UNDERSTANDING THE EFFECT OF ECONOMIC INDICATORS ON THE STOCK MARKET RETURNS

ADVANCED DATA SCIENCE AND PYTHON FOR FINANCE PROJECT - GROUP 7

Javeria Malik, Varad Gandhi, Kesava Ram Raghavan, Manas Tokale, Tarun Raj

Aim of the Project

- The aim of the Project is to understand and analyze the effect of macroeconomic factors on the US stock market.
- The US Market has been quite sensitive to financial indicators of the economy and is heavily reliant on the combined performance of diversified sectors.
- Our aim in this project is to develop a model for analyzing the effect of these factors and come to a conclusion as to which market is better for investing based on different risk appetites.

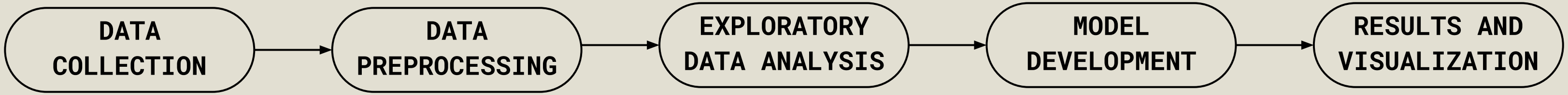
For convenience, we are going to work on the following data -

Stock Markets
S&P 500
NASDAQ COMPOSITE
DOW JONES INDUSTRIAL AVERAGE
RUSSELL 3000

Economic Indicators
Consumer Price Index, Personal Consumption Expenditures, Economy money supply, Unemployment rate, Inflation, GDP

ANALYSIS WORKFLOW AND CODE

Data Analysis Workflow



Major sources -

Yahoo finance

Federal Reserve Economic Development

Major operations -

Cleaning Data

Merging data

Major operations -

Correlation Matrices

Data viz

Major operations -

Multiple Linear Regression model

Major analysis -

Impact analysis

Significant effect analysis

Data Collection

The data is collected from two sources -

STOCK DATA - Yahoo finance

ECONOMIC DATA - FRED website

The extracted data is then read and stored into variables.

For the stock market data, the data right from its date of availability to the current available date is extracted from Yahoo finance.

Importing the necessary libraries

```
In [4]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import yfinance as yf
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Loading the data from the individual datasets

```
In [6]: cpi = pd.read_csv('CPI.csv', parse_dates=['DATE'])
gdp = pd.read_csv('GDP.csv', parse_dates=['DATE'])
inflation = pd.read_csv('Inflation.csv', parse_dates=['DATE'])
pce = pd.read_csv('PCE.csv', parse_dates=['DATE'])
unemployment = pd.read_csv('Unemployment RATE.csv', parse_dates=['DATE'])
wm2ns = pd.read_csv('WM2NS.csv', parse_dates=['DATE'])
```

Downloading the S&P 500 data from yahoo finance

```
In [8]: # Downloading the Adj Close prices of S&P 500 data
stock_data = yf.download('^GSPC', start=cpi['DATE'].min(), end=cpi['DATE'].max())
stock_data
```

Preprocessing the Data

```
stock_data.drop(['Open','High','Low','Close','Volume'],axis=1,inplace=True)
```

```
stock_data
```

Calculating the returns from the data

```
# Use adjusted close prices to calculate returns
stock_data['Returns'] = stock_data['Adj Close'].pct_change() * 100
```

```
stock_data
```

Now merging the data together

```
data = cpi.merge(gdp, on='DATE', how='outer')
data = data.merge(inflation, on='DATE', how='outer')
data = data.merge(pce, on='DATE', how='outer')
data = data.merge(unemployment, on='DATE', how='outer')
data = data.merge(wm2ns, on='DATE', how='outer')
```

```
# Merge the datasets on the date
merged_data = pd.merge(stock_data, data, left_on='Date', right_on='DATE', how='inner')
```

Exploratory Data Analysis

Creating a correlation matrix

```
# Convert the dataframe to numeric, forcing errors to NaN
df2 = df.apply(pd.to_numeric, errors='coerce')

# Now calculate the correlation matrix
corr_matrix = df2.corr()

# Visualize the correlation matrix
plt.figure(figsize=(12, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix of Economic Indicators and Stock Returns')
plt.show()
```

In order to visualize the data, the correlation matrix is used and above is the code for the same. The seaborn functionality is utilized to visualize it and understand the effect of the economic indicators on these markets and their performance.

Model Development

Splitting the data into train and test sets for generating the linear regression model

```
X = df[['CPIAUCSL', 'GDP', 'CORESTICKM159SFRBATL', 'PCE', 'UNRATE', 'WM2NS']]  # Economic indicators
y = df['Adj Close']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Developing the Linear regression model

```
# Create and train the model
model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse}')
```

```
Mean Squared Error: 43909.76583857999
```

Identifying the coefficients for the regression model

```
coefficients = pd.DataFrame(model.coef_, X.columns, columns=['Coefficient'])
coefficients
```

```
print(f"R-squared: {model.score(X, y)}")
```

Results and Visualization

Plotting the actual vs Predicted stock returns

```
plt.scatter(y_test, y_pred)
plt.xlabel('Actual Stock Returns')
plt.ylabel('Predicted Stock Returns')
plt.title('Actual vs Predicted Stock Returns')
plt.show()
```

Plotting the actual vs predicted values

```
# Plot the original vs predicted values
plt.figure(figsize=(10, 6))

# Plotting the actual stock prices
plt.plot(y_test.values, label='Original', color='blue', linewidth=2)

# Plotting the predicted stock prices
plt.plot(y_pred, label='Predicted', color='red', linestyle='--', linewidth=2)

# Adding labels and title
plt.title('Original vs Predicted Stock Prices')
plt.xlabel('Index')
plt.ylabel('Stock Price')
plt.legend()

# Display the plot
plt.show()
```

OUTPUTS AND VISUALIZATION

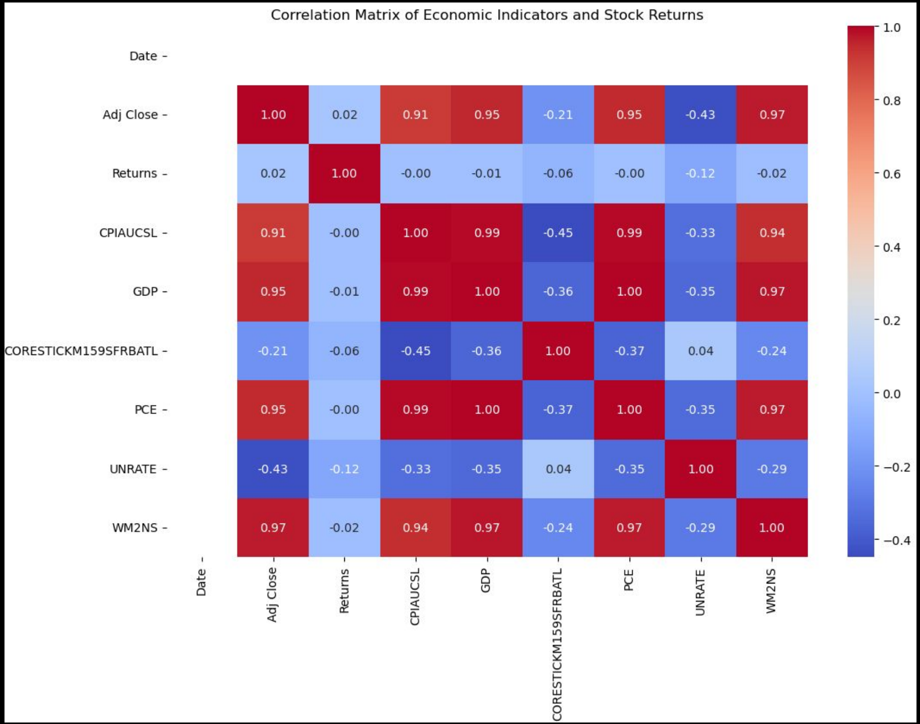
Correlation Analysis of Economic Indicators and Stock Returns

Strong Correlations:

- **GDP** and **PCE** are very strongly correlated with stock prices - directly impact corporate profits, influencing stock prices.
- **Money Supply** is another important indicator with strong positive correlations - Increased money supply tends to boost stock prices and consumption.

Weak or Negative Correlations:

- **Unemployment Rate** has weak negative correlations with stock returns - indicating a less direct impact on stock performance.
- **Core Inflation** shows a weak negative correlation with stock returns, suggesting limited influence on short-term market movements.



Model Performance Evaluation: Actual vs Predicted Stock Returns and Prices

Model Performance Metrics:

- **Mean Squared Error (MSE):** 4,309.77 - A moderate MSE indicates that the model's predictions are somewhat accurate, but there are areas where the model's prediction can be improved. Lower values of MSE would indicate better model performance.
- **R-squared:** 0.97 - The model explains 97% of the variance in the stock returns, suggesting excellent fit and that the economic indicators used are strong predictors of stock returns.

Actual vs Predicted Stock Returns (First Graph):

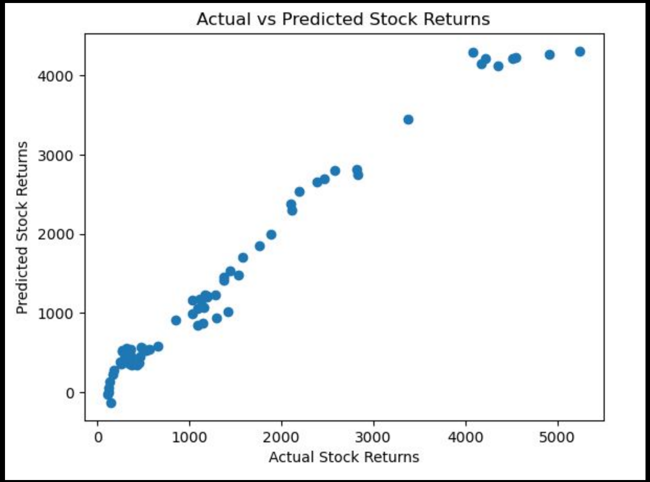
- Points show a strong linear relationship between actual and predicted stock returns.
- Some outliers are observed, with predicted values being higher than actual in certain instances, which is expected for stock market predictions.

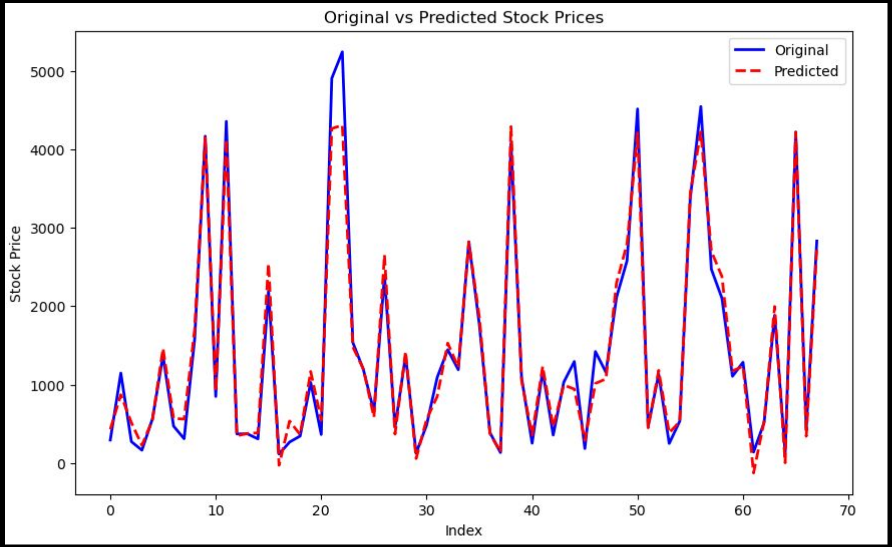
Original vs Predicted Stock Prices (Second Graph):

- Both the actual and predicted stock prices follow similar trends with minimal divergence, indicating the model’s good predictive capability.

Mean Squared Error: 4309.76583

R-squared: 0.967827645





Outputs and Visualization

Analysis of Predictive Model Performance Across Market Indices

S&P 500 (GSPC):

- Achieved the best overall performance with the highest score (0.858).
- Balanced metrics: Low Normalized MSE (0.014) and High R-squared (0.976).
- Indicates strong alignment between the model and S&P 500 behavior.

Dow Jones (DJI):

- Highest R-squared (0.979), showing strong predictive power.
- High MSE (2.24M) led to the lowest overall score (0.000).
- Suggests potential for improvement in error reduction.

NASDAQ (IXIC):

- Lowest R-squared (0.956), indicating weaker explanatory power.
- Moderate performance with Normalized MSE (0.386).
- Highlighted need for model refinement for tech-heavy indices.

Russell 3000 (RUA):

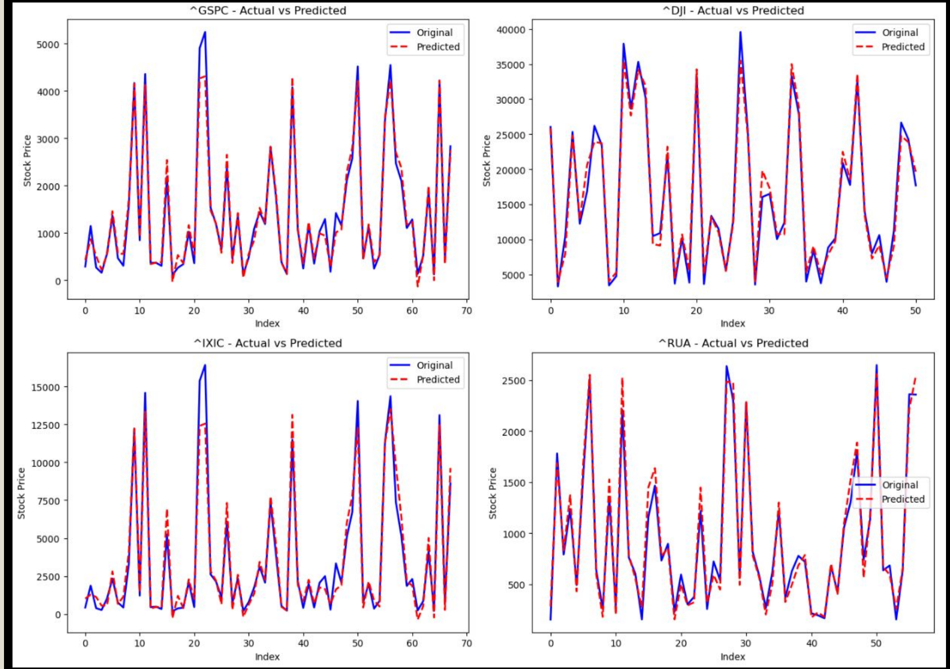
- Lowest Normalized MSE (0.000), but slightly lower R-squared (0.975).
- Ranked second overall with a strong score (0.818).

```
Results DataFrame:
   Index           MSE  R-squared  Normalized_MSE  Normalized_R_squared
0  ^GSPC  4.390977e+04   0.976395        0.014005              0.872259
1   ^DJI  2.236757e+06   0.979453        1.000000              1.000000
2  ^IXIC  8.707351e+05   0.955521        0.385780              0.000000
3   ^RUA  1.276217e+04   0.975093        0.000000              0.817820

      Score
0  0.858254
1  0.000000
2 -0.385780
3  0.817820

Index with the best combination of lowest MSE and highest R-squared:
Index                          ^GSPC
MSE                     43909.765839
R-squared                   0.976395
Normalized_MSE              0.014005
Normalized_R_squared        0.872259
Score                       0.858254
Name: 0, dtype: object
```

Comparison of Model Suitability

Aspect	S&P 500 (^GSPC)	Dow Jones (^DJI)
Metrics	- Low Normalized MSE (0.014)	- High Normalized MSE (1.000)
	- High R-squared (0.976)	- Highest R-squared (0.979)
Representation	- Includes 500 companies, highly diversified	- Includes only 30 large-cap companies
	- Captures a comprehensive market snapshot	- Limited market snapshot, focused on large caps
	- Lower sensitivity to outliers	- Higher sensitivity to outliers
Economic Indicator Alignment	- Strong correlation with macroeconomic variables	- Slightly better correlation with GDP and CPI
	- Balanced representation across multiple sectors	- Sector representation may skew macro alignment

- The **S&P 500** is better suited for the model due to its diversified composition, reduced sensitivity to outliers, and strong balance between metrics.
- While the **Dow Jones** excels in R-squared, its higher MSE and limited representation make it less reliable overall.



MODEL ANALYSIS AND CONCLUSION

Understanding the Magnitude effect

S&P 500

	coef	std err	t	P>|t|	[0.025	0.975]
const	1409.1128	230.760	6.106	0.000	955.167	1863.059
CPIAUCSL	-15.4949	2.673	-5.796	0.000	-20.754	-10.236
GDP	0.4962	0.089	5.584	0.000	0.321	0.671
CORESTICKM159SFRBATL	-12.9688	10.017	-1.295	0.196	-32.674	6.737
PCE	-0.4537	0.115	-3.949	0.000	-0.680	-0.228
UNRATE	-87.6493	7.703	-11.379	0.000	-102.802	-72.497
WM2NS	0.1289	0.015	8.416	0.000	0.099	0.159

Omnibus:	58.713	**Durbin-Watson:**	0.198
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	120.305
Skew:	0.913	**Prob(JB):**	7.52e-27
Kurtosis:	5.288	**Cond. No.**	3.76e+05

- As per the results obtained, overall considering the significance of 5%, the Inflation rate is not significantly affecting the outcome of the market.
- The Unemployment Rate and the CPI numbers however tend to have a strong negative relationship with the overall stock market performance outcome.

Understanding the Magnitude effect

NASDAQ Composite

	coef	std err	t	P>|t|	[0.025	0.975]
const	6108.9135	973.384	6.276	0.000	4194.093	8023.734
CPIAUCSL	-69.8982	11.276	-6.199	0.000	-92.080	-47.716
GDP	1.3946	0.375	3.721	0.000	0.657	2.132
CORESTICKM159SFRBATL	-77.6665	42.254	-1.838	0.067	-160.788	5.455
PCE	-1.3566	0.485	-2.799	0.005	-2.310	-0.403
UNRATE	-246.9426	32.491	-7.600	0.000	-310.859	-183.026
WM2NS	0.7596	0.065	11.759	0.000	0.632	0.887

Omnibus:	90.794	**Durbin-Watson:**	0.154
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	312.917
Skew:	1.158	**Prob(JB):**	1.12e-68
Kurtosis:	7.113	**Cond. No.**	3.76e+05

- As per the results obtained, overall considering the significance of 5%, the Inflation rate is not significantly affecting the outcome of the market.
- The Unemployment Rate and the CPI numbers however tend to have a strong negative relationship with the overall stock market performance outcome, even more against S&P 500 thus showing sector concentration.

Understanding the Magnitude effect

Dow Jones Index

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.267e+04	2983.845	7.598	0.000	1.68e+04	2.85e+04
CPIAUCSL	-234.0347	36.129	-6.478	0.000	-305.197	-162.872
GDP	3.7764	0.554	6.821	0.000	2.686	4.867
CORESTICKM159SFRBATL	-370.1245	103.067	-3.591	0.000	-573.132	-167.117
PCE	-1.9175	0.755	-2.541	0.012	-3.404	-0.431
UNRATE	-573.5617	81.876	-7.005	0.000	-734.829	-412.295
WM2NS	0.8704	0.110	7.903	0.000	0.653	1.087

Omnibus:	2.609	**Durbin-Watson:**	0.395
Prob(Omnibus):	0.271	**Jarque-Bera (JB):**	2.560
Skew:	-0.091	**Prob(JB):**	0.278
Kurtosis:	3.458	**Cond. No.**	7.94e+05

- As per the results obtained, overall considering the significance of 5%, every economic factor is significantly affecting the index since it being the industrial average.
- In this case, almost every factor is affecting the index heavily thus making it having a strong relationship with every factor.

Understanding the Magnitude effect

Russell 3000

	coef	std err	t	P>|t|	[0.025	0.975]
const	786.1723	199.406	3.943	0.000	393.635	1178.710
CPIAUCSL	-6.2911	2.439	-2.580	0.010	-11.092	-1.490
GDP	0.2211	0.051	4.344	0.000	0.121	0.321
CORESTICKM159SFRBATL	-26.7633	8.432	-3.174	0.002	-43.361	-10.165
PCE	-0.2259	0.066	-3.411	0.001	-0.356	-0.096
UNRATE	-64.1188	6.367	-10.070	0.000	-76.653	-51.585
WM2NS	0.0986	0.010	10.003	0.000	0.079	0.118

Omnibus:	53.887	Durbin-Watson:	0.247
Prob(Omnibus):	0.000	Jarque-Bera (JB):	142.859
Skew:	0.861	Prob(JB):	9.52e-32
Kurtosis:	6.010	Cond. No.	5.79e+05

- The Russell 3000 shows the power of a diversified index , not getting affected much in magnitude to the other indexes to the economic factors.
- In this case, The Unemployment factor seems to have a greater impact on the index performance than the other indicators.

Conclusion

- As far as the analysis is concerned, Dow Jones seems to have the strongest relationship with the economic indicators since it is an industrial average index

- In most of the models developed, Unemployment and CPI numbers seem to have the largest impact on the stock performance

- Money Supply in the economic intends to have the least impact

- For a risk averse individual, buying a Russell 3000 ETF would be advisable given the less exposure to the magnitude of change

- For a risk loving individual, Dow Jones would give out the best returns since it is heavily dependent on uncertain economic activity

Contribution

- Equal Contribution towards developing the problem and charting the path for applying the model.

- Interpretation of results with the analysis was done in a collaborative manner.