Harvard Extension Data Science

Dynamic Modeling and Forecasting in Big Data

Instructor: William Yu

Assignment 6

Predicting Macroeconomic Variables and VAR

Part A. Running Three Models: VAR Level, VAR Growth and ARIMA

- In H06a_VAR.R, we learned how to use the VAR model to forecast for the quarterly macroeconomic data. Now, let's use VAR model to model and forecast a different set of macroeconomic variables. Following these steps:
- Import the following data via quantmod library.

```
getSymbols("DCOILWTICO", src="FRED")    # Oil price
getSymbols("NASDAQCOM", src="FRED")     # NASDAQ stock price
getSymbols("PAYEMS", src="FRED")        # Nonfarm payroll jobs
getSymbols("CPILFESL", src="FRED")      # Core Consumer Price Index
getSymbols("DFF", src="FRED")           # Federal Fund Rate
```

- Set the full sample period from 1986-02-01 to 2025-1-31 and convert the whole series to monthly frequency. Note: use na.rm=T if there are missing values.
- Plot the series and convert them to the time series object using "ts" function.
- Create a separate time series data frame (data1) that turns the first four variables into a more stationary series, e.g. continuously compound growth rates. Note: (diff(log(series)). Let's keep the federal fund rate (level) as it is.

```
data0 = cbind(oil, stock, job, cpi, frr)
data1 = cbind(oil_g, stock_g, job_g, cpi_g, frr)
```

- Create the train sets for the above two data frames using

```
data00 = window(data0, start=c(1986, 3), end=c(2024, 11))
```

- Run two VAR models of these 5 variables on the trainsets using the lag of 12 months. (Note: if you find a better lag length, free feel to choose so). For the level model (data0), use type="both". Both means both of constant and trend. For the growth model (data1), use type ="const".
- Plot the forecasts of these two models for the next 12 months (December 2024 to November 2025) for all five variables. Compare the results and briefly explain the differences.
- Use auto.arima to forecast the five series in the trainset individually and plot their forecasts and compare its results to the VAR models results.
- Plot the impulse response functions for the following for both VAR models. And briefly explain them.

```
imp01 = irf(var1, impulse="oil_g", response="cpi_g", n.ahead=36, ortho=F, runs=1000)
imp02 = irf(var1, impulse="cpi_g", response="frr", n.ahead=36, ortho=F, runs=1000)
imp03 = irf(var1, impulse="frr", response="job_g", n.ahead=36, ortho=F, runs=1000)
imp04 = irf(var1, impulse="frr", response="stock_g", n.ahead=36, ortho=F, runs=1000)
```

Part B. Calculating Test Errors

Part B is optional for students pursuing undergraduate credit.

- Calculate the test errors for these three models (2 VAR models and 1 ARIMA model) for the out-of-sample part (December 2024 and January 2025). Which model is the best?
- Play around with the setting of lag term of VAR models and see if you can improve the result of testset errors.
  `var1 = VAR(data10, p = ?, type="const")`
- Hints:
  - To compare apple to apple, the test errors should be all based on the level forecast and actuals. That said, for the VAR growth model, you would need to convert the growth rate forecasts to the level forecasts.
    Note: Use "exp" function.
  - Since five series have different units, you would like to calculate the sum of test errors for all five series with the following way. For example:
    `(log(series A forecast) - log(series A testset actual))^2`
    This is so called Mean Squared Logarithmic Error (MSLE), which was discussed on Page 20 of the HV_slides02.