Problem Set 5: Computer Exercises

Jenny Petrova

April 18, 2025

5. The following exercises examine the dynamic relationship between U.S. quarterly real GDP rates and U.S. underemployment rates over the time period from 1960Q1 to 2012Q4. This data is obtained from a spreadsheet of quarterly data ('quarterly.xls'). The plots and test results presented below are generated in the R programming language.

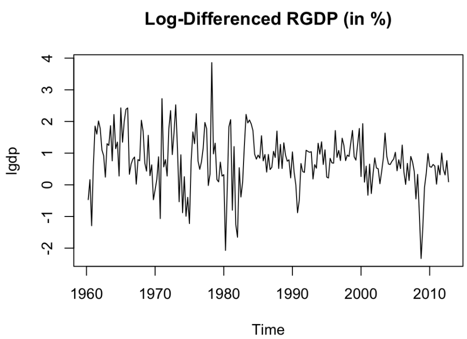
   (a) Let $\{rgdp_t\}$ denote the series of real GDP rates, and let $\{unemp_t\}$ denote the series of unemployment rates. We define the growth rate of real GDP by

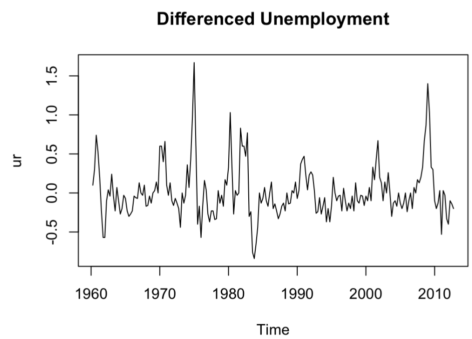
   $$\Delta lgdp_t = lgdp_t - lgdp_{t-1} = \log(rgdp_t) - \log(rgdp_{t-1}) \tag{1}$$

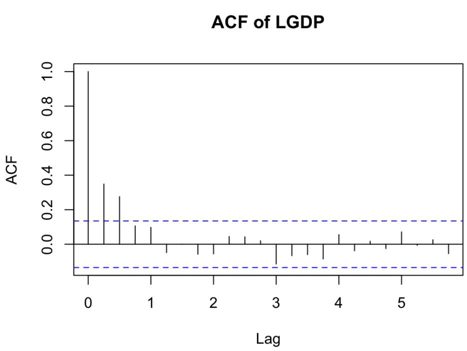
   and the difference in the unemployment rate as

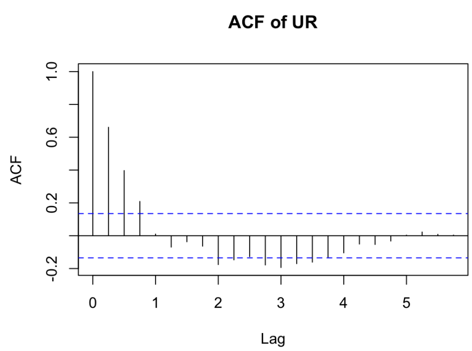
   $$\Delta ur_t = unemp_t - unemp_{t-1}. \tag{2}$$

   Examine the series and ACF plots of $\Delta lgdp_t$ and Δur_t:









   We observe the ACF plots for $\Delta lgdp_t$ and Δur_t decay to zero by lag 1. In the time series plots for $\Delta lgdp_t$ and Δur_t, both series fluctuate around mean zero (and therefore there is not trend in the data), though there exist several spikes in the volatility. Overall, the plots suggest the series are stationary (or, at the very least, near-stationary).

(b) Let $Y_t = (\Delta lgdp_t, \Delta ur_t)'$. We wish to construct a VAR model for the multivariate series Y_t. We select a sample period of the data, from 1973Q1 to 1999Q4. Applying the 'VARselect' command in R to the sample data, we examine the Akaike Information Criterion (AIC) to select the lag length. The results are

```
        $selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     2      2      1      2

$criteria
                 1          2           3           4           5           6
AIC(n) -3.54420958 -3.5968429 -3.54297380 -3.48311998 -3.41209100 -3.36663482
HQ(n)  -3.48094803 -3.4914070 -3.39536350 -3.29333531 -3.18013196 -3.09250142
SC(n)  -3.38789937 -3.3363259 -3.17824997 -3.01418935 -2.83895356 -2.68929058
FPE(n)  0.02889249  0.0274147  0.02894045  0.03074145  0.03303118  0.03460783
                7          8
AIC(n) -3.3396679 -3.35645538
HQ(n)  -3.0233602 -2.99797323
SC(n)  -2.5581169 -2.47069751
FPE(n)  0.0356108  0.03509177
```

Therefore, from the AIC results, we select lag order = 2. We use the 'VARfit' command to fit the model with lag order 2. The relevant resulting output is

```
        VAR Estimation Results:
Estimation results for equation lgdp:
=====================================
lgdp = lgdp.l1 + ur.l1 + lgdp.l2 + ur.l2 + const

        Estimate Std. Error t value Pr(>|t|)
lgdp.l1  0.05921    0.12265   0.483 0.630318
ur.l1   -1.31718    0.34580  -3.809 0.000240 ***
lgdp.l2  0.16750    0.12210   1.372 0.173168
ur.l2    0.97634    0.32904   2.967 0.003752 **
const    0.58682    0.15126   3.880 0.000187 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.7726 on 101 degrees of freedom
Multiple R-Squared: 0.2432, Adjusted R-squared: 0.2133
F-statistic: 8.116 on 4 and 101 DF,  p-value: 1.025e-05


Estimation results for equation ur:
===================================
ur = lgdp.l1 + ur.l1 + lgdp.l2 + ur.l2 + const

        Estimate Std. Error t value Pr(>|t|)
lgdp.l1 -0.06517    0.04280  -1.523   0.1310
ur.l1    0.54683    0.12067   4.532 1.61e-05 ***
lgdp.l2 -0.12447    0.04261  -2.921   0.0043 **
ur.l2   -0.29431    0.11482  -2.563   0.0118 *
const    0.13652    0.05279   2.586   0.0111 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2696 on 101 degrees of freedom
Multiple R-Squared: 0.4563, Adjusted R-squared: 0.4348
F-statistic: 21.19 on 4 and 101 DF,  p-value: 1.036e-12


Covariance matrix of residuals:
        lgdp       ur
lgdp  0.5968 -0.12802
ur   -0.1280  0.07268

Correlation matrix of residuals:
        lgdp      ur
lgdp  1.0000 -0.6147
ur   -0.6147  1.0000
```

From the above output, we find the equations for the fitted models to be

$$lgdp_t = 0.05921\, lgdp_{t-1} - 1.31718\, ur_{t-1} + 0.16750\, lgdp_{t-2} + 0.97634\, ur_{t-2} + 0.58682 + \varepsilon_{1t} \quad (3)$$

and

$$ur_t = -0.06517\, lgdp_{t-1} + 0.54683\, ur_{t-1} - 0.12447\, lgdp_{t-2} - 0.29431\, ur_{t-2} + 0.13652 + \varepsilon_{2t}. \quad (4)$$

The above models can be rewritten in the form

$$\begin{aligned} Y_t = \begin{pmatrix} \Delta lgdp_t \\ \Delta u_t \end{pmatrix} = \begin{pmatrix} 0.58682 \\ 0.13652 \end{pmatrix} &+ \begin{pmatrix} 0.05921 & -1.31718 \\ -0.06517 & 0.54683 \end{pmatrix} \begin{pmatrix} \Delta lgdp_{t-1} \\ \Delta u_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 0.16750 & 0.97634 \\ -0.12447 & -0.29431 \end{pmatrix} \begin{pmatrix} \Delta lgdp_{t-2} \\ \Delta u_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \end{aligned}$$

where $\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim WN(0, \Sigma)$, for

$$\Sigma = \begin{pmatrix} 0.5968 & -0.12802 \\ -0.1280 & 0.07268 \end{pmatrix}.$$

To assess the adequacy of the model, we first check the eigenvalues of the companion coefficient matrix of Y_t:

```
> roots(VAR_fit)
[1] 0.6171136 0.6171136 0.4354861 0.4354861
```

Since each the eigenvalues are < 1 in modulus (inside the unit circle), then the given model satisfies the stability condition. The given VAR process is stable and, therefore, stationary. We also perform residual diagnostics to check the behavior of the residuals:

```
    Portmanteau Test (adjusted)

data:  Residuals of VAR object VAR_fit
Chi-squared = 21.665, df = 24, p-value = 0.5993
```

Our null hypothesis is no autocorrelation among the residuals. The p-value of 0.5993 indicates we *fail* to reject the null hypothesis, meaning there is no evidence of serial correlation in the residuals of the VAR model.

(c) Using the fitted model in part (b), we compute the 1- to 4-step ahead forecasts for $\Delta lgdp_t$ and Δur_t:

```
$lgdp
          fcst      lower    upper       CI
[1,] 1.0756185 -0.4385542 2.589791 1.514173
[2,] 1.0722874 -0.6194204 2.763995 1.691708
[3,] 0.9182895 -0.8071077 2.643687 1.725397
[4,] 0.7695782 -0.9697261 2.508883 1.739304

$ur
              fcst      lower     upper        CI
[1,] -0.2125189441 -0.7409172 0.3158793 0.5283983
[2,] -0.2241942530 -0.8625395 0.4141510 0.6383453
[3,] -0.1272929595 -0.8244409 0.5698550 0.6971479
[4,] -0.0604174745 -0.7751215 0.6542866 0.7147041
```

(d) We check whether the two series, $\Delta lgdp_t$ and Δur_t, Granger cause each other. Granger causality exists when $\Delta lgdp_t$ helps forecast Δur_t, given past Δur_t (and vice versa). The result of the causality test for $\Delta lgdp_t$ is

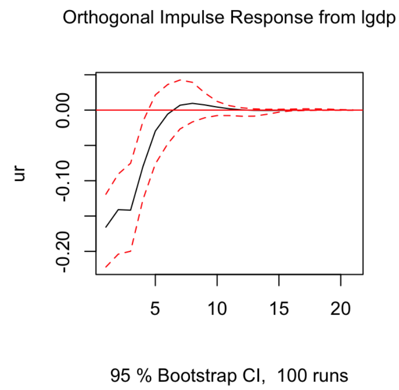
```
> causality(VAR_fit, cause="lgdp")
$Granger

 Granger causality H0: lgdp do not Granger-cause ur

data:  VAR object VAR_fit
F-Test = 5.2493, df1 = 2, df2 = 202, p-value = 0.005991
```

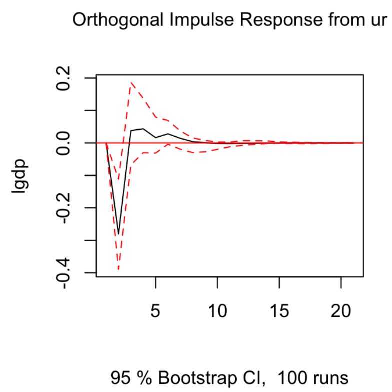
The null hypothesis is $\Delta lgdp_t$ fails to Granger-cause Δur_t. The p-value is $0.005991 < 0.05$, meaning we reject the null. There is statistically significant evidence to suggest $\Delta lgdp_t$ Granger-causes Δur_t. We also examine the result of the causality test for Δur_t:

```
> causality(VAR_fit, cause="ur")
$Granger

 Granger causality H0: ur do not Granger-cause lgdp

data:  VAR object VAR_fit
F-Test = 8.826, df1 = 2, df2 = 202, p-value = 0.0002115
```

The null hypothesis is Δur_t fails to Granger-cause $\Delta lgdp_t$. The p-value is $0.0002115 < 0.05$, meaning we reject the null. There is statistically significant evidence to suggest Δur_t Granger-causes $\Delta lgdp_t$. Therefore both series Granger cause each other.

(e) We plot the orthogonalized impulse response functions of the fitted VAR model:

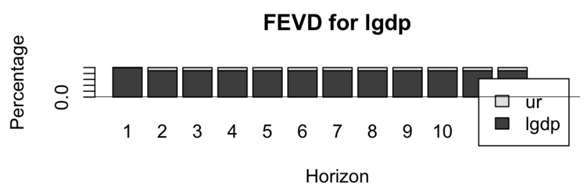


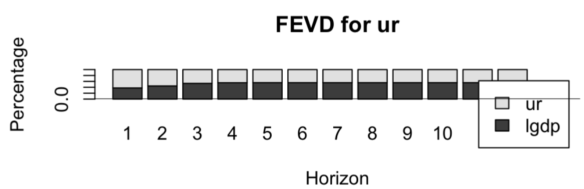


In the top plot, we observe the response of Δur_t to a one-standard-deviation positive shock in $\Delta lgdp_t$. The plot suggests a shock to GDP growth induces lower unemployment rates, but the effect decays to zero after about 8 quarters (i.e. 2 years). In the center plot, we observe the response of $\Delta lgdp_t$ to a one-standard-deviation positive shock in Δur_t. The plot suggests a shock to unemployment rates temporarily lowers GDP growth rates, within the first 2 quarters, after which they decay to zero. The effect of the shock in Δur_t on $\Delta lgdp_t$ lasts for about 4 quarters. For both series, we examine a shock in one series causes temporary decrease in the value of the other. The effects are temporary, and therefore there is no long-term effect on the series resulting from the shocks.

(f) Lastly, we perform forecast error variance decomposition (FEVD) to better understand which shocks drive the variability of $\Delta lgdp_t$ and Δur_t over time. The resulting values and corresponding plots are:

```
$lgdp                                $ur
          lgdp        ur                       lgdp        ur
 [1,] 1.0000000 0.0000000             [1,] 0.3778377 0.6221623
 [2,] 0.8946910 0.1053090             [2,] 0.4462287 0.5537713
 [3,] 0.8969129 0.1030871             [3,] 0.5327767 0.4672233
 [4,] 0.8961527 0.1038473             [4,] 0.5548794 0.4451206
 [5,] 0.8958522 0.1041478             [5,] 0.5575188 0.4424812
 [6,] 0.8950303 0.1049697             [6,] 0.5568498 0.4431502
 [7,] 0.8948726 0.1051274             [7,] 0.5566835 0.4433165
 [8,] 0.8949072 0.1050928             [8,] 0.5568744 0.4431256
 [9,] 0.8949220 0.1050780             [9,] 0.5570291 0.4429709
[10,] 0.8949220 0.1050780            [10,] 0.5570901 0.4429099
[11,] 0.8949182 0.1050818            [11,] 0.5570999 0.4429001
[12,] 0.8949166 0.1050834            [12,] 0.5570984 0.4429016
```





Observing the FEVD for $\Delta lgdp_t$, we see that 100% of the one-quarter-ahead forecast error in GDP growth comes from it's own shock (equivalently, none of the forecast error in GDP growth comes from a shock in unemployment rates). At two-quarters-ahead and beyond, the FEVD of $\Delta lgdp_t$ and Δur_t remains roughly constant. Roughly $89 - 90\%$ of the GDP growth comes from past it's own past shocks, and only about $10 - 11\%$ of the GDP growth comes from past shocks to unemployment rates. Observing the FEVD for Δur_t, we see that about 38% of one-quarter-ahead forecast error in unemployment rates comes from a GDP growth shock, and about 62% of the forecast error unemployment rates comes from it's own shock. For the two-quarters-ahead forecast error in unemployment rates, about 45% comes from GDP growth shocks and about 55% comes from it's own past unemployment shocks. At three-quarters-ahead and beyond, the split roughly evens out to above 56% forecast error from GDP growth shocks and about 44% forecast error from unemployment shocks. Overall, from the results above, we see that GDP growth is largely driven by it's own past shocks, and unemployment rate is influenced more by GDP growth shocks than it's own past shocks.

6. The following exercises examine the cointegrating relationships among five U.S. Treasury rates: 3-month (TB3MS), 6-month (TB6MS), 1-year (GS1), 5-year (GS5), and 10-year (GS10). We focus on the sample period from 1981-01-01 to 2024-12-31. This data is obtained from FRED using the '(quantmod)' library in *R*. The plots and test results presented below are generated in the *R* programming language.

(a) Assume all the series are I(1). We determine the number of cointegrating relationships among the five series using Johansen's trace and maximum eigenvalue tests. The relevant results for Johansen's trace test are:

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.16169038 0.10662464 0.04224824 0.01872348 0.01541880

Values of teststatistic and critical values of test:

           test 10pct  5pct  1pct
r <= 4 |   8.17  6.50  8.18 11.65
r <= 3 |  18.12 15.66 17.95 23.52
r <= 2 |  40.82 28.71 31.52 37.22
r <= 1 | 100.13 45.23 48.28 55.43
r = 0  | 192.90 66.49 70.60 78.87

Eigenvectors, normalised to first column:
(These are the cointegration relations)

            TB3MS.l1  TB6MS.l1     GS1.l1     GS5.l1   GS10.l1
TB3MS.l1  1.00000000  1.000000   1.000000  1.0000000  1.000000
TB6MS.l1 -1.46304272 -3.616003   8.977669 -2.3211131 -3.214741
GS1.l1    0.43053966  2.493695 -16.580218  0.9557054  2.506886
GS5.l1    0.08946576  0.625419  17.187173 -1.0324030 -2.755318
GS10.l1  -0.09364004 -0.762665 -10.489190  1.5882451  2.160374
```

At the 5% confidence level, we reject $r = 0$, $r \leq 1$, $r \leq 2$, and $r \leq 3$. We fail to reject $r \leq 4$, where the critical value is greater than the test statistic ($23.52 > 18.12$). There is significant evidence to suggest the number of cointegrating relations is $r = 4$.

The relevant results for Johansen's maximum eigenvalue tests are:

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
[1] 0.16169038 0.10662464 0.04224824 0.01872348 0.01541880

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 4 |  8.17  6.50  8.18 11.65
r <= 3 |  9.94 12.91 14.90 19.19
r <= 2 | 22.71 18.90 21.07 25.75
r <= 1 | 59.31 24.78 27.14 32.14
r = 0  | 92.77 30.84 33.32 38.78

Eigenvectors, normalised to first column:
(These are the cointegration relations)

            TB3MS.l1  TB6MS.l1     GS1.l1     GS5.l1   GS10.l1
TB3MS.l1  1.00000000  1.000000   1.000000  1.0000000  1.000000
TB6MS.l1 -1.46304272 -3.616003   8.977669 -2.3211131 -3.214741
GS1.l1    0.43053966  2.493695 -16.580218  0.9557054  2.506886
GS5.l1    0.08946576  0.625419  17.187173 -1.0324030 -2.755318
GS10.l1  -0.09364004 -0.762665 -10.489190  1.5882451  2.160374
```

In contrast to our previous results, we only reject up to $r \leq 2$ at the 5% confidence level. We fail to reject $r \leq 3$, where the critical value is greater than the test statistic ($25.75 > 22.71$). There is significant evidence to suggest the number of cointegrating relations among the five series is $r = 3$. We set the number of cointegrating relations to $r = 4$, and construct a (4×5) matrix β, where each of the 4 columns are a cointegrating vector of the five series. From both tests, the resulting estimated cointegrating vectors (normalized to the first column, 'TB3MS.l2') are:

$$\beta_1 = \begin{pmatrix} 1.00000 \\ -1.46304 \\ 0.43054 \\ 0.08947 \\ -0.09364 \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} 1.00000 \\ -3.61600 \\ 2.49370 \\ 0.62542 \\ -0.76267 \end{pmatrix}, \quad \beta_3 = \begin{pmatrix} 1.00000 \\ 8.97767 \\ -16.58022 \\ 17.18717 \\ -10.48919 \end{pmatrix}, \quad \beta_4 = \begin{pmatrix} 1.00000 \\ -2.32111 \\ 0.95570 \\ -1.03240 \\ 1.58825 \end{pmatrix}.$$

Corresponding Code

```r
library(readxl)
library(tseries)
library(quantmod)
library(vars)
library(urca)
library(tsDyn)
library(forecast)

# Problem 5
data <- read_excel("/Users/jennypetrova/Desktop/Time Series Analysis/Codes/
    quarterly.xls")
head(data)

# 5(a)
rgdp_ts  <- ts(data$RGDP,  start = c(1960, 1), frequency = 4)
unemp_ts <- ts(data$Unemp, start = c(1960, 1), frequency = 4)

plot(rgdp_ts,  main = "Real GDP (Levels)")
plot(unemp_ts, main = "Unemployment (Levels)")

lgdp <- diff(log(rgdp_ts)) * 100
ur <- diff(unemp_ts)

plot(lgdp,  main = "Log-Differenced RGDP (in %)")
plot(ur,    main = "Differenced Unemployment")

acf(lgdp, main = "ACF of LGDP")
acf(ur,   main = "ACF of UR")

# 5 (b)
# Merge data
merged_data <- cbind(lgdp, ur)
colnames(merged_data) <- c("lgdp", "ur")
head(merged_data)
plot(merged_data, main = "LGDP and UR Combined", plot.type = "single", col =
    1:2)
legend("topright", legend = c("LGDP","UR"), col = 1:2, lty = 1)

# VAR
y <- window(merged_data, start = c(1973, 1), end = c(1999, 4))
plot(y)
VARselect(y, lag.max = 8, type = "const")
VAR_fit <- VAR(ts(y), p=2, type="const")
summary(VAR_fit)
roots(VAR_fit, modulus=FALSE)
roots(VAR_fit)

res <- serial.test(VAR_fit, lags.pt=8, type="PT.adjusted")
res
plot(res)
plot(res, names="lgdp")
plot(res, names="ur")

# 5 (c)
# Forecasting
(fcast <- predict(VAR_fit, n.ahead=8))
plot(fcast)
fanchart(fcast)
```

```r
# 5 (d)
# Granger causality
causality(VAR_fit, cause="lgdp")
causality(VAR_fit, cause="ur")

# 5 (e)
# Impulse response function
irf_gdp <- irf(VAR_fit, impulse="ur", response="lgdp", n.ahead=20)
plot(irf_gdp, ylab="lgdp")
irf_spread <- irf(VAR_fit, impulse="lgdp", response="ur", n.ahead=20)
plot(irf_spread, ylab="ur")
plot(irf(VAR_fit))

# 5(f)
# FEVD
fevd_result <- fevd(VAR_fit, n.ahead=12)
fevd_result
plot(fevd_result)

# Problem 6
# Define symbols
start_date <- "1981-01-01"
end_date   <- "2024-12-31"
symbols <- c("TB3MS", "TB6MS", "GS1", "GS5", "GS10")

# Retrieve data from FRED
getSymbols(symbols, src = "FRED", from = start_date, to = end_date)

# Merge data
rates <- merge(TB3MS, TB6MS, GS1, GS5, GS10)
colnames(rates) <- c("TB3MS", "TB6MS", "GS1", "GS5", "GS10")
head(rates)
tail(rates)

# Optimal Lag
lag_sel <- VARselect(rates, lag.max = 12, type = "const")
lag_sel$selection

# Johansen Trace Test
ctest_tr <- ca.jo(rates, type  = "trace", ecdet = "none", K = 2,
                      spec = "transitory", season = 4)
summary(ctest_tr)

# Johansen Maximum Eigenvalue Test
ctest_eig <- ca.jo(rates, type  = "eigen", ecdet = "none", K = 2,
                      spec = "transitory", season = 4)
summary(ctest_eig)

VECM_model <- VECM(rates, lag=1, r=3, estim = "ML", LRinclude = "none")
summary(VECM_model)
```