Final Project - Bayesian Analysis

```
library(readxl)
library(dplyr)
library(tidyr)
library(lubridate)
library(stringr)
```

```
load("beatspy.RData")
```

```
library(brms)
library(tidybayes)
library(bayesplot)
library(posterior)
```

Model 1 (Logistic)

$$Y_{i,t} \sim Bernoulli(p_{i,t})$$
$$logit(p_{i,t}) = \beta_0 + \beta_1 + log(PE_{i,t}) + \beta_2 DivYield_{i,t}$$

```
bayes_model1 <- brm(
  beat_spy ~ log_pe + div_yield,
  data = m3_df,
  family = bernoulli(link = "logit"),
  seed = 123
)
```

```
## Compiling Stan program...
```

```
## Trying to compile a simple C file
```

```
## Start sampling
```

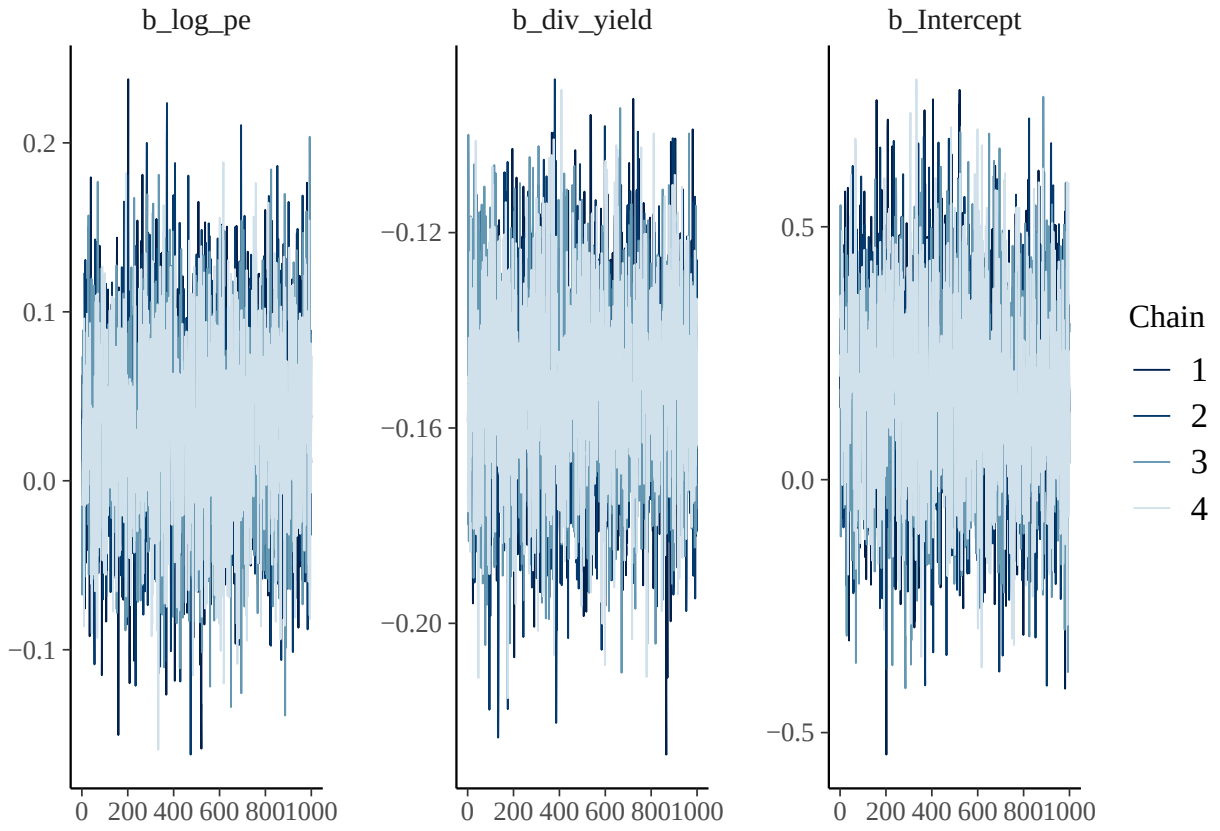
```
summary(bayes_model1)$fixed
```

```
##              Estimate  Est.Error   l-95% CI   u-95% CI     Rhat Bulk_ESS
## Intercept  0.17273278 0.18791917 -0.1830719  0.5376506 1.001122 3007.931
## log_pe     0.03149262 0.05432937 -0.0760189  0.1343352 1.001685 3368.702
## div_yield -0.15189077 0.01951898 -0.1906835 -0.1138507 1.000259 2819.099
##           Tail_ESS
## Intercept 2906.451
## log_pe    3373.531
## div_yield 2796.484
```

Rhats are ~1 and effective sample sizes » 100

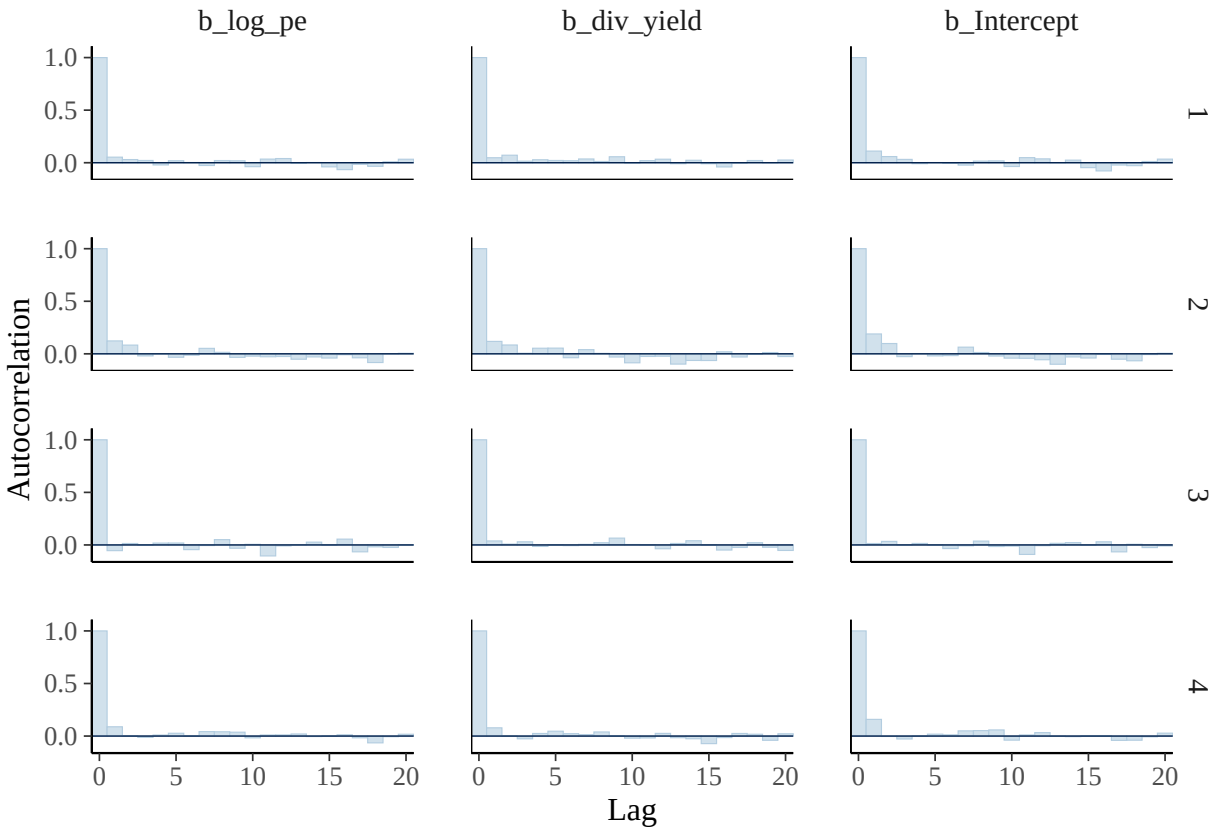
```
draws_model1 = bayes_model1 |>
  as_draws_array()

mcmc_trace(draws_model1,
           pars = c("b_log_pe", "b_div_yield", "b_Intercept"))
```



No discernable pattern from trace plots

```
mcmc_acf_bar(
  draws_model1,
  pars = c("b_log_pe", "b_div_yield", "b_Intercept")
)
```



acfs fall off quickly

Model 2 (Hierarchical)

(including both sector and firm-level random intercepts)

$$logit(p_{i,t}) = \beta_0 + \beta_1 log(PE_{i,t}) + \beta_2 DivYield_{i,t} + u_j + v_i$$

```
bayes_model2 <- brm(
  beat_spy ~ log_pe + div_yield + (1 | gics_sector_name/Ticker),
  data = m3_df,
  family = bernoulli(link = "logit"),
  seed = 123
)
```

```
## Compiling Stan program...

## Trying to compile a simple C file

## Start sampling
```

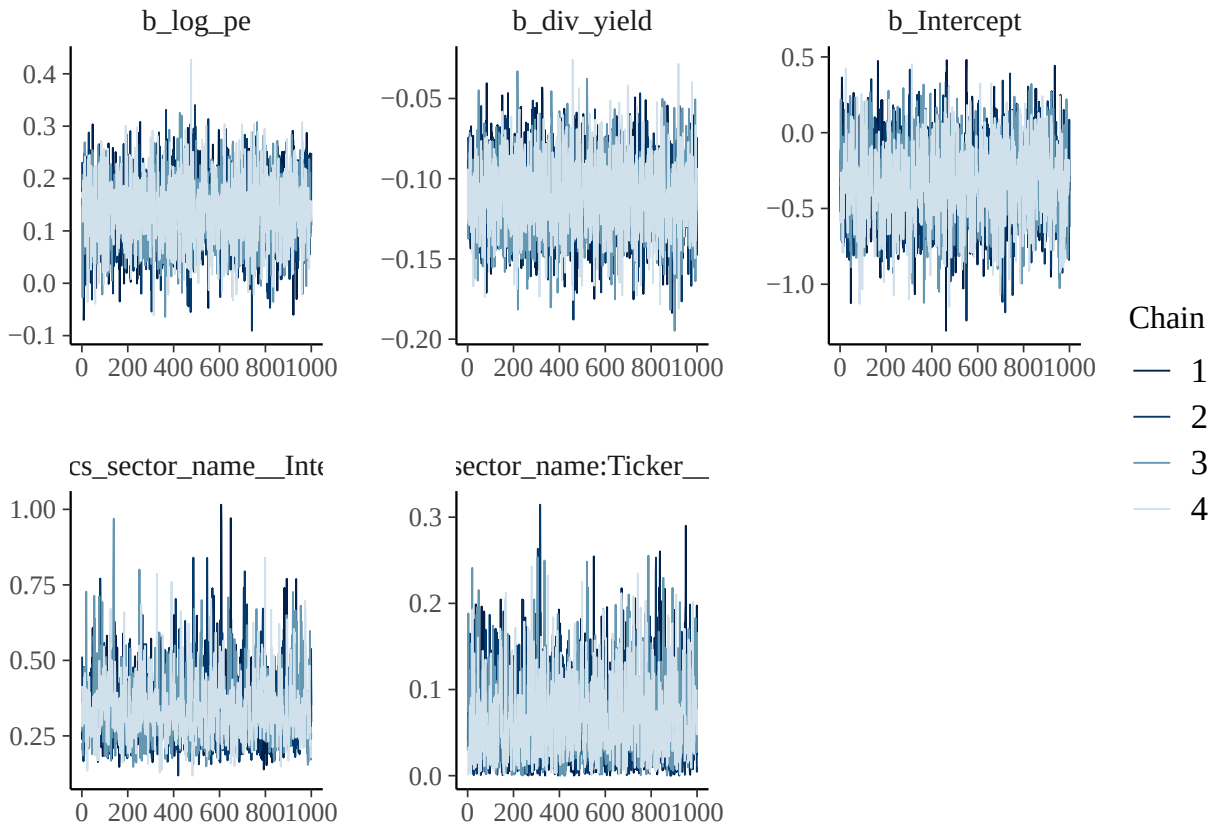
```
summary(bayes_model2)
```

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: beat_spy ~ log_pe + div_yield + (1 | gics_sector_name/Ticker)
##    Data: m3_df (Number of observations: 3400)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Multilevel Hyperparameters:
## ~gics_sector_name (Number of levels: 11)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.34      0.10     0.19     0.59 1.00     1348     2025
##
## ~gics_sector_name:Ticker (Number of levels: 406)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.06      0.05     0.00     0.17 1.00     1607     1590
##
## Regression Coefficients:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    -0.32      0.26    -0.82     0.17 1.00     2110     2593
## log_pe        0.14      0.06     0.02     0.26 1.00     3616     2987
## div_yield    -0.11      0.02    -0.15    -0.07 1.00     3638     2948
##
## 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).
```

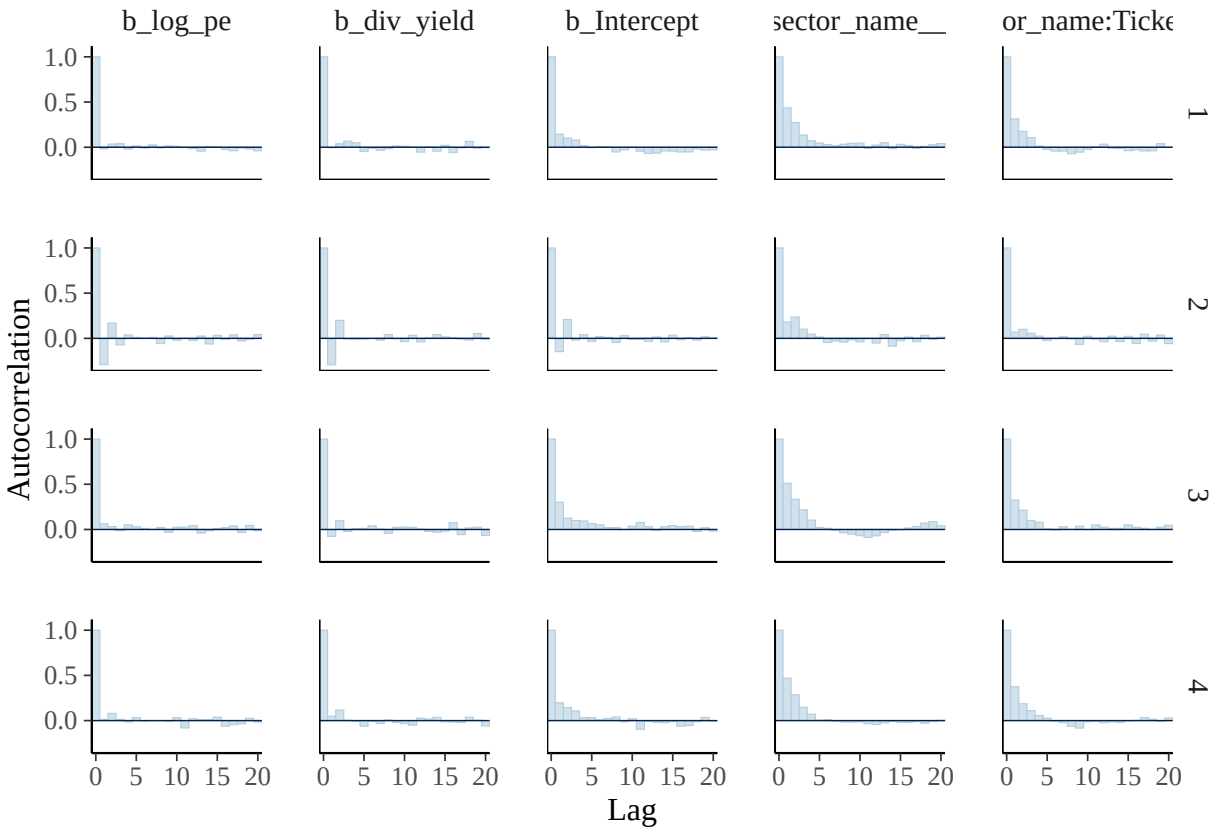
Rhats are all ~1, effective sample sizes » 100

```
draws_model2 = bayes_model2 |>
  as_draws_array()

mcmc_trace(
  draws_model2,
  pars = c("b_log_pe",
           "b_div_yield",
           "b_Intercept",
           "sd_gics_sector_name__Intercept",
           "sd_gics_sector_name:Ticker__Intercept")
)
```




```
mcmc_acf_bar(
  draws_model2,
  pars = c("b_log_pe", "b_div_yield", "b_Intercept", "sd_gics_sector_name__Intercept", "sd_gics_sector_
)
```



acfs fall off quickly

Model Comparison

```
loo_compare(loo(bayes_model1),
            loo(bayes_model2))
```

```
##              elpd_diff se_diff
## bayes_model2   0.0       0.0
## bayes_model1 -22.2       7.0
```

The difference in expected log predictive density is 22.2 (SE = 7.0). LOOCV provides strong evidence that the multilevel specification improves out-of-sample predictive performance.