Modeling and Forecasting the NASDAQ Composite Index: An ARMA and VAR Approach with EGARCH Analysis on NVDA and MSFT

Bo Wu

Beijing International Studies University

Abstract

This study examines the performance and predictability of the NASDAQ Composite Index (COMP) from 2019 to 2024. First, we calculate the growth rate and volatility of the index based on a rolling 30-day window. We apply the ARMA model to conduct both in-sample and out-of-sample forecasting for the index's growth rate and volatility, covering the forecast period of 2024 to 2025. We assess the accuracy of these forecasts and include time series plots of the index, growth rate, and volatility, as well as descriptive statistics for returns and volatility.

Stationarity tests confirm that the index series is non-stationary, while the growth rate is stationary at the 1% significance level, suggesting it follows an $I(0)$ process. Model order selection based on the Bayesian Information Criterion (BIC) indicates that the growth rate follows an ARMA(4,4) model. For volatility, the auto.arima function suggests an ARIMA(2,1,2) model.

Furthermore, we select two representative stocks listed on the NASDAQ—NVIDIA (NVDA) and Microsoft (MSFT)—to investigate leverage effects using the EGARCH model. We find significant evidence of asymmetric volatility in their return series.

Using daily return data for NVDA and MSFT from 2019 to 2024, we construct a bivariate Vector Autoregression (VAR) model to analyze the dynamic relationship between the two stocks. We perform Granger causality tests and analyze the impulse response functions to understand the interdependencies and shock transmissions within the system.

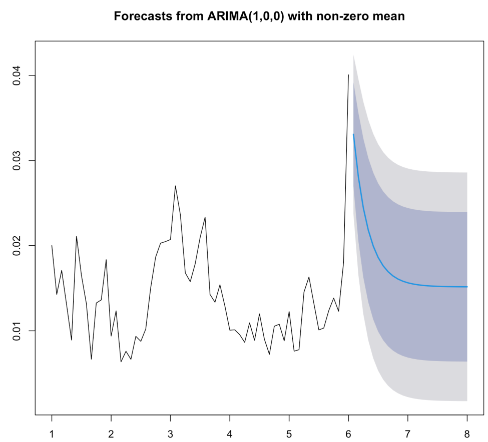

Figure 1: **NASDAQ Composite Index**

1 Preliminary Data Analysis

- Time series plots of NASDAQ Composite Index(Ni_t), growth rate (Gr_t), and volatility (Vol_t) are presented to visually examine their behavior over time.
- Descriptive statistics are reported for the growth rate (Gr_t) and volatility(Vol_t) series, including measures such as mean, standard deviation, skewness, and kurtosis.
- The NASDAQ Composite Index series(Ni_t) exhibits clear non-stationarity, as observed from its trending behavior in the **Figure 1** and confirmed by statistical tests. The closing price series follows the **Formula 1**:

$$\mathrm{Ni}_t = \phi_1 \mathrm{Ni}_{t-1} + \varepsilon_t \tag{1}$$

- For the return series Gr_t, the Augmented Dickey-Fuller (ADF) test rejects the null hypothesis of a unit root at the 1% significance level. Therefore, we conclude that the return series is stationary, i.e., **Formula 2**

$$\mathrm{Gr}_t \sim I(0) \tag{2}$$

  Since the return series (Gr_t) is stationary, it can be modeled by a covariance-stationary autoregressive process such as AR(1) or AR(p)[SSSS17], where the autoregressive coefficients satisfy the stationarity condition.**Code A**

2 ARMA Model Lag Order Determination

2.1 ARIMA Model for Growth Rate:

The lag order for the return series is determined using the Bayesian Information Criterion (BIC), which selects an ARMA(4,4) model (see **Formula 3**), as shown in **Figure 2**.[HX98]

$$\mathrm{Gr}_t = \phi_1 \mathrm{Gr}_{t-1} + \phi_2 \mathrm{Gr}_{t-2} + \phi_3 \mathrm{Gr}_{t-3} + \phi_4 \mathrm{Gr}_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} \tag{3}$$

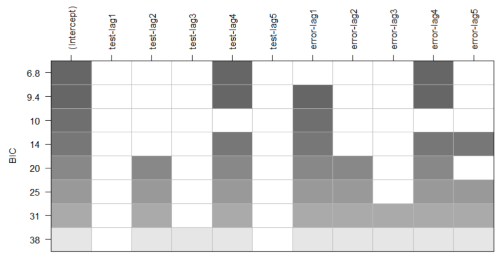

Figure 2: **Growth Rate**

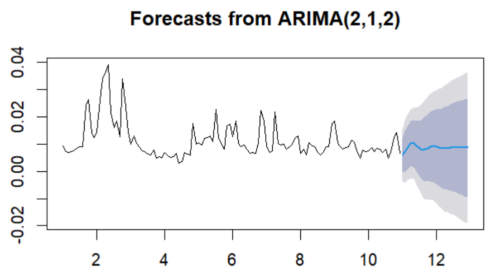

Figure 3: **Volatility**

2.2 ARIMA Model for Volatility:

The volatility series is analyzed using the `auto.arima` function, as shown in the **Figure 3**, which selects an ARIMA(2,1,2) **Formula 4**:

$$(1 - \phi_1 L - \phi_2 L^2)(1 - L)\,\mathrm{Vol}_t = (1 + \theta_1 L + \theta_2 L^2)\varepsilon_t \tag{4}$$

3 EGARCH(1,1) Estimation of Stock Returns

We selected two representative stocks on the Nasdaq exchange: NVDA and MSFT, and carried out EGARCH estimation.

3.1 EGARCH(1,1) Estimation of NVDA Stock Returns

The following equation represents the estimated EGARCH(1,1) model for NVDA daily stock returns from June 24, 2019, to June 21, 2024:

3.1.1 Mean Equation Formula 5

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = z_t \sigma_t, \quad z_t \sim \mathcal{N}(0, 1) \tag{5}$$

The estimated constant mean return is **Formula 6**:

$$\mu = 0.0038046 \tag{6}$$

Here, r_t is the daily return of NVDA stock at time t, and ε_t is the innovation term driven by volatility σ_t and standardized shock z_t.

3.1.2 Variance Equation (EGARCH) Formula 7

$$\log(\sigma_t^2) = 20.1265 + 30.8534 \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + 3.7004 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} - 1590.493 \log(\sigma_{t-1}^2) \tag{7}$$

The term $\sqrt{\frac{2}{\pi}}$ is the expected absolute value of a standard normal distribution, i.e., $\mathbb{E}|z_t| = \sqrt{\frac{2}{\pi}} \approx 0.7979$.

3.1.3 Brief Description:

This EGARCH(1,1) model captures the volatility dynamics of NVDA stock returns. The model allows for asymmetric responses to past shocks:

- The leverage effect term ($\gamma_1 \approx 3.70$) is marginally significant ($p \approx 0.058$), suggesting that positive and negative shocks may affect volatility differently.
- The ARCH term ($\alpha_1 \approx 30.85$) is statistically significant ($p = 0.010$), indicating that past absolute shocks have a strong impact on current volatility.
- The GARCH term ($\beta_1 \approx -1590.49$) is not statistically significant, and its large negative value may reflect instability or overfitting in the variance equation.[BJ06]

The model emphasizes the importance of recent shocks in explaining NVDA's return volatility and suggests potential asymmetry in how shocks influence volatility.

3.2 EGARCH(1,1) Estimation of MSFT Stock Returns

The EGARCH(1,1) model for Microsoft (MSFT) daily stock returns from June 24, 2019 to June 21, 2024 is composed of a mean equation and a variance equation.

3.2.1 Mean Equation Formula 8

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = z_t \sigma_t, \quad z_t \sim \mathcal{N}(0, 1) \tag{8}$$

Here, r_t is the return, $\mu = 0.0011454$ is the estimated constant return, and ε_t is the innovation term with conditional variance σ_t^2.

3.2.2 Variance Equation (EGARCH) Formula 9

$$\log(\sigma_t^2) = 8.7759 + 127.4672 \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + 2.0895 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} - 575.0156 \log(\sigma_{t-1}^2) \tag{9}$$

The term $\sqrt{\frac{2}{\pi}}$ is the expected absolute value of a standard normal distribution, i.e., $\mathbb{E}|z_t| = \sqrt{\frac{2}{\pi}} \approx 0.7979$.

3.2.3 Brief Description

This EGARCH(1,1) model captures the asymmetric and dynamic nature of MSFT's return volatility. Key findings include:

- The leverage effect term ($\gamma_1 \approx 2.0895$) is statistically significant (p = 0.000), suggesting strong asymmetry in how positive vs. negative shocks affect volatility.
- The ARCH effect ($\alpha_1 \approx 127.47$) is also significant (p = 0.000), indicating a high impact of past absolute shocks on current volatility.
- The GARCH term ($\beta_1 \approx -575.02$) is not statistically significant (p = 0.232), and its negative sign may suggest potential model instability.

The model highlights the high responsiveness of MSFT's volatility to past shocks, especially in terms of shock magnitude and direction.

For the two stocks (NVDA and MSFT), volatility increases more in response to positive news than to negative news. This may reflect the market's high growth expectations for these companies, where good news triggers stronger market reactions and price fluctuations.

The leverage effect term of NVDA ($\gamma_1 \approx 3.70$) is greater than that of MSFT ($\gamma_1 \approx 2.0895$), indicating that NVIDIA's stock price is more sensitive to positive news. Therefore, investors should pay closer attention to positive news in the market when dealing with NVDA, as it may lead to larger price swings.**Code** B

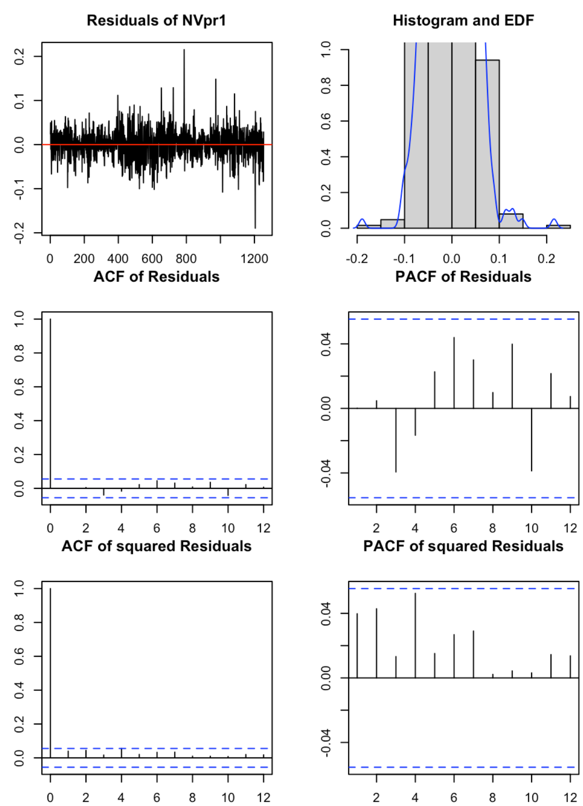

Figure 4: **Correlation test of residuals of NVDA's VAR**

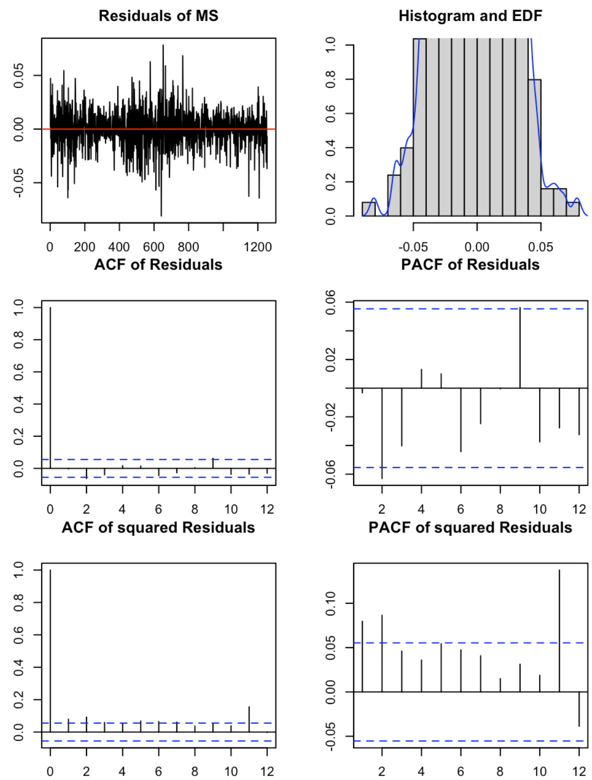

Figure 5: **Correlation test of residuals of MSFT's VAR**

4 VAR(1) Model for NVDA and Microsoft Daily Returns

Let NVpr1_t denote the daily stock return of NVIDIA (NVDA) at time t, and MS_t denote the daily stock return of Microsoft (MSFT) at time t. The estimated VAR(1) **Formula 10** is:

$$\begin{bmatrix} \text{NV}_t \\ \text{MS}_t \end{bmatrix} = \begin{bmatrix} -0.0407 & -0.0210 \\ 0.0214 & -0.0721 \end{bmatrix} \begin{bmatrix} \text{NV}_{t-1} \\ \text{MS}_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{10}$$

where ε_{1t} and ε_{2t} are white noise error terms.

4.1 Interpretation

- The return of NVIDIA today (NV_t) is **negatively affected** by its own past return (-0.0407) and also by Microsoft's past return (-0.0210).
- The return of Microsoft today (MS_t) is **positively influenced** by NVIDIA's past return ($+0.0214$), but **negatively influenced** by its own past return (-0.0721).
- The innovation terms ε_{1t} and ε_{2t} represent shocks that are not explained by the model, assumed to be white noise.

4.2 Correlation Test of Residuals of NVDA's VAR

The residual diagnostics of NVDA's VAR model indicate that the residuals are approximately white noise. The ACF and PACF plots of residuals do not show significant autocorrelations, suggesting that serial correlation has been effectively removed. The histogram and empirical distribution function (EDF) demonstrate approximate normality.

However, the ACF and PACF of squared residuals exhibit some signs of autocorrelation, especially at lag 1, which may suggest potential ARCH effects. Although most spikes lie within the confidence bounds, a few exceed them, warranting further investigation, such as an ARCH-LM test. **Figure 4**

4.3 Correlation Test of Residuals of MSFT's VAR

The residual diagnostics for the MSFT VAR model show that the residuals behave like white noise. Both the ACF and PACF plots of the residuals lack significant spikes, indicating no remaining serial correlation. The histogram and EDF curve also support the assumption of residual normality. **Figure 5**

The squared residuals' ACF and PACF plots show a high value at lag 1, potentially indicating heteroskedasticity. Some minor spikes at higher lags are observed but remain mostly within the 95% confidence bands.[Pfa08]

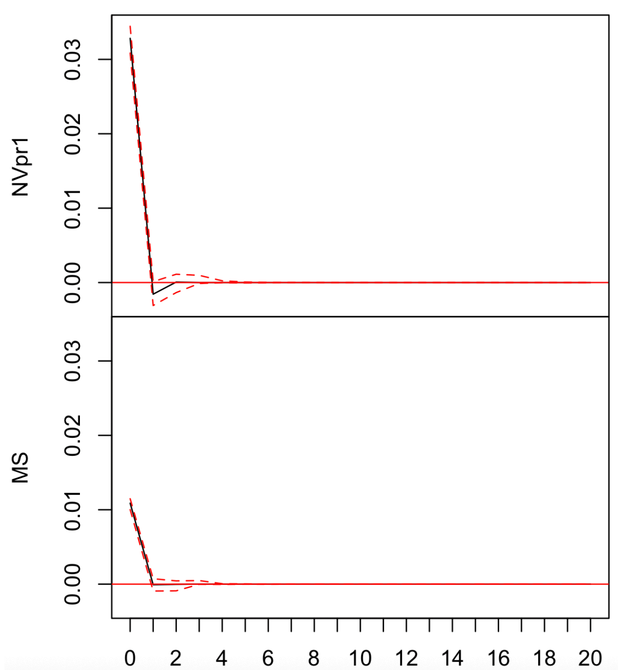

Figure 6: **Orthogonal Impulse Response from NVDA**

5 Orthogonal Impulse Response Analysis

Figure 6 displays the OIRFs to a shock in NVDA (NVpr1). The top panel illustrates the response of NVDA itself, which is strong and positive on impact and dissipates rapidly within a few periods.

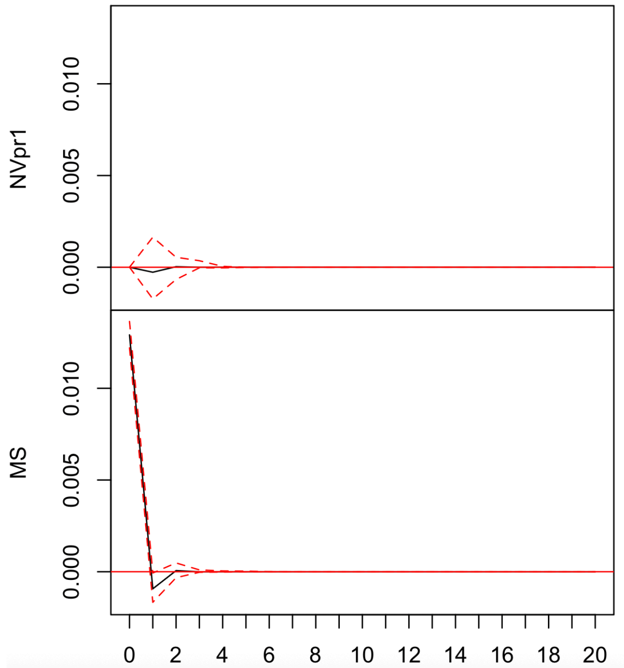

Figure 7: **Orthogonal Impulse Response from MSFT**

The bottom panel shows MSFT's response to a shock in NVDA, which is small and statistically insignificant, suggesting weak spillover effects from NVDA to MSFT.

Figure 7 shows the orthogonal impulse response functions (OIRFs) from a one-standard-deviation shock to MSFT. The top panel shows the response of NVDA (NVpr1), and the bottom panel shows the response of MSFT (MS). The response of MS to its own shock is significantly positive in the first period and quickly dies out, indicating a short-lived effect. The impact on NVDA is negligible and not statistically significant as the confidence bands contain zero throughout the horizon.[Qin11]

The results suggest that both NVDA and MSFT primarily respond to their own innovations, with limited cross-impact effects.**Code C**

6 Granger Causality Test Results

We perform Granger causality tests and instantaneous causality tests between the two variables: `NVpr1` and `MS`, based on a VAR model.

- **Test 1: Does `NVDA` Granger-cause `MSFT`?**
  Null hypothesis: `NVDA` does not Granger-cause `MSFT`.
  F-statistic = 1.266, bootstrap p-value = 0.18.
  Conclusion: We fail to reject the null hypothesis at the 5% significance level. There is no significant Granger causality from `NVDA` to `MS`.

- **Test 2: Does `MSFT` Granger-cause `NVDA`?**
  Null hypothesis: `MSFT` does not Granger-cause `NVDA`.
  F-statistic = 0.08586, bootstrap p-value = 0.74.

Conclusion: We fail to reject the null hypothesis at the 5% significance level. There is no significant Granger causality from `MSFT` to `NVDA`.

- **Instantaneous Causality Test (Symmetric)**
  Null hypothesis: There is no instantaneous causality between `NVpr1` and `MS`.
  Chi-squared $= 368.03$, df $= 1$, p-value $< 2.2 \times 10^{-16}$.
  Conclusion: We strongly reject the null hypothesis. There exists significant instantaneous causality between `NVDA` and `MSFT`.

7 Conclusion

In this study, we analyzed the dynamic relationship between the daily returns of NVIDIA (NVDA) and Microsoft (MSFT) using a VAR(1) model. The estimated coefficients suggest that both firms' returns are influenced primarily by their own past values, with some degree of interdependence. Specifically, NVDA's returns are negatively affected by both its own lag and MSFT's lag, while MSFT's returns are positively influenced by NVDA's lag but negatively influenced by its own lag.

Residual diagnostics confirm that the VAR(1) model captures the serial dependence structure effectively, with residuals resembling white noise for both stocks. However, there is evidence of possible ARCH effects in the squared residuals, particularly for MSFT, which may merit further modeling with GARCH-type specifications.[HS14]

The orthogonal impulse response functions (OIRFs) indicate that shocks to each stock primarily affect themselves, with limited and statistically insignificant spillover effects between the two. This suggests that, while there may be contemporaneous co-movements, innovations in one stock do not meaningfully propagate to the other in future periods.

Granger causality tests reveal no significant predictive causality between NVDA and MSFT in either direction, further supporting the finding that their lead-lag dynamics are weak. Nevertheless, the strong result from the instantaneous causality test highlights a substantial contemporaneous interaction, implying that both stocks react jointly to common shocks or market-wide news.

Overall, our findings suggest that NVDA and MSFT returns are mostly driven by their own dynamics, with limited predictive interaction, but considerable contemporaneous correlation. These insights could be relevant for portfolio diversification and short-term trading strategies involving these two major technology stocks.

References

[BJ06] Michael W Brandt and Christopher S Jones. Volatility forecasting with range-based egarch models. *Journal of Business & Economic Statistics*, 24(4):470–486, 2006. 3

[HS14] Andrew Harvey and Genaro Sucarrat. Egarch models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, 76:320–338, 2014. 8

[HX98] Siu Lau Ho and Min Xie. The use of arima models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2):213–216, 1998. 1

[Pfa08] Bernhard Pfaff. Var, svar and svec models: Implementation within r package vars. *Journal of statistical software*, 27:1–32, 2008. 6

[Qin11] Duo Qin. Rise of var modelling approach. *Journal of Economic Surveys*, 25(1):156–174, 2011. 7

[SSSS17] Robert H Shumway, David S Stoffer, Robert H Shumway, and David S Stoffer. Arima models. *Time series analysis and its applications: with R examples*, pages 75–163, 2017. 1

Part I
Appendix

Table of Contents

A ARIMA Model Estimation

```
library(readxl)
library(forecast)

data <- read_excel("statistics - Copy.xlsx", sheet = 1)
time_series_data <- data[[2]]
Acf(time_series_data, main="Autocorrelation Function")
Pacf(time_series_data, main="Partial Autocorrelation Function")

data <- read_excel("statistics - Copy.xlsx", sheet = 2)
time_series_data <- data[[2]]
Acf(time_series_data, main="Autocorrelation Function")
Pacf(time_series_data, main="Partial Autocorrelation Function")

library(zoo)
library(quantmod)
library(forecast)
library(readxl)

data <- read_excel("statistics - Copy.xlsx", sheet = 1)
ts_data <- ts(data$Log_Return, frequency = 243)
fit <- auto.arima(ts_data)
print(fit)
summary(fit)
plot(forecast(fit))

data <- read_excel("statistics - Copy.xlsx", sheet = 2)
ts_data <- ts(data$Monthly_Vol, frequency = 12)
fit <- auto.arima(ts_data)
print(fit)
summary(fit)
plot(forecast(fit))
```

B GARCH and EGARCH Analysis

```
// Set working directory
cd "/Users/cruiser/Downloads/download/Stata/S3"

// Load CSV file
import delimited "NVDA.csv", clear

// Assume second column is adjclose and generate return
gen close_price = adjclose
gen return = (close_price - close_price[_n-1]) / close_price[_n-1]
replace return = . if _n == 1

// Export date and return to CSV
export delimited date return using "NVDA_return.csv", replace

// Descriptive statistics
summarize return, detail

// Load Excel file
import excel "nvda.xlsx", sheet("Sheet1") firstrow

```

```
// Set time series structure
tsset date

// ARCH effect test
quietly regress return
estat archlm, lags(4)

// Estimate GARCH(1,1) model
arch return, arch(1) garch(1)

// Evaluate model
predict residuals, residuals
predict h, variance
tsset date
gen std_resid = residuals / sqrt(h)
quietly regress std_resid
estat archlm, lags(4)

// Plot conditional variance
tsset date
tsline h, title("Estimated Conditional Variance (h)") ytitle("Conditional Variance
    ")

// Plot standardized residuals
tsline std_resid, title("Standardized Residuals") ytitle("Standardized Residuals")

// Estimate EGARCH(1,1) model
tsset date
arch return, arch(1) garch(1) egarch(1)
```

C VAR Model Analysis with NVIDIA and Microsoft Data

```
# Read in index return data
NVpr <- read.csv("output_returns.csv")

# Ensure the second column is numeric
NVpr[,2] <- as.numeric(NVpr[,2])

# Remove rows with NA, NaN, or Inf values
NVpr <- NVpr[!is.na(NVpr[,2]) & !is.nan(NVpr[,2]) & !is.infinite(NVpr[,2]), ]

# Create a time series
NVpr1 <- ts(NVpr[,2], start=c(2020, 04, 08), freq=245)
ts.plot(NVpr1, main="NVDA")

# Read in turnover rate data
MS_data <- read.csv("output_growth_rate.csv")

# Ensure the second column is numeric
MS_data[,2] <- as.numeric(MS_data[,2])

# Remove rows with NA, NaN, or Inf values
MS_data <- MS_data[!is.na(MS_data[,2]) & !is.nan(MS_data[,2]) & !is.infinite(MS_
    data[,2]), ]

# Create a time series
```

```r
MS <- ts(MS_data[,2], start=c(2020, 04, 08), freq=245)
ts.plot(MS, main="MSFT")

# Ensure equal length of time series
min_length <- min(length(NVpr1), length(MS))
NVpr1 <- NVpr1[1:min_length]
MS <- MS[1:min_length]

# Combine the two time series
y <- cbind(NVpr1, MS)

# Load vars package
library(vars)

# Check for missing values
sum(is.na(y))

# Remove any remaining missing values
y <- na.omit(y)

# Estimate VAR model
z <- VAR(y, type = "const", lag.max = 12, ic = "AIC")

# Display coefficient matrices
Acoef(z)

# Forecast 245 steps ahead with 95% confidence interval
zp = predict(z, n.ahead = 245, ci = 0.95)
par(mar = c(2, 2, 2, 2))
plot(zp)

# Extract residuals
res <- residuals(z)

# Impulse Response Function analysis
ret.MS.irf <- irf(z, n.ahead = 20, ortho = TRUE, cumulative = FALSE, boot = TRUE
    , ci = 0.95)
par(mar = c(4, 4, 2, 2))
quartz()
plot(ret.MS.irf)
```