GROUP WORK PROJECT # 3
GROUP NUMBER: 5338

MScFE 610: FINANCIAL ECONOMETRICS

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Shivansh Kumar	India	shivansh.business23@gmail.com	
Vaibhav Sandeep Somani	Ireland	vaibhavssomani@gmail.com	
Kshitiz Sharma	India	drkshitizsharma@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	**Shivansh Kumar**
Team member 2	**Vaibhav Sandeep Somani**
Team member 3	**Kshitiz Sharma**

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.
N/A

Model Selection

For this group work project, we are going to work on modeling non-stationarity and finding an equilibrium. We are going to apply different methods to model non-stationarity on a multivariate time series dataset, we are going to see an econometric unit root test to show non-stationarity, multiple visualization checks, regression model, co-integration test, equilibrium models, and at last the forecast of the model.

Basic Description of Individual Work

Team Member 1:

Team member 1 have worked on modeling SPY, QQQ, VGT ETF's daily price time series data, in which visualization check like time series plot, ACF plot and econometric unit test like ADF and KPPS has been done apart from that team member 1 have also differenced the data and verified if the data becomes stationary or not after verifying the non-stationarity of the data a VAR model has been modeled to check for lags to use in VEC model. After that a Johansen Trace test was done to check for cointegration and after getting a cointegration relationship a VECM model was modeled to get some linear combination of the variables which indicates a long-term equilibrium then we leverage that linear combination and plot the long-term equilibrium of the time series. After that, we check if the long-term equilibrium is stationary or not using ADF and at last we VEC model result to forecast the time series.

Team Member 2:

Team member 2 has worked on the modeling of the NIFTY 50 Index, INFY, NIFTY IT Index daily price time series. As we can see, there are two indices - NIFTY 50 and NIFTY IT whereas one stock - INFOSYS which is an IT Company. First the ADF Plot for the time series to check for stationarity and the existence of unit roots, after that the KPSS test to confirm for the stationarity. From those tests, it was concluded that the time series is not stationary. After that selection of the Number of Lags for the VAR Model, which suggested 1 lag. After that, the cointegration between three series using the Johnson Trace Test. Later, a VECM model establishes a linear combination of the time series - NIFTY, IT and INFY and the coefficient as well . VECM indicates a long-term equilibrium. ADF test is repeated to test the stationarity of the VECM model. Lastly, using the VEC Model forecasted the future time series.

Team Member 3:

As team member 3, I have worked on time series data on the daily prices of three popular stocks namely Apple Incorp (AAPL), Intel Corporation (INTC), and The Walt Disney Company (DIS). The AAPL and INTC is a NasdaqGS real-time price data, DIS is NYSE delayed price data. All the three categories are different. Apple is in Consumer Electronics, Intel is in semiconductor and Disney is in the entertainment industry. The first step was to plot a line graph of three-time series and to understand the variation in the price over time. Apple and Intel show similarities whereas Disney's price progressed away from these two. The next step is to test the time series for stationarity (one of the mandatory conditions for time series models). Augmented Dickey-Fuller ("ADF") test and Kwiatkowski-Phillips-Schmidt-Shin ("KPSS") test are applied for the stationarity test. Both these tests confirm non-stationarity. Now a differenced method is used to test the convergence to stationarity of the time series. A VAR model has been modeled to confirm the lags to use in the VEC model. The next step is to test the cointegration between three series, Johnson Trace Test confirms cointegration as the mean difference remains constant in the series. A VECM model establishes a linear combination of the variables. VECM indicates a long-term equilibrium. ADF test is repeated to test the stationarity of the VECM model.

For a More detailed report please refer to the notebook submitted or you can also go to this colab link:

Group_5338_Group_Project_Work_3.ipynb

Modeling non-stationarity and finding an equilibrium

Chosen dataset

For our Modeling we are going to use 5 years of daily data from 2019-2024 using Yahoo Finance API, the data we are collecting are from 3 major liquid ETF SPY (SPDR S&P 500 ETF Trust), QQQ (Invesco QQQ Trust), VGT (Vanguard Information Technology Index Fund ETF Shares).

Reasons for selecting these ETFs:

- Market representation
- Liquidity
- Sector Diversification
- Relevance to Economic Trends

Overall, selecting these ETFs for analysis provides a balanced representation of the market, ensures data reliability due to high liquidity, and allows for insights into both broad market trends and sector-specific dynamics, especially in the technology sector, also analyzing stationarity and equilibrium between these ETF's provides long term trends and market behavior.

Definition:

Stationarity is a key property for time series (Yong and Liu). A stationary time series is one whose statistical properties (such as mean, variance, and autocorrelation) remain constant over time.

Stationarity of Time Series:
Let $\{X_t\}$ be a time series with finite variance, $\{X_t\}$ is stationary if

- The mean function $\mu_X(t)=E(X_t)$ is constant and independent of t
- Autocovariance $\gamma_X(t-h,t)$ is independent of t for each h. This means autocovariance is only dependent on the time difference **h**.

Non-stationarity refers to the situation where the statistical properties of a time series change over time.

Two approaches to model non-stationarity are:

- Deterministic Trend with Stationary Disturbances
- Unit Root Process (Non-Stationary Disturbances)

Description:

Modeling non-stationarity and finding an equilibrium

a) Unit root process

A unit root refers to a situation where a shock (change) to a series in one period has a permanent effect on its future values; the presence of a unit root is a non-stationarity of the time series (Nasir Hamid Rao and Abdul). A stationary time series with no unit root is integrated with order zero, or $I(0)$ ((Christopher A. and Sims)). The following tests are available for non-stationarity

- Random Walk
- Dickey-Fuller Test
- Augmented Dickey-Fuller Test
- KPSS Test

First test for weak stationarity or Strict stationarity. For weak stationarity, the mean is constant, and independent of location. For strict stationarity DF test, ADF test, and KPSS test are available based on joint probability (Akeyede and Imam).

b) Nonstationarity to stationarity

- Detrend the time series
- Differences in the time series
- Use ACF and PACF to check stationarity

c) For bivariate or two time series check for Ergodicity.

If a time series xt is stationary and $\sum^{\infty}_{j=0} cov(xt,xt-j)<\infty$, then xt is ergodic for the mean (Rao). Use VAR.

d) Vector Autoregressive Model (VAR) (*Vector Autoregressive Models for Multivariate Time Series.*)

VAR is a time series model that is used to forecast two or more time series ("Vector Autoregressive Models for Multivariate Time Series"). The VAR model can also be used to understand the joint dynamic interaction among time series.

e) Cointegration

Cointegration is used to integrate dependent and independent time series that are non-stationary (Gianluca Cubadda), The following tests are used for cointegration:

- Engle-Granger Test
- Error Correction Model

Demonstration & Diagram:

We will start our modeling by plotting time series to compare their price movement.

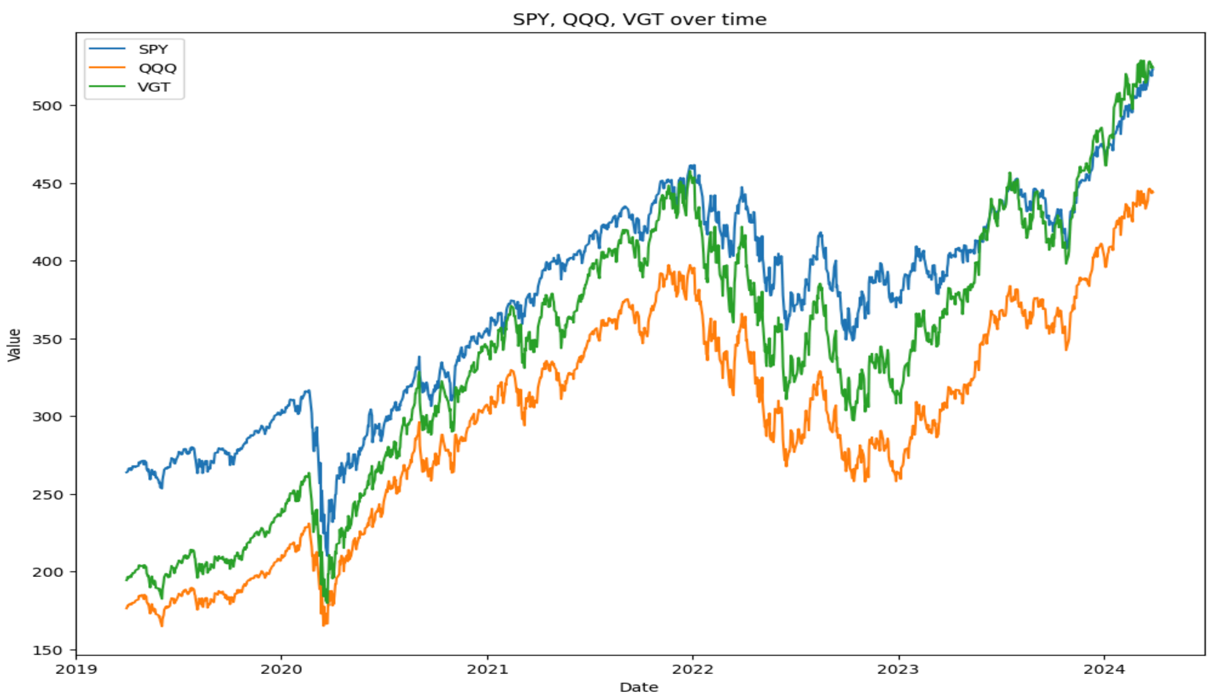

Figure 1: Time Series plot of ETF's

As you can see in Figure 1, All the ETF's are moving together in a similar trend. So we will now first do some visualization check for stationarity and after that we will do some econometric test for stationarity.

In the visualization check we have plotted the ACF plot for each ETF price. And In the ACF diagrams we saw a slow decay suggesting non-stationarity.

Now we have done some Econometrics tests like ADF and KPSS tests to check for stationarity.

	SPY Price	QQQ Price	GLD Price
ADF Test Statistic	1.700658	1.474285	1.467270
5% Critical Value	-1.941216	-1.941216	-1.941216

Figure 2: ADF Test for ETF's

In the ADF results in Figure 2, you can see none of the ETF's prices has a test statistic greater than 5% critical value so we cannot reject the H0(null) Hypothesis and there are unit roots in all three time series.

After that we analyzed the KPSS tests results where we saw all the time series have p-value lower than 0.05, which clearly rejects the H0 hypothesis, which means all these time series are non-stationary and have unit root.

Now to verify our results of non-stationarity we did differencing of the time series data and again did the visualization and econometric tests to check if the data becomes stationary or not. And for the ACF plot the prices dropped to zero quickly suggesting stationarity. Also in the ADF test of differenced data we could see none of the ETF's prices has a test statistic that is lower than 5% critical value so we can reject the H0 (null) Hypothesis and there are no unit roots in all three time series. At last for the KPSS Test we could see all the time series have p-value higher than 0.05, which cannot reject the H0 hypothesis, which means all these time series are stationary and have no unit root.

Now as we can see in the non-differenced data all our test suggests that our dataset is non-stationary but after differencing that data we can see all the tests are suggesting that our data is stationary. This verifies our tests as well as all the assumptions to believe that our data set is non-stationary.

As we have confirmed that our data is non-stationary we will now do some regression modeling before that we need to decide the number of lags in the VEC (regression) model before testing it for cointegration.

VAR Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	17.79	17.80	5.333e+07	17.80
1	4.216*	4.265*	67.76*	4.235*
2	4.217	4.303	67.83	4.250
3	4.217	4.341	67.83	4.263
4	4.225	4.386	68.39	4.286
5	4.233	4.430	68.91	4.307
6	4.240	4.475	69.43	4.329
7	4.245	4.516	69.74	4.347
8	4.232	4.540	68.84	4.348
9	4.233	4.578	68.91	4.363
10	4.221	4.604	68.13	4.365
11	4.227	4.647	68.50	4.385
12	4.236	4.693	69.15	4.408

Figure 3: VAR model Lag selection

In the Figure 3 results we can see all the information criteria suggest we should select lag 1 for the level of the VAR model.

Now as we have decided number of lag, we will now check for cointegration using Johansen Trace test :

Eigenvalues of VECM coefficient matrix : [0.01479121 0.00389051 0.00087087]

	Test statistic	Critical values (90%)	Critical values (95%)	Critical values (99%)
rank=0	24.706822	27.0669	29.7961	35.4628
rank<=1	5.990299	13.4294	15.4943	19.9349
rank<=2	1.094286	2.7055	3.8415	6.6349

Figure 4: Johnsen Trace Test

From the above Johansen Trace test results in Figure 4, we can see three tests: H0 : rank = 0, H1 : rank = 1, H2 : rank = 2. We will use 5% as our decision point.

- For H0 : rank = 0, we can see the test statistic is 24.7068 and the 5% critical value is 29.7961 which means the test statistic is less than the critical value so we do not reject the null hypothesis that there are zero cointegration relationships.
- For H1 : rank = 1, we can see the test statistic is 5.9903 and the 5% critical value is 15.4943 which means the test statistic is less than the critical value so we do not reject the null hypothesis that there are at most one cointegration relationships.
- For H2 : rank = 2, we can see the test statistic is 1.0942 and the 5% critical value is 3.8415 which means the test statistic is less than the critical value so we do not reject the null hypothesis that there are at most two cointegration relationships.

Overall the Johansen Trace Test suggested that there are at most two cointegrating relationships among the variables. This means that some linear combination of the variables is stationary, indicating a long-term equilibrium relationship.

Cointegration relations for loading-coefficients-column 1

	coef	std err	z	P>|z|	[0.025	0.975]
beta.1	1.0000	0	0	0.000	1.000	1.000
beta.2	1.2276	0.401	3.060	0.002	0.441	2.014
beta.3	-1.7958	0.332	-5.408	0.000	-2.447	-1.145
const	-127.9209	13.418	-9.533	0.000	-154.220	-101.621

Figure 5: Cointegration Linear combination

So then we ran a VEC model and using the "Cointegration relations for loading-coefficients-column 1" from Figure 5 we can write the linear combination as follows :

" S = -127.9209 + 1 . SPY + 1.2276 . QQQ - 1.7958 . VGT "

Diagnosis:

The above equation is the deviation from the long-term equilibrium of the three time series. Let's check out the plot to see if this deviation is stationary.

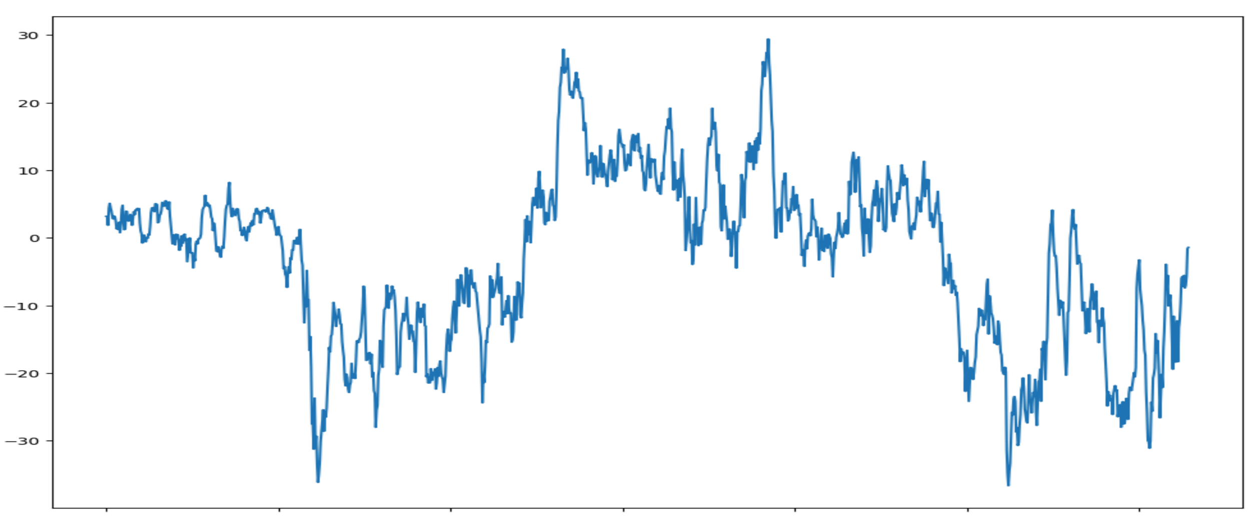

Figure 6: Long-term equilibrium

In this Figure 6 we can see a long-term equilibrium is moving within a channel. Let's take a look at the ADF test for the deviation.

```
Test statistics and critical values:
    Augmented Dickey-Fuller Results
=====================================
Test Statistic                 -3.207
P-value                         0.001
Lags                                0
-------------------------------------

Trend: No Trend
Critical Values: -2.57 (1%), -1.94 (5%), -1.62 (10%)
```

Figure 7: ADF test for equilibrium

Here in these ADF results in figure 7 we can see the P-value is less than 0.05 and the Test statistic is also lower than critical value of 5% so we can easily reject the null hypothesis and that the deviation has a unit root, which makes this deviation stationary.

At last now we will plot the forecast of our model :

In this last figure 8, we can see a time series forecast in which SPY and QQQ are slightly up but VGT remains flat in the forecast.

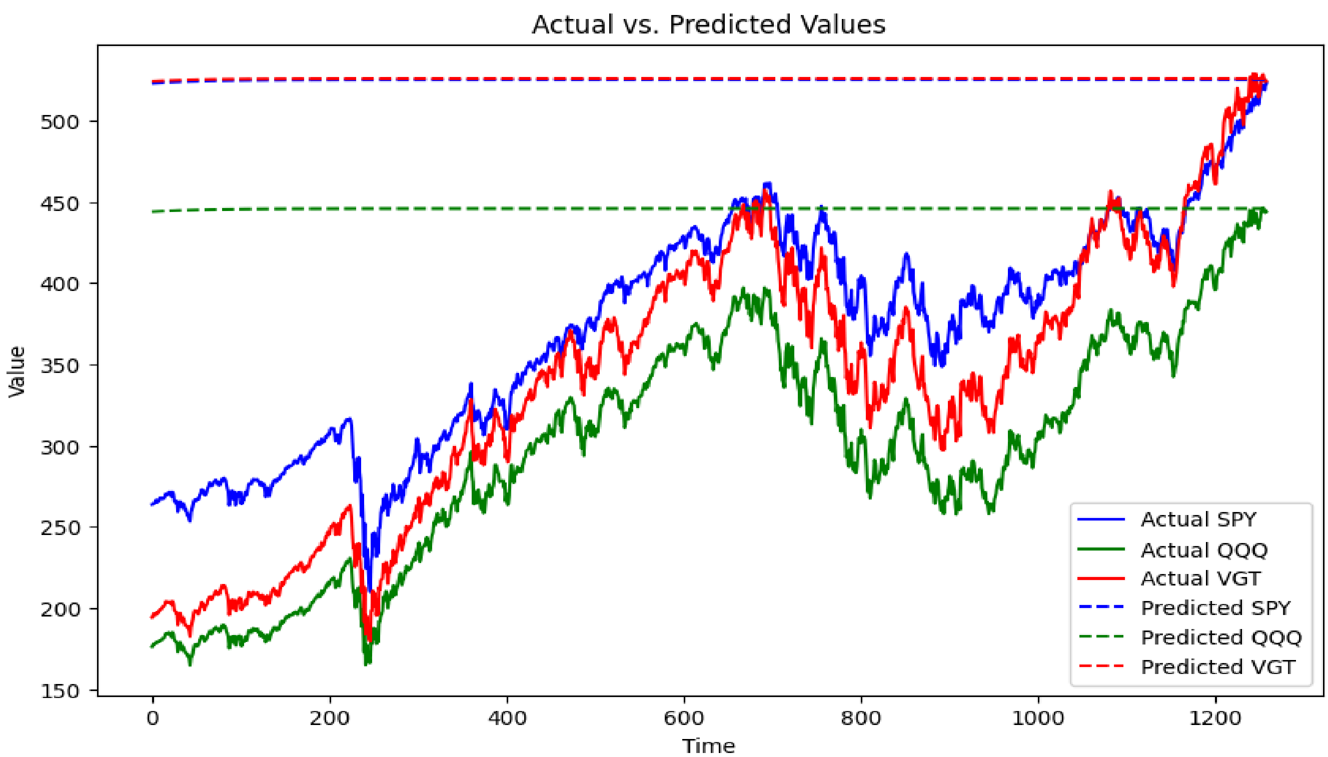

Figure 8: Actual Vs Predicted Value

Now we will discuss the Damage, Direction, and Deployment of our model.

Damage:

The Vector Error Correction Model as used in our above analysis is quite useful for dealing with the multiple time series that influence each other and exhibit a long-term equilibrium relationship, even if they fluctuate in the short term. But, some issues need to be taken care of as it might cause damage to the analysis:

1. **Non-stationarity:** The VECM model assumes that the time series data are stationary.If the time series are stationary, VECM might not be the most suitable model and could produce misleading results.
2. **Cointegration**: As we know, VECM assumes that the time series are cointegrated, if the time series do not meet this criterion, the model may not provide accurate insights.
3. **Impact of Lags Selection:** The number of lags included in the model can impact the results. If there are too few lags, important relationships might be missed, on the other hand, including too many lags can lead to overfitting and unreliable estimates.

Direction:

From the above analysis, here are a few directions that we can take to improve the performance of the VECM Model:

1. **Lags Selection:** As mentioned in the above section of damage, careful selection of the lags is very important. When we run **model.select_order** we get a list of lags with their respective AIC, BIC, FPE, and HQIC values, at times, different models can suggest different lags, so if the model doesn't fit well we can try with different lag.
2. Based on the lag as mentioned above, we get a different set of **coefficients and intercepts for our long-term equilibrium equation**, and thus, we need to meticulously do our analysis.

Deployment:

From the above model description and the code provided in the Google Collab, we can build the model. Once the optimal model is built meticulously we can leverage the model for our further analysis. Here is how we can use the model:

1. **Forecasting future values**: As used in the colab:

   vecm_model.plot_forecast(steps=n)

   This is the function we use to forecast the time series based on the past values of the time series. Here we can specify steps = n, the number of steps ahead we want to forecast. We have used n =100 for the forecasting purpose during our analysis. For the model built for our analysis this is the forecast:

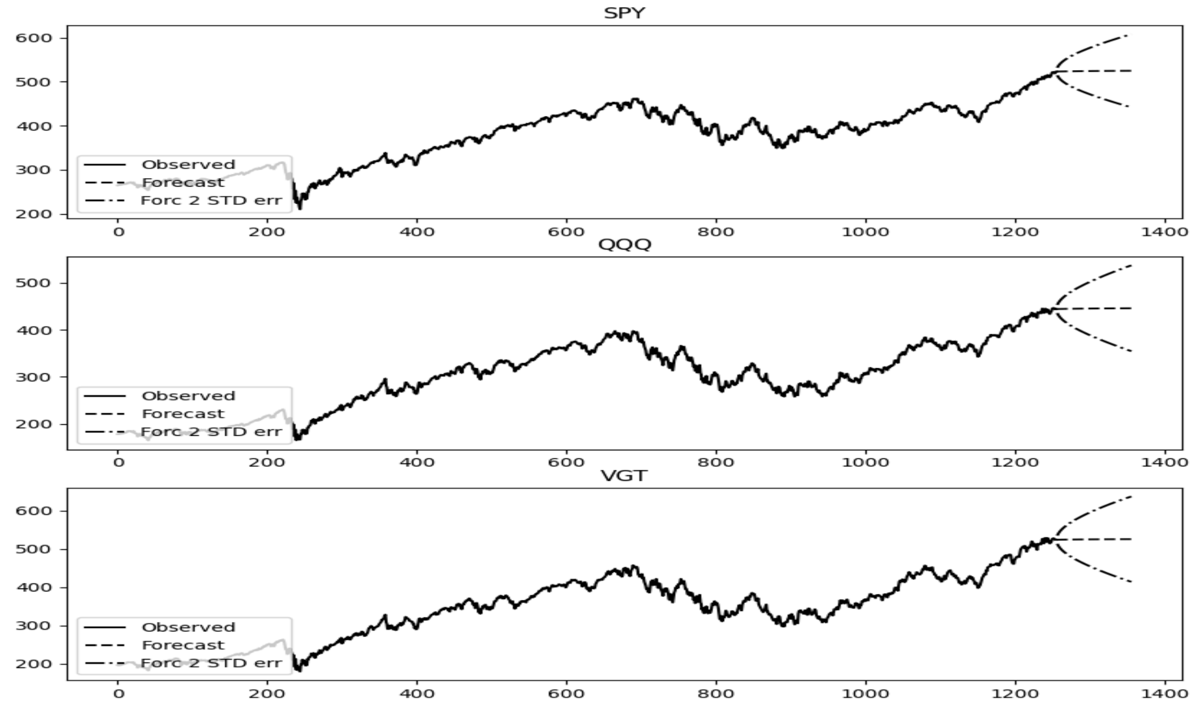

Figure 9 : Forecast Plot

2. **Interpretation:** The VECM model provides several key outputs that are crucial for interpretation:

   Cointegrating Vectors: This is the long-run equilibrium relationship between the variables. It indicates how much of one variable is needed to explain the movements in another variable.
   Adjustments Vectors: These are the Error Correction terms. These terms represent the short-term adjustments that variables make to revert to their long-run equilibrium relationships. (VECM Estimation and Interpretation)

   These vectors help us to analyze how quickly variables adjust back to equilibrium after deviation.

This is how we can use the optimal model built for our analysis.

References

1. Akeyede, and Imam. “On Consistency of Tests for Stationarity in Autoregressive and Moving Average Models of Different Orders.” *American Journal of Theoretical and Applied Statistics*, vol. 5, no. 3, 2016, p. 146, https://doi.org/10.11648/j.ajtas.20160503.20. Accessed 24 feb 2024.
2. Christopher A., and Sims. “Inference in Linear Time Series Models with Some Unit Roots.” *Econometrica*, vol. 50, no. 1, 1990, p. 113, https://doi.org/10.2307/2938337. Accessed 7 Apr 2024.
3. Gianluca Cubadda. “Common Cycles in Seasonal Non‐Stationary Time Series.” *Journal of Applied Econometrics (Chichester, England)*, vol. 14, no. 3, 1999, pp. 273-291, https://doi.org/10.1002/(sici)1099-1255(199905/06)14:3%3C273::aid-jae498%3E3.3.co;2-e. Accessed 7 Apr 2024.
4. Nasir Hamid Rao, and Abdul. “Time Series Analysis (Stationarity,Cointegration, and Causality).” *Environmental Kuznets Curve (EKC)*, 2019, pp. 85-99.
5. Rao, M.M. “Covariance Analysis of Nonstationary Time Series.” *Developments in Statistics*, 1978, pp. 171-225, https://doi.org/10.1016/b978-0-12-426601-8.50010-3. Accessed 7 Apr 2024.
6. Rehal, Viren. “VECM Estimation and Interpretation.” *SPUR ECONOMICS*, https://spureconomics.com/vecm-estimation-and-interpretation/. Accessed 8 April 2024.
7. *Vector Autoregressive Models for Multivariate Time Series.* Modeling Financial Time Series with S-PLUS®, pp. 385-429.
8. Yong, and Liu. “Non-Stationary Transformers: Exploring the Stationarity in Time Series Forecasting.” 2023, arxiv.org/abs/2205.14415.