GROUP WORK PROJECT # 3
GROUP NUMBER: 3997

MScFE 610: FINANCIAL ECONOMETRICS

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Bharat Swami	India	bharatswami1299@gmail.com	
Ka Man Lui	United Kingdom	thomaslui.0924@gmail.com	
Daxin Niu	United States	daxinniu.work@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	Bharat Swami
Team member 2	Ka Man Lui
Team member 3	Daxin Niu

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.

GROUP WORK PROJECT # 3 MScFE 610: FINANCIAL ECONOMETRICS
GROUP NUMBER: 3997

Step 1

As a group we have decided we are going to work upon the Modeling of the Non-Stationary and finding an equilibrium.

First we are going to choose the dataset individually and demonstrate why we have chosen this particular dataset, what type of property this time series data shows and how we are going to use the selected model for it. And then we are going to choose a dataset from these three dataset as a group and work on it for modeling the non-stationary time series and finding its equilibrium.

After selecting a dataset as a group we are going to run different tests on that dataset which we already ran in GWP1 and GWP2, and finally we are going to run UECM to find the coefficient for equilibrium.

Step 2

Individually:

Dataset 1: Stock price of Microsoft and Apple

First dataset we were suggesting was the daily stock price of Microsoft and Apple data from 01/01/2000 to 12/31/2022 from Yahoo api. To test non-stationary, we could use the commonly used ADF test.

The reason why we recommended this dataset is that Microsoft and Apple are the well-known technology companies in the world and we want to test if there's any relationship in the stock price time series model.

We have conducted the test to check if these time series are non-stationary or not by Augmented Dickey–Fuller test and KPSS test. From those tests, we can conclude that they are non-stationary. After confirming they are non-stationary, we can conduct different tests to see if there're any relationship between these two time series. We can see if there is any arbitrage opportunities by exploiting the statistical relationship between the two stock prices.

Dataset 2: SPY

One dataset we were suggesting was the SPY daily price data from 01/01/2022 to 12/31/2022. To test non-stationary, we could use the commonly used ADF test.

The reason why we recommended this dataset is that the SPY is a very commonly used dataset. It is also essentially tracking the overall market. Therefore, it is a quite credible source of data.

There are some potential issues as to why we didn't pick this to be our final dataset. It is too generic and we might want to use something that is more interesting. Furthermore, it is too commonly used in the market making it a bit more difficult to profit off when compared to other datasets.

Dataset 3: NSE NIFITY50

NSE NIFITY50 Index is a very popular index in NSE Exchange in India. It is a weighted average of 50 most popular or biggest companies listed with NSE. We are going to use the daily Adjusted Close of Nifty50 for a time period of 10 years from 2012-02-01 and 2022-02-01.

Since this index contains the biggest companies listed with exchange it is very stable with time. We may find trend and non-stationary properties in this time series data. As we conducted tests on this dataset in GWP2, we found that we have trend and non-stationary properties in the dataset. After different tests on the dataset we concluded we need to take differences of values in the dataset. After taking the first difference we are able to see that the trend and non-stationary properties are diminished from the dataset, which makes the dataset Stationary and we can use this transformed dataset for our future use i.e., to model the index close price.

Main reason to recommend this index to include in our study or any portfolio is that it is very stable as compared to other time series data in NSE, and being an index gives an advantage from a risk point of view.

Step 3

Definition: As we mentioned before we are going to choose the dataset and will try to model the non-stationary time series dataset and find the equilibrium for it. For this we are going to use the dataset 1 which is the stock price of Microsoft and Apple.

In particular we are using the UECM model* which is the Unrestricted error correction model. If we have Y as dependent variable and X as independent variable, following is the equation for modeling the model (Hardy ,2017) -

$$\Delta Y_t = \alpha_0 + \sum_{n=1}^{N} \beta_1 \Delta y_{t-i} + \sum_{n=1}^{N} \beta_2 \Delta x_{t-i} + \gamma_1 y_{t-1} + \gamma_2 x_{t-1}.$$

And finally, what do we call it unrestricted ECM, because we are specifying the long run relationship in the model with the help of β_1 and β_2, if the both are zero we are not doing much with the model, we are just going to relate the first difference of dependent variable with previous value. But by putting both beta parameters non-zero we are taking long-term and short-term relation of previous values of both dependent and non-dependent variables into account.

**For a working model please refer to the python code.*

Description: Now we are going to assume that the dependent variable is Microsoft stock value in the first equation and Apple in the second equation. We are only going to find one parameter for both stock price, which is calculated in the demonstration below.

Demonstration:

Stock price of Microsoft and Apple from 2000 to 2022 is downloaded from the yahoo finance API. After identifying if there're any unit roots for the two stock price time series, cointegration was tested for these two time series.

After the two time series are confirmed cointegrated, A Error Correction Model was run using the time series of return difference in stock price.

UECM Model Results

Dep. Variable:	D.MSFT	No. Observations:	5787
Model:	UECM(1, 1)	Log Likelihood	-9365.433
Method:	Conditional MLE	S.D. of innovations	98.323
Date:	Mon, 25 Sep 2023	AIC	18740.867
Time:	19:11:25	BIC	18774.183
Sample:	1	HQIC	18752.458
	5787		

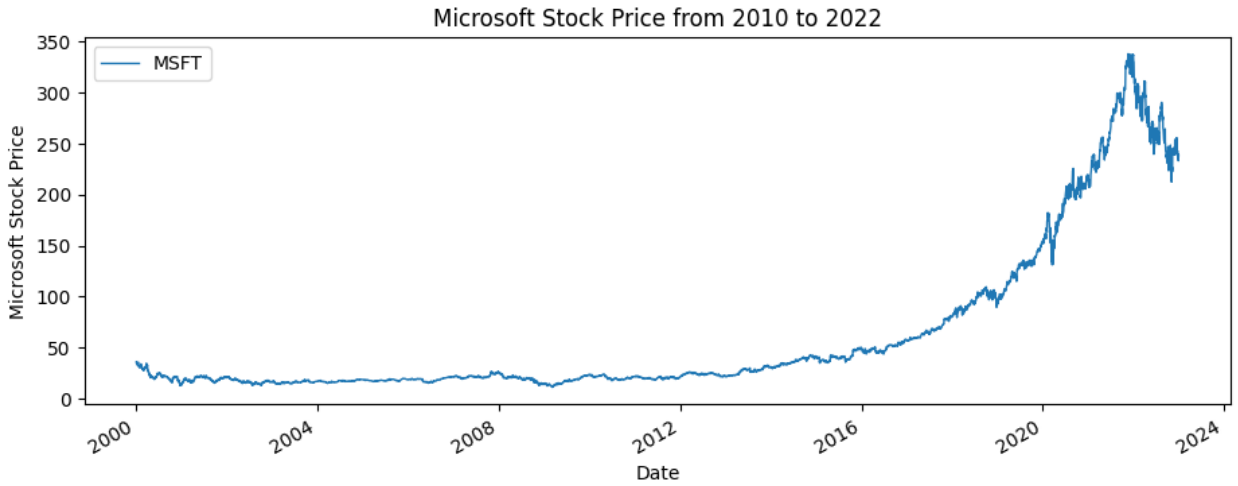
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0441	0.024	1.827	0.068	-0.003	0.091
MSFT.L1	-0.0031	0.001	-2.395	0.017	-0.006	-0.001
AAPL.L1	0.0053	0.002	2.304	0.021	0.001	0.010
D.AAPL.L0	1.3185	0.015	85.261	0.000	1.288	1.349

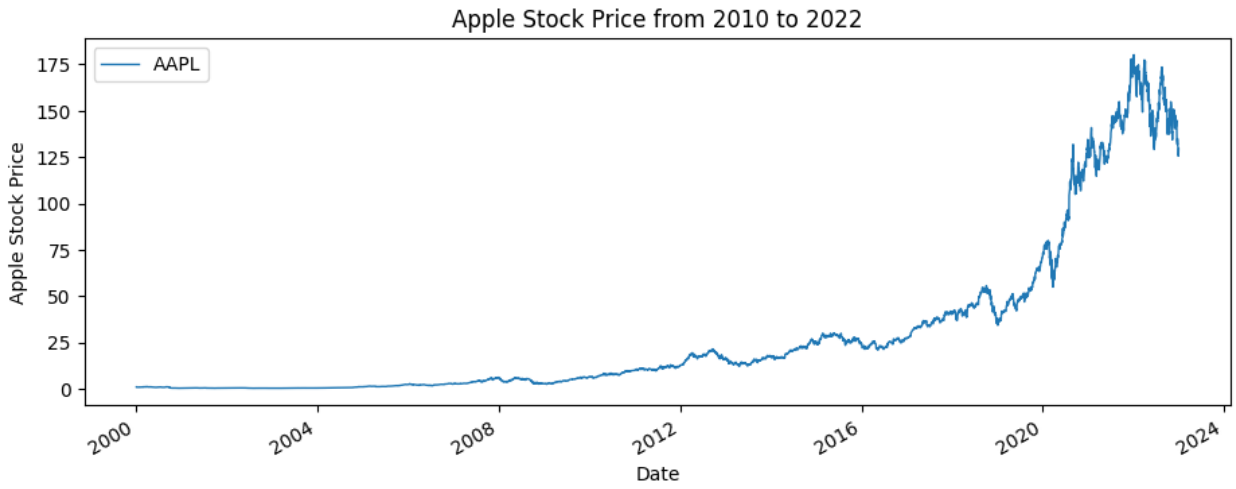
From the ECM result above for stock MSFT and AAPL, we can see all coefficient estimates are significant. The coefficient of the error correction term (i.e. MSFT.L1) is -0.031. It means that if the stock price MSFT moves 1 unit from the long-term equilibrium with AAPL at time t-1, MSFT will decrease by 0.0031 unit at time t to return to the long-term equilibrium.

The term AAPL.L1 and it's coefficient 0.0053 captures the short-term impact of the movement of AAPL to MSFT

Diagram:

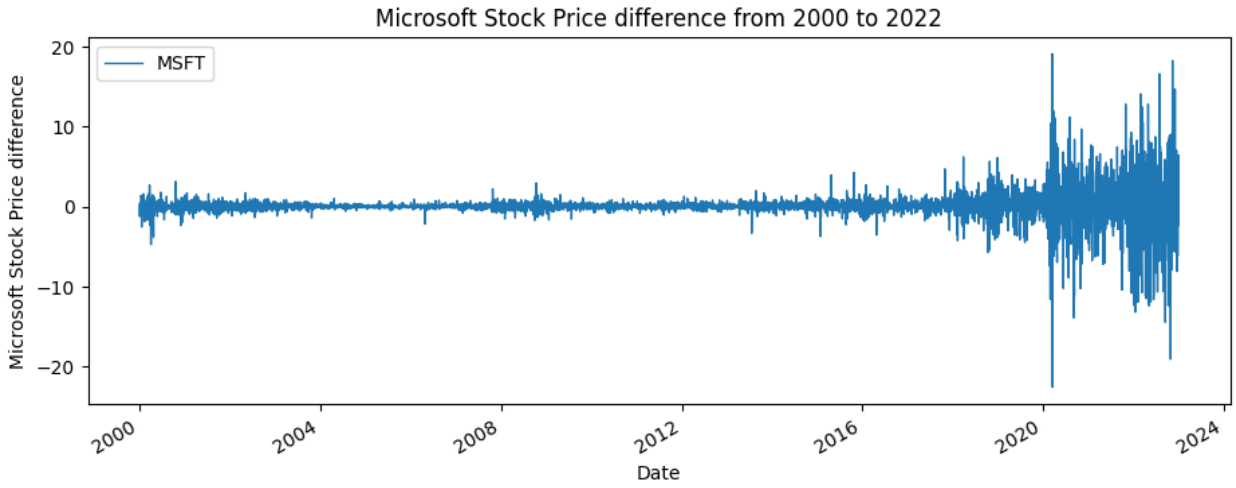
We first need to check whether the stock price of MSFT and AAPL are stationary or not and whether they have unit roots or not.

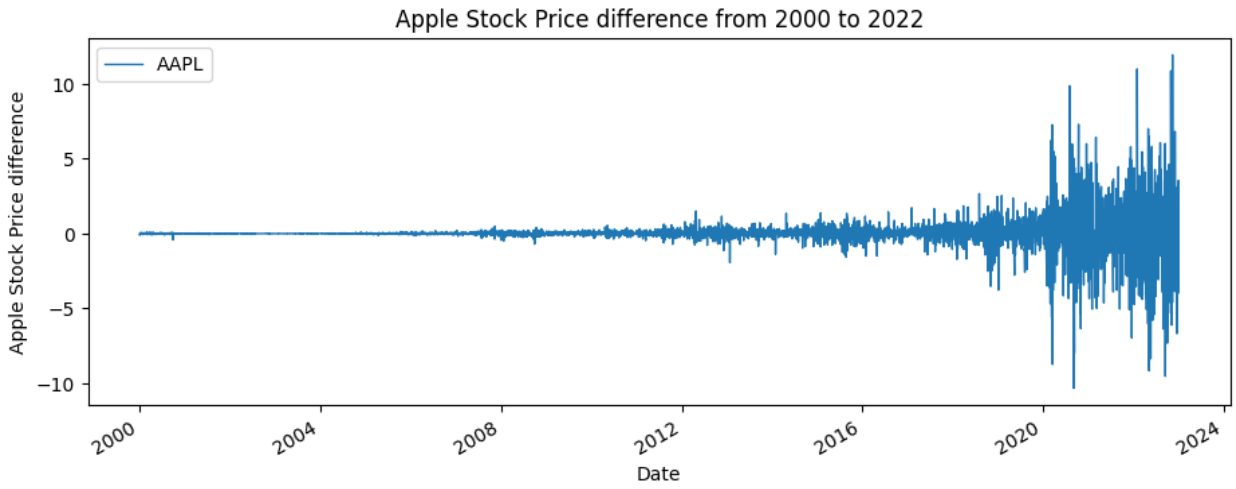




From the plots on the stock price above, we can see that the time series of MSFT and AAPL stock prices are non-stationary. For detecting unit roots, Augmented Dickey–Fuller test and KPSS test were used. Next, we conduct the Phillips and Ouliaris method to check if cointegration exists between the two data series and result shows that there exists cointegration between the stock price of the two companies.

The difference of the stock prices has also been plotted as below:





The mean of the difference for both stock price is around zero. The large fluctuation of the difference mainly cluster between 2020 and 2022 but they are still finite. It looks like they fulfill the requirement of weak stationarity.

Diagnosis:

The Augmented Dickey–Fuller test and KPSS test were conducted for stock price of Microsoft and Apple to check whether there is any unit root for the time series:

Augmented Dickey–Fuller test Without drift and trend

	Microsoft Stock Price	Apple Stock Price
ADF Test Statistic	1.819197	1.446088
5% Critical Value	-1.941047	-1.941047

KPSS Test

```
    KPSS Stationarity Test Results
====================================
Test Statistic                  3.152
P-value                         0.000
Lags                               34
------------------------------------

Trend: Constant and Linear Time Trend
Critical Values: 0.22 (1%), 0.15 (5%), 0.12 (10%)
Null Hypothesis: The process is weakly stationary.
Alternative Hypothesis: The process contains a unit root.
    KPSS Stationarity Test Results
====================================
Test Statistic                  2.767
P-value                         0.000
Lags                               34
------------------------------------

Trend: Constant and Linear Time Trend
Critical Values: 0.22 (1%), 0.15 (5%), 0.12 (10%)
Null Hypothesis: The process is weakly stationary.
Alternative Hypothesis: The process contains a unit root.
```

From both Augmented Dickey–Fuller test and KPSS test, we can see that both time series have unit roots and hance they are non-stationary.

After the check of unit roots and stationarity, Phillips and Ouliaris method was conducted to see if there's any cointegration between two time series.

```
Phillips-Ouliaris Za Cointegration Test
Statistic: -30.839740215222147
P-value: 0.005486115936825
Null: No Cointegration, Alternative: Cointegration
Kernel: Bartlett
Bandwidth: 8.1576
Trend: c
Distribution Order: 3
```

From the result, cointegration exists between the two stock price data series.

The Augmented Dickey–Fuller test and KPSS test were conducted for stock price difference of Microsoft and Apple to check whether there is any unit root for the time series:

Augmented Dickey–Fuller test Without drift and trend

	Microsoft Stock Price	Apple Stock Price
ADF Test Statistic	-25.757806	-22.437883
5% Critical Value	-1.941047	-1.941047

KPSS Test

```
    KPSS Stationarity Test Results
=====================================
Test Statistic                  0.068
P-value                         0.369
Lags                               34
-------------------------------------

Trend: Constant and Linear Time Trend
Critical Values: 0.22 (1%), 0.15 (5%), 0.12 (10%)
Null Hypothesis: The process is weakly stationary.
Alternative Hypothesis: The process contains a unit root.
    KPSS Stationarity Test Results
=====================================
Test Statistic                  0.052
P-value                         0.550
Lags                               34
-------------------------------------

Trend: Constant and Linear Time Trend
Critical Values: 0.22 (1%), 0.15 (5%), 0.12 (10%)
Null Hypothesis: The process is weakly stationary.
Alternative Hypothesis: The process contains a unit root.
```

From both Augmented Dickey–Fuller test and KPSS test, we can see that both time series don't have unit roots and they are weakly stationary.

Durbin-Watson Test Result was conducted to see if the residual residuals from the ECM model result are serially correlated. The test statistics is 2.113 and it suggests the residual residuals from the result are serially uncorrelated.

Damage:

From our experiment with the MSFT and AAPL data, we were able to identify that cointegration exists between these two stocks. We performed ridge experiments following several steps. We first checked that both stocks have unit roots and are non-stationary. After that, experimented with the difference in their returns. In the end, we rejected the null hypothesis and concluded that there is cointegration

between their returns. Given these results, we decided to use the error correction model since they are cointegrated.

These procedures are quite standard and usually should not bring too much damage. Nonetheless, one important idea is that we should keep checking out assumptions and making sure they hold across time. Theoretically and intuitively, it makes sense that MSFT and AAPL are cointegrated since they are both big tech companies with many similarities in between. Nonetheless, suppose some event happens in the future and one of them changes their business model completely, it is possible that they are no longer cointegrated. In a situation like that, there could be some damage happening if we stick to our old model. Suppose we have a long/short strategy that profits when these two stocks diverge and revert back to their original positions. This strategy would work when the stocks are cointegrated. We know that temporary divergence of the stocks will eventually turn back to their original conditions due to the fact that they are cointegrated. But if one of them changed their business model, the two companies might no longer be cointegrated. In a situation like this, there is a possibility that they will diverge forever. That could result in big losses from our strategy.

Directions:

It is possible that the model doesn't fit well with our data. In a situation like this, it is dangerous to manipulate the data and make it fit to the model. A more reasonable way is to change to a different model and see if it could fit.

When a model doesn't fit the data perfectly, sometimes, it is okay to keep using it. It might be more well generalized and prevent the issue of overfitting. Nonetheless, if the data is far off from the model, it might simply indicate that the model cannot fit the data or predict the data well. In this situation, we should not be using the same model anymore. Instead, finding another model that can better explain our observation is a more logical approach.

Manipulating data is something we should be very cautious about. In a lot of situations, data are real-world observations that happened in the past. If we decided to input our bias and modify the data, the model could fit the manipulated data temporarily. Nonetheless, when we are using it for prediction purposes, it will not do a good job since it has introduced bias in its training data. Therefore, it can be quite dangerous to manipulate data to make it fit the model.

Deployment:

There are many different ways that we can use our whole model. The whole testing pipeline can be used to check whether two stocks are cointegrated. This is important for us to keep track of our assumptions. After that, we can use the error correlation model to help us check for cointegration. If cointegration exists, we should be able to form a strategy out of it. One common strategy is long-short equity. We can make predictions on when stocks could diverge. In that situation, we can long one and short one. When the stocks converge back at the end, we will be able to profit from it. Our current model will be an essential part of the process.

Reference:

1. Hardy Richard (2017), ARDL/Error Correction Model: long vs. short run, restricted vs. unrestricted, https://stats.stackexchange.com/questions/135616/ardl-error-correction-model-long-vs-short-run-restricted-vs-unrestricted