```
> POLS = plm(lnGDP ~ D, data = panel_dataFrame, model = "pooli
ng")> summary(POLS)

Pooling Model

Call:
plm(formula = lnGDP ~ D, data = panel_dataFrame, model = "pool
ing")

Balanced Panel: n = 86, T = 51, N = 4386

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-365.6384  -97.2957    5.5567  115.1742  379.9918

Coefficients:
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 659.0082     2.9815 221.033 < 2.2e-16 ***
D           175.6802     4.1673  42.156 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    117260000
Residual Sum of Squares: 83436000
R-Squared:      0.28845
Adj. R-Squared: 0.28828
F-statistic: 1777.16 on 1 and 4384 DF, p-value: < 2.22e-16
```

```
> FE   = plm(lnGDP ~ D, data = panel_dataFrame, model = "withi
n")> summary(FE)

Oneway (individual) effect Within Model

Call:
plm(formula = lnGDP ~ D, data = panel_dataFrame, model = "with
in")

Balanced Panel: n = 86, T = 51, N = 4386

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-176.3868  -18.4393    2.2376   20.2982  200.6431

Coefficients:
  Estimate Std. Error t-value  Pr(>|t|)
D  18.7419     1.6888  11.098 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    5888000
Residual Sum of Squares: 5724000
R-Squared:      0.027852
Adj. R-Squared: 0.0084043
F-statistic: 123.165 on 1 and 4299 DF, p-value: < 2.22e-16
```

```
> RE   = plm(lnGDP ~ D, data = panel_dataFrame, model = "rando
m")> summary(RE)

Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = lnGDP ~ D, data = panel_dataFrame, model = "rand
om")

Balanced Panel: n = 86, T = 51, N = 4386

Effects:
                   var  std.dev share
idiosyncratic  1331.47    36.49  0.09
individual    13437.73   115.92  0.91
theta: 0.956

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-183.2031  -19.8349    1.0469   20.4884  193.5872

Coefficients:
            Estimate Std. Error z-value  Pr(>|z|)
(Intercept) 738.9733    12.6379  58.473 < 2.2e-16 ***
D            19.4544     1.6995  11.447 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    6103900
Residual Sum of Squares: 5926800
R-Squared:      0.029023
Adj. R-Squared: 0.028802
Chisq: 131.042 on 1 DF, p-value: < 2.22e-16
```

```
> pFtest(FE, POLS)

                 F test for individual effects

data:  lnGDP ~ D
F = 686.66, df1 = 85, df2 = 4299, p-value < 2.2e-16
alternative hypothesis: significant effects
```

```
> phtest(FE,RE)

                         Hausman Test

data:  lnGDP ~ D
chisq = 14.007, df = 1, p-value = 0.0001821
alternative hypothesis: one model is inconsistent
```

```
>pbgtest(FE, order = 1)

    Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  lnGDP ~ D
chisq = 3638, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors
```

What is the dynamic impact of a change in $D_{i,t}$ on $lny_{i,t}$?

Since this is a static model and no lag term, so the coefficient on Dit would be the **long run permanent impact.** with one unit shock on Dit, the response of lnyit would be 19.4544 on random effect estimation, 18.7419 on fixed effect estimation, and 175.6802 on pooled OLS estimation, which is significant different than others very likely incorrect. These three estimations need to be tested to choose.

What are the statistical properties of each of the estimators ?

By theory, The fixed effect estimator is the "within estimator" and we do this by the so-called "demean approach". **For consistency**, when N goes to infinity, it need the Dit is strictly exogenous which means Dit need to be independent to error term with any i and t. when T goes to infinity, the mean of Dit is 0 so it just need the Dit is weakly exogenous which means Dit still need to be independent to error term w.r.t any i but just contemporaneously independent with respect to t. **For biased or not,** It's unbiased if it's strictly exogenous. **For efficiency**, because we only use information from within dimension, so it's not efficient. And it's **asymptotically normal** under some regularity conditions.

The Random effect estimator is the combination of "between estimator and within estimator". **For consistency**, when N goes to infinity, we need consistency for both between estimator and within estimator. Dit still need to be strictly exogenous but also need to be uncorrelated with the individual effect. when T goes to infinity, the between variation has been dropped and random effect same as fixed effect, so we just need weak exogenous. **For efficiency**, we both use between and within dimension,so it's efficient.

The pooled OLS estimator is the OLS version of random effect estimator ignoring the individual effect. So no demean approach there. **For consistency** we just need Dit weak exogenous both for N goes to infinity and T goes to infinity and we still need it uncorrelated with error term. **For efficiency**, it is not **efficient** because of ignoring the individual effect.

And these statistical properties are supported by our test. Base on my result, from the **F-test** we reject the H0 of there is no individual effect and from the **Hausman test** we reject the H0 of the two estimators are no significant differences. So econometrically, we can not ignore the individual effect and the individual effect is correlated with Dit. **so the pooled OLS and random effect is inconsistent** and we are going to choose fixed effect estimator which are in line of the economical view that the countries as individuals are "one of a kind"

In addition, if the Dit is not exogenous and it is also determined by the lnGDPit, then all the pooled OLS, fixed effect estimator and random effect estimator is not consistent because of endogenous problem. We then need an instrument ot panel VAR to solve it. Or the Dit can be determined by the lag term of lnGDPit. then it is fine with our model.

What are the implication of your conclusion of autocorrelation test for the statistical properties of your preferred estimator ?

From the autocorrelation test, we reject the H0 of there is no autocorrelation in the error term. But it's still unbiased and consistent for the explanatory variable Dit is still independent with the error term. But it is not efficient because there is still information left in the error term and the variance of the coefficient is not minimum anymore, changing to dynamic panal model may solve this problem.

```
> POLS_inGrowth = plm(GDPgrowth ~ D, data = panel_dataFrame, m
odel = "pooling")> summary(POLS_inGrowth)

Pooling Model

Call:
plm(formula = GDPgrowth ~ D, data = panel_dataFrame, model = "
pooling")

Balanced Panel: n = 86, T = 50, N = 4300

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-71.32422  -2.02422   0.30578   2.40833  64.11578

Coefficients:
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  1.42422    0.11574 12.3057 < 2.2e-16 ***
D            0.66745    0.16151  4.1325 3.657e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    120920
Residual Sum of Squares: 120440
R-Squared:      0.0039576
Adj. R-Squared: 0.0037259
F-statistic: 17.0775 on 1 and 4298 DF, p-value: 3.6568e-05
```

```
> FE_inGrowth   = plm(GDPgrowth ~ D, data = panel_dataFrame, m
odel = "within")> summary(FE_inGrowth)

Oneway (individual) effect Within Model

Call:
plm(formula = GDPgrowth ~ D, data = panel_dataFrame, model = "
within")

Balanced Panel: n = 86, T = 50, N = 4300

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-67.00164  -1.89850   0.24409   2.29066  68.43836

Coefficients:
  Estimate Std. Error t-value Pr(>|t|)
D  0.34546    0.24004  1.4392   0.1502

Total Sum of Squares:    111200
Residual Sum of Squares: 111140
R-Squared:      0.00049139
Adj. R-Squared: -0.019912
F-statistic: 2.07124 on 1 and 4213 DF, p-value: 0.15017
```

```
> RE_inGrowth   = plm(GDPgrowth ~ D, data = panel_dataFrame, m
odel = "random")> summary(RE_inGrowth)

Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = GDPgrowth ~ D, data = panel_dataFrame, model = "
random")

Balanced Panel: n = 86, T = 50, N = 4300

Effects:
                 var std.dev share
idiosyncratic 26.381   5.136 0.941
individual     1.667   1.291 0.059
theta: 0.5096

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-69.15798  -1.88010   0.31068   2.37279  66.28202

Coefficients:
            Estimate Std. Error z-value  Pr(>|z|)
(Intercept)  1.51909    0.19234  7.8979 2.836e-15 ***
D            0.48269    0.20864  2.3135    0.0207 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Total Sum of Squares:    113530
Residual Sum of Squares: 113390
R-Squared:      0.0012437
Adj. R-Squared: 0.0010113
Chisq: 5.35216 on 1 DF, p-value: 0.020697
```

```
> pFtest(FE_inGrowth, POLS_inGrowth)

                F test for individual effects

data:  GDPgrowth ~ D
F = 4.1468, df1 = 85, df2 = 4213, p-value < 2.2e-16
alternative hypothesis: significant effects
```

```
> phtest(FE_inGrowth,RE_inGrowth)

                        Hausman Test

data:  GDPgrowth ~ D
chisq = 1.3368, df = 1, p-value = 0.2476
alternative hypothesis: one model is inconsistent
```

```
> pbgtest(RE_inGrowth, order = 1)

   Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  GDPgrowth ~ D
chisq = 261.18, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors
```

What is the dynamic impact of a change in $D_{i,t}$ on $\Delta \ln y_{i,t}$ and $\ln y_{i,t}$?

This is still a static model with no lag terms so the response of △lnyit to the shock of Dit is still long run and permanent, it is the coefficient of Dit which is 0.66745 by pooled OLS estimator, 0.34546 by fixed effect estimator and 0.48269 on random effect. And these three estimations still need to be tested to choose.

The response of lnyit to Dit now is the cumultaive impulse response of △lnyit, so now the estimation result of this specification indicates that the response of the lnyit is also permanent and it is upper trend because the impact of Dit on GDP growth rate now is permanent. This is no t really possible so we may need to extend the static model to dynamic model to capture more dynamic behavior.

What are the statistical properties of each of the estimators ?

As theory is mentioned above, **For consistency,** with N goes to infinity we need Dit to be strict exogenous both fixed effect estimator and random effect estimator and one more condition for random effect estimator is Dit need to be uncorrelated. when T goes to infinity, the fixed effect and random effect is pretty much same, both of this two just need to be weakly exogenous. pooled OLS ignore the individual effect so just need Dit to be weakly exogenous and uncorrelated with error term. **For efficiency,** fixed effect estimator and pooled OLS is still inefficient and random effect estimator is efficient.

By statistical test, it is still in line with theory. From the F-test we reject the H0 of there is no individual effect. From the Hausman test, the p-value is 0.2476 and we can not reject the H0 of there is no significant difference between this two estimations under 95% significance level or even 90% significance level. **So pooled OLS estimator is consistent but not efficient, (clearly it's not that different with the fixed effect and random effect unlike the first model in level)** fixed effect estimator is consistent but not efficient, and random effect estimator is consistent and efficient. so we are going to choose random effect estimations.

By economic view, the reason might be there is no very big changing of GDP growth rate within one country over time. But the variations between different countries are relevant big. So there might be more information on between dimension rather than within dimension.

In addition, like above, it still need to be considered if the Dit is endogenous.

What are the implication of your conclusion of autocorrelation test for the statistical properties of your preferred estimator ?

From the autocorrelation test, we still reject the H0 of there is no autocorrelation in the error term. And also it is still unbiased and consistent, but it is inefficient for there is information still left in the error term and its variance is not minimum now. We need to change our specification might to dynamic model.

```
> POLS__inlevel_lag <- plm(lnGDP ~ D+lag1+lag2+lag3+lag4, data
 = panel_dataFrame, model = "pooling") > summary(POLS__inlevel
_lag)

Pooling Model
```

```
Call:
plm(formula = lnGDP ~ D + lag1 + lag2 + lag3 + lag4, data = pa
nel_dataFrame,
    model = "pooling")

Unbalanced Panel: n = 86, T = 43-44, N = 3783

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-542.5601   -4.5566    2.8309    8.4306  399.4034

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 26.653660   2.841798  9.3792 < 2.2e-16 ***
D           12.529858   1.332774  9.4013 < 2.2e-16 ***
lag1         1.608782   0.069454 23.1633 < 2.2e-16 ***
lag2        -0.727217   0.112891 -6.4418  1.33e-10 ***
lag3         0.135656   0.111079  1.2213    0.2221
lag4        -0.062679   0.068757 -0.9116    0.3620
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    103530000
Residual Sum of Squares: 4663800
R-Squared:      0.95495
Adj. R-Squared: 0.95489
F-statistic: 16013 on 5 and 3777 DF, p-value: < 2.22e-16
```

```
> FE_inlevel_lag <- plm(lnGDP ~ D+lag1+lag2+lag3+lag4, data =
panel_dataFrame, model = "within")> summary(FE_inlevel_lag)

Oneway (individual) effect Within Model

Call:
plm(formula = lnGDP ~ D + lag1 + lag2 + lag3 + lag4, data = pa
nel_dataFrame,
    model = "within")

Unbalanced Panel: n = 86, T = 43-44, N = 3783

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-248.1745   -8.6546    1.7160   11.4553  128.8817

Coefficients:
      Estimate Std. Error t-value  Pr(>|t|)
D    10.438922   1.206476  8.6524 < 2.2e-16 ***
lag1  0.922245   0.050345 18.3184 < 2.2e-16 ***
lag2 -0.471716   0.077406 -6.0941 1.214e-09 ***
lag3  0.042053   0.076057  0.5529    0.5804
lag4 -0.060180   0.048802 -1.2331    0.2176
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    3819600
Residual Sum of Squares: 2130700
R-Squared:      0.44218
Adj. R-Squared: 0.42858
```

```
F-statistic: 585.314 on 5 and 3692 DF, p-value: < 2.22e-16


> RE_inlevel_lag <- plm(lnGDP ~ D+lag1+lag2+lag3+lag4, data =
panel_dataFrame, model = "random")> summary(RE_inlevel_lag)

Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = lnGDP ~ D + lag1 + lag2 + lag3 + lag4, data = pa
nel_dataFrame,
    model = "random")

Unbalanced Panel: n = 86, T = 43-44, N = 3783

Effects:
                  var std.dev share
idiosyncratic 577.111  24.023 0.995
individual      2.671   1.634 0.005
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.08674 0.08850 0.08850 0.08848 0.08850 0.08850

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-537.43   -4.90    2.88    0.00    8.82  394.84

Coefficients:
             Estimate Std. Error z-value  Pr(>|z|)
(Intercept) 33.370159   3.053057 10.9301 < 2.2e-16 ***
D           13.374865   1.374072  9.7337 < 2.2e-16 ***
lag1         1.590397   0.069436 22.9045 < 2.2e-16 ***
lag2        -0.719355   0.112230 -6.4096 1.459e-10 ***
lag3         0.132355   0.110420  1.1987    0.2307
lag4        -0.058292   0.068694 -0.8486    0.3961
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    86676000
Residual Sum of Squares: 4606700
R-Squared:      0.94685
Adj. R-Squared: 0.94678
Chisq: 67277.4 on 5 DF, p-value: < 2.22e-16



> pFtest(FE_inlevel_lag, POLS__inlevel_lag)

                 F test for individual effects

data:  lnGDP ~ D + lag1 + lag2 + lag3 + lag4
F = 51.639, df1 = 85, df2 = 3692, p-value < 2.2e-16
alternative hypothesis: significant effects



> phtest(FE_inlevel_lag, RE_inlevel_lag)

                         Hausman Test
```

```
data:  lnGDP ~ D + lag1 + lag2 + lag3 + lag4
chisq = 4389.5, df = 5, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

```
> FE_inlevel_lag_adjust <- plm(lnGDP ~ D+lag1+lag2, data = pan
el_dataFrame, model = "within")> summary(FE_inlevel_lag_adjust)

Oneway (individual) effect Within Model

Call:
plm(formula = lnGDP ~ D + lag1 + lag2, data = panel_dataFrame,

    model = "within")

Unbalanced Panel: n = 86, T = 47-48, N = 4127

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-282.4355   -8.2967    2.1184   11.6292  162.9124

Coefficients:
      Estimate Std. Error t-value  Pr(>|t|)
D    10.590942   1.175124  9.0126 < 2.2e-16 ***
lag1  0.948318   0.048088 19.7205 < 2.2e-16 ***
lag2 -0.451443   0.047620 -9.4801 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    4910200
Residual Sum of Squares: 2428800
R-Squared:      0.50537
Adj. R-Squared: 0.49459
F-statistic: 1375.22 on 3 and 4038 DF, p-value: < 2.22e-16
```

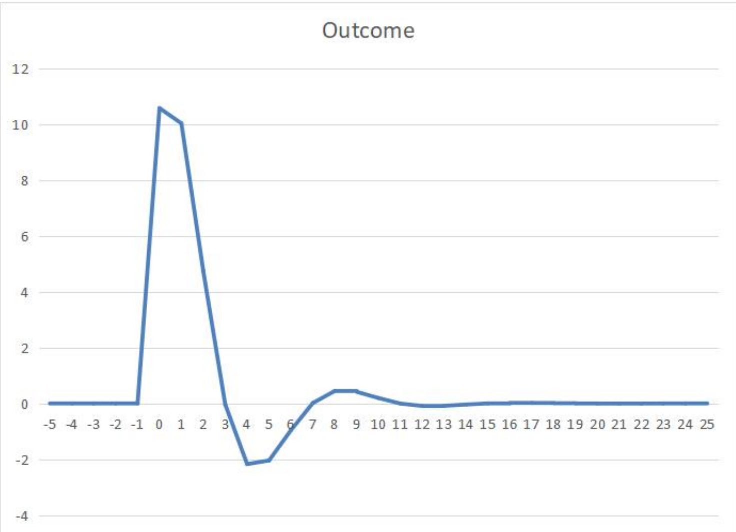
```
> pbgtest(FE_inlevel_lag_adjust, order = 1)

   Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  lnGDP ~ D + lag1 + lag2
chisq = 883.53, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors
```

```
> pbgtest(FE_inlevel_lag, order = 1)

   Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  lnGDP ~ D + lag1 + lag2 + lag3 + lag4
chisq = 936.9, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors
```

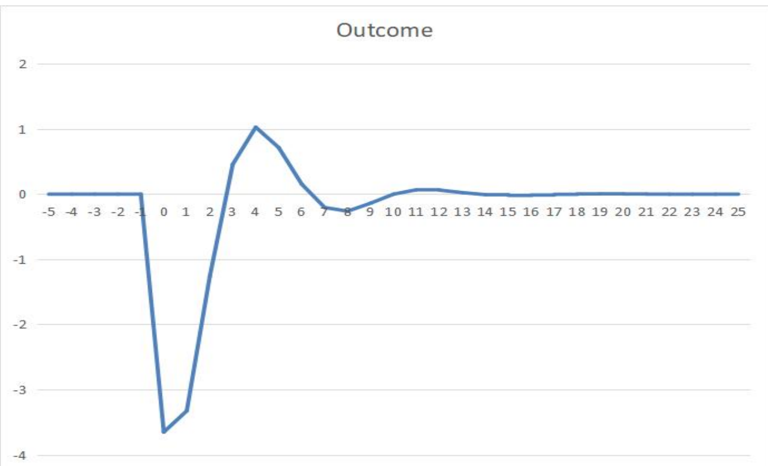
```
> TFE_inlevel_lag_adjust <- plm(lnGDP ~ D+lag1+lag2, data = pa
nel_dataFrame, model = "within", effect= "twoways")> summary(T
FE_inlevel_lag_adjust)

Twoways effects Within Model

Call:
plm(formula = lnGDP ~ D + lag1 + lag2, data = panel_dataFrame,

    effect = "twoways", model = "within")

Unbalanced Panel: n = 86, T = 47-48, N = 4127

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-182.404110   -7.405776   -0.077214    7.750292  196.750535

Coefficients:
      Estimate Std. Error t-value  Pr(>|t|)
D    -3.653130   1.079523  -3.384 0.0007212 ***
lag1  0.910859   0.040995  22.219 < 2.2e-16 ***
lag2 -0.483570   0.040844 -11.840 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    3094100
Residual Sum of Squares: 1618900
R-Squared:      0.47676
Adj. R-Squared: 0.45906
F-statistic: 1212.17 on 3 and 3991 DF, p-value: < 2.22e-16



> pFtest(TFE_inlevel_lag_adjust, FE_inlevel_lag_adjust)

                  F test for twoways effects

data:  lnGDP ~ D + lag1 + lag2
F = 42.477, df1 = 47, df2 = 3991, p-value < 2.2e-16
alternative hypothesis: significant effects



> pbgtest(TFE_inlevel_lag_adjust, order = 1)

   Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  lnGDP ~ D + lag1 + lag2
chisq = 1363.7, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors




> CCEP<- pcce(lnGDP ~ D+lag1+lag2, data = panel_dataFrame, mod
el = "p")> summary(CCEP)
```

```
Common Correlated Effects Pooled model

Call:
pcce(formula = lnGDP ~ D + lag1 + lag2, data = panel_dataFrame, 

    model = "p")

Unbalanced Panel: n = 86, T = 47-48, N = 4127

Residuals:
        Min.      1st Qu.       Median      3rd Qu.          Ma
x. 
-62.83554146  -3.02347397   0.06535095   3.27259385  60.386926
34 

Coefficients:
      Estimate Std. Error z-value  Pr(>|z|)    
D     0.696297   1.009302  0.6899   0.49027    
lag1  0.727282   0.094967  7.6582 1.885e-14 ***
lag2 -0.299151   0.085811 -3.4861   0.00049 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares: 111410000
Residual Sum of Squares: 169500
HPY R-squared: 0.96224


> pbgtest(CCEP, order = 1)

   Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  lnGDP ~ D + lag1 + lag2
chisq = 3227.4, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors




> CCEMG<- pcce(lnGDP ~ D+lag1+lag2, data = panel_dataFrame, mo
del = "mg")> summary(CCEMG)Common Correlated Effects Mean Grou
ps model

Call:
pcce(formula = lnGDP ~ D + lag1 + lag2, data = panel_dataFrame, 

    model = "mg")

Unbalanced Panel: n = 86, T = 47-48, N = 4127

Residuals:
        Min.      1st Qu.       Median      3rd Qu.          Ma
x. 
-60.47077338  -2.51719699   0.09631358   2.79365840  54.526641
71 

Coefficients:
      Estimate Std. Error z-value Pr(>|z|)    
D     0.057980   0.774774  0.0748   0.9403    
lag1  0.624131   0.033152 18.8262   <2e-16 ***
lag2 -0.233573   0.025709 -9.0854   <2e-16 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares: 111410000
Residual Sum of Squares: 132200
HPY R-squared: 0.96835


> pbgtest(CCEMG, order = 1)

	Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  lnGDP ~ D + lag1 + lag2
chisq = 3227.4, df = 1, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors
```



FE_inlevel_lag_adjust



TFE_inlevel_lag_adjust

What is your choice of p?

From the economic intuition, I firstly choose to set the p=4, because most of countries' presidentil term is 4 years. So there may be a economic cycle due to this.the president may impose policies with a 4-year goal. And with F-test and Hausman test, I choose fixed effect to estimate my model. However, From the estimation result, we can see that the lag3 and lag4 is not significant. So I change my specificiation to the order of p=2. And all the coefficients now are significant.

Then I do the autocorrealtion test, it shows that there is still autocorrelation in the error term on p=2, which means that the model may not adequate. then I do the autocorrealtion test of p=4, but it is still strongly reject the H0 of no autocorrelation in error term. So adding more lag term may not the way to solve autocorrelation problem and may bring overfitting problem.

So I try to adjust the specification by considering time effect. By Twoway fixed effect model, all the coefficient is significant and the F-test show that the we reject the H0 of no significant time effect. So the Twoway fixed effect is better. But as for autocorrelation test it's still reject the H0 of there is no autocorrelation when p=2 and even in p=4. And same when I change it to common correlated effects model and when I add slope heterogeneity. So we may need another approach like GMM to estimate this model.

What is the dynamic impact of Dit on lnGDPit ?

From the impulse response function above, we can know that when considering time effect, the response of lnGDPit to one transitory shock of Dit firstly be negative and then the negative impact gradually decrease, about 3 years after, the impact become positive and increasing till 4 years after, then it start to decrease and become negative a bit again at 7 years after the shock, then back to positive and finally gradually goes to 0.

And if we just use fixed effect do not consider the time effect, it is opposite with the twoway fixed effect, the impact firstly be positive and then start to decrease, at 3 years after the shock, the impact become negative and back to postive at 7 years and finally gradually goes to 0.

What are the statistical properties of each of the estimators ?

While for Dit, it is still same with the static model, **For consistency**, weakly exogenous for the pooled OLS estimator both for N and T goes to infinity and uncorrelated with the error term. fixed effect estimator and random effect estimator need both strictly exogenous when N goes to infinity and weakly exogenous when T goes to infinity. And also, random effect need Dit is not correlated with the individual effect. That is for Dit. and **now we also have the lag term of lnyit, so we also need lnyit-1 and lnyit-2 is not correlated with the error term. For efficiency,** still only random effect estimator is efficient because it use information from both within dimension and between dimension. **For unbiasedness,** the model is biased because the explanatory is not completely uncorrelated with error term.

It is supported by the statistical test, from the F-test, we reject H0 of there is no individual effect, and the pooled OLS is inconsistent. from the Huasman test, we reject the H0 of there is no significant difference between fixed effect estimations and random effect estimations. So t he Dit is correlated with the individual effect. The random effect is not consistent. And we ch oose fixed effect. from economic view, the country as individuals is one of a kind, not arbitrar y random draws.

Also, like mentioned above, if the Dit is endogenous and determined by the lnyit, the model w ill not be consistent. And if the Dit is determined just by the lag term of lnyit, there might be multicolinearity in model but since all the coefficient is significant, and the model is still cons istent, so it is still valid.

In addition, we further add the time effect and do the F-test and we reject H0 of there is no ti me effect. So we choose twoway fixed effect finally.

```
> POLS_GDPgrowth_lag <- plm(GDPgrowth ~ D+lag_1+lag_2, data =
panel_dataFrame, model = "pooling")> summary(POLS_GDPgrowth_la
g)

Pooling Model

Call:
plm(formula = GDPgrowth ~ D + lag_1 + lag_2, data = panel_data
Frame,
    model = "pooling")

Balanced Panel: n = 86, T = 46, N = 3956

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-70.72353  -1.92882   0.31762   2.25130  66.88238

Coefficients:
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 1.060179   0.123727  8.5687 < 2.2e-16 ***
D           0.552004   0.165610  3.3332 0.0008666 ***
lag_1       0.164564   0.015754 10.4460 < 2.2e-16 ***
lag_2       0.032148   0.015457  2.0799 0.0375968 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    110450
Residual Sum of Squares: 106710
R-Squared:      0.033879
Adj. R-Squared: 0.033146
F-statistic: 46.1954 on 3 and 3952 DF, p-value: < 2.22e-16
```

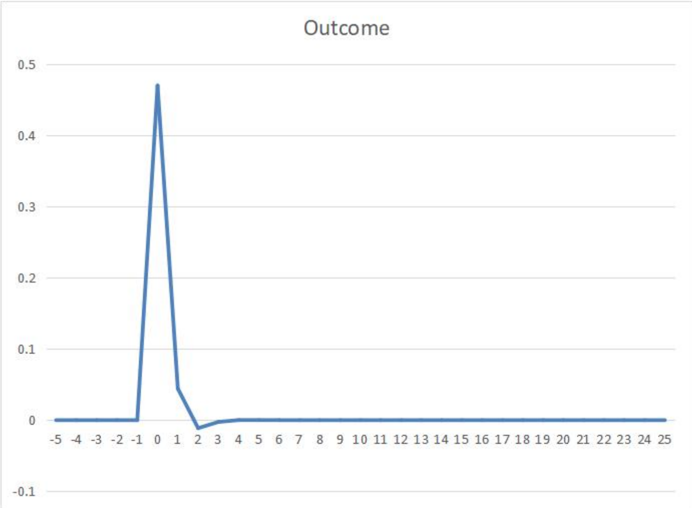
```
> FE_GDPgrowth_lag <- plm(GDPgrowth ~ D+lag_1+lag_2, data = pa
nel_dataFrame, model = "within")> summary(FE_GDPgrowth_lag)

Oneway (individual) effect Within Model

Call:
plm(formula = GDPgrowth ~ D + lag_1 + lag_2, data = panel_data
Frame,
```

```
    model = "within")

Balanced Panel: n = 86, T = 46, N = 3956

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-66.86243  -1.79309   0.29183   2.15417  67.93701

Coefficients:
       Estimate Std. Error t-value  Pr(>|t|)
D      0.470333   0.247211  1.9026   0.05717 .
lag_1  0.094269   0.015960  5.9066 3.793e-09 ***
lag_2 -0.032831   0.015635 -2.0998   0.03581 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    100490
Residual Sum of Squares: 99437
R-Squared:      0.010502
Adj. R-Squared: -0.012016
F-statistic: 13.6805 on 3 and 3867 DF, p-value: 7.1032e-09
```

```
> RE_GDPgrowth_lag <- plm(GDPgrowth ~ D+lag_1+lag_2, data = pa
nel_dataFrame, model = "random")> summary(RE_GDPgrowth_lag)

Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = GDPgrowth ~ D + lag_1 + lag_2, data = panel_data
Frame,
    model = "random")

Balanced Panel: n = 86, T = 46, N = 3956

Effects:
                 var std.dev share
idiosyncratic 25.714   5.071     1
individual     0.000   0.000     0
theta: 0

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-70.72353  -1.92882   0.31762   2.25130  66.88238

Coefficients:
            Estimate Std. Error z-value  Pr(>|z|)
(Intercept) 1.060179   0.123727  8.5687 < 2.2e-16 ***
D           0.552004   0.165610  3.3332 0.0008587 ***
lag_1       0.164564   0.015754 10.4460 < 2.2e-16 ***
lag_2       0.032148   0.015457  2.0799 0.0375325 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    110450
Residual Sum of Squares: 106710
R-Squared:      0.033879
Adj. R-Squared: 0.033146
```

```
Chisq: 138.586 on 3 DF, p-value: < 2.22e-16


> pFtest(FE_GDPgrowth_lag, POLS_GDPgrowth_lag)

                 F test for individual effects

data:  GDPgrowth ~ D + lag_1 + lag_2
F = 3.3282, df1 = 85, df2 = 3867, p-value < 2.2e-16
alternative hypothesis: significant effects



> phtest(FE_GDPgrowth_lag, RE_GDPgrowth_lag)

                         Hausman Test

data:  GDPgrowth ~ D + lag_1 + lag_2
chisq = 391.95, df = 3, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent


> pbgtest(FE_GDPgrowth_lag)

   Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  GDPgrowth ~ D + lag_1 + lag_2
chisq = 372.98, df = 46, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors




> FE_GDPgrowth_lag_adjust <- plm(GDPgrowth ~ D+lag_1+lag_2+lag
_3+lag_4, data = panel_dataFrame, model = "within")> summary(F
E_GDPgrowth_lag_adjust)

Oneway (individual) effect Within Model

Call:
plm(formula = GDPgrowth ~ D + lag_1 + lag_2 + lag_3 + lag_4,
    data = panel_dataFrame, model = "within")

Balanced Panel: n = 86, T = 42, N = 3612

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-66.47239  -1.78733   0.28224   2.18673  66.90491

Coefficients:
       Estimate Std. Error t-value  Pr(>|t|)
D      0.549837   0.256032  2.1475  0.031819 *
lag_1  0.095684   0.016348  5.8529 5.272e-09 ***
lag_2 -0.039753   0.016258 -2.4451  0.014530 *
lag_3 -0.042268   0.016022 -2.6381  0.008373 **
lag_4 -0.035776   0.015805 -2.2636  0.023658 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Total Sum of Squares:    88647
Residual Sum of Squares: 87166
R-Squared:      0.016715
Adj. R-Squared: -0.0084187
F-statistic: 11.9708 on 5 and 3521 DF, p-value: 1.6354e-11


> pbgtest(FE_GDPgrowth_lag_adjust)

	Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  GDPgrowth ~ D + lag_1 + lag_2 + lag_3 + lag_4
chisq = 335.21, df = 42, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors



> TFE_GDPgrowth_lag_adjust <- plm(GDPgrowth ~ D+lag1+lag2, dat
a = panel_dataFrame, model = "within", effect= "twoways")> sum
mary(TFE_GDPgrowth_lag_adjust)

Twoways effects Within Model

Call:
plm(formula = GDPgrowth ~ D + lag1 + lag2, data = panel_dataFr
ame,
    effect = "twoways", model = "within")

Balanced Panel: n = 86, T = 47, N = 4042

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-64.45748  -1.83081   0.21081   2.09663  63.12833

Coefficients:
      Estimate Std. Error t-value  Pr(>|t|)
D     0.526150   0.264344  1.9904   0.04662 *
lag1  0.024659   0.010011  2.4633   0.01381 *
lag2 -0.054865   0.009978 -5.4985 4.074e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    95414
Residual Sum of Squares: 92130
R-Squared:      0.034416
Adj. R-Squared: 0.001299
F-statistic: 46.4187 on 3 and 3907 DF, p-value: < 2.22e-16


> pbgtest(TFE_GDPgrowth_lag_adjust)

	Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  GDPgrowth ~ D + lag1 + lag2
chisq = 291.69, df = 47, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic er
rors
```



FE_GDPgrowth_lag

What is your choice of p?

As for the GDP in level, the choice of p=2, so for GDP growth I also set my specification starting from p=2 and the pooled OLS, fixed effect and random effect all shows that the lag1 is significant, and the lag 2 term is only significant at 0.01significant level. So I think p=2 may already adequate.

Then I do the autocorrelation test, it reject the H0 of no autocorrelation in the error term. So I extend the model to p=4 and test for autocorrelation again. the Dit become less significant and there is still autocorrelation problem.

So I believe like GDP in level, there may be the issue that cannot eliminate the autocorrelation by just extending the lag which will also bring overfitting problem. Then I try to adjust my specification to a more proper model by adding time effect. I use the twoway fixed effect to estimate and test for autocorrelation again, and it still reject H0 of no autocorrelation, so we may need other more valid approach.

What is the dynamic impact of Dit on ΔlnGDPit and lnGDPit?

From the impulse response function above, the dynamic response of ΔlnGDPit to the one unit transitory shock of Dit will be first positive and then after two years the influence become near 0. it is negative very little and then goes back to positive also very little and then converge to 0.

And the response of the lnGDPit is the cumulative response of the ΔlnGDPit to the impact of the Dit now. and **from the impulse responese function above, it show that the cumulative response of ΔlnGDPit is large than 0. put differently, even the unit shock of Dit is transitory, it still have a permanent impact on lnGDPit.**

This is in line with our expectation because by this model, it impose that the lnyit have a unit root. So it will not converge to 0.

What are the statistical properties of each of the estimators ?

In theory, always same story, **For consistency,** pooled OLS need Dit weakly exogenous for both N goes to infinity and T goes to infinity and uncorrelated with the error term. The lag term need to be contemporaneously uncorrelated with the error term. fixed effect estimator and random effect estimator both need to be weakly exogenous when T goes to infinity and strictly exogenous when N goes to infinity and random effect estimator also need Dit uncorrelated with the individual effect, the lag term contemporaneously uncorrelated. **For efficiency,** only the random effect estimator is efficient. **For unbiasedness,** this is dynamic model so it is all biased.

By statistical test, it's in line of the theory, from the F-test we reject the H0 of no individual effect, from the Hausman test we reject the H0 of there is no significant difference between fixed effect estimator and random effect estimator. so we know the pooled OLS is not consistent and randome effect is not consistent. We choose fixed effect estimator. From the economic view, the country as individuals is one of a kind, not arbitrary random draws.

Further questions:

Compute the dynamic impact of democratization on both economic growth and the level of GDP over a 30 years horizon after the shock towards a democracy.

See the impulse response function above.

Which of the above model do you prefer(and why)?

I would choose the model of lnyit's dynamic model on Dit (model(3)). static model can not be the case because GDP can always be influenced by the lag term and indeed it is significant in the dynamic model. static model will miss many dynamic behavior.

And if I use GDP in first difference, I impose the lnGDPit have a unit root without test if it is stationary or not. And even let say it is I(1), **it still might be misspecification if there is cointegration relationship between lnGDPit and Dit, which can not be reparameterization to error correction model, so it miss the error correction term.** Also, **Δ lnGDPit will be I(0), and Dit is I(1), in general this cannot be estimated, the error term will be I(1).**

However, as a matter of fact, the Δ lnGDPit is non-stationary by construction (true reason might be structural break), **but we can never know,** the unit root test now is not reliable for it can not identify the structural break.

And also the dynamic impact of transitory shock on Dit will be permanent on lnGDPit in level which would not really happening in the real world.

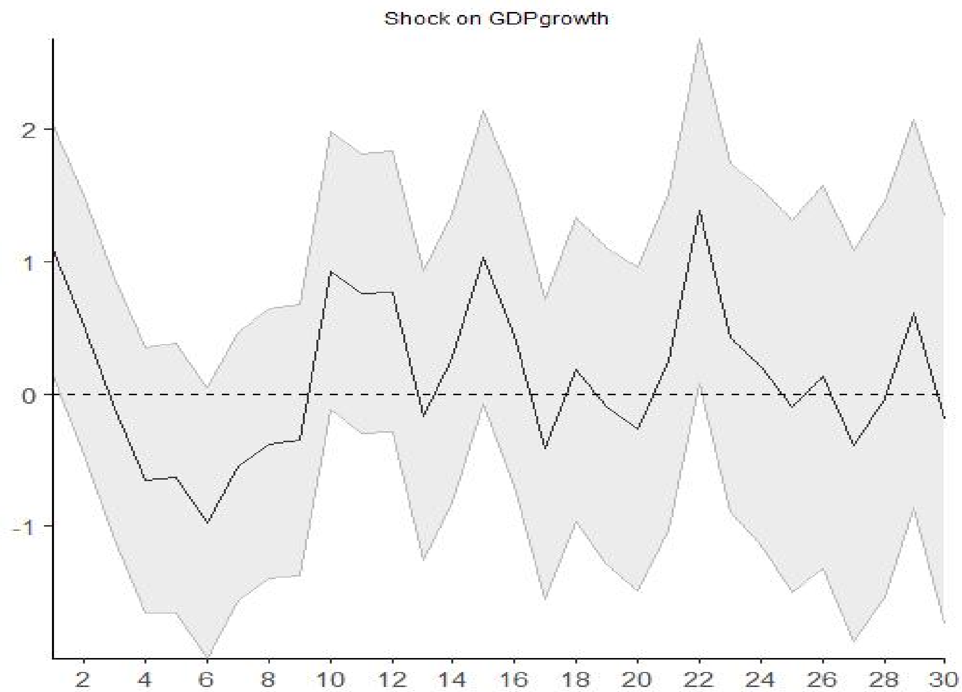
By using lnGDP in level, the real DGP will nested in my specification, I will get consistent estimates. **and in panel model, my model will be consistent even if the variable is non-stationary and no cointegration by adding common correlated fator to absorb the cross-sectional dependence.**

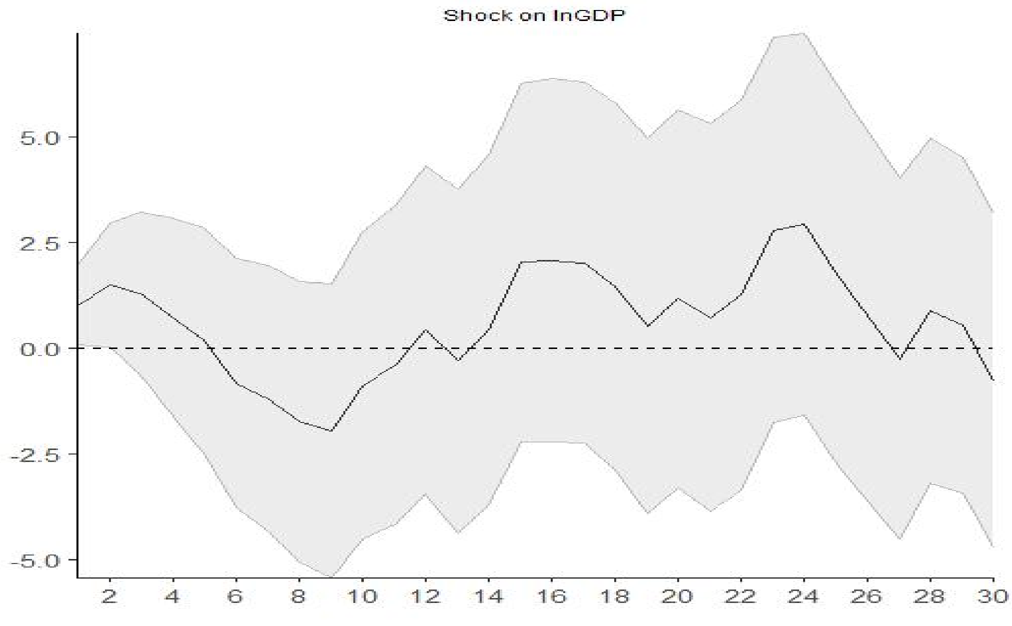
In addition, I choose fixed effect estimator both this two dynamic model which will use only information from within dimension. And the lnGDP in first difference may reduce lots of within information for there may not be much variation in GDP growth within one country.

Suppose we replace Dit by ΔDit in eqs. (1)-(4). How does this change the dynamic impact of democratization? Does this make sense?

The model become to in first differences in (2) and (4) and it only valid when there is no cointegration between lnGDPit and Dit. And it only reflect the short run dynamics.

Also the meaning of the coefficient will change, as the Dit is a dummy variable, its in differences **capture more changes in democratic status rather than the democratic status.** So this will be the dynamic impact of the democratic changes to the lnyit or Δlnyit instead of how the democratic status impact the lnyit or Δlnyit. If we use Dit, it is more capturing the **impact of being democracy on lnyit or Δlnyit,** it is typically considered **long run and persistent.** And if we use ΔDit, it explained how the **change from democratic status to non democracy** or other way round impact economic and it is usually short run and temporary.





What are the statistical properties of the FE estimator ?

For consistency, when N goes to infinity, the Dit need to be strictly exogenous and when T goes to infinity, the Dit need to be weakly exogenous, and also the lag term of Δ lnyit need to be contemporaneously uncorrelated with the error term. Also, **this is the GDP growth with Di**

t in first difference, so this is all stationary and local projection even if we miss the error term, it also can be consistent.

And for efficiency, Fixed effect estimator is not efficient because it only use information fro m within dimension. And this is **biased** because in dynamic model explanatory variable is not completely independent from error term.

What is the dynamic impact of democratization on G DP ?

From the impulse response function of lnGDP above, we know that it is the cumulated impuls e response of the GDP growth, and it's firstly positive to the five years after the change to the democratization, and then it become negative till near 12 years after, with a small fluctuation, starting from 14 years it become positive again, then, it basically keep positive to the 30 years after.

Would you rather use Dit (instead of Δ Dit) in equatio n (5) ?

No, because the local projection is more robust in first difference form, even if there is cointer gration relationship between lnyit and Dit. It is consistent and I can do standard inference.

But if I switch to Dit, then I can not do standard inference and the bias will lead it to inc onsistent. The performance of local projection will deteriorates with the h increases for t here might be spurious-regression-like problem in error term if the Dit have a unit root.

Consider the following alternative models, which mod el do you prefer and why ?

I prefer the model 2, because from the left hand side, it is cumulative lnyit in first difference(i. e. cumulative GDP growth rate). and on the right hand side, they are Dit in difference and lag term of Dit in first difference and lnyit in first difference. **All the variables are stationary.** S o compared to model 1 and 3 in level. **model 2 is consistent even if there is cointegration re lationship.** What is more, model 2 also contain the lag term of Dit in first difference, and the l nyit may also be impacted with that. **so it can capture more comprehensive dynamic impac t compared to model 4.** only problem is there by doing so there might be some multicolineari ty in the model, but it need to be test with that and if all the parameters are significant, then it can be ignored.

```
> purtest(panel_dataFrame$lnGDP, exo= "trend",test = "levinlin
")

  Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts and Trend)

data:  panel_dataFrame$lnGDP
z = -3.22, p-value = 0.000641
alternative hypothesis: stationarity
```

```
> purtest(panel_dataFrame$lnGDP, exo= "trend", test = "ips",in
dex = c ("CountryID","TimeID"))

 Im-Pesaran-Shin Unit-Root Test (ex. var.: Individual Intercepts and Trend)

data:  panel_dataFrame$lnGDP
Wtbar = 0.66795, p-value = 0.7479
alternative hypothesis: stationarity
```

```
> purtest(panel_dataFrame$lnGDP,pmax = 4, exo = "trend", test
= "madwu")

    Maddala-Wu Unit-Root Test (ex. var.: Individual Intercepts and Trend)

data:  panel_dataFrame$lnGDP
chisq = 161.48, df = 172, p-value = 0.7065
alternative hypothesis: stationarity
```

```
> purtest(panel_dataFrame[, "GDPgrowth"],exo = "intercept", te
st = "levinlin")

       Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts)

data:  panel_dataFrame[, "GDPgrowth"]
z = -42.532, p-value < 2.2e-16
alternative hypothesis: stationarity
```

```
> purtest(panel_dataFrame[, "GDPgrowth"], test = "ips", exo =
"intercept")

      Im-Pesaran-Shin Unit-Root Test (ex. var.: Individual Intercepts)

data:  panel_dataFrame[, "GDPgrowth"]
Wtbar = -43.857, p-value < 2.2e-16
alternative hypothesis: stationarity
```

```
> purtest(panel_dataFrame[, "GDPgrowth"],pmax = 4, exo = "inte
rcept", test = "madwu")

        Maddala-Wu Unit-Root Test (ex. var.: Individual Intercepts)

data:  panel_dataFrame[, "GDPgrowth"]
chisq = 2648.4, df = 172, p-value < 2.2e-16
alternative hypothesis: stationarity
```

```
. xtunitroot llc d,  demean lags(bic 1)

Levin-Lin-Chu unit-root test for d
-------------------------------------------
Ho: Panels contain unit roots              Number of panels  =     86
Ha: Panels are stationary                  Number of periods =     51

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included                 Cross-sectional means removed

ADF regressions: 0.02 lags average (chosen by BIC)
LR variance:     Bartlett kernel, 11.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                      Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t          -1.8265
 Adjusted t*            6.6896        1.0000
------------------------------------------------------------------------------
```

```
. xtunitroot ips d,  demean lags(bic 1)

Im-Pesaran-Shin unit-root test for d
-------------------------------------------
Ho: All panels contain unit roots          Number of panels  =     86
Ha: Some panels are stationary             Number of periods =     51

AR parameter: Panel-specific               Asymptotics: T,N -> Infinity
Panel means:  Included                                      sequentially
Time trend:   Not included                 Cross-sectional means removed

ADF regressions: 0.02 lags average (chosen by BIC)
------------------------------------------------------------------------------
                      Statistic      p-value
------------------------------------------------------------------------------
 W-t-bar                6.0960        1.0000
------------------------------------------------------------------------------
```

```
. xtunitroot fisher d,  dfuller demean lags(1)

Fisher-type unit-root test for d
Based on augmented Dickey-Fuller tests
--------------------------------------
Ho: All panels contain unit roots           Number of panels  =     86
Ha: At least one panel is stationary        Number of periods =     51

AR parameter: Panel-specific                Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included                  Cross-sectional means removed
Drift term:   Not included                  ADF regressions: 1 lag
-------------------------------------------------------------------------------
                                  Statistic      p-value
-------------------------------------------------------------------------------
 Inverse chi-squared(172)  P       150.3921       0.8813
 Inverse normal            Z         5.0010       1.0000
 Inverse logit t(434)      L*        4.8840       1.0000
 Modified inv. chi-squared Pm       -1.1650       0.8780
-------------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
-------------------------------------------------------------------------------
```

```
. xtunitroot llc diff_d ,  demean lags(bic 1)

Levin-Lin-Chu unit-root test for diff_d
---------------------------------------
Ho: Panels contain unit roots               Number of panels  =     86
Ha: Panels are stationary                   Number of periods =     50

AR parameter: Common                        Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included                  Cross-sectional means removed

ADF regressions: 0.05 lags average (chosen by BIC)
LR variance:     Bartlett kernel, 11.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -62.1569
 Adjusted t*         -57.4304       0.0000
------------------------------------------------------------------------------
```

```
. xtunitroot ips diff_d ,  demean lags(bic 1)

Im-Pesaran-Shin unit-root test for diff_d
------------------------------------------
Ho: All panels contain unit roots            Number of panels  =     86
Ha: Some panels are stationary               Number of periods =     50

AR parameter: Panel-specific                 Asymptotics: T,N -> Infinity
Panel means:  Included                                      sequentially
Time trend:   Not included                   Cross-sectional means removed

ADF regressions: 0.05 lags average (chosen by BIC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 W-t-bar             -54.6772       0.0000
------------------------------------------------------------------------------
```

```
. xtunitroot fisher diff_d , dfuller demean lags(1)

Fisher-type unit-root test for diff_d
Based on augmented Dickey-Fuller tests
--------------------------------------
Ho: All panels contain unit roots            Number of panels  =     86
Ha: At least one panel is stationary         Number of periods =     50

AR parameter: Panel-specific                 Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included                   Cross-sectional means removed
Drift term:   Not included                   ADF regressions: 1 lag
------------------------------------------------------------------------------
                                  Statistic      p-value
------------------------------------------------------------------------------
 Inverse chi-squared(172)   P     1915.0303       0.0000
 Inverse normal             Z      -37.9527       0.0000
 Inverse logit t(434)       L*     -56.8834       0.0000
 Modified inv. chi-squared Pm       93.9779       0.0000
------------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
------------------------------------------------------------------------------
```

```
> pcdtest(lnGDP)

        Pesaran CD test for cross-sectional dependence in panels

data:  lnGDP
z = 187.73, p-value < 2.2e-16
alternative hypothesis: cross-sectional dependence
```

```
> pcdtest(GDPgrowth)
```

```
          Pesaran CD test for cross-sectional dependence in panels

data:  GDPgrowth
z = 41.304, p-value < 2.2e-16
alternative hypothesis: cross-sectional dependence
```

```
> CCEP_forcointegration<- pcce(lnGDP ~ D, data = panel_dataFra
me, model = "p")> summary(CCEP_forcointegration)

Common Correlated Effects Pooled model

Call:
pcce(formula = lnGDP ~ D, data = panel_dataFrame, model = "p")

Balanced Panel: n = 86, T = 51, N = 4386

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-94.8164036  -4.6671682   0.0574923   4.7848192  57.3585377

Coefficients:
  Estimate Std. Error z-value Pr(>|z|)
D   1.9877    38.3573  0.0518   0.9587
Total Sum of Squares: 117260000
Residual Sum of Squares: 411960
HPY R-squared: 0.9271
```

```
> paneldata_residuals_cce <- pdata.frame(residuals_df_cce, ind
ex = c ("individual_id","time_id"))

> purtest(paneldata_residuals_cce$residuals, test = "madwu")

               Maddala-Wu Unit-Root Test (ex. var.: None)

data:  paneldata_residuals_cce$residuals
chisq = 725.72, df = 172, p-value < 2.2e-16
alternative hypothesis: stationarity
```

What do the results of your cross-sectional dependenc e test imply for the properties of the considered panel unit root tests ?

From the cross-sectional dependence test, it reject the H0 of there is no cross-sectional dependence. And this will lead all the first generation unit root test(LLC test, IPS test and Maddla Wu test have a nontrivial size distortion. There will always be type I error for the significance level has been distorted.

What do the results of your unit root tests imply for the properties of the estimates of equations (1)-(5) ?

From the unit root test, I reject that all the △Di and △lnGDPi contain at least one unit root and I accept that all the lnGDPi and Di contain at least one unit root. Put differently, lnGDPit is I(1) and Dit is also I(1), Also there is cross-sectional dependence in lnGDPit and △lnGDPit.

For model (1), it may be spurious regression like in time series if there is no cointegration between lnGDPit and Dit bring the unbiased but inconsistent result. However,we can add common correlated factor to absorb the cross-sectional dependence, then we will get consistent result.
For model (2), because △lnGDPit is 'stationary' by test (in fact they may not hold because there might be strucutral break) and Dit is non-stationary so the error term will also be non-stationary. The model would be inconsistent it may also be the spurious regression.
For model (3), it is consistent if there is cointegration relationship, if there is no cointegration, I can also add common correlated fator can absorb the cross-sectional dependence which will give me consistent result. Or I can add the lag term of Dit then my model will be consistent because the real DGP will be nested in my model so in general my model will be right.
For model (4), it is not consistent because △lnGDPit is I(0) and Dit is I(1), and this will bring I(1) in error term. Also same problem like model (2), we can not identify the structural break by unit root test.
For model (5), it is consistent, because all the variable is stationary and local projection is robust to misspecification if there is cointegration relationships.