JK Team: Ting wei, Yiqing Hu

S&P 500 time series analysis: does Money Supply affect stock market quotations?

Introduction

The US stock market is indeed the largest and most influential stock market in the world, and what happens on the American stock exchanges can have significant impacts not only on the US economy but also on global financial markets and economic activity worldwide. One crucial factor that influences the economy and the stock market is the money supply, which is one of the components of monetary policy used by the Federal Reserve to manage the economy and stabilize financial markets.

Money supply refers to the total volume of currency held by the public at a particular point in time, and it can be measured using various money aggregates, including M0, M1, M2, and M3. In this paper, we will focus on the M2 indicator, which includes all the elements of M1 (currency in circulation, demand deposits, and other checkable deposits), as well as time deposits, savings deposits, and money market mutual funds. M2 is considered a representation of the total amount of money available in the economy that can be easily accessed by individuals and businesses.

One of the key roles of the Federal Reserve is to manage the money supply to achieve its dual mandate of price stability and maximum employment. Through various monetary policy tools, such as open market operations, discount rate changes, and reserve requirements, the Federal Reserve can influence the money supply and thereby impact credit availability, interest rates, and economic activity.

The money supply also has a significant impact on the stock market. An increase in the money supply can potentially lead to a higher amount of money available for investment in the stock market. When there is more money available to invest, it can increase demand for stocks, leading to higher stock prices. This can be due to increased buying pressure from investors who have more cash on hand to invest or from increased lending by financial institutions due to higher levels of available funds.

Additionally, an increase in the money supply can also impact stock prices through the effect on inflation expectations. According to the Keynesian economic theory, an increase in the money supply may lead to higher inflation expectations, as more money in the economy can potentially lead to higher demand for goods and services, pushing up prices. If investors expect higher inflation, they may demand higher returns on their investments, including stocks, to compensate for the eroding purchasing power of money. This increased demand for higher returns may lead to lower stock prices, as investors require higher expected returns to compensate for the expected inflation.

On the other hand, a decrease in the money supply can have the opposite effect on the stock market. A decrease in the money supply can potentially reduce the amount of money available for investment in the stock market, leading to lower demand for stocks and lower stock prices. It can also lead to tighter credit conditions, higher borrowing costs, and reduced economic activity, which can negatively impact corporate earnings and investor sentiment, further affecting stock prices.

It's important to note that the relationship between the money supply and the stock market can be complex, and other factors such as economic fundamentals, market sentiment, geopolitical events, and global economic conditions can also significantly influence stock market movements. Additionally, the impact of changes in the money supply on the stock market may not be immediate, and there may be time lags before the effects are fully reflected in stock prices.

Furthermore, the impact of the money supply on the stock market may vary depending on the overall economic and market conditions. For example, in times of economic expansion and low inflation, an increase in the money supply may lead to higher stock prices due to increased liquidity and lower borrowing costs. However, in times of economic contraction or high inflation, the relationship may be different, and changes in the money supply may have a different impact on the stock market.

In conclusion, the money supply is a crucial factor that can influence both the economy and the stock market. Changes in the money supply, as managed

Literature Review

Monetary policy, implemented by central banks, is widely recognized as one of the most effective tools for influencing real economic activities. As such, changes in monetary policy, including adjustments in the money supply, are believed to have an impact on the stock market. The relationship between money supply and stock prices has been a subject of significant analysis and research, with different theories and findings.

B. Maskay, in his study on the relationship between money supply and stock prices, differentiated between anticipated and unanticipated changes in money supply and their impact on the stock market. According to Maskay's "Theory and Review of Literature," the price of a stock is determined by the present value of its future cash flows, which is calculated by discounting the future cash flows at a discount rate. This theory suggests that money supply has a substantial influence on the discount rate, and therefore, M2, which is perceived as a reliable predictor of inflation, may also be a useful predictor of stock prices.

The results of Maskay's study confirmed that positive money supply shocks increase stock prices. Furthermore, the findings were consistent with the Efficient Market Hypothesis, which posits that anticipated changes in the money supply matter more than unanticipated changes in determining stock prices. Maskay also highlighted potential improvements to his study, such as changing the frequency of variables used, as his analysis was based on quarterly data.

However, it is important to note that money supply is not the only factor influencing stock prices. L. Shiblee, in her article, identified four other significant factors that can affect stock prices, namely inflation (CPI), GDP, unemployment, and money supply. The relationship between stock prices and GDP growth can be described as follows: higher stock prices imply higher investment costs, as investors would need to pay more for a

given number of shares. Therefore, there is a negative correlation between stock prices and future growth, as higher stock prices may imply lower expected returns on investment. S. Sharpe also highlighted a negative correlation between equity valuations and expected inflation, which may be attributed to lower expected real earnings growth and higher required real returns. Additionally, changes in the unemployment rate can provide valuable information about future interest rates, equity risk premium, and corporate earnings and dividends, although the impact of unemployment on stock prices may vary depending on whether the economy is in an expansionary or contractionary phase.

In conclusion, money supply is recognized as an important factor that can influence stock prices. Anticipated changes in money supply, in particular, are believed to have a significant impact on stock prices, as they can affect the discount rate used to calculate the present value of future cash flows. However, it is important to consider that money supply is not the sole determinant of stock prices, and other factors such as GDP, inflation, and unemployment, among others, can also play a crucial role in shaping stock market dynamics. Further research and analysis are warranted to better understand the complex relationship between money supply and stock prices, and to account for other relevant factors that may influence stock market movements.

Research hypotheses

Our hypothesis is that positive money supply shocks, which refer to an increase in the M2 supply by a central bank, can have an impact on stock prices. When the money supply increases, it typically leads to lower interest rates, increased liquidity in the financial system, and more money available for investment. This can create a favorable environment for businesses and investors, which can in turn lead to an increase in stock prices.

One of the key factors that can impact stock prices is inflation, specifically the Consumer Price Index (CPI). Inflation refers to the rate at which the general level of prices for goods and services in an economy is rising over time. If inflation is anticipated to be high in the future, it can lead to concerns about reduced purchasing power of future cash flows and eroded real returns on investments, including stocks. This can lead to a decrease in demand for stocks, which may negatively impact the stock market, including the S&P 500 index.

Another factor that can impact stock prices is the unemployment rate. The unemployment rate is often seen as an indicator of the overall health of the economy. When unemployment is high, it may signal a weaker labor market, reduced consumer spending, and lower economic growth. In turn, lower economic growth can impact corporate earnings, which are a key driver of stock prices. If companies face reduced demand for their products or services due to high unemployment, it could lead to lower corporate earnings and potentially impact stock prices.

Interest rates are another important factor that can impact stock prices. Higher interest rates can increase borrowing costs for businesses, which can impact their profitability and potentially lower stock prices. Additionally, higher interest rates can make bonds and other fixed-income investments relatively more attractive compared to stocks, which could lead to a shift in investment preferences and impact stock prices.

The relationship between stock prices and GDP growth is also worth considering. The stock market is often seen as a reflection of the overall health of the economy, and GDP growth is a key indicator of economic performance. The relationship between stock prices and GDP growth can be described as follows: higher stock prices mean higher investment costs. If an investor wants to buy a certain number of shares, they have to pay more. This can impact the demand for stocks and potentially impact stock prices.

In conclusion, the relationship between money supply and stock prices is a complex and multifaceted topic. Positive money supply shocks, which refer to an increase in the M2 supply by a central bank, can impact stock prices by influencing interest rates, liquidity in the financial system, and investment preferences. In addition, factors such as inflation, unemployment, interest rates, and GDP growth can also impact stock prices. It is important to consider these various factors when analyzing the relationship between money supply and stock prices, as well as other factors that can impact the stock market. Further research and analysis are needed to gain a deeper understanding of the complex dynamics between money supply and stock

Data description

The data used for analysis contained the time-series quotations of the S&P 500 index and Money Supply, GDP, Unemployment Rate, CPI, Commercial Bank Borrowing and Read Bank Asset from the US Federal Reserve.

1. Money Supply: This indicator measures the total amount of money in circulation within an economy. Changes in money supply can impact the overall liquidity and purchasing power in the economy, which can influence the performance of the stock market.
2. GDP (Gross Domestic Product): GDP is a measure of the total value of goods and services produced within a country's borders. It reflects the overall health and growth of the economy and can provide insights into the strength of the stock market.
3. Unemployment Rate: This indicator measures the percentage of the labor force that is unemployed. High unemployment rates may signal a weaker economy, which can negatively impact the stock market.
4. CPI (Consumer Price Index): CPI is a measure of the average change in prices of a basket of goods and services over time. It reflects inflation or deflation trends in the economy, which can affect the purchasing power of consumers and impact stock market performance.

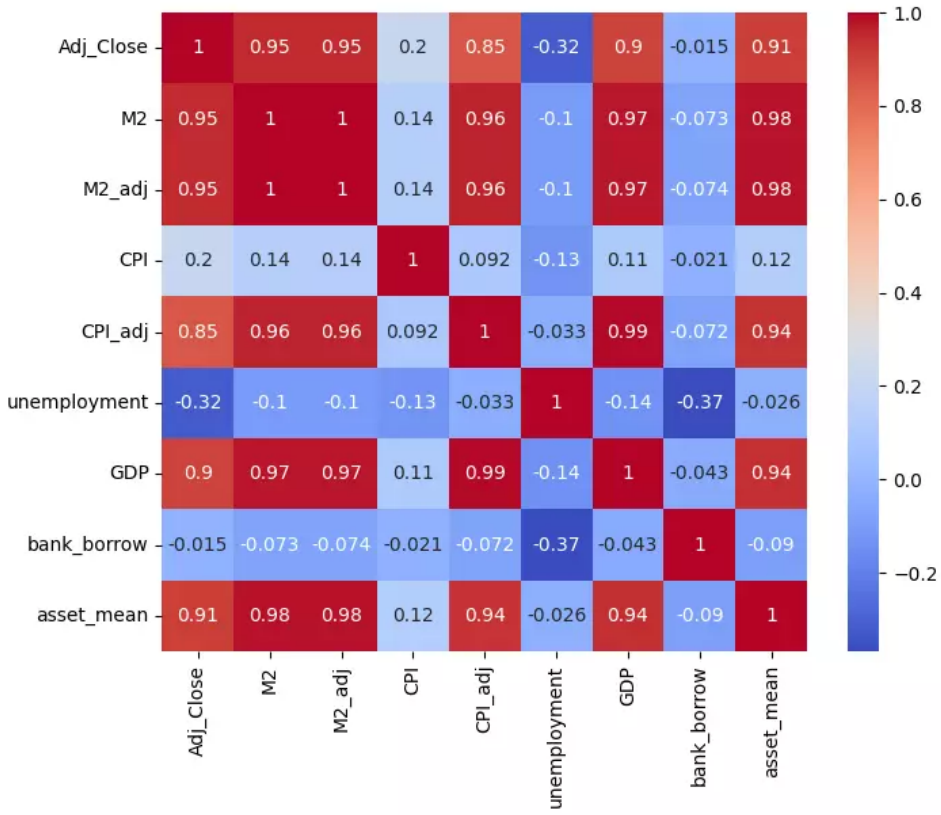
5. Commercial Bank Borrowing: This indicator measures the amount of borrowing that commercial banks engage in. Changes in bank borrowing can reflect changes in credit availability and overall lending conditions, which can impact market liquidity and investor sentiment.
6. Real Bank Assets: This indicator measures the total assets held by banks, adjusted for inflation. Changes in real bank assets can reflect changes in lending activities, credit conditions, and overall banking sector health, which can have implications for stock market performance.

Data Pre-Processing

First, the data available to us includes M2 monthly, M2 Monthly seasonal adjusted, M2 weekly, P&S500 index daily and monthly.

To start with, we fill in the missing dates in the P&S500 index daily data using data from the following day. We use data from the next working day after weekends and holidays, as P&S500 index is only available on US business days. The filled data is then merged with the M2 weekly data by date, resulting in a combined weekly data in a CSV format.

Next, we merge the M2 monthly, M2 Monthly seasonal adjusted, P&S500 index daily and monthly data by date. Additionally, we integrate the monthly data of GDP, unemployment rate, CPI, CPI seasonal adjusted, and Borrow from Commercial Bank into a single monthly table. Since the GDP data is quarterly, we allocate the value of each quarter to all three months within the quarter, making the quarterly GDP data into monthly data. Finally, we select the time window from January 1, 2000 to October 1, 2020 as the period for our analysis. The correlations value between every each pair features are showed by the heat map below:



Feature Engineering

We only perform feature engineering on the monthly data.

1. First, we create Lag Features using CPI seasonal adjusted and M2 Monthly seasonal adjusted. Lag Features use the current time step's feature values as lagged features for predicting the target value of the next time step. We create lag–1, lag–2, and lag–3 features for the target feature, representing the feature values from the previous one, two, and three time steps. However, due to the lagged nature of the data, the latest three data points do not have lagged data, resulting in NA values. We fill in the lagged data for dates after these three points.
2. We create Moving Averages features using M2 Monthly seasonal adjusted. This feature calculates the moving average value for a certain time window, capturing the trend of the data. We calculate the average value of M2 over windows of 3, 6, and 9 months, respectively.
3. We calculate time–series statistical features for each feature using a quarterly time window, including maximum, minimum, mean, and standard deviation.

Methodology

We will go over a time series regression model called the DL and ARDL model. We will prefer using the ARDL library to implement this model and use Okun's law as an

example. The ARDL model is actually two time series regression models combined, so we will briefly cover the **Autoregressive (AR)** portion of the model, as well as the **Distributed Lag** component of the model.

Given the dynamic relationship of time series data, there are three different ways of modeling these relationships.

1. The dependent variable y is a function of current and past values of an explanatory variable x, that is,

$$y_t = f(x_t, x_{t-1}, x_{t-2}, \ldots)$$

We can think of (y_t, x_t) as denoting the values of y and x in the current period; x_{t-1} denotes teh value of x in the previous period, x_{t-2} denotes the value of x in the previous period, and so on. Because of the lagged effects, the above equation is called a distributed lag (DL) model.

2. Specify a modelw with a **lagged dependent variable** as one of the explanatory variables.

$$y_t = y(y_{t-1}, x_t)$$

We can also combine the first two features of the above and previous equation so that we have a dynamic model with lagged values of both the dependent and independent variables, such as

$$y_t = f(y_{t-1}, x_t, x_{t-1}, x_{t-2}, \ldots)$$

Such models are called autogregressive distributed lag models (ARDL), which is the model we are interested in implementing. As we have shown, the ARDL model is composed of an autoregressive component, which is the dependent variable regressed on one or more of its past values, and a distributed lag component, which is the independent variable and one or more of its lagged components.

3. A third way of modeling the continuing change over several periods is via an error term. we can write

$$y_t = f(x_t) + e_t \qquad e_t = g(e_{t-1})$$

where the function $e_t = g(e_{t-1})$ is used to denote the dependence of the error term on its value in the previosu period. In this case, e_t is correlated with e_{t-1}, and in such a scenario, we say the errors are serially correlated or autocorrelated.

4. An assumption that we will maintain throughout this exercise is that variables in our equation are **stationary**, which means that a variable is one that is not explosive, nor trending, and nor wandering aimlessly without returning to its mean. You can learn more on stationarity here, but a stationary variable simply means a variable whose mean, variance and other statistical properties remain constant over time.
5. Finite Distributed Lags

The first dynamic relationship we consider is the first model that we introduced, which took the form of $y_t = f(x_t, x_{t-1}, x_{t-2}, \ldots)$, with the additional assumptions that the relationship is linear, and after q time periods, changes in x no longer have an impact on y. Under these conditions, we ahve the multiple regression model

$$y_t = \alpha + \beta_0 x_t + \beta 1 x_{t-1} + \beta 2 x_{t-2} + \ldots + \beta q x_{t-q} + e_t$$

The above model can be treated in the same way as a multiple regression model. Instead of having a number of different explanatory variables, we have a number of different *lags* of the *same* explanatory variable. This equation can be very useful in two ways:

- Forecasting future values of y. To introduce notation for future values, suppose our sample period is t=1,2,...,T. We use t for the index and T for the sample size to emphasize the time series nature of the data. Given the last observation in our sample is at t=T, the first postsample observation that we wish to forecast is at t=T+1. The equation for this observation can be given by

$$y_{T+1} = \alpha + \beta_0 x_{T+1} + \beta_1 x_T + \beta 2 x_{T-1} + \ldots + \beta q x_{T-q+1} + e_{T+1}$$

- Strategic Analysis. For example, to use an economic example, understanding the effects of the change in interest rate on unemployment and inflation, or the effect of advertising on sales on a firm's products. The coefficient βs gives the change in $E(y_t)$ when x_{t-s} s changes by one unit, but x is held constant in other periods. Alternatively, if we look forward rather than backward, βs gives the changes in $E(y_{t+s})$ when x_t changes by one unit, but x in other periods is held constant. In terms fo derivatives

$$\frac{\partial E(y_t)}{\partial x_{t-s}} = \frac{\partial E(y_{t+s})}{\partial x_t} = \beta_s$$

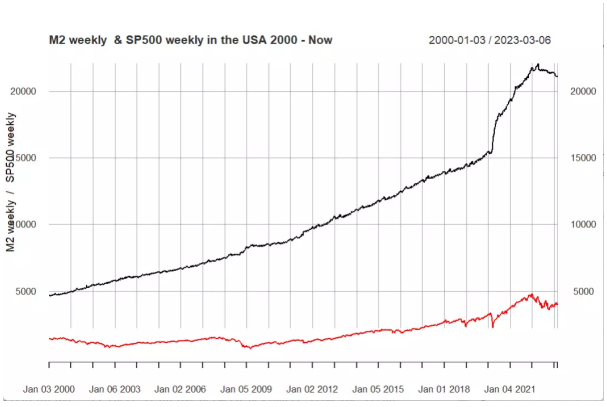
DL Model and ARDL Model

Model1– Distirbuted Lagged Model (DL Model) based on weekly data

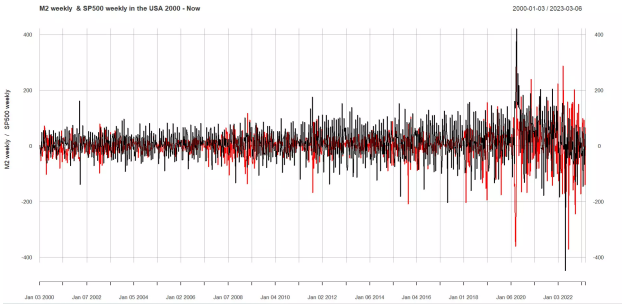
In distributed lagged models, both y(SP500) and x(M2) are typically random. That is, we don't know their values until we sample them. For example, we don't "s SP500 index and then observe the resulting level of M2. To fit this random process, we assume M2 is random, the error term is independent of all x's in the sample–past, current, and future.

1. $y_t = \alpha + \beta_0 x_t + e_t$
2. y and x are stationary random variables, and e_t is independent of current, past, and future values of x.
3. $E(e_t) = 0$
4. $\text{var}(e_t) = \sigma^2$
5. $\text{cov}(e_t, e_s) = 0 \qquad t \neq s$
6. $e_t \sim N(0, \sigma^2)$

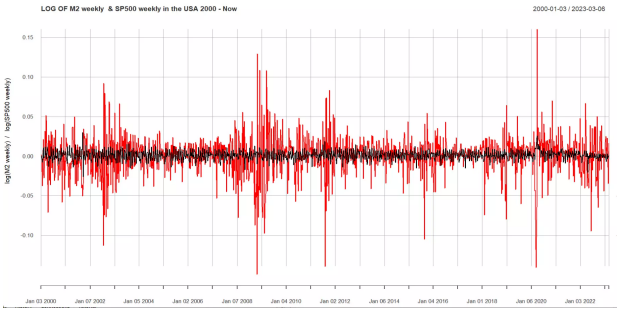
- Step 1: plot the SP500 and M2 in order to see the distribution and whether they are stationary.



From this plot, the intuition is non–stationary, and they can be random walk with a drift or random walk around quadratic trend, then plot the difference of SP500 and M2 and difference of logarithm of SP500 and M2 , to compare the stationarity.



diff(SP500) & diff(M2)



diff(log(SP500)) & diff(log(M2))

- Step 2: Stationarity testing, we can choose Dickey–Fuller test(DF test), Augmented Dickey–Fuller Test(ADF test) or Breusch-Godfrey to test the stationarity. In our case, the BGtest is used due to this method gives us a hint, which number of augmentations to choose
  - BG test to check the stationarity of dependent variable SP500

```
> testdf2(variable = data_weekly[,c("SP_500_W")], # vector tested
+         test.type = "c", # test.type = "nc",
+         max.augmentations = 5, # maximum number of augmentations added
+         max.order=5)
  augmentations       adf     p_adf p_bg.p.value.1 p_bg.p.value.2
1             0 0.3356303 0.9788798      0.6015279      0.8016741
2             1 0.3665967 0.9801352      0.9950544      0.9068608
3             2 0.3899564 0.9810822      0.9792526      0.9991066
4             3 0.5083444 0.9858817      0.9817933      0.9880554
5             4 0.5243934 0.9865324      0.9823067      0.9866659
6             5 0.7164697 0.9900000      0.9388241      0.9947357
  p_bg.p.value.3 p_bg.p.value.4 p_bg.p.value.5
1      0.2276157      0.3524349     0.02578184
2      0.2527350      0.3828064     0.02764947
3      0.2565648      0.3863521     0.02876150
4      0.9980143      0.9944399     0.10810946
5      0.9976614      0.9995360     0.11404659
6      0.9979581      0.9996590     0.99990224
```

In 3th row, the p.value.1–5 are all above significant level (0.05), so we can trust the 3th adf value which is 0.508 larger than 0.05, so we can't reject the Null hypothesis which means not stationary.

- BG test for diff(SP500)

```
> testdf2(variable = diff(data_weekly[,c("SP_500_W")]), # vector tested
+         test.type = "c", # test.type = "nc",
+         max.augmentations = 5, # maximum number of augmentations added
+         max.order=5) # maximum order of residual lags for BG test
  augmentations       adf p_adf p_bg.p.value.1 p_bg.p.value.2 p_bg.p.value.3
1             0 -35.21034  0.01      0.9955186      0.9164063      0.2659815
2             1 -24.97920  0.01      0.9804459      0.9992099      0.2697458
3             2 -21.42749  0.01      0.9833717      0.9888720      0.9982512
4             3 -18.35156  0.01      0.9849044      0.9876579      0.9979527
5             4 -17.59055  0.01      0.9430002      0.9955405      0.9983248
6             5 -16.13784  0.01      0.9687301      0.9960152      0.9996239
  p_bg.p.value.4 p_bg.p.value.5
1      0.4012694     0.03124976
2      0.4049207     0.03251600
3      0.9960796     0.11924105
4      0.9996184     0.12492116
5      0.9997456     0.99993239
6      0.9998978     0.99997904
```

In 3th row, the p.value.1–5 are all above significant level (0.05), so we can trust the 3th adf value which is 0.01 below 0.05, so we can reject the Null hypothesis which means the difference of SP500 is stationary.

- Repeat these process we can get difference of M2 is also stationary

```
> testdf2(variable = diff(data_weekly[,c("M2_W")]), # vector tested
+         test.type = "c", # test.type = "nc",
+         max.augmentations = 14,  # maximum number of augmentations added
+         max.order=5) # maximum order of residual lags for BG test
   augmentations       adf p_adf p_bg.p.value.1 p_bg.p.value.2 p_bg.p.value.3
1              0 -31.731960  0.01     0.79518752   4.600671e-02   1.252626e-10
2              1 -25.459574  0.01     0.68305189   1.533930e-01   1.493741e-08
3              2 -23.559226  0.01     0.01416362   9.271811e-03   1.164706e-03
4              3 -10.986455  0.01     0.08566743   6.621481e-10   1.579346e-09
5              4  -9.484800  0.01     0.26210194   2.214745e-08   1.864729e-08
6              5 -12.940476  0.01     0.74067372   9.340620e-01   8.072275e-02
7              6 -11.743612  0.01     0.99343497   9.439024e-01   1.521559e-02
8              7 -11.166762  0.01     0.94981356   9.524063e-01   1.564748e-02
9              8  -8.079537  0.01     0.06773374   1.614062e-01   2.772216e-01
10             9  -9.677230  0.01     0.67982800   7.055402e-01   3.673717e-02
11            10  -8.755138  0.01     0.94942361   6.292339e-01   2.690402e-02
12            11  -8.740222  0.01     0.69722152   6.811079e-01   5.863795e-02
13            12  -5.704481  0.01     0.57108741   7.926855e-01   8.512720e-01
14            13  -5.882415  0.01     0.95757217   9.342315e-01   9.107143e-01
15            14  -5.988971  0.01     0.96393239   9.985146e-01   9.367058e-01
   p_bg.p.value.4 p_bg.p.value.5
1    1.030171e-71   4.435388e-76
2    3.675579e-64   6.577345e-70
3    6.279115e-63   4.252312e-66
4    4.936634e-09   1.032143e-08
5    3.650953e-08   2.598997e-09
6    6.058585e-03   1.993868e-04
7    9.082433e-04   3.317044e-05
8    6.690403e-04   3.172190e-05
9    1.255197e-06   1.389256e-06
10   4.183186e-04   8.989036e-04
11   4.717540e-04   1.181783e-03
12   8.253507e-04   2.003623e-03
13   7.281821e-01   8.174838e-01
14   8.688386e-01   9.385984e-01
15   8.306543e-01   9.128331e-01
```

- Step 3: Build the DL model using dynlm() function and then check the autocorrelation of model's residuals.

```
dl01 <- dynlm(d(SP_500_W) ~ d(M2_W) + L(d(M2_W)), data = data_weekly)
```

```
Time series regression with "zoo" data:
Start = 2000-01-17, End = 2023-03-06

Call:
dynlm(formula = d(SP_500_W) ~ d(M2_W) + L(d(M2_W)), data = data_weekly02)

Residuals:
    Min      1Q  Median      3Q     Max
-375.49  -18.26    2.36   23.16  355.10

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.02762    1.56608   0.656   0.5118
d(M2_W)      0.04939    0.02478   1.993   0.0465 *
L(d(M2_W))   0.03286    0.02478   1.326   0.1851
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 52.09 on 1205 degrees of freedom
Multiple R-squared:  0.005164,  Adjusted R-squared:  0.003513
F-statistic: 3.128 on 2 and 1205 DF,  p-value: 0.04417
```

so try to drop Lag, because it’s insignificant.

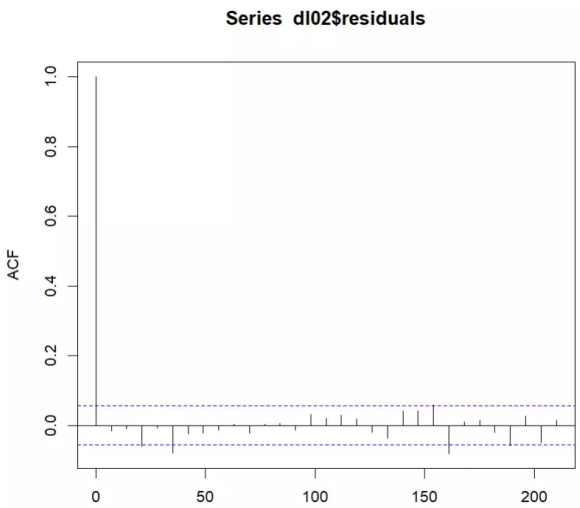
```
dl02 <- dynlm(d(SP_500_W) ~ d(M2_W) , data = data_weekly)
```

```
Time series regression with "zoo" data:
Start = 2000-01-10, End = 2023-03-06

Call:
dynlm(formula = d(SP_500_W) ~ d(M2_W), data = data_weekly)

Residuals:
    Min      1Q  Median      3Q     Max
-372.00  -18.39    2.51   22.90  365.77

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.43384    1.53512   0.934   0.3505
d(M2_W)      0.05234    0.02468   2.121   0.0341 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 52.09 on 1207 degrees of freedom
Multiple R-squared:  0.003713,  Adjusted R-squared:  0.002887
F-statistic: 4.498 on 1 and 1207 DF,  p-value: 0.03414
```

The p–value of F–statistic is 0.03(<0.05 significant level), all variables are jointly significant. The difference of M2 increase 1billion dollars leads to the difference of SP500 increase 0.052 unit which means positive money supply will increase the SP500.

We can then compute the correlogram of our least squares residuals as follows

```
acf(dl02$residuals, type='correlation')
```



Based on the correlogram, we can see that the autocorrelations of the residuals are not statistically significant from 0, suggesting that there is no evidence of autocorrelation. Then we will check first 5 lags to make sure there is no autocorrelation.

```
> bgtest(res~1,data=resids_, order = 1)

        Breusch-Godfrey test for serial correlation of order up to 1

data:  res ~ 1
LM test = 0.2599, df = 1, p-value = 0.6102

> bgtest(res~1,data=resids_, order = 2)

        Breusch-Godfrey test for serial correlation of order up to 2

data:  res ~ 1
LM test = 0.38435, df = 2, p-value = 0.8252

> bgtest(res~1,data=resids_, order = 3)

        Breusch-Godfrey test for serial correlation of order up to 3

data:  res ~ 1
LM test = 4.7939, df = 3, p-value = 0.1875

> bgtest(res~1,data=resids_, order = 4)

        Breusch-Godfrey test for serial correlation of order up to 4

data:  res ~ 1
LM test = 4.9151, df = 4, p-value = 0.2961

> bgtest(res~1,data=resids_, order = 5)

        Breusch-Godfrey test for serial correlation of order up to 5

data:  res ~ 1
LM test = 12.89, df = 5, p-value = 0.02443
```

Unfortunately, notice the 5th lag(0.02) is smaller than 0.05, suggesting that the AR(1)assumption may not be adequate. Next, we can try the robust matrix model or directly use the ARDL model. We choose to use the ARDL model, because we think the SP500 can also have correlation with history value.

Model2– Autoregressive Distributed Lagged Model (ARDL Model)based on weekly data

$$y_t = \delta + \theta y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

The above equation is autoregressive distributed lag model, the dependent variable(SP500) is regressed on its own lagged value (the autoregressive component), and it also includes explanatory variable(M2) and its lagged values (the distributed lag component). Using dynlm() function to build the ARDL model.

```
ardl01 <- dynlm(d(SP_500_W) ~ L(d(SP_500_W),c(1:5)) + d(M2_W) + L(d(M2_W)),
summary(ardl01)
```

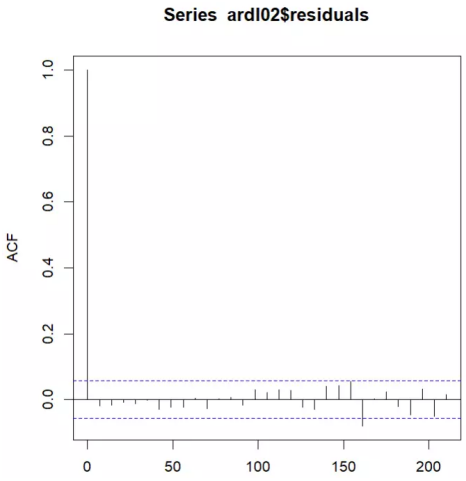
```
Call:
dynlm(formula = d(SP_500_W) ~ L(d(SP_500_W), c(1:5)) + d(M2_W)
+
    L(d(M2_W)), data = data_weekly)

Residuals:
    Min      1Q  Median      3Q     Max
-388.45  -17.89    3.18   23.02  320.08

Coefficients:
                       Estimate Std. Error t value
(Intercept)             1.52405    1.57763   0.966
L(d(SP_500_W), c(1:5))1 -0.01620    0.02884  -0.562
L(d(SP_500_W), c(1:5))2 -0.01109    0.02886  -0.384
L(d(SP_500_W), c(1:5))3 -0.05342    0.02888  -1.850
L(d(SP_500_W), c(1:5))4 -0.00664    0.02887  -0.230
L(d(SP_500_W), c(1:5))5 -0.08149    0.02890  -2.820
d(M2_W)                 0.04692    0.02485   1.888
L(d(M2_W))              0.02787    0.02489   1.120
                       Pr(>|t|)
(Intercept)             0.33422
L(d(SP_500_W), c(1:5))1  0.57444
L(d(SP_500_W), c(1:5))2  0.70101
L(d(SP_500_W), c(1:5))3  0.06461 .
L(d(SP_500_W), c(1:5))4  0.81812
L(d(SP_500_W), c(1:5))5  0.00488 **
d(M2_W)                 0.05926 .
L(d(M2_W))              0.26303
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 52.01 on 1196 degrees of freedom
Multiple R-squared:  0.01461,   Adjusted R-squared:  0.008841
F-statistic: 2.533 on 7 and 1196 DF,  p-value: 0.01372
```

From the ARDL model results, we see lots insignificant varialbes, so we drop it, and try again.

```
ardl02 <- dynlm(d(SP_500_W) ~ L(d(SP_500_W),c(3,5)) + d(M2_W) , data = data_
summary(ardl02)
```

```
Coefficients:
                       Estimate Std. Error t value
(Intercept)             1.78376    1.53755   1.160
L(d(SP_500_W), c(3, 5))3 -0.05530    0.02877  -1.922
L(d(SP_500_W), c(3, 5))5 -0.08166    0.02881  -2.835
d(M2_W)                 0.05074    0.02463   2.060
                       Pr(>|t|)
(Intercept)             0.24622
L(d(SP_500_W), c(3, 5))3  0.05483 .
L(d(SP_500_W), c(3, 5))5  0.00466 **
d(M2_W)                 0.03964 *
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 51.96 on 1200 degrees of freedom
Multiple R-squared:  0.0132,    Adjusted R-squared:  0.01073
F-statistic:  5.35 on 3 and 1200 DF,  p-value: 0.001161
```

Then we still need to check the autocorrelation problem, It can be seen from below 2 plots, we can’t reject the bgtest’s null hypothesis, implying that our model is fine and aren’t autocorrelated.




```
> bgtest(res~1,data=resids_, order = 1)

        Breusch-Godfrey test for serial correlation of order up to 1

data:  res ~ 1
LM test = 0.37734, df = 1, p-value = 0.539

> bgtest(res~1,data=resids_, order = 2)

        Breusch-Godfrey test for serial correlation of order up to 2

data:  res ~ 1
LM test = 0.70332, df = 2, p-value = 0.7035

> bgtest(res~1,data=resids_, order = 3)

        Breusch-Godfrey test for serial correlation of order up to 3

data:  res ~ 1
LM test = 0.80296, df = 3, p-value = 0.8488

> bgtest(res~1,data=resids_, order = 4)

        Breusch-Godfrey test for serial correlation of order up to 4

data:  res ~ 1
LM test = 0.99996, df = 4, p-value = 0.9098

> bgtest(res~1,data=resids_, order = 5)

        Breusch-Godfrey test for serial correlation of order up to 5

data:  res ~ 1
LM test = 1.0113, df = 5, p-value = 0.9616
```

By now, we are able to interpret the results.

△SPFt=1.783+ 0.039 △M2t + 0.054 △SPFt−3+0.004 △SPFt−5

- In short term

  The difference of M2 increases 1 unit leads to leads to 0.039 unit increase in the difference of SP500. Looking at lag q=3, an increase in the difference of SP500 will lead to a increase in the difference of SP500 by 0.05 unit, and 1unit increase of diff(SP500) in the 5th lag will increase only 0.004 unit of diff(SP500). But we can see that the M2 and historical value of SP500 are jonitly significant and postive increase the SP500 current value.

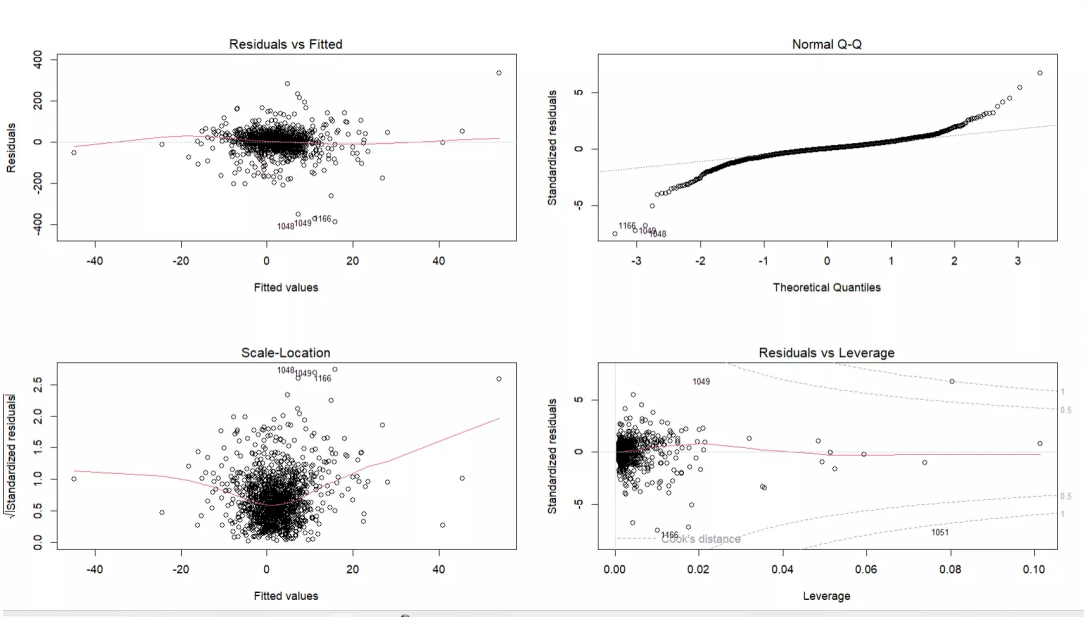
- In middle or long−term

$$LTM = \sum_{i=1}^{+\infty} \beta_i$$ =0.039+0.054+0.004=0.097

if there is a one−perid unsteady shock in △M2 and △SPF , then we expect the distortion from the steady state will sum up to 0.097

So the real mid−and long−term stock value of SP500 = 0.097/sum(1+r)^i, which r present the future discount rate, If the future discount rate r is positive, the denominator is greater than 1, and the longer the period and bigger the r, the stock real value still increase but smaller(<0.097), otherwise, when r is negative, and the longer period, the stock real value will increase more and more (>0.097). Which means when interest rates decrease in the future, people will reduce their deposits and increase their stock investment to obtain higher returns, so SP500 will increase。 On the contrary, if interest rates increase in the future, people will increase their deposits and reduce their stock investment.

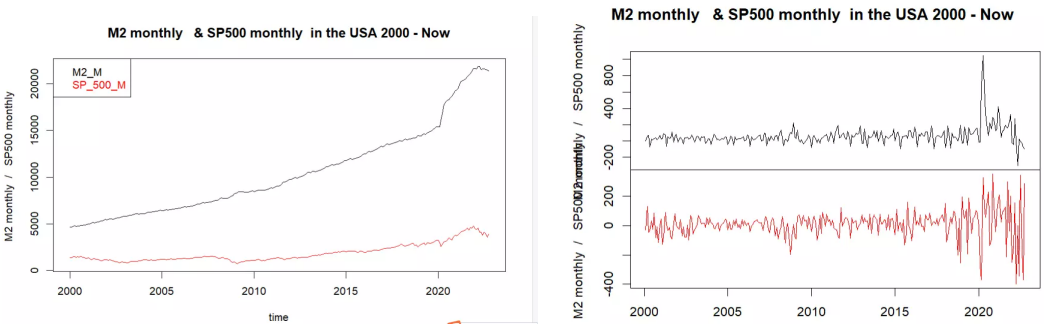
- Diagnostics plots



- Residuals vs Fitted: This plot shows residuals exhibit non–linear patterns, which may indicate that the model does not capture the non–linear relationship between predictor variables and the outcome variable.
- Normal Q–Q: This plot assesses the normality of residuals by comparing them to a straight dashed line. The residuals fall closely along the line, it indicates that they are normally distributed.
- Scale–Location: Also known as the Spread–Location plot, this plot helps to check the assumption of equal variance (homoscedasticity) of residuals along the ranges of predictors. A horizontal line with equally (randomly) spread points indicates homoscedasticity although this line is not quite horizontal.
- Residuals vs Leverage: This plot helps identify influential cases (i.e., subjects) that may significantly affect the regression results. Outlier values at the upper right or lower right corners of the plot, beyond the dashed line representing Cook's distance, in our plot, it only 1 outlier.

Model 3–monthly ARDL model–compare monthly and weekly

As above, these 2 model based on weekly data, now try to build monthly model to compare the results.



So plot the monthly SP500 and M2, and then plot the diff(SP500) and diff(M2), we can see from thses 2 plots, the difference of SP500 and M2 seems to be stationary. Then

check the BGtest. The below plot shows the difference of SP500 and difference of M2 monthly are stationary.

```
> testdf2(variable = diff(data_monthly[,c("SP_500_M")]), # vector tested
+         test.type = "c", # test.type = "nc",
+         max.augmentations = 5, # maximum number of augmentations added
+         max.order=5) # maxi
  augmentations       adf p_adf p_bg.p.value.1 p_bg.p.value.2
1             0 -18.597070  0.01      0.9589855      0.8984259
2             1 -12.719365  0.01      0.9460817      0.9754340
3             2  -9.299620  0.01      0.9229218      0.9943218
4             3  -7.383419  0.01      0.8604694      0.9762453
5             4  -5.809581  0.01      0.6747274      0.8591653
6             5  -6.644118  0.01      0.3661934      0.5821073
  p_bg.p.value.3 p_bg.p.value.4 p_bg.p.value.5
1      0.6161935      0.3924033      0.2777066
2      0.6623447      0.4584994      0.3055726
3      0.9798417      0.7413749      0.5294984
4      0.9965550      0.9990973      0.6686001
5      0.9436531      0.9776367      0.9896743
6      0.6878882      0.8207479      0.9090779
```

Then build monthly ARDL model

```
ardl03 <- dynlm(d(SP_500_M) ~ L(d(SP_500_M),c(1:5)) + d(M2_M) + L(d(M2_M)),
summary(ardl03)
```

```
Time series regression with "zoo" data:
Start = 2000-07-01, End = 2022-10-01

Call:
dynlm(formula = d(SP_500_M) ~ L(d(SP_500_M), c(1:5)) + d(M2_M) +
    L(d(M2_M)), data = data_monthly)

Residuals:
    Min      1Q  Median      3Q     Max
-457.44  -33.28    9.35   46.05  319.25

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -8.03498    6.89615  -1.165 0.245030
L(d(SP_500_M), c(1:5))1 -0.20961    0.06237  -3.361 0.000894 ***
L(d(SP_500_M), c(1:5))2 -0.02142    0.06311  -0.339 0.734604
L(d(SP_500_M), c(1:5))3  0.12013    0.06348   1.892 0.059563 .
L(d(SP_500_M), c(1:5))4  0.09975    0.06586   1.515 0.131078
L(d(SP_500_M), c(1:5))5  0.09046    0.06701   1.350 0.178232
d(M2_M)                   0.02345    0.06546   0.358 0.720388
L(d(M2_M))                0.23196    0.06527   3.554 0.000451 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 92.9 on 260 degrees of freedom
Multiple R-squared:  0.1167,    Adjusted R-squared:  0.09295
F-statistic: 4.909 on 7 and 260 DF,  p-value: 0.00003171
```

Then drop insignificant variables and check the autocorrelation

```
ardl04 <- dynlm(d(SP_500_M) ~ L(d(SP_500_M),c(1,3)) + L(d(M2_M)) , data = da
summary(ardl04)
```

```
Time series regression with "zoo" data:
Start = 2000-05-01, End = 2022-10-01

Call:
dynlm(formula = d(SP_500_M) ~ L(d(SP_500_M), c(1, 3)) + L(d(M2_M)),
    data = data_monthly)

Residuals:
    Min      1Q  Median      3Q     Max
-424.59  -32.78   11.17   48.12  323.58

Coefficients:
                          Estimate Std. Error t value   Pr(>|t|)
(Intercept)               -6.45741    6.58071  -0.981     0.3274
L(d(SP_500_M), c(1, 3))1 -0.20012    0.06062  -3.301     0.0011
L(d(SP_500_M), c(1, 3))3  0.12116    0.06204   1.953     0.0519
L(d(M2_M))                0.25388    0.05302   4.789 0.00000279

(Intercept)
L(d(SP_500_M), c(1, 3))1 **
L(d(SP_500_M), c(1, 3))3 .
L(d(M2_M))                ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 92.78 on 266 degrees of freedom
Multiple R-squared:  0.09951,   Adjusted R-squared:  0.08935
F-statistic: 9.798 on 3 and 266 DF,  p-value: 0.000003751
```

Compare AIC, BIC, R-squared and Adjusted R-Squard

- The R-Squared of Weekly ARDL model is 0.01, and Adjusted R-square is 0.01, but comapre to the monthly ARDL model which is 0.10 and 0.08, so monthly is better to explain the dependent variable SP500
- AIC, BIC： ardl02 is weekly model, ardl04 is monthly model, so still monthly is better

	df	AIC
ardl02	5	12935.517
ardl04	5	3218.515

	df	BIC
ardl02	5	12960.984
ardl04	5	3236.507

By now, we are able to interpret the results.

△SPF_MONTHt=−6.457+ 0.254 △M2_MONTHt −0.200 △SPF_MONTHt−1+0.121△SPF_MONTHt−3

- In short term

  The difference of M2 monthly increases 1 unit leads to leads to 0.254 unit increase in the difference of SP500. Looking at lag q=1, an unit increase in the difference of SP500 will lead to an decrease in the difference of SP500 by 0.2 unit, and 1unit increase of diff(SP500) in the 3th lag will increase 0.121 unit of diff(SP500). And postive increase M2 the SP500 current value montly, however, the previous diff(SP500) increase will decrease the SP500 now, while the 3th lag increase will increase the SP500. The possible reason is that the stock growth last month in the short term will make short−term investors take a wait−and−see attitude, thereby reducing current investment.

- In middle or long−term

$LTM = \sum_{i=1}^{+\infty} \beta_i$ =−0.200+0.121+0.254=0.175

if there is a one−perid unsteady shock in △M2 and △SPF , then we expect the distortion from the steady state will sum up to 0.175

So the real mid−and long−term stock value of SP500 = 0.097/sum(1+r)^i, which r present the future discount rate, If the future discount rate r is positive, the denominator is greater than 1, and the longer the period and bigger the r, the stock real value still increase but smaller(<0.175), otherwise, when r is negative, and the longer period, the stock real value will increase more and more (>0.175). Which means when interest rates decrease in the future, people will reduce their deposits and increase their stock investment to obtain higher returns, so SP500 will increase。 On the contrary, if interest rates increase in the future, people will increase their deposits and reduce their stock investment.

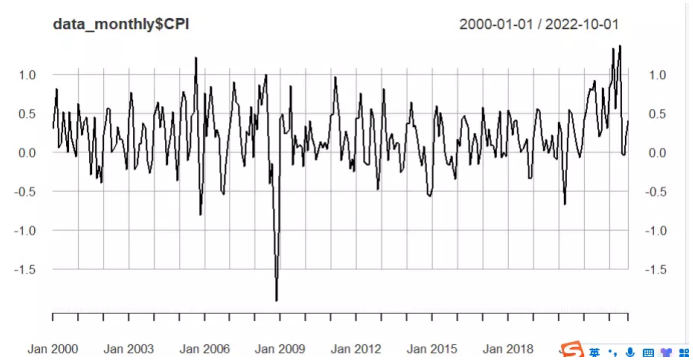
Model 4−Mutiple Variables ARDL model

EDA Analysis

```
     Index              SP_500_M          M2_M            CPI
 Min.   :2000-01-01   Min.   : 735.1   Min.   : 4673   Min.   :-1.915290
 1st Qu.:2005-09-08   1st Qu.:1180.8   1st Qu.: 6595   1st Qu.:-0.007703
 Median :2011-05-16   Median :1423.5   Median : 9085   Median : 0.202501
 Mean   :2011-05-17   Mean   :1860.1   Mean   :10440   Mean   : 0.209516
 3rd Qu.:2017-01-24   3rd Qu.:2341.8   3rd Qu.:13304   3rd Qu.: 0.468495
 Max.   :2022-10-01   Max.   :4766.2   Max.   :21856   Max.   : 1.373608
  unemployment         GDP         bank_borrow          asset_mean
 Min.   : 3.500   Min.   :10002   Min.   :-167.1000   Min.   : 454032
 1st Qu.: 4.400   1st Qu.:13188   1st Qu.:  -9.4500   1st Qu.: 821014
 Median : 5.400   Median :15558   Median :   1.6500   Median :2783582
 Mean   : 5.875   Mean   :16252   Mean   :   0.3993   Mean   :2976470
 3rd Qu.: 6.700   3rd Qu.:19148   3rd Qu.:  12.3500   3rd Qu.:4453681
 Max.   :14.700   Max.   :26138   Max.   : 134.4000   Max.   :8949532
```

- SP_500_M: monthly SP500, when there are lots small area data, handle it with log

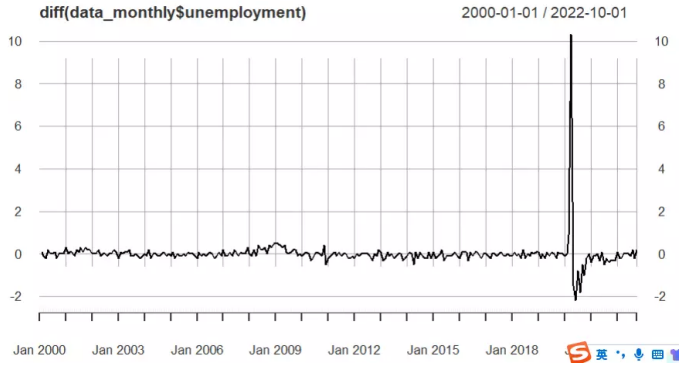
- M2_M: monthly M2, handle it with log
- CPI: Consumer Price Index: Total All Items for the United States,Not Seasonally Adjusted



https://fred.stlouisfed.org/series/CPALTT01USM657N

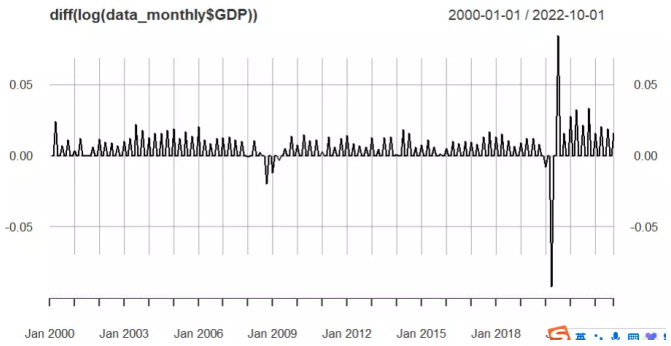
- Unemployment: Unemployment Rate–Percent,Seasonally Adjusted

using diff() function to see the distribution



https://fred.stlouisfed.org/series/UNRATE

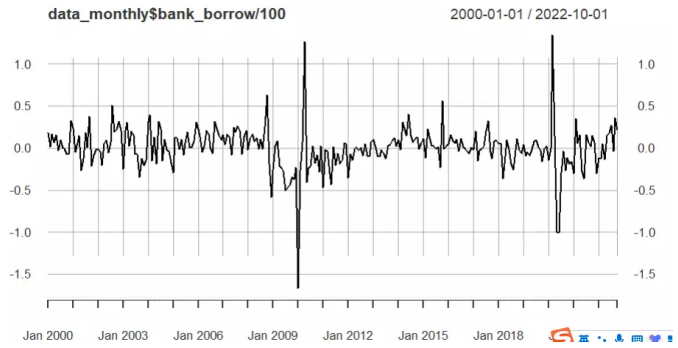
- GDP: Gross Domestic Product (GDP)–Seasonally Adjusted Annual Rate, handle it with log



https://fred.stlouisfed.org/series/GDP

- Bank_borrow: Borrowings, All Commercial Bank–Percent Change at Annual Rate

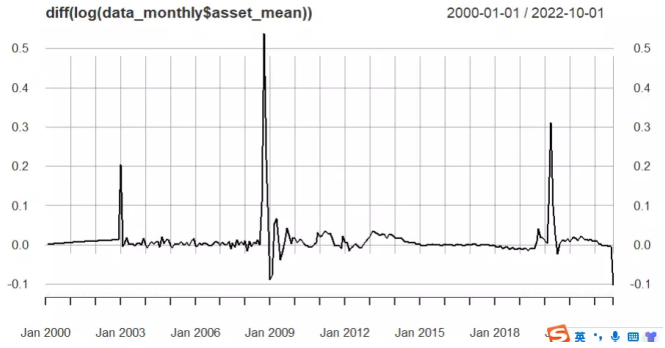
log and diff



https://fred.stlouisfed.org/series/H8B3094NCBCMG

- asset_mean: Total Assets (Less Eliminations from Consolidation): Wednesday Level

  diff and log



https://fred.stlouisfed.org/series/WALCL

Build ARDL Model with Additional Variables

```
ardl05 <- dynlm(d(log(SP_500_M)) ~ L(d(log(SP_500_M)),c(1:5)) +
                  d(log(M2_M)) + L(d(log(M2_M)))+
                  CPI  + L(CPI)+
                  d(unemployment) + L(d(unemployment))+
                  d(log(GDP )) + L(d(log(GDP) ))+
                  bank_borrow + L(bank_borrow)+
                  d(log(asset_mean )) + L(d(log(asset_mean) )),
                data = data_monthly)
summary(ardl05)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.130221 -0.024415  0.005987  0.026654  0.104695

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -0.0069343  0.0049722  -1.395  0.16438
L(d(log(SP_500_M)), c(1:5))1 -0.0068414  0.0650061  -0.105  0.91627
L(d(log(SP_500_M)), c(1:5))2 -0.0242758  0.0658979  -0.368  0.71290
L(d(log(SP_500_M)), c(1:5))3  0.0830023  0.0641821   1.293  0.19712
L(d(log(SP_500_M)), c(1:5))4  0.0536766  0.0623925   0.860  0.39045
L(d(log(SP_500_M)), c(1:5))5  0.0699337  0.0617353   1.133  0.25838
d(log(M2_M))                  0.5859963  0.4278526   1.370  0.17203
L(d(log(M2_M)))               0.6590087  0.4401385   1.497  0.13558
CPI                          -0.0086231  0.0085108  -1.013  0.31195
L(CPI)                        0.0027220  0.0085134   0.320  0.74944
d(unemployment)               0.0166219  0.0056159   2.960  0.00337 **
L(d(unemployment))            0.0106040  0.0056890   1.864  0.06350 .
d(log(GDP))                   0.9802152  0.3941787   2.487  0.01355 *
L(d(log(GDP)))                0.9744486  0.3751828   2.597  0.00995 **
bank_borrow                  -0.0001751  0.0001205  -1.454  0.14727
L(bank_borrow)               -0.0001039  0.0001187  -0.875  0.38229
d(log(asset_mean))           -0.2053055  0.0779377  -2.634  0.00896 **
L(d(log(asset_mean)))        -0.0360870  0.0803300  -0.449  0.65365
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04324 on 250 degrees of freedom
Multiple R-squared:  0.133,     Adjusted R-squared:  0.07407
F-statistic: 2.256 on 17 and 250 DF,  p-value: 0.003651
```

Drop insigificant variables one by one and check with anova()

```
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -0.0084935  0.0045705  -1.858  0.06428 .
L(d(log(SP_500_M)), c(1:5))1 -0.0110225  0.0638075  -0.173  0.86299
L(d(log(SP_500_M)), c(1:5))2 -0.0193422  0.0637387  -0.303  0.76179
L(d(log(SP_500_M)), c(1:5))3  0.0811447  0.0635983   1.276  0.20316
L(d(log(SP_500_M)), c(1:5))4  0.0546547  0.0612218   0.893  0.37285
L(d(log(SP_500_M)), c(1:5))5  0.0708051  0.0610968   1.159  0.24759
d(log(M2_M))                  0.6364414  0.4222362   1.507  0.13297
L(d(log(M2_M)))               0.5872098  0.4320473   1.359  0.17531
d(unemployment)               0.0159281  0.0053916   2.954  0.00343 **
L(d(unemployment))            0.0095002  0.0055553   1.710  0.08847 .
d(log(GDP))                   0.9954568  0.3863099   2.577  0.01054 *
L(d(log(GDP)))                0.9810994  0.3700575   2.651  0.00852 **
bank_borrow                  -0.0001908  0.0001123  -1.698  0.09064 .
d(log(asset_mean))           -0.2069584  0.0687902  -3.009  0.00289 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04306 on 254 degrees of freedom
Multiple R-squared:  0.1264,    Adjusted R-squared:  0.08174
F-statistic: 2.828 on 13 and 254 DF,  p-value: 0.0008262

> anova(ard105,ard106)
Analysis of Variance Table

Model 1: d(log(SP_500_M)) ~ L(d(log(SP_500_M)), c(1:5)) + d(log(M2_M)) +
    L(d(log(M2_M))) + CPI + L(CPI) + d(unemployment) + L(d(unemployment)) +
    d(log(GDP)) + L(d(log(GDP))) + bank_borrow + L(bank_borrow) +
    d(log(asset_mean)) + L(d(log(asset_mean)))
Model 2: d(log(SP_500_M)) ~ L(d(log(SP_500_M)), c(1:5)) + d(log(M2_M)) +
    L(d(log(M2_M))) + d(unemployment) + L(d(unemployment)) +
    d(log(GDP)) + L(d(log(GDP))) + bank_borrow + d(log(asset_mean))
  Res.Df     RSS Df  Sum of Sq      F Pr(>F)
1    250 0.46738
2    254 0.47093 -4 -0.0035468 0.4743 0.7546
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.129336 -0.023899  0.007239  0.028007  0.098238

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.0061437  0.0043263  -1.420  0.15676
d(log(M2_M))   0.8582515  0.3993495   2.149  0.03253 *
d(unemployment)  0.0169603  0.0051535   3.291  0.00113 **
L(d(unemployment))  0.0108438  0.0051226   2.117  0.03521 *
d(log(GDP))    1.0856440  0.3559591   3.050  0.00252 **
L(d(log(GDP)))  1.0035723  0.3569775   2.811  0.00530 **
bank_borrow   -0.0002292  0.0001071  -2.139  0.03337 *
d(log(asset_mean)) -0.1924112  0.0665405  -2.892  0.00415 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04324 on 264 degrees of freedom
Multiple R-squared:  0.1013,    Adjusted R-squared:  0.07746
F-statistic: 4.251 on 7 and 264 DF,  p-value: 0.0001815

> anova(ard107,ard108)
Analysis of Variance Table

Model 1: d(log(SP_500_M)) ~ d(log(M2_M)) + L(d(log(M2_M))) + d(unemployment) +
    L(d(unemployment)) + d(log(GDP)) + L(d(log(GDP))) + bank_borrow +
    d(log(asset_mean))
Model 2: d(log(SP_500_M)) ~ d(log(M2_M)) + d(unemployment) + L(d(unemployment)) +
    d(log(GDP)) + L(d(log(GDP))) + bank_borrow + d(log(asset_mean))
  Res.Df     RSS Df  Sum of Sq      F Pr(>F)
1    263 0.49262
2    264 0.49360 -1 -0.00097899 0.5227 0.4703
```

Finally, Variables M2, Umemployment Rate, GDP, Bank Borrowing Percent Change at Annual Rate, Average asset are jointly significant, then check bgtest(), and there's no autocorrelation problem.

```
> bgtest(res~1,data=resids_, order = 1)

        Breusch-Godfrey test for serial correlation of order up to 1

data:  res ~ 1
LM test = 0.1787, df = 1, p-value = 0.6725

> bgtest(res~1,data=resids_, order = 2)

        Breusch-Godfrey test for serial correlation of order up to 2

data:  res ~ 1
LM test = 1.0385, df = 2, p-value = 0.595

> bgtest(res~1,data=resids_, order = 3)

        Breusch-Godfrey test for serial correlation of order up to 3

data:  res ~ 1
LM test = 1.2166, df = 3, p-value = 0.749

> bgtest(res~1,data=resids_, order = 4)

        Breusch-Godfrey test for serial correlation of order up to 4

data:  res ~ 1
LM test = 1.2695, df = 4, p-value = 0.8665

> bgtest(res~1,data=resids_, order = 5)

        Breusch-Godfrey test for serial correlation of order up to 5

data:  res ~ 1
LM test = 4.6786, df = 5, p-value = 0.4563
```

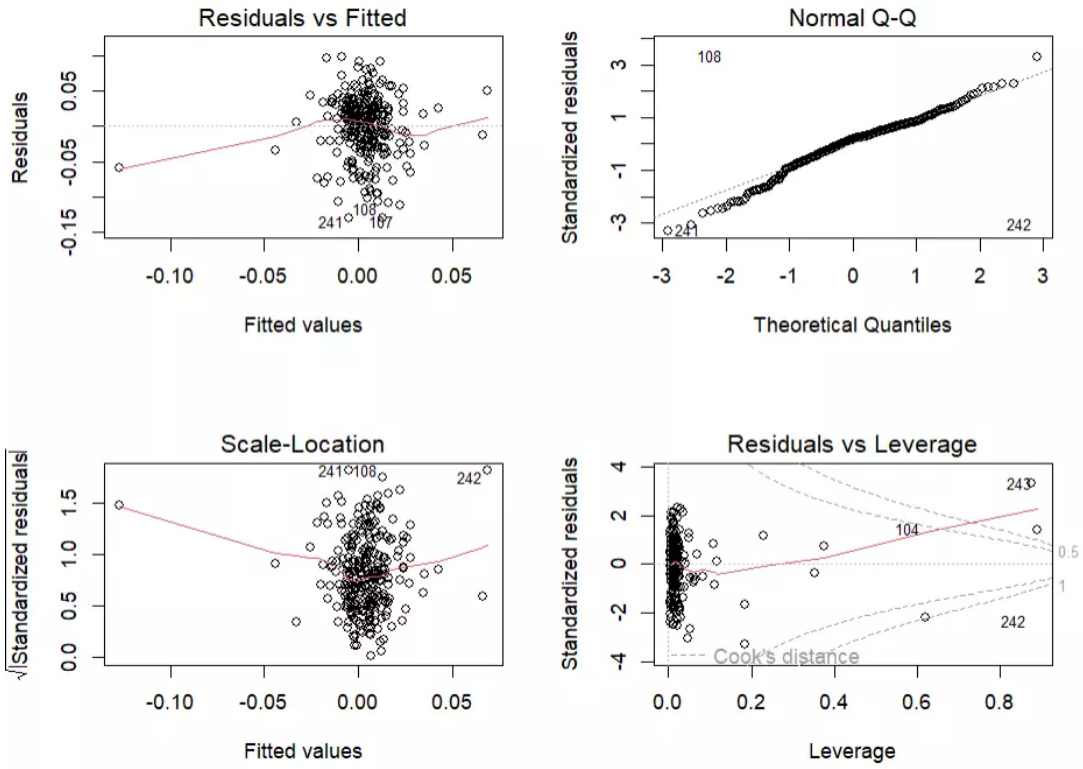
By now, we are able to interpret the results.

△log(SPFt)=−0.006+ 0.858 △log(M2)t + 0.016 △UNEMPLOYMENTt+0.01 △UNEMPLOYMENT(t−1)+1.08 △log(GDP)t+1.0△log(GDP)(t−1)−0.0002 BANKt−△log(ASSET)t

- We can see M2, Unemployment rate, and GDP has positive influence on the SP500, when these 3 variables increases , it will increase the stock market.
  - For GDP, GDP is a measure of the overall economic activity in a country, and a positive increase in GDP generally indicates that the economy is growing. When the economy is expanding, it can lead to increased corporate profits for companies listed in the S&P 500, which can positively impact stock prices and drive the index higher.
  - But for the unemployment rate, an increase in the unemployment rate would lift stocks is surprising. However, there are some possible explanations for why the unemployment rate may increase while the S&P 500 also increases. For examples, Interest rates and monetary policy: Unemployment can also impact monetary policy decisions made by central banks, which can in turn influence stock prices. For example, during periods of high unemployment, central banks may implement monetary policies, such as lower interest rates or quantitative easing, to stimulate economic growth and reduce unemployment. These policies can impact borrowing costs for businesses, influence consumer spending, and impact overall market sentiment, which can potentially impact the performance of the S&P 500.
- Bank Borrowing Percent Change at Annual Rate and average value of asset has negative influence on the SP500.

  A decrease in bank borrowing could indicate reduced credit availability for businesses and consumers. This can lead to decreased borrowing and spending by businesses, which may result in reduced investments, lower production levels, and potentially lower corporate profits. Reduced credit availability can also impact consumer spending, as consumers may have less access to credit for making purchases, leading to decreased consumer spending. Lower corporate profits and reduced consumer spending can negatively impact the overall economic activity, which can, in turn, lead to lower stock prices, including those of companies listed in the S&P 500.

Diagnostics plot



Long-term Prediction

Model Description

LightGBM (Light Gradient Boosting Machine) is a popular gradient boosting framework used for solving regression and classification problems. It is chosen to handle the time series prediction of S&P index stock prices due to its efficient and effective performance in handling large datasets with high-dimensional features, such as financial time series data.

One of the main reasons for using LightGBM in this context is its ability to handle complex relationships and interactions among features in the data, which are common in financial time series data. LightGBM uses a tree-based boosting algorithm, where each tree is built sequentially to correct the errors made by the previous trees. This allows the model to capture non-linear patterns and dependencies in the data, which can be crucial in predicting stock prices.

Tunning

The parameters used in the LightGBM model for this time series prediction task are as follows:

- 'boosting_type': 'gbdt': This specifies the type of boosting algorithm used, which is Gradient Boosting Decision Trees (gbdt) in this case.

- 'objective': 'regression': This specifies the type of problem being solved, which is regression in this case, as we are predicting stock prices.
- 'metric': 'rmse': This specifies the evaluation metric used during model training, which is Root Mean Squared Error (RMSE) in this case, a common metric for regression problems.
- 'max_depth': 7: This specifies the maximum depth of each tree in the boosting process, which is set to 7 in this case. A higher value allows the model to capture more complex patterns in the data, but may also result in overfitting.
- 'num_leaves': 30: This specifies the maximum number of leaves in each tree, which is set to 30 in this case. A higher value allows the model to capture more detailed interactions among features, but may also result in overfitting.
- 'learning_rate': 0.15: This specifies the learning rate, which controls the step size in updating the model weights during training. A higher value allows the model to learn faster, but may also result in overshooting the optimal weights.
- 'n_estimators': 70: This specifies the number of boosting rounds or iterations, which is set to 70 in this case. A higher value allows the model to learn more complex patterns in the data, but may also increase the risk of overfitting.
- 'reg_alpha': 0.12: This specifies the L1 regularization term, which helps in controlling the complexity of the model by adding a penalty for large weights. A higher value of reg_alpha results in more regularization, which can prevent overfitting.
- 'reg_lambda': 0.1: This specifies the L2 regularization term, which helps in controlling the complexity of the model by adding a penalty for large squared weights. A higher value of reg_lambda results in more regularization, which can prevent overfitting.
- 'feature_fraction': 0.98: This specifies the fraction of features used for training each tree, which is set to 0.98 in this case. A lower value allows the model to use a smaller subset of features, which can reduce the risk of overfitting, but may also result in loss of important information.

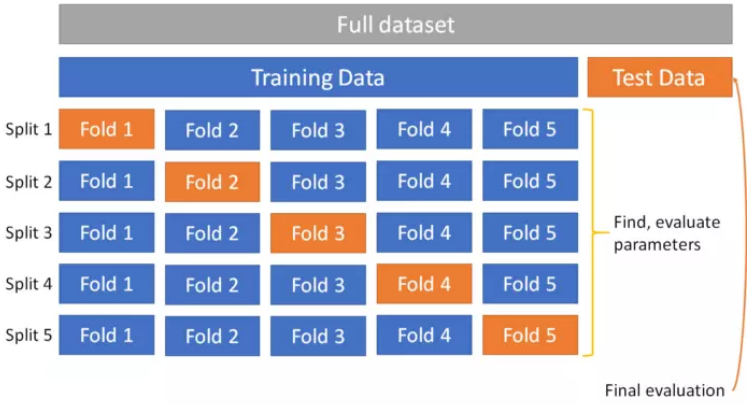
The process of tuning the LightGBM parameters involves experimenting with different values for each parameter and evaluating the model performance using appropriate evaluation metrics, such as RMSE in this case. The goal of parameter tuning is to find the optimal combination of parameter values that results in the best performance of the model on the specific time series prediction task.

Time Series Cross-validation

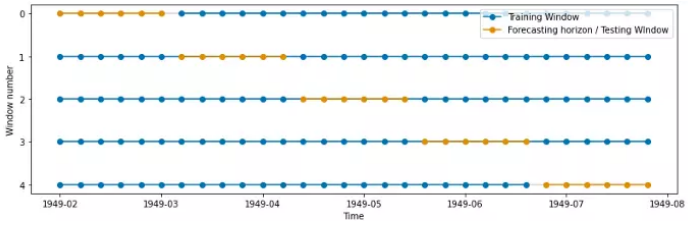
Cross-validation is a useful procedure for helping machine learning models to select optimal hyper parameters. It is particularly beneficial for smaller datasets that do not have enough data to create representative training, validation, and

test sets. In simple terms, cross–validation involves splitting a single training dataset into multiple subsets for training and testing purposes.

The simplest form of cross–validation is k–fold cross–validation, which involves dividing the training set into k smaller sets. For each split, the model is trained on k–1 sets as the training data, and the remaining set is used for validation. Then, for each split, the model is scored on the remaining set. The scores are averaged across all the splits to obtain the overall performance of the model.



However, this hyper parameter tuning approach is not applicable to time series forecasting! The figure below illustrates why standard k–fold cross–validation (and other non–time–based data splits) are not suitable for time series machine learning. The figure shows a univariate sequence divided into five windows, indicating which dates in the sequence are assigned to which fold.



There are three prominent issues:

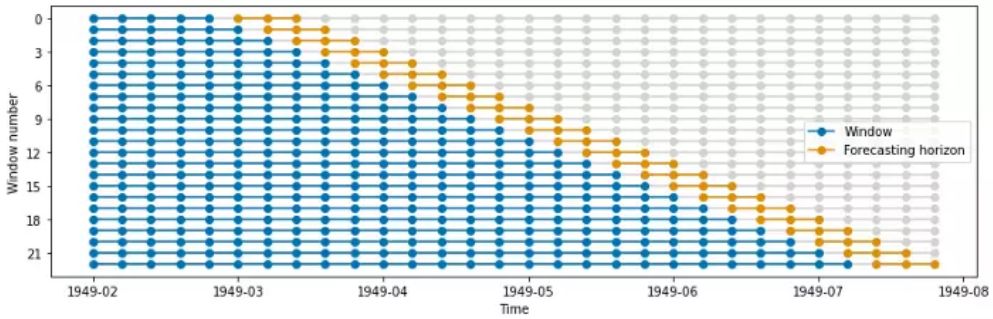
1. Prediction/test data occurs before training data. In fold 0, the test data occurs before the training data.
2. Data leakage. In windows 2–4, some training data occurs after the test data. This is problematic as the model is able to see into the ”future”.
3. Gap in the sequence. In windows 2–4, there is a gap in the training sequence due to the test data being taken from the middle portion of the sequence.

So we use TimeSeriesSplit to valid the parameters crossly, it is a cross–validation method available in scikit–learn, a popular machine learning library in Python. It is specifically designed for time series data and allows for the evaluation of time–dependent models using time–based folds.

TimeSeriesSplit in scikit–learn works by splitting the time series data into multiple sequential folds or windows, where each fold serves as both training and

testing data in a rotating manner. It ensures that the test data only comes after the training data, preserving the temporal order of the data.
we test TimeSeriesSplit and demonstrake with plot as below:

```
from sklearn.model_selection import TimeSeriesSplit
X = np.array([[1, 2], [3, 4], [1, 2], [3, 4], [1, 2], [3, 4]])
y = np.array([1, 2, 3, 4, 5, 6])
tscv = TimeSeriesSplit(n_splits=3)
for train, test in tscv.split(X):
     print("%s %s" % (train, test))
[0 1 2] [3]
[0 1 2 3] [4]
[0 1 2 3 4] [5]
```



We set the 10 splits for TimeSeriesSplit in our task to fulfil the validation needs:

```
tscv = TimeSeriesSplit(n_splits=10)
for train_index, test_index in tscv.split(X):
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]
```

Result

We utilized lightGBM to predict the S&P index prices based on time series data. In addition to the original features, we performed feature engineering by appending new features to optimize the model. This process involved adding new relevant features to the dataset to potentially capture additional patterns and relationships in the data.

The feature engineering step was crucial in enhancing the performance of the model. By carefully selecting and engineering new features, we aimed to improve the model's ability to capture the underlying patterns and trends in the time series data. Tuning while adding new features into our model.

After applying feature engineering, we were able to effectively control the Root Mean Squared Error (RMSE) to around 300, which indicates the accuracy of our predictions. RMSE is a commonly used evaluation metric in time series forecasting, and a lower RMSE value indicates better predictive performance of the model.

In summary, we utilized lightGBM for time series prediction of S&P index prices, and through the process of feature engineering, we appended new relevant features to the original dataset to optimize the model's performance, ultimately achieving an RMSE of around 300.

Conclusion

1. Is M2 money supply a useful predictor of stock prices (USA M2 supply vs S&P500)?

YES.

From the Model2 weekly data results

```
Coefficients:
                          Estimate Std. Error t value
(Intercept)                1.78376    1.53755   1.160
L(d(SP_500_W), c(3, 5))3 -0.05530    0.02877  -1.922
L(d(SP_500_W), c(3, 5))5 -0.08166    0.02881  -2.835
d(M2_W)                   0.05074    0.02463   2.060
                          Pr(>|t|)
(Intercept)               0.24622
L(d(SP_500_W), c(3, 5))3  0.05483 .
L(d(SP_500_W), c(3, 5))5  0.00466 **
d(M2_W)                   0.03964 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.96 on 1200 degrees of freedom
Multiple R-squared:  0.0132,    Adjusted R-squared:  0.01073
F-statistic:  5.35 on 3 and 1200 DF,  p-value: 0.001161
```

By now, we are able to interpret the results.

△SPFt=1.783+ 0.039 △M2t + 0.054 △SPFt−3+0.004 △SPFt−5

- In short term

  The difference of M2 increases 1 unit leads to leads to 0.039 unit increase in the difference of SP500. Looking at lag q=3, an increase in the difference of SP500 will lead to a increase in the difference of SP500 by 0.05 unit, and 1unit increase of diff(SP500) in the 5th lag will increase only 0.004 unit of diff(SP500). But we can see that the M2 and historical value of SP500 are jonitly significant and postive increase the SP500 current value.

- In middle or long-term

  $LTM = \sum_{i=1}^{+\infty} \beta_i$ =0.039+0.054+0.004=0.097

if there is a one−perid unsteady shock in △M2 and △SPF , then we expect the distortion from the steady state will sum up to 0.097

So the real mid−and long−term stock value of SP500 = 0.097/sum(1+r)^i, which r present the future discount rate, If the future discount rate r is positive, the denominator is greater than 1, and the longer the period and bigger the r, the stock real value still increase but smaller(<0.097), otherwise, when r is negative, and the longer period, the

stock real value will increase more and more (>0.097). Which means when interest rates decrease in the future, people will reduce their deposits and increase their stock investment to obtain higher returns, so SP500 will increase。 On the contrary, if interest rates increase in the future, people will increase their deposits and reduce their stock investment.

From the Model3 monthly data results

```
Time series regression with "zoo" data:
Start = 2000-05-01, End = 2022-10-01

Call:
dynlm(formula = d(SP_500_M) ~ L(d(SP_500_M), c(1, 3)) + L(d(M2_M)),
    data = data_monthly)

Residuals:
    Min      1Q  Median      3Q     Max
-424.59  -32.78   11.17   48.12  323.58

Coefficients:
                          Estimate Std. Error t value   Pr(>|t|)
(Intercept)               -6.45741    6.58071  -0.981     0.3274
L(d(SP_500_M), c(1, 3))1 -0.20012    0.06062  -3.301     0.0011
L(d(SP_500_M), c(1, 3))3  0.12116    0.06204   1.953     0.0519
L(d(M2_M))                0.25388    0.05302   4.789 0.00000279

(Intercept)
L(d(SP_500_M), c(1, 3))1 **
L(d(SP_500_M), c(1, 3))3 .
L(d(M2_M))               ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 92.78 on 266 degrees of freedom
Multiple R-squared:  0.09951,   Adjusted R-squared:  0.08935
F-statistic: 9.798 on 3 and 266 DF,  p-value: 0.000003751
```

- In short term

  The difference of M2 monthly increases 1 unit leads to leads to 0.254 unit increase in the difference of SP500. Looking at lag q=1, an unit increase in the difference of SP500 will lead to an decrease in the difference of SP500 by 0.2 unit, and 1unit increase of diff(SP500) in the 3th lag will increase 0.121 unit of diff(SP500). And postive increase M2 the SP500 current value montly, however, the previous diff(SP500) increase will decrease the SP500 now, while the 3th lag increase will increase the SP500. The possible reason is that the stock growth last month in the short term will make short-term investors take a wait-and-see attitude, thereby reducing current investment.

- In middle or long-term

$LTM = \sum_{i=1}^{+\infty} \beta_i$ =-0.200+0.121+0.254=0.175

if there is a one-perid unsteady shock in △M2 and △SPF , then we expect the distortion from the steady state will sum up to 0.175

So the real mid-and long-term stock value of SP500 = 0.097/sum(1+r)^i, which r present the future discount rate, If the future discount rate r is positive, the denominator is greater than 1, and the longer the period and bigger the r, the stock real value still increase but smaller(<0.175), otherwise, when r is negative, and the longer period, the stock real value will increase more and more (>0.175). Which means when interest rates

decrease in the future, people will reduce their deposits and increase their stock investment to obtain higher returns, so SP500 will increase。 On the contrary, if interest rates increase in the future, people will increase their deposits and reduce their stock investment.

2. At which frequency (weekly or monthly) the relationship is stronger?

Compare AIC, BIC, R-squared and Adjusted R-Squard fromweekly and monthly model, the monthly model is better.

- The R-Squared of Weekly ARDL model is 0.01, and Adjusted R-square is 0.01, but comapre to the monthly ARDL model which is 0.10 and 0.08, so monthly is better to explain the dependent variable SP500
- AIC, BIC： ardl02 is weekly model, ardl04 is monthly model, so still monthly is better

	df	AIC
ardl02	5	12935.517
ardl04	5	3218.515

	df	BIC
ardl02	5	12960.984
ardl04	5	3236.507

By now, we are able to interpret the results.

△SPF_MONTHt=−6.457+ 0.254 △M2_MONTHt −0.200 △SPF_MONTHt−1+0.121△SPF_MONTHt−3

3. Which other predictors, except M2, play an important role in a predictive model?

From model 4 result

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.129336 -0.023899  0.007239  0.028007  0.098238

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.0061437  0.0043263  -1.420  0.15676
d(log(M2_M))        0.8582515  0.3993495   2.149  0.03253 *
d(unemployment)     0.0169603  0.0051535   3.291  0.00113 **
L(d(unemployment))  0.0108438  0.0051226   2.117  0.03521 *
d(log(GDP))         1.0856440  0.3559591   3.050  0.00252 **
L(d(log(GDP)))      1.0035723  0.3569775   2.811  0.00530 **
bank_borrow        -0.0002292  0.0001071  -2.139  0.03337 *
d(log(asset_mean)) -0.1924112  0.0665405  -2.892  0.00415 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04324 on 264 degrees of freedom
Multiple R-squared:  0.1013,    Adjusted R-squared:  0.07746
F-statistic: 4.251 on 7 and 264 DF,  p-value: 0.0001815

> anova(ardl07,ardl08)
Analysis of Variance Table

Model 1: d(log(SP_500_M)) ~ d(log(M2_M)) + L(d(log(M2_M))) + d(unemployment) +
    L(d(unemployment)) + d(log(GDP)) + L(d(log(GDP))) + bank_borrow +
    d(log(asset_mean))
Model 2: d(log(SP_500_M)) ~ d(log(M2_M)) + d(unemployment) + L(d(unemployment)) +
    d(log(GDP)) + L(d(log(GDP))) + bank_borrow + d(log(asset_mean))
  Res.Df     RSS Df   Sum of Sq      F Pr(>F)
1    263 0.49262
2    264 0.49360 -1 -0.00097899 0.5227 0.4703
```

△log(SPFt)=−0.006+ 0.858 △log(M2)t + 0.016 △UNEMPLOYMENTt+0.01 △UNEMPLOYMENT(t−1)+1.08 △log(GDP)t+1.0△log(GDP)(t−1)−0.0002 BANKt−△log(ASSET)t

- We can see M2, Unemployment rate, and GDP has positive influence on the SP500, when these 3 variables increases , it will increase the stock market.
    - For GDP, GDP is a measure of the overall economic activity in a country, and a positive increase in GDP generally indicates that the economy is growing. When the economy is

expanding, it can lead to increased corporate profits for companies listed in the S&P 500, which can positively impact stock prices and drive the index higher.

    - But for the unemployment rate, an increase in the unemployment rate would lift stocks is surprising. However, there are some possible explanations for why the unemployment rate may increase while the S&P 500 also increases. For examples, Interest rates and monetary policy: Unemployment can also impact monetary policy decisions made by central banks, which can in turn influence stock prices. For example, during periods of high unemployment, central banks may implement monetary policies, such as lower interest rates or quantitative easing, to stimulate economic growth and reduce unemployment. These policies can impact borrowing costs for businesses, influence consumer spending, and impact overall market sentiment, which can potentially impact the performance of the S&P 500.

- Bank Borrowing Percent Change at Annual Rate and average value of asset has negative influence on the SP500.

A decrease in bank borrowing could indicate reduced credit availability for businesses and consumers. This can lead to decreased borrowing and spending by businesses, which may result in reduced investments, lower production levels, and potentially lower corporate profits. Reduced credit availability can also impact consumer spending, as consumers may have less access to credit for making purchases, leading to decreased consumer spending. Lower corporate profits and reduced consumer spending can negatively impact the overall economic activity, which can, in turn, lead to lower stock prices, including those of companies listed in the S&P 500.