Final Exam

Tao Wu

December 23, 2023
Professor Suleyman Taspinar
ECON 387: Advanced Econometrics

1 Question 1

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

(a) Dynamic causal effects is about how a change in one variable can affect the other variable over time. The word "dynamic" indicates temporal aspects of the relationship between two variables. The impact of the change, after the initial period, will continue to have delayed consequences as time goes on. Due to the dynamic nature of such an effect, an economic model called the Distributed lag model selected to estimate and study dynamic causal effects should capture the lagged effects of changes in the independent variable on the dependent variable. This way, we can analyze how shocks or random changes affect the outcome variable over several time periods.

(b) Past and present exogeneity is an updated definition when we deal with a distributed lag model where the variables' values both in the past and currently do not correlate with the error term, regardless of the number of lags. Such an assumption needs to be made to avoid bias in parameter estimates. $E(u_t|X_t, X_{t-1}, X_{t-2}, ...) = 0$

(c) Strict exogeneity is a more stringent form of past and present exogeneity which basically extends its zero condition mean to the future values of independent variables in the model, hence its other name "Past, present, and future exogeneity". Strict exogeneity implies past and present exogeneity.

(d) The dynamic multiplier, more specifically, the h-period dynamic multiplier is like the ceteris paribus effect in OLS regression analysis where we hold everything else constant besides the variable we are studying. The h-period dynamic multiplier is the effect β_{h+1} of a change in the independent variable X_{t-h} on the dependent variable Y_t after h periods. The reason why we have a +1 in the subscript of β is because β_1 itself has a special name to it. It's called impact effect rather than a 1st-period dynamic multiplier. Keep this detail in mind.

(e) The cumulative dynamic multiplier is the cumulative sum of all dynamic multipliers of however many periods plus the impact effect.

(f) The long-run cumulative dynamic multiplier is a concept related to the cumulative dynamic multiplier. The difference between them is that The long-run cumulative dynamic multiplier includes and considers the effects over an extended period just to be more comprehensive.

(g) In the presence of issues such as heteroskedasticity and autocorrelation of u_t, normal standard errors may still be biased. To counter these issues, HAC standard errors are the solution, and below is its formula.

$$\sqrt{Var(\hat{\beta}_1) = [\frac{1}{T}\frac{\sigma_v^2}{(\sigma_x)^2}] \times f_T}$$

Heteroskedasticity and autocorrelation consistent (HAC) standard errors are also called heteroskedasticity and autocorrelation robust (HAR) standard errors.

2 Question 2

Suppose you want to estimate the effect of monetary policy on inflation. That is, you want to estimate the dynamic causal effect of a change in short-run interest rate on inflation. To that end, you estimate a distributed lag model relating inflation to current and lags of short-run interest rates by the OLS estimator. Under which assumptions will the OLS estimator deliver the causal effect of short-run interest rates on inflation? Are these assumptions likely to be satisfied for a modern economy? Explain.

The key assumptions that make the OLS estimator unbiased and deliver the dynamic causal effect of short-run interest rate on inflation in the context of distributed lag models are the followings:

1. the strict exogeneity at best or past and present exogeneity for all lagged values of short-run interest rates.

2. both short-run interest rates and inflation as varibales have a stationary distribution. Plus, as the amount of time separating each variable and their time lags becomes large, both of these variables become independently distributed to their respective lags value. If j is sufficiently large, then

$$(Y_t, X_t) \sim iid(Y_{t-j}, X_{t-j})$$

3. there is no perfect multicollinearity meaning no two or more varibales are extremely highly correlated that they virtually can be expressed as one another.

Generally speaking, whether a variable is exogenous or not depends on the context and the context for our problem here makes these assumptions **unlikely** to be satisfied for a modern economy. Since the world economy is integrated, there is something called simultaneous causality according to the textbook on a case study about export and import. We can extend that thought when considering monetary policy and inflation. The short-term interest rate is set by the central bankers who deliberately take into account the current and the expected future economic condition as well as the inflation rate. Moreover, since we all know the idea of supply and demand in the modern economy, the rate of inflation will be influenced and affected by the interest rate which could reduce aggregate demand as it goes up.

3 Question 3

For the questions below, you will again use the sample period 1963Q1–2019Q4, but data before 1963 may be used, if necessary, as initial values for lags in regressions. Let PCEP denote Personal Consumption Expenditures. Compute the inflation rate,

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

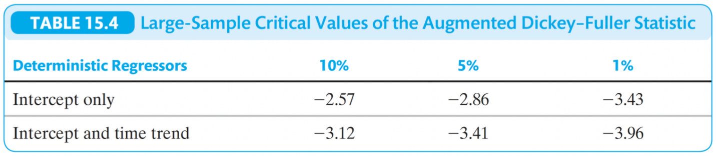


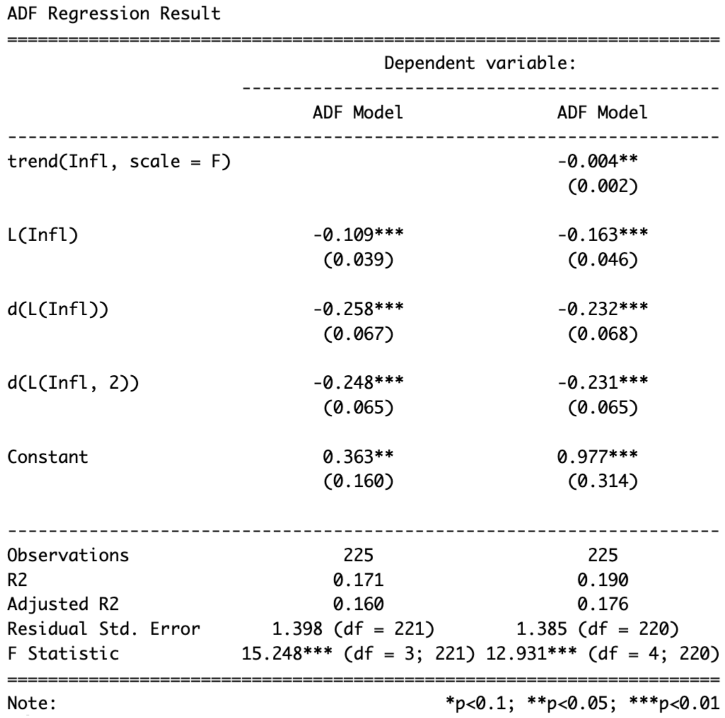
TABLE 15.4 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

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

(a) Use the augmented Dickey-Fuller (ADF) test for a stochastic trend in Infl and the ADF test statistic is $-0.109 \div 0.039 \approx -2.795$. The corresponding critical value from Table 15.4 above is -2.86 at 5% level on the 1st row where it says "Intercept only". So we fail to reject the null hypothesis that Infl has a **stochastic** trend.

$$H_o : \delta = 0 \ (random\ walk)$$

$$H_1 : \delta < 0 \ (stationary)$$



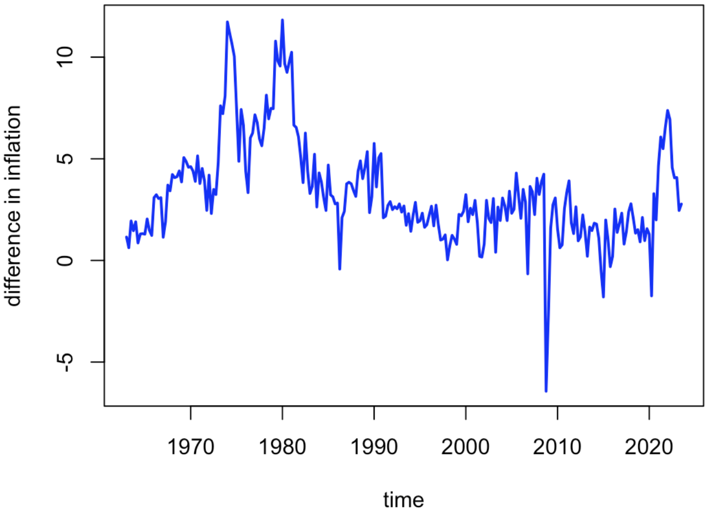
```
ADF Regression Result
=====================================================================
                                     Dependent variable:
                        ---------------------------------------------
                              ADF Model              ADF Model
---------------------------------------------------------------------
trend(Infl, scale = F)                               -0.004**
                                                      (0.002)

L(Infl)                       -0.109***              -0.163***
                               (0.039)                (0.046)

d(L(Infl))                    -0.258***              -0.232***
                               (0.067)                (0.068)

d(L(Infl, 2))                 -0.248***              -0.231***
                               (0.065)                (0.065)

Constant                       0.363**                0.977***
                               (0.160)                (0.314)

---------------------------------------------------------------------
Observations                     225                    225
R2                              0.171                  0.190
Adjusted R2                     0.160                  0.176
Residual Std. Error       1.398 (df = 221)       1.385 (df = 220)
F Statistic           15.248*** (df = 3; 221) 12.931*** (df = 4; 220)
=====================================================================
Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

(b) Use the augmented Dickey-Fuller (ADF) test for a **deterministic** trend in Infl and the ADF test statistic is $-0.163 \div 0.046 \approx -3.543$. The corresponding critical value from Table 15.4 above is -3.41 at 5% level on the 2nd row where it says "Intercept and time trend". So we can reject the null hypothesis that Infl has a stochastic trend in favor of the alternative that it is stationary around a deterministic linear time trend.

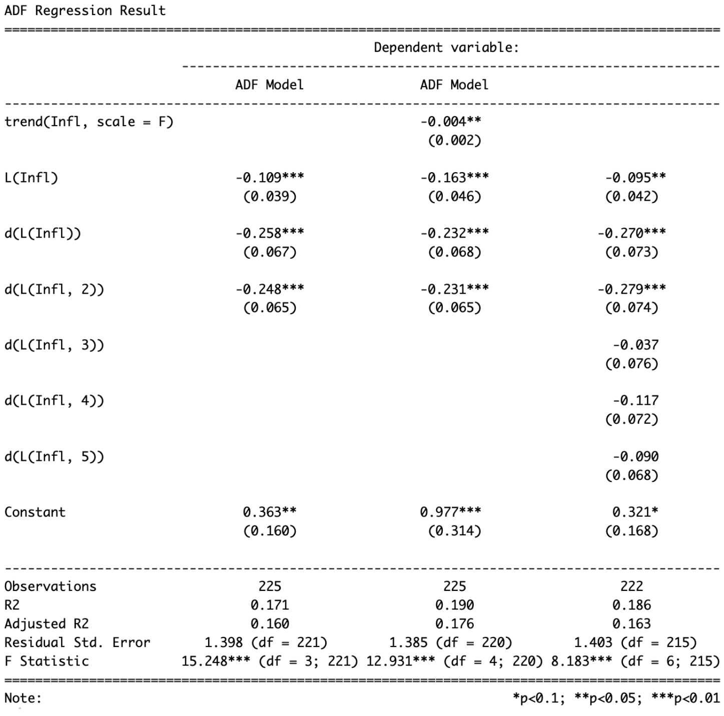
$$H_o : \delta = 0 \; (random\ walk)$$

$$H_1 : \delta < 0 \; (stationary)$$

Conclusion: the ADF test based on Equation 2 is **not** preferred to the test based on Equation 1 because the alternative hypothesis of stationarity is appropriate for series such as the U.S. inflation rate that does not exhibit growth over the long run. It makes less sense to test for stationarity around a deterministic trend. Plus, I didn't see any indication of a deterministic trend in $Infl$ when I took a look at its plot which is attached here.



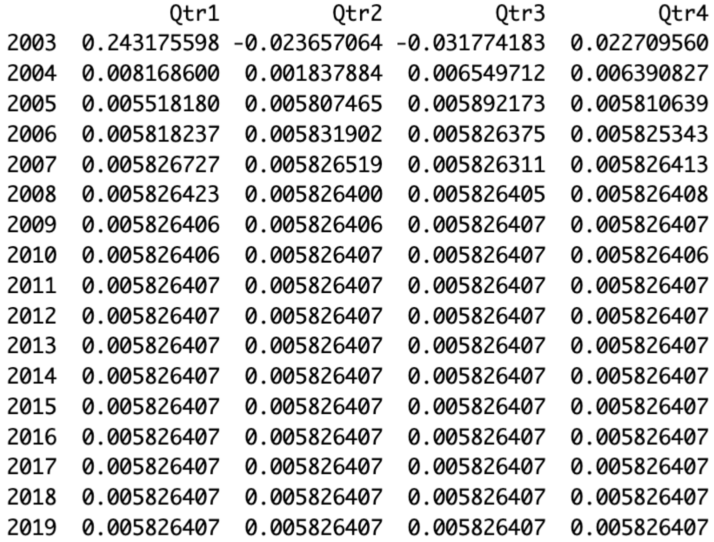
(c) My result becomes less significant statistically when I use more lags in the ADF test used in part a). What remains unchanged is with the slight increase of ADF test statistic -2.26, I still fail to reject the H_0 stating that $Infl$ has a stochastic trend (unit root).



```
ADF Regression Result
==========================================================================================
                                          Dependent variable:
                       -------------------------------------------------------------------
                              ADF Model               ADF Model
------------------------------------------------------------------------------------------
trend(Infl, scale = F)                                -0.004**
                                                       (0.002)

L(Infl)                       -0.109***               -0.163***               -0.095**
                               (0.039)                 (0.046)                 (0.042)

d(L(Infl))                    -0.258***               -0.232***               -0.270***
                               (0.067)                 (0.068)                 (0.073)

d(L(Infl, 2))                 -0.248***               -0.231***               -0.279***
                               (0.065)                 (0.065)                 (0.074)

d(L(Infl, 3))                                                                  -0.037
                                                                               (0.076)

d(L(Infl, 4))                                                                  -0.117
                                                                               (0.072)

d(L(Infl, 5))                                                                  -0.090
                                                                               (0.068)

Constant                       0.363**                 0.977***                0.321*
                               (0.160)                 (0.314)                 (0.168)

------------------------------------------------------------------------------------------
Observations                     225                     225                     222
R2                              0.171                   0.190                   0.186
Adjusted R2                     0.160                   0.176                   0.163
Residual Std. Error        1.398 (df = 221)        1.385 (df = 220)        1.403 (df = 215)
F Statistic            15.248*** (df = 3; 221) 12.931*** (df = 4; 220) 8.183*** (df = 6; 215)
==========================================================================================
Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```

(d) Based on the test I carried out in part a), the AR model for Infl **may likely** contain a unit root even though the failure to reject a null hypothesis **does not** mean that the null hypothesis - the presence of a unit root - is true. Let me explain. First, based on the plot of $Infl$ above, the inflation series looks like it **may have** a stochastic trend due to the upward trends from the '60s until the '70s and the persistent downward trend from the '90s to 2000s. I decided to use the version of the ADF test without a time trend regressor in it and its test statistics give us failure to reject the null $H_0 : \delta = 0$ (*random walk*), which gives credence to the conclusion that Infl indeed has a random walk stochastic trend. Finally, based on this conclusion, we now know to use $\Delta Infl$ instead of Infl for regression analysis and forecasting, paving the way for the following questions.

(e) Using the AR(2) model for $\Delta Infl$ to compute pseudo out-of-sample forecasts $\widehat{\Delta Infl}_{2003Q1|2002Q4}$, $\widehat{\Delta Infl}_{2003Q2|2003Q1}$, ..., $\widehat{\Delta Infl}_{2019Q4|2019Q3}$



```
            Qtr1         Qtr2         Qtr3         Qtr4
2003  0.243175598 -0.023657064 -0.031774183  0.022709560
2004  0.008168600  0.001837884  0.006549712  0.006390827
2005  0.005518180  0.005807465  0.005892173  0.005810639
2006  0.005818237  0.005831902  0.005826375  0.005825343
2007  0.005826727  0.005826519  0.005826311  0.005826413
2008  0.005826423  0.005826400  0.005826405  0.005826408
2009  0.005826406  0.005826406  0.005826407  0.005826407
2010  0.005826406  0.005826407  0.005826407  0.005826406
2011  0.005826407  0.005826407  0.005826407  0.005826407
2012  0.005826407  0.005826407  0.005826407  0.005826407
2013  0.005826407  0.005826407  0.005826407  0.005826407
2014  0.005826407  0.005826407  0.005826407  0.005826407
2015  0.005826407  0.005826407  0.005826407  0.005826407
2016  0.005826407  0.005826407  0.005826407  0.005826407
2017  0.005826407  0.005826407  0.005826407  0.005826407
2018  0.005826407  0.005826407  0.005826407  0.005826407
2019  0.005826407  0.005826407  0.005826407  0.005826407
```

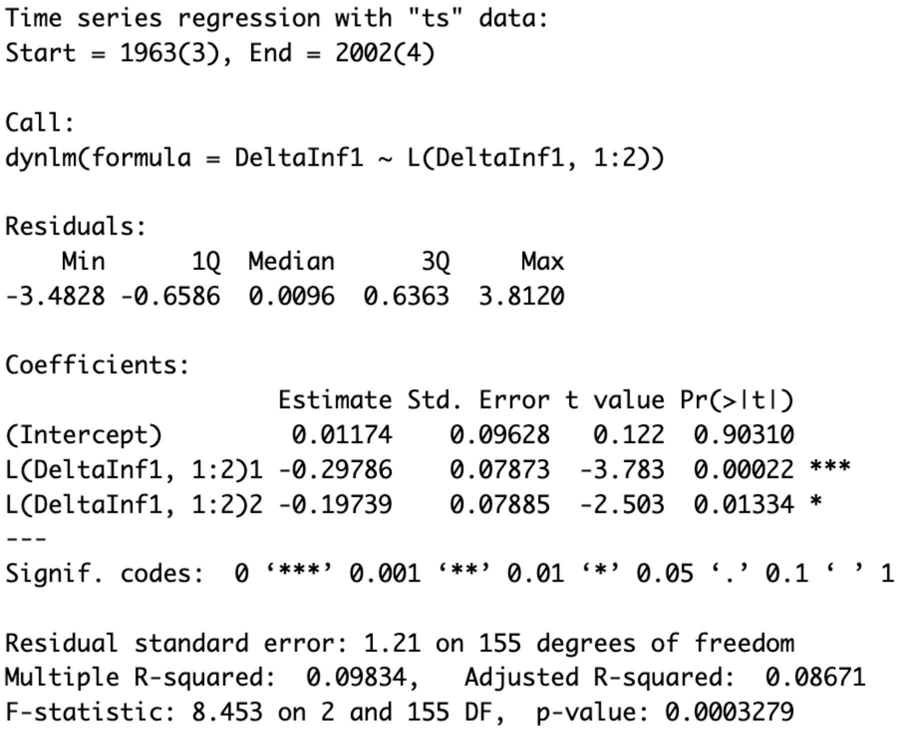
(f) Too many number of periods (68 in this case) into the future for which I am making predictions, destroy forecast accuracy for sure. Since the pseudo out-of-sample (POOS), incorporates the estimation error and requires neither stationarity nor assumptions like homoskedasticity. Thus, the pseudo-out-of-sample forecasts are likely to be biased and undoubtedly inaccurate.

$$Forecast\ error = Y_{T+1} - \hat{Y}_{T+1|T}$$

$$\widehat{MSFE}_{POOS} = \frac{1}{P} \sum_{s=T-P+1}^{T} \hat{u}_s^2$$

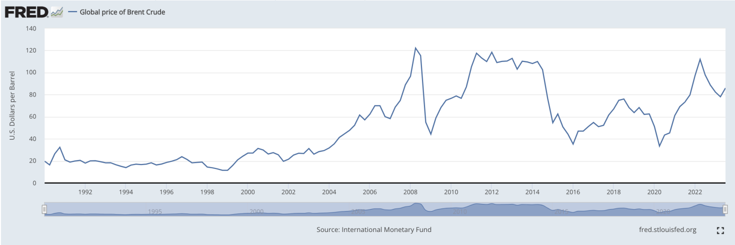
For forecast error, we know what it is and the formula to calculate MSFE. So to answer the question of if the forecast errors have a non-zero mean, we need to take the root mean squared forecast error (RMSFE) which is the square root of the MSFE. Finally, we found that the forecast errors **do have** a non-zero mean.

(g) The root mean squared forecast error (RMSFE) is about 1.995264 compared to the residual standard error 1.21 of the AR(2) model for $\Delta Infl$ estimated over the 1963Q1–2002Q4 sample period. So, it is **not** consistent with the AR(2) model for $\Delta Infl$ estimated over the 1963Q1–2002Q4 sample period.



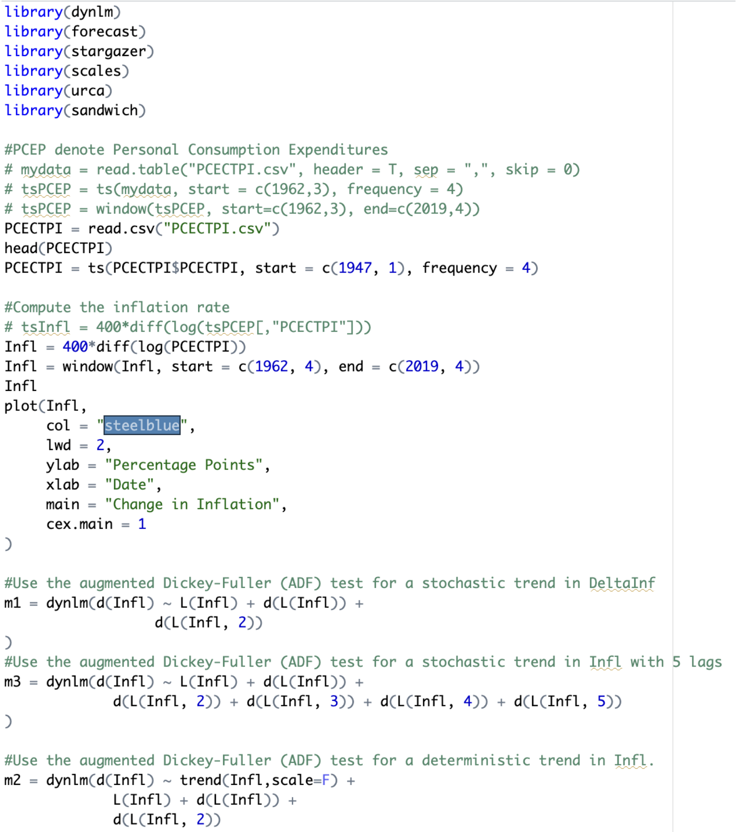
```
Time series regression with "ts" data:
Start = 1963(3), End = 2002(4)

Call:
dynlm(formula = DeltaInf1 ~ L(DeltaInf1, 1:2))

Residuals:
    Min      1Q  Median      3Q     Max 
-3.4828 -0.6586  0.0096  0.6363  3.8120 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)         0.01174    0.09628   0.122  0.90310    
L(DeltaInf1, 1:2)1 -0.29786    0.07873  -3.783  0.00022 ***
L(DeltaInf1, 1:2)2 -0.19739    0.07885  -2.503  0.01334 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.21 on 155 degrees of freedom
Multiple R-squared:  0.09834,	Adjusted R-squared:  0.08671 
F-statistic: 8.453 on 2 and 155 DF,  p-value: 0.0003279
```

(h) Judging from the graph on FRED website, during the first half of 2008, oil prices experienced a surge to a very high level. In mid-2008, the price of Brent crude oil reached a historic high of over $145 per barrel. Within the same year, it also experienced a dramatic price decline, thus providing some contextual background for the subsequent fall of the inflation rate in the 4th quarter of 2008.

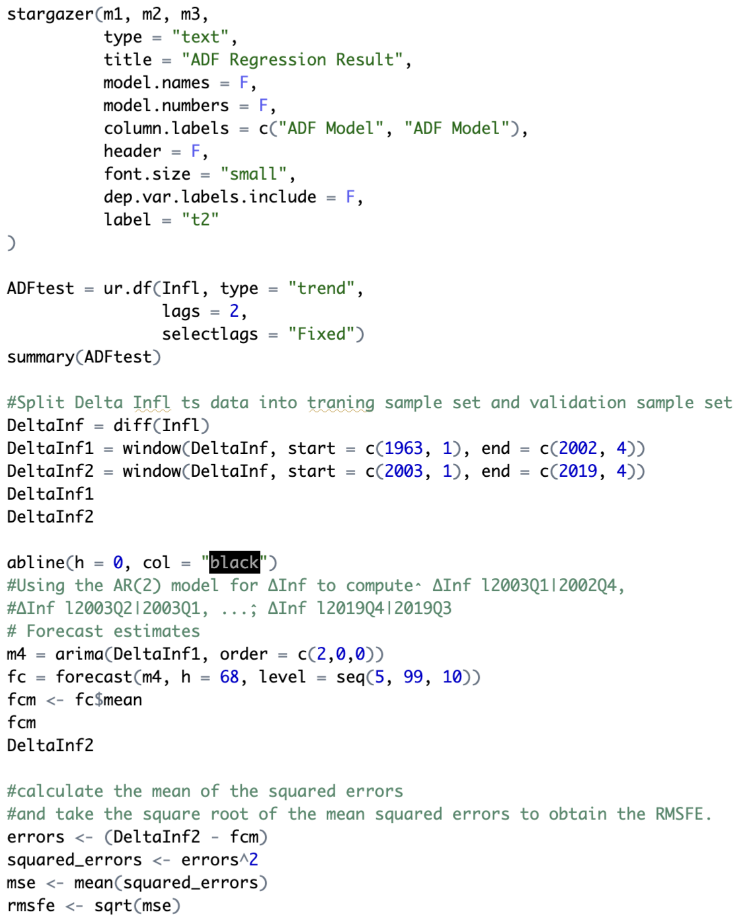


Source: International Monetary Fund

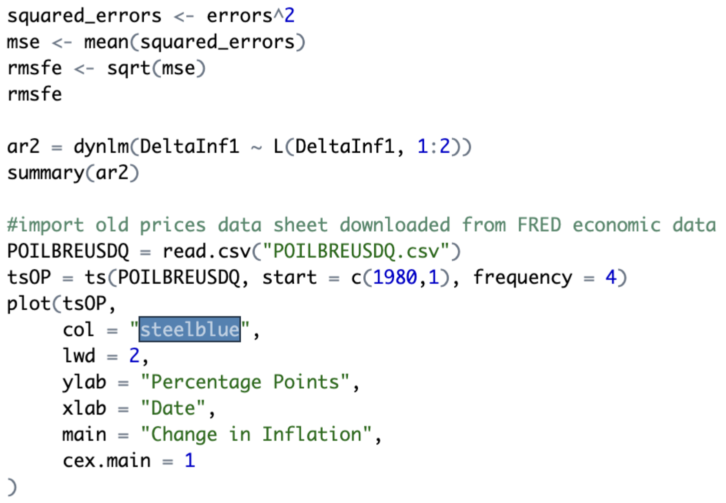
Energy costs are a major component of production expenses. When production costs decrease, it results in deflation. Pressure on businesses to pass on costs to consumers ease; transportation costs go down contributing to lower overall price levels, all of which suffice to be the reasons why we see a big fall of inflation in 2008Q4 in our plot of the U.S inflation rate.



```
library(dynlm)
library(forecast)
library(stargazer)
library(scales)
library(urca)
library(sandwich)

#PCEP denote Personal Consumption Expenditures
# 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))
PCECTPI = read.csv("PCECTPI.csv")
head(PCECTPI)
PCECTPI = ts(PCECTPI$PCECTPI, start = c(1947, 1), frequency = 4)

#Compute the inflation rate
# tsInfl = 400*diff(log(tsPCEP[,"PCECTPI"]))
Infl = 400*diff(log(PCECTPI))
Infl = window(Infl, start = c(1962, 4), end = c(2019, 4))
Infl
plot(Infl,
     col = "steelblue",
     lwd = 2,
     ylab = "Percentage Points",
     xlab = "Date",
     main = "Change in Inflation",
     cex.main = 1
)

#Use the augmented Dickey-Fuller (ADF) test for a stochastic trend in DeltaInf
m1 = dynlm(d(Infl) ~ L(Infl) + d(L(Infl)) +
                  d(L(Infl, 2))
)
#Use the augmented Dickey-Fuller (ADF) test for a stochastic trend in Infl with 5 lags
m3 = dynlm(d(Infl) ~ L(Infl) + d(L(Infl)) +
             d(L(Infl, 2)) + d(L(Infl, 3)) + d(L(Infl, 4)) + d(L(Infl, 5))
)

#Use the augmented Dickey-Fuller (ADF) test for a deterministic trend in Infl.
m2 = dynlm(d(Infl) ~ trend(Infl,scale=F) +
             L(Infl) + d(L(Infl)) +
             d(L(Infl, 2))
```



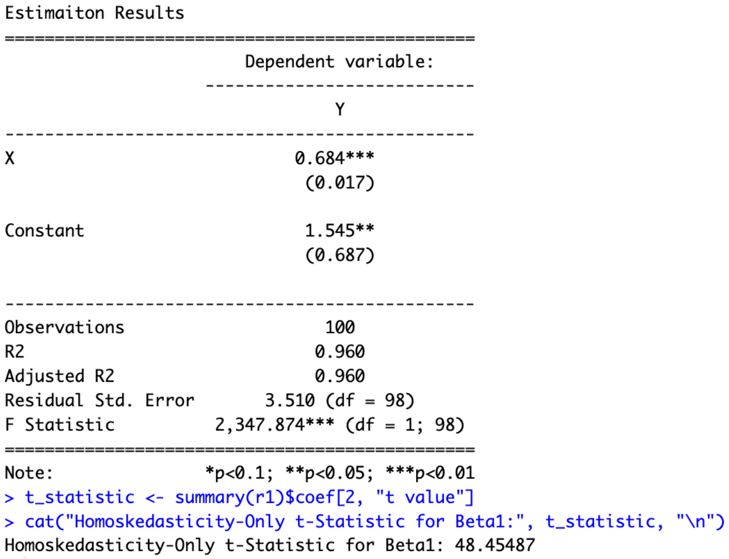
```
stargazer(m1, m2, m3,
          type = "text",
          title = "ADF Regression Result",
          model.names = F,
          model.numbers = F,
          column.labels = c("ADF Model", "ADF Model"),
          header = F,
          font.size = "small",
          dep.var.labels.include = F,
          label = "t2"
)

ADFtest = ur.df(Infl, type = "trend",
                lags = 2,
                selectlags = "Fixed")
summary(ADFtest)

#Split Delta Infl ts data into traning sample set and validation sample set
DeltaInf = diff(Infl)
DeltaInf1 = window(DeltaInf, start = c(1963, 1), end = c(2002, 4))
DeltaInf2 = window(DeltaInf, start = c(2003, 1), end = c(2019, 4))
DeltaInf1
DeltaInf2

abline(h = 0, col = "black")
#Using the AR(2) model for ΔInf to compute^ ΔInf l2003Q1|2002Q4,
#ΔInf l2003Q2|2003Q1, ...; ΔInf l2019Q4|2019Q3
# Forecast estimates
m4 = arima(DeltaInf1, order = c(2,0,0))
fc = forecast(m4, h = 68, level = seq(5, 99, 10))
fcm <- fc$mean
fcm
DeltaInf2

#calculate the mean of the squared errors
#and take the square root of the mean squared errors to obtain the RMSFE.
errors <- (DeltaInf2 - fcm)
squared_errors <- errors^2
mse <- mean(squared_errors)
rmsfe <- sqrt(mse)
```



```
squared_errors <- errors^2
mse <- mean(squared_errors)
rmsfe <- sqrt(mse)
rmsfe

ar2 = dynlm(DeltaInf1 ~ L(DeltaInf1, 1:2))
summary(ar2)

#import old prices data sheet downloaded from FRED economic data
POILBREUSDQ = read.csv("POILBREUSDQ.csv")
tsOP = ts(POILBREUSDQ, start = c(1980,1), frequency = 4)
plot(tsOP,
     col = "steelblue",
     lwd = 2,
     ylab = "Percentage Points",
     xlab = "Date",
     main = "Change in Inflation",
     cex.main = 1
)
```

4 Question 4

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. Independently distributed stochastic data tend to move together over time, even if there is no causal relationship. In this problem, I think we will learn about it and how to discern such a situation.



```
Estimaiton Results
===============================================
                         Dependent variable:
                    ---------------------------
                                  Y
-----------------------------------------------
X                              0.684***
                               (0.017)

Constant                       1.545**
                               (0.687)

-----------------------------------------------
Observations                     100
R2                              0.960
Adjusted R2                     0.960
Residual Std. Error        3.510 (df = 98)
F Statistic          2,347.874*** (df = 1; 98)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
> t_statistic <- summary(r1)$coef[2, "t value"]
> cat("Homoskedasticity-Only t-Statistic for Beta1:", t_statistic, "\n")
Homoskedasticity-Only t-Statistic for Beta1: 48.45487
```

(a) Based on the homoskedasticity-only t-statistic that rejects the null hypothesis of the coefficient on X being zero, while using the usual 5% critical value of 1.96, my conclusion is that $\beta_1 \neq 0$. Moreover, the R^2 of my regression is 0.960.

(b)

$$5th\ percentile\ of\ R^2 : 0.9096226$$

$$50th\ percentile\ of\ R^2 (Median) : 0.9689477$$

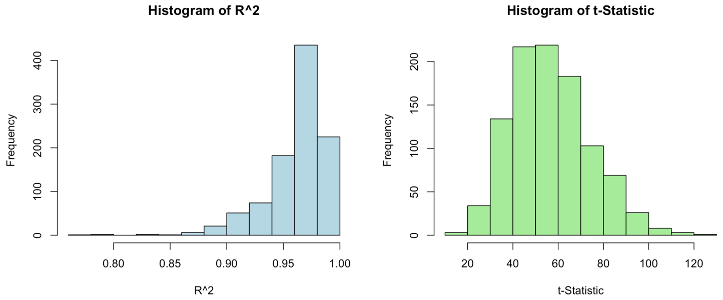
$$95th\ percentile\ of\ R^2 : 0.9873279$$

$$5th\ percentile\ of\ t - Statistic : 31.40616$$

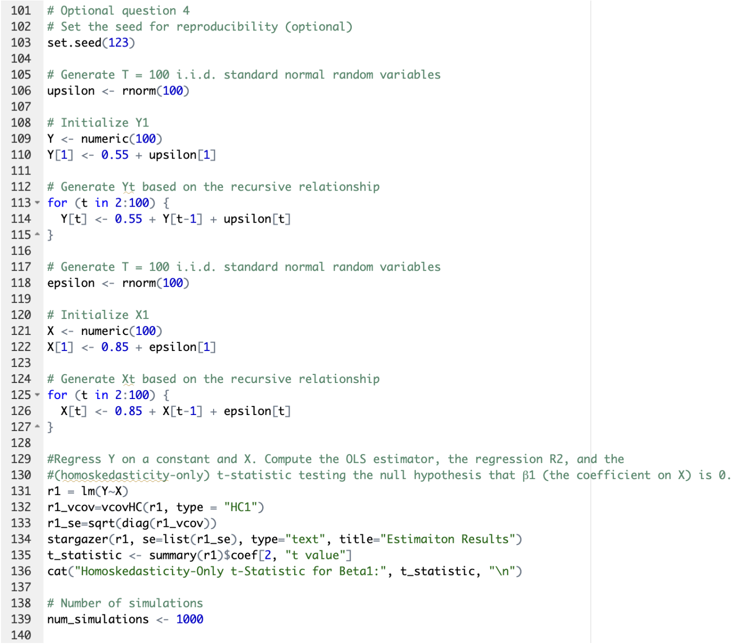
$$50th\ percentile\ of\ t - Statistic\ (Median) : 55.29884$$

$$95th\ percentile\ of\ t - Statistic : 87.38148$$

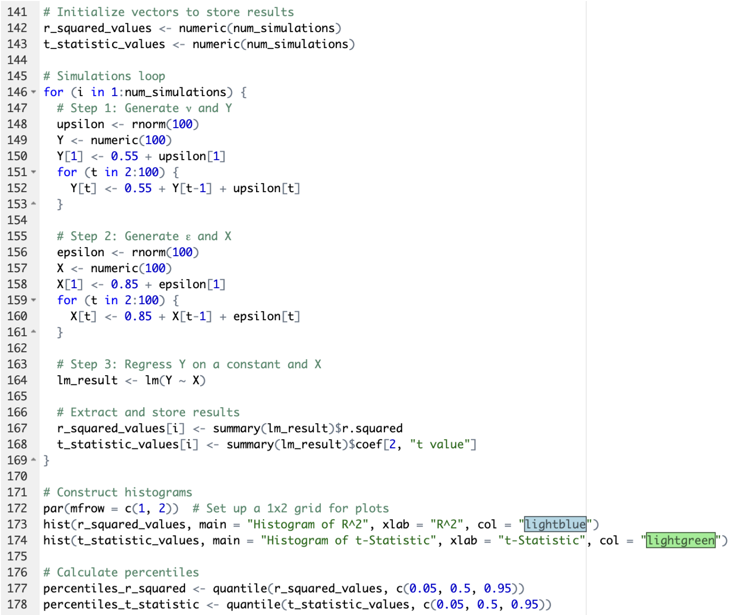
Fraction exceeding 1.96 in absolute value is 100%. This is where I believe my code goes wrong. Looking at the above t-statistics, all of them are too ”large”.



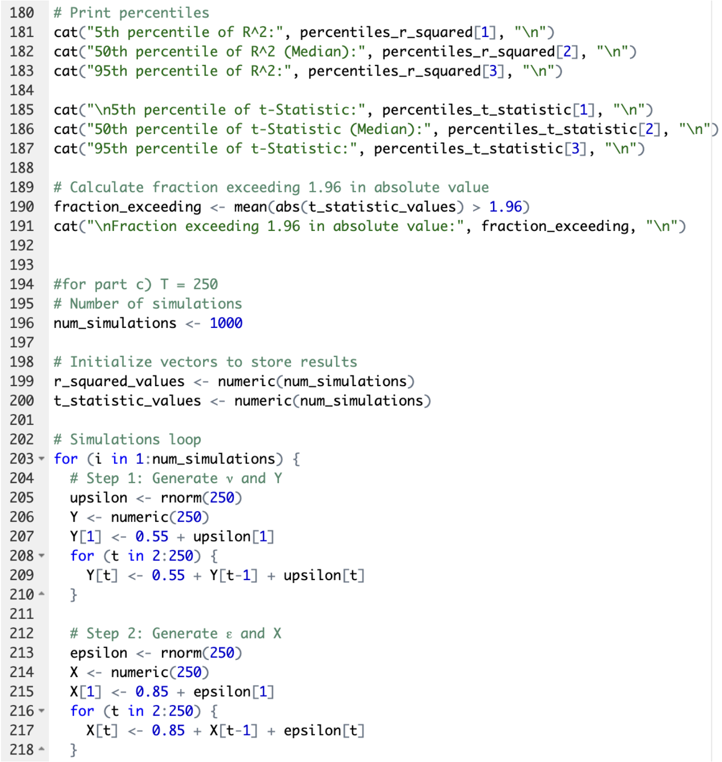
(c) I repeated (b) for T = 250 and T = 500. As the sample size increases, I keep getting larger and larger t-statistics on 5th, 50th, and 95th percentile. However, although I was not able to get at the correct t-statistics, I learned that as sample size increases, the fraction of times that you reject the null hypothesis should increase as well.



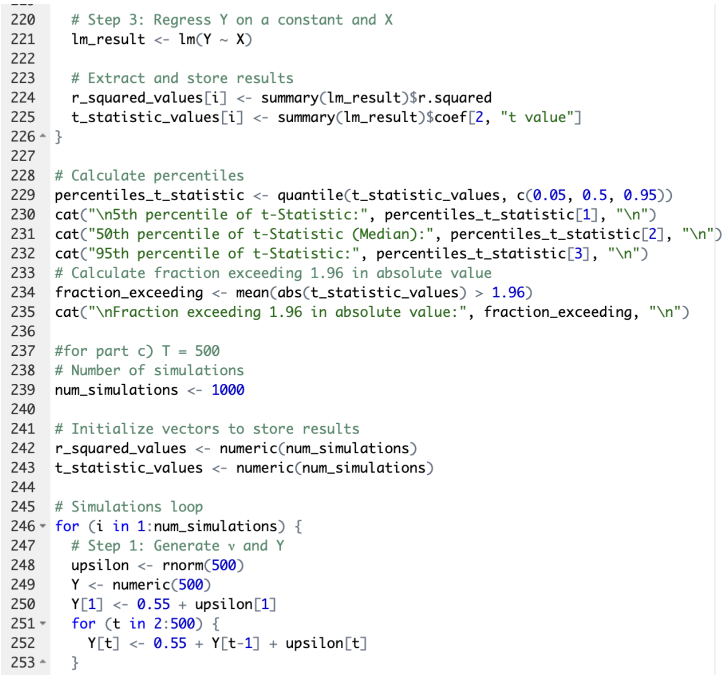
```
# Optional question 4
# Set the seed for reproducibility (optional)
set.seed(123)

# Generate T = 100 i.i.d. standard normal random variables
upsilon <- rnorm(100)

# Initialize Y1
Y <- numeric(100)
Y[1] <- 0.55 + upsilon[1]

# Generate Yt based on the recursive relationship
for (t in 2:100) {
  Y[t] <- 0.55 + Y[t-1] + upsilon[t]
}

# Generate T = 100 i.i.d. standard normal random variables
epsilon <- rnorm(100)

# Initialize X1
X <- numeric(100)
X[1] <- 0.85 + epsilon[1]

# Generate Xt based on the recursive relationship
for (t in 2:100) {
  X[t] <- 0.85 + X[t-1] + epsilon[t]
}

#Regress Y on a constant and X. Compute the OLS estimator, the regression R2, and the
#(homoskedasticity-only) t-statistic testing the null hypothesis that β1 (the coefficient on X) is 0.
r1 = lm(Y~X)
r1_vcov=vcovHC(r1, type = "HC1")
r1_se=sqrt(diag(r1_vcov))
stargazer(r1, se=list(r1_se), type="text", title="Estimaiton Results")
t_statistic <- summary(r1)$coef[2, "t value"]
cat("Homoskedasticity-Only t-Statistic for Beta1:", t_statistic, "\n")

# Number of simulations
num_simulations <- 1000

```



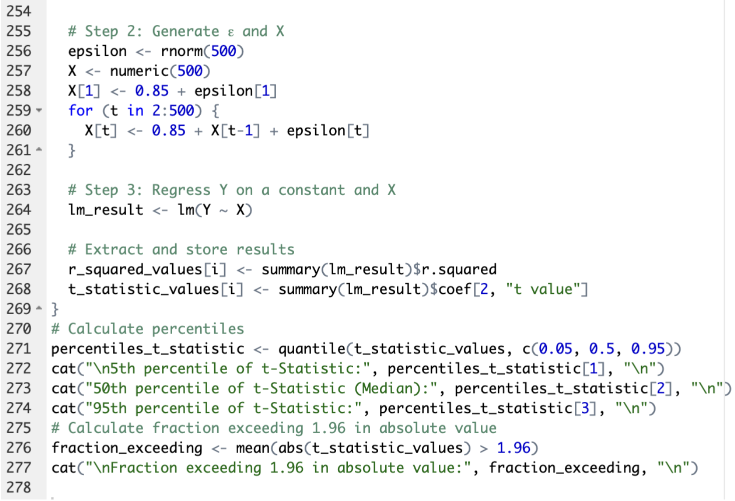
```r
# Initialize vectors to store results
r_squared_values <- numeric(num_simulations)
t_statistic_values <- numeric(num_simulations)

# Simulations loop
for (i in 1:num_simulations) {
  # Step 1: Generate ν and Y
  upsilon <- rnorm(100)
  Y <- numeric(100)
  Y[1] <- 0.55 + upsilon[1]
  for (t in 2:100) {
    Y[t] <- 0.55 + Y[t-1] + upsilon[t]
  }

  # Step 2: Generate ε and X
  epsilon <- rnorm(100)
  X <- numeric(100)
  X[1] <- 0.85 + epsilon[1]
  for (t in 2:100) {
    X[t] <- 0.85 + X[t-1] + epsilon[t]
  }

  # Step 3: Regress Y on a constant and X
  lm_result <- lm(Y ~ X)

  # Extract and store results
  r_squared_values[i] <- summary(lm_result)$r.squared
  t_statistic_values[i] <- summary(lm_result)$coef[2, "t value"]
}

# Construct histograms
par(mfrow = c(1, 2))  # Set up a 1x2 grid for plots
hist(r_squared_values, main = "Histogram of R^2", xlab = "R^2", col = "lightblue")
hist(t_statistic_values, main = "Histogram of t-Statistic", xlab = "t-Statistic", col = "lightgreen")

# Calculate percentiles
percentiles_r_squared <- quantile(r_squared_values, c(0.05, 0.5, 0.95))
percentiles_t_statistic <- quantile(t_statistic_values, c(0.05, 0.5, 0.95))
```



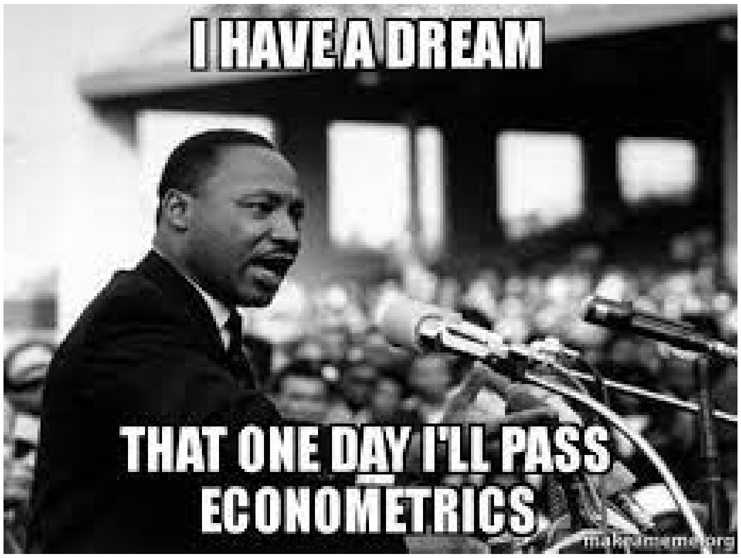
```r
# Print percentiles
cat("5th percentile of R^2:", percentiles_r_squared[1], "\n")
cat("50th percentile of R^2 (Median):", percentiles_r_squared[2], "\n")
cat("95th percentile of R^2:", percentiles_r_squared[3], "\n")

cat("\n5th percentile of t-Statistic:", percentiles_t_statistic[1], "\n")
cat("50th percentile of t-Statistic (Median):", percentiles_t_statistic[2], "\n")
cat("95th percentile of t-Statistic:", percentiles_t_statistic[3], "\n")

# Calculate fraction exceeding 1.96 in absolute value
fraction_exceeding <- mean(abs(t_statistic_values) > 1.96)
cat("\nFraction exceeding 1.96 in absolute value:", fraction_exceeding, "\n")


#for part c) T = 250
# Number of simulations
num_simulations <- 1000

# Initialize vectors to store results
r_squared_values <- numeric(num_simulations)
t_statistic_values <- numeric(num_simulations)

# Simulations loop
for (i in 1:num_simulations) {
  # Step 1: Generate ν and Y
  upsilon <- rnorm(250)
  Y <- numeric(250)
  Y[1] <- 0.55 + upsilon[1]
  for (t in 2:250) {
    Y[t] <- 0.55 + Y[t-1] + upsilon[t]
  }

  # Step 2: Generate ε and X
  epsilon <- rnorm(250)
  X <- numeric(250)
  X[1] <- 0.85 + epsilon[1]
  for (t in 2:250) {
    X[t] <- 0.85 + X[t-1] + epsilon[t]
  }
```



```r
  # Step 3: Regress Y on a constant and X
  lm_result <- lm(Y ~ X)

  # Extract and store results
  r_squared_values[i] <- summary(lm_result)$r.squared
  t_statistic_values[i] <- summary(lm_result)$coef[2, "t value"]
}

# Calculate percentiles
percentiles_t_statistic <- quantile(t_statistic_values, c(0.05, 0.5, 0.95))
cat("\n5th percentile of t-Statistic:", percentiles_t_statistic[1], "\n")
cat("50th percentile of t-Statistic (Median):", percentiles_t_statistic[2], "\n")
cat("95th percentile of t-Statistic:", percentiles_t_statistic[3], "\n")
# Calculate fraction exceeding 1.96 in absolute value
fraction_exceeding <- mean(abs(t_statistic_values) > 1.96)
cat("\nFraction exceeding 1.96 in absolute value:", fraction_exceeding, "\n")

#for part c) T = 500
# Number of simulations
num_simulations <- 1000

# Initialize vectors to store results
r_squared_values <- numeric(num_simulations)
t_statistic_values <- numeric(num_simulations)

# Simulations loop
for (i in 1:num_simulations) {
  # Step 1: Generate ν and Y
  upsilon <- rnorm(500)
  Y <- numeric(500)
  Y[1] <- 0.55 + upsilon[1]
  for (t in 2:500) {
    Y[t] <- 0.55 + Y[t-1] + upsilon[t]
  }
```



```r

  # Step 2: Generate ε and X
  epsilon <- rnorm(500)
  X <- numeric(500)
  X[1] <- 0.85 + epsilon[1]
  for (t in 2:500) {
    X[t] <- 0.85 + X[t-1] + epsilon[t]
  }

  # Step 3: Regress Y on a constant and X
  lm_result <- lm(Y ~ X)

  # Extract and store results
  r_squared_values[i] <- summary(lm_result)$r.squared
  t_statistic_values[i] <- summary(lm_result)$coef[2, "t value"]
}
# Calculate percentiles
percentiles_t_statistic <- quantile(t_statistic_values, c(0.05, 0.5, 0.95))
cat("\n5th percentile of t-Statistic:", percentiles_t_statistic[1], "\n")
cat("50th percentile of t-Statistic (Median):", percentiles_t_statistic[2], "\n")
cat("95th percentile of t-Statistic:", percentiles_t_statistic[3], "\n")
# Calculate fraction exceeding 1.96 in absolute value
fraction_exceeding <- mean(abs(t_statistic_values) > 1.96)
cat("\nFraction exceeding 1.96 in absolute value:", fraction_exceeding, "\n")

```


I HAVE A DREAM
THAT ONE DAY I'LL PASS ECONOMETRICS
makeameme.org