Music and Markets: Exploring the Impact of Public Mood on Stock Performance

This project investigates the potential connection between public sentiment (music) and stock market performance. By analyzing music sentiment and their impact on stock returns, we shed light on the possible role of it in financial markets. Our analysis encompasses major indices like the **S&P 500 and Russell 3000**.

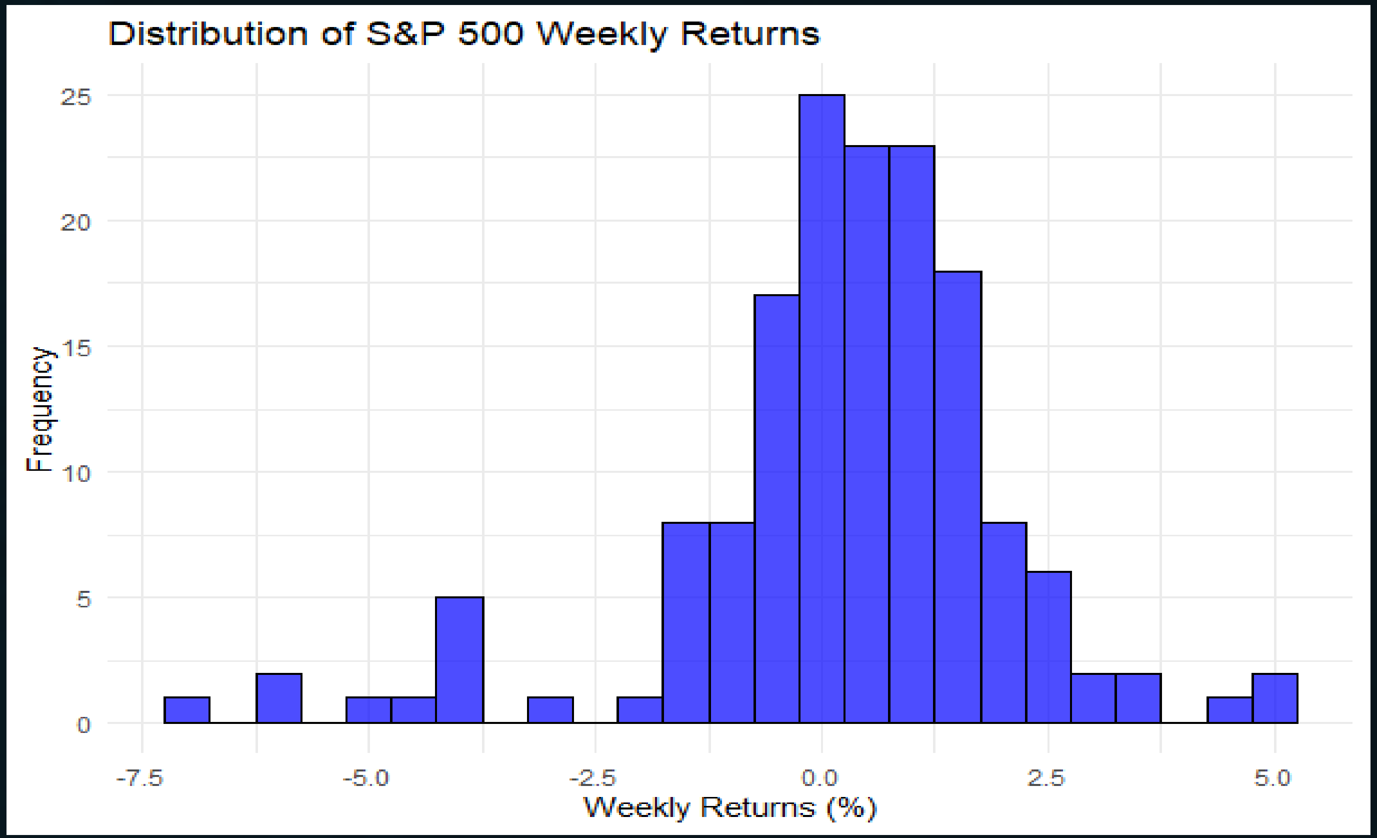
by ABHISHEK KUMAR SINGH

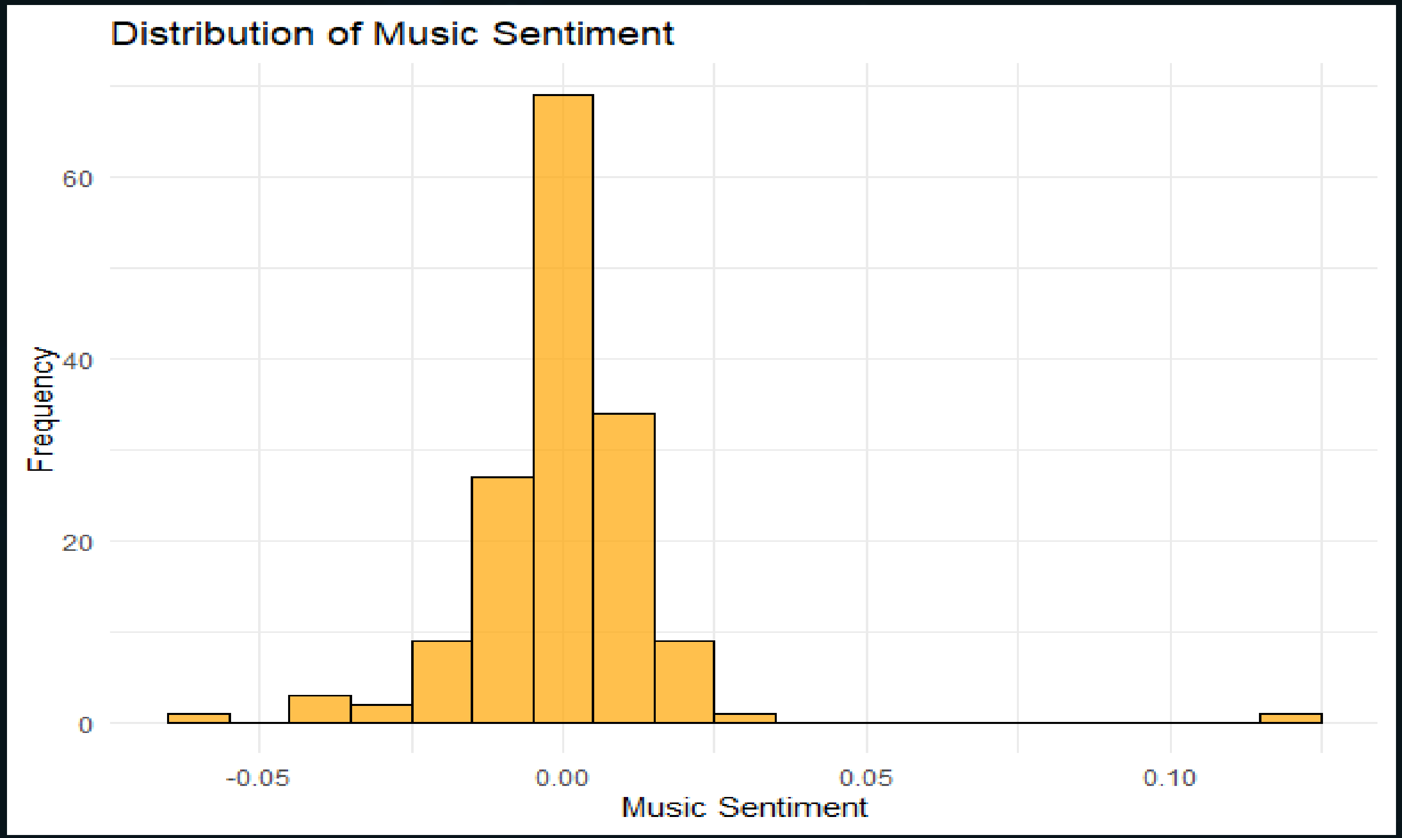
MIST.6160 Advanced Data Mining Project

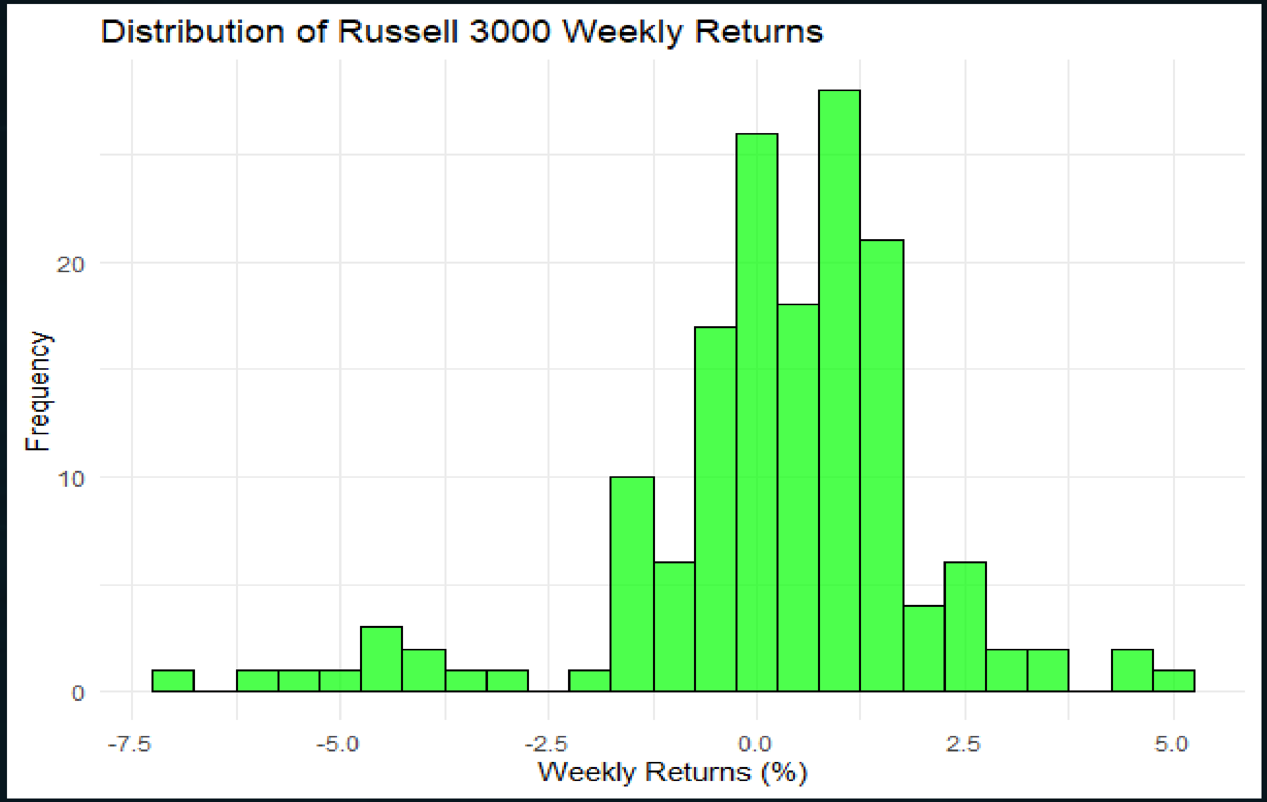
Article Insights

1. Music sentiment shows a positive correlation with same-week stock market returns and a negative correlation with next week's returns, suggesting sentiment-driven mispricing.

2. During times when certain trading activities were restricted (like during COVID-19 when some countries banned short-selling), the relationship between the music sentiment and stock prices became stronger, likely because these restrictions made it harder for markets to correct sentiment-driven price movements


Distribution of S&P 500 Weekly Returns
Frequency
0
5
10
15
20
25
-7.5
-5.0
-2.5
0.0
2.5
5.0
Weekly Returns (%)


Distribution of Music Sentiment
Frequency
0
20
40
60
-0.05
0.00
0.05
0.10
Music Sentiment


Distribution of Russell 3000 Weekly Returns
Frequency
0
10
20
-7.5
-5.0
-2.5
0.0
2.5
5.0
Weekly Returns (%)

Data Pre-processing

Feature engineering & Data Cleaning

- Calculated columns for weekly returns for S&P 500 and Russell 3000 markets, with & without risk-adjustment, Lagged returns & Lagged sentiments.

- After Pre-processing, there were **154 rows** and **16 columns** in the dataset.

- Since, there were very less NA values, Removed "NA" values because they could interfere with the modeling results

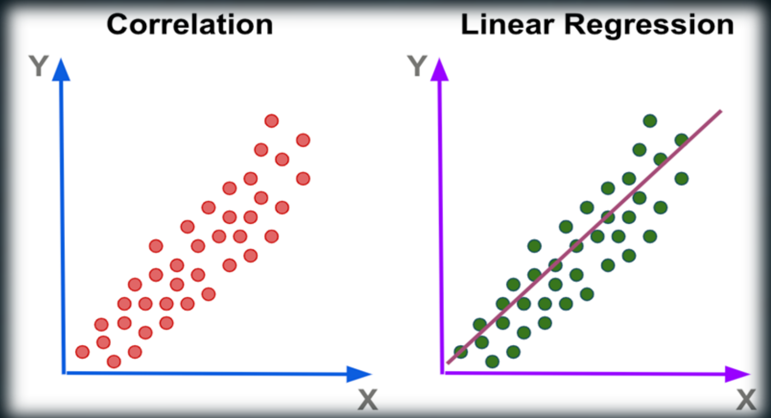
- Converted IRX rates using compounding interest formula to replicate real world analysis.

Date	Adj.Close_.GSPC	Adj.Close_.IRX	Adj.Close_.RUA	SWAV	Music_Sentiment	GSPC_Returns	RUA_Returns	Weekly_Risk_Free_Rate	GSPC_Risk_Adjusted	RUA_Risk_Adjusted	Lagged_GSPC_Returns	Lagged_RUA_Returns	Lagged_Music_Sentiment	Lagged_GSPC_RA	Lagged_RUA_RA
01-15-2017	2271.310059	0.479999989	1350.099976	0.4469045	-0.003949192	-0.1463895	-0.2806742	0.00920911	-0.155598611	-0.289883305	-0.102771523	-0.011079723	0.002033302	-0.112612173	-0.020920374
01-22-2017	2294.689941	0.497999996	1364.48999	0.4469491	0.0000446	1.029356688	1.06584803	0.009553612	1.019803076	1.056294419	-0.146389501	-0.280674195	-0.003949192	-0.155598611	-0.289883305
01-29-2017	2297.419922	0.488000005	1367.290039	0.4446552	-0.002293891	0.118969493	0.205208468	0.009362229	0.109607264	0.195846238	1.029356688	1.06584803	0.0000446	1.019803076	1.056294419
02-05-2017	2316.100098	0.523000002	1378.579956	0.4372419	-0.007413329	0.813093672	0.825714858	0.010031986	0.803061686	0.815682872	0.118969493	0.205208468	-0.002293891	0.109607264	0.195846238
02-12-2017	2351.159912	0.508000016	1398.040039	0.4382885	0.001046627	1.51374347	1.411603507	0.009744976	1.503998495	1.401858531	0.813093672	0.825714858	-0.007413329	0.803061686	0.815682872
02-19-2017	2367.340088	0.497999996	1405.599976	0.4390204	0.000731866	0.688178457	0.540752538	0.009553612	0.678624845	0.531198927	1.51374347	1.411603507	0.001046627	1.503998495	1.401858531
02-26-2017	2383.120117	0.683000028	1413.550049	0.4388264	-0.000193966	0.666572119	0.565599967	0.013090821	0.653481298	0.552509146	0.688178457	0.540752538	0.000731866	0.678624845	0.531198927
03-05-2017	2372.600098	0.725000024	1404.040039	0.4559315	0.017105104	-0.44143889	-0.67277491	0.013892975	-0.455331866	-0.68666788	0.666572119	0.565599967	-0.000193966	0.653481298	0.552509146
03-12-2017	2378.25	0.708000004	1410.380005	0.4523829	-0.003548607	0.238131239	0.451551653	0.013568333	0.224562906	0.43798332	-0.44143889	-0.672774905	0.017105104	-0.455331866	-0.68666788
03-19-2017	2343.97998	0.748000026	1387.959961	0.4286518	-0.023731065	-1.44097635	-1.58964562	0.014332111	-1.455308459	-1.603977733	0.238131239	0.451551653	-0.003548607	0.224562906	0.43798332
03-26-2017	2362.719971	0.737999976	1401.5	0.4351395	0.006487656	0.799494499	0.975535273	0.014141194	0.785353305	0.96139408	-1.440976348	-1.589645622	-0.023731065	-1.455308459	-1.603977733
04-02-2017	2355.540039	0.797999978	1395.630005	0.4376093	0.002469836	-0.30388417	-0.4188366	0.015286414	-0.319170585	-0.434123018	0.799494499	0.975535273	0.006487656	0.785353305	0.96139408

Ordinary Least Squares (OLS)

OLS Regression:

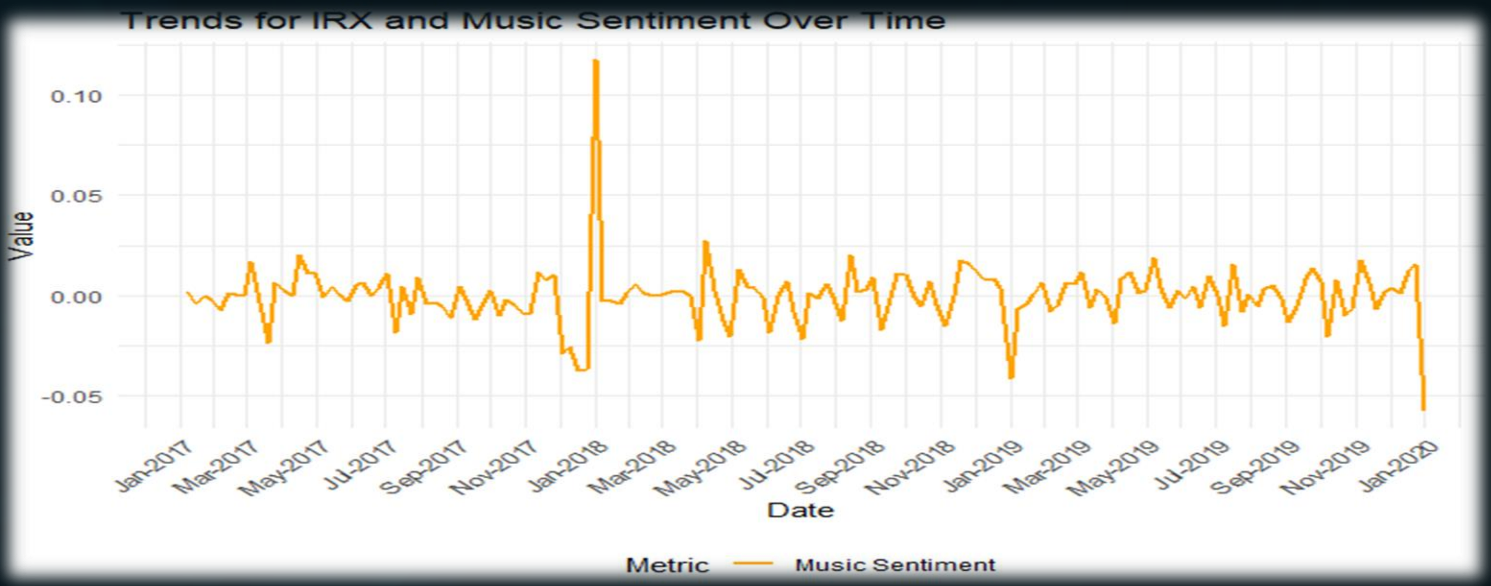
OLS regression is a fundamental technique for estimating relationships between variables. It finds the line that best fits a set of data points, minimizing the sum of squared errors between predicted and actual values.

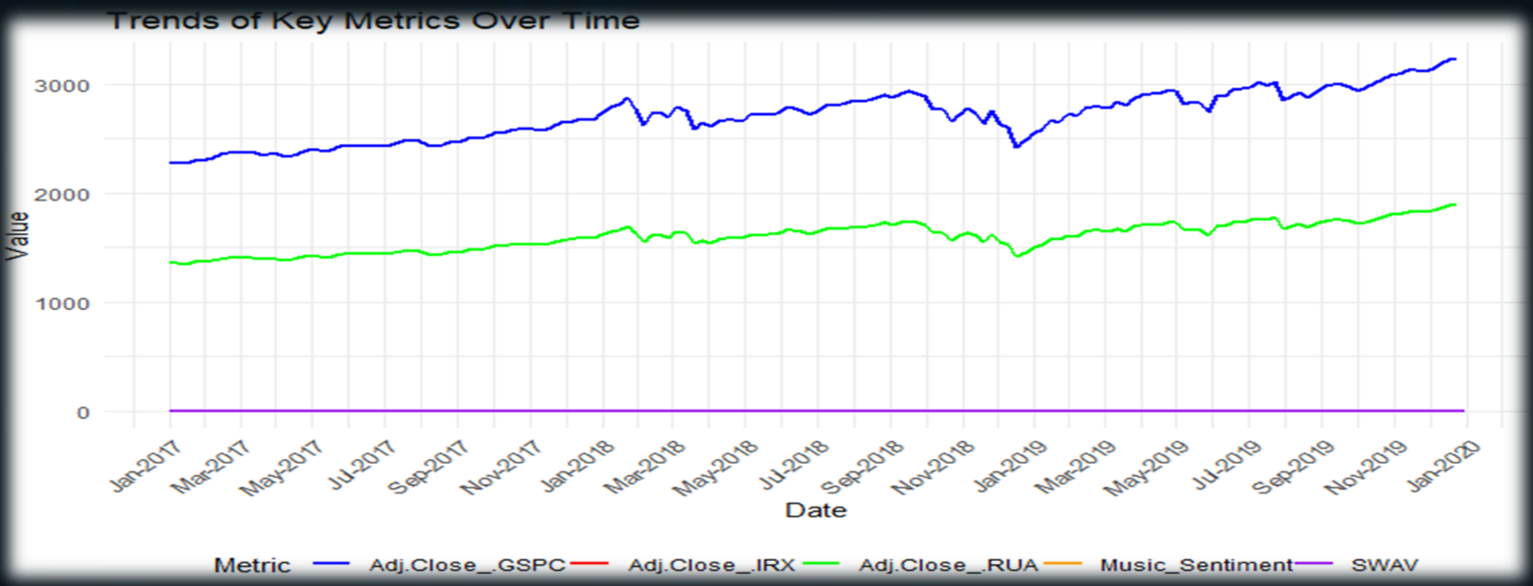


OLS regression models:

8 OLS models

4 Models related to S&P500 returns & 4 related to Rusell 3000 returns. These 4 models include each model with simple returns, risk-adjusted returns for the same week and then 2 models for sentiment from the previous weeks (lagged music sentiments).





OLS Model Results: Music Sentiment and Stock Returns

S&P 500

The analysis shows that music sentiment has a coefficient of -8.919 with a standard error of 13.649 and a **p-value of 0.513. These results indicate that music sentiment does not significantly predict S&P 500 returns for both the normal returns & risk-adjusted returns.**

Russell 3000

Similarly, here, we found a music sentiment coefficient of -10.13, a standard error of 10.44, and a **p-value of 0.333**. These findings **reinforce the conclusion that music sentiment lacks a significant predictive power on Russell 3000 returns for both the normal returns & risk-adjusted returns.**

If we don’t consider the p-values, for a moment. We can see that returns are this week’s music sentiment has more profound effect on the stock returns than the sentiment last week, i.e. Lagged Music sentiment.

Model	Music Sentiment Estimate	Music Sentiment P-Value	Lagged Returns Estimate	Lagged Returns P-Value
GSPC Returns	-8.91936	0.3842	-0.17644	0.0297
RUA Returns	-10.45457	0.3082	-0.15055	0.0638
GSPC Risk-Adjusted Returns	-8.93661	0.3835	-0.17576	0.0303
RUA Risk-Adjusted Returns	-10.47156	0.308	-0.14987	0.065
GSPC Returns (Lagged Music)	-1.7980	0.8610	-0.1704	0.0355
RUA Returns (Lagged Music)	-1.97045	0.8482	-0.14389	0.0767
GSPC RA (Lagged Music)	-1.81739	0.8595	-0.16968	0.0363
RUA RA (Lagged Music)	-1.98796	0.8469	-0.14319	0.0781

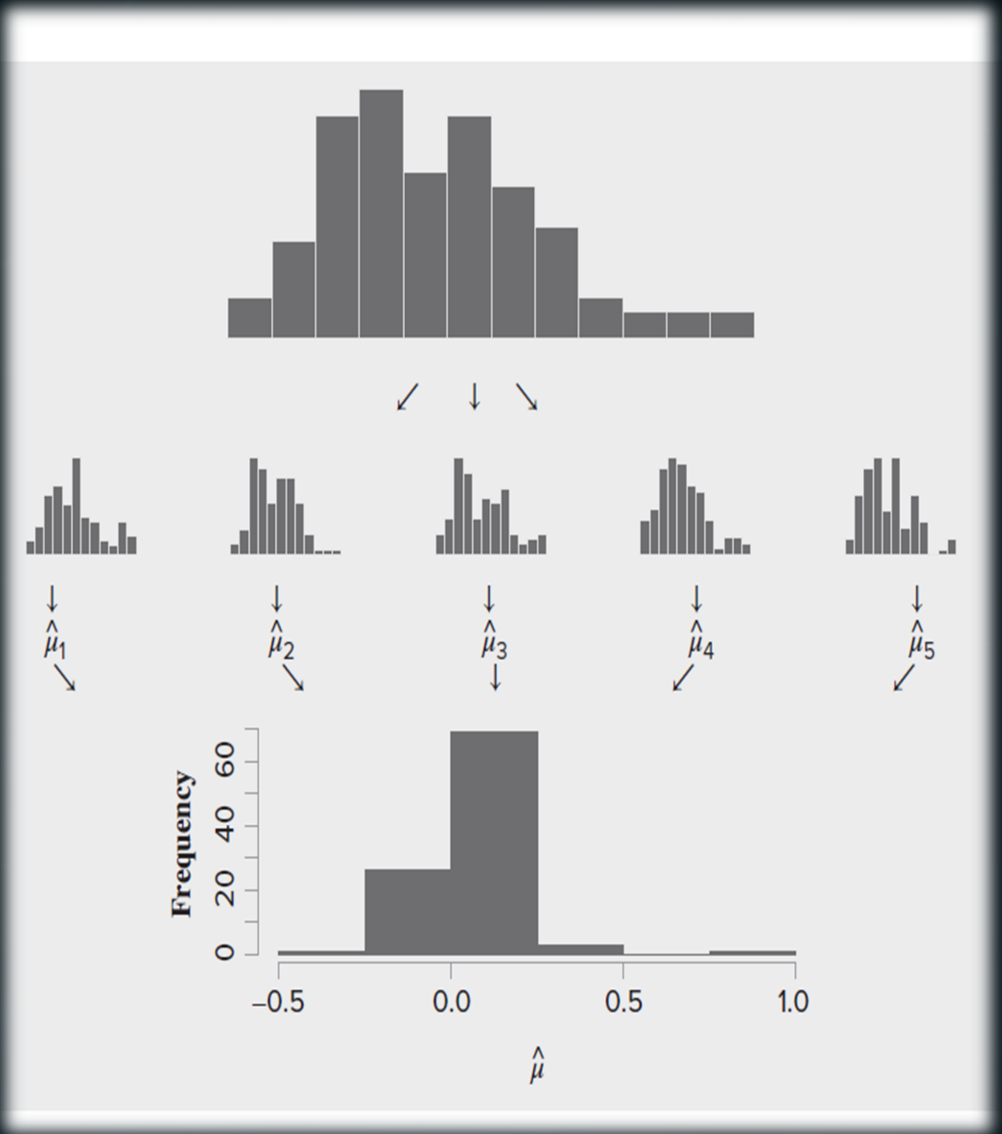
Robust standard errors: Bootstrapping and HC0 Standard Errors

Bootstrapping

Bootstrapping is a resampling technique that generates multiple datasets by randomly sampling with replacement from the original dataset.

HC0

HC0 standard errors address heteroscedasticity, a condition where the variance of the errors is not constant across all observations.



Comparing Standard Errors: OLS, Bootstrap, and HC0

10.2257

OLS

The standard error for music sentiment in the OLS model is 10.2257.

15.52354

Bootstrap

The bootstrap method produces a higher standard error of 15.52354, indicating greater uncertainty in the estimate.

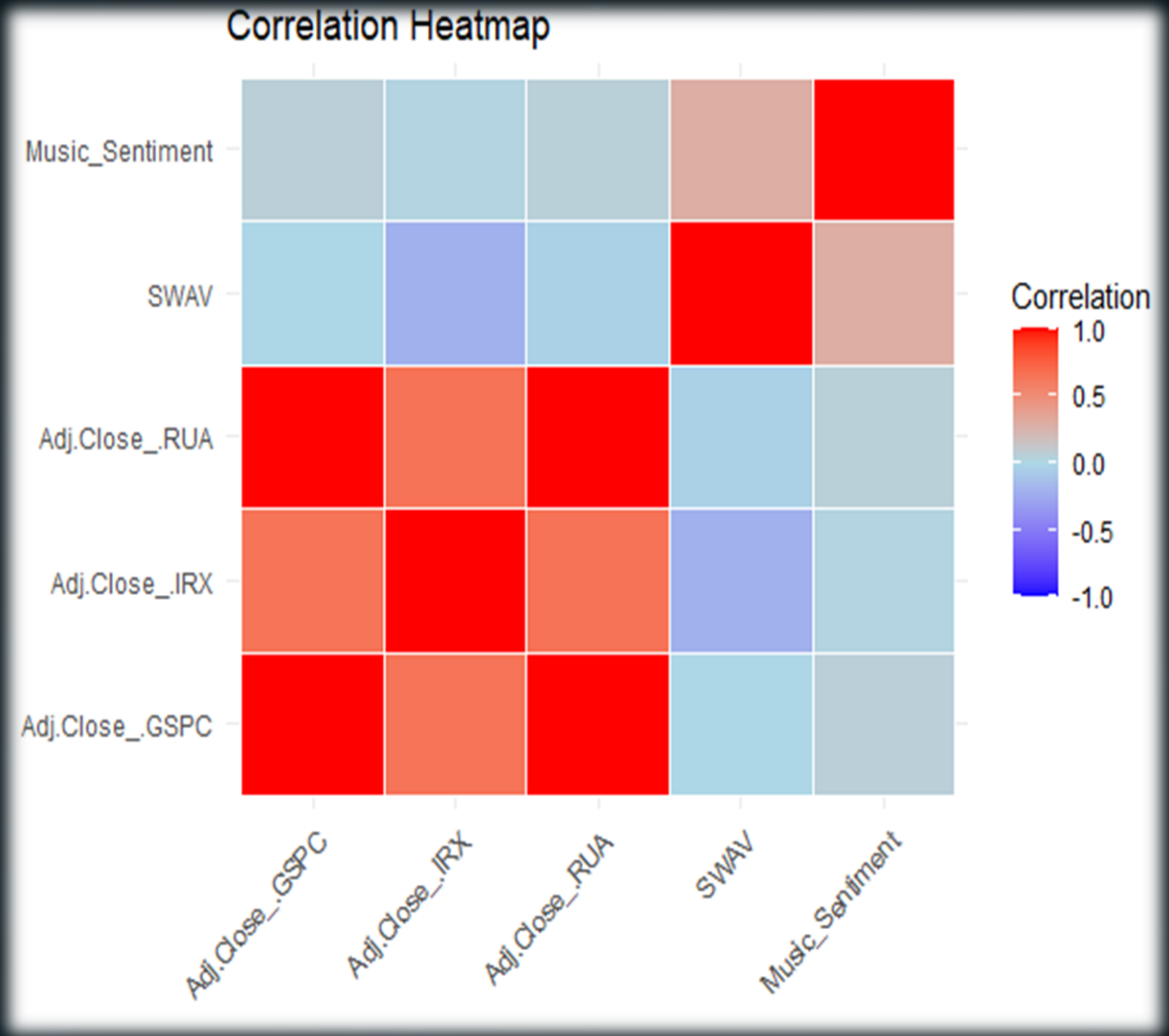
13.67312

HC0

The HC0 method provides a standard error of 13.67312, which is also higher than OLS, suggesting potential underestimation of standard errors by the original OLS model.

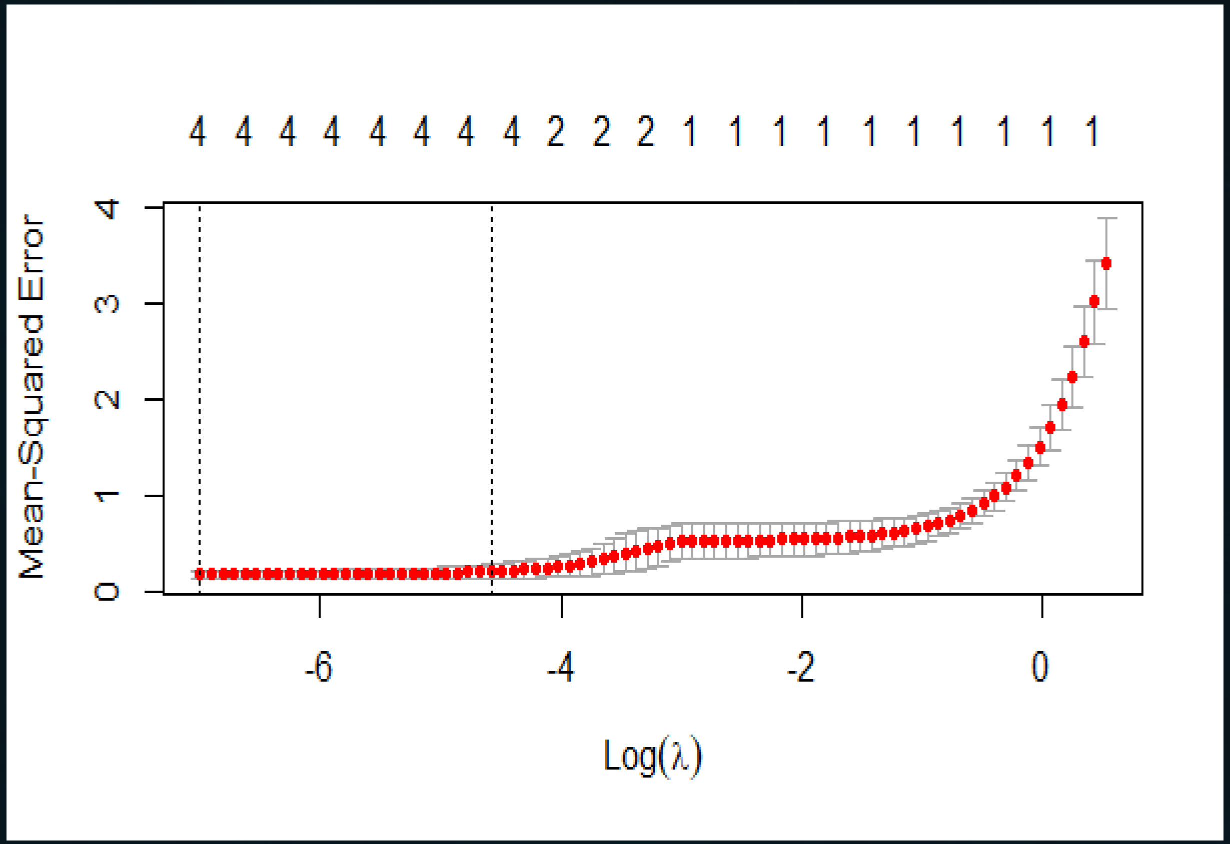
Reasons why OLS could have underestimated the standard errors:

1. Small sample size.
2. Autocorrelation.



Model	Standard Error (OLS)	Bootstrap Std. Error	Standard Error (HC0)	P-Value (OLS)	P-Value (HC0)
GSPC Returns (Contemp)	10.22095	15.91051	13.64994	0.3842	0.51348
RUA Returns (Contemp)	10.22519	15.63012	13.87782	0.3082	0.45125
GSPC Risk-Adjusted Returns (Contemp)	10.22568	15.60611	13.67312	0.3835	0.51338
RUA Risk-Adjusted Returns (Contemp)	10.22971	16.00841	13.90015	0.3080	0.45124
GSPC Returns (Lagged)	10.2475	11.32415	9.43220	0.8610	0.84882
RUA Returns (Lagged)	10.27400	11.07718	9.40695	0.8482	0.83408
GSPC Risk-Adjusted Returns (Lagged)	10.25226	11.20277	9.44853	0.8595	0.84747
RUA Risk-Adjusted Returns (Lagged)	10.27855	11.27499	9.42257	0.8469	0.83290

Lasso Model Selection Results:



1 S&P 500

Lasso analysis resulted in a lambda.min value of **0.313**, which led to the music sentiment coefficient being shrunk to zero. This indicates that music sentiment is not a significant predictor of S&P 500 returns.

2 Russell 3000

Similar results were observed, with a lambda.min of **0.262**. The Lasso algorithm reduced the music sentiment coefficient to zero, confirming the insignificance of music sentiment in explaining Russell 3000 stock returns.

Robustness Checks: Impact of Control Variables

1

Control Variables

We conducted robustness checks by adding control variables (lagged Music sentiment)to the same model, which account for previous week’s sentiment that may influence stock returns.

2

S&P 500 with Controls

The S&P 500 model with control variables yielded a music sentiment coefficient of -9.066, while the model without controls had a coefficient of -6.8015. This difference is not statistically significant, suggesting that the inclusion of control variables does not substantially alter the results.

Double Machine Learning: Causal Inference

Causal Inference

Double Machine Learning is a powerful framework for causal inference, allowing for robust estimation of treatment effects while controlling for potential confounding variables. It utilizes machine learning algorithms to estimate the treatment and control groups, reducing bias and improving accuracy.

Robustness Across Models

In the S&P 500 model, Double Machine Learning produced a music sentiment coefficient of **-9.096**, with a standard error of **10.340** and a **p-value of 0.38**. This result further supports the conclusion that music sentiment has a non-significant impact on stock returns.

1

2

3

Variables

Treatment variable: Music Sentiment

Outcome variable: Returns

Confounders: Lagged Returns, Lagged Music Sentiment.

Conclusion: Music Sentiment's Negligible Predictive Power

This analysis using multiple methods, including OLS, Lasso, and Double Machine Learning, consistently demonstrates that **music sentiment has negligible predictive power** regarding stock returns across the indices, S&P 500 and Russell 3000.

Our findings does not align with the results for same week in the research paper. Although the coefficient being negative for previous week aligns with the paper, the **p-values in our analysis clearly shows that our results are clearly insignificant. Maybe due to small sample size, autocorrelation and due to various factors not being in our dataset, the results do not significantly match with the results from the paper except for a miniscule similarity in previous weeks returns findings.**

Thank you

Prof. Hongwei (Harry) Zhu – Fall 2024