Draft Preview Continue

Style Button More

Edit email header / footer

A look at AI's impact on stocks

Add a subtitle...

jarvit

Introduction

Nvidia's growth at the beginning of the year has been interesting to observe; reaching all-time highs and continuing to grow, while even some of the giants of the S&P 500 have failed to perform even half as well (and that's being generous).

Alongside this, developments in the field of artificial intelligence (AI) have been astonishing, and progress seems to be occurring exponentially. As you may know, technology has been at the forefront of major developments for the past decade, and we do not expect this fact to change anytime soon. Moreover, AI could likely be the next breakthrough in technology which is why it is such an interesting topic to investigate.

In this blog, we will attempt to analyse AI's impact on stocks, namely Nvidia. Going further, we will seek to find what benefits this may have, whether that be predicting future stock performance.

Initial Overview

First, we need to understand both variables individually. To do this, we need to find a measure for both, which will represent them. For Nvidia, this was simple: you just look at the stock's close price to find how much the stock traded for at the end of a trading session. This gives a good indication of how much the stock was valued at while also enabling us to observe how this value changed over time. AI, on the other hand, is far more challenging as we need to find a metric that will allow us to accurately determine the growth of AI over time while at the same time being something we could measure. The metric chosen here was Google search volume for the term ‘AI’, hoping that there was a correlation between people's interest over time and advancements in the field.

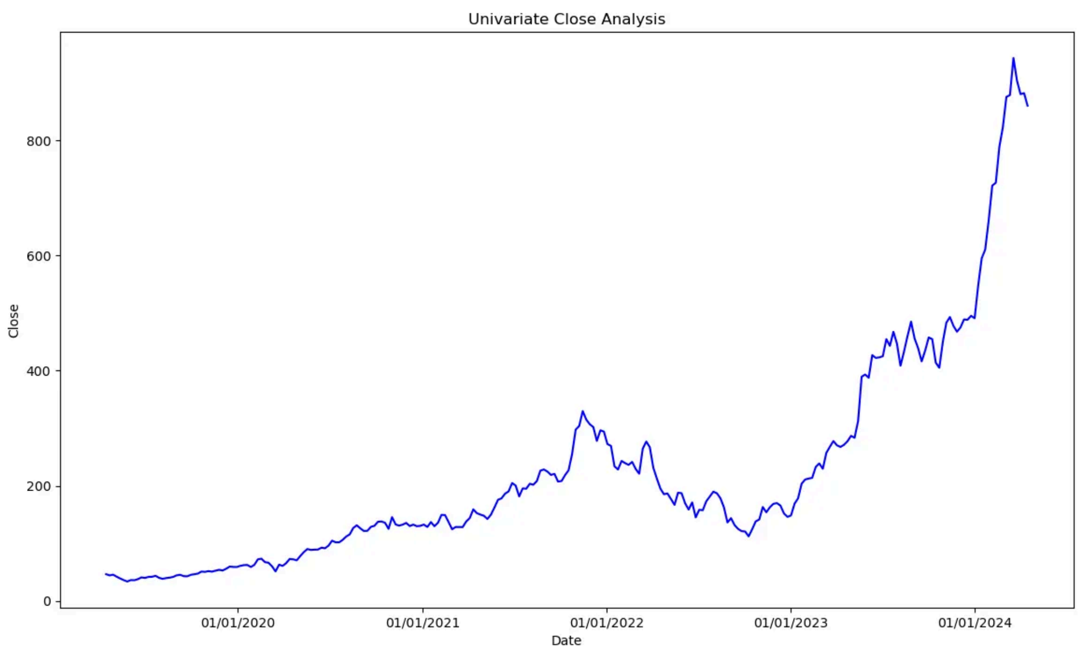

Figure.1 Line chart of a weekly average stock close price for 5 years.

We notice an overall l upward trend mid 2019 all the way up until 2022, where Nvidia had a poor year with a negative slope. That is until around 2023 when the stock price bounced back and grew rapidly, and then even more so in the early months of 2024.

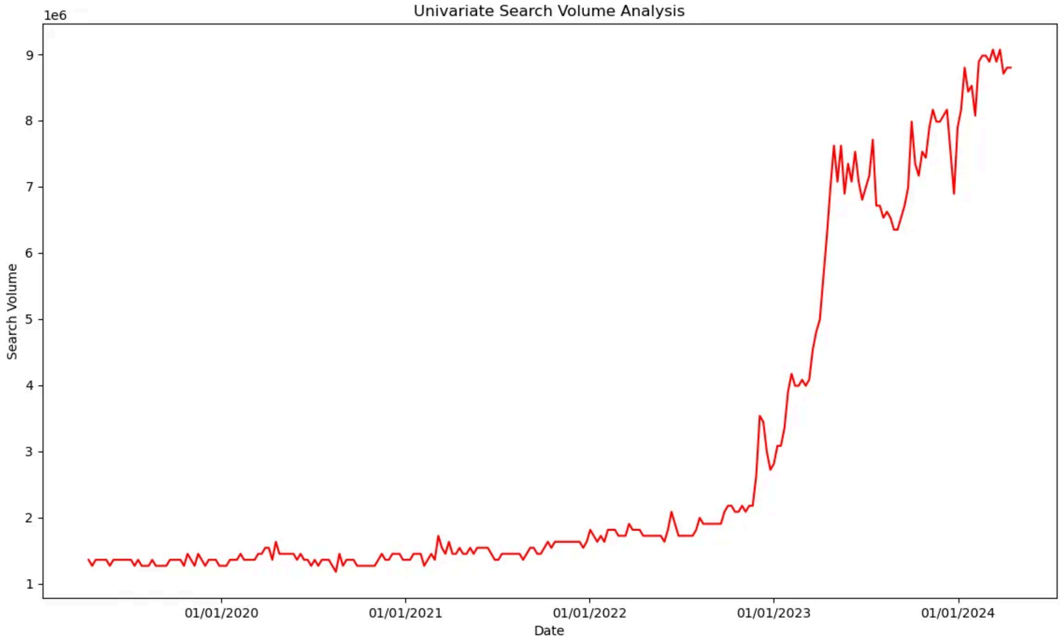

Figure.2 Line chart of a weekly average Google search volume data for the term ‘AI’.

For the first 3 years of data here, we have a relatively flat curve until early 2023, where we have a large spike resulting in search volume jumping from a low 2 million weekly average searches to a high of 7 million average.

With an initial observation of both plots, we notice a clear similarity between them. Starting with low slopes for the first 3 years, then large spikes in 2023.

AI and Nvidia Comparisons

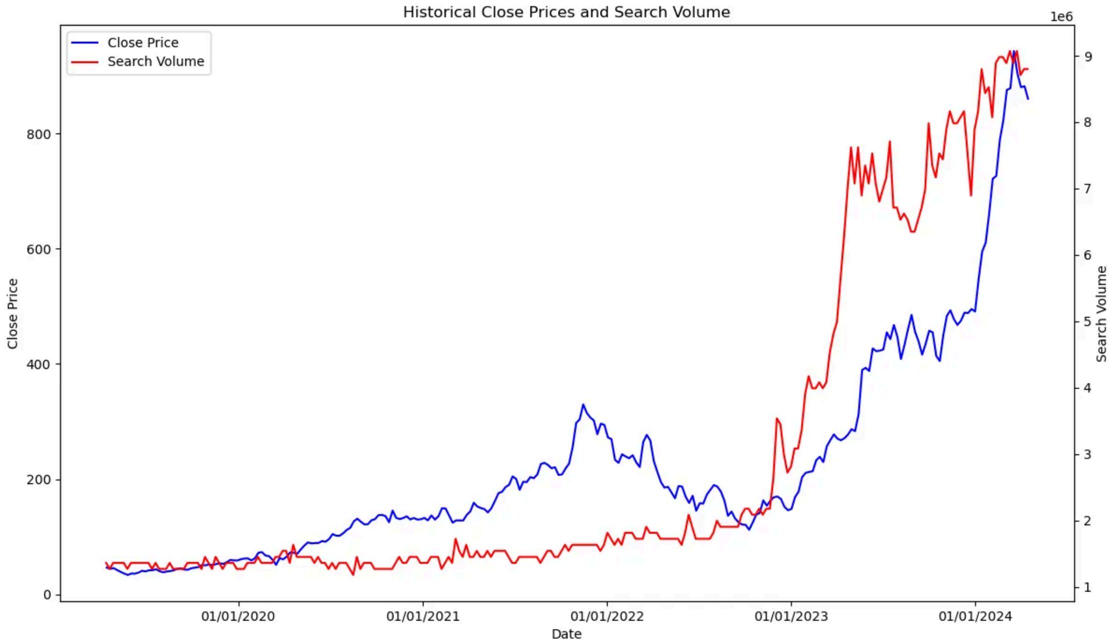

Figure.3: Side-by-side comparison of Nvidia close price and Google search volume.

Figure 3 clearly shows a high correlation between the two variables. Although at the start, there doesn't appear to be much correlation, as we approach 2023, we immediately see a spike in 'AI' interest followed by an increase in Nvidia's stock price. Looking even closer, we notice that the patterns are even more similar, with drops and flat periods occurring at almost the same time in both.

Let us dive in deeper at some of these dates:

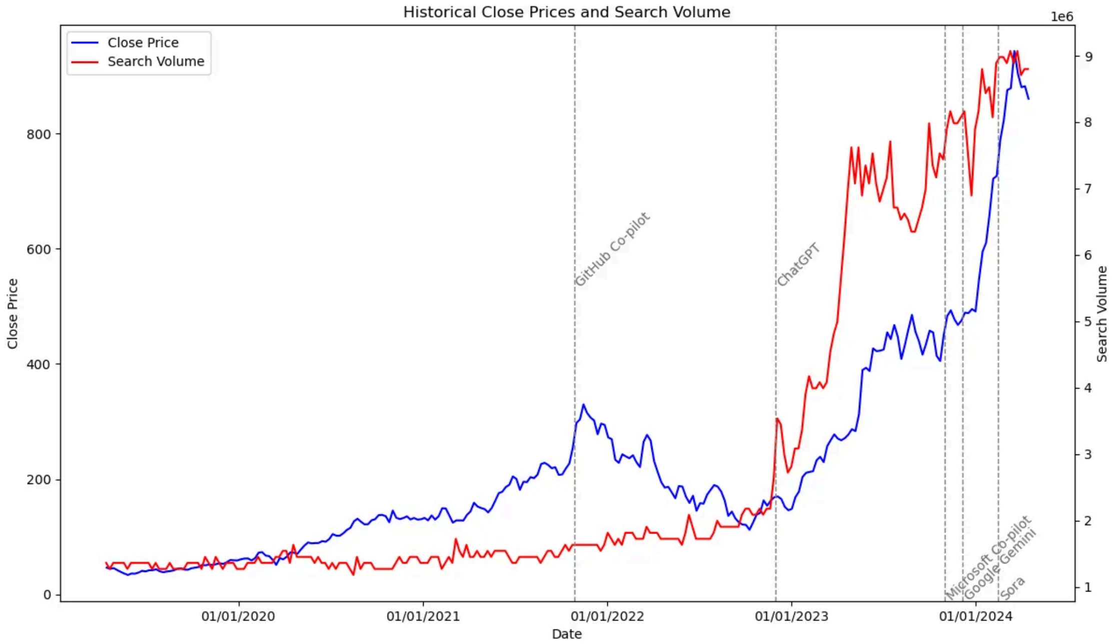

Figure.4: Shows dates when major commercial AI tools were released on the line chart.

We can see that many of these release dates coincide with some of the major spikes in both search volume and stock price, further indicating the effects that developments in the field of AI have.

The first event was the release of GitHub Co-pilot, which occurred right before the stock price peaked. Surprisingly, we did not observe any significant change in search volume as a result. This might lead us to believe that there isn't a correlation between the two variables. However, it is more likely that we don't see a correlation because our metric for AI growth is flawed. In our analysis, we opted for a basic measure to keep things simple. Therefore, it should be within our expectations that there will be limitations to this metric. This is something that we will need to account for in our analysis and consider for future research.

The most prominent of these is the release of ChatGPT in late 2022. The release of this tool was a major success in the industry and was a key component that made so many aware of the power of AI. Since then, there has been far more interest in the field of AI. Considering that AI has been growing, this means that companies like Nvidia, who now have a substantial focus on AI, will benefit from this. With this, we would expect then that companies that work with AI would have more interest in them, causing demand for the stock to increase, thus leading to an increase in the price of the stock.

Level Regression Model

We can analyse this further by running an OLS regression:

$$Close = \beta_0 + \beta_1 Search + \mu$$

Where:

· Close is the price of the stock at the end of the trading session

· β0 is the slope intercept of the regression

· β1is the coefficient of the independent variable Search

· Search is the weekly average absolute search volume of the term AI

· μ is the unexplained error term of the regression.

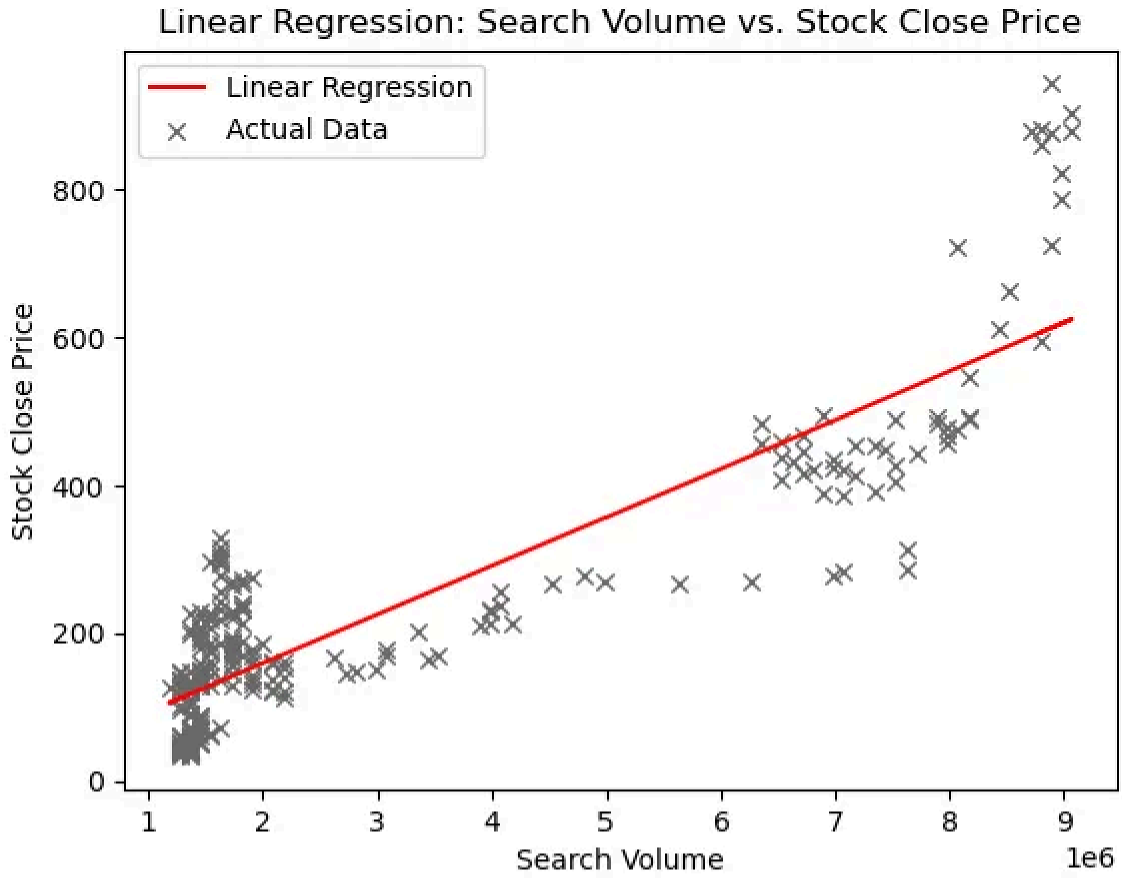

Figure.5 shows the data points and the least squares regression line that runs through them.

We can see that our data is heavily skewed right, with most of the data being on the left. Despite this, we can still see a clear positive slope and correlation of the data.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.779
Model:                            OLS   Adj. R-squared:                  0.778
Method:                 Least Squares   F-statistic:                     914.8
Date:                Wed, 24 Apr 2024   Prob (F-statistic):           3.96e-87
Time:                        20:06:22   Log-Likelihood:                -1543.0
No. Observations:                 262   AIC:                             3090.
Df Residuals:                     260   BIC:                             3097.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         28.4688      8.366      3.403      0.001      11.995      44.943
x1          6.572e-05   2.17e-06     30.246      0.000    6.14e-05       7e-05
==============================================================================
Omnibus:                       40.594   Durbin-Watson:                   0.064
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               63.567
Skew:                           0.903   Prob(JB):                     1.57e-14
Kurtosis:                       4.600   Cond. No.                     5.94e+06
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.94e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure.6: Result of the level linear regression.

From Figure 5, it is hard to tell what the actual results of the regression are and to interpret them in any meaningful way as the coefficient is so small. This is because the independent variable is very large.

To change this, we only have to make a small change to our regression formula, simply transforming from a level-level regression to a level-log regression:

Level-Log Model

$$NVDA = \beta_0 + \beta_1 \ln(Search) + \mu$$

Which means that the interpretation is slightly different. Now is interpreted as a 1% increase in the search volume is expected to increase the close price by $ (β1/100).

Changing the regression as such slightly changes the graphs.

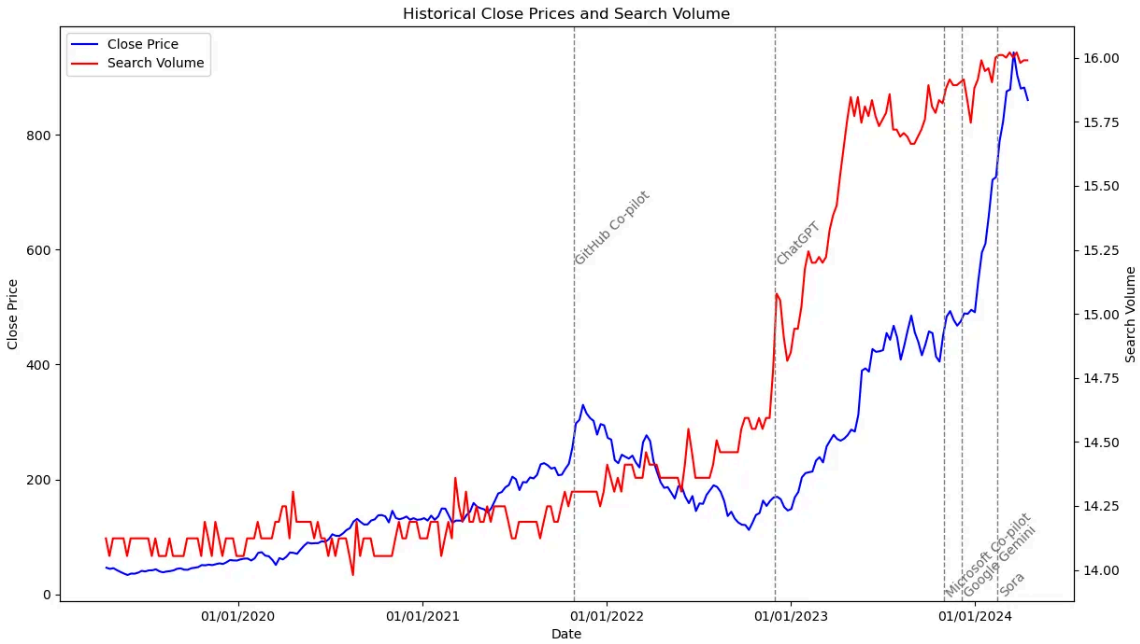

Figure.7: Comparison line plot of close price and search volume which has been transformed by natural log.

Figure 7 depicts a very similar plot to the absolute search volume plot we saw in Figure 4. However, we do notice a lot more variation in the independent variable now.

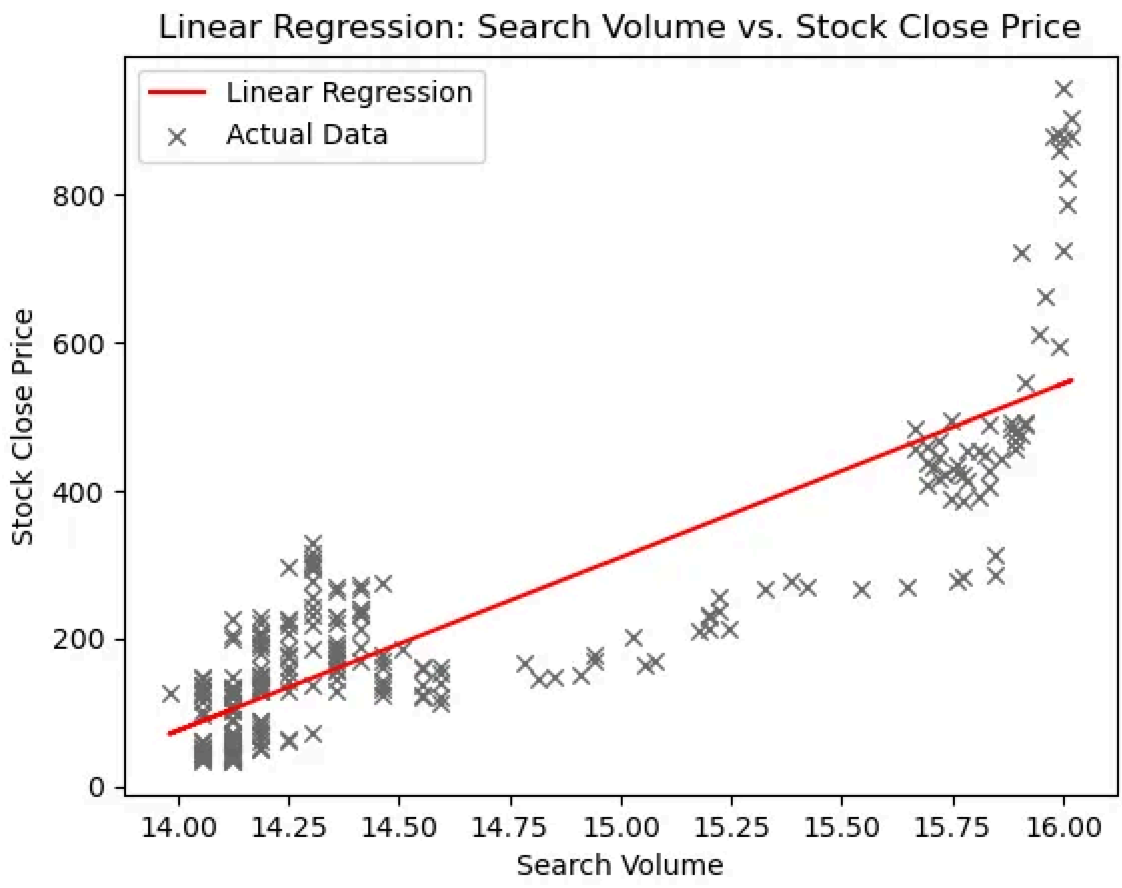

Figure.8: Level-log plot of the OLS regression.

Just looking at the data points in Figure 8, the positive slope is even more distinct, and the positive correlation is more easily visible, even without the regression line.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.725
Model:                            OLS   Adj. R-squared:                  0.724
Method:                 Least Squares   F-statistic:                     684.5
Date:                Wed, 24 Apr 2024   Prob (F-statistic):           8.54e-75
Time:                        20:18:26   Log-Likelihood:                -1571.6
No. Observations:                 262   AIC:                             3147.
Df Residuals:                     260   BIC:                             3154.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -3201.6927    130.967    -24.447      0.000   -3459.584   -2943.801
x1           234.1190      8.948     26.164      0.000     216.499     251.739
==============================================================================
Omnibus:                       66.055   Durbin-Watson:                   0.052
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              153.230
Skew:                           1.194   Prob(JB):                     5.33e-34
Kurtosis:                       5.887   Cond. No.                         319.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure.9: OLS regression results of the level-log model.

Additionally, reading and interpreting the results of the new model is far easier. Looking at the ‘x1’ coefficient or, we get a value of 234.4. This means we expect a 1% increase in search volume to increase the close price by $2.34. This is calculated with an R-squared of 0.723, which is very high, meaning 72.3% of the variation in close price can be explained by our independent search volume.

Finally, we can use a few metrics to determine whether our result is statistically significant. But let us just focus on one, the 95% confidence interval, which has our in the bounds. Meaning we are 95% confident that the true population mean lies within this interval.

With that said, we cannot say that the growth of AI directly caused Nvidia stock prices to rise. However, looking into Nvidia here, we see that in the last few years, their focus has shifted from gaming services to data centres which they sell to large companies, many of which are now used to train AI models. This has resulted in the data centre revenues far surpassing any other streams for them. Major increases in revenue like this are exactly what investors like to see and are likely why we have seen such a large increase in stock price.

Improved Model (with the market)

The question then becomes: Can we use this to predict future stock performance?

Using multiple linear regression, we can further understand the impact that AI has on Nvidia's stock by adding another independent variable. We will use the S&P 500 index price to represent the market, allowing us to capture the effects the market has on Nvidia's stock price:

$$NVDA = \beta_0 + \beta_1 \ln(Search) + \beta_2 S\&P500 + \mu$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.848
Model:                            OLS   Adj. R-squared:                  0.846
Method:                 Least Squares   F-statistic:                     720.5
Date:                Wed, 24 Apr 2024   Prob (F-statistic):         1.50e-106
Time:                        20:20:29   Log-Likelihood:                -1494.1
No. Observations:                 262   AIC:                             2994.
Df Residuals:                     259   BIC:                             3005.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -2566.9480    107.041    -23.981      0.000   -2777.730   -2356.166
x1           156.3712      8.568     18.250      0.000     139.499     173.243
x2             1.2919      0.089     14.456      0.000       1.116       1.468
==============================================================================
Omnibus:                      103.416   Durbin-Watson:                   0.057
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              446.523
Skew:                           1.590   Prob(JB):                     1.09e-97
Kurtosis:                       8.549   Cond. No.                     9.39e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.39e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure.10 OLS summary statistics from Nvidia's stock price being regressed on 'AI' search volume and the S&P 500 index close price.

Adding this market variable increases our adjusted R-squared, while also having a coefficient that is not zero, indicating significance and the importance of including it in the model. Our model now explains nearly 85% of the variation in Nvidia's stock price.

The introduction of the market variable has reduced the coefficient of the search volume variable in the regression, suggesting some overlap between search volume and the S&P 500. However, is still very significant.

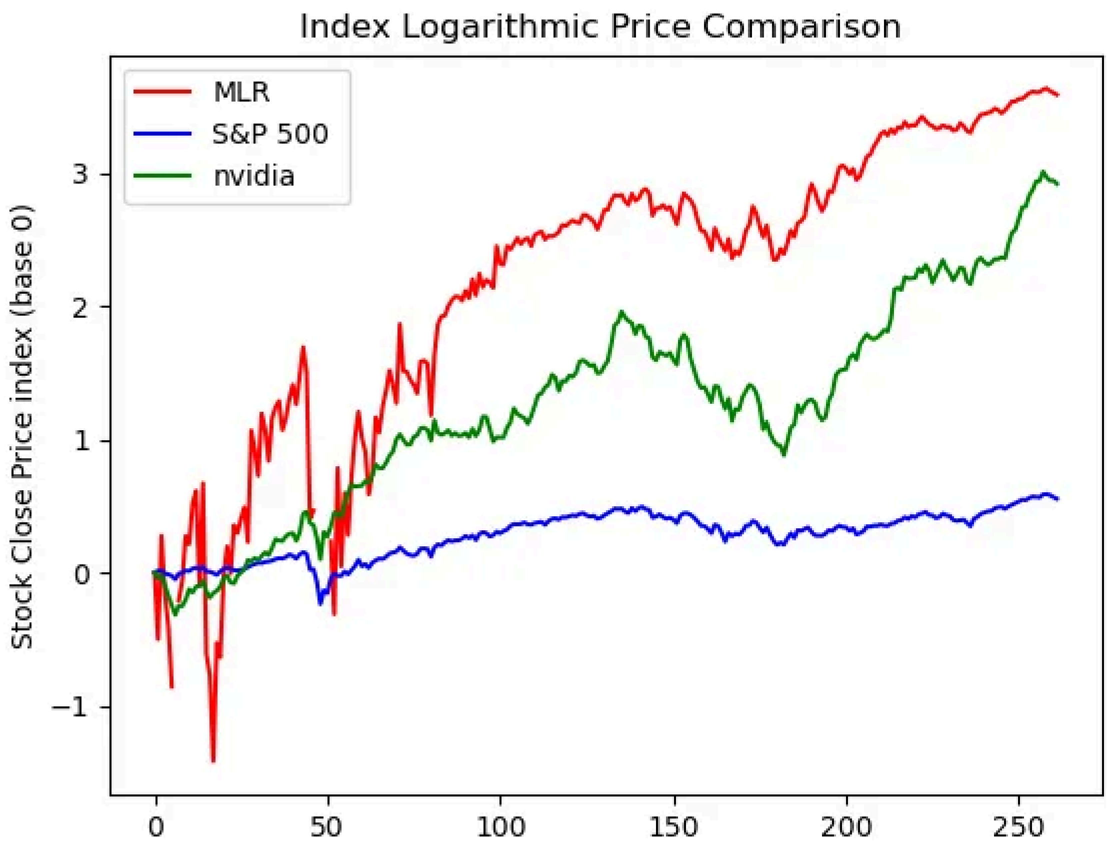

Figure.11 Logarithmic index comparison of Nvidia's close price data, S&P 500 close price data, and a multilinear regression of S&P 500 and AI to predict Nvidia's stock price.

Upon closer inspection, we can see how our new model performs against the actual data and also against the market. There is a considerable amount of variation at the start of our data; however, later on, we notice that the predicted values closely follow the trend of the real data. Nonetheless, the regression seems to systematically overpredict the price.

While there is high volatility at the start in the regressed model, which may not promise immediate results, the latter section of the graph indicates the possibility that we could use this regression to predict changes in the price of Nvidia stock, albeit with consideration of the model's tendency to overpredict.