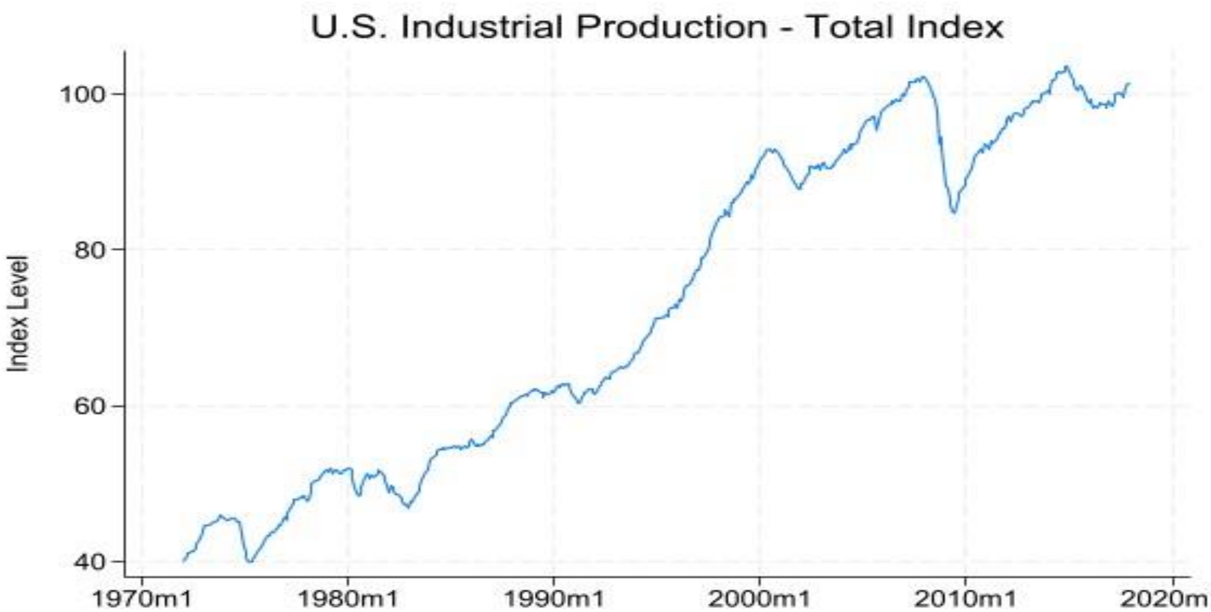
Group: Jayesh Jain, Rishi Kamath, Ethan Pinto

FINANCIAL ECONOMETRICS

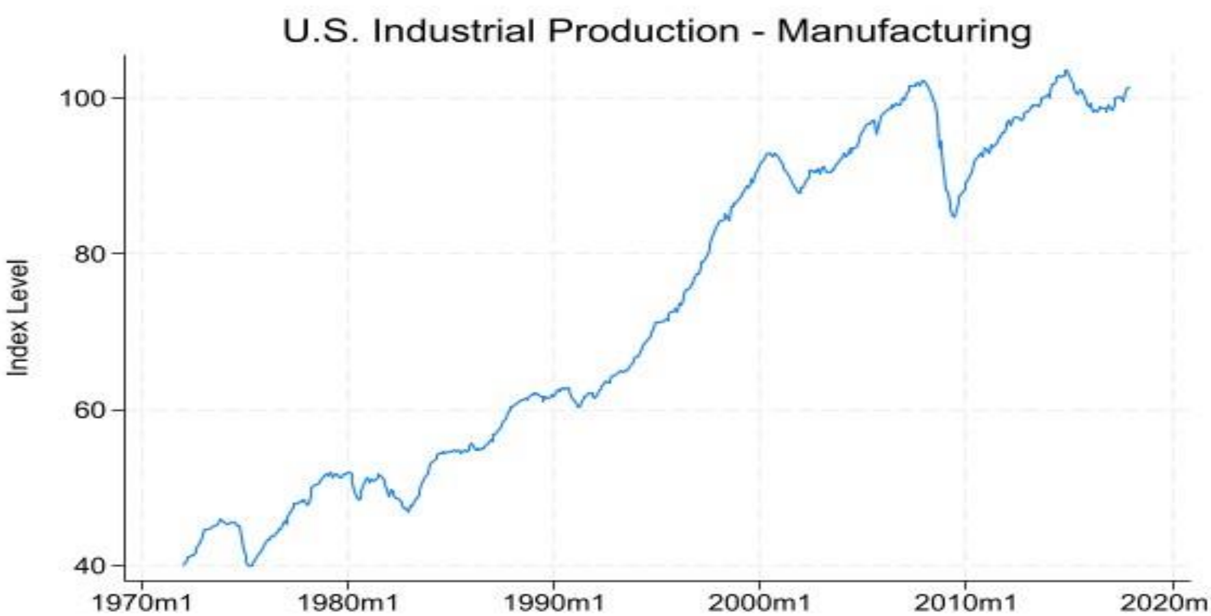
Problem Set-1

Q1.

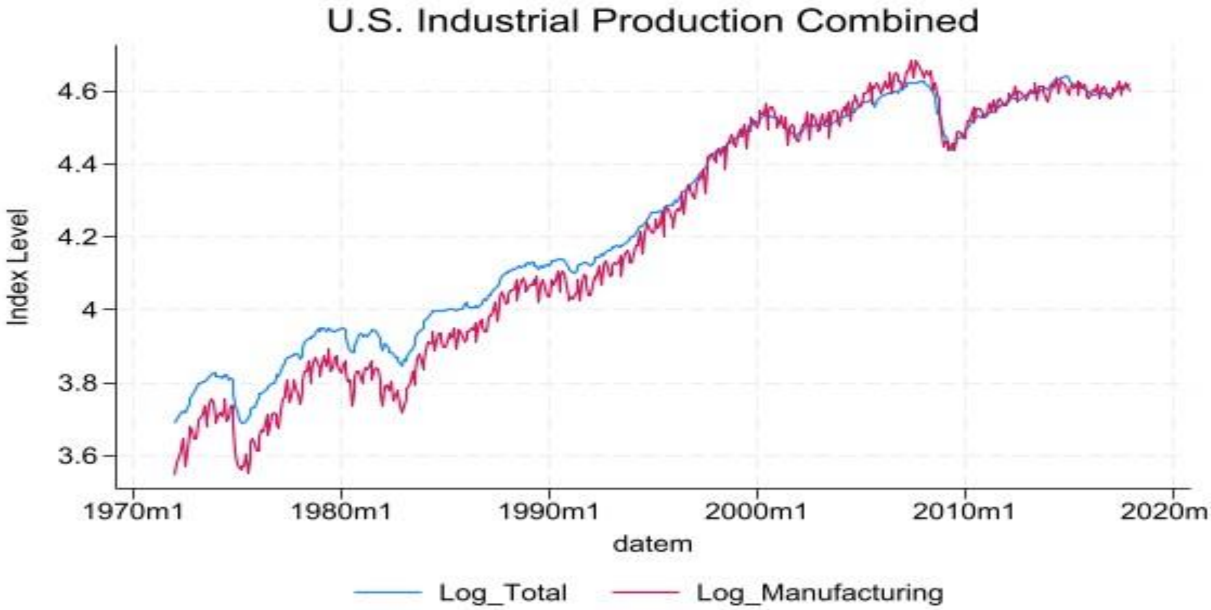
a.



b.



c.


U.S. Industrial Production Combined
Index Level
4.6
4.4
4.2
4
3.8
3.6
1970m1
1980m1
1990m1
2000m1
2010m1
2020m1
datem
Log_Total
Log_Manufacturing

Q2.

a.

```
. summarize, detail

                              r(m1)
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.336852      -1.550657
 5%    -1.231915      -1.502916
10%    -1.178755      -1.479418       Obs               5,000
25%     -1.09399      -1.464529       Sum of wgt.       5,000

50%    -1.000471                      Mean          -1.000328
                        Largest       Std. dev.      .1405898
75%    -.9061102      -.5926968
90%    -.8186465      -.5832513       Variance       .0197655
95%    -.7702168      -.5598732       Skewness      -.0640003
99%    -.6814846      -.4621855       Kurtosis       3.002239
```

Sample Variance = 0.0197655

Sample Bias: $E[\hat{\beta}] - \beta = -1.000328 + 1 = -0.000328$

b.

```
. summarize, detail

                            r(median)
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.409499      -1.637296
 5%    -1.292199      -1.629403
10%    -1.227982      -1.557177       Obs               5,000
25%    -1.122236      -1.540094       Sum of wgt.       5,000

50%    -1.003142                      Mean          -1.002358
                        Largest       Std. dev.      .1755834
75%    -.8770337      -.4682474
90%    -.7797345      -.4534649       Variance       .0308295
95%    -.7179845      -.4252262       Skewness       -.024258
99%    -.5962797      -.3268192       Kurtosis       2.899953
```

Sample Variance = 0.0308295

Sample Bias: $E[\hat{\beta}] - \beta = -1.002358 + 1 = -0.002358$

Clearly, the mean has both a lower variance and bias than the median, so we should use the mean as an estimator.

c.

```
. summarize, detail

                           r(mean)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .6109588        .500175
 5%     .6896025       .5037712
10%     .7376716       .5083175       Obs               5,000
25%     .8345756        .509377       Sum of wgt.       5,000

50%     .9593815                      Mean            1.00567
                        Largest       Std. dev.      .2859106
75%     1.118677       4.254641
90%     1.310578       4.340088       Variance       .0817449
95%     1.454491       6.150612       Skewness       5.402133
99%     1.824515       7.498726       Kurtosis       88.94567
```

Sample Variance = 0.0817449
Sample Bias: $E[\hat{\beta}] - \beta = 1.00567 - 1 = 0.00567$

```
. summarize, detail

                          r(median)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .2442765       .1945058
 5%     .2748112       .1960465
10%     .2930143       .2107337       Obs               5,000
25%     .3256424       .2144704       Sum of wgt.       5,000

50%     .3668547                      Mean           .3728189
                        Largest       Std. dev.      .0657745
75%     .4162321       .6261464
90%     .4587524       .6359828       Variance       .0043263
95%     .4878296       .6536595       Skewness       .4858539
99%     .5509161       .7216309       Kurtosis       3.354239

.
```

Sample Variance = 0.0043263
Sample Bias: $E[\hat{\beta}] - \beta = .3728189 - 1 = -0.6271811$

In this case, the variance is lower for the median estimator, but the bias is substantially higher. It is not exactly clear which estimator would be better because both low bias and low variance are desirable features of a good estimator. Low variance implies that, on repeated sampling, the estimate will not change much while low bias is important too. Thus, it is unclear which estimator to use.

Q3.

a. The findings we obtained align with the values documented in the article.

```
regress var1 L(1).var1 if tin(1953m4, 2013m12)

      Source |       SS           df       MS      Number of obs   =       729
-------------+----------------------------------   F(1, 727)       >  99999.00
       Model |   181189993         1   181189993   Prob > F        =    0.0000
    Residual |  453508.784       727  623.808506   R-squared       =    0.9975
-------------+----------------------------------   Adj R-squared   =    0.9975
       Total |   181643502       728  249510.305   Root MSE        =    24.976

------------------------------------------------------------------------------
        var1 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        var1 |
         L1. |   1.003282   .0018616   538.94   0.000     .9996273    1.006937
             |
       _cons |   .9466052   1.255869     0.75   0.451    -1.518957    3.412167
------------------------------------------------------------------------------
```

```
regress D(1).var1 L(1).var1 if tin(1953m4, 2013m12)

      Source |       SS           df       MS      Number of obs   =       729
-------------+----------------------------------   F(1, 727)       =      3.11
       Model |  1938.99747         1  1938.99747   Prob > F        =    0.0783
    Residual |  453508.784       727  623.808506   R-squared       =    0.0043
-------------+----------------------------------   Adj R-squared   =    0.0029
       Total |  455447.781       728  625.615084   Root MSE        =    24.976

------------------------------------------------------------------------------
      D.var1 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        var1 |
         L1. |    .003282   .0018616     1.76   0.078    -.0003727    .0069368
             |
       _cons |   .9466052   1.255869     0.75   0.451    -1.518957    3.412167
------------------------------------------------------------------------------
```

b. The findings we obtained align with the values documented in the article.

```
regress GS10 L(1/3).GS10 if tin(1953m4, 2013m12)
```

Source	SS	df	MS
Model	5526.82921	3	1842.2764
Residual	46.9884088	722	.065080899
Total	5573.81762	725	7.6880243

Number of obs	=	726
F(3, 722)	=	28307.48
Prob > F	=	0.0000
R-squared	=	0.9916
Adj R-squared	=	0.9915
Root MSE	=	.25511

GS10	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
GS10						
L1.	1.370355	.0363631	37.69	0.000	1.298965	1.441745
L2.	-.5889166	.0586691	-10.04	0.000	-.704099	-.4737343
L3.	.2135858	.0363561	5.87	0.000	.1422096	.284962
_cons	.0302642	.023024	1.31	0.189	-.0149378	.0754662

```
regress D(1).GS10 L(1/3).GS10 if tin(1953m3, 2013m12)
```

Source	SS	df	MS
Model	7.62063045	3	2.54021015
Residual	46.9884088	722	.065080899
Total	54.6090393	725	.075322813

Number of obs	=	726
F(3, 722)	=	39.03
Prob > F	=	0.0000
R-squared	=	0.1395
Adj R-squared	=	0.1360
Root MSE	=	.25511

D.GS10	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
GS10						
L1.	.3703547	.0363631	10.18	0.000	.2989648	.4417447
L2.	-.5889166	.0586691	-10.04	0.000	-.704099	-.4737343
L3.	.2135858	.0363561	5.87	0.000	.1422096	.284962
_cons	.0302642	.023024	1.31	0.189	-.0149378	.0754662

c.

```
summarize
```

Variable	Obs	Mean	Std. dev.	Min	Max
r2	1,000	.0098684	.0130894	7.45e-09	.0871385
r2diff	1,000	.5067238	.0495061	.3497049	.6443531

The first R-squared value, which looks at the relationship between a variable "y" and its past value (lag), is extremely low, around 0.99%. This low R-squared suggests that knowing the past value of "y" doesn't help much in predicting its current value because "y" is essentially random with a mean of 0 and a variance of 1. In simpler terms, there's no useful information in past values of "y" to explain its current value.

On the other hand, when we examine the regression of the difference between "y" and its lagged value against the lagged value itself, the R-squared is high, approximately 50%. This means that there's a strong relationship between the change in "y" from one period to the next and its lagged value. In other words, the past value of "y" does contain important information for predicting how "y" will change over time.

STATA CODE

Q1.

a. clear all
   ```
   freduse INDPRO
   gen datem = mofd(daten)
   format datem %tm
   tsset datem
   tsline INDPRO, xtitle("") ytitle("Index level") title("U.S. Industrial Production - Total Index"), if tin(1972m1,2017m12)
   ```

b. clear all
   ```
   freduse IPGMFN
   gen datem = mofd(daten)
   format datem %tm
   tsset datem
   tsline IPGMFN, xtitle("") ytitle("Index level") title("U.S. Industrial Production - Manufacturing"), if tin(1972m1,2017m12)
   ```

c. clear all
   ```
   freduse INDPRO IPGMFN
   gen datem = mofd(daten)
   format datem %tm
   tsset datem
   gen Log_Total = ln(INDPRO)
   gen Log_Manufacturing = ln(IPGMFN)
   tsline INDPRO, xtitle("") ytitle("Index level") title("U.S. Industrial Production - Total Index"), if tin(1972m1,2017m12)
   tsline IPGMFN, xtitle("") ytitle("Index level") title("U.S. Industrial Production - Manufacturing"), if tin(1972m1,2017m12)
   tsline Log_Total Log_Manufacturing,ytitle("Index Level") title("U.S. Industrial Production Combined"), if tin(1972m1,2017m12)
   ```

Q2.

a.
```
clear all
program findmean, rclass
   drop _all
   set obs 100
   gen x = rnormal(-1, sqrt(2))
   quietly summarize x, detail
   return scalar m1 = r(mean)
end

findmean
display r(m1)
set seed 6219
simulate xmean=r(m1), reps(5000): findmean
summarize, detail
```

b.
```
clear all
program findmedian, rclass
   drop _all
   set obs 100
   gen x = rnormal(-1, sqrt(2))
   quietly summarize x, detail
   return scalar median = r(p50)
end

findmedian
display r(median)
set seed 6219
simulate xmedian=r(median), reps(5000): findmedian
summarize, detail
```

c.
```
clear all
program findMeanForExp, rclass
   drop _all
   set obs 100
   gen x = rnormal(-1, sqrt(2))
   gen y = exp(x)
   quietly summarize y, detail
   return scalar mean = r(mean)
end

set seed 6219
simulate ymean=r(mean), reps(5000): findMeanForExp
summarize, detail
```

```
clear all
program findMedianForExp, rclass
   drop _all
   set obs 100
   gen x = rnormal(-1, sqrt(2))
   gen y = exp(x)
   quietly summarize y, detail
   return scalar median = r(p50)
end

set seed 6219
simulate ymedian=r(median), reps(5000): findMedianForExp
summarize, detail
```

Q3.

a.
```
clear all
use SPShiller.dta
gen time=tm(1871m1)+_n-1
format time %tm
tsset time
describe
regress var1 L.var1 if tin(1953m4, 2013m12)
regress D.var1 L.var1 if tin(1953m4, 2013m12)
```

b.
```
clear all
freduse GS10
describe
tsmktim time, start(1953m4)
tsset time
describe
regress GS10 L(1/3).GS10 if tin(1953m4, 2013m12)
regress D.GS10 L(1/3).GS10 if tin(1953m4, 2013m12)
```

c.
```
program findrsq, rclass
  drop _all
  set obs 100
  gen time= _n
  tsset time
  gen y = rnormal(0,1)
  quietly regress y L.y
  return scalar rsq = e(r2)
  quietly regress D.y L.y
  return scalar rsqdiff = e(r2)
end

set seed 6219
simulate r2=r(rsq) r2diff=r(rsqdiff), reps(1000) nodots: findrsq
summarize
```