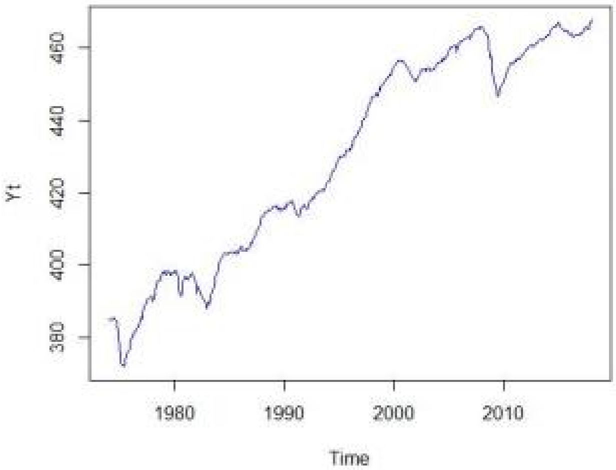
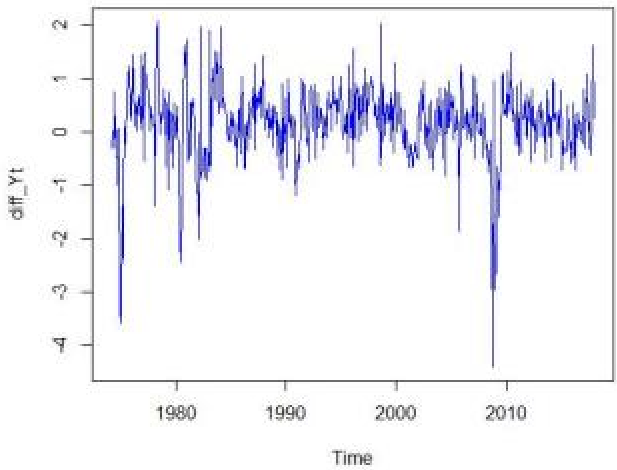
1.

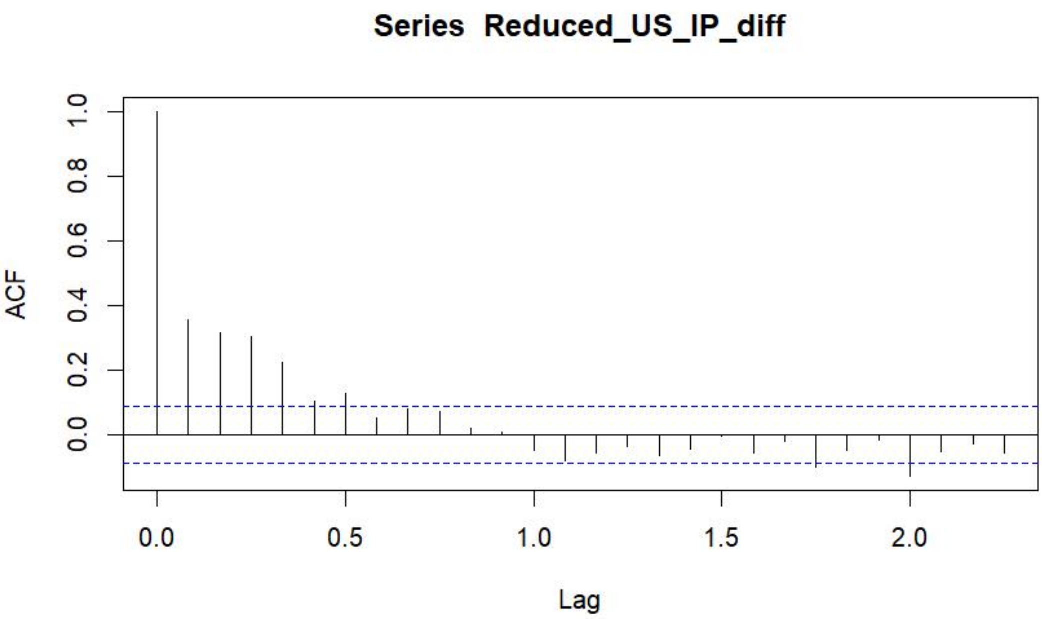


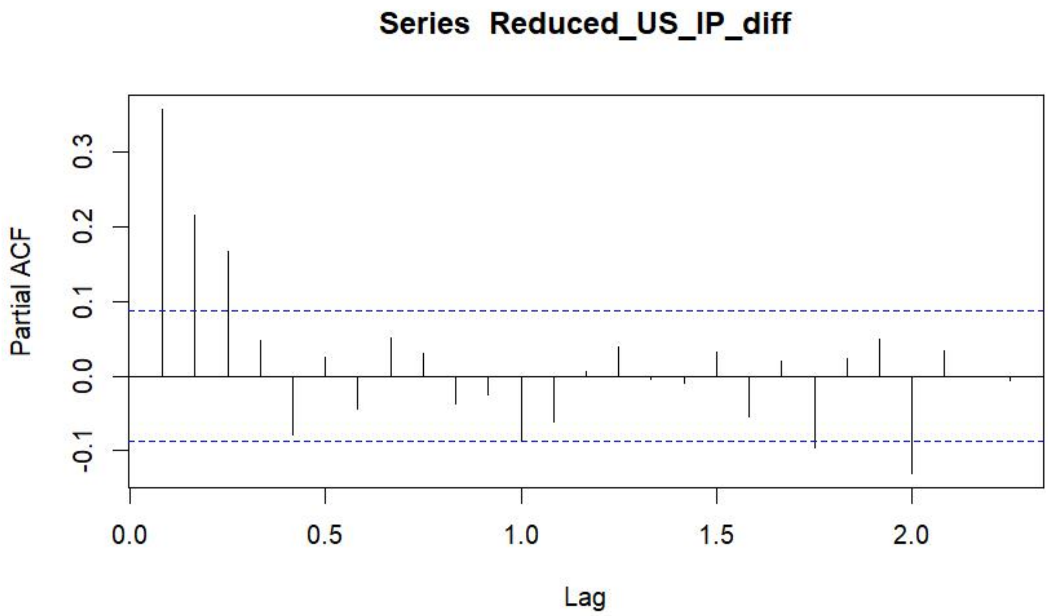


Do the series appear stationary? If not what might be the reason ?

The lnYt looks non-stationary, it seems have the upperward trend and the lnYt in differences l ooks stationary with a 0 mean.

2.





Suggest possible orders for an ARMA model

It can be AR(3), for the PACF cut at the 3 lag and ACF can be deemed as gradually dies out.

3.

```
> print(AIC_Reduced_US_IP_diff)
```

	MA0	MA1	MA2	MA3	MA4	MA5	MA6
AR0	1114.817	1069.108	1053.142	1032.096	1014.635	1016.116	1014.608
AR1	1048.176	1016.530	1016.575	1014.416	1015.311	1015.372	1015.937
AR2	1026.427	1017.152	1019.745	1016.292	1014.723	1015.533	1017.533
AR3	1014.231	1015.648	1015.080	1010.326	1012.147	1012.506	1014.408
AR4	1015.104	1013.727	1014.806	1016.521	1013.356	1013.145	1016.067
AR5	1013.962	1014.871	1016.809	1015.781	1014.114	1006.368	1008.265
AR6	1015.656	1016.864	1015.865	1009.725	1012.422	1006.503	1009.759

```
> print(BIC_Reduced_US_IP_diff)
```

	MA0	MA1	MA2	MA3	MA4	MA5	MA6
AR0	1123.262	1081.776	1070.033	1053.209	1039.970	1045.674	1048.388
AR1	1060.844	1033.420	1037.688	1039.751	1044.869	1049.153	1053.941
AR2	1043.318	1038.265	1045.080	1045.850	1048.504	1053.536	1059.759
AR3	1035.344	1040.983	1044.638	1044.106	1050.151	1054.732	1060.856
AR4	1040.440	1043.285	1048.586	1054.524	1055.582	1059.594	1066.738
AR5	1043.520	1048.652	1054.812	1058.007	1060.562	1057.039	1063.158
AR6	1049.436	1054.868	1058.091	1056.173	1063.093	1061.397	1068.875

```
> LR_test_ARMA1_1_ARMA5_5 <- lrtest(ARMA1_1, ARMA5_5)> print(L
R_test_ARMA1_1_ARMA5_5)

                          Likelihood ratio test

Model 1: arima(x = Reduced_US_IP_diff, order = c(1, 0, 1))
Model 2: arima(x = Reduced_US_IP_diff, order = c(5, 0, 5))
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   4 -504.27
2  12 -491.18  8 26.162  0.0009852 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

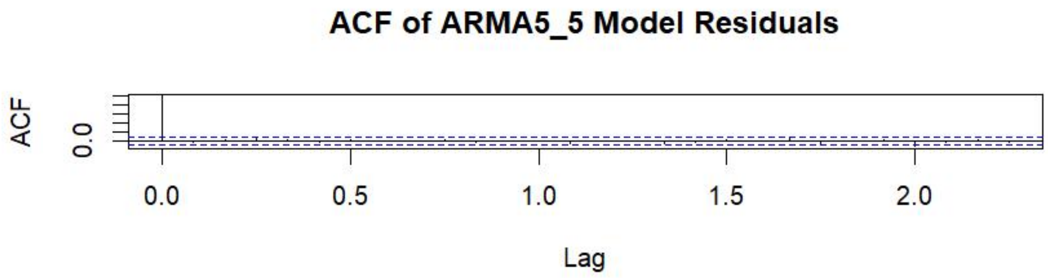
```
> LR_test_AR3_ARMA5_5 <- lrtest(AR3, ARMA5_5)> print(LR_test_A
R3_ARMA5_5)

                          Likelihood ratio test

Model 1: arima(x = Reduced_US_IP_diff, order = c(3, 0, 0))
Model 2: arima(x = Reduced_US_IP_diff, order = c(5, 0, 5))
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -502.12
2  12 -491.18  7 21.863   0.002682 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

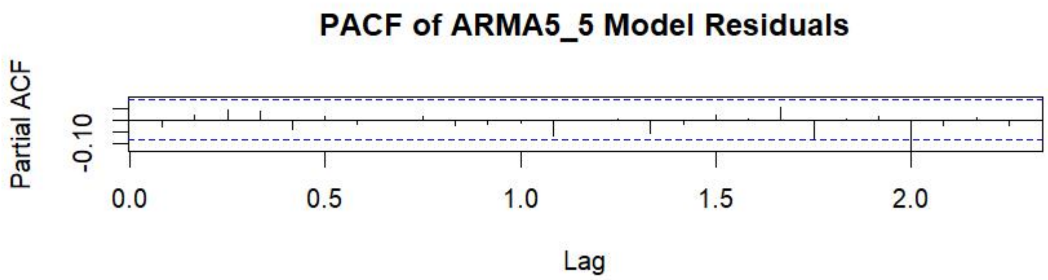
Compare with what you expected in the previous ques tion

From the AIC table, the smallest one is ARMA(5,5), and from the BIC table, the smallest one is ARMA(1,1), as the BIC embodies much stiffer penalty of losing the degrees of freedom.

Then for the diagnostic checking of the joint significance for the added parameters by LR-test. The added parameters is significance and the ARMA(5,5) is chosen.

4.






```
> Qstatistic_arma5_5=LjungBox(ARMA5_5$residuals, 30, 10)> prin
t(Qstatistic_arma5_5)
```

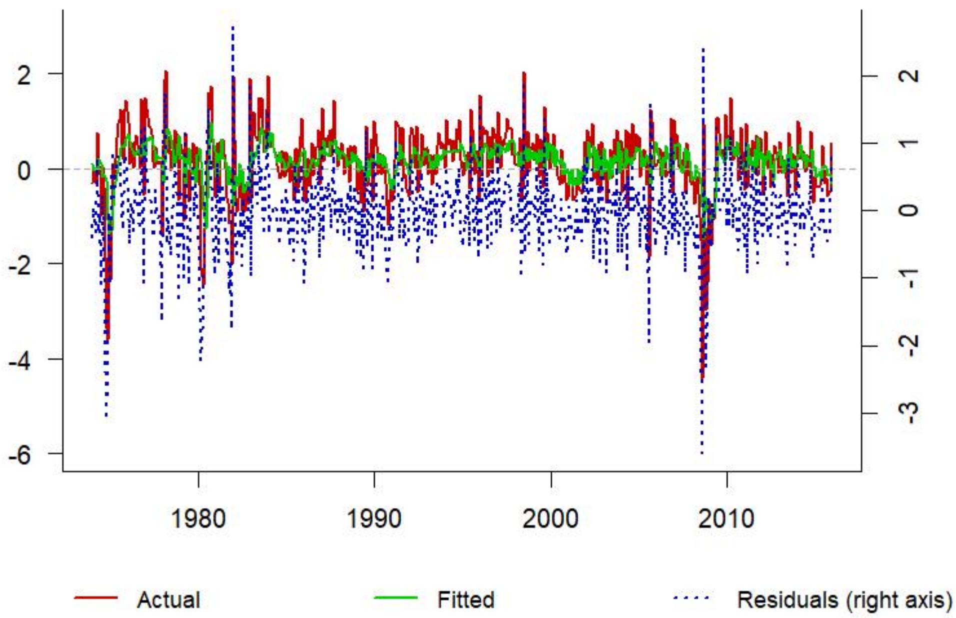
	lag	Ljung-Box	p-value
1	1	0.4217	NA
2	2	0.6513	NA
3	3	1.5811	NA
4	4	2.1243	NA
5	5	3.0725	NA
6	6	3.3440	NA
7	7	3.5036	NA
8	8	3.5060	NA
9	9	3.6296	NA
10	10	3.9730	NA
11	11	4.1298	0.9660
12	12	4.2110	0.9793
13	13	6.7396	0.9151
14	14	6.7428	0.9442
15	15	6.7434	0.9644
16	16	8.9598	0.9151
17	17	9.1786	0.9345
18	18	9.5439	0.9458
19	19	9.5701	0.9628
20	20	10.8149	0.9509
21	21	14.0596	0.8670
22	22	14.0888	0.8983
23	23	14.2145	0.9206
24	24	22.3035	0.5612
25	25	22.6613	0.5973
26	26	22.8723	0.6402
27	27	23.4448	0.6609
28	28	23.7326	0.6956
29	29	26.9411	0.5749
30	30	27.0550	0.6204

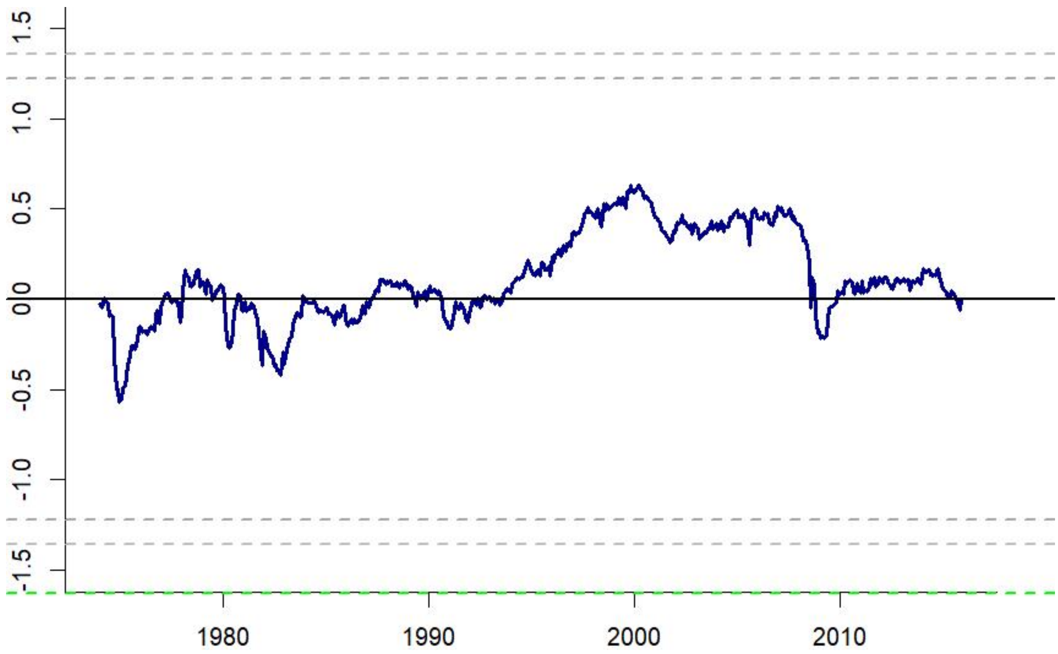
Are the errors white noise ? what is the implication if they are not ?

From the Autoregressive function and partial autoregressive function of the residuals, except l ag=24, the left are all in the confidence bound. and the lag=24 is out of confidence bounce be cause of error. and then from the ljung box test, it accept the H0 of there is no autocorrelation in the residuals till lag=30. so the error term now is white noise.

If the error term is not the white noise, then it means there are still dynamic patterns left in the error term the model didnt capture. and this persistence in the error term may cause your esti mations inconsistent. So it need to extend the model by adding lags to capture the left dynami cs.

5.

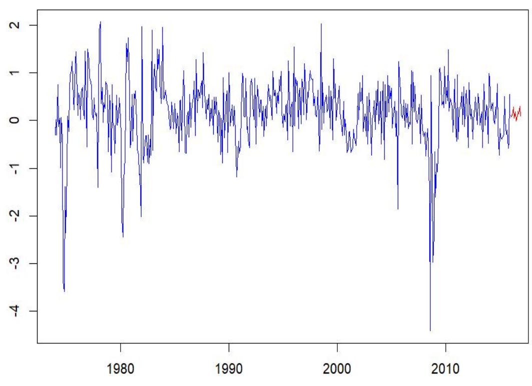


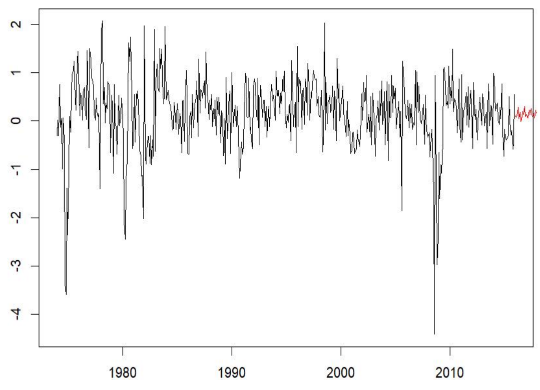


Is your preferred ARMA model stable over time ?

Yes, the cumulated sum of standardized error term(CUSUM) is under the confidence bound, so the parameter of the ARMA model is stable over time.

6.






```
> accuracy(pred_2016, actual_2016)
```

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	-0.9360966	1.616367	1.013763	34.68024	143.7229	0.5579314	1.136502

```
> accuracy(pred_2017, actual_2017)
```

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	-0.9404522	1.60084	1.027565	22.12659	165.017	0.552183	1.107627

7.

```
> print(AR_AIC_US_IP_diff)

          MA0

AR0  1152.957
AR1  1086.550
AR2  1064.170
AR3  1051.919
AR4  1052.470
AR5  1051.376
AR6  1052.730
AR7  1053.377
AR8  1054.024
AR9  1055.574
AR10 1057.234
```

```
> print(AR_BIC_US_IP_diff)

          MA0

AR0  1161.491
AR1  1099.352
AR2  1081.239
AR3  1073.255
AR4  1078.073
AR5  1081.246
AR6  1086.868
AR7  1091.782
AR8  1096.696
AR9  1102.513
AR10 1108.441
```

```
> Qstatistic_ar3=LjungBox(AR3$residuals, 30, 3)> print(Qstatis
tic_ar3)
```

	lag	Ljung-Box	p-value
1	1	0.0380	NA
2	2	0.0406	NA
3	3	0.0597	NA
4	4	2.7443	0.6015
5	5	5.2583	0.3852
6	6	5.7297	0.4541
7	7	7.4960	0.3791
8	8	8.1493	0.4190
9	9	9.1979	0.4192
10	10	9.4372	0.4912
11	11	9.5550	0.5708
12	12	10.7128	0.5537
13	13	14.2899	0.3537
14	14	14.6375	0.4034
15	15	14.6623	0.4760
16	16	15.1759	0.5118
17	17	15.3051	0.5735
18	18	16.8417	0.5340
19	19	17.3868	0.5637
20	20	17.8882	0.5948
21	21	20.8420	0.4686
22	22	20.8422	0.5305
23	23	21.7201	0.5372
24	24	29.9609	0.1861
25	25	30.2099	0.2165
26	26	30.2100	0.2590
27	27	31.7279	0.2423
28	28	32.0745	0.2715
29	29	34.3935	0.2251
30	30	34.4688	0.2625

```
> adf2_trend = ur.df (US_IP_diff, lags= 2, type= "trend")> sum
mary(adf2_trend)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9855 -0.3553 -0.0050  0.3500  2.7439

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.000e-02  5.862e-02   1.535 0.125307
z.lag.1     -4.295e-01  5.324e-02  -8.066 5.06e-15 ***
tt          -7.719e-05  1.893e-04  -0.408 0.683589
z.diff.lag1 -3.308e-01  5.278e-02  -6.268 7.67e-10 ***
z.diff.lag2 -1.640e-01  4.347e-02  -3.772 0.000181 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6545 on 519 degrees of freedom
Multiple R-squared:  0.3736,	Adjusted R-squared:  0.3687
F-statistic: 77.37 on 4 and 519 DF,  p-value: < 2.2e-16


Value of test-statistic is: -8.0661 21.6909 32.5313

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

What is the link between your ADF specification for Δ lnYt and your ARMA model for Δ lnYt ?

The ADF specification is actually reparameterization of the AR model. And the order of the ADF is p-1, which p is the order of the AR model. In this case, switch the ARMA(5,5) model to the pure AR model and select the order by AIC/BIC which suggest the AR(3) model. So it i s the ADF(2) model and thus the error term is white noise by Ljung-Box test.

What do you conclude from the ADF test (order of int egration, drift...) ? is the result in line with what you e xpected from the visual inspection of the series ?

From the Augmented Dicky Fuller test, it reject the H0 of there is at least one unit root in Δ ln Yt, so Δ lnYt is I(0) and this is in line with the previous visual inspection of the series.

8.

```
> Qstatistic_ar4_level =LjungBox(AR4_level$residuals, 30, 4)>
print(Qstatistic_ar4_level )

                     lag Ljung-Box p-value

                   1    1    0.2026      NA
                   2    2    0.2749      NA
                   3    3    0.5262      NA
                   4    4    2.4730      NA
                   5    5    5.5345  0.3542
                   6    6    5.8237  0.4432
                   7    7    7.6551  0.3640
                   8    8    8.1598  0.4180
                   9    9    9.1131  0.4269
                   10  10    9.6396  0.4727
                   11  11    9.9530  0.5346
                   12  12   11.2064  0.5113
                   13  13   14.5512  0.3362
                   14  14   14.9108  0.3843
                   15  15   14.9221  0.4570
                   16  16   15.4109  0.4948
                   17  17   15.5696  0.5545
                   18  18   17.0991  0.5163
                   19  19   17.6708  0.5445
                   20  20   18.1350  0.5785
                   21  21   20.9111  0.4644
                   22  22   20.9217  0.5256
                   23  23   21.7608  0.5347
                   24  24   30.3574  0.1732
                   25  25   30.6222  0.2018
                   26  26   30.6271  0.2424
                   27  27   32.1235  0.2275
                   28  28   32.4733  0.2557
                   29  29   34.9033  0.2077
                   30  30   34.9675  0.2438
```

```
> level_adf3_trend = ur.df (US_IP, lags= 3, type= "trend")> su
mmary(level_adf3_trend)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9495 -0.3516 -0.0205  0.3405  2.7085

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.2222048  1.4595536   2.208 0.027705 *
z.lag.1     -0.0082234  0.0038290  -2.148 0.032202 *
tt           0.0014423  0.0007322   1.970 0.049397 *
```

```
z.diff.lag1  0.2408522  0.0431474   5.582 3.84e-08 ***
z.diff.lag2  0.1718692  0.0438683   3.918 0.000101 ***
z.diff.lag3  0.1737455  0.0435603   3.989 7.60e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6523 on 518 degrees of freedom
Multiple R-squared:  0.1914,      Adjusted R-squared:  0.1836
F-statistic: 24.53 on 5 and 518 DF,  p-value: < 2.2e-16


Value of test-statistic is: -2.1477 3.4186 2.39

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34





> level_adf3_drift = ur.df (US_IP, lags= 3, type= "drift")> su
mmary(level_adf3_drift)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9682 -0.3628 -0.0097  0.3413  2.7215

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4721243  0.4267207   1.106  0.26907
z.lag.1     -0.0009358  0.0009891  -0.946  0.34456
z.diff.lag1  0.2393762  0.0432605   5.533 4.99e-08 ***
z.diff.lag2  0.1670450  0.0439213   3.803  0.00016 ***
z.diff.lag3  0.1645102  0.0434273   3.788  0.00017 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6541 on 519 degrees of freedom
Multiple R-squared:  0.1854,      Adjusted R-squared:  0.1791
F-statistic: 29.53 on 4 and 519 DF,  p-value: < 2.2e-16


Value of test-statistic is: -0.9461 3.1704

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

```
> level_adf3_none = ur.df (US_IP, lags= 3, type= "none")> summ
ary(level_adf3_none)


###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9977 -0.3539 -0.0049  0.3469  2.7688

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
z.lag.1     1.559e-04  6.895e-05   2.262 0.024141 *
z.diff.lag1 2.405e-01  4.326e-02   5.561  4.3e-08 ***
z.diff.lag2 1.676e-01  4.393e-02   3.815 0.000153 ***
z.diff.lag3 1.648e-01  4.344e-02   3.794 0.000166 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6542 on 520 degrees of freedom
Multiple R-squared:  0.221, Adjusted R-squared:  0.215
F-statistic: 36.88 on 4 and 520 DF,  p-value: < 2.2e-16


Value of test-statistic is: 2.2615

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

What is the link between your ADF specification for lnYt and your ARMA model for Δ lnYt ?

By construction, the AR(p) model for Δ lnYt can be reparameterized to AR(p+1) model for lnYt, and AR(p+1) for lnYt can also be reparameterized to the ADF(p) for lnyt. in this case, it is ADF(3) same with the AR(3) for Δ lnYt.

What do you conclude from the ADF test (order of integration, drift...) ? is the result in line with what you expected from the visual inspection of the series ?

From the augmented dicky fuller test by a dynamic procedure, It firstly accept H0 of there is at least one unit root under the specification with trend term. then it accept H0 of there is no joint significance of trend term and gamma. so that change the specification and it accept H0 of there is at least one unit root under the specification with drift. then it accept H0 of there is no joint significance of drift and gamma. so that change the specification and it accept H0 of ther e is at least one unit root under the specification of no trend term and drift.

So the lnYt is I(1) with no trend and no drift, not completely follow the previous visual inspec tion for it more likely there is trend term.