Do Firms Truly Outperform the Market? A Bayesian Perspective

Daniel Fu, Jack Wong

Winter 2026

1 Introduction

The primary question of interest is whether individual firms outperform the market over time. To this end, we use the SPDR S&P 500 ETF Trust (SPY) as a benchmark and define outperformance at the annual horizon. Specifically, we study (i) the probability that a firm's annual return exceeds SPY and (ii) whether differences across stocks, specifically in outperformance can be explained by firm characteristics such as valuation and income measures using price-to-earnings (P/E) ratio and dividend yield as proxies.

This distinction is hard to make in practice since returns are noisy, especially over short horizons (we only have access to 10 years worth of data, and only 9 full calendar years), so even large wins can be driven by luck rather than persistent skill. Focusing on annual returns helps smooth out some of this volatility and gives a more economically meaningful measure of performance. At the same time, valuation and dividend measures provide predictors that may help explain why some firms appear to beat the market more often than others.

Bayesian methods are especially useful here because the data are limited at the firm level: each firm only has about ten annual observations. By treating a firm's probability of beating the market as unknown and combining the data with reasonable prior beliefs, Bayesian inference allows us to express uncertainty directly and reduces the risk of over-interpreting extreme outcomes that may simply be noise.

Table 1: Data dictionary

Variable	Description
`ticker`	Unique trading symbol used to identify the firm's equity security
`name`	Legal name of the firm
`gics_sector_name`	Global Industry Classification Standard (GICS) sector of the firm, of which there are 11
`mk cap`	Market capitalization of the firm
`date`	Observation date for the reported variables
`PX_LAST`	Last traded (closing) price of the firm's equity on the observation date
`EQY_DVD_YLD_IND`	Equity dividend yield (dividends per share divided by current share price) measured at date
`PE_RATIO`	Price-to-earnings ratio (share price divided by earnings per share) measured at observation date

2 Exploratory Data Analysis

In Figure 1 we derive the annual stock returns for firm i by calculating each weekly growth rates r_{it}:

$$r_{it} = \frac{PX\_LAST_{it}}{PX\_LAST_{i,t-1}}$$

... and compounding over all weeks in the year.

$$a_i = \prod_{t \in year} r_{it} - 1$$

Figure 1: Distribution of Annual Stock Returns

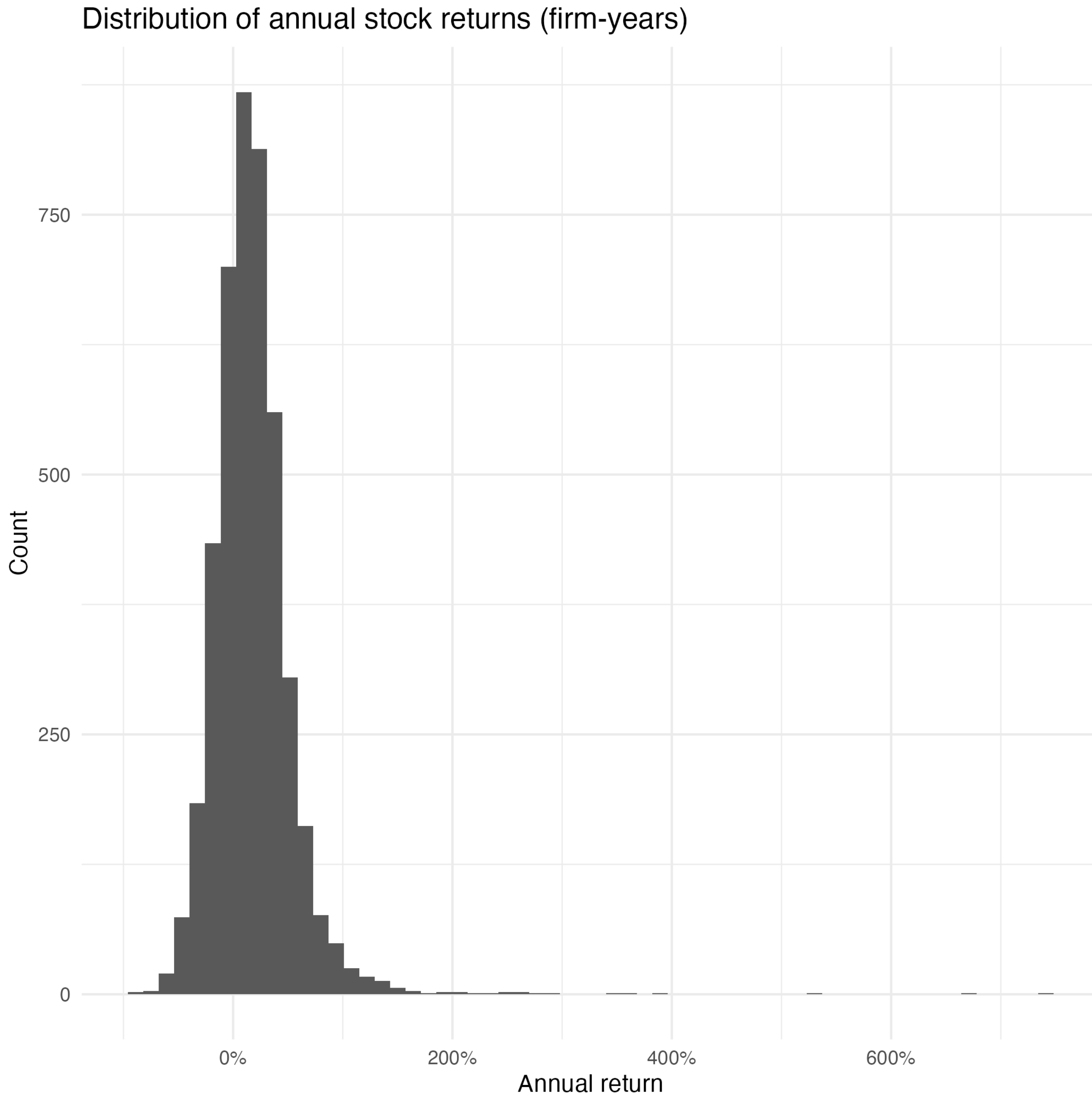
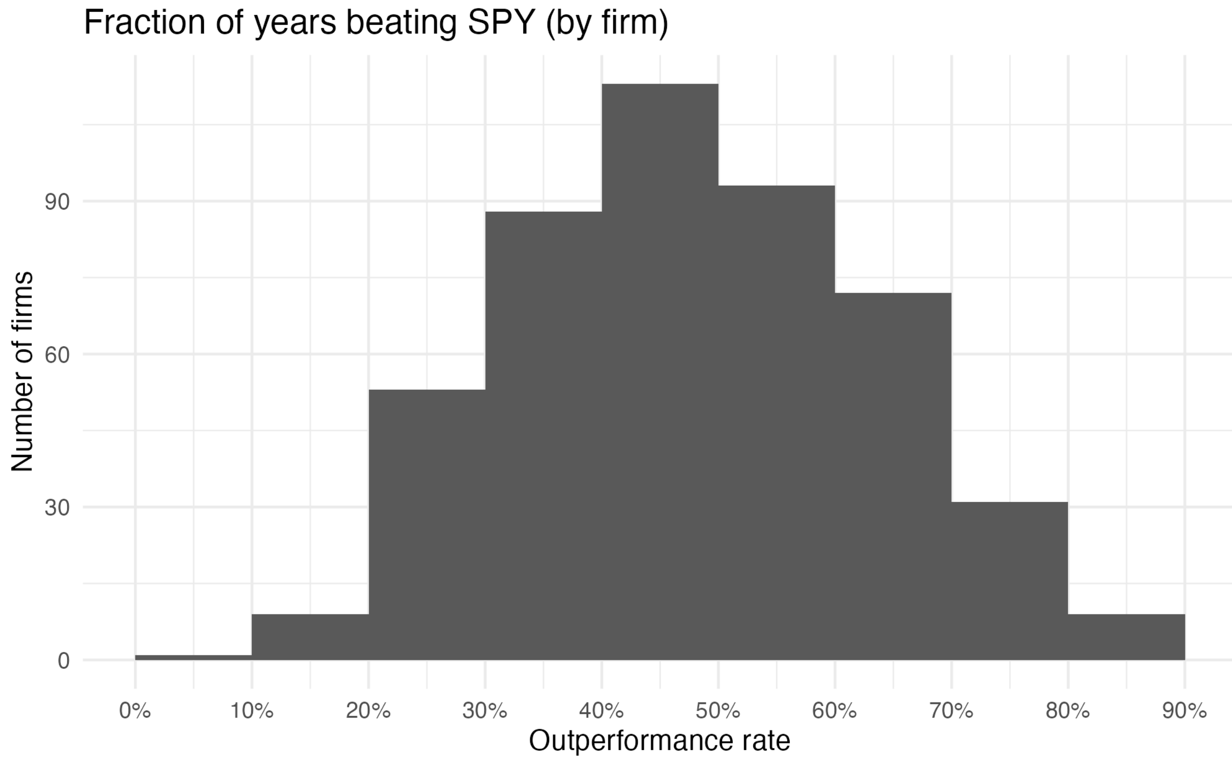


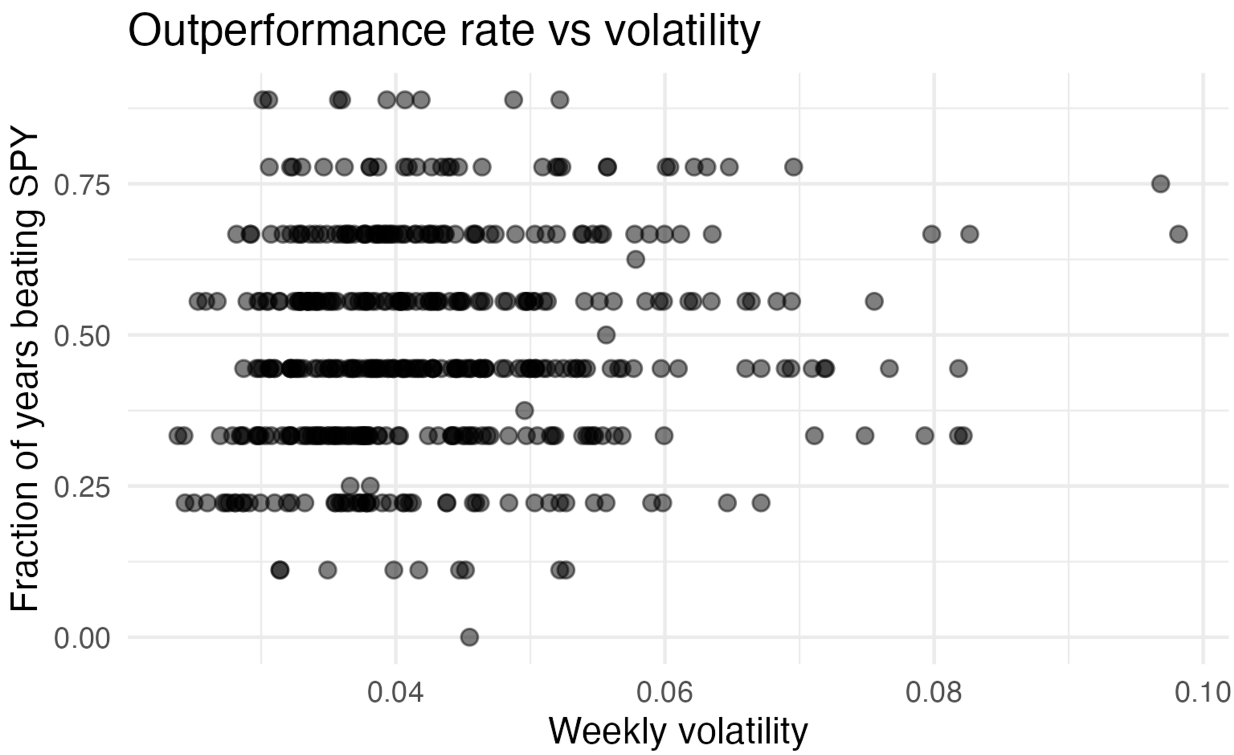
Figure 2 displays the distribution of firms' outperformance rates relative to SPY over the sample period as a bar plot. For each firm, the outperformance rate is calculated as the fraction of years in which the firm's annual return exceeded that of SPY. We consider only full calendar years and firms with at least 8 full calendar years' worth of data. For ease of interpretation, we bin by increments of 10% to approximate one-year increments.

Figure 2: Distribution of Annual Stock Returns



In our initial data set, when we look at weekly or even monthly data data, stock performance varies a lot from firm to firm, but much of this variation appears to be noisy rather than meaningful.

Figure 3: Volatility vs. Outperformance



We measure volatility using the standard deviation of weekly returns because the underlying price data are weekly and weekly returns provide many observations per firm, yielding a more stable estimate of risk than annual volatility (which would be based on only 9 data points). This measure is directly comparable to annual risk because weekly volatility can be annualized.

Additionally, annual returns are very spread out and skewed by a small number of huge winners, which makes it clear that extreme outcomes play an outsized role and that it makes more sense to look at performance over longer horizons. Most firms beat SPY in about half of the years in the sample, and only a small handful do so consistently. This pattern suggests that many “top performers” are likely the result of luck rather than persistent skill. We also see little connection between how often a firm beats the market and how volatile its weekly returns are, implying that higher outperformance is not simply compensation for taking more risk. Overall, these patterns caution against taking raw performance rankings at face value and point to the value of a Bayesian approach that explicitly accounts for uncertainty when estimating which firms truly outperform the market.

Figure 4: Distribution of Average Dividend Yield

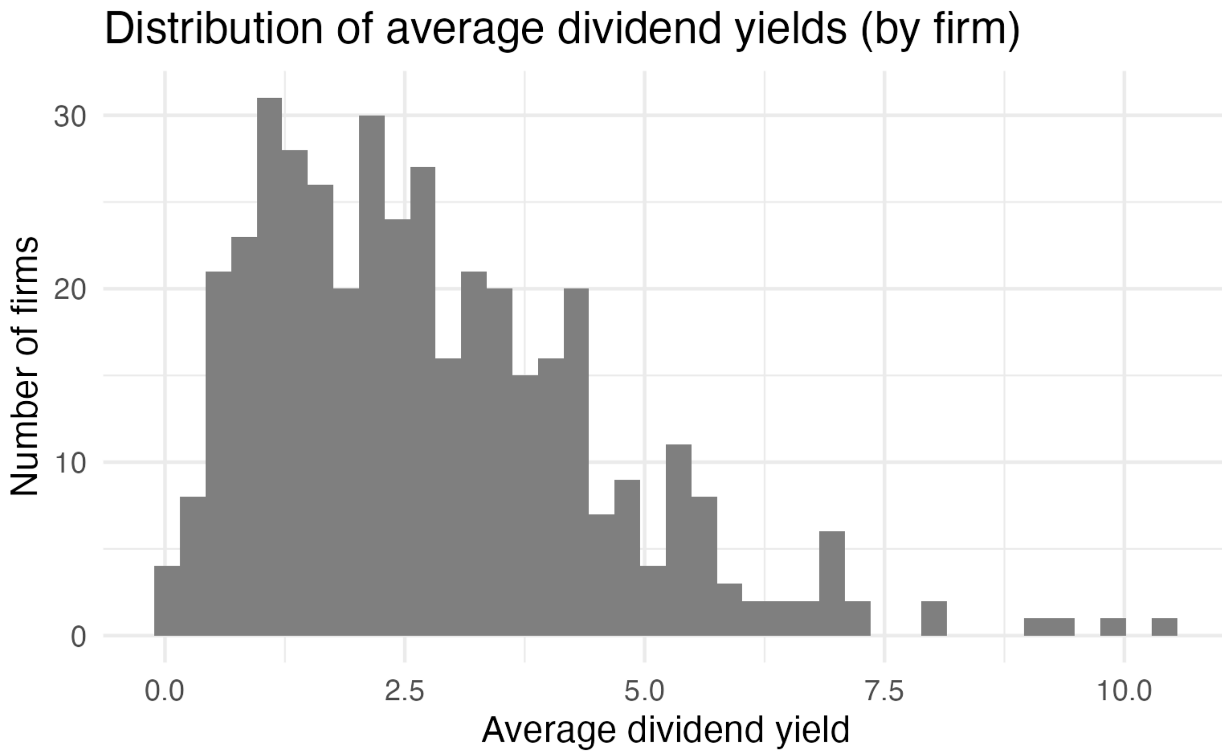


Figure 5: Dividend Yield vs. Outperformance

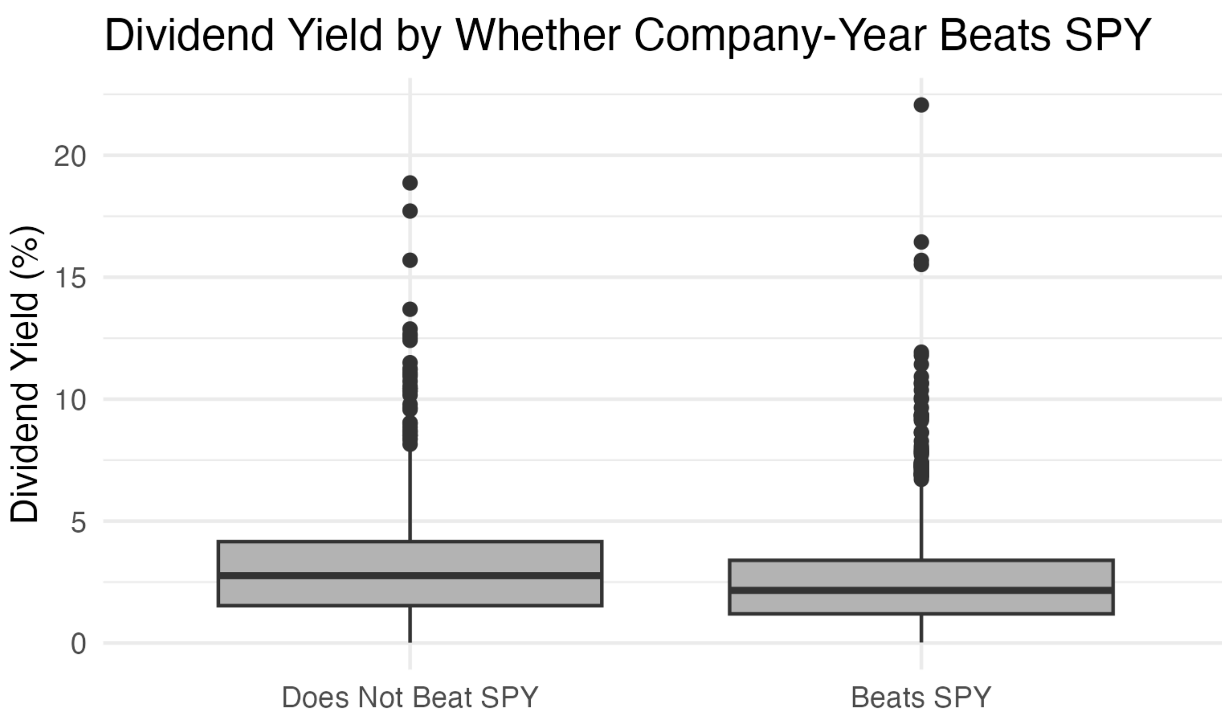


Figure 6: Distribution of P/E Ratio

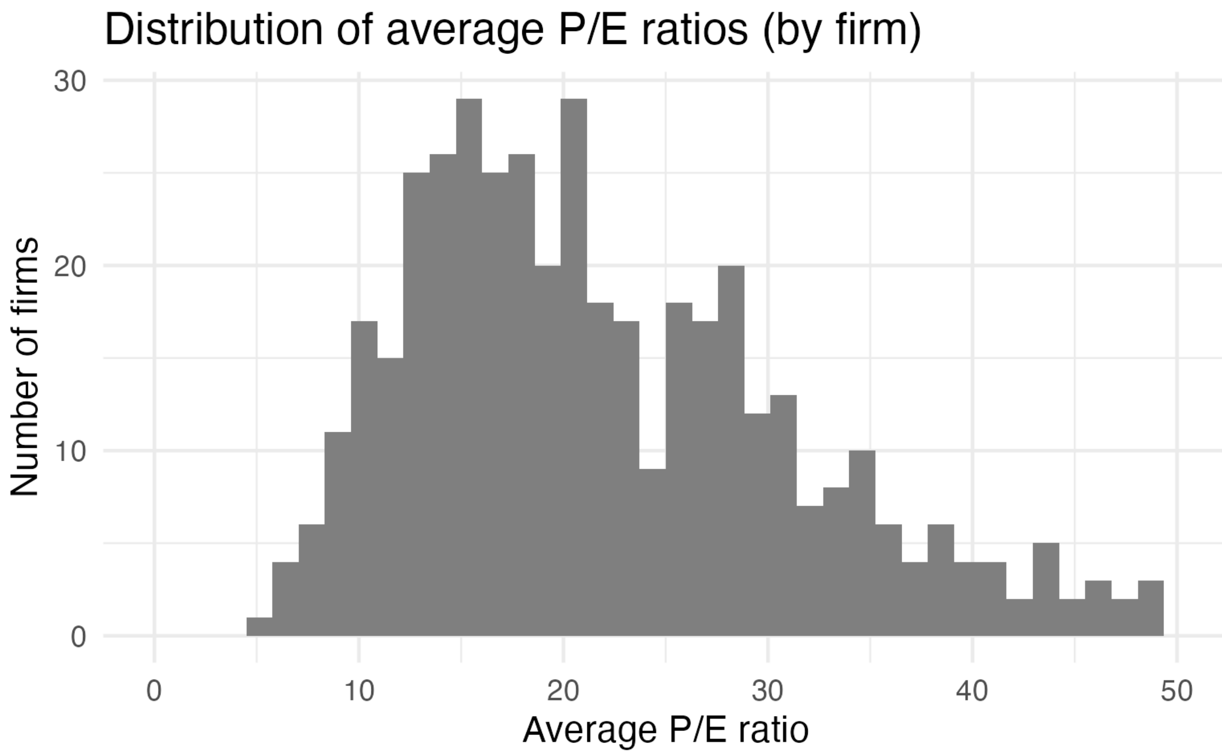
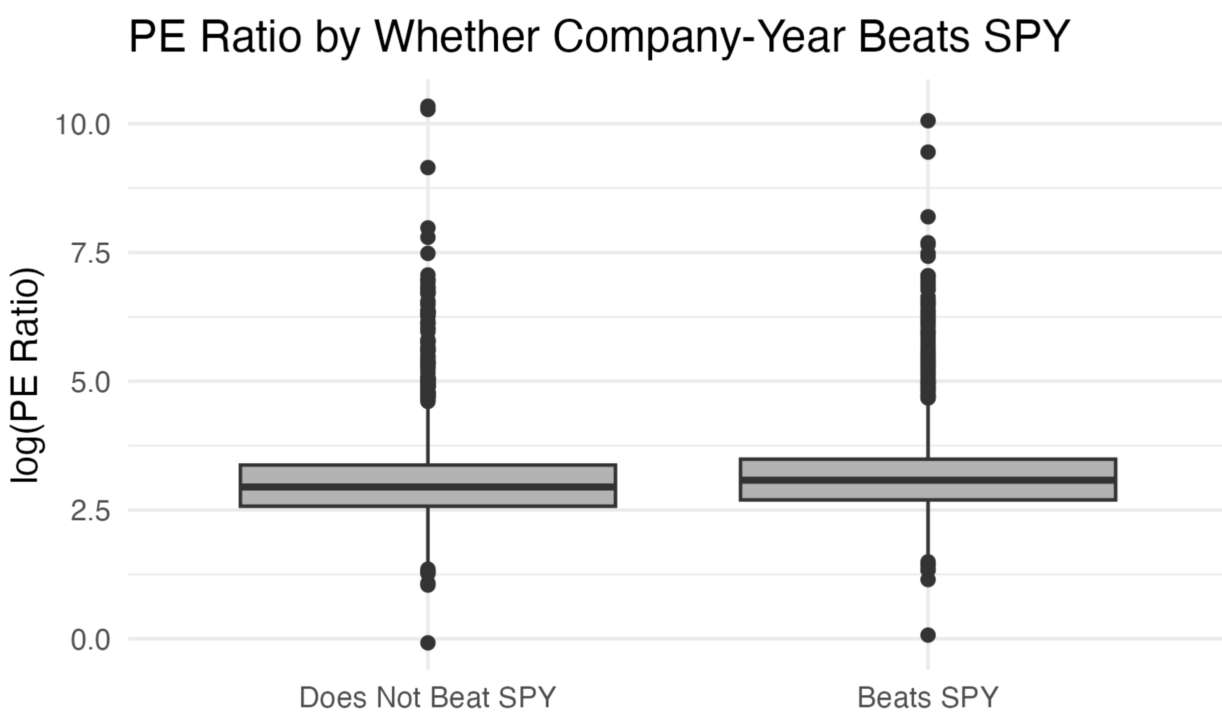


Figure 7: P/E Ratio vs. Outperformance



When we expand the exploratory analysis in terms of risk, valuation, and dividend policies, these differences do not translate into clear differences in how often they beat the market. P/E ratios and dividend yields vary widely across firms and have long right tails, which tells us there is real dispersion in fundamentals. When we plot these characteristics by firm-year, categorized by whether or not the firm-year beat the SPY (Figs. 5, 7), we see that firm-years which beat the SPY see lower median dividend yields and higher PE ratios versus firm-years that did not beat the SPY. This gap appears consistent even when only comparing firm-years in a shared sector. However, the differences appear small and noisy. Higher volatility does not reliably lead to more frequent outperformance, suggesting that simply taking on more risk does not mechanically increase the chances of beating the market.

What stands out most across all the plots is how tightly outperformance rates cluster around 50%. Since we only observe a small number of annual outcomes per firm (9 full calendar years), this discreteness makes it easy for firms to look unusually good or bad just by chance. Overall, the exploratory data analysis suggests that apparent differences in outperformance across firms are largely driven by noise rather than persistent skill, and that simple, realized comparisons can be very misleading without a framework that explicitly accounts for uncertainty. Thus, these findings motivate the use of a Bayesian framework.

3 Frequentist Analysis

3.1 Model Specifications

To study market outperformance at the firm level, we have adopted a binary approach to define if a firm did/did not outperform SPY in a given year. Using this approach, we assume that the outcomes follow a Bernoulli distribution where the probability of firm i beats the market benchmark in a given year t from 0 to 1. Additionally, we treat each firm's performance in different years as independent observations. This means that, once returns are aggregated to the annual level, how a firm performs in one year is assumed not to directly determine how it performs in the next. This is a reasonable simplifying assumption because annual returns already smooth out short-term momentum, reversals, and other high-frequency dynamics that operate at the weekly or monthly level. Further, we take the log of P/E ratio since some companies have extremely high ratios that would dominate our regression and skew our results.

3.1.1 Baseline Logistic Regression

Given $Y_{it} \sim Bernoulli(p_{it})$ where p_{it} is the probability that firm i outperformed SPY in year t,

$$logit(p_{it}) = \beta_0 + \beta_1 log(PE_{it}) + \beta_2 DivYield_{it}$$

Table 2: Frequentist Baseline Logistic Model Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.18	0.19	-0.92	0.36
log_pe	-0.03	0.06	-0.55	0.58
div_yield	0.15	0.02	7.68	0.00

Given this baseline pooled model, we are interested in more company-specific terms.

3.1.2 Random Intercept and Coefficient for Firm

We are also interested in firm-specific terms. We fit a mixed effects logistic regression model with firm-level random intercepts and slopes.

$$logit(p_{it}) = \beta_0 + \beta_1 log(PE_{it}) + \beta_2 DivYield_{it} + \beta_{0i} + \beta_{1i} log(PE_{it}) + \beta_{2i} DivYield_{it}$$

Table 3: Frequentist Mixed Effect Model for Firm

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.25	0.20	-1.22	0.22
log_pe	-0.02	0.06	-0.38	0.70
div_yield	0.17	0.02	8.06	0.00

An important result from this model comes from the covariance matrix for random effects.

Table 4: Random Effect Covariance Matrix

	grp	var1	var2	vcov	sdcor
1	Ticker	(Intercept)		0.29	0.54
2	Ticker	log_pe		0.01	0.11
3	Ticker	div_yield		0.01	0.09
4	Ticker	(Intercept)	log_pe	-0.06	-1.00
5	Ticker	(Intercept)	div_yield	-0.05	-1.00
6	Ticker	log_pe	div_yield	0.01	1.00

We observe a singular covariance matrix. This indicates we have a multicollinearity issue. Conceptually, this makes sense–out model is likely overparametrized since we have 400+ firms but only 8 useable annual observations for each firm. We thus move away from looking at very fine firm-specific effects and instead look to firm sectors.

3.1.3 Sector Fixed Effects

As a "middle ground" between the baseline and mixed models, we fit a logistic regression model with sector fixed effects.

$$logit(p_{it}) = \beta_0 + \beta_1 log(PE_{it}) + \beta_2 DivYield_{it} + \eta_1 + \eta_2 + \cdots + \eta_{10}$$

Table 5: Frequentist Sector FE Model Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.51	0.31	1.65	0.10
log_pe	-0.16	0.06	-2.49	0.01
div_yield	0.10	0.02	4.37	0.00
gics_sector_nameConsumer Discretionary	-0.26	0.25	-1.05	0.29
gics_sector_nameConsumer Staples	0.26	0.25	1.08	0.28
gics_sector_nameEnergy	0.03	0.26	0.11	0.91
gics_sector_nameFinancials	-0.43	0.23	-1.85	0.06
gics_sector_nameHealth Care	0.13	0.24	0.52	0.61
gics_sector_nameIndustrials	-0.31	0.23	-1.35	0.18
gics_sector_nameInformation Technology	-0.67	0.24	-2.75	0.01
gics_sector_nameMaterials	0.01	0.26	0.04	0.97
gics_sector_nameReal Estate	0.40	0.27	1.49	0.14
gics_sector_nameUtilities	-0.13	0.25	-0.51	0.61

3.2 Model Comparison

We compare information criteria between the three frequentist models.

Table 6: AIC Comparison

	df	AIC
Sector Fixed Effects	13.00	4579.02
Baseline Logistic	3.00	4624.37
Firm Mixed Effects	9.00	4632.86

Table 7: BIC Comparison

	df	BIC
Baseline Logistic	3.00	4642.76
Firm Mixed Effects	13.00	4658.73
firm_re	9.00	4688.05

To gauge predictive performance, we also perform 10-fold CV and compare AUC scores.

Table 8: AUC from 10-fold CV

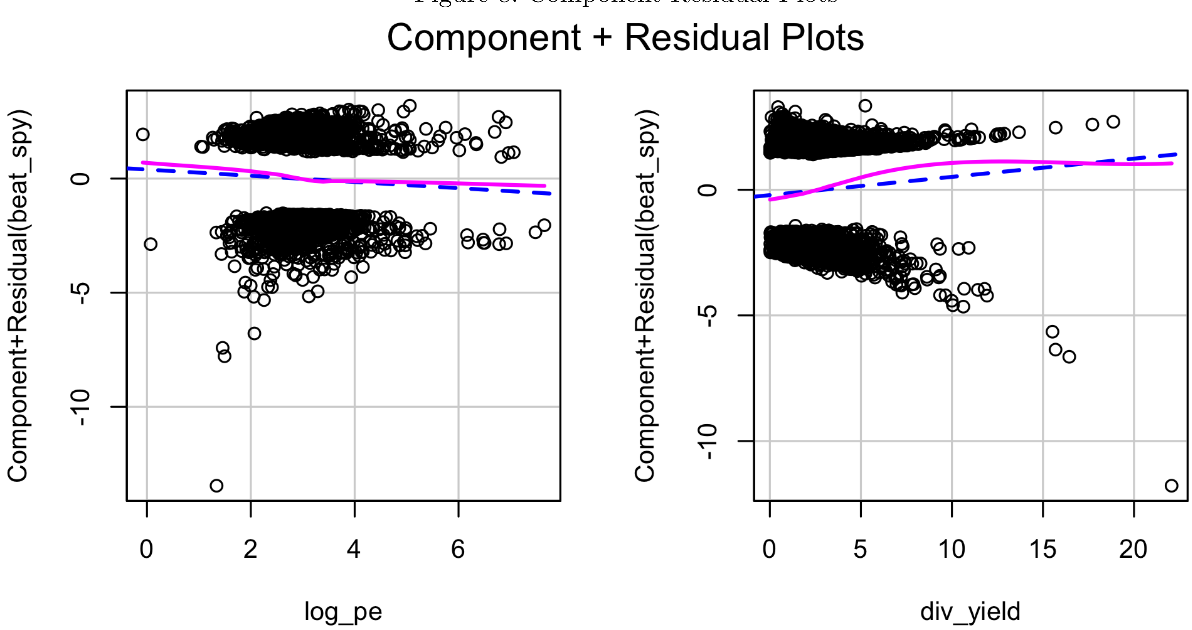
	model	auc
1	Sector Fixed Effects	0.61
2	Firm Mixed Effects	0.59
3	Baseline Logistic	0.59

We observe that the sector fixed effects model has lower AIC and marginally higher AUC. Since we want to produce a Bayesian analog for predictive purposes, we proceed with the fixed effects model.

3.3 Model Diagnostics

We perform model checks on the frequentist fixed effects model. To diagnose linearity violations we plot component-residual plots.

Figure 8: Component Residual Plots



Although these is slight curvature on the plot for dividend yield, it is minor. We argue that linearity assumptions reasonably hold.

We then use the `vif` function to test for multicollinearity.

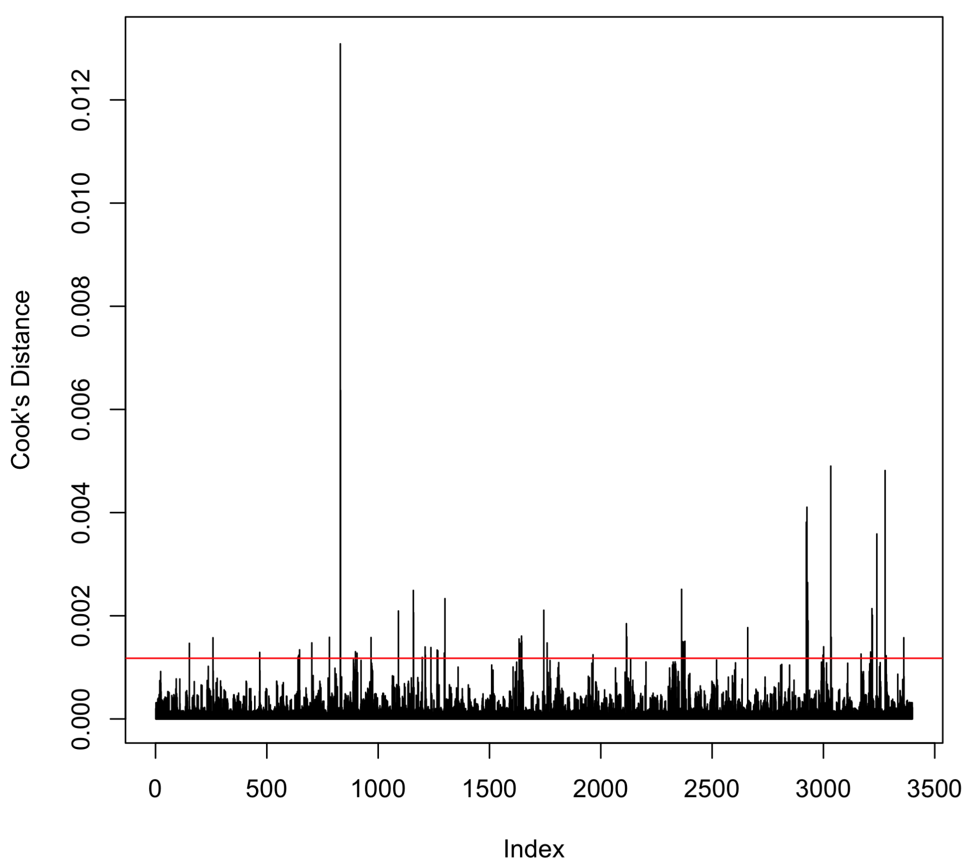
Table 9: VIF Results

	GVIF	Df	GVIF^(1/(2*Df))
log_pe	1.48	1.00	1.22
div_yield	1.52	1.00	1.23
gics_sector_name	1.65	10.00	1.03

Since our GVIF scores are low, we conclude we do not have multicollinearity issues.

To test for influential points, we plot Cook's distances.

Figure 9: Cook's Distances



We observe a few influential points. We remove the outliers and try refitting the model.

Table 10: Frequentist Firm Fixed Effects w/o Influential Points

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.25	0.38	3.28	0.00
log_pe	-0.22	0.07	-2.90	0.00
div_yield	0.17	0.03	6.44	0.00
gics_sector_nameConsumer Discretionary	-1.02	0.30	-3.41	0.00
gics_sector_nameConsumer Staples	-0.49	0.30	-1.65	0.10
gics_sector_nameEnergy	-0.61	0.32	-1.91	0.06
gics_sector_nameFinancials	-1.20	0.29	-4.20	0.00
gics_sector_nameHealth Care	-0.58	0.30	-1.95	0.05
gics_sector_nameIndustrials	-1.03	0.29	-3.62	0.00
gics_sector_nameInformation Technology	-1.42	0.30	-4.77	0.00
gics_sector_nameMaterials	-0.69	0.31	-2.24	0.03
gics_sector_nameReal Estate	-0.36	0.32	-1.13	0.26
gics_sector_nameUtilities	-1.00	0.30	-3.28	0.00

We observe some differences from the original frequentist sector fixed effects model. Notably, the slope coefficients become more extreme. It might be reasonable to remove these influential observations before fitting our Bayesian models. However, since we already have few annual observations per firm and Bayesian models inherently utilize difference model assumptions than frequentist ones, we opt to use the original dataset.

4 Bayesian Analysis

We fit two Bayesian models corresponding to the baseline and sector fixed effect models using the `brms` package in R. We use the default `brms` priors, although we later perform sensitivity analysis to show that our model results are not overshadowed by the choice of priors.

4.1 Model Fitting

4.1.1 Baseline Logistic Regression

We check that our baseline Bayesian model has converged by analyzing Rhat and ESS values, as well as trace and ACF plots.

The Rhat values for the estimated parameters are 1, and the effective sample sizes are greater than 100. The trace plots also do not display drift, and the ACF decays quickly. This indicates successful MCMC convergence.

Table 11: Summary of Bayesian Baseline Logistic Model

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-0.18	0.19	-0.55	0.18	1.00	3151.58	2686.70
log_pe	-0.03	0.05	-0.13	0.08	1.00	3418.78	2877.85
div_yield	0.15	0.02	0.11	0.19	1.00	3661.87	2709.49

Figure 10: Trace Plots for Baseline Bayesian Model

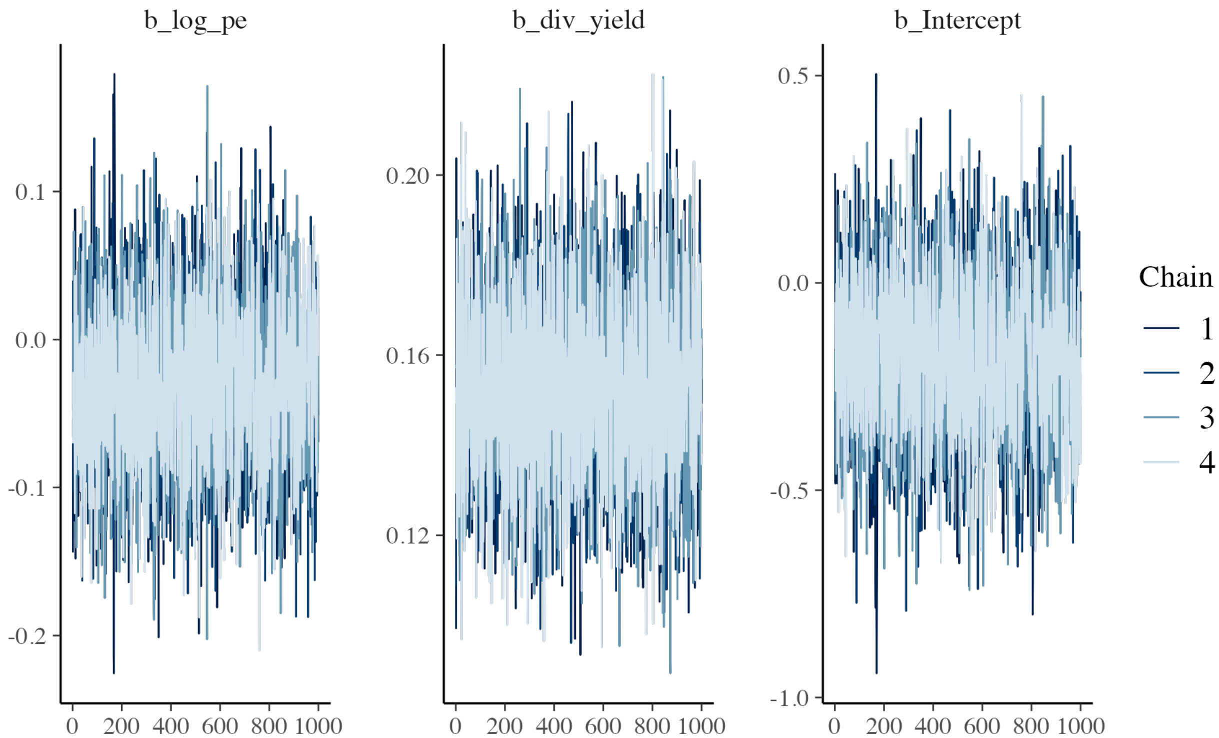
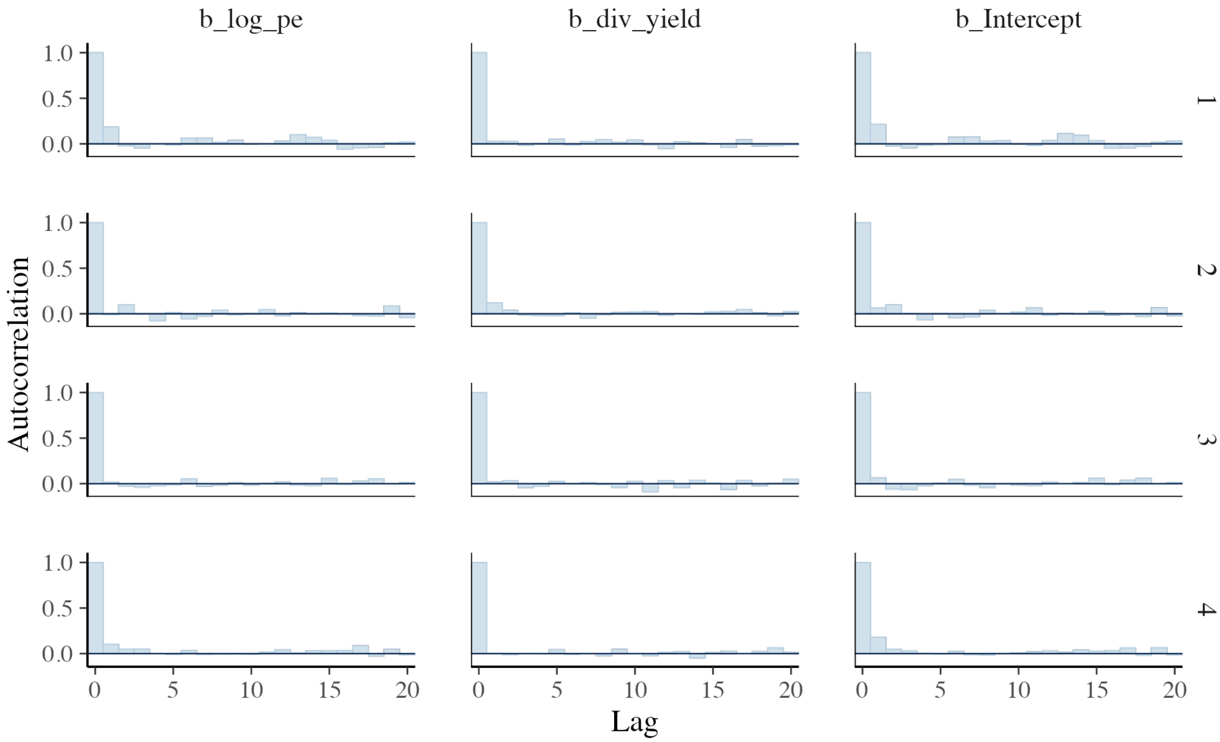


Figure 11: ACF Plots for Baseline Bayesian Model



4.1.2 Sector Fixed Effects

We similarly check for MCMC convergence of our Bayesian model with sector fixed effects. For each parameter, Rhat values are equal to 1, effective sample sizes are well over 100, trace plots do not show drift, and ACF decay quickly. For sake of space, we only include plots for one of the sector indicators corresponding to the Information Technology sector.

Figure 12: Trace Plots for Sector FE Bayesian Model

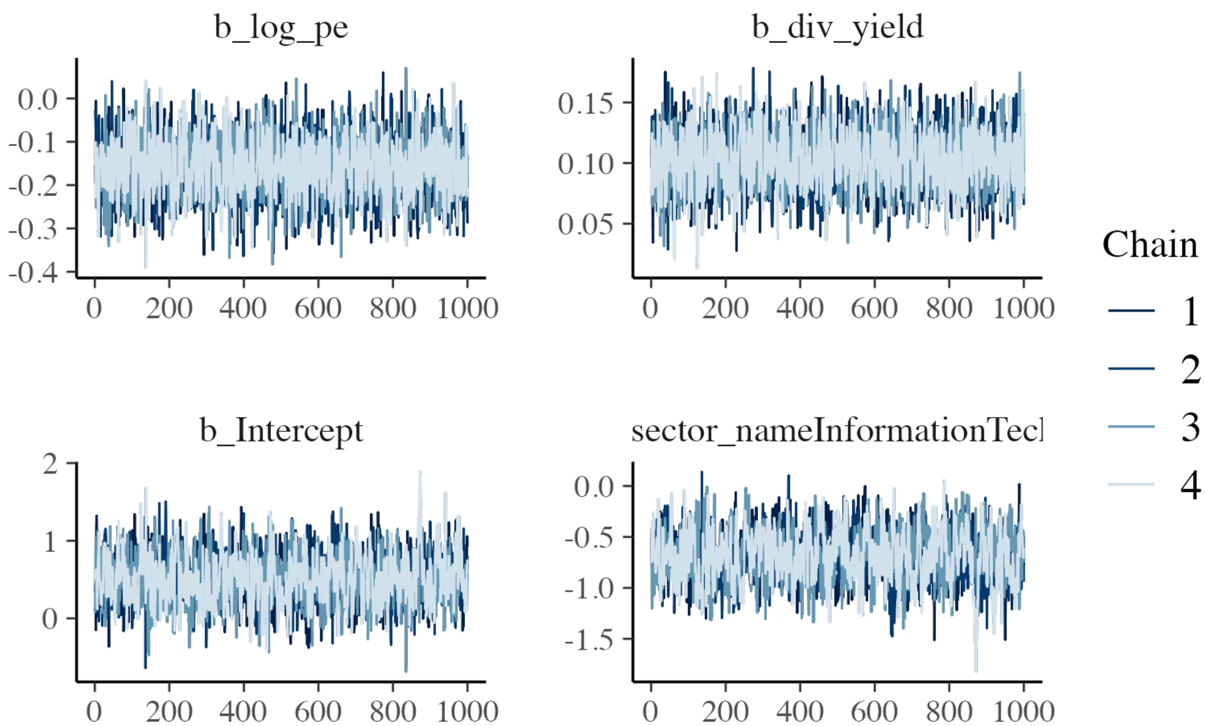


Figure 13: ACF Plots for Sector FE Bayesian Model

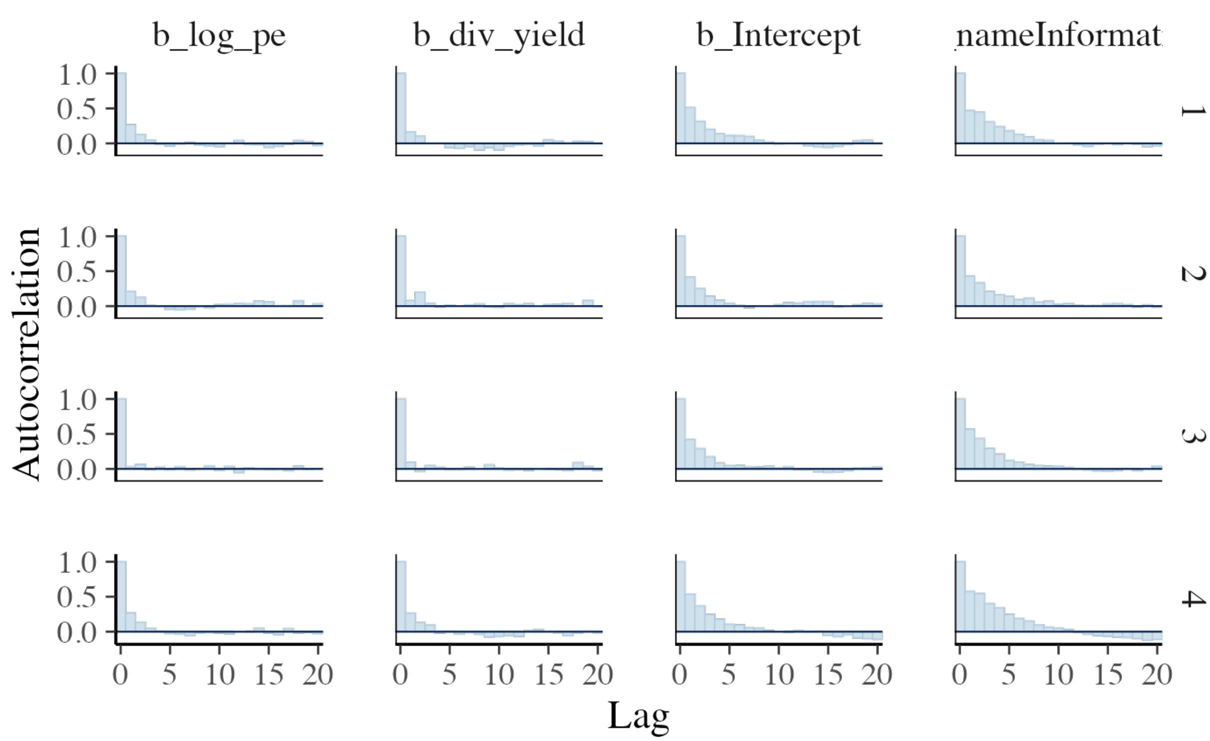


Table 12: Summary of Bayesian Sector FE Model

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.51	0.31	-0.10	1.14	1.00	1099.51	1778.43
log_pe	-0.16	0.07	-0.29	-0.03	1.00	2424.67	2577.72
div_yield	0.10	0.02	0.06	0.15	1.00	2519.43	2880.30
gics_sector_nameConsumerDiscretionary	-0.27	0.25	-0.77	0.22	1.00	795.27	1172.71
gics_sector_nameConsumerStaples	0.26	0.25	-0.23	0.75	1.00	813.66	1443.16
gics_sector_nameEnergy	0.02	0.27	-0.52	0.53	1.00	913.25	1455.72
gics_sector_nameFinancials	-0.43	0.23	-0.90	0.00	1.00	750.67	1247.95
gics_sector_nameHealthCare	0.12	0.25	-0.38	0.59	1.00	796.48	1405.77
gics_sector_nameIndustrials	-0.32	0.23	-0.79	0.12	1.01	725.87	1348.52
gics_sector_nameInformationTechnology	-0.68	0.25	-1.17	-0.21	1.00	792.79	1264.93
gics_sector_nameMaterials	0.00	0.26	-0.51	0.53	1.00	907.25	1453.87
gics_sector_nameRealEstate	0.39	0.27	-0.16	0.91	1.00	871.04	1413.98
gics_sector_nameUtilities	-0.14	0.25	-0.64	0.34	1.00	807.86	1184.87

4.2 Model Comparison

We use the `loo` function to compare the expected log predictive density (ELPD) of our two Bayesian models.

Table 13: Leave-one-out Comparison of Bayesian Models

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
bayes_fe	0.00	0.00	-2289.83	12.82	13.50	0.28	4579.67	25.63
bayes_baseline	-22.44	8.04	-2312.28	10.22	3.23	0.18	4624.56	20.44

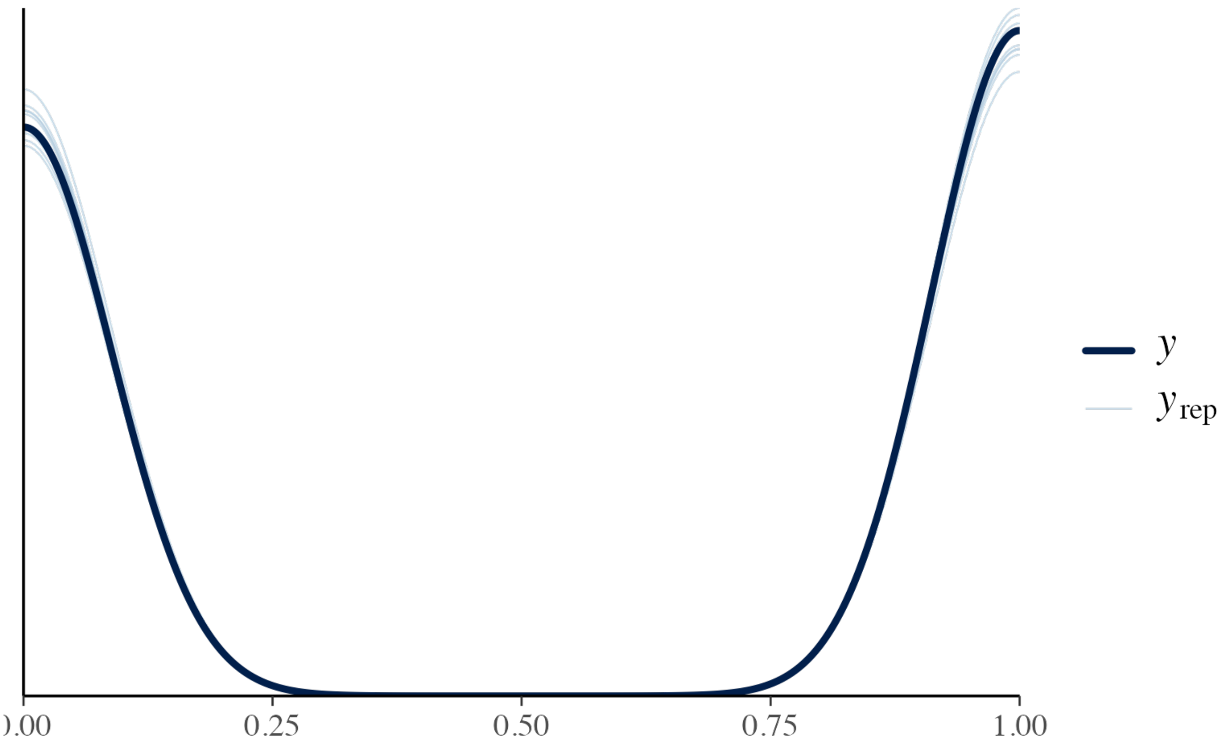
The ELPD for our sector fixed effects model is substantially higher than that of the baseline model ($\Delta = 22.44$) with a significant difference in standard error (8.04). This indicates that the sector fixed effects model has significantly better predictive performance than the baseline model. Therefore, we use the sector fixed effects model for the following model checking and prediction analysis.

4.3 Model Checks

4.3.1 Posterior Predictive Checks

We perform posterior predictive checks to show that our model can produce replicated data that reasonably resembles our observed data using the `pp_check` function.

Figure 14: PPC for Sector FE Model (default prior)



The observed data resembles the U-shaped curve produced by data simulated from our model's posterior distribution. We conclude our model gives us reasonable predictions.

4.3.2 Sensitivity Analysis

As mentioned before, we use the default `brm` priors when fitting our models. In particular, the priors for slope coefficients are flat, and the prior and the intercept is a studentized t-distribution with degrees of freedom, location, and scale parameters equal to (3, 0, 2.5). Here, we perform a sensitivity analysis of our sector fixed effects model to the choice of prior. Ideally, our choice of prior should not produce a major change in the posterior results.

We first try weakly informative priors, using a $t(3, 0, 2.5)$ distribution for slope coefficients and $t(3, 0, 5)$ distribution for intercepts. We then try slightly regularizing priors, with a $N(0, 1)$ prior for slopes and a $N(0, 2)$ distribution for intercepts.

These two sets of priors result in the following posterior estimates.

Table 14: Sector FE model using Weakly Informative Priors

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.51	0.31	-0.07	1.12	1.00	1058.89	1521.78
log_pe	-0.16	0.06	-0.29	-0.04	1.00	2424.54	2642.91
div_yield	0.10	0.02	0.06	0.15	1.00	2249.90	2325.14
gics_sector_nameConsumerDiscretionary	-0.25	0.24	-0.73	0.21	1.00	798.26	1187.32
gics_sector_nameConsumerStaples	0.27	0.24	-0.22	0.73	1.00	765.05	1314.17
gics_sector_nameEnergy	0.04	0.27	-0.48	0.55	1.00	882.46	1617.17
gics_sector_nameFinancials	-0.42	0.23	-0.87	0.01	1.00	721.29	1084.08
gics_sector_nameHealthCare	0.13	0.24	-0.34	0.59	1.00	766.88	1139.73
gics_sector_nameIndustrials	-0.31	0.23	-0.75	0.12	1.00	693.95	1205.59
gics_sector_nameInformationTechnology	-0.66	0.24	-1.15	-0.21	1.00	777.80	1181.99
gics_sector_nameMaterials	0.02	0.26	-0.48	0.50	1.00	816.27	1434.06
gics_sector_nameRealEstate	0.40	0.27	-0.11	0.94	1.00	849.89	1613.50
gics_sector_nameUtilities	-0.12	0.25	-0.61	0.36	1.00	803.12	1359.53

Table 15: Sector FE model using Weakly Informative Priors

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.48	0.29	-0.08	1.04	1.00	1276.40	1792.04
log_pe	-0.16	0.06	-0.28	-0.03	1.00	2419.47	2574.31
div_yield	0.10	0.02	0.06	0.15	1.00	2468.07	2717.92
gics_sector_nameConsumerDiscretionary	-0.23	0.22	-0.66	0.19	1.00	1034.60	1745.07
gics_sector_nameConsumerStaples	0.28	0.22	-0.14	0.69	1.00	967.56	1458.58
gics_sector_nameEnergy	0.05	0.23	-0.40	0.51	1.00	1128.68	1732.61
gics_sector_nameFinancials	-0.40	0.20	-0.79	-0.01	1.00	876.13	1537.23
gics_sector_nameHealthCare	0.15	0.21	-0.27	0.55	1.00	974.85	1526.90
gics_sector_nameIndustrials	-0.29	0.20	-0.67	0.09	1.00	879.02	1515.31
gics_sector_nameInformationTechnology	-0.64	0.22	-1.06	-0.23	1.00	1006.10	1594.49
gics_sector_nameMaterials	0.03	0.23	-0.41	0.47	1.00	1118.07	1536.79
gics_sector_nameRealEstate	0.42	0.24	-0.06	0.89	1.00	1110.45	1849.02
gics_sector_nameUtilities	-0.11	0.22	-0.53	0.32	1.00	1050.07	1605.28

Figure 15: PPC of Sector FE Model w/ weakly informative priors

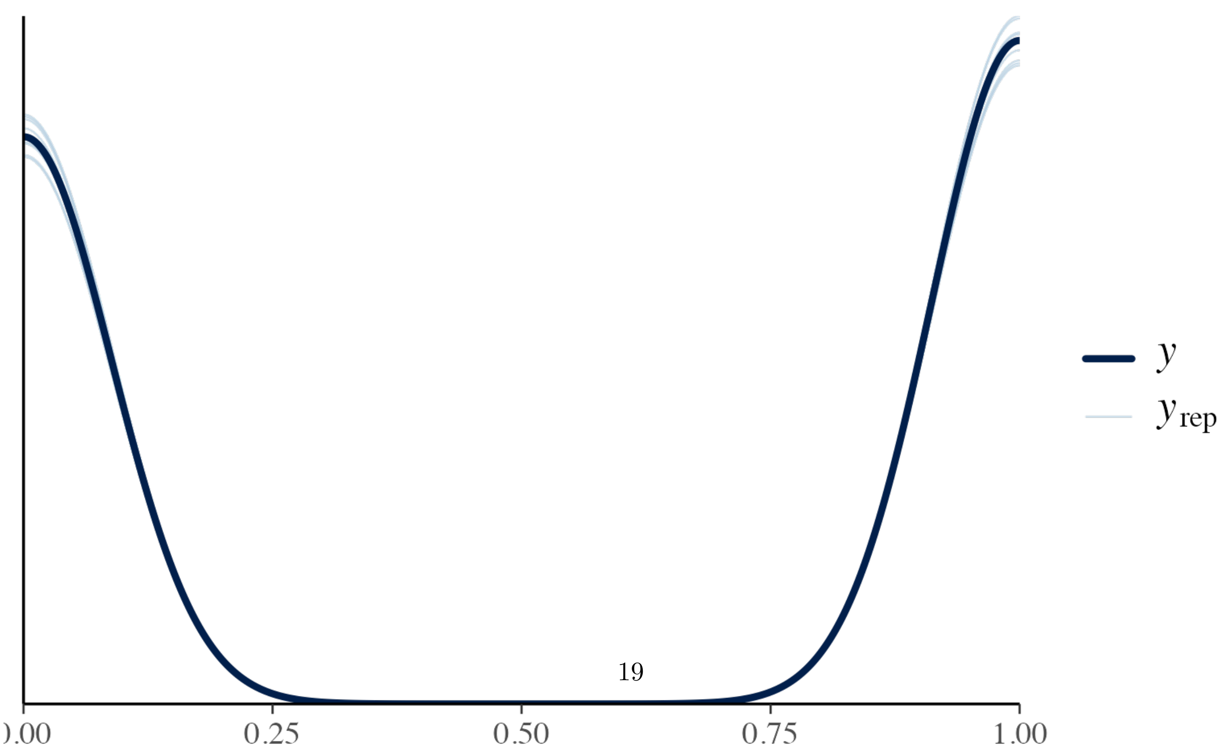
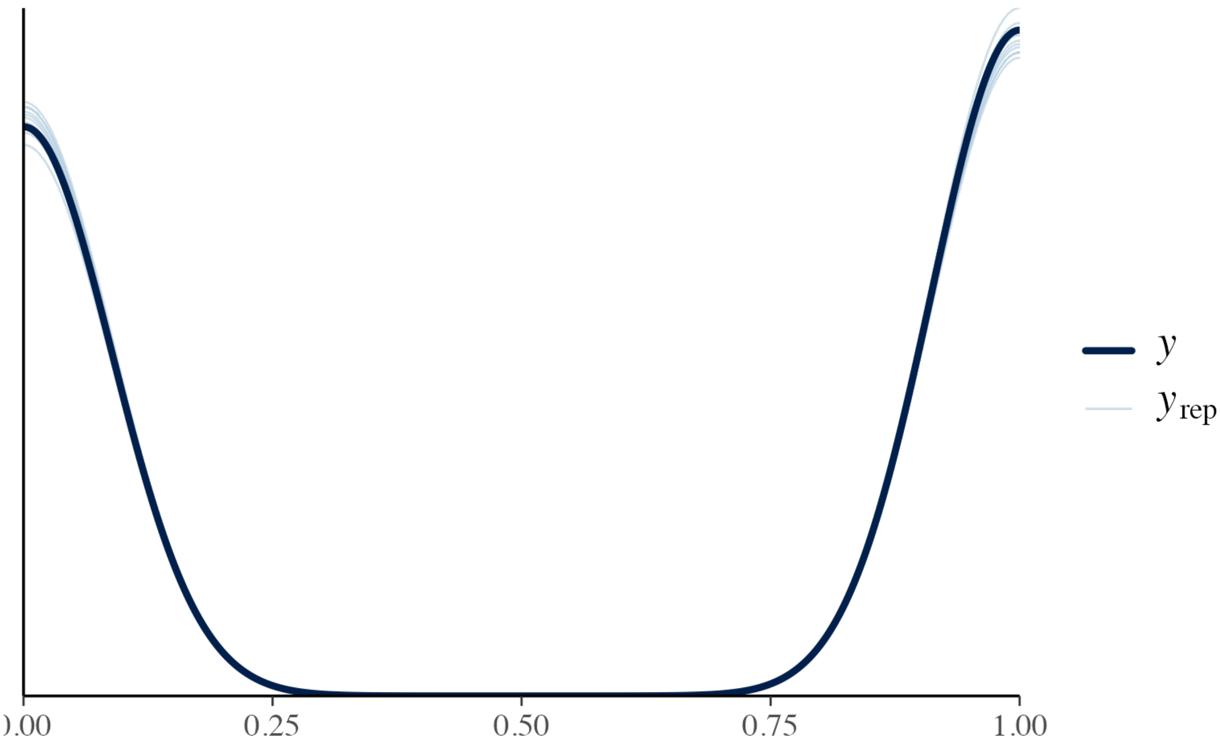


Figure 16: PPC of Sector FE Model w/ slightly regularizing priors



Although the point estimates differ slightly, they are quite similar to those from our original sector fixed effects model that uses the default `brm` priors. The posterior predictive checks also show that simulations from the posterior distributions are aligned with the observed data. This indicates our model is robust against the choice of prior.

4.4 Prediction

To demonstrate how to use the Bayesian sector fixed effects model to predict new observations, we consider an out-of-sample prediction. Paramount (ticker PSKY) is in the Communication Services sector. As of late February 2026 it has a P/E ratio of 450.33 with a dividend yield of 1.48%. Using the Bayesian sector fixed effects model, there is a 95% chance that the probability that IBM beats the SPY is between 35.4% and 46.3%.

5 Discussion