ~ loggpa1 * ethnic1final,
    data = psyc573,
    prior = prior(normal(0.1, 0.05), class = "b") +
        prior(normal(0.1, 0.05), class = "b", coef = "ethnic1finalAsian") +
        prior(normal(0, 1), class = "Intercept") +
        prior(student_t(3, 0, 2.5), class = "sigma"),
    seed = 940,
    iter = 4000
)
```

```
m1
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: pop_d ~ loggpa1 * ethnic1final
##    Data: psyc573 (Number of observations: 335)
##   Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup draws = 8000
##
## Population-Level Effects:
##                           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                    -0.17      0.08    -0.33    -0.00 1.00     9661
## loggpa1                       0.11      0.05     0.01     0.21 1.00     9524
## ethnic1finalAsian             0.10      0.05     0.01     0.19 1.00     9199
## loggpa1:ethnic1finalAsian     0.09      0.04     0.01     0.17 1.00     9696
##                           Tail_ESS
## Intercept                     6097
## loggpa1                       5926
## ethnic1finalAsian             5993
## loggpa1:ethnic1finalAsian     6135
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.77      0.03     0.72     0.84 1.00     8740     5688
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

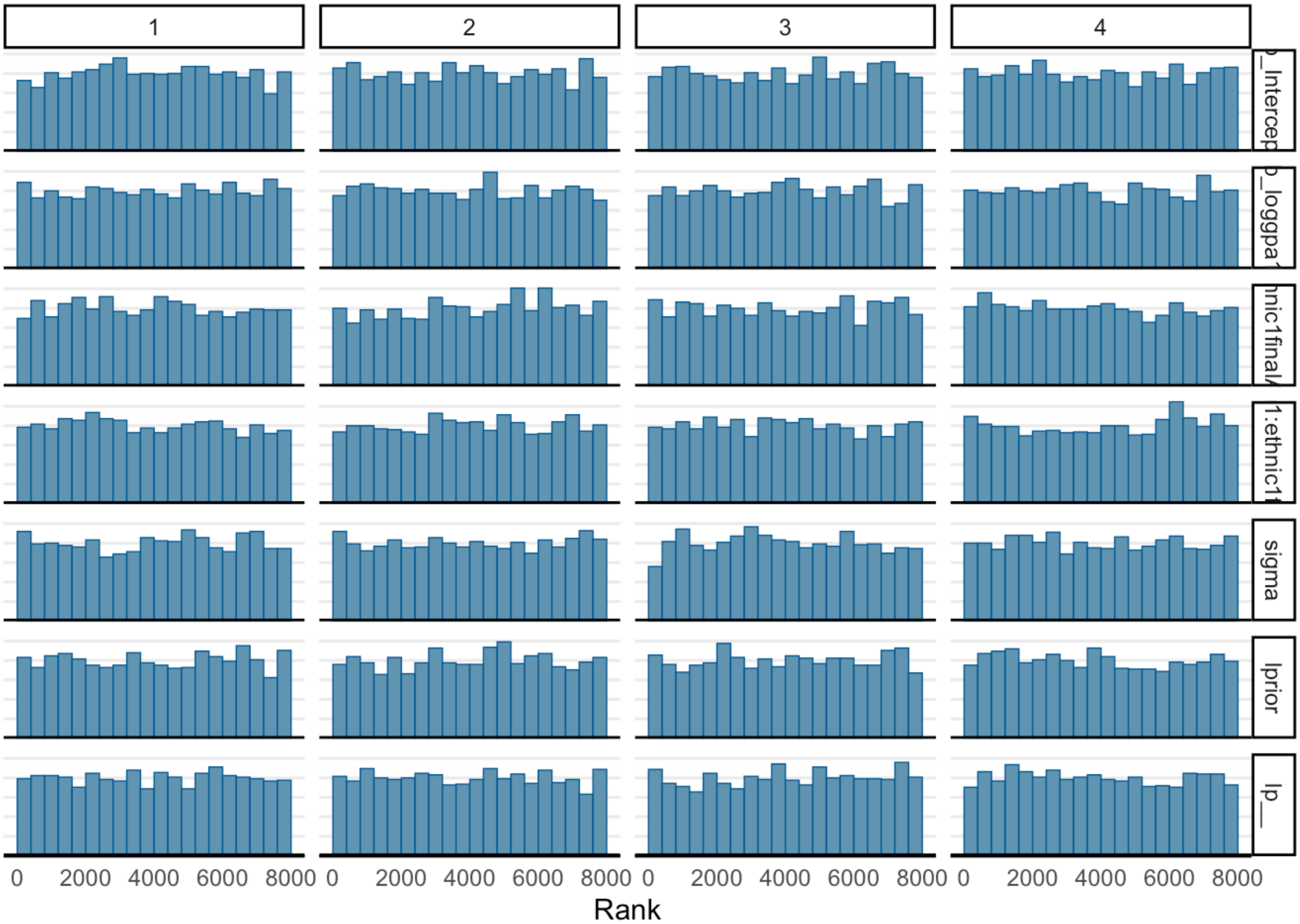
```
pp_check(m1, type = "dens_overlay_grouped", group = "ethnic1final")
```



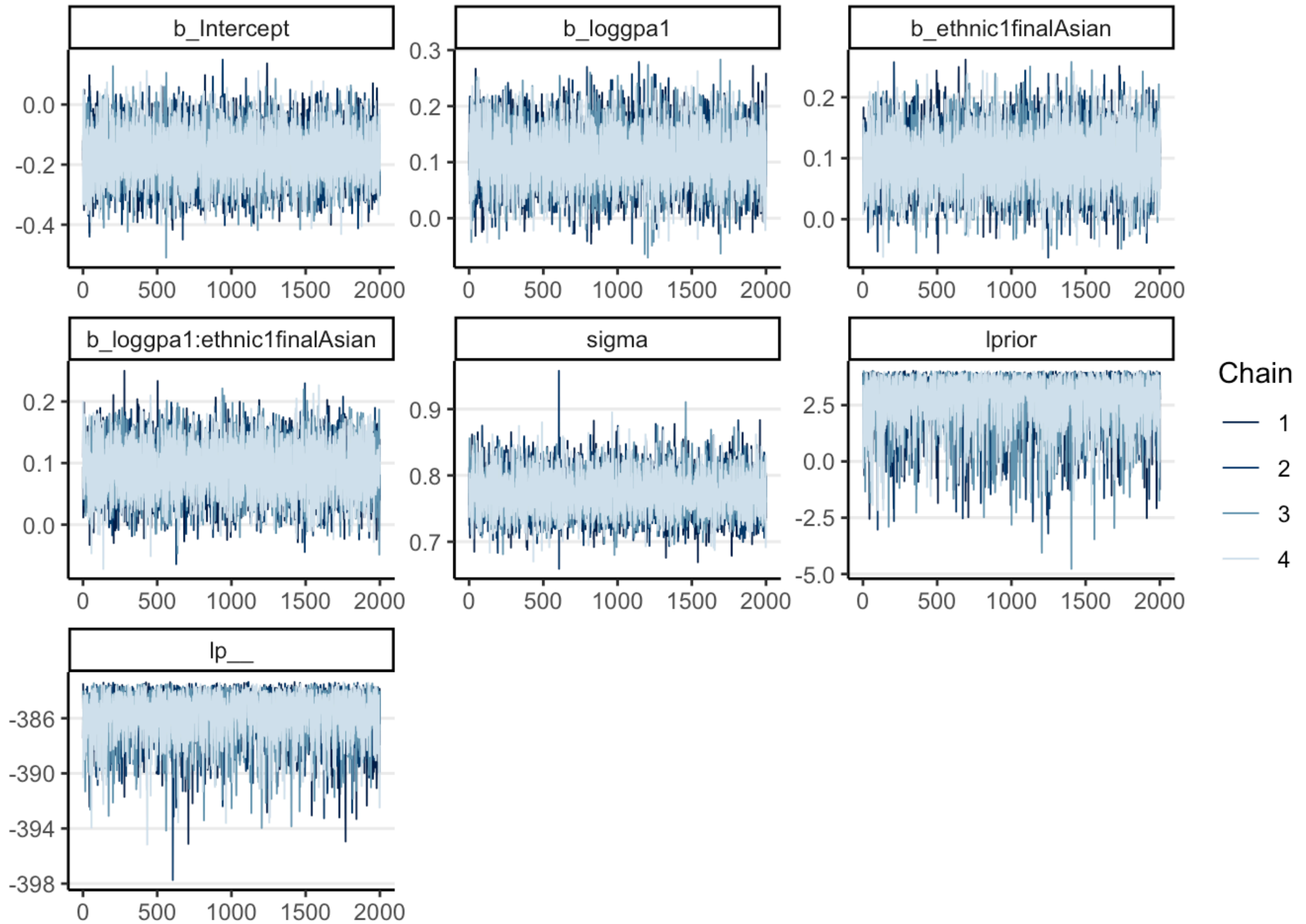
Results:

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m1)
```




```
mcmc_trace(as.array(m1))
```

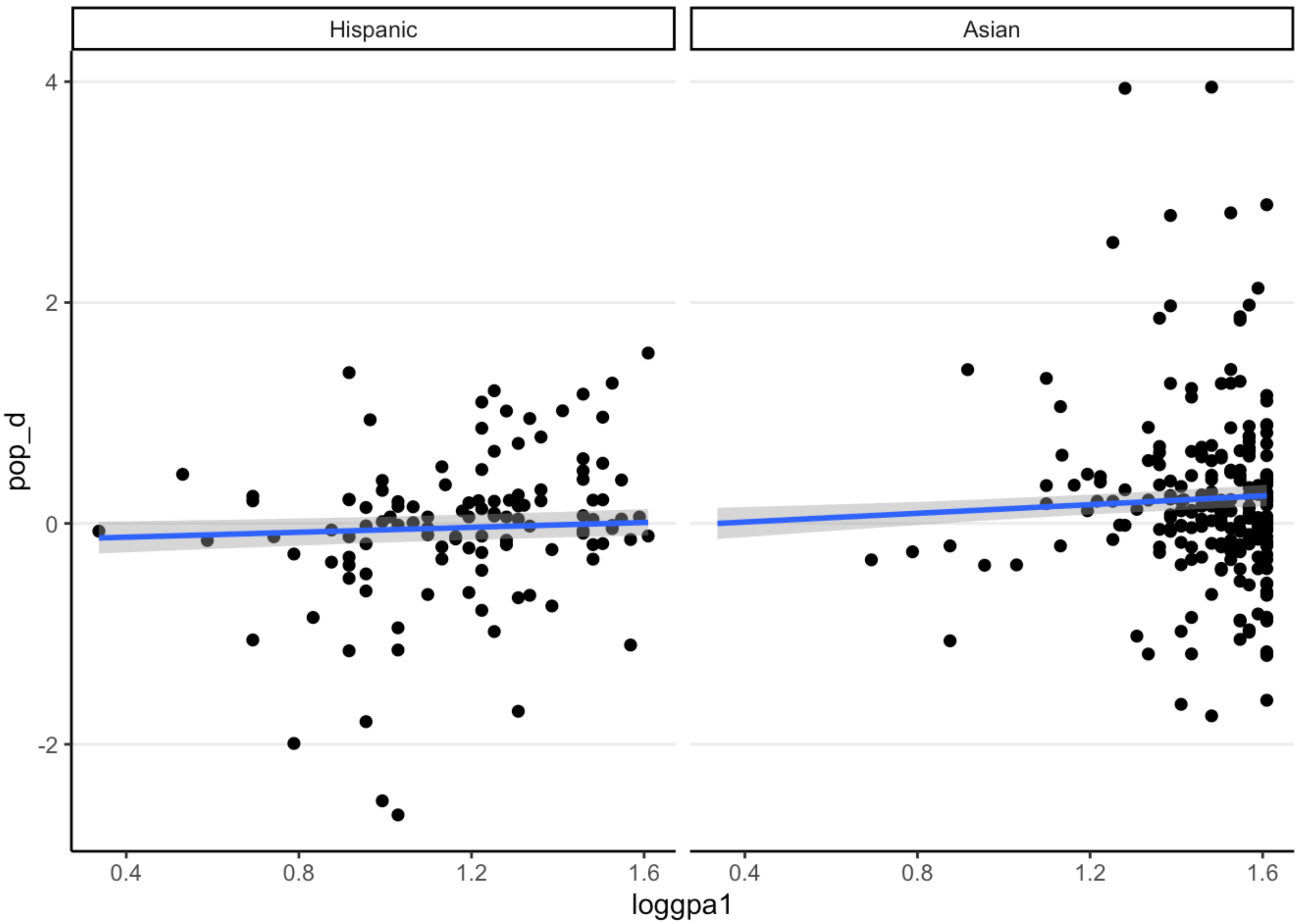


The following table and graph show the posterior distributions of b_hispanic and b_asian b_loggpa1:ethnic1finalAsian

```
as_draws(m1) %>%
    mutate_variables(
        b_hisp = b_loggpa1,
        b_asian = b_loggpa1 + `b_loggpa1:ethnic1finalAsian`
    ) %>%
    posterior::subset_draws(
        variable = c("b_hisp", "b_asian")
    ) %>%
    summarize_draws()
```

```
## # A tibble: 2 × 10
##   variable  mean median     sd    mad     q5   q95  rhat ess_bulk ess_tail
##   <chr>    <dbl>  <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>    <dbl>    <dbl>
## 1 b_hisp   0.110  0.110 0.0486 0.0490 0.0294 0.190  1.00    9524.    5926.
## 2 b_asian  0.199  0.199 0.0593 0.0605 0.101  0.297  1.00    9212.    6200.
```

```
plot(
    conditional_effects(m1,
        effects = "loggpa1",
        conditions = data.frame(ethnic1final = c("Hispanic", "Asian"),
                                cond__ = c("Hispanic", "Asian"))
    ),
    points = TRUE
)
```



Interpretation:

The analysis showed that on average, the patterns for gpa at time 1 predicting changes in levels of popularity differed for Asian and Hispanic American middle school students.

Specifically, gpa at time1 is more strongly associated with an increase in popularity from time 1 to time2 for Asian American students (posterior mean= 0.11, 90%CI [0.10, 0.30]) than for Hispanic American students (posterior mean = 0.11, 90% CI [0.03, 0.19]).

Model for gpa time1 predicting likability change score:

Models

Let G = loggpa1, L = like_d, E = ethnic1final

$$
\begin{aligned}
L_i &\sim N(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i \times E_i
\end{aligned}
$$

Prior:

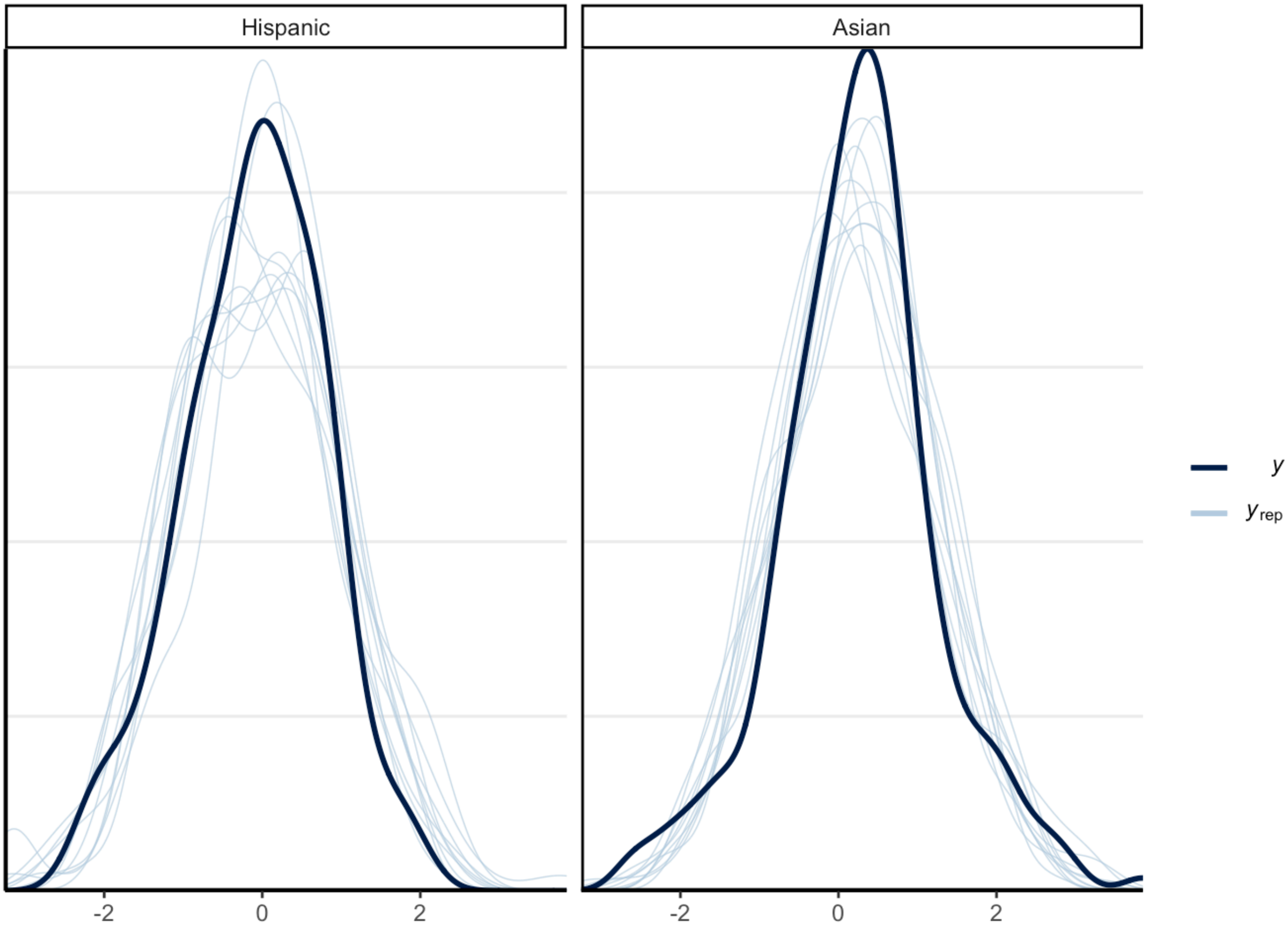
$$
\begin{aligned}
\beta_0 &\sim N(0, 1) \\
\beta_1 &\sim N(0.1, 0.05) \\
\beta_2 &\sim N(0.1, 0.05) \\
\beta_3 &\sim N(0.1, 0.05) \\
\sigma &\sim t_3^+(0, 2.5)
\end{aligned}
$$

```
# gpa1 predicts likability change score
m2 <- brm(
    like_d ~ loggpa1 * ethnic1final,
    data = psyc573,
    prior = prior(normal(0.1, 0.05), class = "b") +
        prior(normal(0.1, 0.05), class = "b", coef = "ethnic1finalAsian") +
        prior(normal(0, 1), class = "Intercept") +
        prior(student_t(3, 0, 2.5), class = "sigma"),
    seed = 941,
    iter = 4000
)
```

```
summary(m2)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: like_d ~ loggpa1 * ethnic1final
##    Data: psyc573 (Number of observations: 335)
##   Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup draws = 8000
##
## Population-Level Effects:
##                           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                    -0.19      0.09    -0.37    -0.00 1.00    10719
## loggpa1                       0.10      0.05     0.01     0.19 1.00    10002
## ethnic1finalAsian             0.11      0.05     0.02     0.20 1.00    10220
## loggpa1:ethnic1finalAsian     0.11      0.04     0.03     0.20 1.00     9301
##                           Tail_ESS
## Intercept                     6355
## loggpa1                       6334
## ethnic1finalAsian             6160
## loggpa1:ethnic1finalAsian     6574
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.97      0.04     0.90     1.05 1.00     9024     5927
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

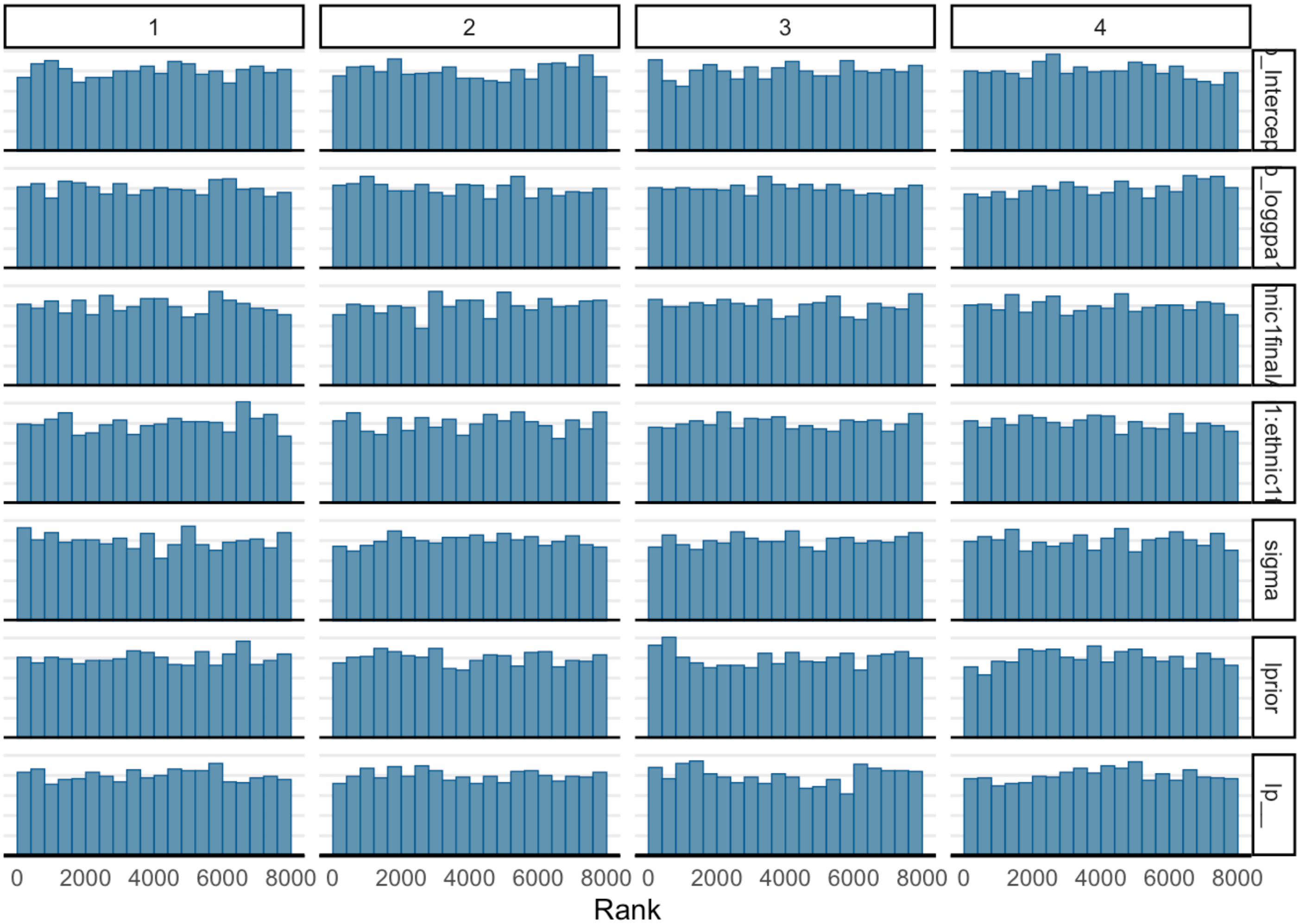
```
pp_check(m2, type = "dens_overlay_grouped", group = "ethnic1final")
```



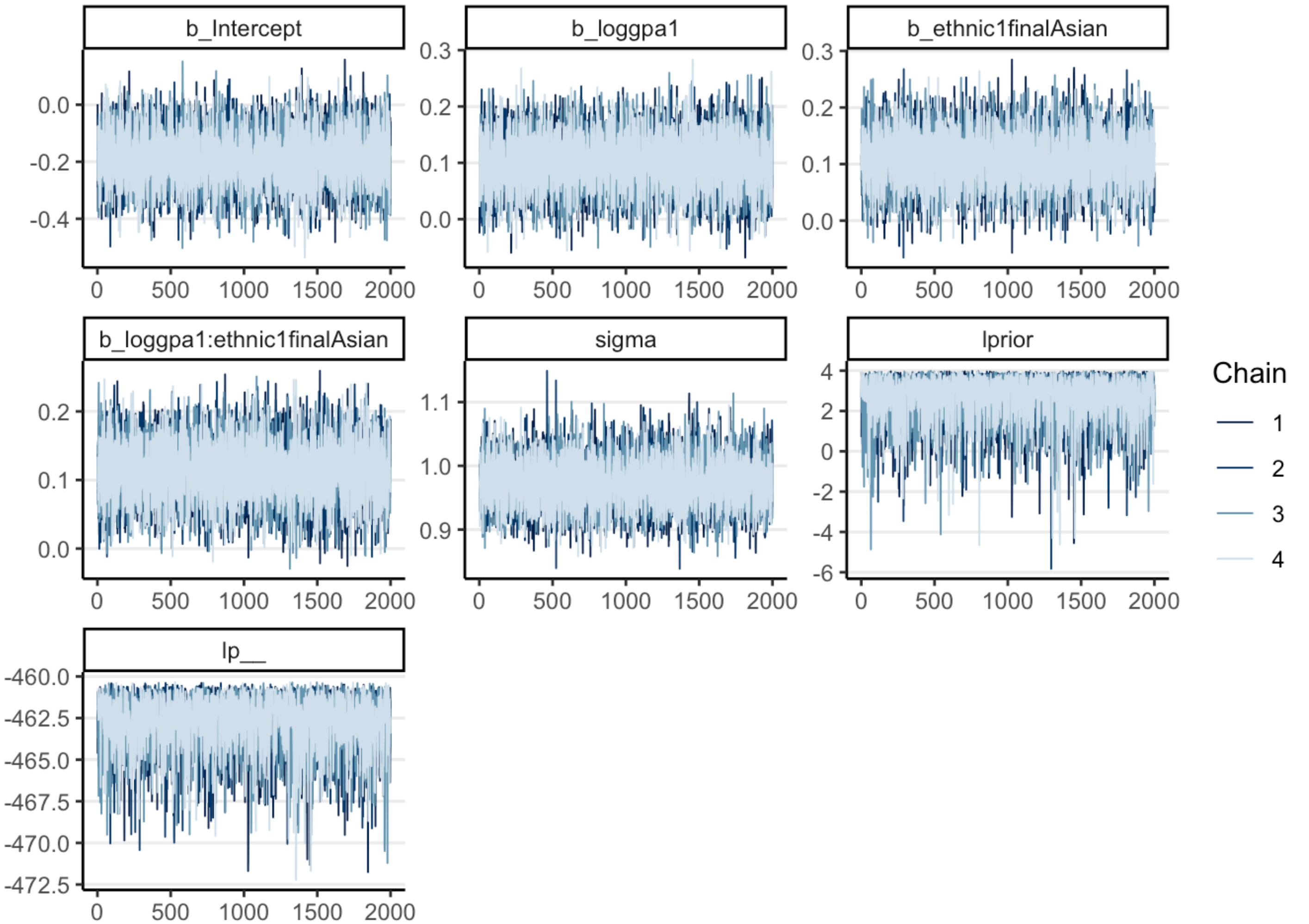
Results:

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m2)
```




```
mcmc_trace(as.array(m2))
```

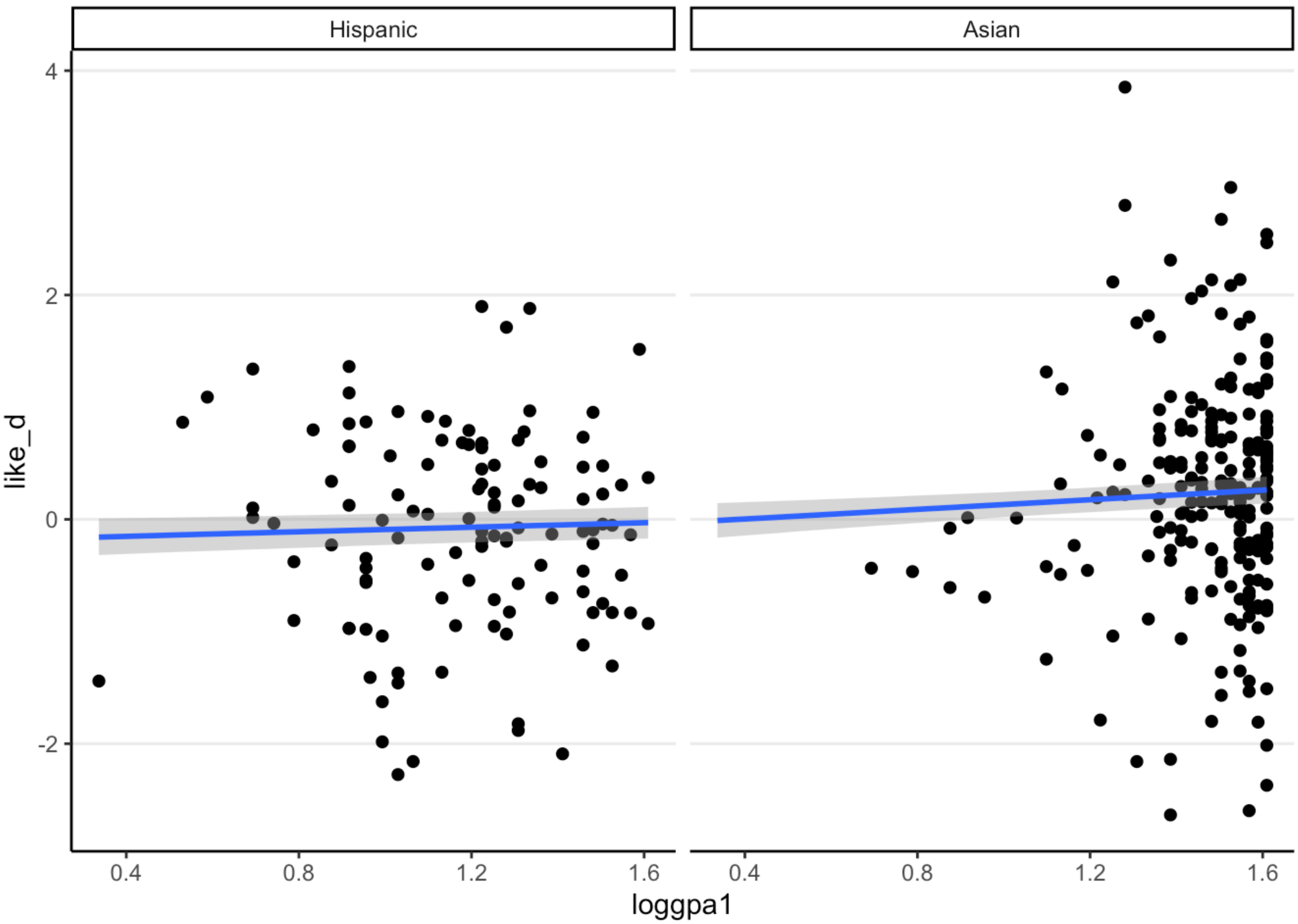


The following table and graph show the posterior distributions of b_hispanic and b_asian

```
as_draws(m2) %>%
    mutate_variables(
        b_hisp = b_loggpa1,
        b_asian = b_loggpa1 + `b_loggpa1:ethnic1finalAsian`
    ) %>%
    posterior::subset_draws(
        variable = c("b_hisp", "b_asian")
    ) %>%
    summarize_draws()
```

```
## # A tibble: 2 × 10
##   variable   mean median     sd    mad     q5   q95  rhat ess_bulk ess_tail
##   <chr>     <dbl>  <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>    <dbl>    <dbl>
## 1 b_hisp   0.0999  0.100 0.0488 0.0497 0.0184 0.179  1.00   10002.    6334.
## 2 b_asian  0.214   0.213 0.0625 0.0626 0.111  0.315  1.00   10008.    6406.
```

```
plot(
    conditional_effects(m2,
        effects = "loggpa1",
        conditions = data.frame(ethnic1final = c("Hispanic", "Asian"),
                                cond__ = c("Hispanic", "Asian"))
    ),
    points = TRUE
)
```



Interpretation:

The analysis showed that on average, the patterns for gpa at time 1 predicting changes in levels of likability differed for Asian and Hispanic American middle school students.

Specifically, gpa at time1 is more strongly associated with an increase in likability from time 1 to time2 for Asian American students (posterior mean= 0.21, 90%CI [0.11, 0.32]) than for Hispanic American students (posterior mean = 0.10, 90% CI [0.02 0.18]).

Model for pop1 predicting gpa change score :

Model:

Let G = gpa_d, P = logpop1, E = ethnic1final

$$
\begin{aligned}
G_i &\sim N(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 P_i + \beta_2 E_i + \beta_3 P_i \times E_i
\end{aligned}
$$

Prior:

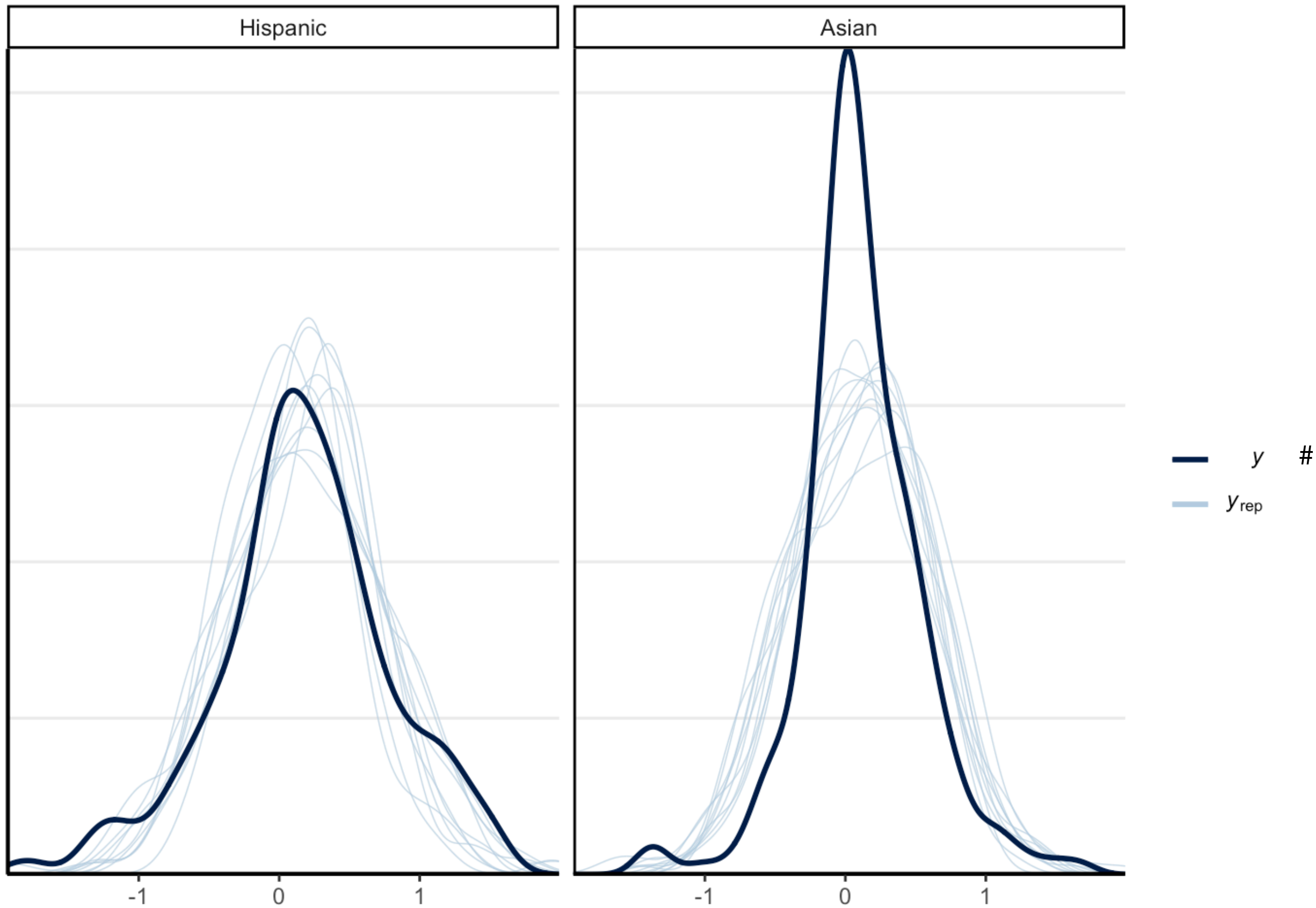
$$
\begin{aligned}
\beta_0 &\sim N(1, 0.5) \\
\beta_1 &\sim N(0.5, 0.5) \\
\beta_2 &\sim N(0.5, 0.5) \\
\beta_3 &\sim N(0.5, 0.5) \\
\sigma &\sim t_3^+(0, 2.5)
\end{aligned}
$$

```
# pop1 predicts gpa change score
m3 <- brm(
    gpa_d ~ logpop1 * ethnic1final,
    data = psyc573,
    prior = prior(normal(0.5, 0.5), class = "b") +
        prior(normal(0.5, 0.5), class = "b", coef = "ethnic1finalAsian") +
        prior(normal(1, 0.5), class = "Intercept") +
        prior(student_t(3, 0, 2.5), class = "sigma"),
    seed = 939,
    iter = 4000
)
```

```
summary(m3)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: gpa_d ~ logpop1 * ethnic1final
##    Data: psyc573 (Number of observations: 335)
##   Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup draws = 8000
##
## Population-Level Effects:
##                           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                     0.17      0.05     0.08     0.26 1.00     7367
## logpop1                       0.08      0.05    -0.02     0.18 1.00     4356
## ethnic1finalAsian            -0.05      0.06    -0.16     0.07 1.00     8470
## logpop1:ethnic1finalAsian    -0.07      0.06    -0.18     0.05 1.00     4296
##                           Tail_ESS
## Intercept                     5381
## logpop1                       5080
## ethnic1finalAsian             5385
## logpop1:ethnic1finalAsian     4785
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.50      0.02     0.47     0.54 1.00     8024     5871
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

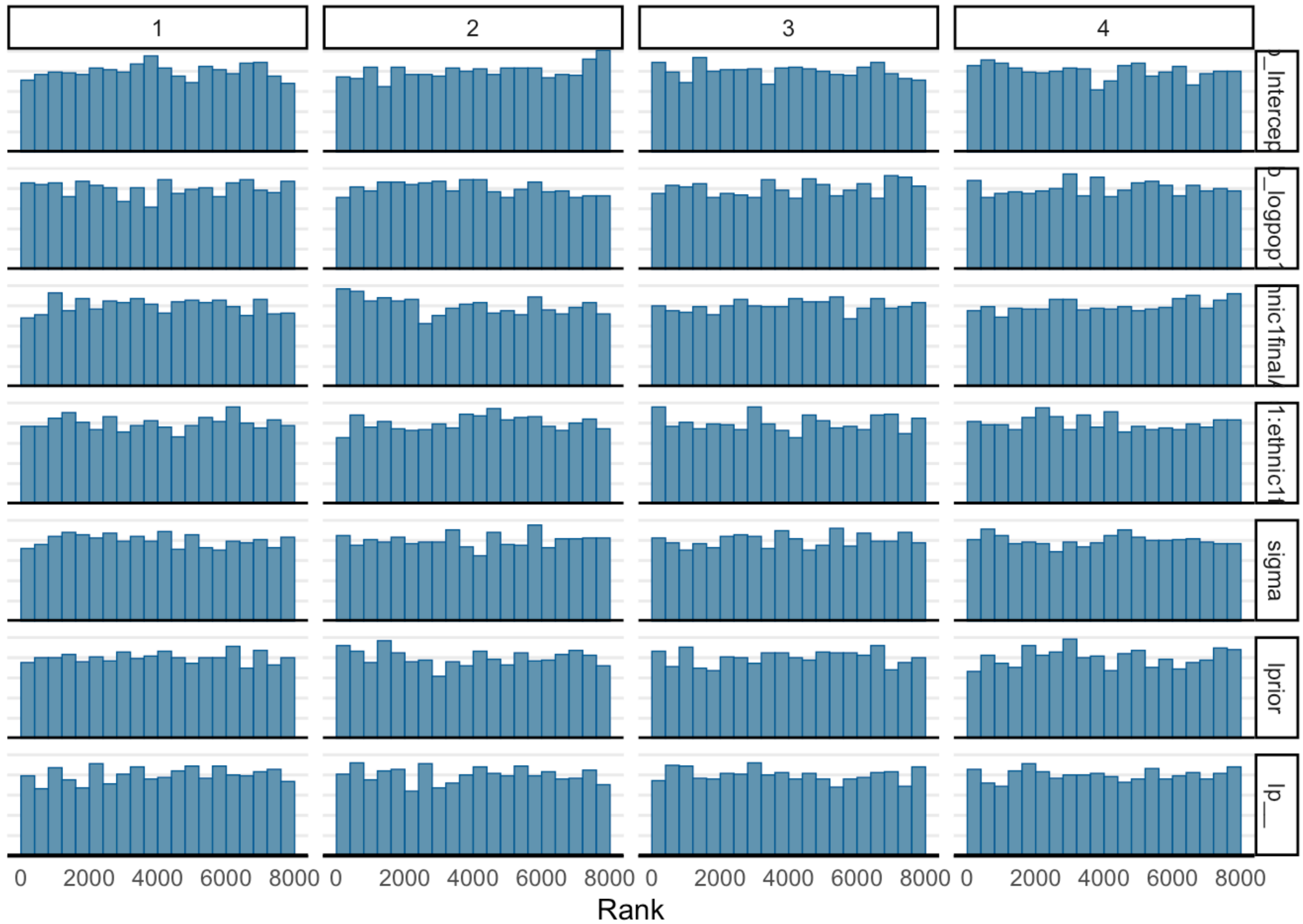
```
pp_check(m3, type = "dens_overlay_grouped", group = "ethnic1final")
```



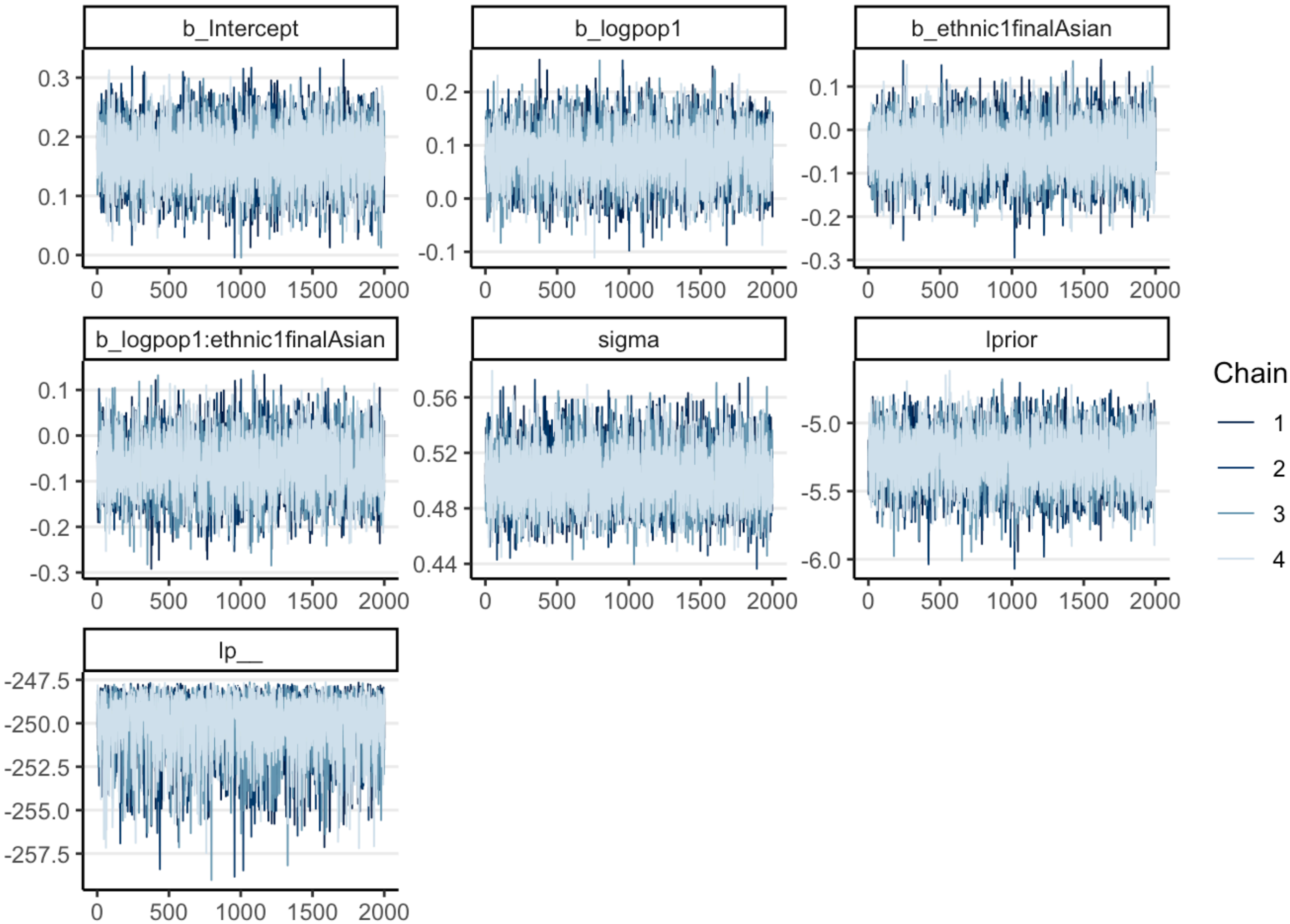
Results:

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m3)
```




```
mcmc_trace(as.array(m3))
```

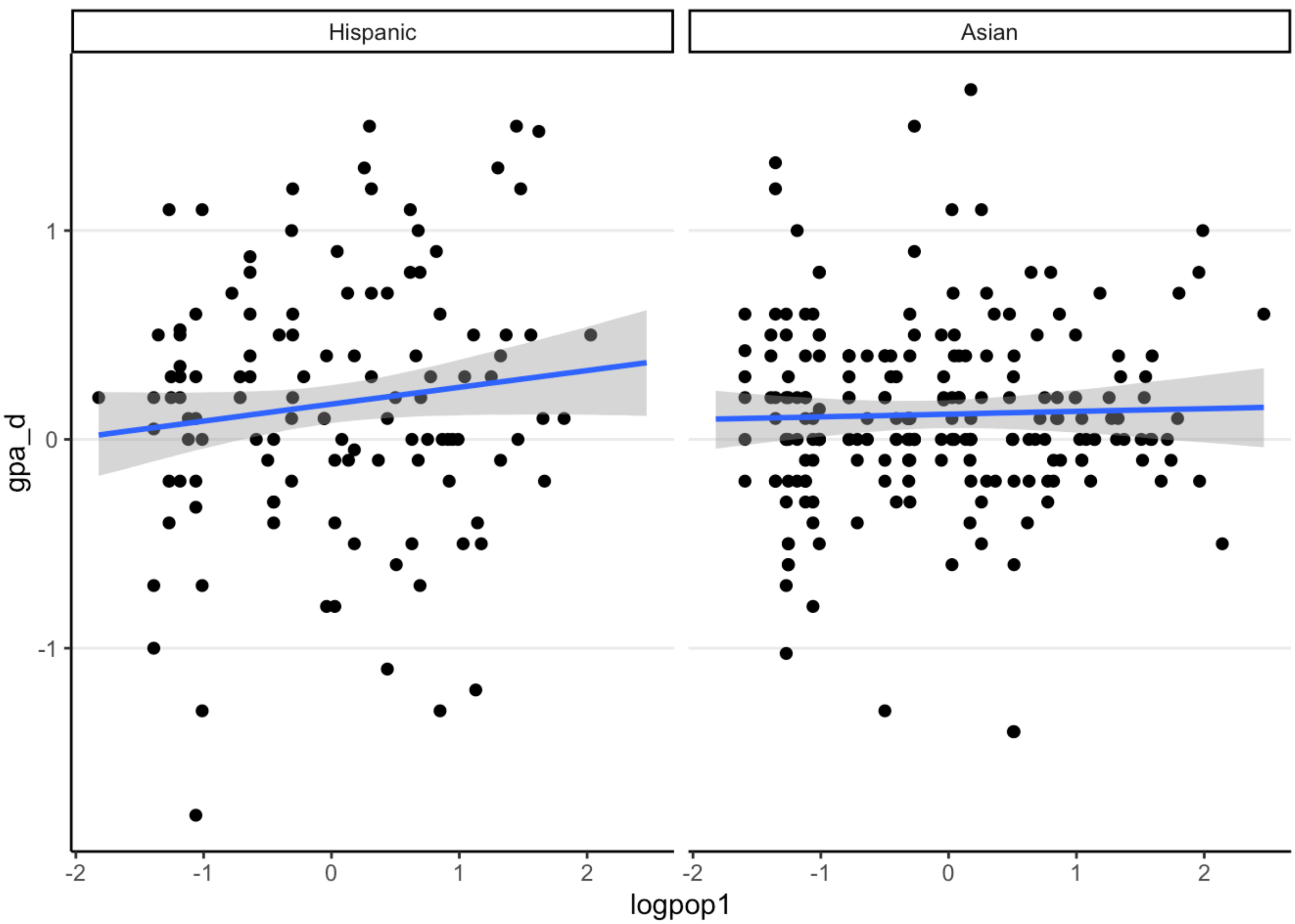


The following table and graph show the posterior distributions of b_hispanic and b_asian

```
as_draws(m3) %>%
    mutate_variables(
        b_hisp = b_logpop1,
        b_asian = b_logpop1 + `b_logpop1:ethnic1finalAsian`
    ) %>%
    posterior::subset_draws(
        variable = c("b_hisp", "b_asian")
    ) %>%
    summarize_draws()
```

```
## # A tibble: 2 × 10
##   variable   mean median     sd    mad        q5    q95  rhat ess_bulk ess_tail
##   <chr>     <dbl>  <dbl>  <dbl>  <dbl>     <dbl>  <dbl> <dbl>    <dbl>    <dbl>
## 1 b_hisp   0.0808 0.0809 0.0491 0.0488  0.0000931 0.161   1.00    4356.    5080.
## 2 b_asian  0.0132 0.0131 0.0355 0.0356 -0.0448    0.0724  1.00    9242.    6223.
```

```
plot(
    conditional_effects(m3,
        effects = "logpop1",
        conditions = data.frame(ethnic1final = c("Hispanic", "Asian"),
                                cond__ = c("Hispanic", "Asian"))
    ),
    points = TRUE
)
```



#

Interpretation:

The analysis indicate that the patterns for popularity at time 1 predicting increase in gpa from time1 and time2 did not differ for Hispanic (posterior mean= 0.08, 90 CI [0.00, 0.16]) and Asian (posterior mean= 0.01, 95% CI[-0.00, 0.07]) American students (β_3 (the beta for pop1 x gpa_d)=-0.017, 95% CI [-0.18 0.05])

Model for like1 predicting gpa change score :

Model

Let G = gpa_d, P = loglike1, E = ethnic1final

$$G_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \beta_0 + \beta_1 L_i + \beta_2 E_i + \beta_3 L_i \times E_i$$

Prior:

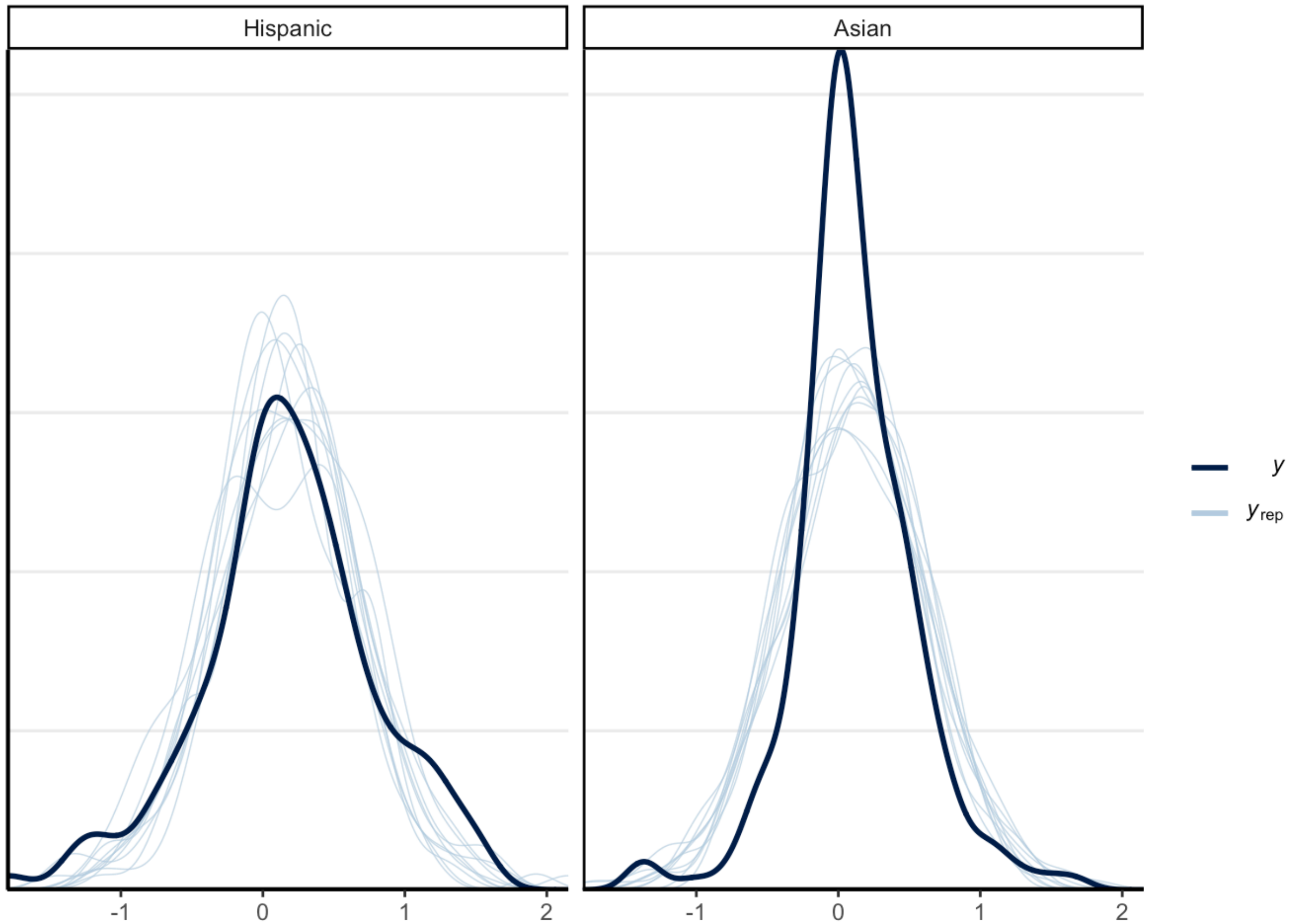
$$\beta_0 \sim N(1, 0.5)$$
$$\beta_1 \sim N(0.5, 0.5)$$
$$\beta_2 \sim N(0.5, 0.5)$$
$$\beta_3 \sim N(0.5, 0.5)$$
$$\sigma \sim t_4^+(0, 3)$$

```
# like1 predicts gpa change score
m4 <- brm(
    gpa_d ~ loglike1 * ethnic1final,
    data = psyc573,
    prior = prior(normal(0.5, 0.5), class = "b") +
        prior(normal(0.5, 0.5), class = "b", coef = "ethnic1finalAsian") +
        prior(normal(1, 0.5), class = "Intercept") +
        prior(student_t(3, 0, 2.5), class = "sigma"),
    seed = 938,
    iter = 4000
)
```

```
summary(m4)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: gpa_d ~ loglike1 * ethnic1final
##    Data: psyc573 (Number of observations: 335)
##   Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup draws = 8000
##
## Population-Level Effects:
##                             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                       0.17      0.05     0.08     0.26 1.00     8019
## loglike1                       -0.05      0.05    -0.15     0.04 1.00     4012
## ethnic1finalAsian              -0.05      0.06    -0.17     0.06 1.00     7944
## loglike1:ethnic1finalAsian      0.03      0.06    -0.09     0.15 1.00     3814
##                             Tail_ESS
## Intercept                       5221
## loglike1                        5048
## ethnic1finalAsian               5180
## loglike1:ethnic1finalAsian      5018
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.50      0.02     0.47     0.54 1.00     7934     5507
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

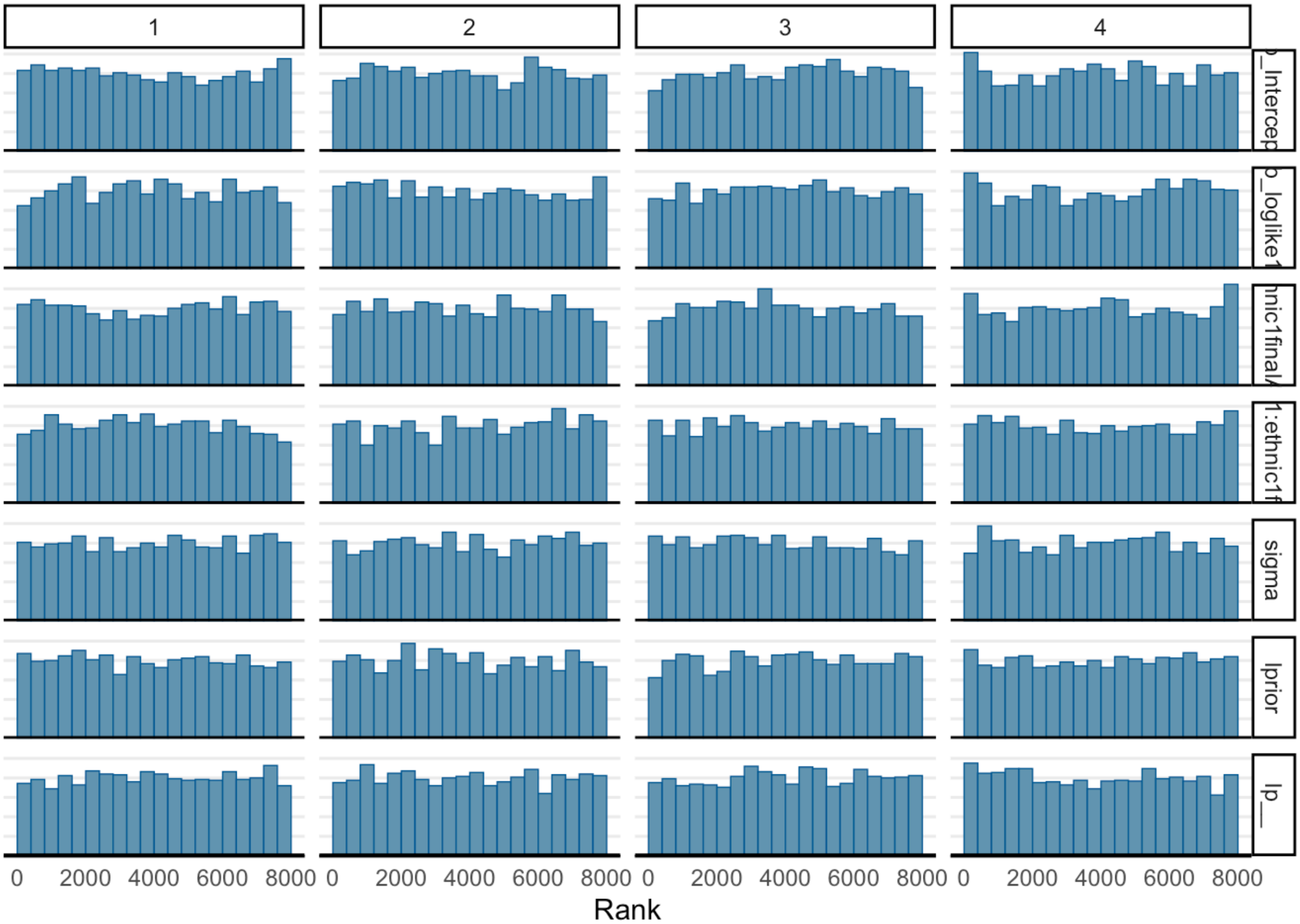
```
pp_check(m4, type = "dens_overlay_grouped", group = "ethnic1final")
```



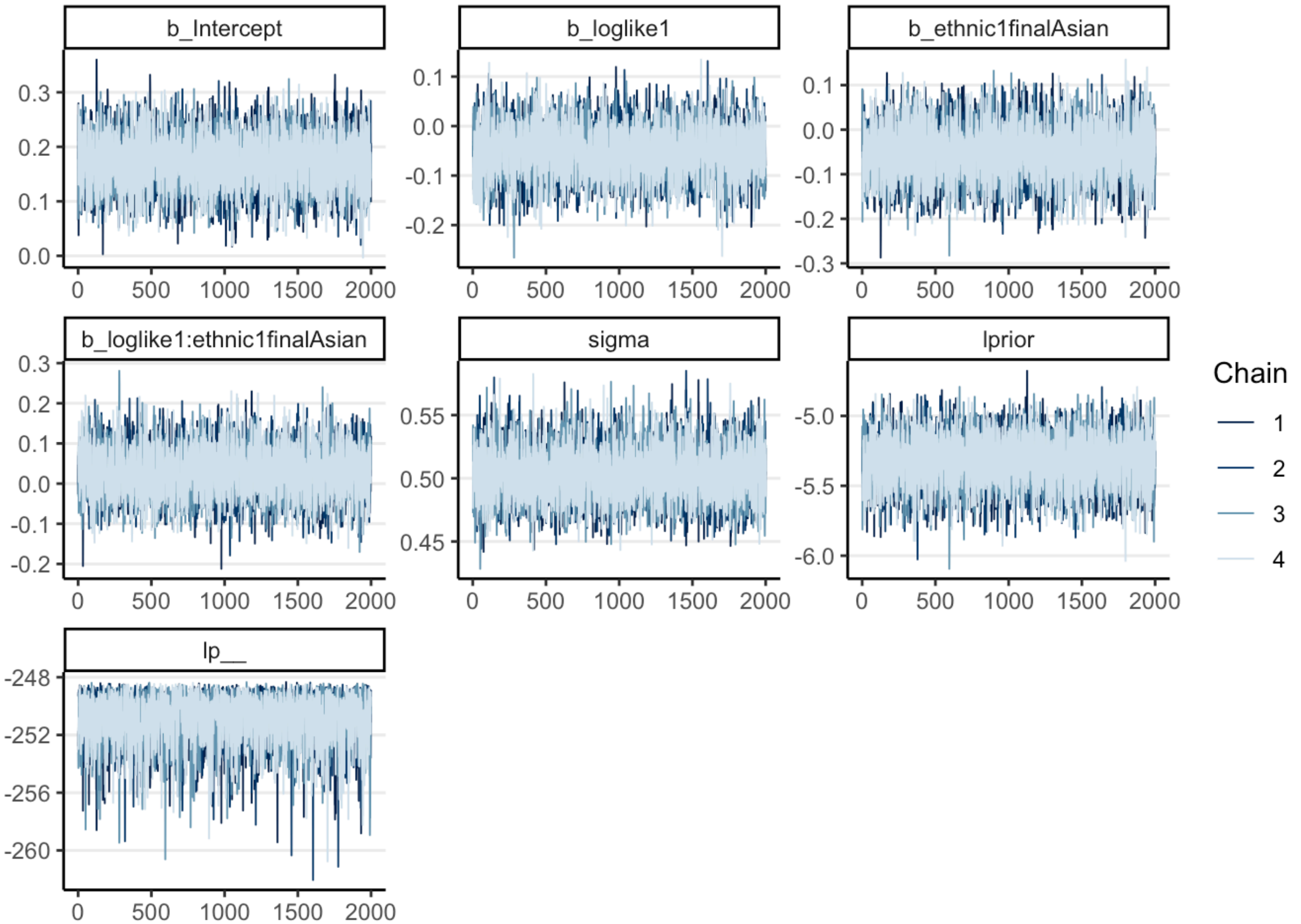
Results:

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m4)
```




```
mcmc_trace(as.array(m4))
```

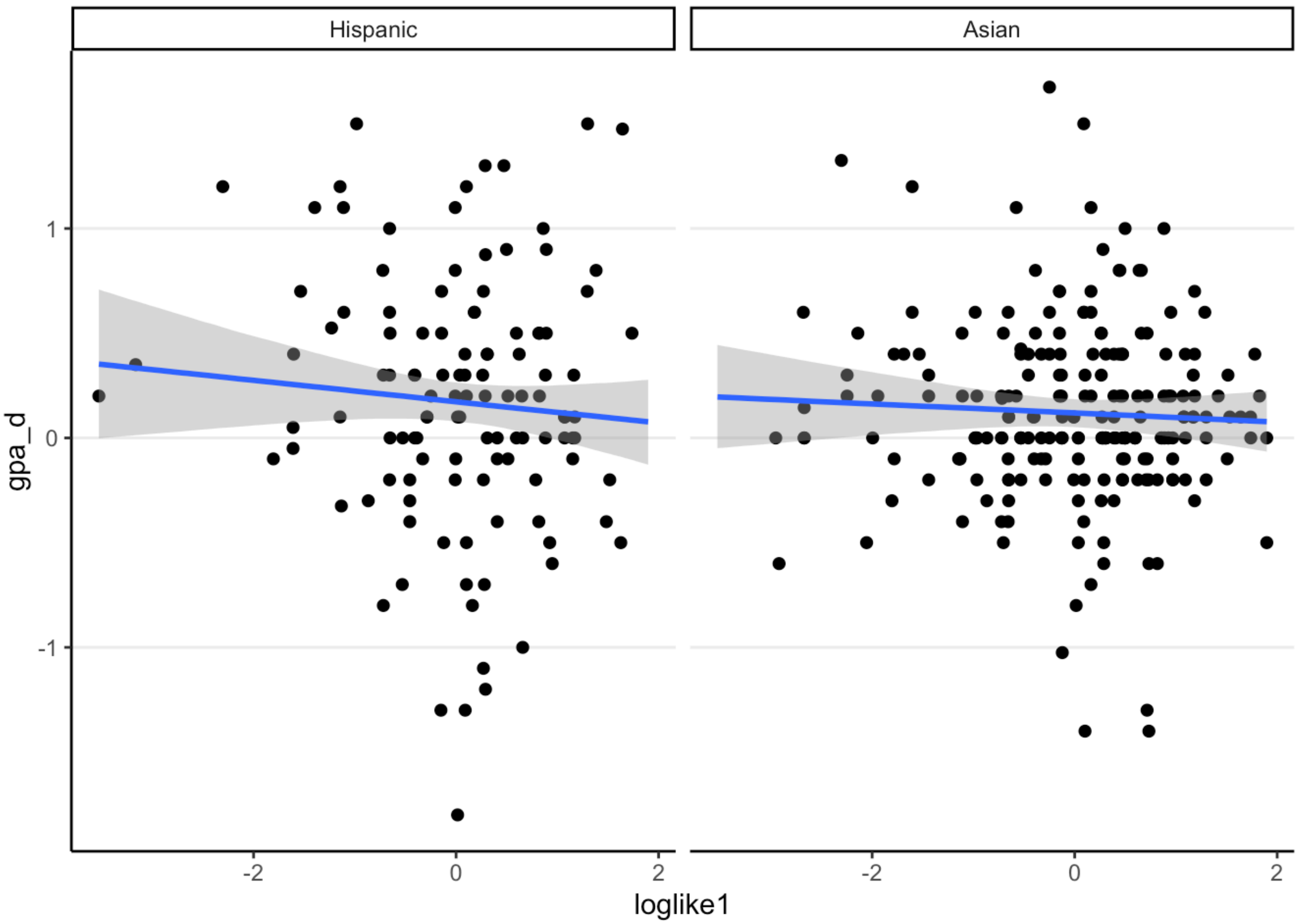


The following table and graph show the posterior distributions of b_hispanic and b_asian

```
as_draws(m4) %>%
    mutate_variables(
        b_hisp = b_loglike1,
        b_asian = b_loglike1 + `b_loglike1:ethnic1finalAsian`
    ) %>%
    posterior::subset_draws(
        variable = c("b_hisp", "b_asian")
    ) %>%
    summarize_draws()
```

```
## # A tibble: 2 × 10
##   variable    mean  median     sd    mad      q5    q95  rhat ess_bulk ess_tail
##   <chr>      <dbl>   <dbl>  <dbl>  <dbl>   <dbl>  <dbl> <dbl>    <dbl>    <dbl>
## 1 b_hisp   -0.0510 -0.0507 0.0491 0.0495 -0.133  0.0275  1.00    4012.    5048.
## 2 b_asian  -0.0220 -0.0217 0.0348 0.0354 -0.0788 0.0345  1.00    8768.    6339.
```

```
plot(
    conditional_effects(m4,
        effects = "loglike1",
        conditions = data.frame(ethnic1final = c("Hispanic", "Asian"),
                                cond__ = c("Hispanic", "Asian"))
    ),
    points = TRUE
)
```



Interpretation:

The analysis indicate that the patterns for likability at time 1 predicting increase in gpa from time1 and time2 did not differ for Hispanic (posterior mean= -0.05, 90 CI [-0.13, 0.03]) and Asian (posterior mean= -0.02, 95% CI[-0.08, 0.03]) American students (β_3 (the beta for like1 x gpa_d)=0.03, 95% CI [-0.09 0.15])