ANALYSIS OF ETF WITH UNDERLYING ASSET USING TRACKING ERROR

Sudhir PR[*]
Debasis Patnaik[**]

ABSTRACT

Exchange Traded Funds have grown tremendously in recent times, although they have been in India for over sixteen years now, with the first ETF, Nifty Bees listed in 2002. As the ETFs track benchmark Index, their price movements are expected to track these indices or any other assets they represent. Using the latest data from NSE, the relationship between prices of ETFs in India and those of their underlying assets of the previous two years using the econometric techniques of Vector Auto regression are studied. The paper also focuses on understanding tracking error across the cross-section of ETF's present in the Indian Markets. The key findings of the study are a lag of 2 days between the price change in index and ETFs. Further, the tracking error is significantly decreased when the number of days for rolling standard deviation are increased.

***Keywords:** Exchange Traded Funds, Vector Auto Regression (VAR), Vector Error Correction (VEC), Tracking Error, Stationarity, Correlation, Time Series Analysis.*

_____

Introduction

An ETF is a fund, which is traded on the stock exchanges. ETFs are innovative in nature and provide exposure to an index or a basket of securities and can be traded on the exchange.The ETFs scores over traditional open-ended index funds in that they can be traded on the exchange during trading hours at NAV of the Scheme. The listed close-ended funds, trade at premiums or at discounts to NAV. The ETFs are structured in a way which facilitate " creation and redemption of units directly with the fund by authorized participants i.e., where a pre-defined number of shares for units are exchanged allowing them to take advantage of any mispricing. Indices which are commonly tracked by most of the ETFs are S&P 500 or MSCI EAFE or NSE Nifty or BSE Sensex. Due to their Low costs, tax efficiency, and stock-like features makes ETFs an attractive investment. Exchange traded fund is a very popular exchange-traded product. An ETF has the characteristics of a mutual fund and a unit investment trust, and similar to close ended funds which is traded on the exchange which is traded throughout the day at prices that may be equal, greater or less than its net asset value. Closed-end funds do not fall into the category of or are not perceived as "ETFs", even they are mutual funds and have a feature of tradability on an stock exchange. ETFs have been available in the US since 1993 and in Europe since 1999. There are various types of ETFs that are traded on the stock exchanges, like equity index ETFs, liquid ETFs, and the gold ETFs. The paper is divided into four parts. Firstly, the literature review, followed by Research methodology, Empirical results and finally conclusion.

Literature Review

The authors are of the opinion that reasons behind the Growth of ETFs is the advantages of ETF off liquidity, low fees, transparency, and potential tax advantages. Secondly, the universe of ETFs has been expanding beyond the traditional equity-based funds, including funds providing access to fixed income, commodities, currency, volatility, multi-asset class structures, and "smart beta".[1]

Harip R khanapuri, opines that Equity based ETFs exhibit stronger correlation whereas and nearly no or weak correlation is observed in commodity based ETF markets. The investors will have to understand the implications of these findings while investing in an ETF market. The author states that underlying asset contributes to price volatility ETF market. There is a marked difference between equity-based ETFs and Gold based ETFs [2]

_____

[*]   PhD Student, Department of Economics & Finance, BITS Pilani, Goa Campus, Goa, India.
[**]   PhD, Associate Professor, Department of Economics& Finance, BITS Pilani, Goa Campus, Goa, India.

Stock et al, employed vector auto-regressions (VARs) to analyze: data description, forecasting, structural inference, and policy analysis. They stated that VARs is powerful tool for data description and forecasting [3].

Maysami et al, studied long-term equilibrium conditions between the Singapore stock index and various macroeconomic variables by vector error-correction models [4].

Avellaneda et al, found that highly leveraged exchange-traded funds (LETFs) are incapable of producing multiple of index returns over a longer period of time. The return of a leveraged fund was related to the corresponding multiple of the return of the unleveraged fund and its realized variance by the use of a model.By employing dynamic rebalancing strategy, the leveraged funds can be made to replicate the returns of the benchmark index or indices [5].

In this paper the author studied the market efficiency of sectoral ETFs indices in the short- and long-run horizons, small and large fluctuations, and pre and post global financial crisis (GFC). They opined that there is a adequate data to support that the sector ETF markets are multifractal in nature and exhibited a crossover in the multifractality of sector ETF market dynamics. They also observed that FMCG sector ETF markets have outperformed with respect to financial services and telecommunications sector. The author concluded that discrepancies which with respect to market efficiency, among the short- and long-term horizons and the global financial crisis adversely affected ETF market efficiency [6].

Zhou et al, cross-correlation relationship between Chinese treasury futures contracts and treasury ETF was studied by the author. They came out with the conclusion there is lack of normality between the series, and statistically significant correlation exists in the futures markets have greater price volatility. The said relationship was verified by using DCCA coefficient and the DMCA coefficient [7].

In this paper, the authors examined the pricing deviations of the three most liquid Exchange Traded Funds and they observed that price deviation was predictable, nonzero. They concluded that deviations in prices could be valued as an additional cost of administering an Exchange Traded Fund and stationarity induces price deviations. The Nonzero price deviation stems from the specific price discovery processes and dividend accumulation and distribution for this asset class.[8]

The authors examined model for multivariate time series data analysis and outlined the basic design concepts. The authors cited the application of model to a database of ambient air pollution measurements in the United States and to a hypothetical portfolio of stocks [9]

The authors further extended the work of DeFusco et al, which included pricing deviation of DIA, SPY and QQQQ on intraday basis. The study concluded that prices of DIA and QQQQ are on an average less than the underlying indexes. SPY was higher than the theoretical price of the S&P 500 index [10].

The author, by using VAR approach examined the U.S. listed currency hedged ETF with corresponding changes in the benchmark and foreign exchange fluctuations between July 2011 to November 2015. The author concluded that investors anticipated changes in future exchange rates and migrated to currency hedged ETFs. [11].

A significant gap exists in studying cross section of ETFs tracking the same asset offered by different financial firms. Further, while a significant number of researches, analyses ETFs and related assets, there is scope for conducting econometric analysis of past ETF data and understand the relationships in greater detail. Finally, the temporal aspect of relationship between ETF and underlying assets, especially using rolling tracking error and vector auto regression, have significant scope for analysis. We examined these areas in this research paper

Rsearch Methodology

- **Data extraction and Pre-processing**

This paper studies 6 ETFs that follow NIFTY50 index namely: Edelweiss, Kotak Mahindra, Motilal Oswal, SBI, Birla Capital and Reliance Capital. The daily open, low, high and close data for all the ETFs is extracted in Python from the application program interface of the data provider. Price data of Nifty50 index is extracted from NSE website. The extracted data preprocessed to obtain uniformity of data of all ETFs in a specified time period. The time period is chosen as the time period of the ETF for which minimum data is available, in the above case: Edelweiss ETF.

- **Analysis of Tracking Error**

The literature review provided tracking error [3] as an important parameter for analysis of ETF and the corresponding underlying. The paper utilizes tracking error to measure the dependence of ETF

returns on NIFTY50 index returns. Log returns are calculated for all the aforementioned ETFs and NIFTY50 index on Closing Price of the particular day. Error term is subsequently calculated for each ETF and above calculated time period by subtracting NIFTY50 index returns from that particular ETF's returns for the particular day. The paper further analysis Tracking error of each ETF by the use of rolling standard deviation of the Error term with rolling windows of 10, 20, 30 and 40 days.

- **Analysis through Econometric Techniques**

This paper analysis the interdependence between ETF and the underlying asset using Vector Auto-Regression (VAR) and Vector Error Correction (VEC). In order to conduct the analysis, we use MATLAB's econometric toolbox and its VAR and VECM model.The hypothesis for the analysis is to find whether a causality exists between daily returns of Nifty 50 index and daily returns of the ETFs that track this index. For the purpose of the analysis, default VAR (4) model was used with Nifty and only one ETF as estimation parameters.
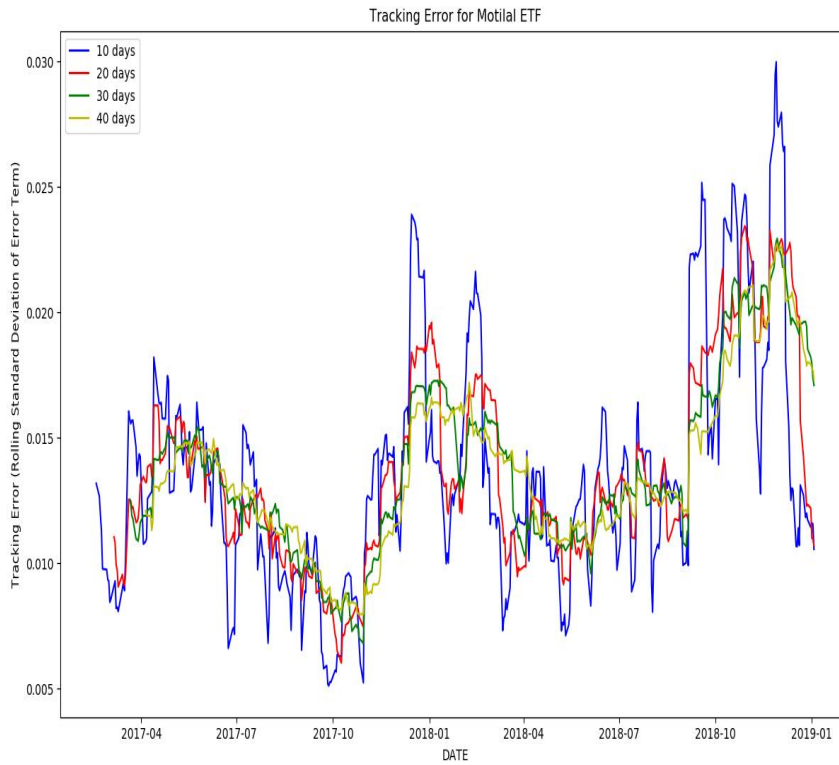
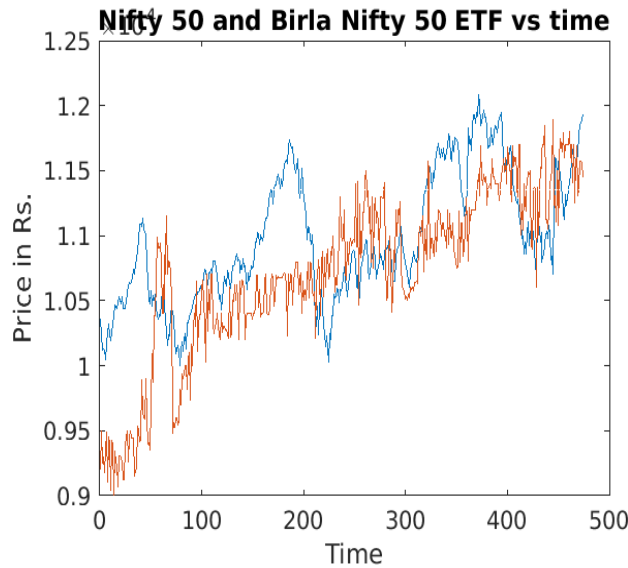**Empirical Observation**
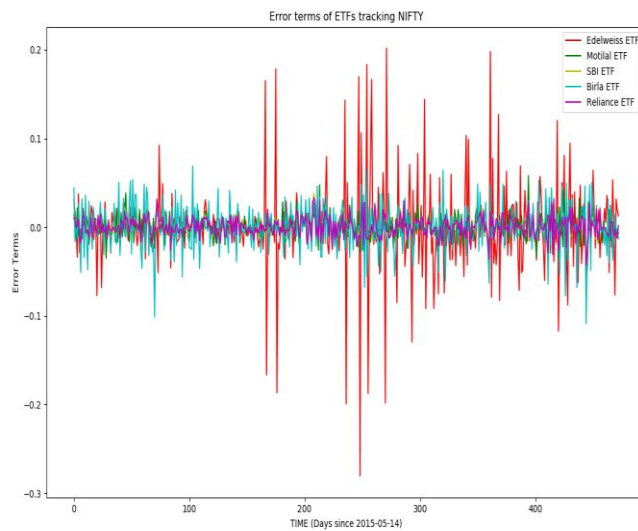
(A)   Error Terms (Difference in the returns of ETFs and NIFTY50) of ETFs over time.
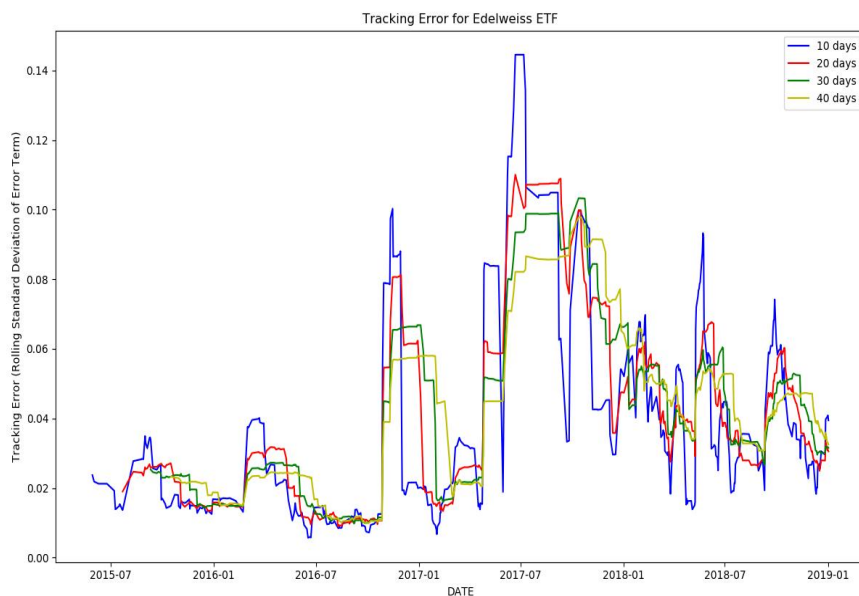
(B)   Tracking Error of Edelweiss ETF over time

(C)   Tracking Error of Kotak ETF over time

(D)   Nifty 50 and Birla Nifty 50 ETF price levels over time

Empirical Results

Tracking Error Analysis

The graph of error terms clearly shows that the difference in the returns of ETF and NIFTY Index are not uniform across ETFs, but vary considerably across different ETFs. The paper finds that the Edelweiss ETF has the highest deviation. Further, we take Edelweiss and Motilal ETFs for analysis of tracking error while taking rolling standard deviation across various days.

From the graphs B and C, it can be seen that when the lag is minimum, the tracking error is maximum. Subsequently, the finding can be reinforced with other ETFs. Also, across time, the peaks become sharper, which mean that the tracking error starts decreasing considerably with time. This effect can be attributed to the rebalancing of the ETF following the rebalancing of indexes.

- **Results for Econometric Analysis**

The analysis VAR model shows p-values of 0.35 for lag 1, 0.03 for lag 2, 0.09 for lag 3 and 0.20 for lag 4. This suggests that the null hypothesis of the existence of causality is strongly true for lag 2 while it is not true for all other lags. The above results indicate that the causality of index returns, and ETF returns are caused by lag of 2 suggesting that the returns from the index take 2 days for realization the ETF.

It also suggests that no causality is present between ETFs and index returns when the lag exceeds 2. Similar results are observed when VEC model is used instead of VAR model and the results are consistent among all ETFs that track Nifty50.

Conclusion

Our paper identifies the lag between ETFs and underlying assets through econometric method of vector autoregression. Further, this paper successfully proves high degree of causality between the same. Also, the analysis of rolling tracking error leads to fundamental insights about how the ETFs and underlying asset often diverge significantly in their returns often during times of rebalancing. Lastly, we explore the temporal aspect of rolling standard deviation and conclude the lag in the number of days and tracking error have an inverse relationship.

References

1. Lettau, Martin, and Ananth Madhavan. "Exchange-Traded Funds 101 for Economists." *Journal of Economic Perspectives*, vol. 32, no. 1, 2018, pp. 135–154., doi:10.1257/jep.32.1.135

2. Harip R. Khanapuri. "Examining the Relationship between ETFS and Their Underlying Assets in Indian Capital Market".2012 2nd International Conference on Computer and Software Modeling (IC CSM 2012)

3. Stock, James H, and Mark W Watson. "Vector Autoregressions." *Journal of Economic Perspectives*, vol. 15, no. 4, 2001, pp. 101–115., doi:10.1257/jep.15.4.101.

4. Maysami, Ramin Cooper, and Tiong Sim Koh. "A Vector Error Correction Model of the Singapore Stock Market." *International Review of Economics & Finance*, vol. 9, no. 1, 2000, pp. 79–96., doi:10.1016/s1059-0560(99)00042-8.

5. Avellaneda, Marco, and Stanley Zhang. "Path-Dependence of Leveraged ETF Returns." *SIAM Journal on Financial Mathematics*, vol. 1, no. 1, 2010, pp. 586–603., doi:10.1137/090760805.

6. Tiwari, Aviral Kumar, et al. "A Multifractal Detrended Fluctuation Analysis of Financial Market Efficiency: Comparison Using Dow Jones Sector ETF Indices." *Physica A: Statistical Mechanics and Its Applications*, vol. 483, 2017, pp. 182–192., doi:10.1016/j.physa.2017.05.007.

7. Zhou, Yu, and Shi Chen. "Cross-Correlation Analysis between Chinese TF Contracts and Treasury ETF Based on High-Frequency Data." *Physica A: Statistical Mechanics and Its Applications*, vol. 443, 2016, pp. 117–127., doi:10.1016/j.physa.2015.09.078.

8. Defusco, Richard A., et al. "The Exchange Traded Funds' Pricing Deviation: Analysis and Forecasts." *Journal of Economics and Finance*, vol. 35, no. 2, 2009, pp. 181–197., doi:10.1007/s12197-009-9090-6.

9. Peng, Roger D. "A Method for Visualizing Multivariate Time Series Data." *Journal of Statistical Software*, vol. 25, no. Code Snippet 1, 2008, doi:10.18637/jss.v025.c01.

10. Ivanov, Stoyu I. "High- Frequency Analysis of Exchange Traded Funds' Pricing Deviation." *Managerial Finance*, vol. 39, no. 5, 2013, pp. 509–524., doi:10.1108/03074351311313834.

11. Shank, Corey A. "Are US-Dollar-Hedged-ETF Investors Aggressive on Exchange Rates? A Panel VAR Approach." *SSRN Electronic Journal*, 2016, doi:10.2139/ssrn.2782082.

Appendix A: Data Extraction Code

```
import pandas as pd#LIBRARY TO HANDLE DATA

import quandl#FOR GETTING DATA

quandl.ApiConfig.api_key = "input_api_key_here"
```

```
import matplotlib.pyplot as plt
import numpy as np


#GETTING DATA
edelweiss_data=quandl.get("NSE/NIFTYEES")
kotak_data=quandl.get("NSE/KOTAKNIFTY")
motilal_data=quandl.get("NSE/M50")
sbi_data=quandl.get("NSE/SETFNIF50")
birla_data=quandl.get("NSE/BSLNIFTY")
reliance_data=quandl.get("NSE/NIFTYBEES")
nifty=pd.read_csv("nifty_full.csv",header=None)
```

Appendix B: Tracking Error Calculation Code

```
kotak_data["log_returns"]=np.log(kotak_data["Close"].shift(-1))-np.log(kotak_data["Close"])
edelweiss_data["log_returns"]=np.log(edelweiss_data["Close"].shift(-1))-np.log(edelweiss_data["Close"])
motilal_data["log_returns"]=np.log(motilal_data["Close"].shift(-1))-np.log(motilal_data["Close"])
sbi_data["log_returns"]=np.log(sbi_data["Close"].shift(-1))-np.log(sbi_data["Close"])
birla_data["log_returns"]=np.log(birla_data["Close"].shift(-1))-np.log(birla_data["Close"])
reliance_data["log_returns"]=np.log(reliance_data["Close"].shift(-1))-np.log(reliance_data["Close"])
nifty["log_returns"]=np.log(nifty[1].shift(-1))-np.log(nifty[1])


kotak_data["tracking_error"]=kotak_data.loc[:,"log_returns"]-nifty.loc[:,"log_returns"].values
edelweiss_data["tracking_error"]=edelweiss_data["log_returns"]-nifty["log_returns"].values
motilal_data["tracking_error"]=motilal_data["log_returns"]-nifty["log_returns"].values
sbi_data["tracking_error"]=sbi_data["log_returns"]-nifty["log_returns"].values
birla_data["tracking_error"]=birla_data["log_returns"]-nifty["log_returns"].values
reliance_data["tracking_error"]=reliance_data["log_returns"]-nifty["log_returns"].values


kotak_data["rolling_std_10"]=kotak_data["tracking_error"].rolling(10).std()
edelweiss_data["rolling_std_10"]=edelweiss_data["tracking_error"].rolling(10).std()
motilal_data["rolling_std_10"]=motilal_data["tracking_error"].rolling(10).std()
sbi_data["rolling_std_10"]=sbi_data["tracking_error"].rolling(10).std()
birla_data["rolling_std_10"]=birla_data["tracking_error"].rolling(10).std()
reliance_data["rolling_std_10"]=reliance_data["tracking_error"].rolling(10).std()


kotak_data["rolling_std_20"]=kotak_data["tracking_error"].rolling(20).std()
edelweiss_data["rolling_std_20"]=edelweiss_data["tracking_error"].rolling(20).std()
motilal_data["rolling_std_20"]=motilal_data["tracking_error"].rolling(20).std()
sbi_data["rolling_std_20"]=sbi_data["tracking_error"].rolling(20).std()
birla_data["rolling_std_20"]=birla_data["tracking_error"].rolling(20).std()
reliance_data["rolling_std_20"]=reliance_data["tracking_error"].rolling(20).std()


kotak_data["rolling_std_30"]=kotak_data["tracking_error"].rolling(30).std()
edelweiss_data["rolling_std_30"]=edelweiss_data["tracking_error"].rolling(30).std()
motilal_data["rolling_std_30"]=motilal_data["tracking_error"].rolling(30).std()
```

```
sbi_data["rolling_std_30"]=sbi_data["tracking_error"].rolling(30).std()
birla_data["rolling_std_30"]=birla_data["tracking_error"].rolling(30).std()
reliance_data["rolling_std_30"]=reliance_data["tracking_error"].rolling(30).std()

kotak_data["rolling_std_40"]=kotak_data["tracking_error"].rolling(40).std()
edelweiss_data["rolling_std_40"]=edelweiss_data["tracking_error"].rolling(40).std()
motilal_data["rolling_std_40"]=motilal_data["tracking_error"].rolling(40).std()
sbi_data["rolling_std_40"]=sbi_data["tracking_error"].rolling(40).std()
birla_data["rolling_std_40"]=birla_data["tracking_error"].rolling(40).std()
reliance_data["rolling_std_40"]=reliance_data["tracking_error"].rolling(40).std()
```

Appendix C: Code for generating Plots

```
plt.plot(list(edelweiss_data["tracking_error"]),'r',label='Edelweiss ETF')
plt.plot(list(motilal_data["tracking_error"]),'g',label='Motilal ETF')
plt.plot(list(sbi_data["tracking_error"]),'y',label='SBI ETF')
plt.plot(list(birla_data["tracking_error"]),'c',label='Birla ETF')
plt.plot(list(reliance_data["tracking_error"]),'m',label='Reliance ETF')
plt.title('Error terms of ETFs tracking NIFTY')
plt.legend()
plt.xlabel('TIME (Days since 2015-05-14)')
plt.ylabel('Error Terms')
plt.show()
plt.plot(edelweiss_data["rolling_std_10"],'b',label='10 days')
plt.plot(edelweiss_data["rolling_std_20"],'r',label='20 days')
plt.plot(edelweiss_data["rolling_std_30"],'g',label='30 days')
plt.plot(edelweiss_data["rolling_std_40"],'y',label='40 days')
plt.title('Tracking Error for Edelweiss ETF')
plt.legend()
plt.xlabel('DATE')
plt.ylabel('Tracking Error (Rolling Standard Deviation of Error Term)')
plt.show()
plt.plot(motilal_data["rolling_std_10"],'b',label='10 days')
plt.plot(motilal_data["rolling_std_20"],'r',label='20 days')
plt.plot(motilal_data["rolling_std_30"],'g',label='30 days')
plt.plot(motilal_data["rolling_std_40"],'y',label='40 days')
plt.title('Tracking Error for Motilal ETF')
plt.legend()
plt.xlabel('DATE')
plt.ylabel('Tracking Error (Rolling Standard Deviation of the Error Term)')
plt.show()
```

Appendix D: VAR Model Code
Matlab Code for developing the VAR model and estimating the results.

```
clear all;
clc;
```

```
load Data_USEconModel
T = readtable('SAPM_.xlsx'); //loads the index and ETF data
t=1:1:474;
nifty=table2array(T(:,20)); //queries for nifty data
birla = table2array(T(:,2)); //queries for Birla ETF data
figure;
plot(t,nifty);
title("Nifty50 vs time"); //plots index data
hold on;
plot(t,birla.*100); //plots Birla ETF data
title("Nifty 50 and Birla Nifty 50 ETF vs time");
niftyrate = price2ret(nifty); // finds the daily returns of index
birlarate = price2ret(birla); //finds the daily returns of ETF
Mdl = varm(2,4); //VAR model
EstMdl = estimate(Mdl,[niftyratebirlarate]); //fitting the model
summarize(EstMdl) // Getting the p-values.
```

❖◆❖