Problem Set 7

Tao Wu

December 3, 2023
Professor Suleyman Taspinar
ECON 387: Advanced Econometrics

1 Question 1

(20 pts) Define the following terms in your own words.

(a) Strict Stationarity is a stringent form of the stationarity assumption in time series analysis. It requires all the moments including autocovariance to be the same (no heterogeneity).

(b) Compared to strict stationarity, weak stationarity is a more commonly assumed property in many time series analyses. It only requires constant mean and variance and weak stationarity implies strict stationarity, but not the other way around.

(c) A weak dependence means that the observations in time series data are not strongly correlated with each other. In other words, as the time lags between observations increase to a sufficiently large number, their autocorrelation decreases, hence weak dependence.

(d) The mean squared forecast error (MSFE) is the expected value of the square of the forecast error, a measure that tells us about the accuracy of our time series prediction. One of the methods to get at this measurement is through pseudo-out-of-sample forecasting (POOS).

$$MSFE = E[(Y_{T+1} - \hat{Y}_{T+1|T})^2] = E(u_{T+1}^2)$$

$$Var(u_{T+1}) = E(u_{T+1}^2) - (E(u_{T+1}))^2$$
$$Var(u_{T+1}) = E(u_{T+1}^2) - 0$$
$$MSFE \approx \sigma_u^2$$

This method considers estimation error and does not need assumptions like stationarity and homoskedasticity, which oftentimes is the preferred method in the real world and is feasible in practice.

(e) In the context of time series analysis, a trend is a long-run movement or pattern of the values of a random variable. A deterministic trend is a nonrandom, systematic, and predictable pattern of movement in the form of a mathematical function.

(f) In contrast to a deterministic trend, a stochastic trend is an unpredictable fluctuation of time series data.

(g) A random walk process is a primary example of a stochastic trend. The future value of the trend is unpredictable and random and is calculated by the current value plus random noises. The random walk process is often used to model phenomena where future values are influenced by recent past values, but where there is also an element of randomness or unpredictability.

(h) The presence of a unit root means the graph of a time series variable has a stochastic trend, the opposite of stationarity. A unit root implies that the process does not have a stable mean, and any deviation from the mean is permanent. Think of unit roots when we talk about stochastic trends.

(i) Spurious regression refers to two stochastically trending time series that appear to be highly correlated and share a significant association, but in fact, there is no true underlying relationship between them.

(j) The Dickey-Fuller test is a regression-based test for the presence of a unit root or a stochastic trend using the t-statistic for testing $\delta = 0$ as the null hypothesis.

Nonstationarity

- Consider the following models:

$$\text{Deterministic trend}: \ Y_t = \beta_0 + \beta_1 t^2, \tag{9}$$
$$\text{AR}(1): \ Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t, \tag{10}$$
$$\text{Random walk}: \ Y_t = Y_{t-1} + u_t, \tag{11}$$
$$\text{Random walk with drift}: \ Y_t = \beta_0 + Y_{t-1} + u_t. \tag{12}$$

- It is often helpful to study time series model by simulation.
- This enables the main features of the model to be observed in plots, so that when historical data exhibit similar features, the model may be selected as a potential candidate.

2 Question 2

(20 pts) Suppose Y_t is a random walk process, $Y_t = Y_{t-1} + u_t$, for t = 1, 2, . . . , T, where $Y_0 = 0$ and u_t is i.i.d. with mean 0 and variance σ_u^2.

(a) Given that the random walk under $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$ and u_t is iid and distributed normally.

$$E(Y_t) = E(Y_{t-1} + u_t)$$

$$E(Y_t) = E(Y_{t-1}) + E(u_t)$$

Thus, the mean of Y_t is the sum of the means of the error terms up to time t

$$E(Y_t) = Y_0 + \sum_{i=1}^{t} E(u_i)$$

$$E(Y_t) = 0$$

As for the variance of Y_t, we have the following:

$$Var(Y_t) = Var(Y_{t-1} + u_t)$$

$$Var(Y_t) = Var(Y_{t-1}) + Var(u_t) + 2Cov(Y_{t-1}, u_t)$$

$$Var(Y_t) = Var(Y_{t-1}) + Var(u_t)$$

Thus, the variance of Y_t is the sum of the variances of the error terms up to time t

$$Var(Y_t) = Y_0 + \sum_{i=1}^{t} Var(u_i) = \sigma_u^2$$

$$Var(Y_t) = t \times \sigma_u^2$$

(b)

$$Cov(Y_t, Y_{t-k}) = Cov(Y_{t-1} + u_t, Y_{t-k})$$

By the covariance property,

$$Cov(Y_t, Y_{t-k}) = Cov(Y_{t-1}, Y_{t-k}) + Cov(u_t, Y_{t-k})$$

$$Cov(Y_t, Y_{t-k}) = Cov(Y_{t-2}, Y_{t-k}) + Cov(u_{t-1}, Y_{t-k}) + Cov(u_t, Y_{t-k})$$

$$\vdots$$

$$Cov(Y_t, Y_{t-k}) = Cov(Y_{t-k}, Y_{t-k}) + \sum_{i=0}^{k-1} Cov(u_{t-i}, Y_{t-k})$$

$$Cov(Y_t, Y_{t-k}) = Cov(Y_{t-k}, Y_{t-k}) + 0$$

$$Cov(Y_t, Y_{t-k}) = Var(Y_{t-k}) = (t-k) \times \sigma_u^2$$

(c) Y_t has its variance increasing in a linear fashion and increasing covariance with past values, hence not stationary.

3 Question 3

A researcher tests for a stochastic trend in $ln(IP_t)$, using the AR(5) model including 4 lages of $\Delta ln(IP_t)$ plus additional regressor t that is a linear time trend. We will now use the extension of the Dickey–Fuller test, augmented Dickey-Fuller (ADF) statistic, to the AR(5) model for a stochastic trend (unit root) in $ln(IP_t)$. Since the null distribution of the Dickey-Fuller test is not standard, the critical values need to be calculated using simulation methods and displayed below in a table.

TABLE 15.4 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

For testing the hypothesis that $H_0 : \delta = 0$ with given information of $\delta = 0.007$ and its standard error is equal to 0.0037, we plug them into the formula

$$t = \frac{\overline{x} - \mu_o}{\frac{s}{\sqrt{n}}} \sim t_{n-1} = \frac{-0.007 - 0}{0.0037} = -1.892$$

Since the AR(5) model has intercept as well as time trend regressor in it, we compare our t-statistic to the corresponding one-sided critical values of augmented Dickey-Fuller (ADF) statistic. $-1.892 > -3.12$ even at the 10% significance level, thus failing to reject the null hypothesis. Finally, we conclude that $ln(IP_t)$ has a unit root, and therefore it contains a stochastic trend.

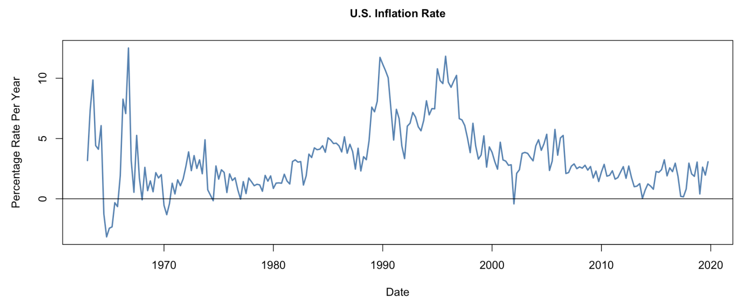
4 Question 4

(a) The units of the inflation rate are defined at an **annual rate** of percentage,

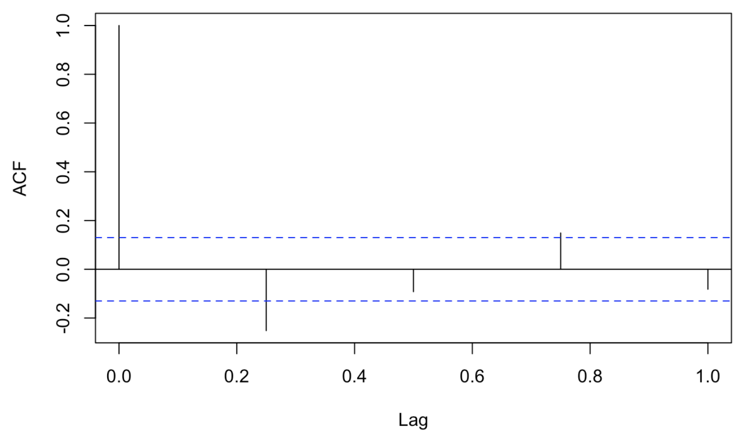
$$Infl_t = 400 \times [ln(PCEP_t) – ln(PCEP_{t-1})]$$

because we are using the percentage change of a time series $Infl_t$: $100 \times \Delta ln(PCEP_t)$. Also, since the time series data sample period is quarterly, to begin with. We multiply the first difference by 4 to get the percentage per year.

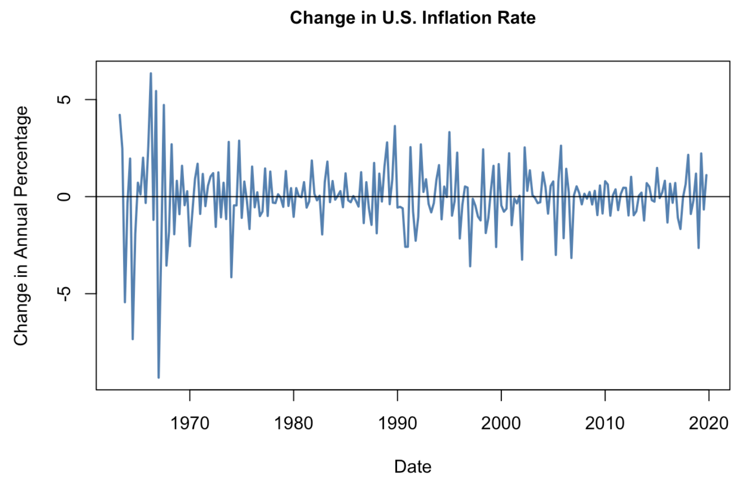
(b) Based on the graph below that shows the time series plot of the percentage inflation rate at an annual level, I think the inflation rate has a stochastic trend because the plot seems to have an unpredictable fluctuation without constant mean and variance.



(c) The first four autocorrelations of $\Delta Infl$ are [1,] 1.00000000 [2,] -0.25171480 [3,] -0.09192252 [4,] 0.14872972 [5,] -0.08173708 respectively.



(d) The $\Delta Infl$ graph below does look choppy and jagged. The plot behavior is consistent with the first autocorrelations we computed in part c because correlations between annual inflation rates traverse above and below the x-axis meaning correlations become positive and negative and every stay still.



(e) Knowing the change in inflation over the current quarter **does not** help predict the change in inflation over the next quarter because the change in inflation rate plot appears to be a random process and stochastic in nature. Even β_1 in AR(1) is tested to be statistically significant, the R^2 and adjusted R^2 are both poorly low. The model's prediction won't be accurate anyway.

(f) The AR(2) model is not necessarily better than the AR(1) model. Even β_1 in AR(2) is tested to be more statistically significant and that of AR(1), the R^2 and adjusted R^2 are both equally poorly low. The model's prediction won't be accurate anyway.

AR(1) and AR(2) Estimation Results

	Dependent variable:	
	AR(1)	AR(2)
L(tsChangeInfl, 1)	-0.252**	-0.320***
	(0.120)	(0.118)
L(tsChangeInfl, 2)		-0.172*
		(0.089)
Constant	-0.021	-0.039
	(0.110)	(0.109)
Observations	227	226
R2	0.065	0.105
Adjusted R2	0.061	0.097
Residual Std. Error	1.662 (df = 225)	1.626 (df = 223)
F Statistic	15.738*** (df = 1; 225)	13.085*** (df = 2; 223)
Note:		*p<0.1; **p<0.05; ***p<0.01

(g) Minimizing BIC(p) or AIC(p) trades off bias and variance to determine a "best" value of p for our forecast is what we need to do here.

So, the lag length suggested by the BIC is 8 lags with the lowest value of 848.0861 from the m8 model. The lag length chosen by the AIC is 8 lags as well with the lowest value being 814.1499 produced by the m8 model (shown in the screenshot below).

```
> rownames(tmp) = c("AIC","BIC")
> tmp
          m1       m2       m3       m4       m5       m6       m7       m8
AIC 878.8990 866.1199 857.0257 854.5520 847.2123 831.8408 818.2978 814.1499
BIC 889.1739 879.8021 874.1062 875.0219 871.0625 859.0622 848.8812 848.0861
>
```

(h) Using the AR(2) model to predict the change in inflation from 2019Q4 to 2020Q1:

$$\Delta Infl_t = \beta_0 + \beta_1 \Delta Infl_{t-1} + \beta_2 \Delta Infl_{t-2} + \epsilon_t$$

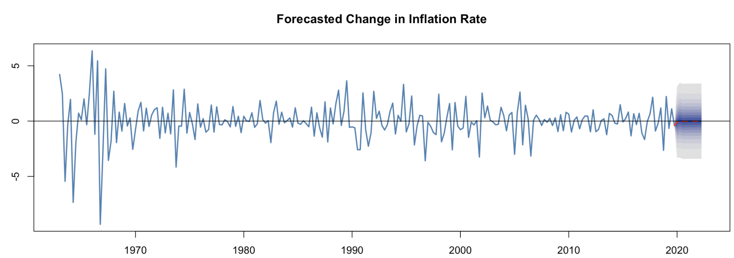
$$\widehat{\Delta Infl_t} = \widehat{\beta_0} + \widehat{\beta_1} \Delta Infl_{t-1} + \widehat{\beta_2} \Delta Infl_{t-2}$$

$$\Delta \widehat{Infl_{2020Q1}} = Infl_{2020Q1} - Infl_{2019Q4}$$

$$Infl_{2020Q1} = \Delta \widehat{Infl_{2020Q1}} + Infl_{2019Q4}$$

$$Infl_{2020Q1} = -0.093137959 + 2.68254333 = 2.589405371$$

```
> fc$mean
             Qtr1         Qtr2         Qtr3         Qtr4
2020 -0.093137959  0.084276664 -0.023425476 -0.022103470
2021 -0.003221430 -0.009260313 -0.010786234 -0.009235847
2022 -0.009438979 -0.009654217
```



```r
library(readr)
library(stargazer)
library(sandwich)
library(lmtest)
library(car)
library(haven)
library(readxl)
library(AER)
library(forecast)
library(scales)
library(urca)
library(dynlm)
library(stats)
mydata = read.table("PCECTPI.csv", header = T,
                    sep = ",", skip = 0)
tsPCEP = ts(mydata, start = c(1962,3), frequency = 4)
tsPCEP = window(tsPCEP, start=c(1962,3), end=c(2019,4))

#Compute the inflation rate
tsInfl = 400*diff(log(tsPCEP[,"PCECTPI"]))

#Plot the value of Infl from 1963Q1 through 2019Q4
plot(400*diff(log(tsPCEP[,"PCECTPI"])),
     col = "steelblue",
     lwd = 2,
     ylab = "Percentage Rate Per Year",
     xlab = "Date",
     main = "U.S. Inflation Rate",
     cex.main=1
)
abline(h=0)

#Compute the first four autocorrelations of ΔInfl.
tsChangeInfl = diff(tsInfl)
acfPlot = acf(tsChangeInfl, main = "",
              lag.max = 4,
              plot = T)
acfPlot$acf
```

```r
#Plot the value of ΔInfl from 1963Q1 through 2019Q4
plot(tsChangeInfl,
     col = "steelblue",
     lwd = 2,
     ylab = "Change in Annual Percentage",
     xlab = "Date",
     main = "Change in U.S. Inflation Rate",
     cex.main=1
)
abline(h=0)

# AR(1) Run an OLS regression of ΔInfl on an intercept and its first lag
tsChangeInfl
r1<- dynlm(tsChangeInfl~L(tsChangeInfl ,1),data = tsChangeInfl,
           start = c(1962,4),end = c(2019,4))
r1Var=vcovHC(r1,type = "HC1")
r1Se=sqrt(diag(r1Var))

# AR(2) Run an OLS regression of ΔInfl on an intercept and its first 2 lags
r2<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r2Var=vcovHC(r2,type = "HC1")
r2Se=sqrt(diag(r2Var))

#estimation results from AR(1) & AR(2)
stargazer(r1, r2, se=list(r1Se,r2Se),
          type ="text",
          title = "AR(1) and AR(2) Estimation Results",
          model.names = F,
          model.numbers = F,
          column.labels = c("AR(1)","AR(2)"),
          header = F,
          font.size = "small",
          dep.var.labels.include = F,
          label = "t1")
```

```r
#Estimate an AR(p) model for p = 1, 2, . . . , 8, for AIC and BIC values
r1<- dynlm(tsChangeInfl~L(tsChangeInfl ,1),data = tsChangeInfl,
           start = c(1962,4),end = c(2019,4))
r2<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r3<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2) + L(tsChangeInfl ,3),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r4<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2) + L(tsChangeInfl ,3)
           + L(tsChangeInfl ,4),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r5<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2) +
             L(tsChangeInfl ,3) + L(tsChangeInfl ,4) + L(tsChangeInfl ,5),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r6<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2) +
             L(tsChangeInfl ,3) + L(tsChangeInfl ,4) + L(tsChangeInfl ,5)
           + L(tsChangeInfl ,6),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r7<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2) +
             L(tsChangeInfl ,3) + L(tsChangeInfl ,4) + L(tsChangeInfl ,5)
           + L(tsChangeInfl ,6) + L(tsChangeInfl ,7),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
r8<- dynlm(tsChangeInfl~L(tsChangeInfl ,1) + L(tsChangeInfl ,2) +
             L(tsChangeInfl ,3) + L(tsChangeInfl ,4) + L(tsChangeInfl ,5)
           + L(tsChangeInfl ,6) + L(tsChangeInfl ,7) + L(tsChangeInfl ,8),
           data = tsChangeInfl,start = c(1962,4),
           end = c(2019,4))
tmp = data.frame("m1"=c(AIC(r1),BIC(r1)),
                 "m2"=c(AIC(r2),BIC(r2)),
                 "m3"=c(AIC(r3),BIC(r3)),
                 "m4"=c(AIC(r4),BIC(r4)),
                 "m5"=c(AIC(r5),BIC(r5)),
                 "m6"=c(AIC(r6),BIC(r6)),
                 "m7"=c(AIC(r7),BIC(r7)),
                 "m8"=c(AIC(r8),BIC(r8)))
rownames(tmp) = c("AIC","BIC")
tmp
```

```r
#Use the AR(2) model to predict the change in inflation from 2019Q4 to 2020Q1
#that is, to predict the value of Infl2020Q1
m1 = arima(tsChangeInfl, order = c(2,0,0))
stargazer(m1,
          type ="latex",
          title = "AR(2) Estimation Results",
          model.names = F,
          model.numbers = F,
          column.labels = "AR(2)",
          header = F,
          font.size = "small",
          dep.var.labels.include = F,
          label = "t1"
)
fc = forecast(m1, h = 10, level = seq(5, 99, 10))
fc$mean
tsInfl
plot(fc,
     main = "Forecasted Change in Inflation Rate",
     showgap = F,
     fcol = "red",
     flty = 2,
     col = "steelblue",
     lwd = 2)
abline(h=0)
```