2025

Shubham Rawat

Correlation of variables



# How variables are co-related:

The listed economic indicators, commodities, and financial metrics are interconnected and collectively influence market growth. Here’s how they relate and contribute to the economy and markets:

**1. M2 Velocity of Money**

* **Definition:** Measures how quickly money is circulating in the economy.
* **Relation to Others:**
  + A high velocity suggests active economic transactions and growth.
  + Can influence **inflation (CPI)** and **stock indices**, as increased spending can lead to higher prices and corporate revenues.
* **Contribution to Growth:** Indicates the economy's efficiency in utilizing money to generate output.

**2. Unemployment Rate**

* **Definition:** Percentage of the labor force that is jobless but actively seeking employment.
* **Relation to Others:**
  + High unemployment may reduce consumer spending, affecting the **velocity of money**, **CPI**, and **stock indices**.
  + Inversely correlated with **Fed Funds Rate** adjustments, as central banks may lower rates to stimulate job creation.
* **Contribution to Growth:** Lower unemployment boosts disposable income and demand, driving market activity.

**3. Consumer Price Index (CPI)**

* **Definition:** Measures inflation by tracking changes in the prices of consumer goods and services.
* **Relation to Others:**
  + High CPI can prompt higher **Fed Funds Rates** to curb inflation.
  + Influences commodity prices like **gold**, **oil**, and **natural gas**, which are components of the index.
* **Contribution to Growth:** Moderate inflation is essential for sustainable growth, as it encourages investment without eroding purchasing power.

**4. Producer Price Index (PPI)**

* **Definition:** Tracks changes in prices received by producers for goods and services.
* **Relation to Others:**
  + Leads **CPI**, as higher producer costs can translate into consumer inflation.
  + Affects corporate profits and **stock indices**.
* **Contribution to Growth:** Reflects cost pressures on businesses, which can affect production and market competitiveness.

**5. Fed Funds Rate**

* **Definition:** Interest rate at which banks lend to each other overnight, set by the Federal Reserve.
* **Relation to Others:**
  + Higher rates can slow growth by increasing borrowing costs (affecting **building permits**, **stock indices**, and **DXY**).
  + Impacts commodity prices like **gold**, which inversely correlates with real interest rates.
* **Contribution to Growth:** Determines liquidity and borrowing conditions, influencing investment and consumer spending.

**6. Building Permits**

* **Definition:** Reflects the number of approved construction projects.
* **Relation to Others:**
  + An indicator of future construction activity and economic confidence.
  + Sensitive to changes in the **Fed Funds Rate** and commodity prices like **natural gas** and **oil**.
* **Contribution to Growth:** Stimulates employment and demand in related industries, boosting GDP.

**7. Financial Conditions Index (FCI)**

* **Definition:** Aggregates data on interest rates, credit spreads, equity prices, and exchange rates.
* **Relation to Others:**
  + A loose financial environment (lower index value) encourages borrowing and investment.
  + Linked to the **Fed Funds Rate** and **DXY**.
* **Contribution to Growth:** Reflects overall financial market health and liquidity, guiding investment decisions.

**8. Dollar Index (DXY)**

* **Definition:** Measures the value of the U.S. dollar against a basket of other currencies.
* **Relation to Others:**
  + A stronger dollar can suppress **oil**, **gold**, and **natural gas prices**, as these are priced in dollars.
  + Impacts export competitiveness, influencing the **PPI** and corporate earnings.
* **Contribution to Growth:** Affects trade dynamics and global investment flows into U.S. markets.

**9. Gold Price**

* **Definition:** Price of gold, often a hedge against inflation and currency devaluation.
* **Relation to Others:**
  + Inversely related to the **DXY** and **Fed Funds Rate**.
  + Moves with inflation expectations (**CPI**) and economic uncertainty.
* **Contribution to Growth:** Indicates investor sentiment and provides a safe haven during market volatility.

**10. Oil Price**

* **Definition:** Price of crude oil, a critical input for global energy.
* **Relation to Others:**
  + Directly impacts **CPI**, **PPI**, and consumer spending.
  + Affects industries reliant on transportation and energy, influencing **stock indices**.
* **Contribution to Growth:** Higher prices can increase production costs but also signal strong demand and economic activity.

**11. Natural Gas Price**

* **Definition:** Cost of natural gas, a key energy source.
* **Relation to Others:**
  + Like **oil**, affects energy costs, **PPI**, and **CPI**.
  + Impacts industrial output and construction (**building permits**).
* **Contribution to Growth:** Supports industrial and residential energy needs, influencing cost structures and economic productivity.

**12. U.S. Stock Indices**

* **Definition:** Represents the performance of major U.S. equities (e.g., S&P 500, Nasdaq).
* **Relation to Others:**
  + Influenced by corporate earnings, which are tied to consumer spending (**CPI**) and production costs (**PPI**).
  + Sensitive to **Fed Funds Rate**, **DXY**, and commodity prices.
* **Contribution to Growth:** Reflects investor confidence and provides capital for corporate expansion.

**Summary of Contributions to Market Growth**

* **Liquidity:** Indicators like the **Fed Funds Rate** and **FCI** govern borrowing costs and investment.
* **Demand-Supply Dynamics:** **CPI**, **PPI**, and commodity prices indicate purchasing power and cost pressures.
* **Economic Confidence:** Metrics like **building permits**, **unemployment rate**, and **stock indices** reflect optimism and resource allocation.
* **Global Interplay:** The **DXY** and commodity prices shape international trade and investment flows.

# Steps to follow:

Listed down are the steps that need to be taken in order to build the model using only random forest

**Objective**

* Develop a Random Forest model to analyze the relationship between fundamental economic indicators and stock market returns.
* Predict stock market returns either as a regression (continuous values) or classification (categories like bullish, neutral, bearish, etc.).

**Data Requirements**

The following columns will be used as independent variables:

1. **M2 Velocity of Money** – Indicates the rate at which money circulates in the economy.
2. **Unemployment Rate** – Reflects economic health and labor market conditions.
3. **Consumer Price Index (CPI)** – Measures inflation by tracking changes in prices of a basket of goods.
4. **Producer Price Index (PPI)** – Tracks price changes from the perspective of producers.
5. **Federal Funds Rate** – Represents monetary policy stance.
6. **Building Permits** – Serves as a proxy for economic activity in the housing market.
7. **Financial Conditions Index (FCI)** – Measures the overall state of financial markets.
8. **Dollar Index (DXY)** – Indicates the strength of the US dollar relative to a basket of currencies.
9. **Gold Price** – Often inversely correlated with market sentiment, acting as a safe haven.
10. **US Stock Indices** – Dependent variable, representing the stock market's performance.

**Methodology**

1. **Data Collection and Preprocessing**
   * Obtain historical data for all independent variables and stock indices.
   * Ensure data consistency (same frequency: daily, weekly, or monthly).
   * Handle missing values using imputation methods.
   * Scale or normalize features if necessary for the model.
2. **Exploratory Data Analysis (EDA)**
   * Examine trends, correlations, and distributions of features.
   * Investigate relationships between economic indicators and stock market returns.
3. **Feature Engineering**
   * Create lagged features to capture temporal dependencies (e.g., lagged CPI, PPI).
   * Transform the dependent variable:
     + **Regression**: Calculate percentage returns (log or simple returns).
     + **Classification**: Define return categories (e.g., >2% as bullish, <-2% as bearish).
4. **Modeling with Random Forest**
   * **Data Splitting**: Split into training, validation, and testing sets.
   * Train a Random Forest model:
     + For **regression**, predict continuous stock returns.
     + For **classification**, predict categories (e.g., strong bullish, bearish).
   * Use grid search or cross-validation to optimize hyperparameters like the number of trees and max depth.
5. **Model Evaluation**
   * For regression:
     + Use metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R².
   * For classification:
     + Evaluate accuracy, precision, recall, F1-score, and confusion matrix.
   * Validate results on unseen test data.
6. **Feature Importance Analysis**
   * Identify which economic indicators contribute most to predicting stock returns.
7. **Visualization**
   * Plot feature importance.
   * Compare predicted vs. actual returns for regression.
   * Create confusion matrices and classification reports for classification.