

Emotional Trends in Popular Music and U.S. Economic Conditions

Jeremy Han, Tianyu Zhao, Junyi (Daniel Tang), Hillary Metcheke, Bryce Grover

1 Abstract

2 Introduction

3 Related Work

4 Data

4.0.1 Data Collection

4.0.2 Data Preprocessing

5 Study I: Longitudinal Analysis of Musical Sentiment

6 Study II: Economic Volatility and the Decoupling of Cultural Sentiment

6.1 Structural Instability in the Streaming Economy: A Time Series Approach

6.1.1 Motivation and Research Questions

While Study I focused on how music changed, this section shifts to the economic side of the music industry. The COVID-19 pandemic was a health crisis, but it also caused a major shock to the economy. We wanted to investigate two main questions regarding the financial performance of the sector:

- How were the stock prices of music streaming companies affected during the COVID-19 period?
- Did Spotify's stock exhibit abnormal volatility or structural changes during the pandemic compared to before?

We hypothesized that even if the content of the music (sentiment) stayed stable, the market valuation (stock price) would show distinct signs of instability and “abnormal volatility” due to the pandemic.

6.1.2 Closing Stock Price Inspection

We first plotted Spotify's monthly closing stock prices from 2018 to 2021 to get a general sense of how the market behaved over time.

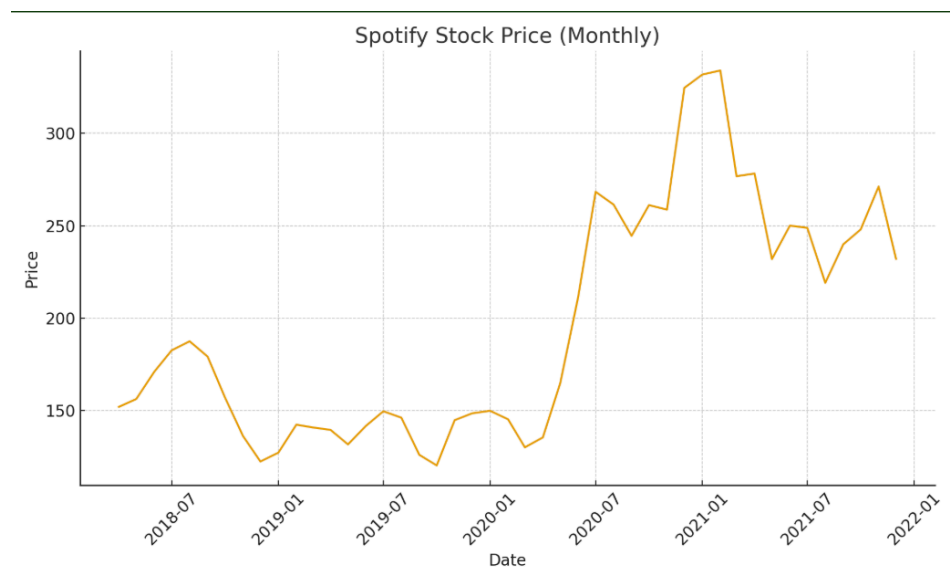


Figure 1: Original Spotify Stock Price Trend (2018-2021)

From the figure Figure 1, it is clear that the price does not follow a smooth or consistent linear trend.

1. We first plotted Spotify's monthly closing stock prices from 2018 to 2021 to get a general sense of how the market behaved over time. From the figure, it is clear that the price does not follow a smooth or consistent linear trend.
2. Around the start of the pandemic in early 2020, the pattern changes noticeably. The stock price increases rapidly and in a non-linear way, rising from roughly \$150 to above \$300 within about a year. This sharp shift suggests a structural break rather than a continuation of the earlier trend.

That said, a line chart alone cannot tell us whether this growth was smooth or whether it came with increased instability or volatility.

6.1.3 Decomposition

After examining the trend, we knew the price had increased, but we needed to know how it went up. Was it just a normal seasonal thing? Or was something broken? To figure this out, we used decomposition.

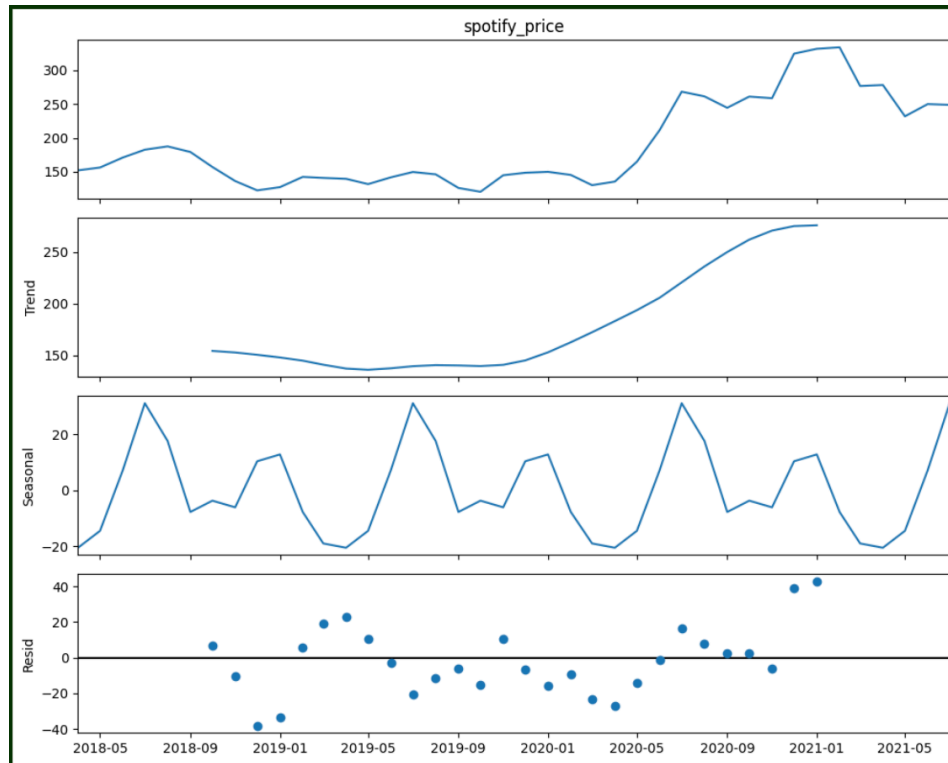


Figure 2: decomposition of Spotify Stock Price Trend (2018-2021)

We plotted these components in Figure Figure 2, and here is what we discovered:

- **Trend:** This part of the decomposition merely illustrates the general direction. It confirms what we already knew—the price started skyrocketing in 2020. It was not a straight line, instead, it curved upwards very fast.
- **Seasonality:** We thought maybe there would be a pattern, like sales going up every Christmas. However, if you examine the y-axis, the numbers are relatively small compared to the trend. The waves are present, but they do not significantly impact the price. This indicates that the substantial price jump was not due to the time of year.

- Residuals: In a healthy market, residual should be small and randomly scattered around zero. However, upon inspecting the bottom panel, we see a distinct cluster of huge spikes starting in 2020.

Seeing those big, clustered spikes in the bottom graph proves that the market was not stable. This gave us the proof we needed to say that the volatility was abnormal during COVID.

6.1.4 Data Transformation

We noticed that in the decomposition section, it creates a problem for our modeling. Models like ARIMA assume the data is stationary, which basically means the average and the spread of the data shouldn't change over time. However, looking at our stock price charts, the data clearly breaks this rule. It goes up and down wildly. So, we need to conduct a preprocessing step to make it stationary

6.1.4.1 Lag & ACF

We first plotted the Lag Plots and the Autocorrelation Function to look at the internal patterns.

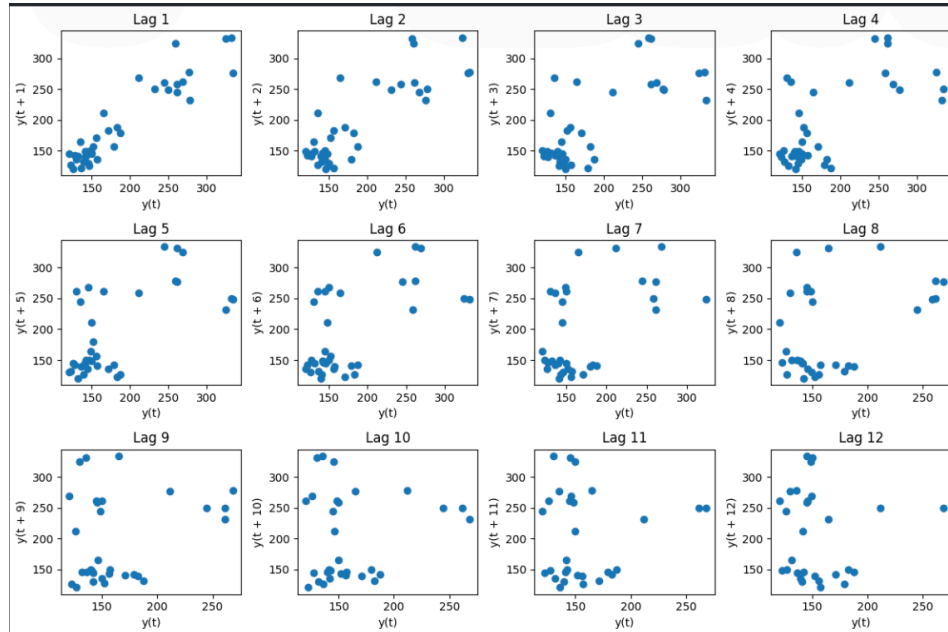


Figure 3: Lag plot of the original time series

Base on the figures we notices:

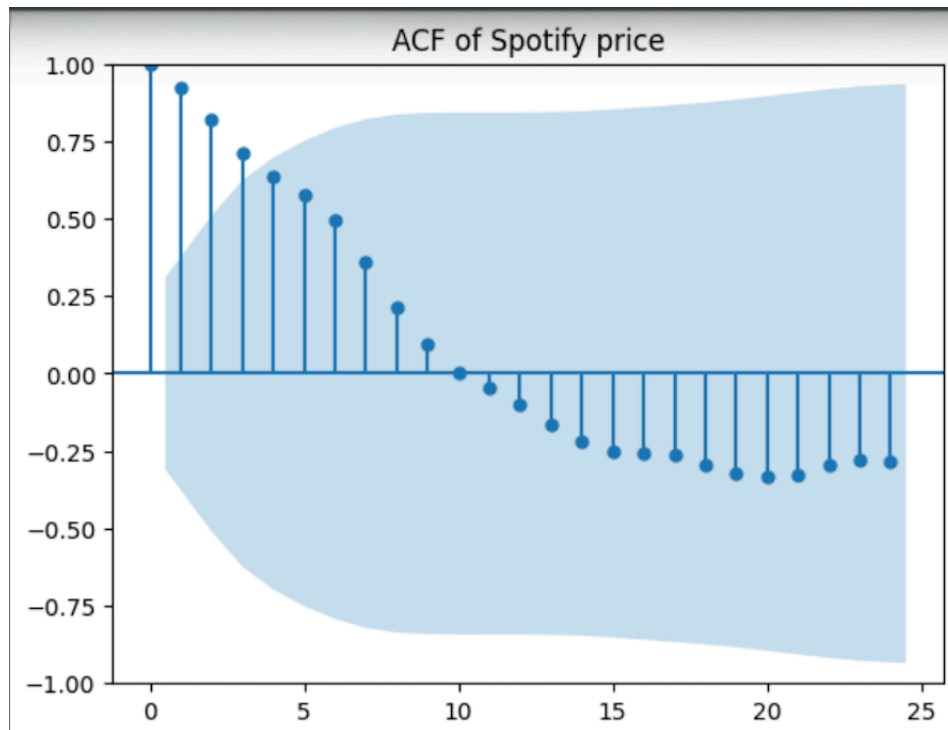


Figure 4: autocorrelation function of original ts

- Lag Plots: Strong positive relationship is observed at lag 1 and lag 2. This means today's price is almost perfectly predicted by yesterday's price.
- ACF: The autocorrelation bars decay very slowly and stay outside the confidence bounds for many lags. This usually happens when the data is not stationary and shows strong persistence. In this case, the slow decay likely means that the overall trend is very strong and is affecting the series, making the short-term changes harder to see.

6.1.4.2 Step 1: Log-Transformation

The first thing we noticed was the variance. There was a strange pattern here. When the stock price was low in 2018, the movements were pretty small. But once the price became much higher around 2020, the fluctuations also became much bigger.

To fix this heteroscedasticity, we applied a Logarithmic Transformation.

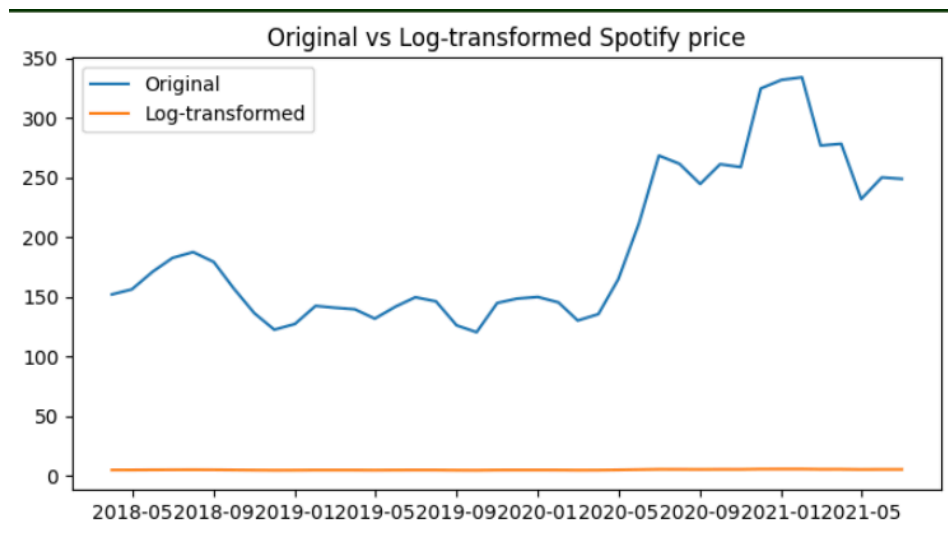


Figure 5: log transformation of original ts

- The blue line is the original price. You can clearly see the huge, messy spike in 2020 where the variance explodes.
- The orange line is the log-transformed price. It looks much smoother and more consistent.

Applying a logarithmic transformation reduces the impact of large values by compressing the scale of the data. Price movements are interpreted in relative terms rather than absolute dollar amounts, making the series more suitable for modeling.

6.1.4.3 Step 2: First-Order Differencing

At first glance (looking back at log transformation), the log-transformed data (orange line) might appear “flat” enough. However, this is deceptive. While the variance was stabilized, the series still retained a deterministic trend—the mean value was drifting upwards over time as the company grew. For an ARIMA model to be valid, the data must not just be stable in width, but also horizontal in direction (stationary mean).

To rigorously remove this remaining trend, we applied First-Order Differencing.

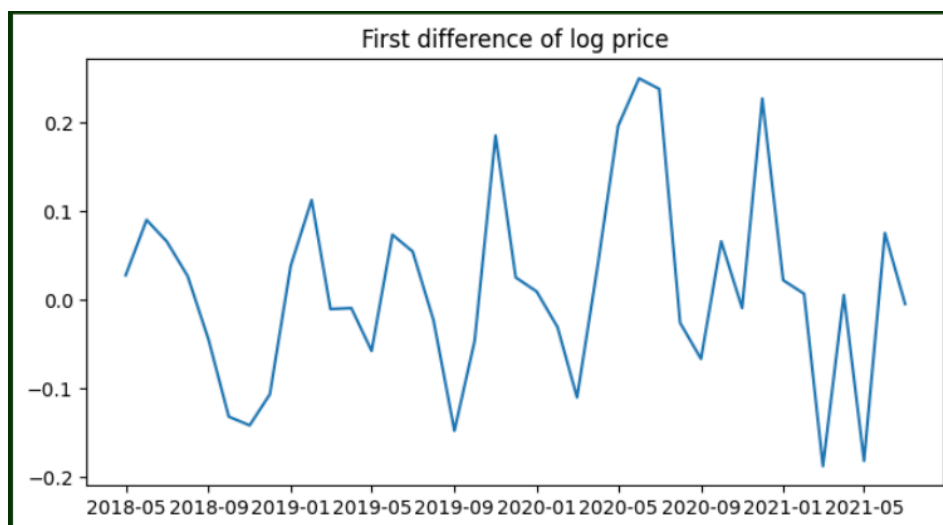


Figure 6: first difference ts

By analyzing the change from one month to the next rather than the raw value, we isolated the stochastic component.

Figure Figure 6 confirms the necessity of this step.

Unlike the log-series which drifted upwards, the differenced series oscillates consistently around zero. This transformation successfully detrended the data, leaving us with a pure “growth rate” metric that satisfies the strict stationarity requirements for autoregressive modeling.

6.1.5 Verifying the Result

We conduct the Augmented Dickey-Fuller (ADF) test on the before and after transformation series:

Table 1: ADF Stationarity Test Comparison

Series	Test Statistic	p-value	Conclusion
Original Stock Price (P_t)	-1.96	0.304	Non-Stationary (Fail to reject H_0)
Transformed Series ($\Delta \ln P_t$)	-4.91	0.00003	Stationary (Reject H_0)

- Before Transformation: The p-value was around 0.30. Since this is way bigger than 0.05, it confirmed that our original data was definitely not stationary.
- After Transformation: After doing the log and the difference, the p-value dropped to 0.00003. Since 0.00003 is tiny, we can say for sure that the data is stationary now. We also checked the ACF plot one last time.

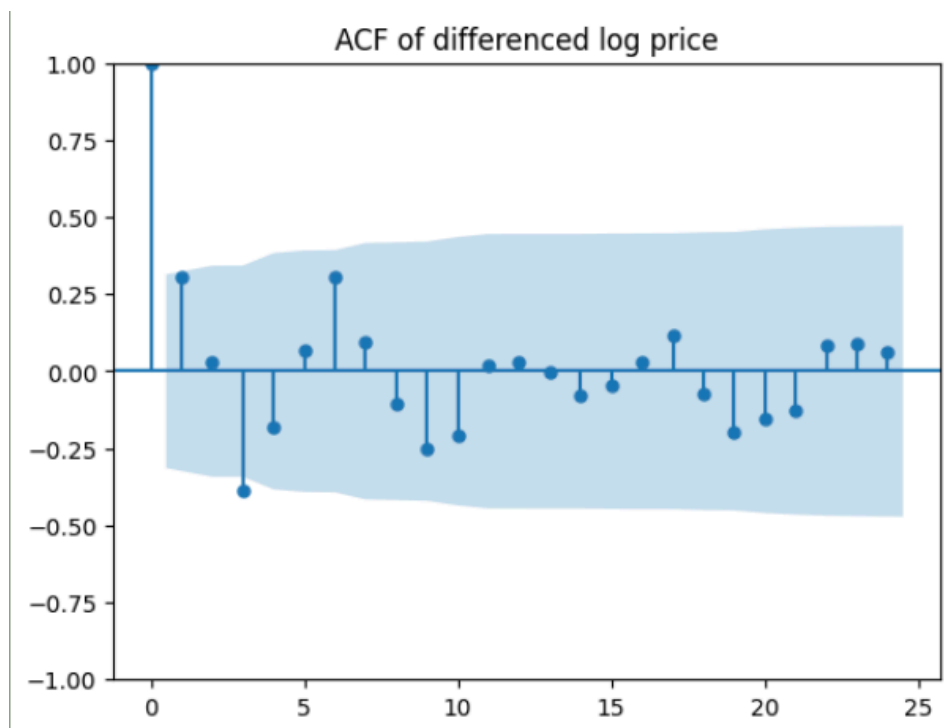


Figure 7: ACF after transformation ts

In Figure Figure 7, the difference is obvious. Unlike the first ACF plot that dragged on forever, this one cuts off really fast after the first lag. Most of the dots are inside the blue shaded area. This confirms that we successfully removed the trend and the unstable variance, so now we are finally ready to put this data into the ARIMA model.

6.1.6 ARIMA Modeling

Now that the data is stationary, we could finally run the ARIMA model. We didn't know which parameters would be perfect, so we ran a grid search to test different combinations. To pick the winner, we just looked at the AIC score—basically, the lower the number, the better the model.

Table 2: ARIMA Model Comparison

Model	AIC
ARIMA(1, 1, 0)	-60.172188
ARIMA(3, 1, 0)	-59.056107
ARIMA(3, 1, 1)	-57.422914

Looking at the table, the ARIMA(1, 1, 0) model gave us the lowest score of -60.17. We tried more complex models like ARIMA(3, 1, 0) and ARIMA(3, 1, 1), but their scores were higher (-59.05 and -57.42), so they weren't worth the extra complexity.

Table 3: SARIMAX Result

Parameter	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.3159	0.156	2.026	0.043	0.010	0.621
sigma2	0.0108	0.003	3.745	0.000	0.005	0.016

Based on the SARIMAX Result It showed an AR coefficient of 0.3159 with a p-value of 0.043, which means the relationship is statistically significant.

6.1.7 Residual Diagnostics

We plotted the residuals to see if the model missed anything.

1. Standardized Residuals: Figure 8 shows the reality. If you look at the top-left graph, the line is messy. You can see huge spikes around 2020 and 2021. This tells us that even with the best model, the variance during the pandemic was just too high to predict perfectly.
2. Q-Q Plot: Next, look at the Q-Q plot. We want the blue dots to sit on the red line. Most of them do, but look at the tails—the dots peel away from the line at both ends. This confirms “Heavy Tails.” It means extreme events happened way more often than a normal statistical model expects.

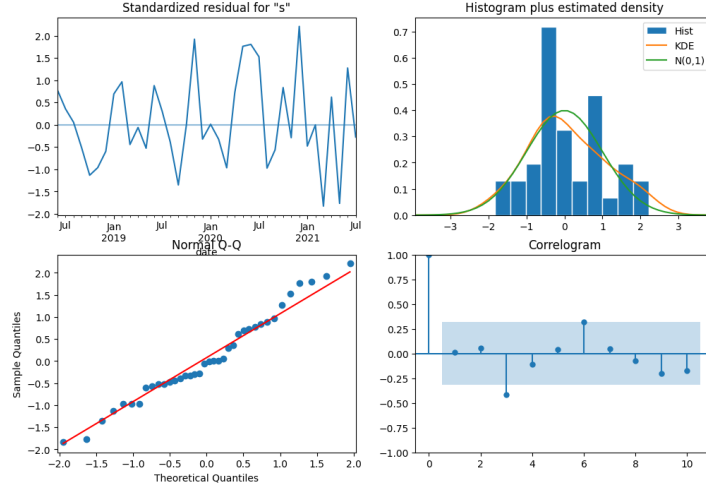


Figure 8: Residual Analysis

3. Correlogram: The only good news is in the bottom-right graph. All the dots stay inside the blue shaded area. This means there's no pattern left in the errors, so mathematically, our model did its job correctly.

6.1.8 Cross-Domain Dynamics: The Resilience of Cultural Sentiment

7 Conclusion

7.1 Key Findings

7.2 Limitations

7.3 Future Directions

8 References

- [1] Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., & Kaiser, Ł. (2018). Universal Transformers. *arXiv preprint arXiv:1807.03819*.